Max-Planck-Institut
für Plasmaphysik

Zerlegung von Massenspektren mit Hilfe der Bayes'schen Datenanalyse

Thomas Schwarz-Selinger,

H.D. Kang, R. Preuss, V. Dose

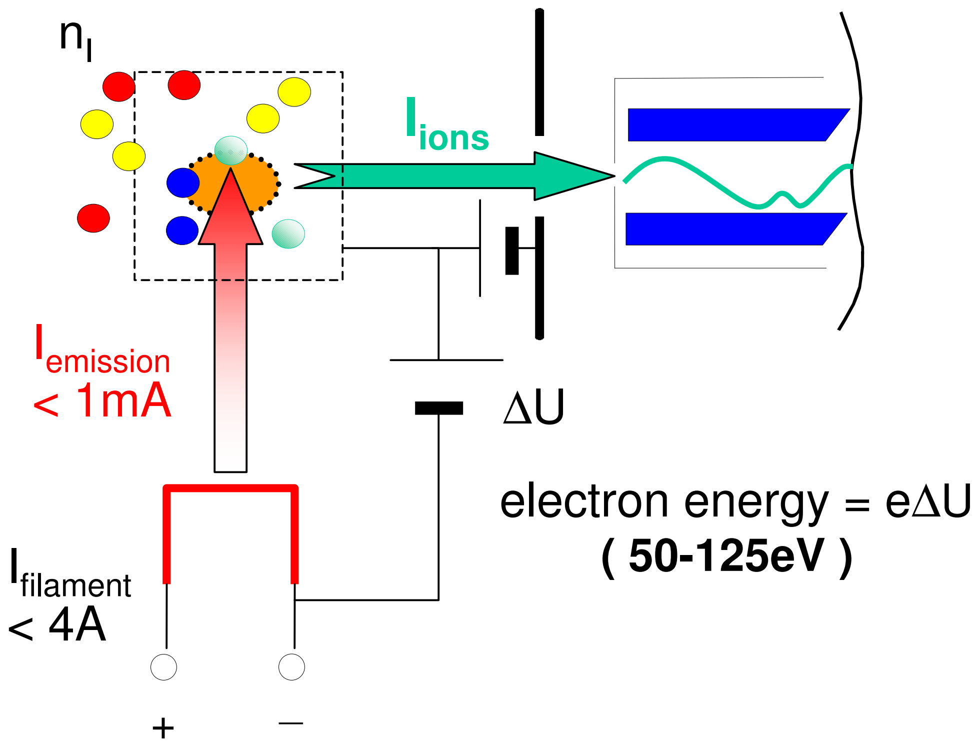
Workshop Oberflächentechnologie mit Plasma- und Ionenstrahlprozessen
Mühlleithen, March 2 nd, 2004

CENTRE FOR INTERDISCIPLINARY PLASMA SCIENCE

detection of neutrals: electron impact ionization first

QMS ionizer:

$A + e^- \text{--->} A^+ + 2e^-$



However:

$AB + e^- \longrightarrow AB^+ + 2e^-$

$A + B^+ + 2e^-$ or $A^+ + B + 2e^-$

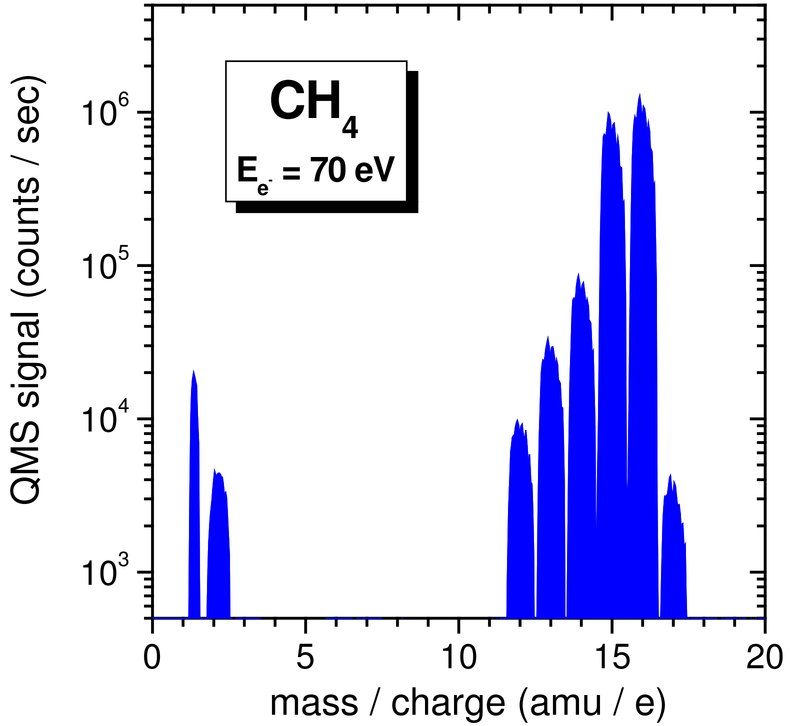
$AB^{2+} + 3e^-$

:

detection of neutrals: electron impact ionization first

QMS ionizer:

$CH_4 + e^- \text{--->} CH_4^+ + 2e^-$



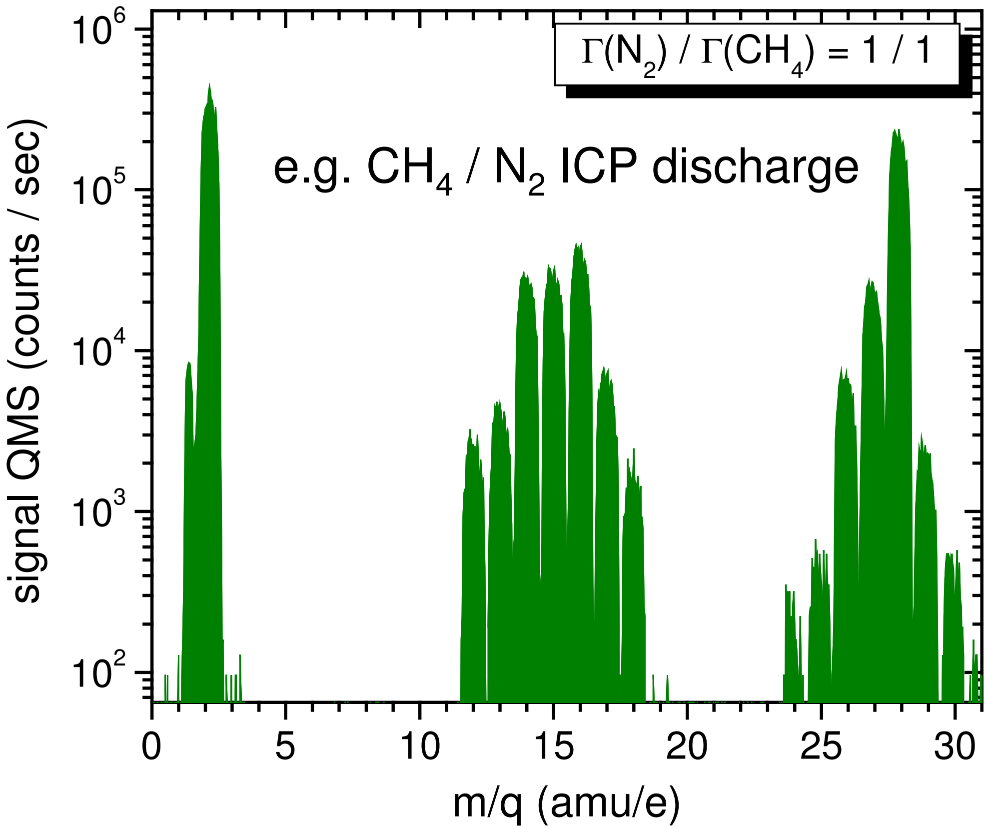
However:

$CH_4 + e^- \text{--->} CH_3^+ + H + 2e^-$, $CH_2^+ + 2H + 2e^-$

$CH^+ + 3H + 2e^-$, $C^+ + 4H + 2e^-$

$H^+ + CH_3 + 2e^-$, ...

What we measure: overlap of different constituents



expected constituents:

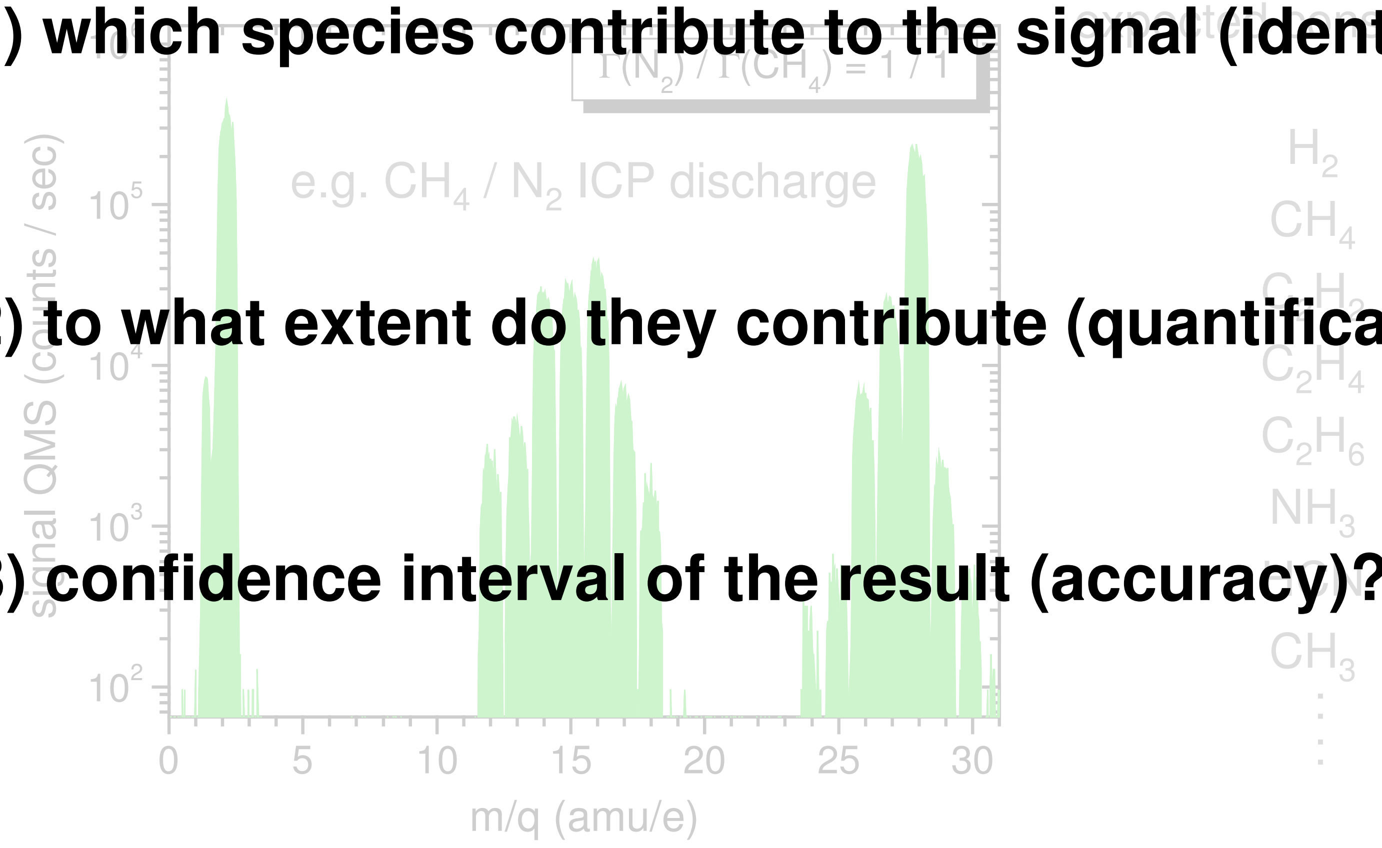
H_2
CH_4
C_2H_2
C_2H_4
C_2H_6
NH_3
HCN
CH_3
⋮

What we want to know:

1) which species contribute to the signal (identification)?

2) to what extent do they contribute (quantification)?

3) confidence interval of the result (accuracy)?



Decomposition of Multicomponent Mass Spectra applying Bayesian Data Analysis

Outline:

- Treating the Inverse Problem of Mass Spectrometry
- Basic Rules of Bayesian Probability Theory
- Nitrogen Containing Methane Plasma
- Summary

inverse problem of quadrupole mass spectrometry

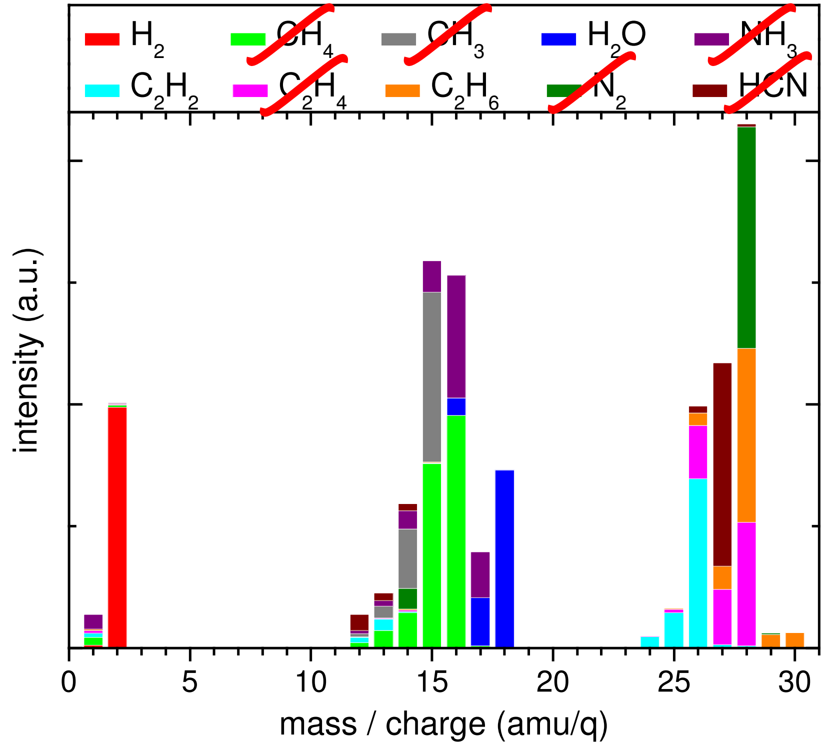
1) successive substraction of constituents signal by signal

⇒ only applicable if there is <u>one</u> channel with <u>no</u> overlap

⇒ exact cracking patterns of constituents are needed

⇒ error propagation

⇒ not applicable for radicals (no cracking pattern, low concentration)



inverse problem of quadrupole mass spectrometry

2) matrix inversion $$\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \varepsilon$$

$$\underline{d} = \begin{pmatrix} d_1 \\ : \\ d_m \end{pmatrix}, \quad \underline{\underline{C}} = \begin{bmatrix} C_{1,1} & \cdot & \cdot & \cdot \\ \cdot & C_{2,2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ C_{m,1} & \cdot & \cdot & C_{m,n} \end{bmatrix}, \quad \underline{x} = \begin{pmatrix} x_1 \\ : \\ x_n \end{pmatrix}, \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ : \\ \varepsilon_n \end{pmatrix}$$

Definitions:	data $\underline{d}$	with *m* elements, representing *m* mass channels
	concentrations $\underline{x}$	with *n* elements, representing *n* species
	cracking matrix $\underline{\underline{C}}$	with *n* columns consisting of *m* elements
	measurement error $\underline{\varepsilon}$	with *m* elements (counting statistics, drifts...)

$\underline{x}$: relative concentrations depending on normalization of cracking matrix

inverse problem of quadrupole mass spectrometry

2) matrix inversion: $\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \varepsilon$

e.g. H_2, CH_4:

$$
\begin{array}{r}
1\,\text{amu} \rightarrow \\ 2\,\text{amu} \rightarrow \\ : \\ 15\,\text{amu} \rightarrow \\ 16\,\text{amu} \rightarrow
\end{array}
\begin{pmatrix} 0.015 \\ 0.691 \\ : \\ 0.12 \\ 0.15 \end{pmatrix}
=
\begin{array}{cc}
\underline{\underline{C}}_{H_2} & \underline{\underline{C}}_{CH_4} \\
\downarrow & \downarrow
\end{array}
\begin{bmatrix} 0.015 & 0.015 \\ 0.985 & 0.004 \\ : & : \\ 0 & 0.4 \\ 0 & 0.5 \end{bmatrix}
\cdot
\begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix}
\begin{array}{l}
\leftarrow x_{H_2} \\ \leftarrow x_{CH_4}
\end{array}
$$

Definitions:

Definitions:	data $\underline{d}$	with *m* elements, representing *m* mass channels
	concentrations $\underline{x}$	with *n* elements, representing *n* species
	cracking matrix $\underline{\underline{C}}$	with *n* columns consisting of *m* elements
	measurement error $\underline{\varepsilon}$	with *m* elements (counting statistics, drifts...)

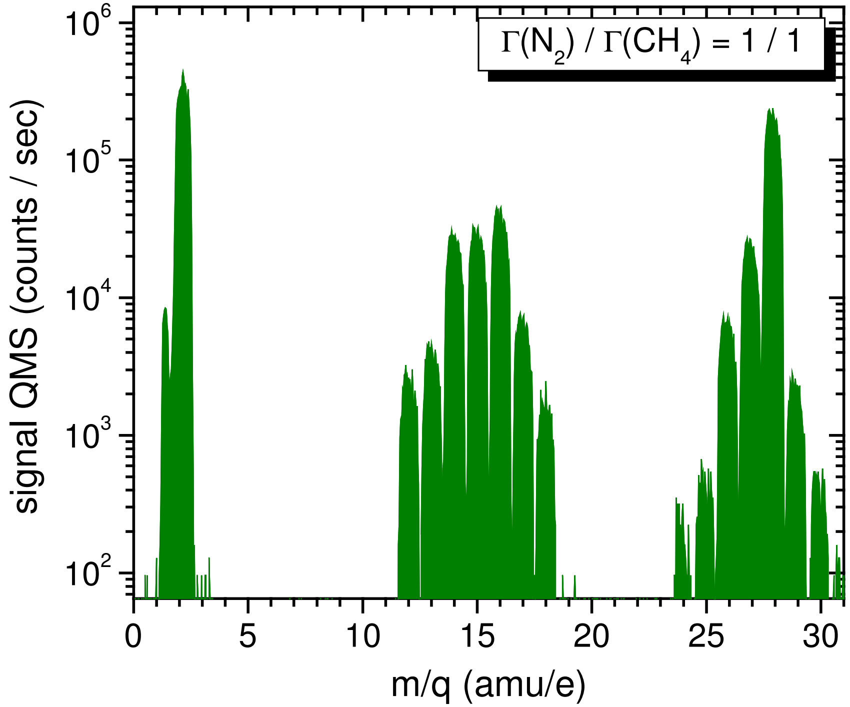
$\underline{x}$: relative concentrations depending on normalization of cracking matrix

2) matrix inversion if: $\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \cancel{\underline{\varepsilon}}$: $\underline{d} = \underline{\underline{C}} \cdot \underline{x}$: $\underline{x} = \underline{\underline{C}}^{-1} \cdot \underline{d}$
(singular value decomposition)

⇒ exact cracking patterns of constituents are needed

⇒ negative concentrations are possible

⇒ not applicable for radicals (no cracking pattern, low concentration)



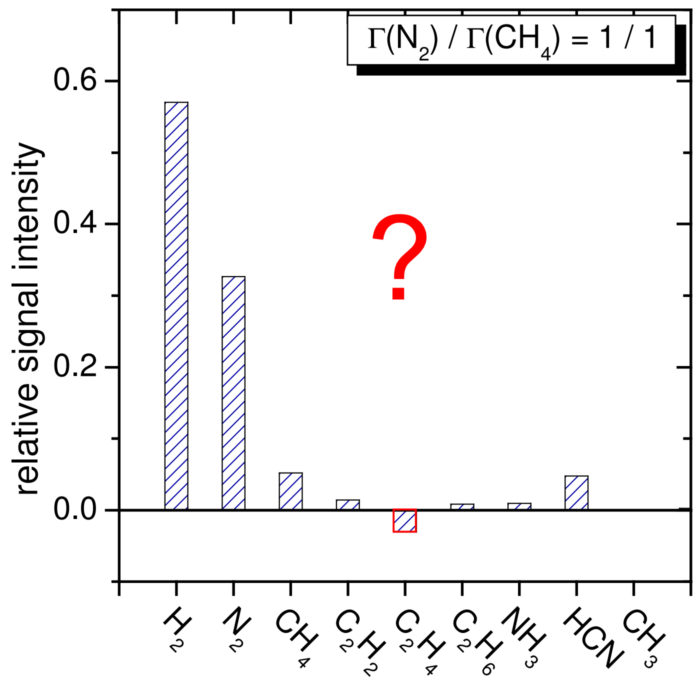
inverse problem of quadrupole mass spectrometry

2) matrix inversion if: $\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \cancel{\varepsilon}$: $\underline{d} = \underline{\underline{C}} \cdot \underline{x}$: $\underline{x} = \underline{\underline{C}}^{-1} \cdot \underline{d}$
(singular value decomposition)

⇒ exact cracking patterns of constituents are needed

⇒ negative concentrations are possible

⇒ not applicable for radicals (no cracking pattern, low concentration)



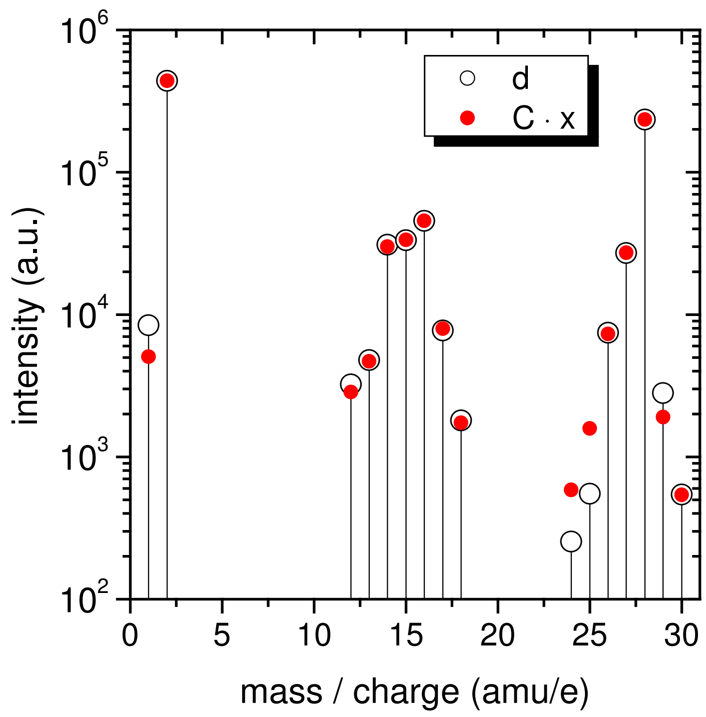
inverse problem of quadrupole mass spectrometry

2) matrix inversion if: $\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \cancel{\varepsilon}$: $\underline{d} = \underline{\underline{C}} \cdot \underline{x}$: $x = \underline{\underline{C}}^{-1} \cdot \underline{d}$
(singular value decomposition)

⇒ exact cracking patterns of constituents are needed

⇒ negative concentrations are possible

⇒ not applicable for radicals (no cracking pattern, low concentration)



3) least square evaluation (χ^2-fits): $\underline{d} = \underline{\underline{C}} \cdot \underline{x} + \cancel{\varepsilon}$:

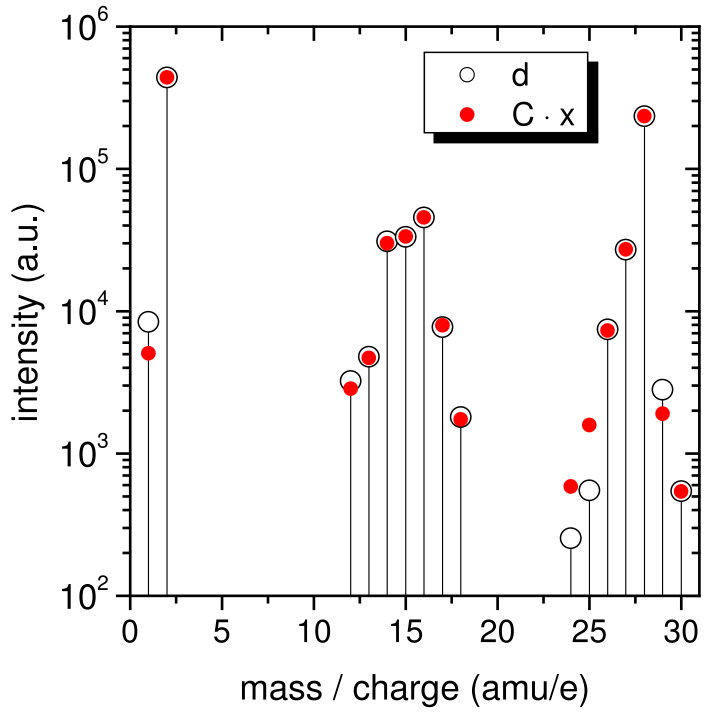
forward calculation: minimize: $\underline{d} - \underline{\underline{C}} \cdot \underline{x}$

$\Rightarrow$ assumes exact measurements

$\Rightarrow$ needs exact cracking patterns of constituents

$\Rightarrow$ best fit $\neq$ best result **fitting noise ?**

$\Rightarrow$ not applicable for radicals
(no cracking pattern, low concentration)



Bayes basics (parameter estimation)

- powerful tool to solve **inverse problems**, incorporating consistently further information
- provides the **most probable result under the current state of knowledge:**

$$< x >= \frac{\int x \cdot p(x)\,dx}{\int p(x)\,dx}$$

(expectation values)

Bayes basics (parameter estimation)

- powerful tool to solve **inverse problems**, incorporating consistently further information

- provides the **most probable result under the current state of knowledge:**

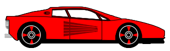
$$< x_k > = \frac{\int x_k \, p(x_k \mid \vec{d}, \sigma_d, I) dx_k}{\int p(x_k \mid \vec{d}, \sigma_d, I) dx_k}$$ (expectation values)

- provides the **confidence interval of the result:**

$$\sigma^2(x_i) = < (x_i - < x_i >)^2 > = < x_i^2 > - < x_i >^2$$ (standard deviation)

Strategy: evaluate $p(x_k \mid \vec{d}, \sigma_d, I)$ with the help of simple rules

Product Rule



$$P(rot,Ferrari|I) = P(rot|I,Ferrari) \cdot P(Ferrari\,|I) = P(Ferrari|I,rot) \cdot P(rot\,|I)$$

Product Rule

$$P(A,B\,|\,I) = P(A\,|\,B,I)\cdot P(B\,|\,I) = P(B\,|\,A,I)\cdot P(A\,|\,I)$$

$\Rightarrow$ **Bayes' theorem**

$$P(A\,|\,B,I) = \frac{P(A\,|\,I)\cdot P(B\,|\,A,I)}{P(B,I)}$$

Bayes basics

A $\rightarrow$ ***H***ypothesis, B $\rightarrow$ ***D***ata

$$P(H \mid D, I) = \frac{P(H \mid I) \cdot P(D \mid H, I)}{P(D \mid I)}$$

tells us how to update our **prior knowledge** $P(H|I)$

about the physical **hypothesis** H in the light of **data** D

which we collected from our experiment

Sum Rule

$$P(A_1 + A_2 \mid I) = P(A_1 \mid I) + P(A_2 \mid I) - P(A_1, A_2 \mid I)$$

$\Rightarrow$ **marginalization**

$$P(A \mid I) = \int P(A, B \mid I)\, dB$$

getting rid of nuisance parameters

Bayes basics

likelihood: (foreward calculation)

$$P(D|H,I)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(D-M)^2}{2\sigma^2}\right)$$

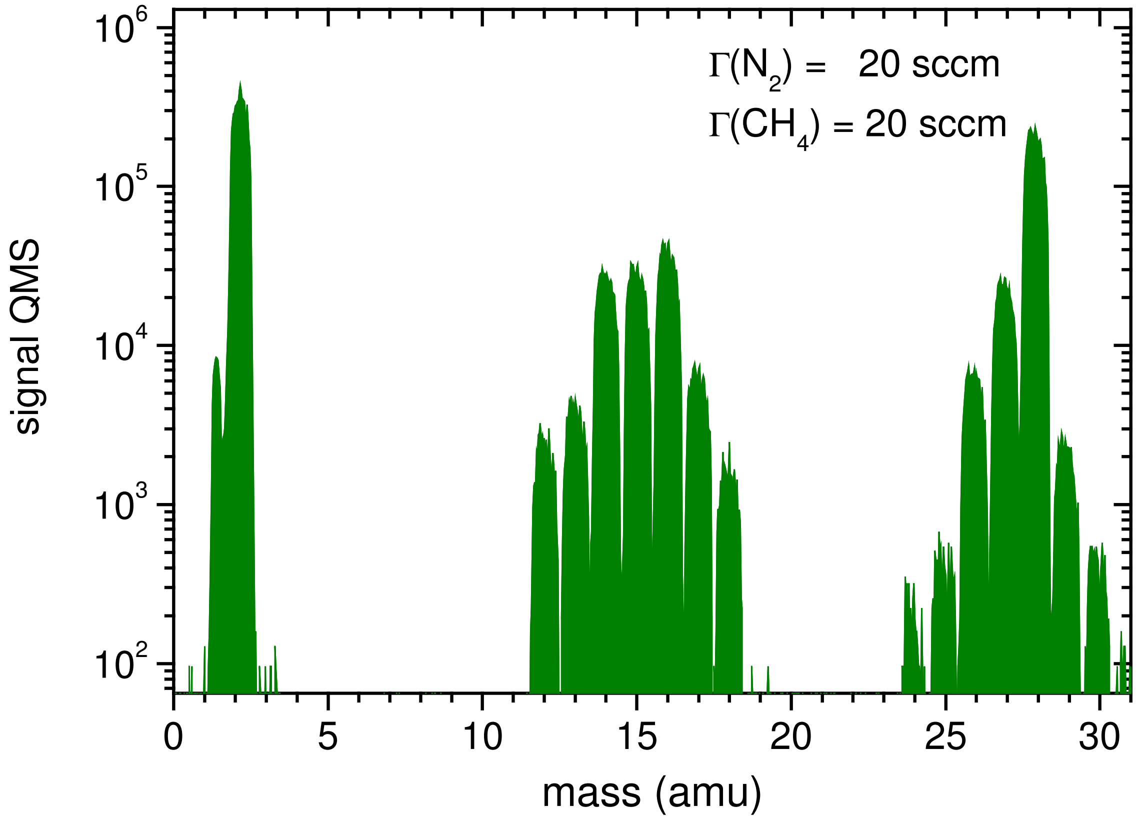
prior: e.g. if you know only a point estimate μ (from maximum Entropy)

$$p(x|\mu,I)=\frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)$$

nitrogen containing methane plasma

final goal: decomposing mass spectra of reactive plasmas

e.g. CH_4 / N_2 ICP discharge



expected constituents:

H_2
CH_4
C_2H_2
C_2H_4
C_2H_6
NH_3
HCN
CH_3
:
:

no calibration available

final goal: decomposing mass spectra of reactive plasmas

input:

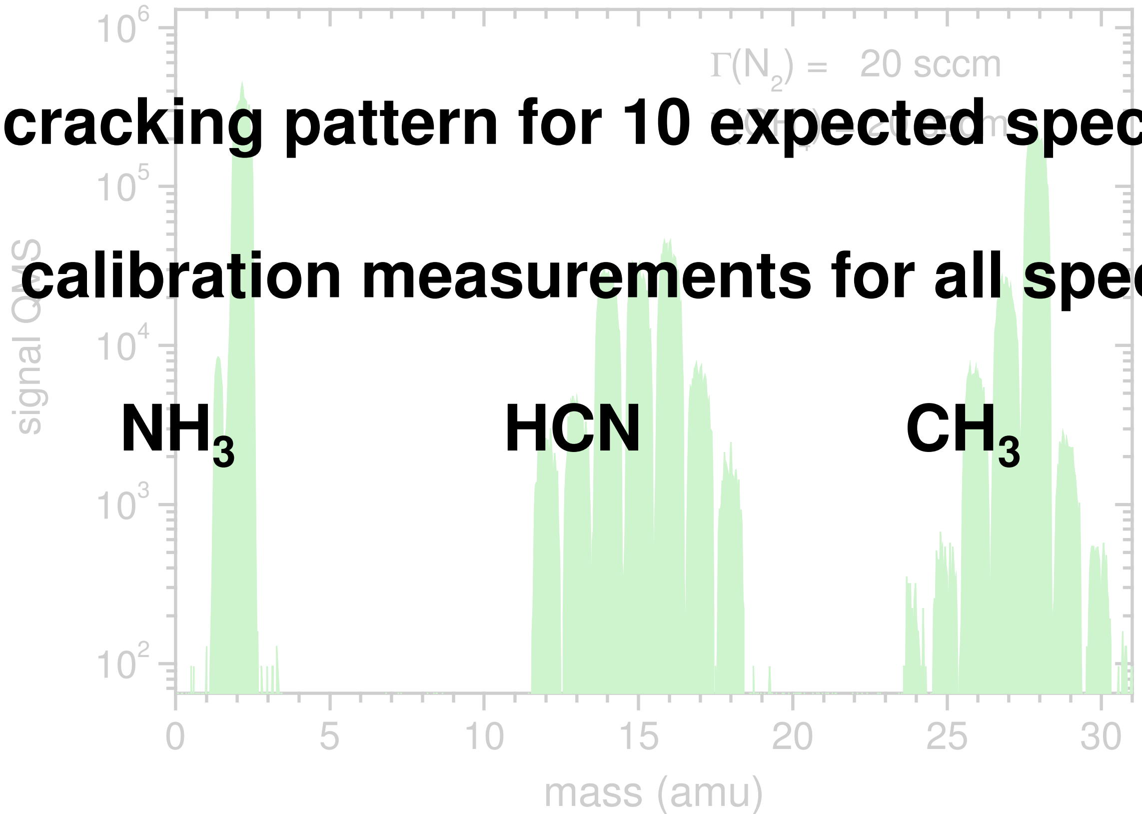
e.g. CH_4 / N_2 ICP discharge

1) intensities of 16 mass channels for 7 mixtures

expected constituents:

2) cracking pattern for 10 expected species (tables)

3) calibration measurements for all species except:

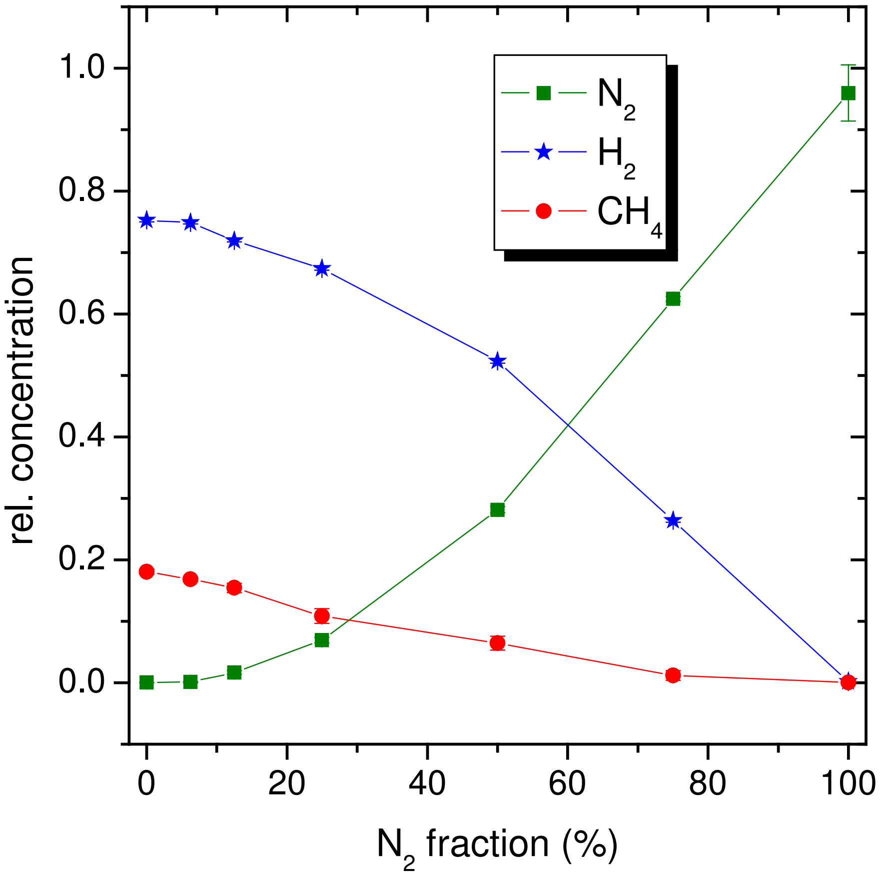


NH_3 **HCN** **CH_3**

H_2
CH_4
C_2H_2
C_2H_4
C_2H_6
NH_3
HCN
CH_3
:

no calibration available

Bayesian analysis



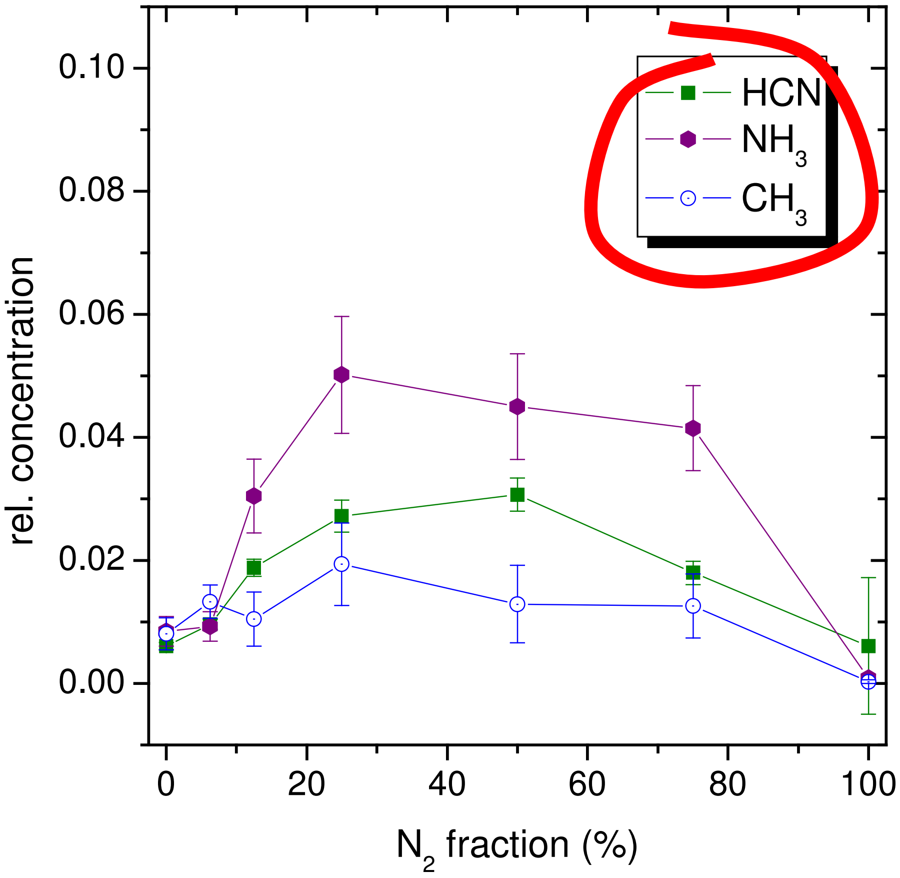
weighting factors:

$H_2 \approx 0.4$

$CH_4 \approx 0.6$

$N_2 \approx 0.6$

nitrogen containing methane plasma



Bayesian analysis

no calibration available!

weighting factors:

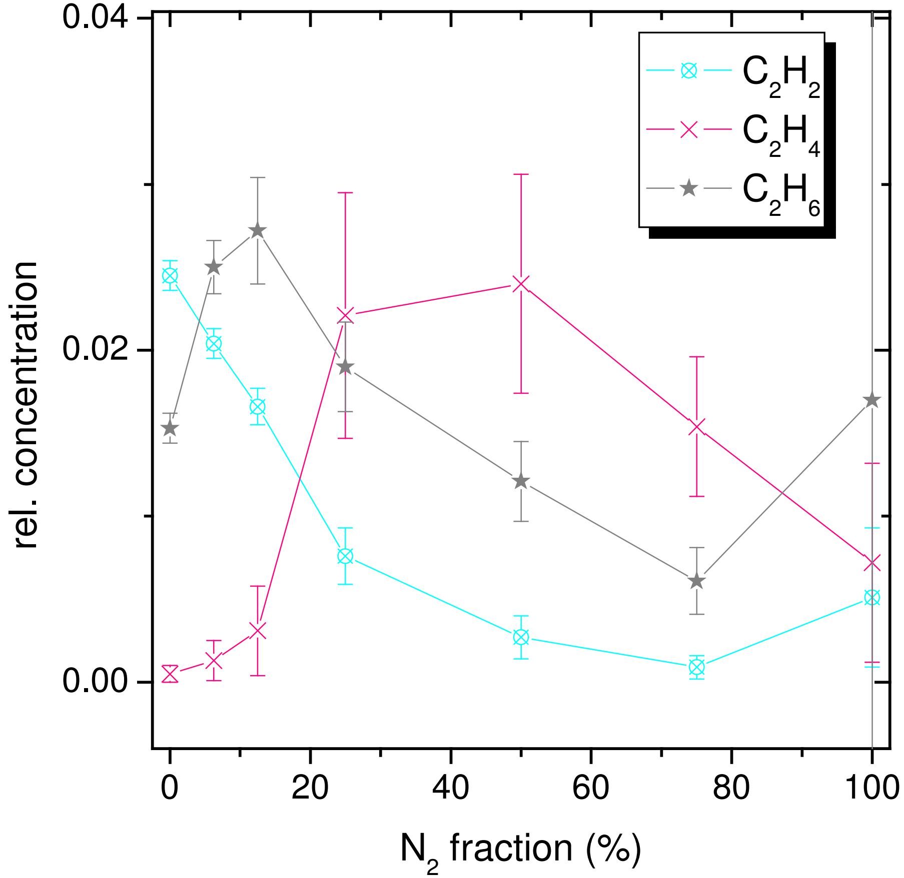
HCN ≈ 0.3

NH_3 ≈ 0.8

CH_3 ≈ 0.6

nitrogen containing methane plasma

Bayesian analysis



weighting factors:

$C_2H_2 \approx 1$

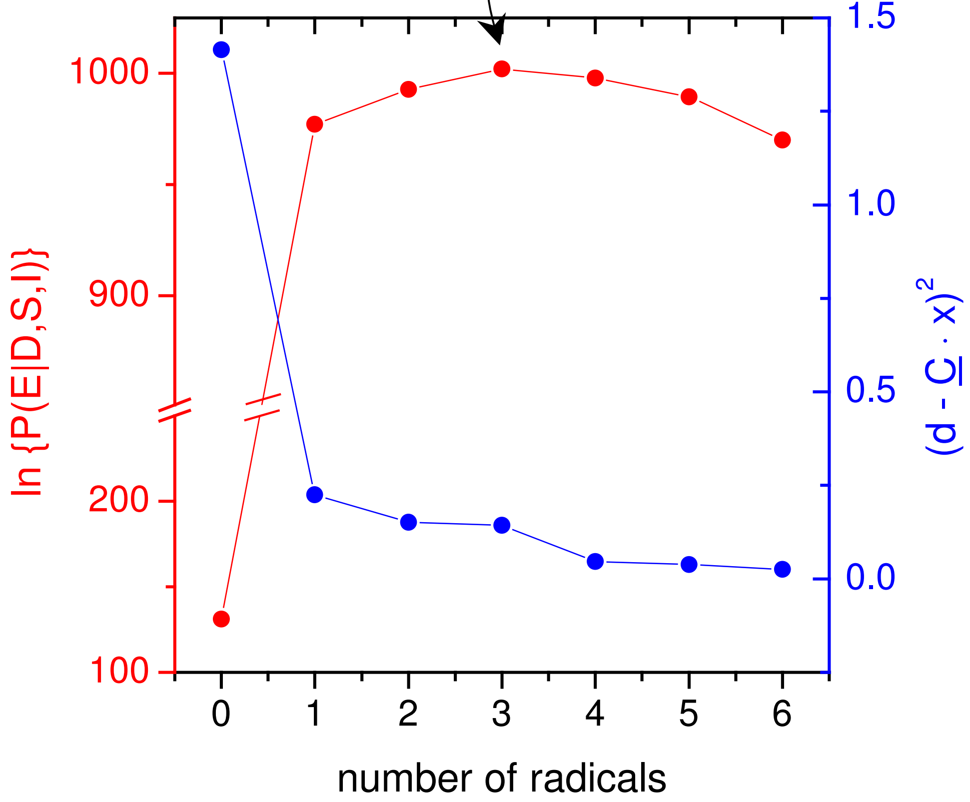
$C_2H_4 \approx 0.6$

$C_2H_6 \approx 0.9$

output:

model comparison with „Occams Razor“

CH_3, C_2H_5, H



conclusion

- Bayesian data analysis allows to decompose multicomponent mass spectra incorporating further information (e.g. calibration measurements)
- parameter estimation delivers the expectation value as well as the confidence interval for the relative signal intensities as well as the cracking patterns
- species with unknown cracking pattern can be included
- model comparison allows to determine the species present in the mixture
- the analysis can help to optimise the experiment

conclusion

- **the method has to fail, when**

  - we under- / overestimate our errors

  - we make wrong assumptions like
  inappropriate calibration measurements
  excluding species

- **the method cannot do wonders**

- **its simply “common sense reduced to calculation”**

general problem of quadrupole mass spectrometry

"a mass spectrometrist is someone, who figures out what something is, by smashing it with a hammer and looking at the pieces"