```
Autosaving every 60 seconds
```

3. more advanced aspects

2nd-AUS-FTICR

Marc-André Delsuc - Prague 26-30 Sept 2021

a developed content of this part can be found on github.com/delsuc (https://github.com/delsuc/Fourier_Transform/blob/master/Definition_Properties.ipynb)

Time resolved MS

All the steps presented so far are producing a final spectrum.

is outside of the presentation:

- peak picking
  - detection
  - filtering
  - centroid fit ⇒ precise position and width estimate
- calibration
- interpretation ...

what about Time resolved MS ?

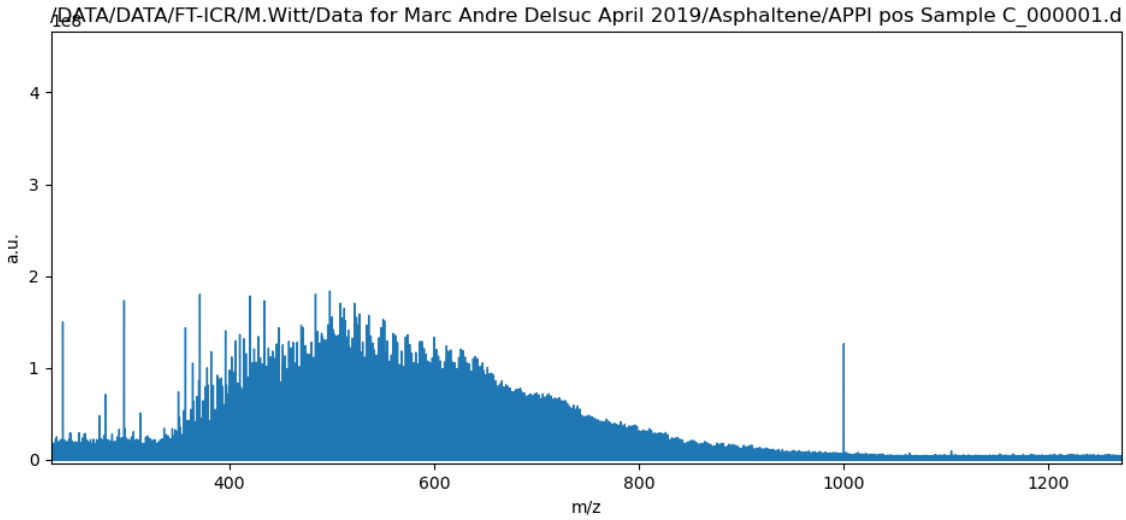
problems with

- loss of information
- artefact cannot be explored
- instanteneous variation lost, improved - slower - method, ...)
- choices made during processing cannot be revisited

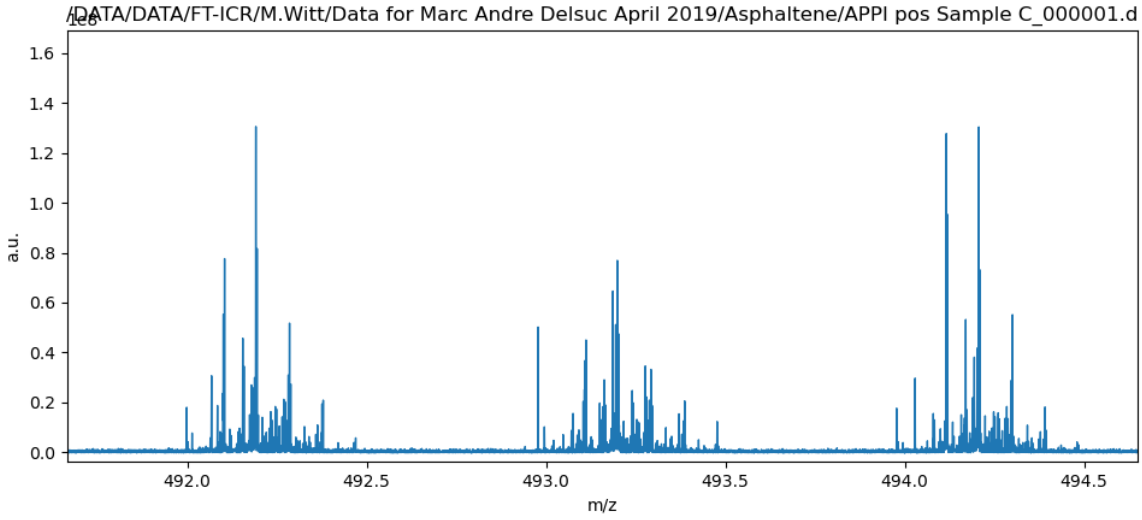
One example

2 years ago Matthias W. sent me some data-sets from Asphaltene samples

I processed them - every thing is normal !

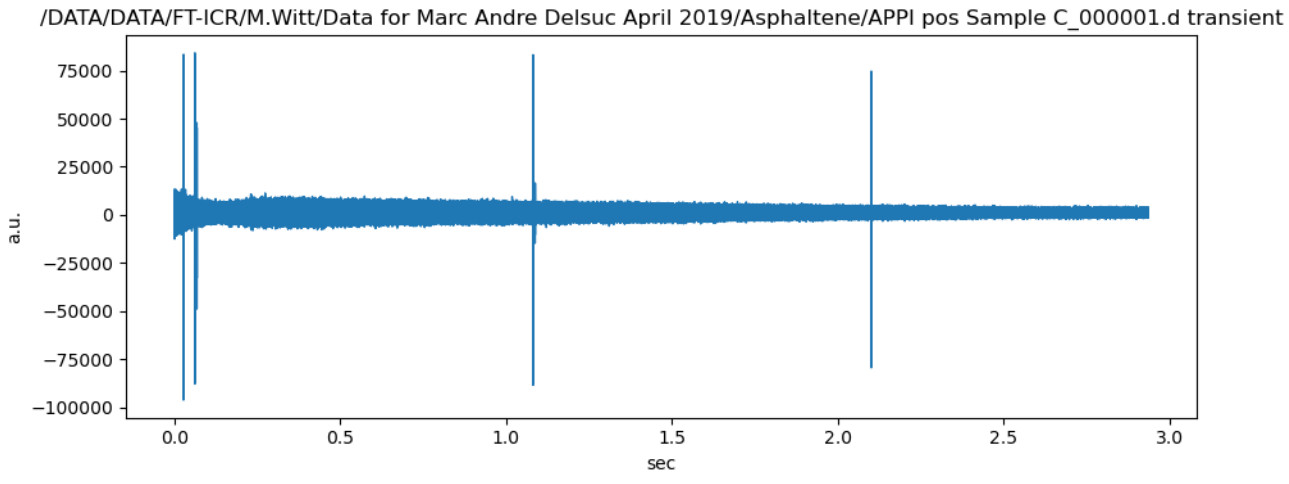


very normal !

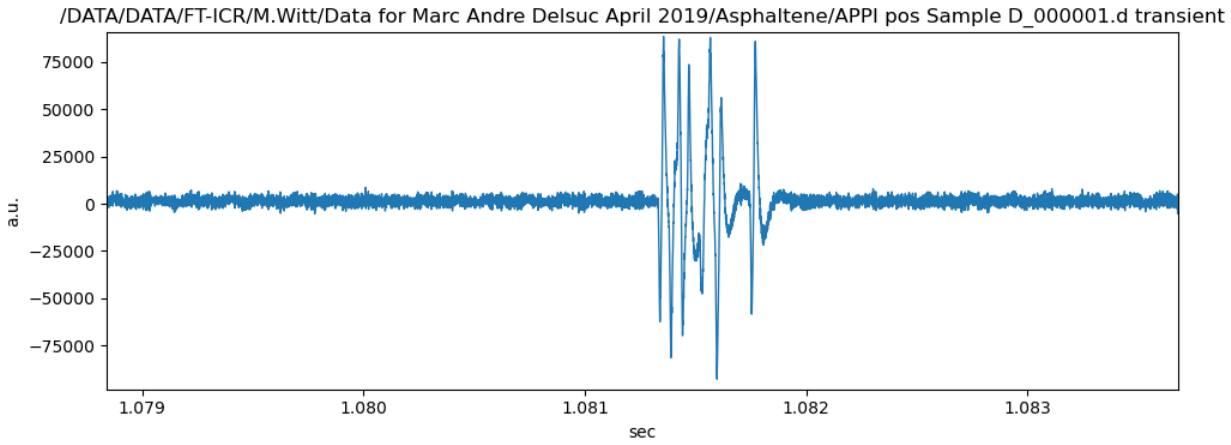


Really ?

looking at the time domain data-set



Really ?



FT stability !

Open-Science trend

Open Science - Open source softwares - Open data

save raw / time domain / data

as early as possible in the analysis pipeline:

- for later (re)-analysis
  - eventually with a newer approach, not available at acquisition time
- for transverse data-mining
- needs to be F.A.I.R.
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable

developping every where

- Genomic data !
- many biological data
- official administration data
- XRay data (Cambridge database / PDB database)
  - similar to MS

enforced by many provider

in particular within EU-FTICR-MS Network

But is it feasible ???

Out[8]:

```
(1535611.754328803, 42.1875)
```

Volume of data to store

standard pipeline

`fid` *(big)* → spectrum *(very big)* → parameter lists *(smaller)* → further analysis *(may grow again)*

how big ?

one example

- 15T (or 7T 2XR) - *m/z* 150 ≈ 1.5 MHz ⇒ 3 MHz sampling
- 4 bytes per point ⇒ 12 MB per second ⇒ **42 GB per hour**
- depends on resolving power - but whatever you you, there is only 3600 seconds per hour !
- big ! **but**
  - doable
  - not *much* bigger than NetFlix ! - and much more informational !!

efficient large data-set storage

on example

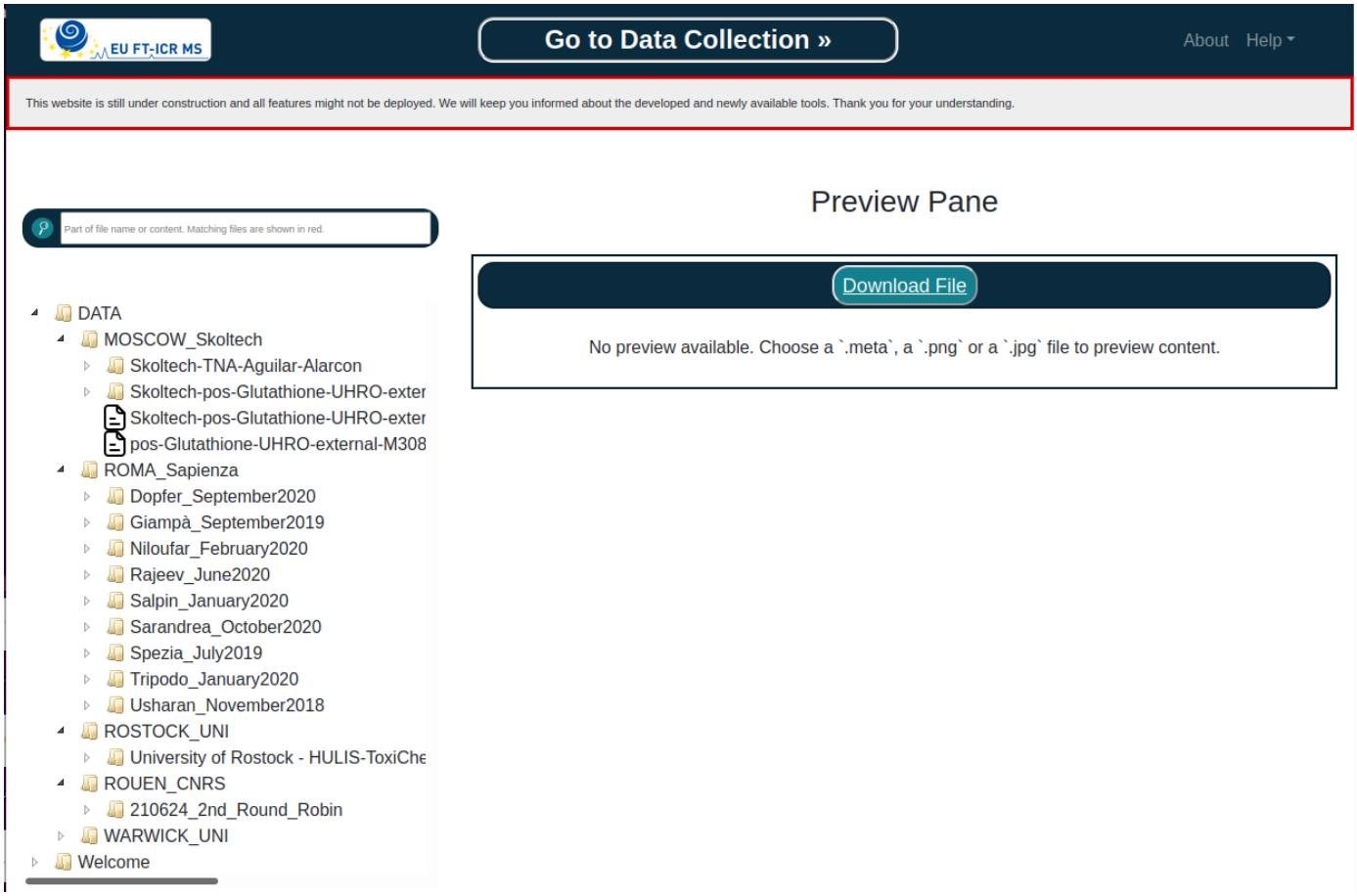
- HDF5 storage developped at CERN for fast access to huge data-sets
- random access along all axes
  - useful in LC-MS / mandatory in 2D-MS
- loss-less compression possibilities on the fly

compression

- spectra are *very* easy to compress efficiently
- time domain are more tricky, and usually poorly compressible
- there are loss-less compression methods for periodic datasets - not used yet -
- only noise is uncompressible

tentative implementation in EU FT-ICR MS network

- alternative analysis pipeline independent of manufacturer software, down to peak-list.
- possibility to display and analyze data-sets
- FAIR **public dataset** deposited and available at [data.eu-fticr-ms.eu/ (https://data.eu-fticr-ms.eu/)](https://data.eu-fticr-ms.eu/)

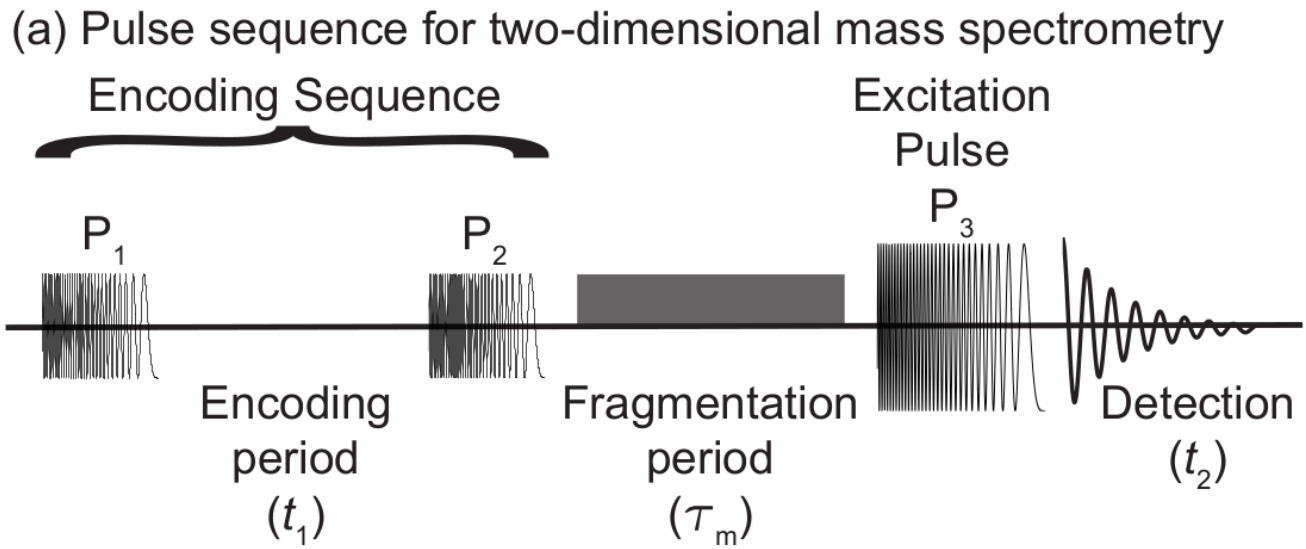


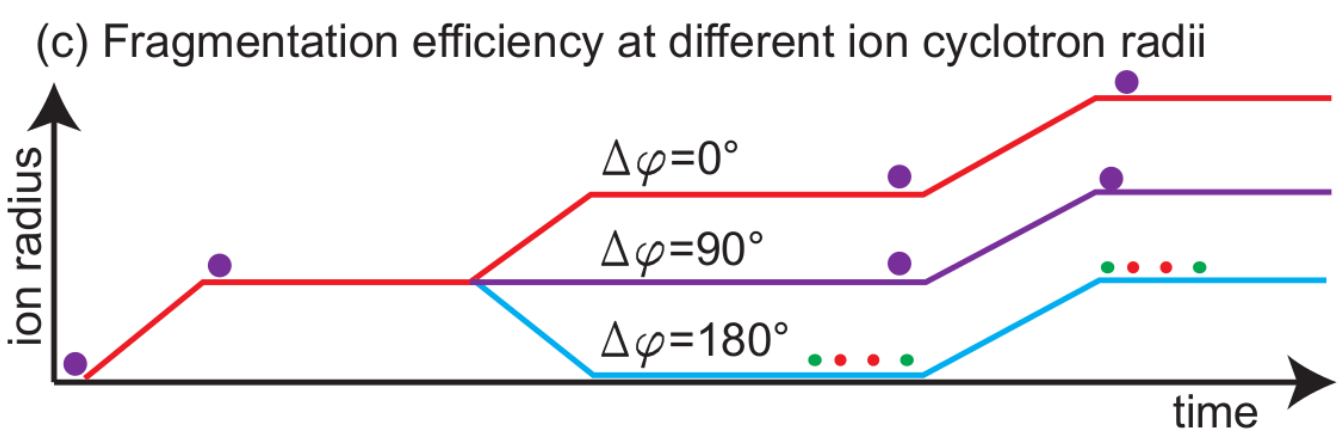
2D experiments

Thematic of the day.

Actually, there is a just a few things which are really special in 2D

2D basic principles





The 3 pulses have different roles

- P_1 excites ions to a first orbit at a given radius,
- P_2 similar to P_1 stops this first evolution
- P_3 is the read pulse which generates the signal that will be acquired

The 3 periods have very different properties

- t_1 is a delay, which will be incremented throughout the experiment
- τ_m is a fixed delay during which fragmentation takes place
- t_2 is a regular acquisition

t1

t_1 is a delay which samples the evolution of the ions between the two pulses P_1 and P_2

- we repeat several time the whole experiment, while incrementing t_1
- we use a constant increment Δt_1 applying Nyquist rule to this period

$$\Delta t_1 = \frac{1}{2F_1^{max}}$$

- the intensition of the fragment ions measure during t_2 is modulated by the evolution of the precursor during t_1

$$S(t_1) \propto \sqrt{1+\cos(\omega_1 t_1)}$$

experimental protocol

Using a setup with a constant (as constant as possible) source of invariant sample

- start with a given t_1 (usually close to 0)
- repeat many time:
  - run the sequence above
  - store the resulting `fid`
  - increment t_1 by Δt_1
  - redo

Then we have a surface $S(t_1, t_2)$ - considered as a big matrix S_{ij}

- process all the `fid` in t_2 :

    - center - apodisation - zerofilling - FT
- process all the column of the data matrix - doing independently real and imaginary parts
    - apodisation - zerofilling - FT
    - combine real and imaginary to compute modulus

specificities of this *t1* modulation

the $S(t_1) \propto \sqrt{1+\cos(\omega_1 t_1)}$ modulation

- is periodic ⇒ FT
- is not pure ⇒ strong harmonics
- the signal along t_1 can be very noisy ⇒ need for a (fast) denoising tool
- because of the frequency generator continuous phase, there is an additional frequency in the t_1 signal which has to be "demodulated"

One additional remark

We have: $S(t_1, t_2) \propto \cos(\omega_2 t_2)\sqrt{1+\cos(\omega_1 t_1)}$

there 2 ways to do a 2D modulation

amplitude modulation : $S(t_1, t_2) \propto \cos(\omega_2 t_2)\cos(\omega_1 t_1)$

- this is the classical modulation in image processing (MRI, Xray, ...)

phase modulation : $S(t_1, t_2) \propto \cos(\omega_2 t_2 + \omega_1 t_1)$

- is specific to spectrometry (2D-NMR - 2D-MS - ...)
- requires special algebra !

Hypercomplex Algebra

you cannot do that in $\mathbb{C}$, we have "only" one phase, one j

We posit a hypercomplex algebra $\mathbb{H}$, a 4 dimensionnal, commutative, non-invertible algebra.

$$z = a + ib + jc + kd \tag{1}$$
$$i^2 = -1 \quad j^2 = -1 \quad k^2 = 1$$
$$ij = ji = k \quad ik = ki = -j \quad jk = kj = -i$$

in $\mathbb{H}$ you can define independent phases: $e^{i\theta}$ and $e^{j\phi}$

and you can write (for a sub class of the element of $\mathbb{H}$) :

$$z = Ae^{i\theta}e^{j\phi} \tag{2}$$

⇒ $|z| = \sqrt{a^2+b^2+c^2+d^2}$

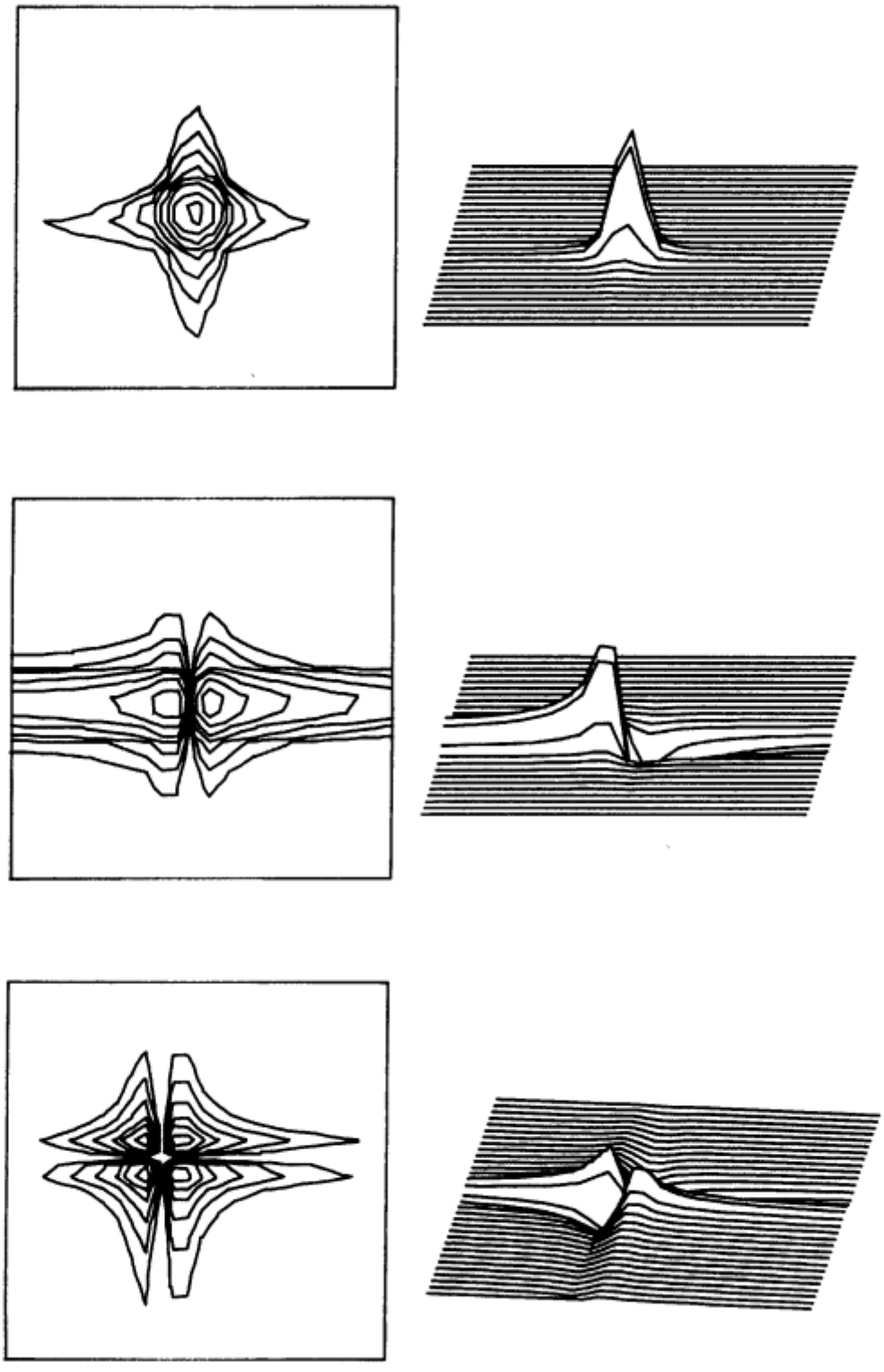
⇒ phases in F1 and in F2 are independent

Absorption mode 2D MS

We will use this this mathematics to phase correct both dimensions and generate 2D peaks in absorption mode in both dimensions

However one thing is simpler

- F2 (horizontal - classical) requires a 2nd order correction
- F1 (vertical - indirect) requires a 1st order correction



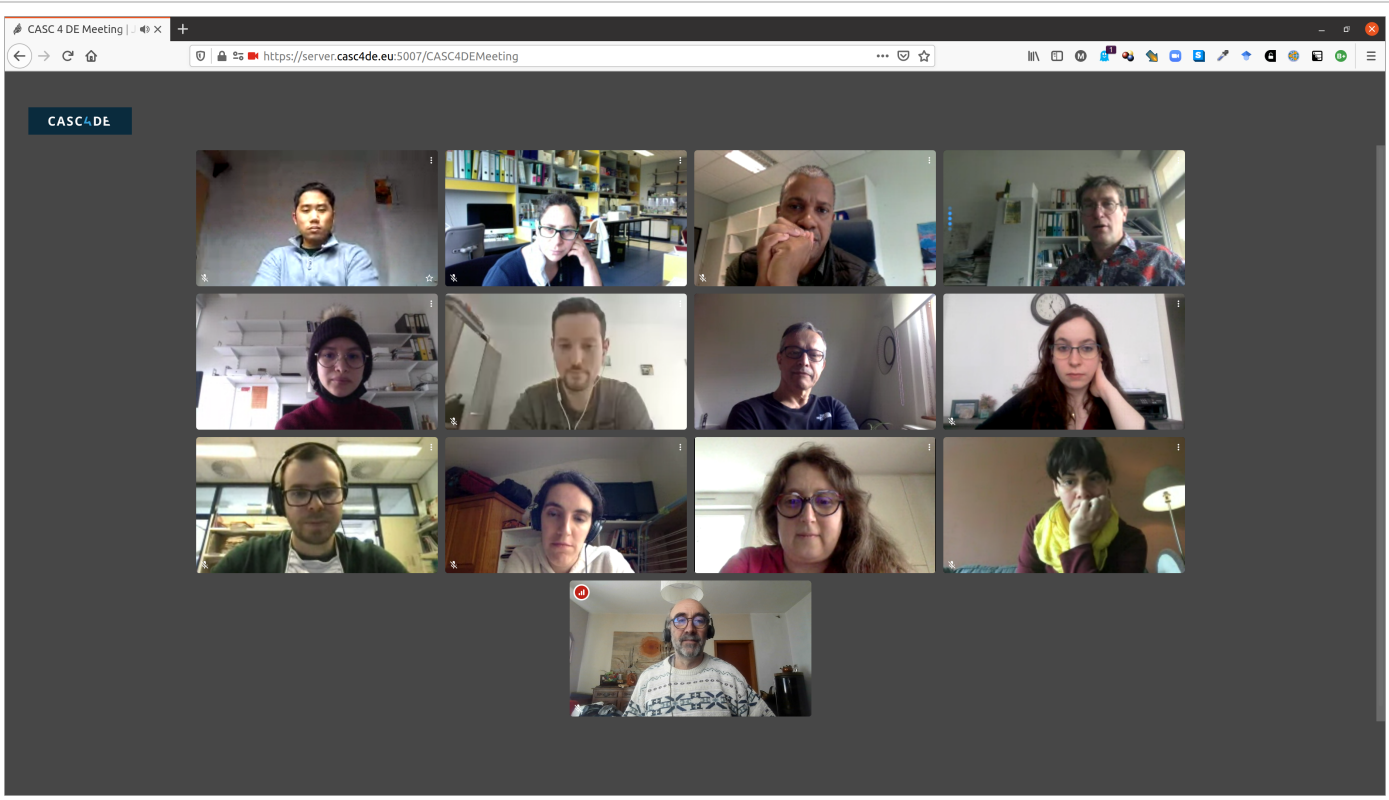
Non-linear Sampling

Many thanks

- Christian Rolando
- Peter O'Connor
- Maria van Agthoven
- Kathrin Breuker

but also

- The CASC4DE team
  - Camille Marin; Lionel Chiron; Laura Duciel; Luis Baptista; Anne Briot Dietsch; Do Manh Dung
- the IGMBC team
  - Bruno Kieffer; Jocelyn Céraline; Celia Deville; Claude Ling; ...



Thank you !