**Resource-Constrained Analysis of Spectrometry Data**

Dominik Kopczynski*  
Sven Rahmann†

*Bioinformatics for High-Throughput Technologies, Algorithm Engineering, Computer Science XI, TU Dortmund, Germany  
dominik.kopczynski@tu-dortmund.de

† Genome Informatics, Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen, Germany  
sven.rahmann@uk-essen.de

- Ion mobility spectrometry (IMS) device coupled with a multi-capillary column (MCC)
- Depending on resolution amount of data points from 3 to 75 million
- Visualisation as heatmap (X-axis: drift-time $d$ in ms; Y-axis: retention-time $r$ in s)
- Regions with high signal intensity called peaks

**Measurement**

- Modeling 1D-IMS spectra as weighted sum of inverse gaussian distributions
  
  \[
  f_G(x; \mu, \lambda, o) = \left( \frac{\lambda}{2\pi(\lambda^2 - \mu^2)} \right)^{1/2} \exp \left( -\frac{\lambda((x - o) - \mu)^2}{2\mu^2(x - o)} \right)
  \]

- Use EM algorithm for estimation of all parameters for all models
- After processing one spectrum connect models to coherent models in previous spectrum
- Use EM algorithm again to estimate parameters for 2D peak model with already processed 1D model chains
- Seven parameters per 2D model

\[
M(r, d, r, \lambda, \mu, r, \lambda, o) = \omega \cdot f_G(r, \mu, \lambda, o) \cdot f_G(d, \mu, \lambda, o)
\]

- Data reduction factor ranges from 10,000 to 250,000.

**Feature Selection**

- Exemplary 2D peak models:

**Exemplary 1D-IMS spectrum:**

**Peak Alignment:**

- Wastage parts, temperature or carrier-/drift gas flow lead to displaced peaks in measurement
- MCC/IMS devices not exactly calibrated
- Necessary for correct peak comparison

**Registrate Metabolites:**

- Completing IMS metabolite database
- Use gas chromatography (GC) data to identify metabolites in IMS measurement with normalized data (alignment)

**Further Work**

- Completing IMS metabolite database
- Use gas chromatography (GC) data to identify metabolites in IMS measurement with normalized data (alignment)

**Prediction:**

- Metabolite calibration very costly in terms of time
- Simulation of MCC/IMS device with unregistrated compounds just by considering chemical structures