Automated Peak Identification in a Time-of-Flight Spectrum

Haijian Chen[1], Eugene R. Tracy[1], William E. Cooke[1], Michael W. Trosset[1], Dasha Malyarenko[1,2], Dennis Manos[1], Maciek Sasinowski[2], O. J. Semmes[3]

1. College of William and Mary, Williamsburg, VA
2. INCOGEN Inc., Williamsburg, VA
3. Eastern Virginia Medical School, Norfolk, VA

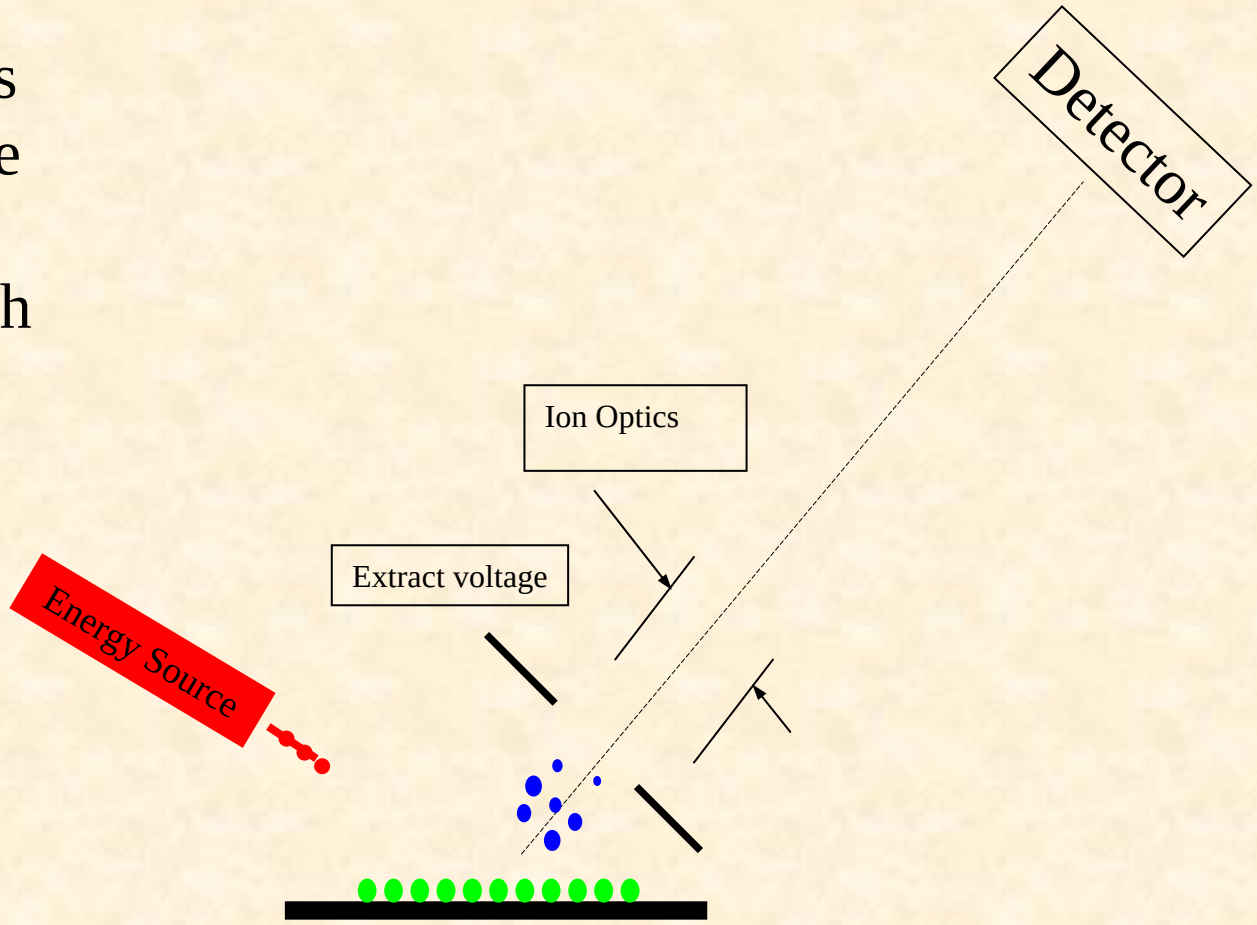
Outline

- **Motivation**
- Methodology with Simulated Data
- Example from Real Data

Motivation

Common use TOF (Time-of-Flight) Mass Spectrometer

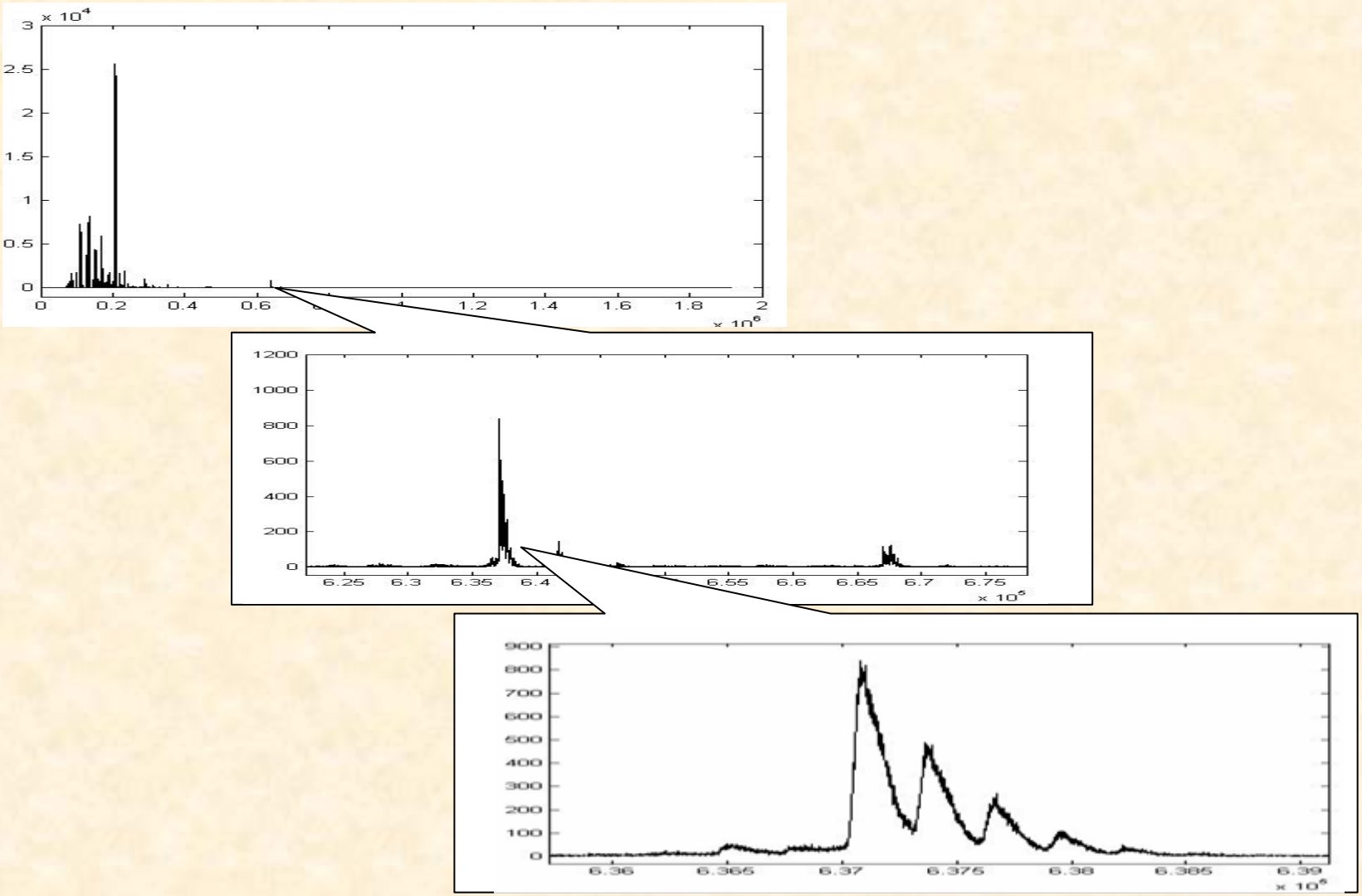
- Surface analysis
- Material science
- Biology
- Clinical research
- ….



MALDI/SELDI
SIMS
ESI
TOF/TOF
…

One thing in common

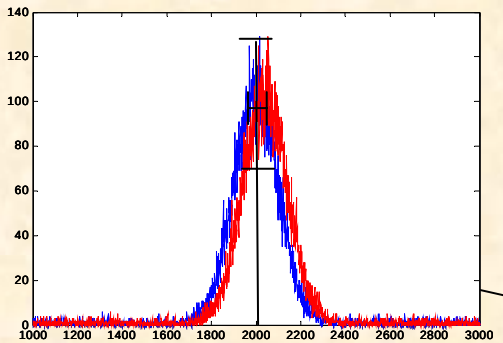
----large number of data points & wealth of information

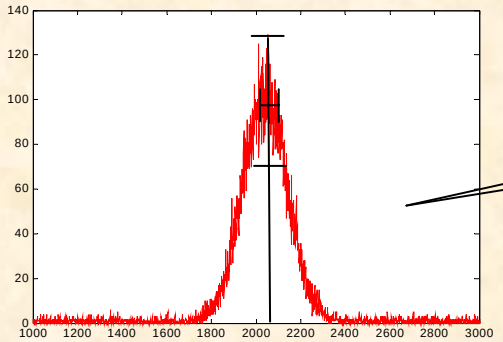


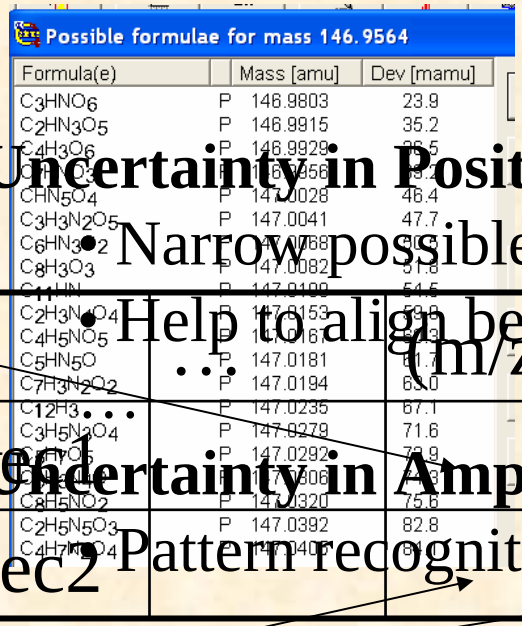
What Information?

A single peak:

- Position and Intensity Accurately
- Uncertainties in Position and Intensity







Uncertainty in Position

- Narrow possible chemical IDs
- Help to align between spectra

Uncertainty in Amplitude

- Pattern recognition

	$(m/z)_1$	$(m/z)_2$	…
spec1			
spec2			
…			

?

In summary, we want:

- Accurate peak position and intensity
- Uncertainties
- Automated

~$800k

+ superficial treatment =

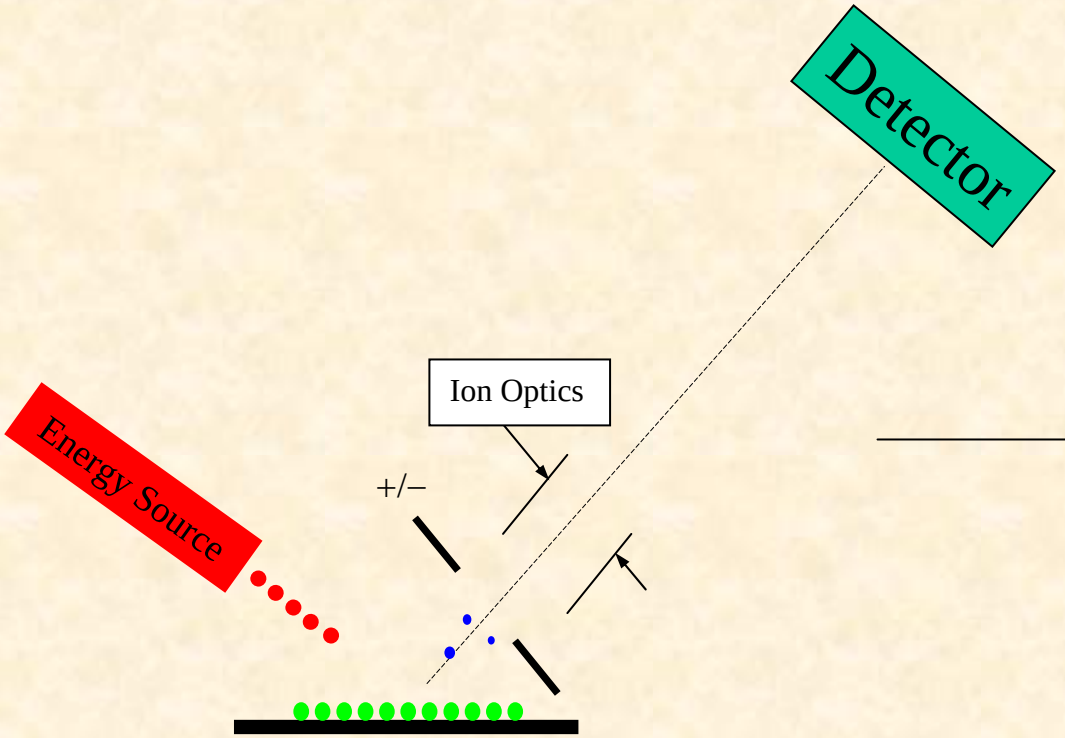
>10^9

<$10

Outline

- Motivation
- Methodology with Simulated Data
- Example from Real Data

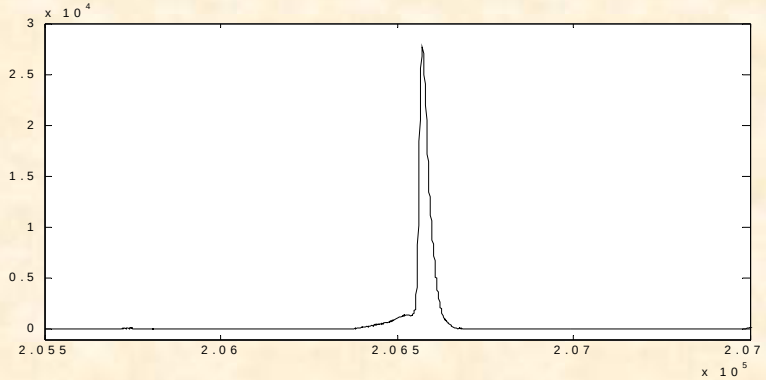
TOF-SIMS: Poisson Noise



0000000000000000001000100000000000100000…
0000000000000000001001000000000000000000…
0000000000000000001000010000000000000000…
0000000000000000001000100000000000000000…
0000000000000000001001000000000000000000…

⋮

0000000000000000001001000000000000000000…



SELDI: Mixed Noise

- Counting
- Poisson noise: $p(n) = \frac{e^r r^n}{n!}$

Maximum Likelihood Method

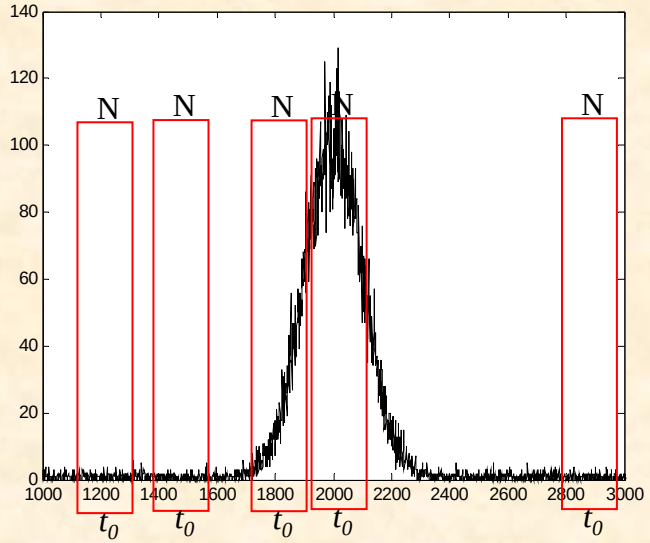
Assumption M_1:

$$r_i = ax_i + \eta_i \qquad i = 1...N$$

x_i : *wavelet shape*

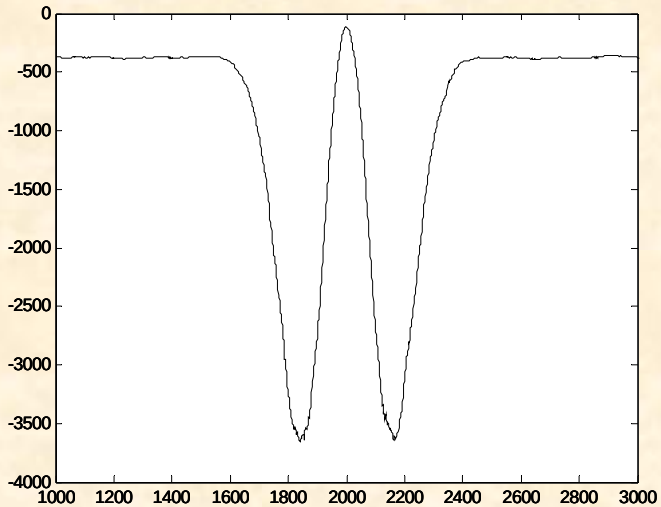
a : *unknown amplitude*

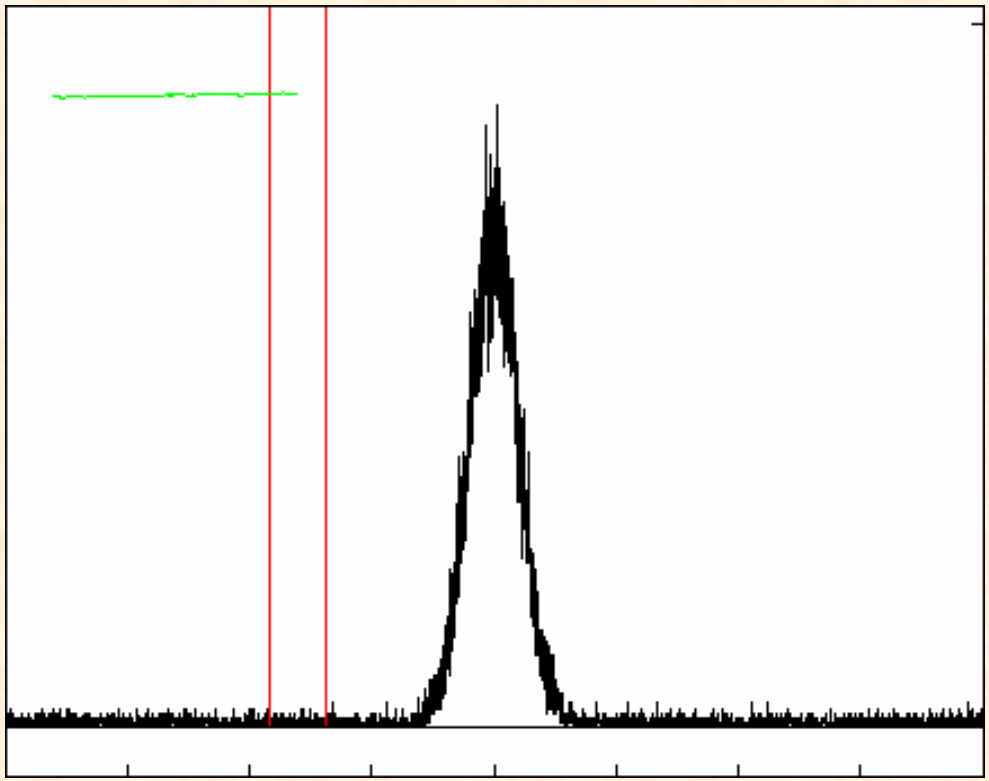
η_i : *noise which can be characterized by parameter* λ



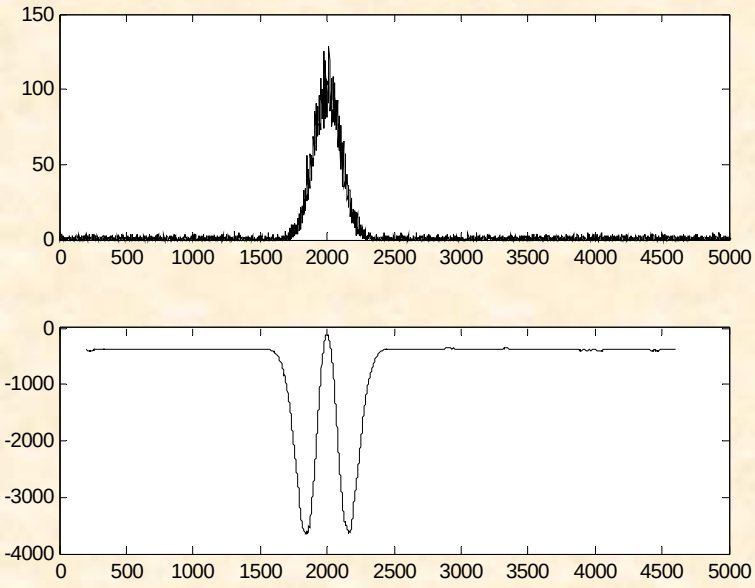
Log of Likelihood Function:

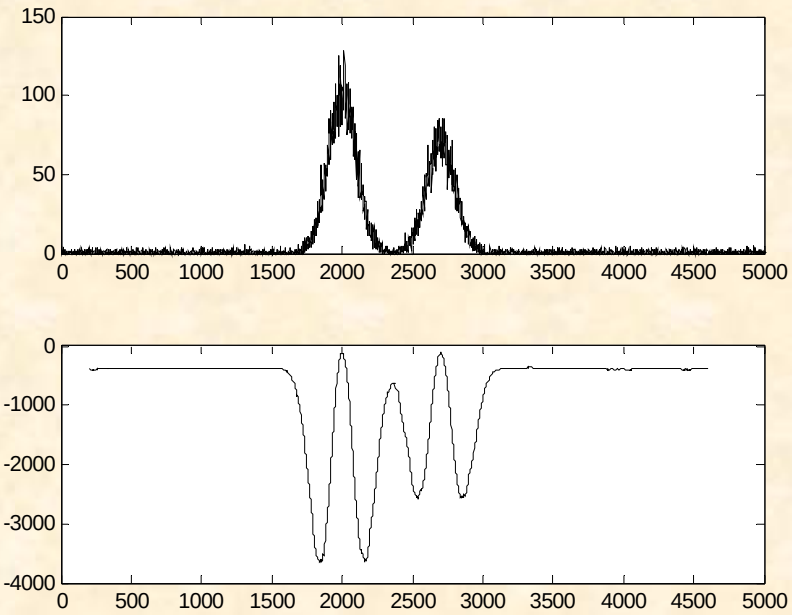
$$\log(p(\{s\}|a,\lambda,t_0,M_1))$$





Where is a peak?



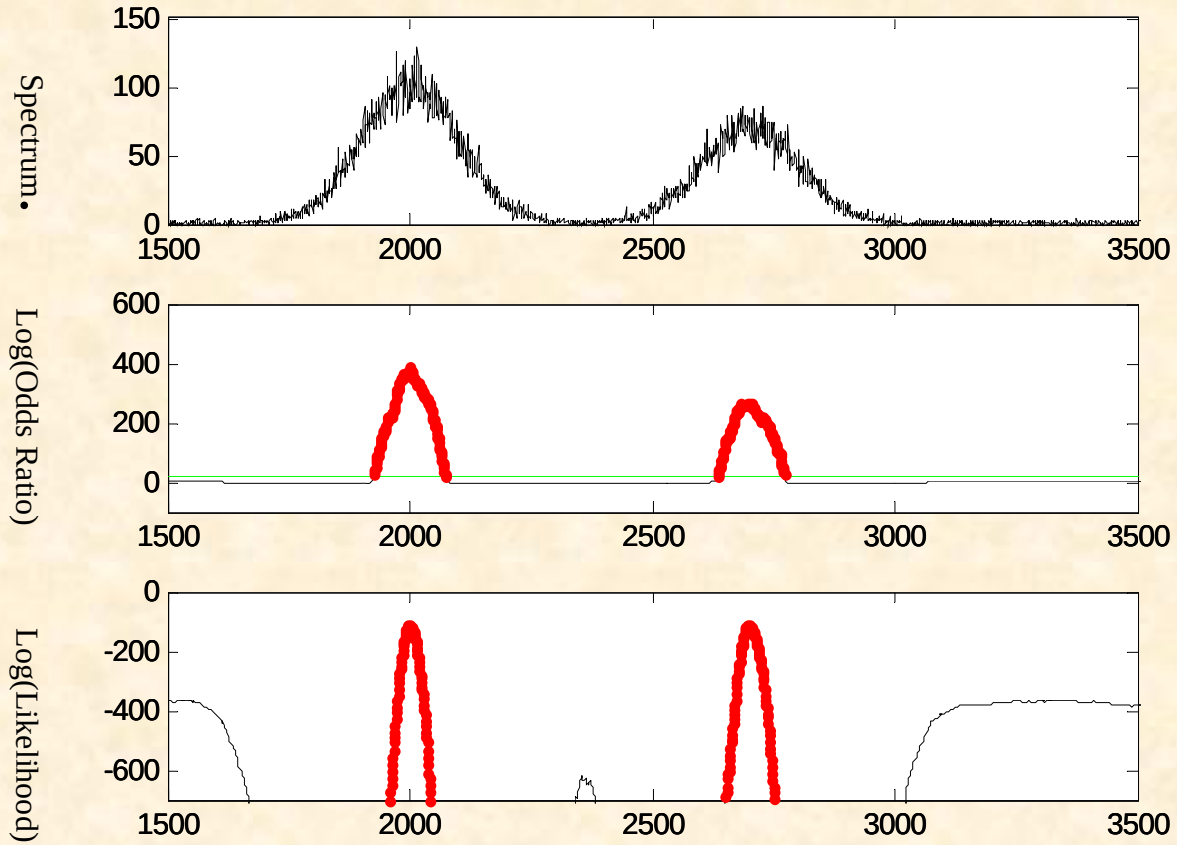


Hypothesis test:

M_1: there is a peak in the window

M_0: there is no peak in the window

$$\Leftrightarrow \quad Odds\ Ratio \equiv \frac{P(M_1 \mid \{data\}, t_0)}{P(M_0 \mid \{data\}, t_0)}$$



$\sigma_0 = 99.$

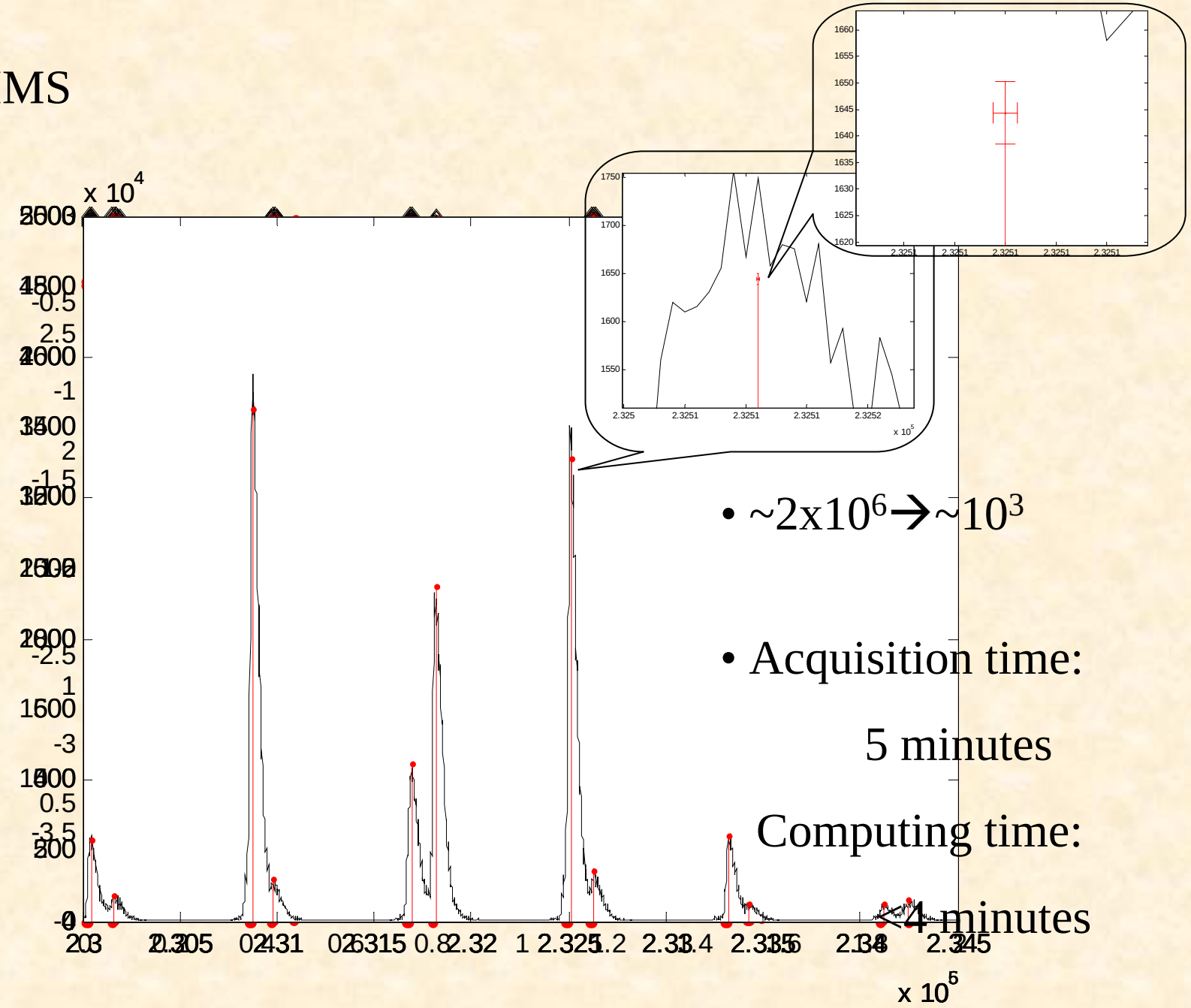
$\sigma_1 = 3.2$

$\sigma_0 = 1.5$

Outline

- Motivation
- Methodology with Simulated Data
- Examples with Real Data

TOF-SIMS



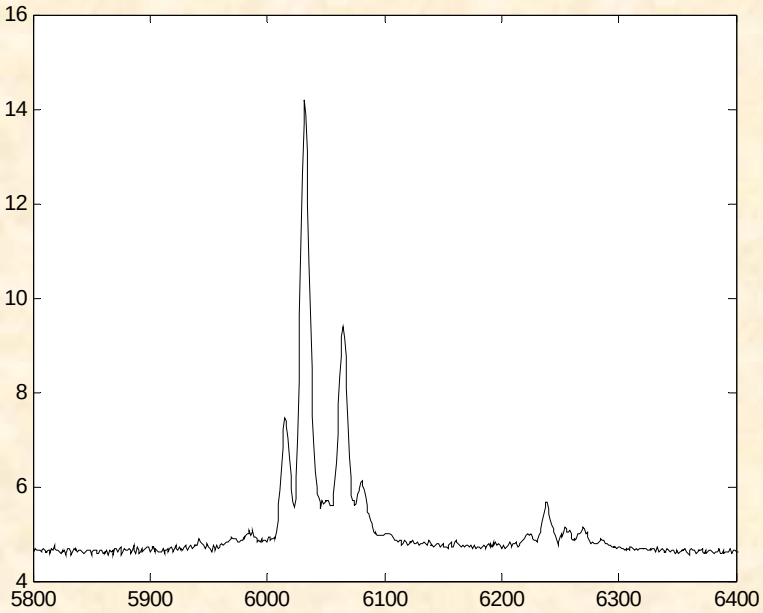
- ~$2x10^6$→~10^3
- Acquisition time:

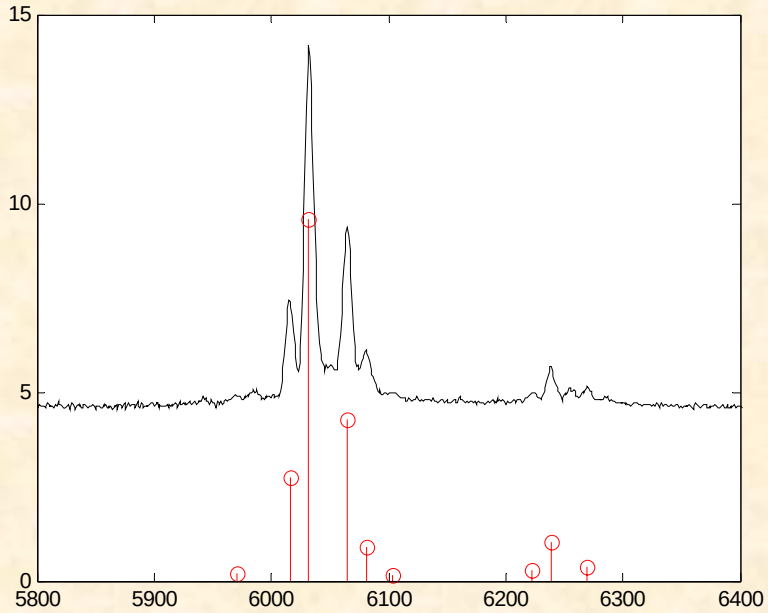
  5 minutes

  Computing time:

  <4 minutes

SELDI


16
14
12
10
8
6
4
5800
5900
6000
6100
6200
6300
6400


15
10
5
0
5800
5900
6000
6100
6200
6300
6400

Summary

An automated peak picking algorithm that:

1. Detects peaks more accurately, approaching instrument resolution
2. Automatically estimates uncertainties
3. Includes appropriate noise models