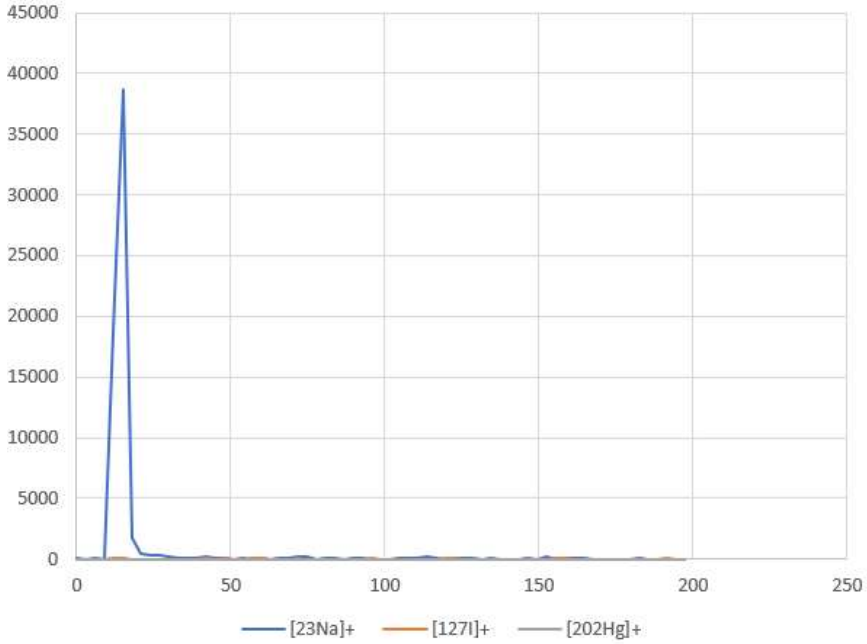
What is this program?

The two phase analyser is a program that does a statistical analysis of chemical data.
The chemical data is the output of a machine called a “mass spectrometer”.
A mass spectrometer is a machine that can analyze a sample and tell you exactly which chemicals (atoms, or more correctly: ions) are in it.
For example: you put a sample in the machine, it runs the test, and then it tells you that the sample contains x amount of sodium, iron, potassium, …

The mass spectrometer used for this study is even more detailed: it outputs time values for every ion. So you can tell not only the total amount of an ion, but also when it got detected. So it generates a table where the rows are time intervals and the columns are the amount of ions detected in that time interval.

index	time (ms)	[26Mg]+ m	C2H2+ ma	[27Al]+ m	C
0	0	0,000345	0	0,000894	
1	2,99	0,00038	0,000383	0,000304	
2	5,98	0	0,000184	0,0009	
3	8,97	0,000231	0	0,00071	
4	11,96	0,004241	0,0007	0,052345	
5	14,95	0,011593	0,000718	0,093615	
6	17,94	0,000403	0	0,003996	
7	20,93	0	9,59E-05	0,003335	
8	23,92	0	0,000174	0,002487	
9	26,91	0	0,000106	0,000709	
10	29,9	0	0,000282	0,001973	
11	32,89	6,57E-07	0,000366	0,001625	
12	35,88	0	0,00034	0,002015	
13	38,87	0,000163	0	0,000837	
14	41,86	0,000397	0	0,000651	
15	44,85	8,80E-05	9,00E-05	0,001858	
16	47,84	0	0,000251	0,000611	
17	50,83	8,93E-05	3,35E-05	0,000542	
18	53,82	0	0,000241	0,000363	
19	56,81	0	0	0,000533	
20	59,8	0,000292	0	0,000795	
21	62,79	0,000168	0,000492	0,000133	
22	65,78	0,000145	0,000207	0,000846	
23	68,77	0	9,78E-05	0,000427	
24	71,76	0,000591	0	0,000667	
25	74,75	0	0,000102	0,000467	
26	77,74	0	0	0,000754	
27	80,73	5,58E-05	0,000219	0,001219	
28	83,72	0,000273	0	0,001089	
29	86,71	0	0,000118	0,001032	

you could use this to make a time graph:



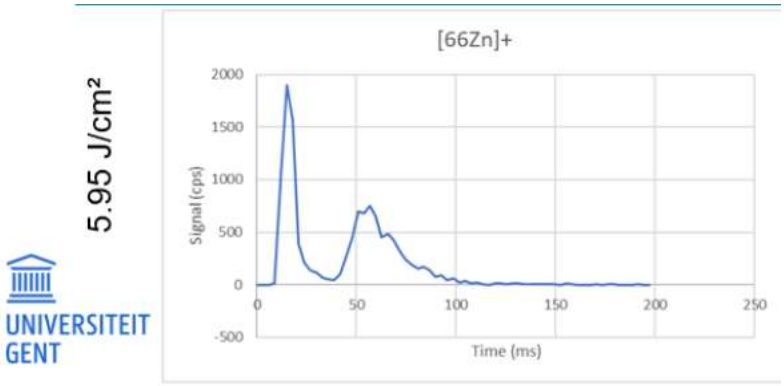
(the horizontal axis is time in ms, vertical axis is the amount of ions).

This graph looks like you would expect the graph to look: all of the ions arrive at approximately the same time. So you see one peak.
But something happened when a laser got added to the machine, sometimes a second peak appeared, and that is what this software is going to analyze.

Second peak phenomenon

A laser got added to the machine on the input-side of the mass spectrometer. So now instead of simply putting the sample in the mass spectrometer, the sample first goes into a laser, gets hit whit the laser beam, and travels to the mass spectrometer through a carbon tube connected to the input side of the mass spec.

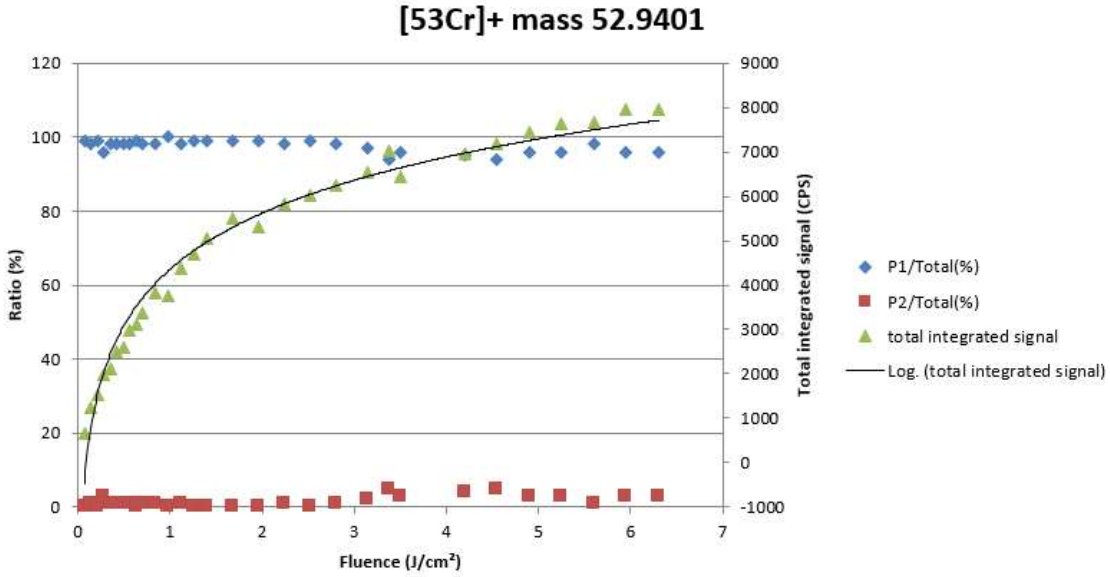
This caused, in some edge cases, an unexpected result: a second peak appeared!



A comprehensive study was done about this, and it concluded that the two peaks are the same ion, but in a different phase (of matter): the first peak is solid-phase (solid phase particles floating in a gas). The second peak is gaseous phase. (so the material in the second peak is a gas: it evaporated).

What does the two phase analyser do?

The program can easily read a large number of output files (output files of the mass spec, in a tsv format). You simply select the directory that holds the files. It will output an excel in which the relationship between laser intensity (energy/cm²) and the distribution of signal in peaks 1 and 2 gets plotted:



This graph is the result of combining 20+ runs off the mass spectrometer, always varying the “fluence” (intensity of the laser). That’s the horizontal axis.
The blue and red lines are peak 1 and peak 2 respectively. As you can see, for this element (53Cr+), the second peak is bigger in the area between 3 and 5 J/cm².

Input and flow of the program

The massaspec outputs a tsv file for every run. It looks like this:

```
Segment profile
==============
Data source: F:\Tom\220902_Two-phase\220831\Raw\E series gel2_20micron\16h17m07s_Eseries_rest_all.h5
Average over all bufs/writes (100 profiles)
Number of profiles (masses): 316
Segment length: 67 points (200.33 ms)

index   time (ms)       [6Li]+ mass 6.01546     [7Li]+ mass 7.01546     [9Be]+ mass 9.01163     [10B]+ mass 10.0124     [11B]+ mass
11.0088 [12C]+ mass 11.9994     [13C]+ mass 13.0028     [14N]+ mass 14.0025     [15N]+ mass 14.9996     [16O]+ mass 15.9944     OH+ mass
17.0022 [18O]+ mass 17.9986     H2O+ mass 18.01 H3O+ mass 19.0178       [23Na]+ mass 22.9892    [24Mg]+ mass 23.9845    [25Mg]+ mass
24.9853 [26Mg]+ mass 25.982     C2H2+ mass 26.0151      [27Al]+ mass 26.981     C2H3+ mass 27.0229      [28Si]+ mass 27.9764    N2+ mass
28.0056 [29Si]+ mass 28.976     HN2+ mass 29.0134       [30Si]+ mass 29.9732    NO+ mass 29.9974        [31P]+ mass 30.9732     [15N]O+
mass 30.9945    NOH+ mass 31.0053       [32S]+ mass 31.9715     O2+ mass 31.9893        [33S]+ mass 32.9709     O2H+ mass 32.9971
        [34S]+ mass 33.9673     O[18O]+ mass 33.9935    [35Cl]+ mass 34.9683    [36S]+ mass 35.9665     [36Ar]+ mass 35.967     HCl+ mass
35.9761 [37Cl]+ mass 36.9654    [36Ar]H+ mass 36.9748   [38Ar]+ mass 37.9622    [39K]+ mass 38.9632     [38Ar]H+ mass 38.97     [40Ar]+
mass 39.9618    [40Ca]+ mass 39.962     [41K]+ mass 40.9613     ArH+ mass 40.9697       [42Ca]+ mass 41.9581    ArH2+ mass 41.9775
        [43Ca]+ mass 42.9582    [44Ca]+ mass 43.9549    CO2+ mass 43.9893       [45Sc]+ mass 44.9554    CHO2+ mass 44.9971      [46Ti]+
mass 45.9521    [46Ca]+ mass 45.9531    NO2+ mass 45.9924       [47Ti]+ mass 46.9512    [48Ti]+ mass 47.9474    [48Ca]+ mass 47.952
        [49Ti]+ mass 48.9473    [50Ti]+ mass 49.9442    [50Cr]+ mass 49.9455    [50V]+ mass 49.9466     [51V]+ mass 50.9434     ClO+ mass
50.9632 [52Cr]+ mass 51.94      [36Ar]O+ mass 51.9619   [53Cr]+ mass 52.9401    [37Cl]O+ mass 52.9603   [54Cr]+ mass 53.9383    [54Fe]+
mass 53.9391    ArN+ mass 53.9649       [55Mn]+ mass 54.9375    [56Fe]+ mass 55.9344    ArO+ mass 55.9567       [57Fe]+ mass 56.9348
        ArOH+ mass 56.9646      [58Fe]+ mass 57.9327    [58Ni]+ mass 57.9348    [59Co]+ mass 58.9327    [60Ni]+ mass 59.9302    [61Ni]+
mass 60.9305    [62Ni]+ mass 61.9278    [63Cu]+ mass 62.9291    [64Ni]+ mass 63.9274    [64Zn]+ mass 63.9286    [65Cu]+ mass 64.9272
        [66Zn]+ mass 65.9255    [67Zn]+ mass 66.9266    [68Zn]+ mass 67.9243    [137Ba]++ mass 68.4524  [69Ga]+ mass 68.925     [138Ba]++
mass 68.9521    [70Ge]+ mass 69.9237    [70Zn]+ mass 69.9248    [71Ga]+ mass 70.9241    [72Ge]+ mass 71.9215    [73Ge]+ mass 72.9229
        [74Ge]+ mass 73.9206    [74Se]+ mass 73.9219    [75As]+ mass 74.9211    ArCl+ mass 74.9307      [76Se]+ mass 75.9187    [76Ge]+
```

that kind of data in a table format (excel) looks like this:

	A	B	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	Segment profile																		
2	==========																		
3	Data source: F:\Tom\220902_Two-phase\220831\Raw\E series gel2_20micron\16h14m21s_Eseries_rest_all.h5																		
4	Average over all bufs/writes (100 profiles)																		
5	Number of profiles (masses): 316																		
6	Segment length: 67 points (200.33 ms)																		
7																			
8	index	time (ms)	[26Mg]+ m	C2H2+ ma	[27Al]+ m	C2H3+ ma	[28Si]+ m	N2+ mass	[29Si]+ m	HN2+ mas	[30Si]+ m	NO+ mass	[31P]+ ma	[15N]O+ n	NOH+ ma	[32S]+ ma	O2+ mass	[33S]+ ma	O2H+ m
9	0	0	0,000345	0	0,000894	0,000716	0,258863	1,1655	0,01022	0,049234	0,009042	0,034269	0,000175	0,000192	0,000213	0,001353	0,001695	0,005423	0,0278
10	1	2,99	0,00038	0,000383	0,000304	0,000884	0,254122	1,1655	0,012088	0,048688	0,008195	0,027995	0,000317	0,000483	0,000326	0,001992	0,002495	0,005396	0,0216
11	2	5,98	0	0,000184	0,0009	0,000984	0,256071	1,16102	0,010581	0,048988	0,009372	0,031426	0,000294	0,000294	9,98E-05	0,001487	0,001659	0,005995	0,0258
12	3	8,97	0,000231	0	0,00071	0,000278	0,260154	1,19502	0,010116	0,050426	0,008163	0,035107	0,00039	0,000659	0,000613	0,002007	0,001844	0,008442	0,026
13	4	11,96	0,004241	0,0007	0,052345	0,000175	0,256221	1,15856	0,013911	0,049345	0,010016	0,029636	0,001506	0,001867	0,001227	0,001328	0,0016	0,006711	0,0270
14	5	14,95	0,011593	0,000718	0,093615	0,00024	0,270894	1,16527	0,010825	0,049512	0,008474	0,031618	0,002521	0,002463	0,000614	0,002575	0,002981	0,008754	0,0265
15	6	17,94	0,000403	0	0,003996	0,001206	0,246722	1,15194	0,009763	0,046319	0,007574	0,030982	0,000584	0,000431	2,89E-05	0,001134	0,001414	0,006107	0,027
16	7	20,93	0	9,59E-05	0,003335	0	0,273714	1,17372	0,013982	0,055572	0,009757	0,030261	0,000875	0,000875	0,000427	0,001963	0,001881	0,006352	0,0326
17	8	23,92	0	0,000174	0,002487	0	0,255527	1,16604	0,011641	0,056188	0,007038	0,026534	0,00016	0,000465	0,000521	0,000969	0,000965	0,006573	0,0274
18	9	26,91	0	0,000106	0,000709	0,00055	0,256313	1,16902	0,012051	0,046437	0,006452	0,027723	0,000342	0,000136	0,000475	0,001468	0,002077	0,009019	0,0292
19	10	29,9	0	0,000282	0,001973	0,00151	0,255155	1,17584	0,010537	0,050084	0,008648	0,0289	0,000316	0,000229	0,000183	0,001407	0,001939	0,007939	0,0247
20	11	32,89	6,57E-07	0,000366	0,001625	0,000379	0,255693	1,17382	0,012916	0,05304	0,009895	0,032933	0,000188	0,00012	1,77E-05	0,000695	0,001032	0,005675	0,0280
21	12	35,88	0	0,00034	0,002015	0,000539	0,257904	1,16847	0,012972	0,052871	0,010178	0,031215	0,000456	0,000456	0,000195	0,001733	0,001998	0,006391	0,0278
22	13	38,87	0,000163	0	0,000837	0,000928	0,253536	1,15893	0,011155	0,051872	0,009131	0,035059	0,000691	0,000691	0,000416	0,001044	0,001204	0,008286	0,0311
23	14	41,86	0,000397	0	0,000651	0,001049	0,26383	1,18692	0,014417	0,05197	0,006577	0,028648	0,000261	0,000611	0,000431	0,001644	0,002148	0,005681	0,0286
24	15	44,85	8,80E-05	9,00E-05	0,001858	0	0,254741	1,1667	0,012519	0,050614	0,009092	0,031868	0,000263	0,000486	0,000468	0,000999	0,001449	0,005104	0,0223
25	16	47,84	0	0,000251	0,000611	0	0,258483	1,15912	0,011089	0,049323	0,008362	0,033486	0	8,93E-05	0,000106	0,000337	0,000239	0,007178	0,0289
26	17	50,83	8,93E-05	3,35E-05	0,000542	0,000283	0,253047	1,18411	0,015703	0,053247	0,012458	0,032085	0,000403	0,000407	0,000222	0,00159	0,001937	0,00464	0,0254
27	18	53,82	0	0,000241	0,000363	0,001114	0,258203	1,17815	0,009098	0,050414	0,007934	0,029317	0,001128	0,000613	0,000515	0,001425	0,001722	0,006914	0,0269
28	19	56,81	0	0	0,000533	0,000367	0,25605	1,14976	0,008761	0,054968	0,008002	0,029739	0,000997	0,000997	0,000413	0,002453	0,002553	0,004726	0,026
29	20	59,8	0,000292	0	0,000795	0,000473	0,271195	1,1814	0,011303	0,053998	0,011217	0,029253	0,000667	0,000554	0,000529	0,000882	0,001129	0,006132	0,0295
30	21	62,79	0,000168	0,000492	0,000133	0,000856	0,253075	1,17509	0,012968	0,047776	0,006707	0,029504	0,000552	0,000553	0,000185	0,001788	0,001815	0,007122	0,0286
31	22	65,78	0,000145	0,000207	0,000846	0,000614	0,250714	1,15774	0,010796	0,051036	0,00975	0,035045	0,000632	0,000636	0,000609	0,001579	0,001711	0,006699	0,0312
32	23	68,77	0	9,78E-05	0,000427	0,000638	0,251517	1,18895	0,014631	0,059127	0,009842	0,029211	0,000351	0,00055	0,0005	0,000902	0,000793	0,007609	0,0283
33	24	71,76	0,000591	0	0,000667	0,00044	0,256798	1,20617	0,011974	0,04515	0,009081	0,030958	0,000281	0,000359	9,78E-05	0,001687	0,001752	0,004238	0,0246
34	25	74,75	0	0,000102	0,000467	0,000376	0,252836	1,19391	0,011515	0,047123	0,008547	0,030152	0,000417	0,000693	0,000555	0,000682	0,000918	0,007675	0,0299

Every column represent data for one ion.
The class “IonClass” is build around that data. Ionclass hold the data values and some other properties that are used later in the analysis.

The whole table represents one experiment.
The class ExperimentData is build around that data. It hold all IonClass objects of the experiments + some other properties (like laser intensity, timestamp, …)

A group of experiments (this study consisted of 90 experiments) is represented by the “MemoryClass”. It holds all the ExperimentData objects that are going to be analysed.

The sample data in this directory holds 90 output files. That amounts to a total of 2+ million data points.

Selecting the best gel for each ion

For this experiment,gels (gelatins) were used to carry the ions. So the chemist takes a gel and injects it with the ions that he wants to measure. Multiple gels had to be used, because only a limited amount of ions can be added to a gel before it becomes too reactive.
The two phase analyser selects, for each ion, the gel that contains the most of that ion.

That data gets stored in the MemoryClass field dictIonGelPeak. The structure is:

```
{ionname : { “GelTotal” : {gelname1  : totalsignal ,
                           gelname2 : totalsignal  ,
                           gelname3 : totalsignal },

             “BestGel” : “gelname” ,

            “IonOfEset” : { Esetvalue : Iondata,
                            Esetvalue : Iondata,
                            Esetvalue : Iondata,
                            ……                  },
ionname : { “GelTotal” : {gelname1  : totalsignal ,
                           gelname2 : totalsignal  ,
                           gelname3 : totalsignal },

             “BestGel” : “gelname” ,

            “IonOfEset” : { Esetvalue : Iondata,
                            Esetvalue : Iondata,
                            Esetvalue : Iondata,
                            ……                  },
                  ……..
```

So the subdictionary “IonOfEset” contains, for each ion, a dictionary linking iondata’s to the Esetvalues.
This dictionary is later transformed into a graph (first into a pandas series, then into a graph).