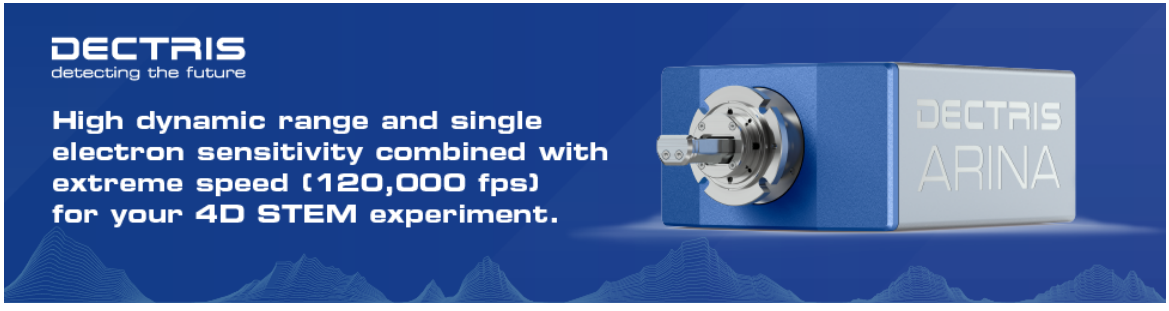
Large-Scale Atom Probe Tomography Data Mining: Methods and Application to Inform Hydrogen Behavior

Martin S Meier, Paul A J Bagot, Michael P Moody, Daniel Haley



Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

Microscopy and Microanalysis, 2023, **29**, 879–889
https://doi.org/10.1093/micmic/ozad027
Advance access publication 15 March 2023
Original Article

Microscopy AND Microanalysis

Large-Scale Atom Probe Tomography Data Mining: Methods and Application to Inform Hydrogen Behavior

Martin S. Meier*, Paul A. J. Bagot, Michael P. Moody, and Daniel Haley

Department of Materials, University of Oxford, Parks Rd, Oxford, Oxfordshire OX1 3PH, UK

***Corresponding author:** Martin S. Meier, E-mail: martin.meier@materials.ox.ac.uk

Abstract

A large number of atom probe tomography (APT) datasets from past experiments were collected into a database to conduct statistical analyses. An effective way of handling the data is shown, and a study on hydrogen is conducted to illustrate the usefulness of this approach. We propose to handle a large collection of APT spectra as a point cloud and use a city block distance–based metric to measure dissimilarity between spectra. This enables quick and automated searching for spectra by similarity. Since spectra from APT experiments on similar materials are similar, the point cloud of spectra contains clusters. Analysis of these clusters of spectra in this point cloud allows us to infer the sample materials. The behavior of contaminant hydrogen is analyzed and correlated with voltage, electric field, and sample base material. Across several materials, the H_2^+/H^+ ratio is found to decrease with increasing field, likely an indication of postionization of H_2^+ ions. The absolute amounts of H_2^+ and H^+ are found to frequently increase throughout APT experiments.

Key words: atom probe tomography, cluster, data analysis, data mining, database, hydrogen, machine learning, mass-to-charge spectrum, t-SNE, zirconium

Introduction

The field of atom probe tomography (APT) has experienced significant expansion in the past 30 years. Modern instruments can routinely produce multiple datasets, incorporating more than 10 M (million) detected ions, every day. This has over time led to the amassing of many APT datasets. With more than 120 atom probes worldwide currently in use, we assume that the number of APT datasets collected to date well exceeds 1 M. We believe that a vast amount of yet undiscovered knowledge exists within these data archives and explore their potential use.

Experiments in current APT research are set up in a targeted way to specifically investigate particular materials or conditions. In order to gain new knowledge about materials, more experiments need to be conducted. However, the availability of large collections of experimental data enables an alternative avenue of research, looking for statistical observations across many prior experiments. This type of analysis is common and established in many other areas of science, for example, drug discovery and high-energy particle physics (Radovic et al., 2018; Zhu, 2020).

Similar studies have not yet been published in the field of APT. This is due to the use of closed file formats for storing raw experimental data (which is forced through restrictions in the control software on the commonly used commercial APT instruments), making them inaccessible to custom software and analysis scripts. While extensive efforts to develop open APT software exist (Kühbach et al., 2021), they cannot escape from the issue that most raw APT data that has been produced to date is only available in proprietary formats. We have developed a simple method, based on user interface (UI) automation, to circumnavigate this problem and automatically collect the data from past experiments in our own research group to firstly build a database. We then conducted various analyses on this new database as a proof of concept, as presented in this paper.

For easy handling of a collection of APT mass spectra, we propose a city block distance–based metric (Craw, 2010) which corresponds to the similarity of two spectra. We can then regard our collection of spectra as a point cloud, enabling an effective search of spectra by similarity. Similar spectra are grouped in this space such that a cluster search can uncover groups of experiments that result in similar mass spectra. More advanced analyses, such as on contaminant hydrogen, can be conducted based on this structure.

Contaminant hydrogen in APT has long been focus of different strands of research (Sundell et al., 2013; Kolli, 2017; Mouton et al., 2019; Jones et al., 2021; Meier et al., 2021; Rigutti et al., 2021). It appears in virtually every APT experiment; however, the source of and contamination mechanism remain controversial despite increasing levels of activity in this area (Felfer et al., 2021; Yoo et al., 2022). Due to this ubiquitous detection of hydrogen, nearly every APT dataset also can be regarded as an experiment investigating the factors affecting the measurement of contaminant hydrogen. Our collection of automatically exported datasets therefore contains a vast amount of data about contaminant hydrogen.

The clearest evidence of contaminant hydrogen in APT mass-to-charge-state ratio (m/n) spectra are the peaks at 1 and 2 Da, corresponding to H^+ and H_2^+, respectively. Due to their low mass, these peaks cannot possibly overlap with any ions that we would realistically expect in APT spectra. This means we can automatically extract the counts in these peaks for a large number of experiments and conduct analyses on these. In this paper, we aim to analyze the field dependence

Received: November 4, 2022. **Revised:** January 17, 2023. **Accepted:** February 20, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of the Microscopy Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

of hydrogen in the contaminant peaks at $m/n = 1,2$ Da across past experiments in our collection.

Background and Theory

Data

The primary data of interest collected in APT experiments are the detector coordinates and time-of-flight information. Data processing is then required in order to convert this data into mass-to-charge spectra and 3D real space atom-by-atom reconstructed datasets. The detector data is complemented with additionally monitored parameters during the experiment, e.g., pressure or sample temperature. On Cameca LEAP instruments, this data is stored in a proprietary file format, and their reconstruction software (IVAS or APSuite) is needed for extraction of spectra and reconstruction data from these files. To date, there exists no direct and fully automatable method for analyzing or exporting batches of many datasets.

To overcome this problem, a Python script for UI automation was devised that can automatically open raw APT experiment files in IVAS (.rhit, .hits) and export the available experimental data into .epos and.csv files. These file formats are open and suitable for use with a wide range of free and existing tools and also allow conducting large-scale data analysis outside using our own tools.

A full spatial reconstruction of APT data, at present, requires a significant amount of information to be manually input by the user, e.g., for assigning ion identities in the mass–charge spectrum and selection of reconstruction parameters. Development of capabilities to automate this is beyond our scope of research. Our UI automation script skips these steps or falls back to default values. This means that the reconstruction is often not constrained with sensible reconstruction parameters, rendering the spatial reconstruction coordinates that we export less meaningful.

City Block Distance of APT Spectra

A crucial capability for successful research on a large library of APT spectra is the ability to search for spectra with certain features or similarities or to discover groups (or clusters) in this collection. We propose to use a method based on the city block distance (Craw, 2010) to achieve this.

We create APT spectra for all experiments in the collection with the same histogram binning and normalize them such that the counts (or intensities) in each spectrum sum up to a fixed value of 1. The collection of spectra obtained is regarded as a high dimensional point cloud. Each dataset corresponds to one data point, and each bin corresponds to one dimension in the point cloud. Thus, m spectra with n bins each amount to a m*n point cloud. We then use the inverse city block distance between two spectra in the point cloud as a measure for their similarity, enabling the retrieval of spectra by similarity through a nearest-neighbor search in the spectral point cloud.

The city block distance between two spectra a and b with n bins is

$$d(a, b) = \sum_{k=1}^{n} abs(a_k - b_k) \quad (1)$$

where a_k and b_k are the measured intensities of the k-th bin in spectrum a and b. d is scalar.

The intensities in APT spectra are measured as a number of detector events. However, we normalize the spectra a_k and b_k such that the intensities in all bins sum up to one for distance calculations. This removes the influence of the total number of counts (ion yield of an experiment) and machine detection efficiency on the calculated distance between any two spectra.

Figure 1 illustrates an example. Figures 1a and 1b each show two normalized example spectra. In the case of Figure 1a, these spectra are similar, while they are dissimilar in Figure 1b. The city block distance between these spectra corresponds to the summed-up differences between intensities per bin, shown as black arrows. For the case of dissimilar spectra, this sum is higher, resulting in a greater city block distance between spectra.

The spectra in our collection are not represented by a uniform distribution of points in the point cloud. Instead, there are groups of experiments that have been run on the same or similar base materials, resulting in groups of similar spectra. In a point cloud where the inverse distance corresponds to similarity, these groups appear as clusters. Cluster search in this point cloud can therefore identify the base materials of the samples in our point cloud, as will be shown in Searching and Clustering in the Data. It can also be used to estimate the base material of a sample, based on the cluster in which the spectrum is found.

Analysis of Hydrogen Behavior

Hydrogen is a common contaminant in APT spectra, appearing in peaks at $m/n = 1$ (H^+), $m/n = 2$ (H_2^+), $m/n = 3$ (H_3^+) and

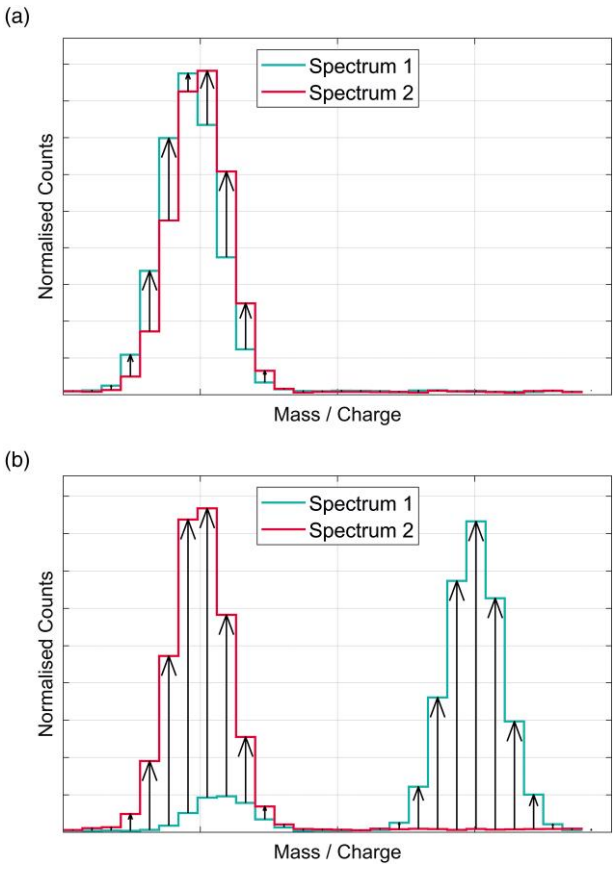


Fig. 1. Examples for two similar (**a**) or dissimilar (**b**) APT spectra (simplistic drawing with one or two peaks only per spectrum). The city block distance is the sum of all black arrows. It is seen that the dissimilarity of the spectra in (**b**) directly leads to a higher distance than in (**a**).

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

Table 1. Reconstruction Steps in IVAS and Approach of our UI Automation Script.

Step in reconstruction wizard	Purpose	UI automation script action
1. Setup	Review experiment data	–
2. Ion sequence range selection	Select part of the raw experiment file to include in reconstruction, select initial voltage for reconstruction, and export graphs: voltage, pressure, laser energy, etc.	Select everything, and export all graphs as .csv format
3. Select detector region of interest (ROI)	Select detector area to include in reconstruction	Leave at preselected default
4. Time-of-flight (TOF) correction	Apply voltage and bowl correction to TOF spectrum	Run with default values
5. Mass calibration	Apply (piecewise) linear correction to *m/n* spectrum to match position of clearly identified peaks with their correct locations	Skip (this is corrected during postprocessing of the data, outside of IVAS)
6. Ranged-ion assignments	Assign peaks in the mass-to-charge-state ratio spectrum to ions for reconstruction	Load a dummy range file containing only one range for H^+ at the m/n^{-1} peak
7. Reconstruction	Adjust reconstruction parameters	Run with default values (voltage-based reconstruction)

more complex hydride ions. The behavior of this contaminant has attracted significant research interest; however, it is still not fully understood, and relevant supply and evaporation mechanisms remain controversial (Tsong & Liou, 1985; Mouton et al., 2019; Felfer et al., 2021; Jones et al., 2021; Rigutti et al., 2021; Yoo et al., 2022). One of the reasons for this is the highly material dependent behavior of contaminant hydrogen in APT. Insights obtained using a particular type of experiment and/or sample may not be fully transferrable to other settings.

Properties of contaminant hydrogen that have been previously investigated are the ratio of the H^+ and H_2^+ peaks, their absolute ionic fractions within the dataset, and their evolution over voltage or electric field strength during the experiment (Sundell et al., 2013; Haley et al., 2014; Jones et al., 2021). With a large collection of APT experiments available, it is possible to study H behavior across many datasets and search for parameters that influence it. A basic approach to conduct this analysis is to automatically extract the height of the H^+ and H_2^+ peaks (at 1 and 2 Da) and to calculate their ratio as a function of the voltage applied during the course of the experiment.

It is however important to keep in mind that the voltage does not directly influence the evaporation at the tip of the APT sample. Instead, the relevant underlying parameter is the strength of the electric field at the apex of the specimen. The relationship between actual voltage and field at the tip surface is dependent on many other factors, including sample geometry, which is difficult to estimate. However, for a range of materials, the approximate field can be estimated from the charge state ratios (CSR) of ions in the spectrum, using look-up from Kingham curves (Kingham, 1982). This approach has some limitations, e.g., reduced accuracy in materials having strong molecular ion evaporation (Schreiber et al., 2014; Gault et al., 2016); however, it is suitable for most metallic materials.

Once the base material of a sample is known (and this can be determined using a cluster analysis, as shown in Searching and Clustering Data in the Data), the CSR of the majority ions can usually be extracted easily and automatically as the ratio of two peaks whose position in the spectrum is known. This means that an analysis of hydrogen over voltage in our collection of APT spectra can be easily extended to an analysis of hydrogen over electric field.

Experimental

Data

We created a UI automation script that automatically operates the commercial APT software IVAS to export data from the proprietary formats into open formats that can be used for our research. To this end, the script needs to load the raw experiment files into IVAS and process it using a "Reconstruction Wizard". This function leads the user step-by-step through reconstruction of the APT data. Our script does not aim to accurately reconstruct a spatial image from the data, but only at exporting it, and hence skips as many of the steps in this process as possible or uses default values. A list of parameters that our script feeds into the reconstruction wizard is given in Table 1.

After completing the reconstruction wizard, the UI automation script saves the APT dataset as an .epos file. This file along with the .csv files of additional sensor data that were exported in step 2 (Table 1) is copied, and the reconstruction is restarted to export the next raw experiment file. An overview showing all of the data that is retrieved per raw experiment file is provided in Supplementary Material S1.

This method has been applied to APT experiment data files (.rhit files) that have been obtained between 2012 and 2015 on a Cameca LEAP 3000X HR atom probe and with a file size above 100 MB. Unlike more recent generations of LEAP instruments, this atom probe does not support changes of pulse frequency and laser energy during experiments, meaning that experiments are at constant pulse rate and laser energy (if laser evaporation is used). The selection by file size leads to the selection of datasets with more than approximately 6.8 M detector events. The reason for this minimum file size is that in short APT experiments, the influence of spurious effects occurring in the beginning of the experiment (e.g., due to evaporation of surface contaminants from the sample, initially poor laser targeting, slow voltage ramp up to the onset of evaporation) on the results may be outsized. It is noted however that this also biases the selection of experiments toward experiments with parameters that typically yield long runs, i.e., laser-assisted evaporation modes and high-yielding materials.

A crucial step skipped during the reconstruction in IVAS is the calibration of the mass–charge spectra. This step typically involves a (piecewise) linear scaling of the *m/n* values of the

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

detected ions, such that prominent and easily identifiable peaks in the spectra align with their expected positions. In our database, we apply a semi-automated mass calibration on the exported data. First, datasets which contain clear isotopic fingerprints of one or more out of Ti^{2+}, U^{3+}, Si^{+}, Mo^{2+}, Cr^{2+}, Zr^{2+}, and Ga^{+} are selected through manual inspection. These ions were chosen as they are commonly present in the experimental data collection and their fingerprints have characteristic shapes that can be easily visually identified by an operator. In addition, we attempt to automatically identify the location of the $m/n = 1$ peak in our datasets. The $m/n = 1$ peak is present in almost every APT experiment due to contaminant H^{+} ions, and as such, this peak was located successfully in 843 out of 848 datasets.

For all datasets where the H^{+} peak and at least one other isotopic fingerprint was identified (by manual inspection), linear correction of the m/n spectrum can be applied directly and automatically by shifting and scaling the spectra such that peak maxima of identified fingerprints align with their expected positions. For all datasets where H^{+}, but no other fingerprint was identified, we apply a shift such that the $m/n = 1$ peak is at its expected position and subsequently determine the scaling factor through optimization.

To this end, we assume that for all of these datasets, there will be at least one dataset amongst the previously corrected ones that at least in parts looks similar, i.e., contains shared ionic fingerprints. This is a reasonable assumption as our spectra are typically dominated by ions from one out of a low number of base materials (as will be shown in Searching and Clustering in the Data, Fig. 3), for which reason at least one dataset with the same main ion should be present for every dataset. For all of the spectra where the scaling is to be determined, we iterate through a range of scaling factors and select the scaling factor where the distance to any of the (corrected) spectra with one or more known isotopic fingerprints is minimal.

No such optimization was conducted for those five datasets where no H^{+} peak was found.

Searching and Clustering in the Data

The distance between respective spectra [Equation (1)] can be used as an effective mean for searching spectra or parts of spectra that are similar to a given input spectrum. To find the spectra in the database that are most similar to a given candidate, the spectra with lowest distance are computed. We demonstrate this capability on three example spectra from outside the database, where we search the database to see if it contains similar-looking spectra. The run numbers and hash values of these search candidates are provided in Supplementary Material S5.

The first example is a Ti^{2+} isotopic fingerprint that overlaps with Cr^{2+} obtained from a nickel-based superalloy that was analyzed at the APT and 3D Nanoanalytics group (P.J. Felfer) in Erlangen, Germany, on a Cameca LEAP 4000X HR. This search candidate combines two favorable aspects that are likely beneficial for the search for similar neighbors. Firstly, Ti and Cr are common alloying elements in Ni-based superalloys and steels, of which many samples have been previously analyzed at the Oxford APT facility, leading to a high number of potentially well-matched candidates. Secondly, Ti^{2+} and Cr^{2+} have been used in the calibration of those m/n spectra where they have been identified, leading to a particularly good calibration of the spectra at the locations of Ti^{2+} and Cr^{2+}.

The second example is a lift-out of a pure tungsten sample that has been run on a Cameca LEAP 5000 XS. Like for the Ti–Cr case, there are many datasets in our collection of spectra that should be a good match. However, the search candidate was intentionally chosen as a spectrum with relatively few counts and a poor peak shape (due to deliberately poor application of the voltage/bowl correction in the IVAS software). This represents an intermediate case, meaning that we expect a nearest-neighbor search to be more prone to pick up datasets that are a poor match to the search candidate.

The third example is a complex peak overlap of SiO_2^{+}, Na_4P^{2+}, PO_2^{+}, and PO_2H^{+} ions in a bioactive glass sample that was analyzed by Ren et al. on a Cameca LEAP 5000 XR (Ren et al., 2021). There are no similar samples in our collection of datasets. This represents a hard, or impossible, case, where we expect our approach will retrieve only "false" leads from a nearest-neighbor search, i.e., spectra that may look similar, but which are chemically different.

As mentioned in City Block Distance of APT Spectra, it is expected that the point cloud of all spectra in the database contains clusters. Due to the high dimensionality of the point cloud, direct visual inspection of these is impossible. We therefore apply t-distributed stochastic neighbor embedding (t-SNE) to embed the high-dimensional point cloud into two-dimensional space while preserving the spatial relationships as much a possible (van der Maaten & Hinton, 2008). t-distributed stochastic neighbor embedding works by calculating a probability distribution based on the point distances in high-dimensional space and then tries to find a distribution of points in low-dimensional space with similar distribution by minimizing the difference between the probability distribution in high-dimensional and low-dimensional space. This enables an approximate overview over the similarity (distance) relations between our spectra. In this work, we use a spectrum bin size of 0.1 Da for the t-SNE embedding to reduce computation time and memory usage. This is lower than the bin sizes commonly used in APT spectra (usually in the range 0.001–0.01 Da), but is sufficient for our application.

To provide additional information about sample chemistry, we fit isotopic fingerprints of common sample base materials from our group (Fe, Ni, Zr, W, Si, Ti, Ga, U) into our spectra, taking charge states +1 and +2 into account. If one of these fingerprints accounts for more than a certain fraction of all detector events in the spectra (critical fractions listed in Fig. 3a, legend), we draw the points in a color specific to that element. This effectively highlights the main elements of the samples in the diagram, where they have been identified. For clarification, no manual identification of peaks was conducted in this step, unlike for the identification of peaks for spectra calibration.

It is noted that visual inspection of t-SNE embeddings is not an ideal method for proving the existence of clusters. This is due to human bias, but also t-SNE possibly introduces artefacts, such as clusters that are not contained in the original data (Wattenberg et al., 2016). To confirm that our observations are not artefacts, we apply the "Ordering Points To Identify the Clustering Structure" algorithm (OPTICS) to our point cloud of spectra (not to the t-SNE coordinates) in order to create a reachability diagram. This algorithm travels through all data points (spectra) in the point cloud by visiting the nearest, not yet visited, neighbor of the current data point (Ankerst et al., 1999; Daszykowski et al., 2002). For each

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

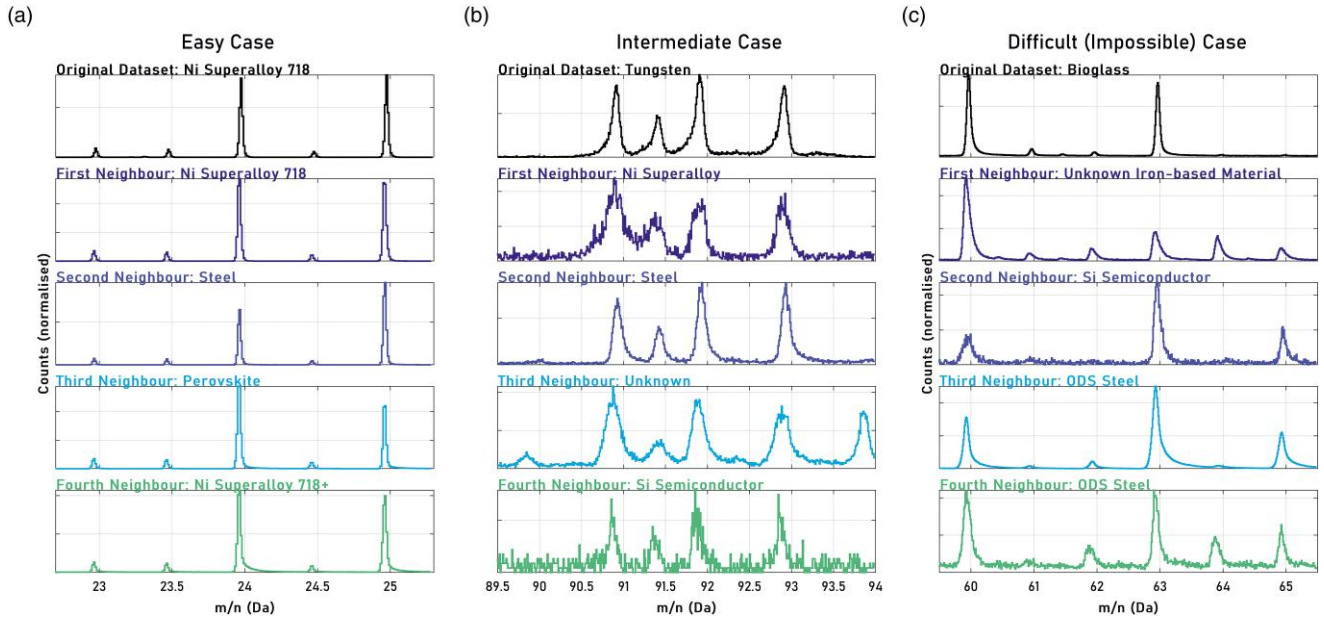


Fig. 2. Three sections from APT spectra that are not contained in our database (top lines, black) and their four nearest neighbors in the Oxford APT database, corresponding to an easy (a), intermediate (b) or difficult (c) case of retrieving spectra by similarity. It is seen that in all three cases, neighbors with similar peak structures are found.

point, the so-called reachability distance is calculated as the maximum out of the distance that needs to be travelled to this point and the radius around this point within which a certain minimum number of Points ("MinPoints") is found. The reachability plot contains the reachability distances for all points, in the order in which they are visited. Spectra in clusters have low-reachability distances, but they are surrounded by distanced points that are not part of the cluster and thus appear as sinks in the reachability plot.

Analysis of Hydrogen Behavior

We analyze the ratio of H_2^+ to H^+ as function of voltage in our collection of spectra. Typically, APT experiments are run using a constant detection rate (ions per applied pulse). The voltage during the experiment is adjusted to achieve a certain target detection rate. This usually leads to the voltage slowly increasing during the experiment as the specimen evolved throughout.

We separate each dataset into voltage bins of 500 V width. For each voltage bin, we fit a constant TOF background to the areas around the hydrogen peaks, where no other ions are expected (Haley et al., 2020). We then extract the amount of H_2^+ and H^+ detector events as the background-corrected counts at $m/n = 0.9 \ldots 1.1$ and $m/n = 1.9 \ldots 2.1$. Bins with less than 50,000 total atoms or less than 500 H_2^+ or H^+ ions are rejected. A side effect of this is the rejection of all datasets where the H_2^+ or H^+ peaks are missing. From this, the evolution of the $H_2^+ H^+$ ratio over voltage can be calculated for each dataset. Since it is possible to infer the base material of samples through association with clusters in our spectra's point cloud (Searching and Clustering in the Data), this enables drawing diagrams of the fraction of H_2^+ and H^+ and the ratio H_2^+/H^+ as function of voltage per material for many experiments.

As will be shown in Searching and Clustering in the Data, some base materials of the samples in our collection can be easily identified, including Fe, Si, Ni, Zr, and W. For these materials, we extract the CSRs of Fe^{2+}/Fe^{+}, Si^{2+}/Si^{+}, Ni^{2+}/Ni^{+}, Zr^{3+}/Zr^{2+}, and W^{3+}/W^{2+} for all voltage bins where the counts of H_2^+ and H^+ were estimated from the counts in the (noise-corrected) spectra. It is noted that this will be inaccurate if the peaks that are used for measuring the CSR overlap with other ions. Given that we will have determined the base material with confidence and that our conclusions will be drawn from averaged data across many experiments, we believe that the errors that this introduces will not critically influence our results. We then use Kingham curves [as calculated by Yao et al. (Gault, 2012)] to estimate electric fields from the CSR. This enables us to analyze the evolution of H^+ and H_2^+ and their ratio H_2^+/H^+ as a function of voltage and electric field in these materials.

Results

Searching and Clustering

For the three example spectra that we use to demonstrate nearest-neighbor search, we show histograms of the spectra as well as of the nearest four neighbors in Figure 2. In the "easy" case of the Ti–Cr overlap (Fig. 2a) it is seen that all of the nearest neighbors are very similar to the search candidates. In fact, all of these are Ti–Cr overlaps which look just like the search candidate, and the closest neighbor is even a sample from exactly the same alloy (Inconel 718) that was analyzed as part of a project in Oxford in 2012 that was unrelated to the search candidate, which was run in 2016 at the APT group of Erlangen (Pröbstle et al., 2016).

For the intermediate case of a W spectrum with few counts and a less well-defined peak shape, all of the four nearest neighbors (Fig. 2b) are tungsten fingerprints too. However, none of these correspond to a sample that is pure tungsten or one of its alloys. All of these samples only contain minor amounts of W, so have small peaks slightly above the noise floor. In the very difficult (or "impossible') case of the bioglass samples (Fig. 2c), all of the four nearest neighbors are spectra showing Ni and Cu fingerprints from APT runs on materials that have no similarity to bioglass (Yeli et al., 2017).

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

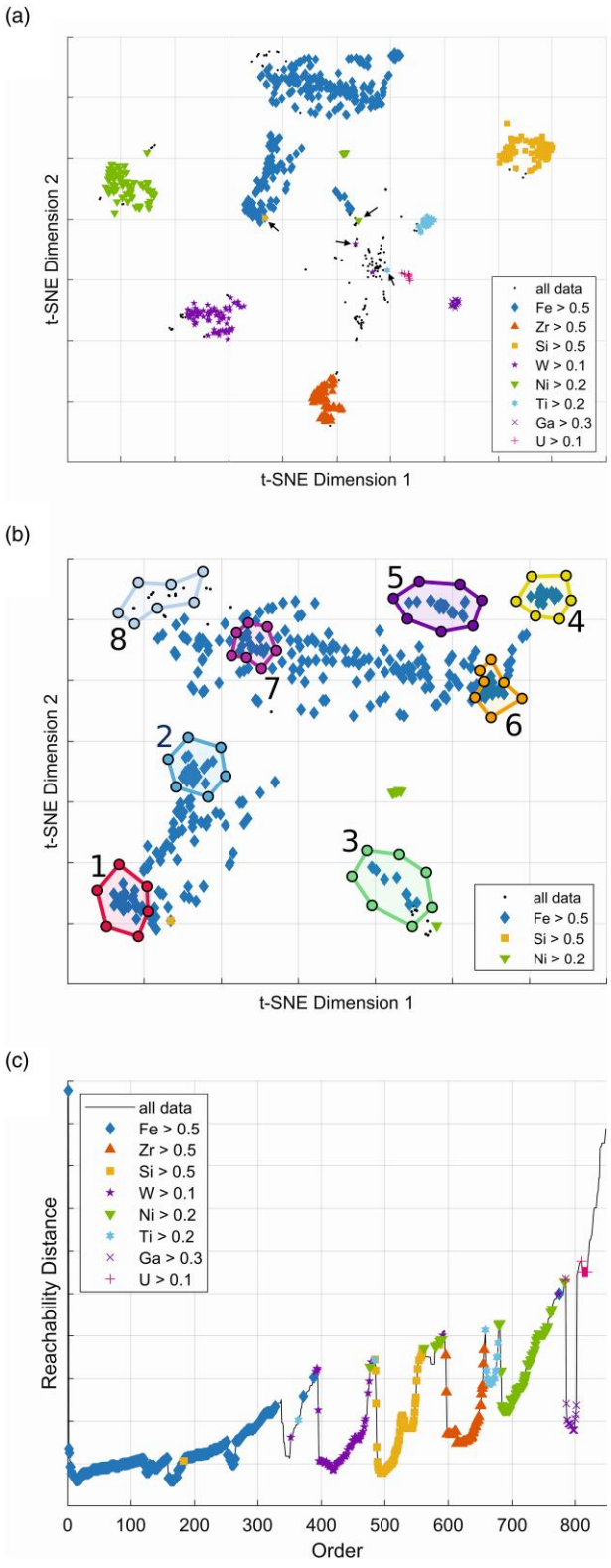


Fig. 3. (**a**) t-SNE of the spectra point cloud with black arrows highlighting outliers, (**b**) zoomed-in iron cluster in the t-SNE of the cloud with the datasets plotted in Figure 4 highlighted, (**c**) OPTICS reachability plot of the point cloud. It is seen that clusters are present in the APT spectra point cloud and that this segregation strongly occurs along the sample base material.

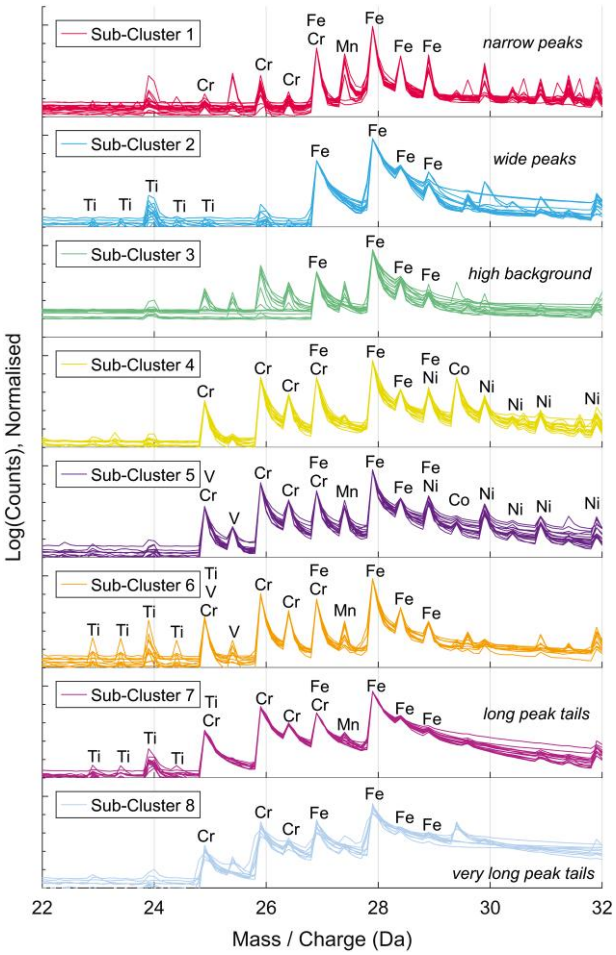


Fig. 4. Sections around the iron isotopic fingerprints in the APT spectra in the eight ROIs highlighted in Figure 3b. It is seen that clear iron fingerprints are contained in all spectra in all ROIs; however all of the ROIs also contain features that set them apart from other ROIs. The bin size (0.1 Da) is relatively wide, as chosen for t-SNE.

A t-SNE embedding of all spectra in our collection of datasets is shown in Figure 3a. Datasets where common isotopic fingerprints (Fe, Zr, Si, W, Ni, Ti, Ga, U) were found to fit well have been highlighted in different colors; these are datasets where these elements were found to account for more than a certain fraction of all counts (thresholds in Fig. 3a, legend). There are distinct clusters in our point cloud, and the clustering occurs mainly in line with the chemical composition of the samples. The cluster of datasets where none of the tested fingerprint fits well (in the center of the map) was found to include mostly spectra with very a high noise floor, i.e., unsuccessful experiments where the noise floor dominates the spectrum.

From visual inspection, the cluster of iron spectra in Figure 3a contains several sub-clusters. This structure inherent to the cluster seems to be due to the peak shape and the alloying elements, such as Mn, Cr, and Ni. This is demonstrated in Figure 3b and Figure 4. In Figure 3b, only the iron cluster is plotted, and eight manually chosen regions of interest (ROI) are highlighted in numbered polyhedrons. Figure 4 shows the spectra around the Fe^{2+} isotopic fingerprints from these eight regions of interest. It is seen that while the spectra in all ROIs contain an iron fingerprint, all of them also have individual features that set them apart from spectra in other ROIs

It is important to note that the t-SNE coordinates are due to the full spectra and not just the regions around the Fe^{2+} ion as

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

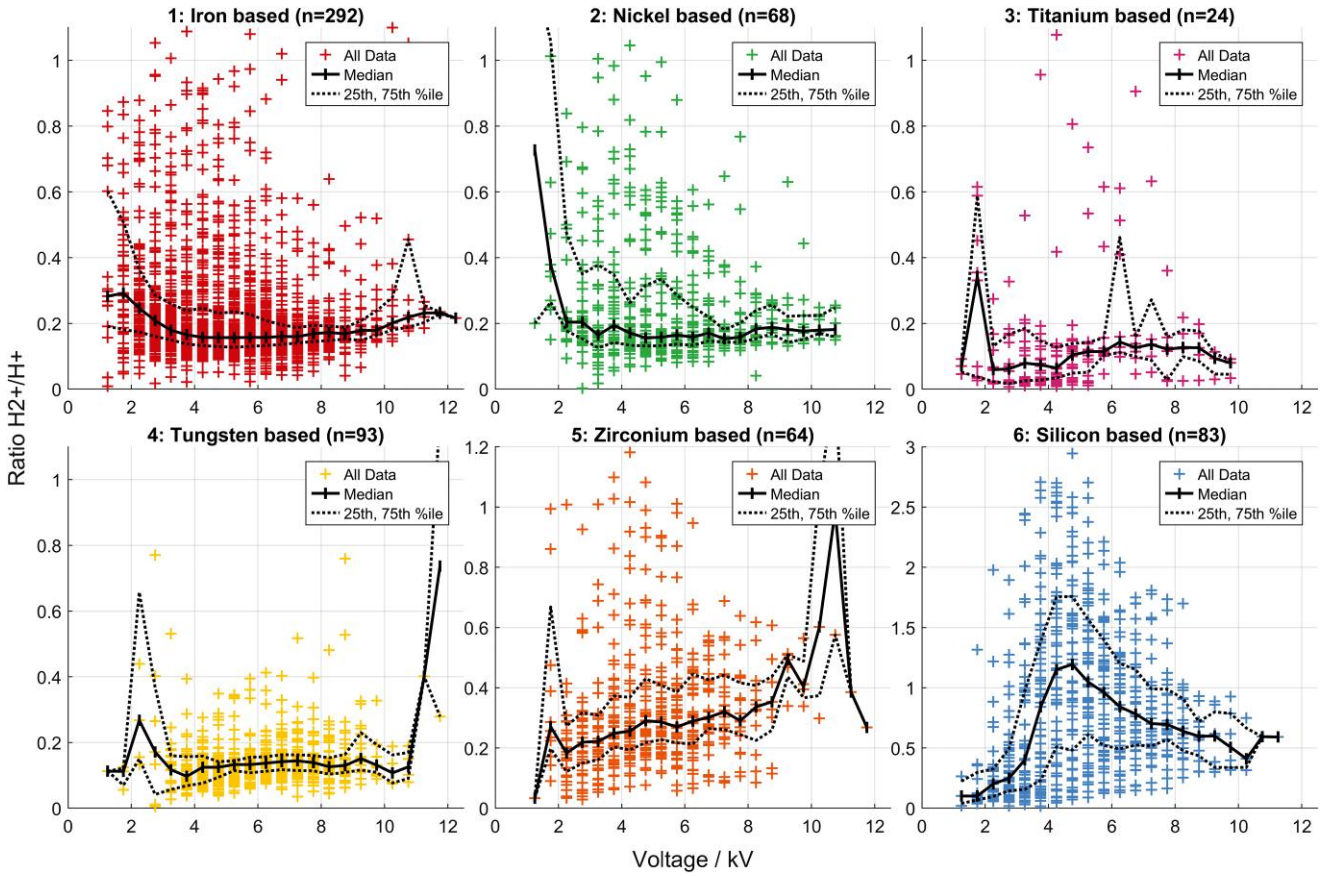


Fig. 5. Evolution of the H_2^+/H^+ ratio as functionof voltage in six base material–related clusters in our collection of spectra. In Fe, Ni, and W, H_2^+/H^+ converges into a relatively small channel during a large number of experiments. In Ti and Zr, there may be different trends. In Si, an interesting maximum at approximately 4.5 kV is observed. Overall, a material dependence of the behavior of contaminant hydrogen becomes evident.

drawn in Figure 4. However, the region around the Fe^{2+} peaks contains most of the largest peaks in APT experiments on typical steels (Fe^{2+}, Cr^{2+}, Mn^{2+}, Ni^{2+}). For this reason, inspection of just this region provides a good clue to understanding the structure of the Fe cluster. The full spectra are shown in the Supplementary Material S2.

The OPTICS reachability diagram, using the same markers as the t-SNE diagram, is shown in Figure 3c. It is seen that there are clear sinks in the plot and that these sinks correspond to the different base materials. This confirms that the point cloud of spectra indeed contains separable clusters along the sample base material, as has been expected from the t-SNE diagram.

Hydrogen Behavior

Figure 5 shows the evolution of the H_2^+/H^+ as function of voltage for the most common materials in our collection, for laser mode experiments only. The curves of the fractions of H_2^+ and H^+ are supplied in Supplementary Materials S3.

It is seen that the H_2^+/H^+ in most experiments on iron- and nickel-based materials lie in a relatively narrow channel around 0.2. However, it is also seen that outliers can extend far away from this value, to below 0.1 and above 1. Similarly narrow channels are observed in titanium and tungsten, however, at lower ratios of approximately 0.1. It is noted that the number of titanium datasets (24) is comparably small. In zirconium, the H_2^+/H^+ ratio lies slightly higher at approximately 0.2–0.3 and increases with rising standing voltage throughout the experiment. There seems to be a number of Zr experiments where the H_2^+/H^+ ratio spikes at a certain point between 3 and 6 kV to clearly above the 75th percentile—an observation that is not made in any tungsten experiment.

A curious observation can be made for silicon. There is a spike in the H_2^+/H^+ ratio at approximately 4–5 kV, which is observed in a high number of experiments as well as in the average across all data points. Inspection of the curves for H_2^+ and H^+ reveals that this is due to an increase of H_2^+ at this voltage (Supplementary Material S3). Both H_2^+ and H^+ increase with voltage in silicon; however, at 4–5 kV, the increase in H_2^+ is much stronger, while the H^+ increases relatively more at higher voltages.

Curves of the absolute fractions of H_2^+ and H^+ throughout the experiments are provided in Supplementary Material S3. It is seen in these curves that both median H_2^+ and H^+ increase with increasing voltage in experiments on Fe, Ni, Si, and W. For W, the increase is however less pronounced, and the scatter at the lower and upper ends (<4 kV and >9 kV) of the voltage range is more pronounced due to poor statistics, presumably arising from the low number of experiments that were analyzed in such voltage ranges. For Zr, the trend is opposite, with H_2^+ and H^+ decreasing throughout the experiment. For Ti, it is possible that a decrease similar to Zr is observed; however, this is not clear due to the high amount of scatter (and poor statistics from only 24 experiments) for this material.

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

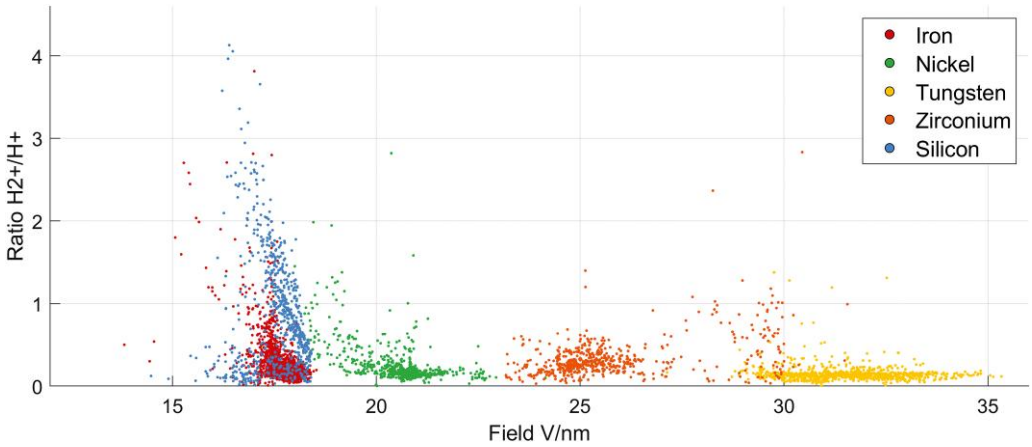


Fig. 6. Hydrogen ratio H_2^+/H^+ for common materials in our collection of spectra as function of field, calculated from CSR. An overarching, decreasing trend of the ratio with increasing field is found for Fe, Ni, Si, and W. This likely indicates that an influence of the electric field on contaminant hydrogen behavior exists across a wide range of material systems.

Figure 6 shows a scatter plot of the same experiments and voltage bins as used in Figure 5, but instead of the voltage, the electric field as calculated from CSRs is used. We were unable to calculate meaningful field estimates for the titanium datasets as most do not have Ti^+ peaks (needed for calculating Ti^{2+}/Ti^+ CSR) or they are overlapped; therefore, these are excluded from the diagram. It is seen that across the base materials iron, nickel, and silicon, a consistent trend of H_2^+/H^+ decreasing with increasing field strength down to approximately 0.1–0.3 is observed.

It is noted that the Si data points (Fig. 6) are bimodal, where there are also a number of data points where a relatively low field is not paired with a high H_2^+/H^+ ratio. Matching up data between Figure 5 and Figure 6 reveals that these Si data points of low fields and low H_2^+/H^+ ratio in Figure 6 correspond to the data points at the lower voltage range in Figures 5–6 (silicon), i.e., before the spike in H_2^+/H^+ that is seen in this diagram. This, along with the fact that such a clear maximum of the H_2^+/H^+ versus voltage curve (Figs. 5–6) is observed at all, could indicate the observation of a field- or voltage-dependent phenomenon in Si that is not made in the other materials.

The calculated fields for tungsten and zirconium are higher than for Si, Fe, and Ni. The ratio of H_2^+/H^+ in tungsten is always low (approximately 0.1), such that it continues the trend of low H_2^+/H^+ ratios at high fields that is observed in Fe, Ni, and Si. Like in tungsten, no decrease or increase of H_2^+/H^+ is seen in zirconium, but the absolute values and scatter are higher (approximately 0.2–0.3). It is noted that, unlike for Fe, Ni, Si, and W, Zr easily forms hydrides during APT sample preparation, which means that a different behavior of hydrogen in this material would not be surprising (also see discussion in Hydrogen Behavior).

Discussion

Searching APT Spectra by Similarity

The three example datasets in Figure 2 illustrate that nearest neighbor search based on city block distance of normalized spectra is well suited for retrieving similar spectra (or sections of spectra) from a database. In particular, the fact that the nearest neighbor of the Inconel 718 Ni superalloy (data from the APT group in Erlangen) in the Oxford database is another Inconel 718 specimen from an unrelated project demonstrates how well this method can match isotopic fingerprints under good conditions.

However, it is also clear that there are limits to this. In the case of the poor quality tungsten fingerprint (Fig. 2b), all of the neighbors are similar to the tungsten fingerprint, but none of the data are from tungsten-based materials. Instead, all of these datasets are from materials that happen to contain a low amount of tungsten as an alloying element, which appears in rather poorly defined peaks that do not stand as well above the noise floor as the peaks in the Inconel 718 case. This shows that nearest neighbor search can in principle retrieve similar datasets when neighbors of low-quality spectra are searched. However, random noise in the retrieved nearest neighbors seems to be a major marker for their similarity, rather than actual similarity of sample or composition. It is therefore believed that sufficient data quality (and counts) is needed for reliable results. Our search method may thus be most applicable to searching through fingerprints above a reasonable threshold number of atoms and with low levels of noise.

The impossible case of bioglass illustrates that (somewhat) similar fingerprints can be found in sets of APT spectra that do not contain any similar materials or fingerprints. Nearest neighbor searches cannot guarantee that the composition of the retrieved spectra is even remotely similar, and additional plausibility checks are needed to confirm that a retrieved spectrum actually originates from a similar experiment.

It is noted that computation time of nearest neighbor searches is of no concern in the current study. This is because computation of nearest neighbor distances on modern hardware is sufficiently fast, such that a brute force approach to nearest neighbor search (for a given candidate, distances to all datasets in the point cloud are calculated and the datasets with lowest distance are selected) concludes in less than 0.1 s for parameters as used in this study (848 datasets, 17,999 spectrum bins, Intel Xeon 4,114 Dual CPU). Potential methods for the speed-up of the search, such as dimensionality reduction (Broderick et al., 2013) or k-d trees (Bentley, 1975), were not required for this study.

Clustering in the APT Spectra Point Cloud

The t-SNE map (Fig. 3a) and the OPTICS reachability distances (Fig. 3c) confirm that the point cloud of APT spectra

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

contains clusters. It is seen that the association of datasets with clusters on the t-SNE map and the OPTICS reachability diagram agrees well with our estimation of sample material, based on the fitting of isotopic fingerprints. This demonstrates that sample base material is the main criterion for spectra similarity in our set of spectra or association with certain clusters. There are only very few cases of outliers where samples with a certain estimated base material are not located in their respective clusters. Four examples are highlighted in Figure 3a with arrows.

It is seen that most clusters do contain some datasets that were not identified as having the base material of the other datasets in the clusters (the black dots in Fe, Si, Ni, Ti, W, and Zr clusters in Fig. 3a). This shows that an estimation of sample base material based on similarity (i.e., association with certain clusters in the point cloud) could possibly be more accurate than a naive estimation based on fitting of fingerprints. In particular, a clustering approach does not rely on fingerprints and fit parameters provided by the operator. In this sense, experiment categorization based on clusters in the point cloud poses an unsupervised machine learning approach, while sample identification based on fingerprints and content thresholds is supervised.

As illustrated by the spectra in the different ROIs in the iron cluster (Fig. 3b), sample similarity is not only guided by base material, but also by peak shape, possible alloying elements, and, if sufficiently high, the noise floor.

Association of peaks with ions, known as ranging, is a fundamental challenge in APT. A high amount of effort from experienced operators is needed for accurate results. Methods for automatic ranging or giving a list of potential ions for given peaks have been developed; however, these are still not sufficiently reliable and accurate for ranging arbitrary spectra with little to no human input (Johnson et al., 2013; Vurpillot et al., 2019; Meisenkothen et al., 2020). Existing methods usually rely on recognizing and fitting isotopic fingerprints into spectra (a supervised approach) (Larson et al., 2013; London, 2019). Our work indicates that an alternative approach where a spectrum with unknown ranging is compared with spectra in an existing database of well-ranged datasets could provide superior potential ionic identities, provided that the database contains sufficient similar datasets. A large database of well-ranged APT spectra could allow for the training of machine learning models that range APT spectra more accurately than current approaches. Unfortunately, we do not currently have high-quality range files for the spectra in our collection, meaning that this exercise will be left for future research.

Hydrogen Behavior

Despite a significant amount of prior research investigating the source and supply route of the contaminant hydrogen observed in APT spectra, the proposed mechanisms still lack consensus (Sundell et al., 2013; Breen et al., 2020; Felfer et al., 2021; Chen et al., 2022; Yoo et al., 2022). Contaminant hydrogen may originate from residual gas in the APT vacuum, the tip or crimp tube/puck, and it may either arrive directly at the tip while the experiment is running or diffuse along the shaft onto the tip apex during the experiment. While our study cannot fully answer this problem, we believe that it can contribute some unique insights into the behavior of contaminant hydrogen.

Figure 5 shows that for Fe, Ni, Ti, and W, in a large number of past experiments, there seem to be relatively narrow ranges toward which the H_2^+/H^+ ratios converge with increasing voltage. However, it is also seen that there is a significant number of outlier data points in these plots. This suggests that there are material-dependent H_2^+/H^+ ratios that are likely to be observed under typical experimental conditions (or conditions that were commonly chosen in past experiments in our group), though this may only apply statistically and not strictly for every experiment.

Figure 6 confirms that for laser evaporation experiments on Fe, Ni, and Si, the H_2^+/H^+ ratio is field dependent. Since the probability of postionization is greater at higher fields, this could indicate that postionization of H_2^+ and subsequent breaking into two H^+ is the underlying cause for this.

The absolute amounts of H_2^+ and H^+ in Fe, Ni, Si, and W clearly increase during the experiments with increasing voltages (Supplementary Fig. S3). Assuming that in Fe, Si, Ni, and W, only a comparably low amount of contaminant hydrogen evaporates as complex molecules (e.g., FeH^+), this indicates that the supply of contaminant H commonly increases throughout experiments on such materials. From our data, it is not directly obvious which mechanism is underlying this increase in supply. It may be due to changes in field or surface area, which we shall briefly discuss, or other yet unknown factors.

In a study on multi-layered semiconductors, it has been found that the amount of detected contaminant hydrogen is negatively correlated with the electric field required to evaporate the respective layers, when several layers are compared. However, inside of the layers, the amount was positively correlated with the field strength (Rigutti et al., 2021). This led to the conclusion that the relationship between field and hydrogen is complex and possibly compounded with other factors, including sample geometry. We have tested the possibility of the supply of contaminant hydrogen being field dependent in our collection of experiments, by plotting the total (ionic) fraction of hydrogen as function of field, similar to Figure 6. The result is shown in Supplementary Material S4. In contrast to the H_2^+/H^+ ratio, no clear trend however emerges for any of the materials or even sub-sets of experiments, and the scatter in the data is high. In particular, even analysis of sub-sets of the iron cluster indicates that particular relationships between H_2^+ or H^+ fractions and field were not reproducible in past experiments within our research group. Our observations therefore confirm that the field is not the only relevant factor for the total amount of contaminant hydrogen, but we cannot show what the exact influence of the field is.

It might be assumed that higher voltages typically correspond to later stages of an experiment, which corresponds to blunter tips. Blunter tips have a higher surface area that is visible to the detector (a wider field of view), so if the supply of contaminant hydrogen per surface area is constant, this could explain the increasing amounts of hydrogen with increasing voltage. If this was to be confirmed, it might suggest arrival of contaminant hydrogen at the tip from the gas phase during the experiment is the dominant supply mechanism. More research is needed to investigate whether or what role the sample surface area plays for hydrogen contamination.

Zr and Ti are known to easily form hydrogen-rich phases during sample preparation and APT experimental procedures (Ding & Jones, 2011; Chang et al., 2018; Mouton et al., 2019). This means that the dominant supply route for

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

contaminant hydrogen into these materials may be different than for Fe, Ni, Si, and W. It is therefore not particularly surprising that the behavior of hydrogen in these materials does not follow the trends observed in Fe, Si, Ni, and W. Also, it is known that Zr and Ti have a high tendency to evaporate in complex hydride ions (e.g., ZrH^+, ZrH_2^+, TiH_2^+) (Chang et al., 2019; Mouton et al., 2019; Jones et al., 2021). Our analysis does not address this aspect of contaminant hydrogen.

It is also very important to keep in mind that the postionization model of Kingham's that was used to convert CSRs to electric field is only approximately accurate and there are known materials where it is inaccurate (Schreiber et al., 2014; Gault et al., 2016). In addition, the CSR across the detector-visible part of the specimen surface in APT is usually not constant, and our approach which is based on voltage bins simply averages across the entire surface (Chang et al., 2019; Rigutti et al., 2021). This means that the absolute values of the fields in Figure 6 should not be over-interpreted.

Conclusion

From our study, we have shown that collections of data from past APT experiments contain valuable and diverse information that can be used for many current challenges in materials science and APT. The main conclusion of our work is that APT experiments should be archived, in open file formats so that the data is accessible to a wide range of analysis software.

Regarding collections of APT spectra as point clouds, using city block distances to measure (dis-) similarity is a suitable way of handling collections of spectra. In particular, automated search of spectra by similarity becomes possible, and analyses on the neighborhood of spectra in the point cloud allow conclusions about their chemical composition and other characteristics. If further observations, such as the fractions of the contaminant H_2^+ and H^+ over voltage, are extracted and compounded with the information about the spectra, this enables search for unknown phenomena that only appear statistically and that cannot be easily observed in few targeted experiments.

Our work in this paper is based on only 848 datasets, and we barely consider any data other than the mass–charge spectra. Collections with more datasets and consideration of further data such as range files or desorption maps, or even access to fully unprocessed raw detector data, will enable considerably more advanced investigations than we show in this work. We encourage APT groups—especially those with large build-up collections of data from past experiments—to explore ways in which these archives could be used for new research.

Availability of Data and Materials

Code for analysis for APT datasets is available at https://github.com/MartinMega/APTmap. Code for UI automation of IVAS is available at https://github.com/MartinMega/AutoRec.

The archival APT data used in this study contains confidential material from various projects and cannot be made available. A list of the experiments used for this study along with hash values is available at the Oxford Research Archive (https://ora.ox.ac.uk/objects/uuid:ed413cc3-1520-42a1-8936-eed47ef3ad5f), as well as a set of 14 APT experiments on Cameca PSM Silicon samples for a simple test of the code (these datasets are only for the purpose of testing and have not been used for the research presented in this paper). Intermediate results which are produced by the code during the analysis of the datasets have been archived to ORA at https://ora.ox.ac.uk/objects/uuid:9eecec05-8004-4348-85bf-c7d913158f7e; however, these cannot be made available at the time due to the confidential material contained in these results.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1093/micmic/ozad027.

Acknowledgments

M.S.M. would like to thank Megan E. Jones (University of Oxford, UK), Christina Hofer (University of Oxford, UK), Yanru Ren (University of Oxford, UK), Peter J. Felfer (FAU Erlangen, Germany), Baptiste Gault (MPIE Duesseldorf, Germany), and Ty J. Prosa (Cameca Instruments, USA) for helpful discussions.

Financial Support

M.S.M. would like to thank CAMECA Instruments / AMETEK Inc. for funding. The authors acknowledge EPSRC funding under grant numbers EP/M022803/1, EP/D077664/1, and EP/L014742/1.

Conflict of Interest

The authors declare that they have no competing interest.

References

Ankerst M, Breunig MM, Kriegel H-P & Sander J (1999). OPTICS: Ordering points to identify the clustering structure. In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, pp. 49–60. Philadelphia, Pennsylvania, USA: ACM.

Bentley JL (1975). Multidimensional binary search trees used for associative searching. *Commun ACM* **18**(9), 509–517. https://doi.org/10.1145/361002.361007

Breen AJ, Stephenson LT, Sun B, Li Y, Kasian O, Raabe D, Herbig M & Gault B (2020). Solute hydrogen and deuterium observed at the near atomic scale in high-strength steel. *Acta Mater* **188**, 108–120. https://doi.org/10.1016/j.actamat.2020.02.004

Broderick SR, Bryden A, Suram SK & Rajan K (2013). Data mining for isotope discrimination in atom probe tomography. *Ultramicroscopy* **132**, 121–128. https://doi.org/10.1016/j.ultramic.2013.02.001

Chang YH, Breen AJ, Tarzimoghadam Z, Kurnsteiner P, Gardner H, Ackerman A, Radecka A, Bagot PAJ, Lu WJ, Li T, Jagle EA, Herbig M, Stephenson LT, Moody MP, Rugg D, Dye D, Ponge D, Raabe D & Gault B (2018). Characterizing solute hydrogen and hydrides in pure and alloyed titanium at the atomic scale. *Acta Mater* **150**, 273–280. https://doi.org/10.1016/j.actamat.2018.02.064

Chang YH, Mouton I, Stephenson L, Ashton M, Zhang GK, Szczpaniak A, Lu WJ, Ponge D, Raabe D & Gault B (2019). Quantification of solute deuterium in titanium deuteride by atom probe tomography with both laser pulsing and high-voltage pulsing: Influence of the surface electric field. *New J Phys* **21**(5), 053025. https://doi.org/10.1088/1367-2630/ab1c3b

Chen Y-S, Liu P-Y, Niu R, Devaraj A, Yen H-W, Marceau RKW & Cairney JM (2022). Atom probe tomography for the observation of hydrogen in materials: A review. *Microsc Microanal* **29**(1), 1–5. https://doi.org/10.1093/micmic/ozac005

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023

Craw S (2010). Manhattan distance. In *Encyclopedia of Machine Learning*, Sammut C & Webb GI (Eds.), pp. 639–639. Boston, MA: Springer US.

Daszykowski M, Walczak B & Massart DL (2002). Looking for natural patterns in analytical data. 2. Tracing local density with OPTICS. *J Chem Inf Comput Sci* **42**(3), 500–507. https://doi.org/10.1021/ci010384s

Ding R & Jones IP (2011). In situ hydride formation in titanium during focused ion milling. *J Electron Microsc (Tokyo)* **60**(1), 1–9. https://doi.org/10.1093/jmicro/dfq066

Felfer P, Ott B, Monajem M, Dalbauer V, Heller M, Josten J & Macaulay C (2021). An atom probe with ultra-low hydrogen background. *Microsc Microanal* **28**, 1–9.

Gault B (2012). *Atom probe microscopy*. New York: Springer.

Gault B, Saxey DW, Ashton MW, Sinnott SB, Chiaramonti AN, Moody MP & Schreiber DK (2016). Behavior of molecules and molecular ions near a field emitter. *New J Phys* **18**(3), 033031. https://doi.org/10.1088/1367-2630/18/3/033031

Haley D, London AJ & Moody MP (2020). Processing APT spectral backgrounds for improved quantification. *Microsc Microanal* **26**(5), 964–977. https://doi.org/10.1017/S1431927620024290

Haley D, Merzlikin SV, Choi P & Raabe D (2014). Atom probe tomography observation of hydrogen in high-Mn steel and silver charged via an electrolytic route. *Int J Hydrogen Energy* **39**(23), 12221–-12229. https://doi.org/10.1016/j.ijhydene.2014.05.169

Johnson LJS, Thuvander M, Stiller K, Odén M & Hultman L (2013). Blind deconvolution of time-of-flight mass spectra from atom probe tomography. *Ultramicroscopy* **132**, 60–64. https://doi.org/10.1016/j.ultramic.2013.03.015

Jones ME, London AJ, Breen AJ, Styman PD, Sikotra S, Moody MP & Haley D (2021). Improving the quantification of deuterium in zirconium alloy atom probe tomography data using existing analysis methods. *Microsc Microanal* **28**(4), 1–10. https://doi.org/10.1017/s1431927621012848

Kingham DR (1982). The post-ionization of field evaporated ions: A theoretical explanation of multiple charge states. *Surf Sci* **116**(2), 273–301. https://doi.org/10.1016/0039-6028(82)90434-4

Kolli RP (2017). Controlling residual hydrogen gas in mass spectra during pulsed laser atom probe tomography. *Adv Struct Chem Imaging* **3**(1), 10. https://doi.org/10.1186/s40679-017-0043-4

Kühbach M, London AJ, Wang J, Schreiber DK, Mendez Martin F, Ghamarian I, Bilal H & Ceguerra AV (2021). Community-driven methods for open and reproducible software tools for analyzing datasets from atom probe microscopy. *Microsc Microanal* **28**, 1–16. https://doi.org/10.1017/S1431927621012241

Larson DJ, Prosa TJ, Ulfig RM, Geiser BP, Kelly TF & Humphreys PSCJ (2013). *Local electrode atom probe tomography: A user's guide*. New York: Springer.

London AJ (2019). Quantifying uncertainty from mass-peak overlaps in atom probe microscopy. *Microsc Microanal* **25**(2), 378–388. https://doi.org/10.1017/S1431927618016276

Meier MS, Jones ME, Felfer PJ, Moody MP & Haley D (2021). Extending estimating hydrogen content in atom probe tomography experiments where H2 molecule formation occurs. *Microsc Microanal* **28**(4), 1–14.

Meisenkothen F, Samarov DV, Kalish I & Steel EB (2020). Exploring the accuracy of isotopic analyses in atom probe mass spectrometry. *Ultramicroscopy* **216**, 113018. https://doi.org/10.1016/j.ultramic.2020.113018

Mouton I, Breen AJ, Wang S, Chang Y, Szczepaniak A, Kontis P, Stephenson LT, Raabe D, Herbig M, Britton TB & Gault B (2019). Quantification challenges for atom probe tomography of hydrogen and deuterium in zircaloy-4. *Microsc Microanal* **25**(2), 481–488. https://doi.org/10.1017/S143192761801615X

Pröbstle M, Neumeier S, Hopfenmüller J, Freund LP, Niendorf T, Schwarze D & Göken M (2016). Superior creep strength of a nickel-based superalloy produced by selective laser melting. *Mater Sci Eng A* **674**, 299–307. https://doi.org/10.1016/j.msea.2016.07.061

Radovic A, Williams M, Rousseau D, Kagan M, Bonacorsi D, Himmel A, Aurisano A, Terao K & Wongjirad T (2018). Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560**-(7716), 41–48. https://doi.org/10.1038/s41586-018-0361-2

Ren Y, Autefage H, Jones JR, Stevens MM, Bagot PAJ & Moody MP (2021). Developing atom probe tomography to characterize Sr-loaded bioactive glass for bone scaffolding. *Microsc Microanal* **28**(4), 1–11.

Rigutti L, Di Russo E, Chabanais F, Blum I, Houard J, Gogneau N, Largeau L, Karg A, Eickhoff M, Lefebvre W & Vurpillot F (2021). Surface microscopy of atomic and molecular hydrogen from field-evaporating semiconductors. *J Phys Chem C* **125**(31), 17078–-17087. https://doi.org/10.1021/acs.jpcc.1c04778

Schreiber DK, Chiaramonti AN, Gordon LM & Kruska K (2014). Applicability of post-ionization theory to laser-assisted field evaporation of magnetite. *Appl Phys Lett* **105**(24), 244106. https://doi.org/10.1063/1.4904802

Sundell G, Thuvander M & Andrén HO (2013). Hydrogen analysis in APT: Methods to control adsorption and dissociation of H-2. *Ultramicroscopy* **132**, 285–289. https://doi.org/10.1016/j.ultramic.2013.01.007

Tsong TT & Liou Y (1985). Cluster-ion formation in pulsed-laser-stimulated field desorption of condensed materials. *Physical Review B* **32**(7), 4340–4357. https://doi.org/10.1103/PhysRevB.32.4340

van der Maaten L & Hinton G (2008). Visualizing data using t-SNE. *J Mach Learn Res* **9**, 2579–2605.

Vurpillot F, Hatzoglou C, Radiguet B, Da Costa G, Delaroche F & Danoix F (2019). Enhancing element identification by expectation–maximization method in atom probe tomography. *Microsc Microanal* **25**(2), 367–377. https://doi.org/10.1017/S1431927619000138

Wattenberg M, Viégas F & Johnson I (2016). How to use t-SNE effectively. *Distill* **1**(10). https://doi.org/10.23915/distill.00002

Yeli G, Auger MA, Wilford K, Smith GDW, Bagot PAJ & Moody MP (2017). Sequential nucleation of phases in a 17-4PH steel: Microstructural characterisation and mechanical properties. *Acta Mater* **125**, 38–49. https://doi.org/10.1016/j.actamat.2016.11.052

Yoo S-H, Kim S-H, Woods E, Gault B, Todorova M & Neugebauer J (2022). Origins of the hydrogen signal in atom probe tomography: Case studies of alkali and noble metals. *New J Phys* **24**(1), 013008. https://doi.org/10.1088/1367-2630/ac40cd

Zhu H (2020). Big data and artificial intelligence modeling for drug discovery. *Annu Rev Pharmacol Toxicol* **60**(1), 573–589. https://doi.org/10.1146/annurev-pharmtox-010919-023324

Downloaded from https://academic.oup.com/mam/article/29/3/879/7078172 by guest on 14 November 2023