Weights : User's guide

D. Haley

January 31, 2018

1 Introduction

This document describes how to use, and interpret the output of *weights* - a part of the APTTools package. The program is designed to assist atom probe users in interpreting mass spectra.

The problem that it specifically solves is the problem of identifying what the normally *unknown* components of a peak's mass *could be*. The core algorithm used in the program is relatively straightforward, and involves brute-forcing all possible combinations of species that could potentially provide the correct mass for a peak, and checking their validity against some known tests in order to reduce the number of probable combinations.

The computational costs are, for sensible analyses, usually very small - taking usually less than a few seconds to complete for most purposes. Note that as the search-space is widened, this can take far (orders of magnitude) longer.

The code is open-source and thus, should you choose to do so, is fully inspectable and furthermore modifiable, as you see fit. This program was intially developed to solve mass identification in atom probe in 2015 [1]. The work has since been considerably extended.

1.1 Installation and running

The procedure for installing the program depends on your system. Binaries ("executables") are made available for the windows platform. For Unix based systems (e.g. Linux, or MacOSX), users are encouraged to compile from the source code. The program is highly portable, and should be able to run on any platform for which it can be compiled.

1.1.1 Linux and Mac OS

For "binary" versions, simply unzip the program, and launch it by clicking on the program icon.

1.1.2 Windows

First unzip the program to a folder. You can launch the program by clicking on it.

Note that Windows force-closes terminals on program termination. Thus, if you wish to use an input file (described later), you must use the command line. Use the keyboard shortcut Win+R, then type "cmd" into the "run" prompt.

Use "cd" to navigate to the folder with weights installed. To run the program, type:

```
weights.exe [INPUTFILE]
```

note that the INPUTFILE is optional, and specifies the name of the file you wish to use as the input. Do not literally type [INPUTFILE].

1.2 Contact

The authors can be contacted through our website, `http://apttools.sourceforge.net`. We are happy to answer any questions, such as (but not limited to) using, compiling, or modifying the program.

Bug reports can be made through the website, and we will attempt to deal with these as promptly as is possible. Please be aware that bug-fixing can be a tricky process, and we need to be able to replicate your problem on our systems, before we can fix bugs. We may need to contact you further for additional information if we cannot reproduce the problem.

1.3 Limitations

The program cannot necessarily provide unique answers for identification of a peak - this is a limitation of the APT method at this time. The program does not take into account any form of molecular stability, and only utilises isotopic mass data, and optionally mass spectrum information. Users may need to consider the output of the program carefully.

1.4 Licence

This program is free and open-source, and is made available under the GPLv3 licence. You are free to copy, modify or redistribute it, with limited restrictions. If you have paid for this program, please request a refund. As explained in the licence, no guarantees are made regarding fitness for your purpose. The full text of the licence is packaged with the program, and is also available at `http://www.gnu.org/licenses/gpl-3.0.html`.

2 Getting started

2.1 Launching the program

Under Unix-like systems, once you navigate to the folder containing the program *via a command line*, you should be able to run the program with the command `./weights`. Under Windows it is also recommended to use the command line - this is available with the "Run" feature in your start menu, or via windows+R then execute the command "cmd". Under more recent versions of windows, you may have to first add the "run" command to your program launcher (*e.g.* start menu, or tile, or whatever)[1][2]. Double clicking the program icon will not allow

---

[1]`http://windows.microsoft.com/en-gb/windows7/customize-the-start-menu`

[2]`http://blogs.msdn.com/b/zxue/archive/2012/03/07/win8-howto-12-show-and-use-run-command.aspx`

you to provide an input file, and will greatly limit the analysis options available to you.

2.2 Basic usage

The program has two operation modes. The simplest mode to use is the so-called *interactive* mode where the program will prompt the operator with questions regarding the analysis, whilst the alternate "batch" mode requires supplying a text file with all the needed parameters predefined. Batch mode is significantly more powerful than interactive mode, as it gives you access to more features of the program. Here we demonstrate the most basic usage of the program. User inputs are coloured in blue.

———————————

```
Enter element symbols (eg H), space separated. Blank for all elements:
 You can also put groups of elements in, using eg [10-100], [-50],[23-]
Fe C Cr Si P B S N
Enter target weight
34.3
Maximal number of elements to combine?
3
Mass tolerance?
0.05
Max charge state?
3
Select post-processing/filter mode:
        0:None
        1:Compositional ranking
        2:Major Isotopes only
        3:Minimum PPM value
0
```

———————————

Here we specify the so-called *search-space* that we wish to examine. The search-space is the possible combinations of parameters that we consider worth searching.

There are 5 parameters that determine the search space, as follows:

- the list of elements that one wishes to consider. For conciseness, in this example contaminant species (*e.g.* H, O) have been omitted.

- The mass of interest, here 34.3 amu.

- The maximum number of elements that can be combined to reach the target mass

- The tolerance on the search - this corresponds to the uncertainty at the *centre of the peak*. **Not** the width of the peak (which is usually far wider). Most atom probes should be able to achieve a mass certainty less than 0.1 amu.

- The maximum charge state allowable for the combined ion. This will very rarely exceed 4, and 3 is normally sufficient.

————————

```
No peak file provided. To limit results, you can supply an observed peaks file...
Target weight :34.3 Tolerance :0.05
Searching...Done
23 solutions
3+ charge state:
34.3021 : 36^S:34^S:33^S
34.3026 : 36^S:36^S:31^P
34.3026 : 52^Cr:36^S:15^N
34.3036 : 53^Cr:36^S:14^N
34.2956 : 53^Cr:50^Cr
34.3023 : 54^Cr:34^S:15^N
34.3031 : 54^Cr:36^S:13^C
34.3025 : 54^Fe:34^S:15^N
34.3034 : 54^Fe:36^S:13^C
34.3024 : 56^Fe:32^S:15^N
34.3032 : 56^Fe:33^S:14^N
34.3021 : 56^Fe:34^S:13^C
34.3038 : 56^Fe:36^S:11^B
34.3031 : 57^Fe:31^P:15^N
34.3035 : 57^Fe:32^S:14^N
34.3034 : 57^Fe:33^S:13^C
34.3011 : 57^Fe:34^S:12^C
34.3051 : 57^Fe:36^S:10^B
34.3024 : 58^Fe:30^Si:15^N
34.3034 : 58^Fe:31^P:14^N
34.3029 : 58^Fe:32^S:13^C
34.3016 : 58^Fe:33^S:12^C
34.3035 : 58^Fe:34^S:11^B
```

————————

To interpret the output, each charge state is listed in turn, if there are solutions - in this example only the 3+ charge state has a solution. The formatting of the output is as follows: `36^S:34^S:33^S` means a combination of three S ions, i.e. S_3. More specifically there are several differing isotopes forming the S_3 molecule : $^{33}S^{34}S^{36}S$. Each line represents a new possible combination that fits the target mas, and is preceded with the combination's respective mass.

The output lists all *possible* combinations of elements that can be emitted - including highly *improbable* ones (such as $^{33}S^{34}S^{36}S$). There are several modes in the program which allow both reducing the number of possible combinations, and furthermore, ranking the output, in order to separate the likely and unlikely combinations.

In most analyses, isotopes are normally have very similar masses. Thus it is most likely that if the largest unknown peak is selected, it will originate as a combination of the most prevalent isotopes. This method, though crude is relatively simple to use, however it works best if there is a large difference in

4

isotopic abundances. Thus the simplest solution is to ignore minor isotopes. Returning to our previous example, we will change the "filtering" mode for the program, but in addition add our "contaminant" species, H and O.

––––––––––––

```
Enter element symbols (eg H), space separated. Blank for all elements:
 You can also put groups of elements in, using eg [10-100], [-50],[23-]
Fe C Cr Si P B S N H O
Enter target weight
34.3
Maximal number of elements to combine?
3
Mass tolerance?
0.05
Max charge state?
3
Select post-processing/filter mode:
        0:None
        1:Compositional ranking
        2:Major natural peak only
        3:Minimum PPM value
2
No peak file provided. To limit results, you can supply an observed peaks file...
Target weight :34.3 Tolerance :0.05
Searching...Done
1 solution
3+ charge state:
34.3012 : 56^Fe:31^P:16^O
```

––––––––––––

Here we select the filter mode "Major natural peak only". *i.e.* only the isotope combinaitons with the largest natural abundances are kept. As such, most of the solutions are now rejected, as they are composed of minor ions. Only one major solution remains - Fe-P-O. Note the drastic reduction ion the number of output combinations! Whilst it is possible that other combinations of species could be a solution - one would expect nearby peaks in the mass spectrum with larger amplitudes in most cases. More advanced analyses are possible and will be covered in subsequent sections.

2.3 Filtering output

2.3.1 Compositional

In the majority of cases, there will be a very large list of possible combinations of species to form a given mass, as the space in which combinations occur is vast. For a simple 2 -species problem, say Fe-Ni, up to and including charge state 3 there are $(3 \times 54) \rightarrow 162$ different peak combinations possible, each combination giving rise to a new mass. The search space explodes factorially with the number of input species (Combination problem), and thus limits the output using any available information is paramount.

This program allows for several methods for reducing the number of output options. Firstly, the simplest method of filtering out the likely, versus the unlikely species, is via providing an approximate compositional weighting for each species. This is shown here for the basic usage example 2.2 in interactive mode. Continuing on with that example, we select the "Compositional Ranking" mode, and enter in approximate fractional values for the given peak.

```
Select post-processing/filter mode:
        0:None
        1:Compositional ranking
        2:Major natural peak only
        3:Minimum PPM value
1
        B
0.001
        C
0.01
        Cr
0.005
        Fe
0.9
        N
1
        P
0.001
        S
0.005
        Si
0.005
No peak file provided. To limit results, you can supply an observed peaks file...
Target weight :34.3 Tolerance :0.05
Searching...Done
23 solutions
3+ charge state:
34.3026 : 36^S:36^S:31^P Score: 2.500000e-16
34.3021 : 36^S:34^S:33^S Score: 3.984375e-15
34.3031 : 54^Cr:36^S:13^C Score: 6.326375e-15
34.3051 : 57^Fe:36^S:10^B Score: 1.897565e-12
34.3034 : 54^Fe:36^S:13^C Score: 2.814367e-12
34.3026 : 52^Cr:36^S:15^N Score: 7.624798e-12
34.3034 : 57^Fe:33^S:13^C Score: 7.652239e-11
34.3023 : 54^Cr:34^S:15^N Score: 9.146638e-11
34.3036 : 53^Cr:36^S:14^N Score: 2.366604e-10
34.3038 : 56^Fe:36^S:11^B Score: 3.307273e-10
34.3035 : 58^Fe:34^S:11^B Score: 4.319994e-10
34.3016 : 58^Fe:33^S:12^C Score: 9.415663e-10
34.3029 : 58^Fe:32^S:13^C Score: 1.289803e-09
34.3024 : 58^Fe:30^Si:15^N Score: 1.428244e-09
```

```
34.3021 : 56^Fe:34^S:13^C Score: 1.877631e-08
34.3011 : 57^Fe:34^S:12^C Score: 4.009225e-08
34.3025 : 54^Fe:34^S:15^N Score: 4.068997e-08
34.3031 : 57^Fe:31^P:15^N Score: 6.941844e-08
34.2956 : 53^Cr:50^Cr Score: 1.032046e-07
34.3034 : 58^Fe:31^P:14^N Score: 2.528762e-06
34.3024 : 56^Fe:32^S:15^N Score: 1.427634e-05
34.3032 : 56^Fe:33^S:14^N Score: 3.085425e-05
34.3035 : 57^Fe:32^S:14^N Score: 9.024800e-05
```

Here we see that our previous *improbable*, but possible, example is now ranked with a very low score. More realistic combinations, such as $^{53}Cr^{53}Cr$ are, comparatively, ranked higher to the top of the list.

The score value is derived from multiplying the compositional value entered for each species by the isotopes natural abundances. This causes unlikely combinations to a low score (because they have very low abundances, they are unlikely to appear), and high scoring values to the top.

As can be seen in the input, the normal compositional values were used, with the exception of N. For N, in this example, the semi-nonsensical value 1 was entered as a relative fraction in the material - this is merely to indicate that the amount of nitrogen is uncertain (*e.g.* as N can be introduced from air) - thus we assume that it is highly prevalent. Thus more complex N containing species score well - this is an artefact of the inherent uncertainty in the compositional ranking scheme. The order in the rank is *indicative only*, as should be used as a *very* rough rule-of-thumb guide. The formation of molecular ions in high fields is significantly more complex than is allowed for in this program - careless use of the program will lead to wrong answers!

Compositional ranking should only be used as a last resort. More robust methods for eliminating impossible peaks such as via the peak file (though this requires more user input), are available, and it is recommended that these be used first when possible.

2.3.2   Mass spectrum

In the previous examples, it may have been noticed that the program emits the following cryptic statement:

```
No peak file provided. To limit results, you can supply an observed peaks file...
```

The option to provide this file is not available in interactive mode at time of writing (but may be available in future versions) – batch mode must be used (See Section 4). The peaks file uses a *complete* list of peak centres observed in the mass spectrum, in order to identify if other peaks, which, if detectable, were present in the original mass spectrum. In the simple case, the file is simply a single-floating point value per line, in ASCII, as follows. No amplitude data is required.
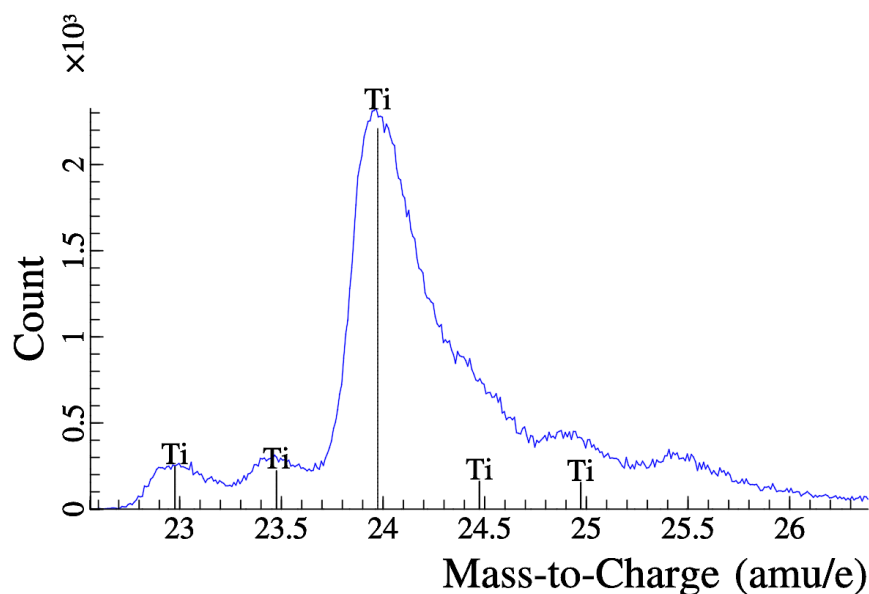
Figure 1: "Fingerprint" pattern of isotopes for titanium

```
1.0
2.0
12.1
13.05
23.445
22.334
```

To understand how this isotopic elimination using the peak position data can work, one must consider that each isotope combination has a given isotopic fingerprint - that is a pattern of peaks of differing amplitudes and masses. If a peak of greater amplitude from a specific isotope combination *should be there and is not*, the isotopic combination *cannot have occurred*, and can be eliminated. To illustrate, Figure 1 shows an example expected isotopic pattern for Ti. If the main peak is not present in the peaks-file, then any side peak (*e.g.* at 24.9) cannot be Ti - the main peak is required.

This reduction is very powerful, but requires a complete list of peaks, which can be tedious to build at this time. An example of such reduction is given (using the example provided with the program):

```
./weights examples/input-example

Target weight :93.3 Tolerance :0.15
Searching...Done
Peakfile filtering result :23 -> 1 solutions
1 solution
3+ charge state:
93.2397 : 94^Zr:93^Nb:93^Nb
```

As can be seen in the output, for this example, the otherwise 23 possible solutions have been reduced to a single solution, even with a very wide tolerance of 0.15 amu, due to the data present in the peaks file.

2.3.3 Intensity analysis

Peak-position analysis is quite powerful, however, it often will leave several possible solutions for a given peak. This is a limitation in APT, and is not fully avoidable. However, with additional information, it is possible to rank the solutions further.

By providing not only the peak positions, but also the intensity of the peaks identified, it is possible to answer a simple question : *What is the maximum extent that a given solution can explain the intensity of a particular peak?*

Such an analysis can be performed by looking at the intensity of other peaks in the dataset, and examining the ratio of the observed and theoretical counts. This ratio limits the maximum possible amplitude of a given isotopic fingerprint, and thus locks in a maximum possible scaling for this peak. Thus we can provide a value, which is the "explained fraction" of a given peak, which varies between 0 (least likely) and 1 (most likely).

Providing this data is very simple, and follows the peak-file method. However, instead of providing a single-column file, the position and intensity data are now provided as a pairing. An example file might look like:

```
1.004 100
2.004 50
6.0    1000
9.0    300000
```

Note that for every peak, a corresponding intensity must be provided.

In this case, the first column is the peak mass-to-charge (in amu/charge), and the second is the number of observed counts (it is recommended to perform background subtraction before filing in the count data). This must be provided wherever possible - note that there is a lower limit below which the count data is ignored (currently 10 counts), in order to prevent spurious calculation of the theoretical to observed ratio.

If peaks are missed, it may generate an erroneous result.

Once calculated, you will see output like this, for a peak near 16.31 (intensities not shown):

```
3+ charge state:
16.3254 : 1^H:1^H:1^H:46^Ti Explained: 1.21e-01
16.3225 : 1^H:1^H:47^Ti Explained: 1.00e-01
16.3297 : 16^O:16^O:17^O Explained: 6.05e-02
16.3345 : 12^C:12^C:12^C:13^C Explained: 1.00e+00
16.3309 : 1^H:16^O:16^O:16^O Explained: 1.00e+00
16.335 : 1^H:14^N:16^O:18^O Explained: 2.52e-06
16.3186 : 1^H:48^Ti Explained: 1.00e+00
16.316 : 49^Ti Explained: 6.58e-01
```

In this case, the intensity data suggests that either the carbon, TiH_x or possibly Ti are suitable candidates solutions. NO2H and O3 are considered less likely, as the side peaks for these do not provide much intensity. Note that the answers sum to greater than one, as these are maximum possible values.

Note that this intensity analysis is a less potent, but more robust analysis than linear fitting of peak intensities (aka decomposition). This is due to the slightly different mathematical problem - the "Explained fraction" question can always be answered. A linear model may not be able to be answered, due to rank-deficiency, however the linear model will provide more accurate results than this method, where it can.

2.3.4   PPM limit

There exists an option to use the natural abundances, in combination with the instrumentation's theoretical detection limit, in order to limit the number of matching species for any given peak. This is achieved by computing the highest possible composition for a given species or molecule (by assuming the molecule is the only species present), and discarding the option if even this molecule's isotope is below the detection limit. Thus the only required input is the maximum PPM (parts-per-million (atomic)), that can be observed by the experimental equipment.

As an example, ^{17}O is a rare oxygen isotope - even in pure oxygen, this isotope only occurs naturally at a concentration of 380 ppm. If your experimental apparatus is only able to detect species down to 1000 ppm, then it is impossible for you to observe this isotope. Thus one can discard this option as matching any experimental peak. In atom probe systems, at time of writing, typical detection limits are around 30 ppm. Thus it is possible to observe ^{17}O, but it is impossible to observe $^{17}O^{18}O$, which has a natural abundance of 0.76 ppm.

The value is given in the input file as `ppmlimit=` *value*, where the value is in atomic ppm. This will only be used if the filtering mode is set to "Minimum PPM value". This feature can be used in conjunction with a peak file, by specifying the correct option in the input file.

3   Ion Database

It is possible to include an "Ion database" of ions that have been seen in experiments, in order to mark ions that have been identified by others previously. The database is built from a series of rangefiles, and at this time, must be supplied by the end user. The database format is very simple, and requires three columns, indicating ion, charge state and the number of observations that exist for that ion. An example is given below, which shows (at first) aluminium being identified in 3 charge states, and this was found by at least 32, 12 and 17 observations from the input used to build the observation database. Multiple ions are shown as per below, e.g. FeO and FeO_2. Note that the mass is not checked, so any FeO_2 ion detected, including a low-count FeO_2 with 18-Oxygen, will show as a result.

```
Al 1 32
Al 2 12
Al 3 17
```

```
FeO 1 32
FeO 2 43
FeO2 1 5
```

To use this, simply add the following to the batch input (see Section 4):

```
ionDatabase=database_filename
```

When using this, the output observations that match an entry in the database will be marked. For example, using a database with VH, in charge state 1, the following output is given as an example.

```
Target weight :51.94 Tolerance :0.05
Searching, with 31 isotopes
Done
Peakfile filtering result :3 -> 1 solutions
1 solution
1+ charge state:
*51.9518 : 1^H:51^V

 * - denotes ions present in previously seen database
```

There are a some example databases given with the program, separated by experiment mode. These have been derived from a limited pool of rangefiles, so will likely be incomplete. We would greatly appreciate any further rangefiles, particularly from unusual materials, which would allow us to extend the list of rangefiles that we can handle.

4   Batch input mode

Batch mode can be used to run the program without it pausing for input from the user. This can be useful if attempting to integrate the program into part of a wider system of analysis, or otherwise attempting to automate analysis of a large number of peaks. Batch mode can be run by listing the program as

4.1   Batch input options

Input in the batch file mode is given using a *key=value* pairing. The definition of each key is listed below, and under what conditions the key is required input.

- **elementComp**/Element : Only required when performing compositional ranking. This is a pseudo-composition to assign this element to determine the sort order of presented solutions.

- **ionDatabase** : A database of known ions, which can be marked up if they are seen in the solution

- **massTolerance** : The tolerance that a solution can deviate from the target weight. This is also used when performing peak file analysis, to determine the width inside which peaks must fall when checking for alternate peaks (peak filtering).

- **maxChargeState** : The maximum charge state to calculate solutions for a given weight. For example, setting a charge state of 3 will check all ions with charge state 1, 2 and 3 inclusive.

- **maxElemCombine** : The maximum number of elements that can be used when searching for an ion solution.

- **peakFile** : This option provides a filename to a file that has a list of peaks that can be experimentally observed. Use this option to enable peak-file filtering

- **postFilterMode** : This option sets the method of post-search filtering to limit the number of observed results

- **ppmLimit** : The value below which a theoretical pure material composed of a fixed species would not be experimentally observable, in atomic ppm. Use in conjunction with "ppm" filtering mode

- **searchElements** : Required. The list of elements that may be present in the analysed material, and that should be used for the peak search.

- **targetWeight** : Required. The mass-to-charge of the unidentified peak. This should be the centre of the peak, and should be reasonably accurate (usually 2 decimal places)

4.1.1 Example

An example input file is provided, and is reproduced below. Note that if the postFilterMode is set to "1" (compositional ranking), you will need to provide a compositional estimate for *every* atom. The peakFile will be used to remove impossible ion combinations if provided.

```
# The "#" symbol means comment.

#Compulsory parameters
#----------
#Elements to search for, case sensitive
searchElements=Zr Nb
#Weight of the peak you want to find
targetWeight=93.30
#Number of elements that can be combined into a molecular ion
maxElemCombine=4
#Maximum charge state for searching.
# Eg, for 4, search for 1->4 as possible solutions
maxChargeState=4

# The tolerance (in amu) for the peak search.
massTolerance=0.15

#0: None, 1:Comp. Ranking, 2: Major Natural peak only
# 3: Minimum PPM value
postFilterMode=0
```

```
#----------


#Optional parameters
#------------
#Experimental peaks file used to aid in culling impossible peaks
# Format of the file is simply a list of peaks one on each line,
# each giving the peak position.
peakFile=./examples/peaks-example

#This is the weighting to give to each element when performing probability
# calculations/compositional ranking. As a rough approximation, use the
# bulk composition. Does not need to sum to one, as it will be normalised.
# Every element must be supplied.
elementComp/Zr=.12
elementComp/Nb=0.03
```

As abundance data is loaded dynamically in the program, non-natural abundances can be dealt with by modifying the `naturalAbundances.xml` file supplied with the program, as long as the synthetic abundance is known. This can be done with any text editor.

5 Copyright

This document is copyrighted under the Creative Commons share-alike licence, 3.0. A full copy of the licence is available at `https://creativecommons.org/licenses/by-sa/3.0/legalcode`.

References

[1] Daniel Haley, P Choi, and Dierk Raabe. Guided mass spectrum labelling in atom probe tomography. *Ultramicroscopy*, 159:338–345, 2015.