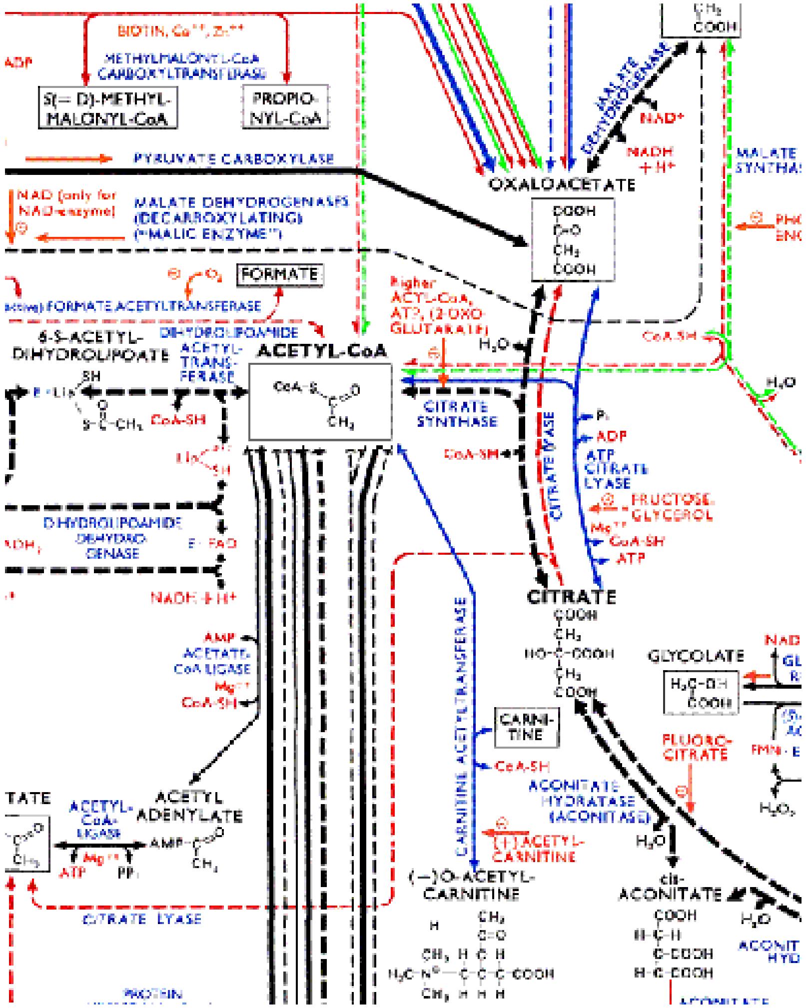
A method for estimating metabolic fluxes from incomplete isotopomer information

Juho Rousu, Ari Rantanen, Esa Pitkänen, Katja Saarela, Esko Ukkonen
Department of Computer Science, University of Helsinki

Hannu Maaheimo
VTT Biotechnology

Challenge: reverse engineering metabolic networks

- Metabolic networks are abstractions of the reaction pathways existing inside each living cell
- Big questions:
  - Which pathways are there?
  - Which pathways are active in given conditions?
  - How is the activity regulated?



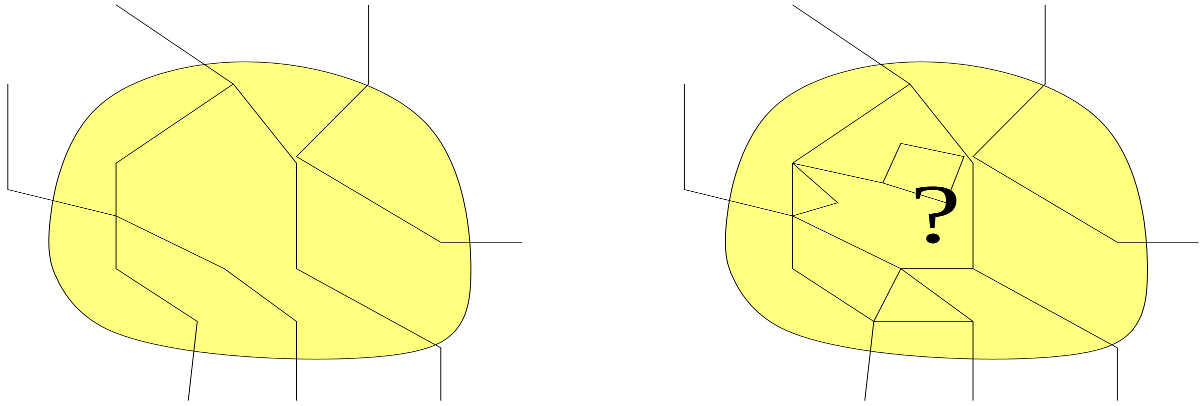
Metabolic flux estimation

Publications:

- A Method for Estimating Metabolic Fluxes from Incomplete Isotopomer Information. Computational Methods in Systems Biology, CMSB'2003, *LNCS* 2602 (2003), 88–103.
- Flux estimation using incomplete isotopomer information. Report C-2002-55, Department of Computer Science, Univ. of Helsinki, 2002.

Flux estimation in metabolic networks

- Given measurement values on (isotopomer) distributions of some metabolites, **infer the steady-state velocities** v_i of the metabolic reactions, the **fluxes**; Marx at al (96), Schmidt et al (97), Stephanopoulos at al (98).
- Kind of an inverse maximum flow problem: given a (maximum) flow, infer the capacities of the edges.
- For **tree-shaped topologies** solution is **trivial**, for **general graphs** (like metabolic networks) **not so**: alternative paths through the network difficult to distinguish.

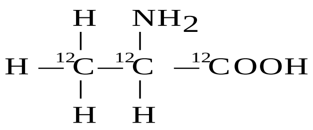


Isotope tracing experiments

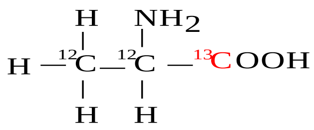
In isotope tracing experiments the cell culture is fed a mixture of natural and ^{13}C substrate (e.g. 90%/10%).

This induces different kinds of ^{13}C labeling patterns, **isotopomers**:

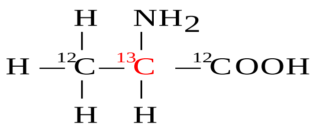
^{000}**Ala**



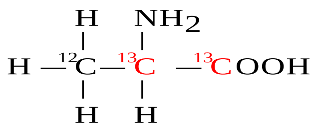
^{001}**Ala**



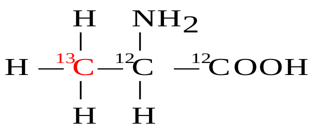
^{010}**Ala**



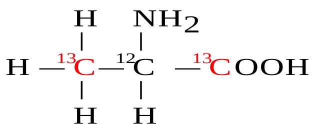
^{011}**Ala**



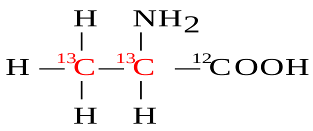
^{100}**Ala**



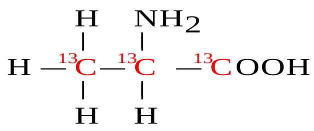
^{101}**Ala**



^{110}**Ala**

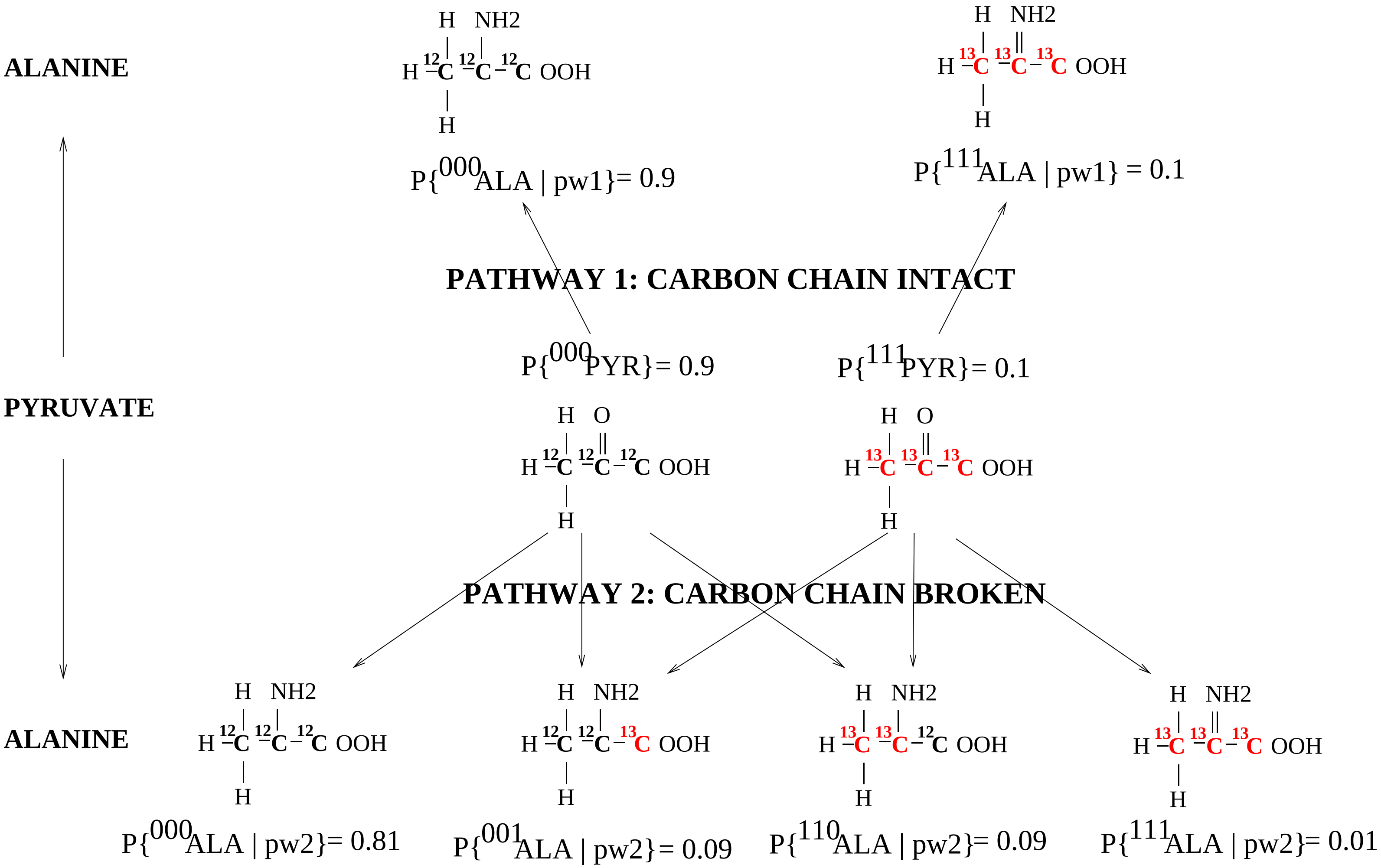


^{111}**Ala**



The vector of relative frequencies of the isotopomers $\mathbb{I}_{Ala} = [\mathbb{P}\{^{000}Ala\}, \mathbb{P}\{^{001}Ala\}, \ldots, \mathbb{P}\{^{111}Ala\}] \in [0,1]^{2^3}$, is called an **isotopomer distribution** of alanine.

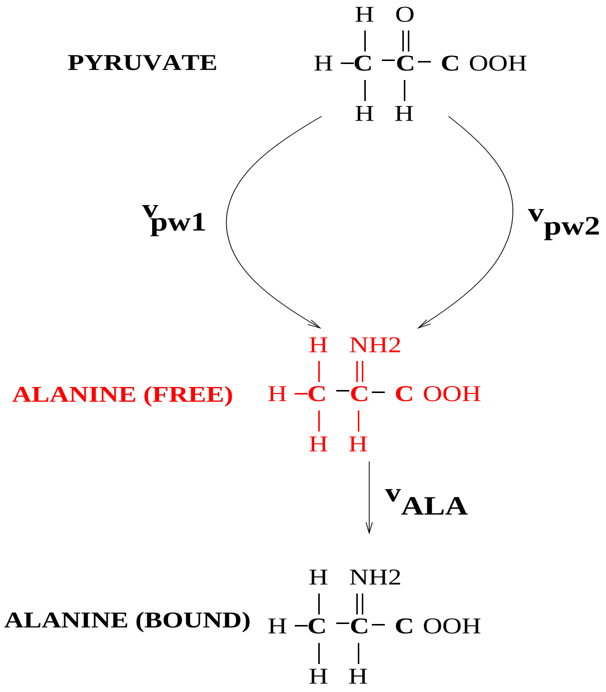
Isotopomer distributions & alternative pathways



Isotopomeric balance equations

- In a steady state, Kirkhoff's law applies: **flow in must match the flow out** of each metabolite:

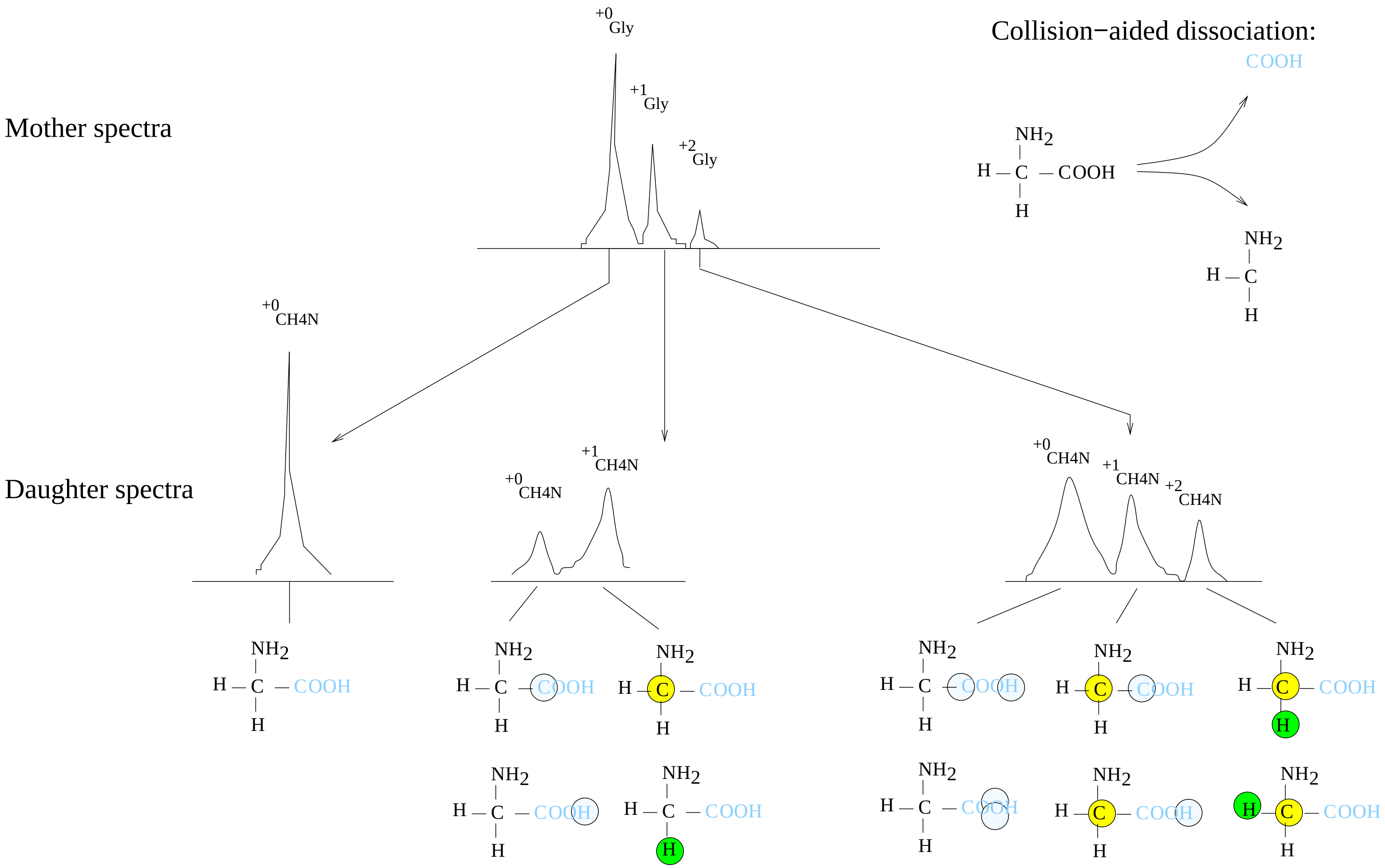
$$v_{pw1} + v_{pw2} = v_{ALA}$$



- If the isotopomer frequencies are also known, additional balance equations in the fluxes v can be written (for **junction** metabolites):

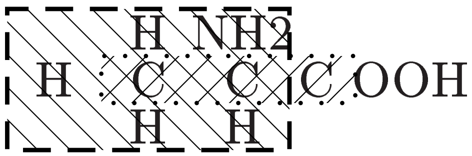
$$P(^{000}ALA|pw1) \cdot v_{pw1} + P(^{000}ALA|pw2) \cdot v_{pw2} = P(^{000}ALA) \cdot v_{ALA}$$
$$P(^{001}ALA|pw1) \cdot v_{pw1} + P(^{001}ALA|pw2) \cdot v_{pw2} = P(^{001}ALA) \cdot v_{ALA}$$
$$P(^{110}ALA|pw1) \cdot v_{pw1} + P(^{110}ALA|pw2) \cdot v_{pw2} = P(^{110}ALA) \cdot v_{ALA}$$
$$P(^{111}ALA|pw1) \cdot v_{pw1} + P(^{111}ALA|pw2) \cdot v_{pw2} = P(^{111}ALA) \cdot v_{ALA}$$

Glycine isotopomers via tandem mass spectrometry



Correction for natural abundance

Consider distributing extra neutrons among the carbons and non-carbons, within and outside the fragment



	C	$\overline{C}$	
F	**+i**	**+k-i**	**+k**
$\overline{F}$	**+j**	**+l-k-j**	**+l-k**
	+i+j	**+l-i-j**	**+l**

We get a linear equation tying the relative height daughter and mother peaks to the carbon isotopomer frequencies:

$$P(^{+k}F|^{+l}Ala) \cdot P(^{+l}Ala) =$$

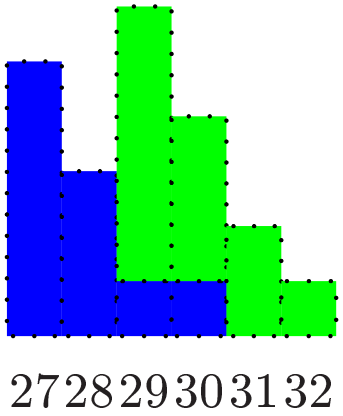
$$= \sum_{i=0}^{k} \sum_{j=0}^{l-k} P(^{+k-i}\{\overline{C} \cap F\} \cap {}^{+l-k-j}\{\overline{C} \cap \overline{F}\}) \cdot P(^{+i}\{C \cap F\} \cap {}^{+j}\{C \cap \overline{F}\})$$

Coefficients are computed from tables of natural isotope abundance .

Overlapping daughter spectra

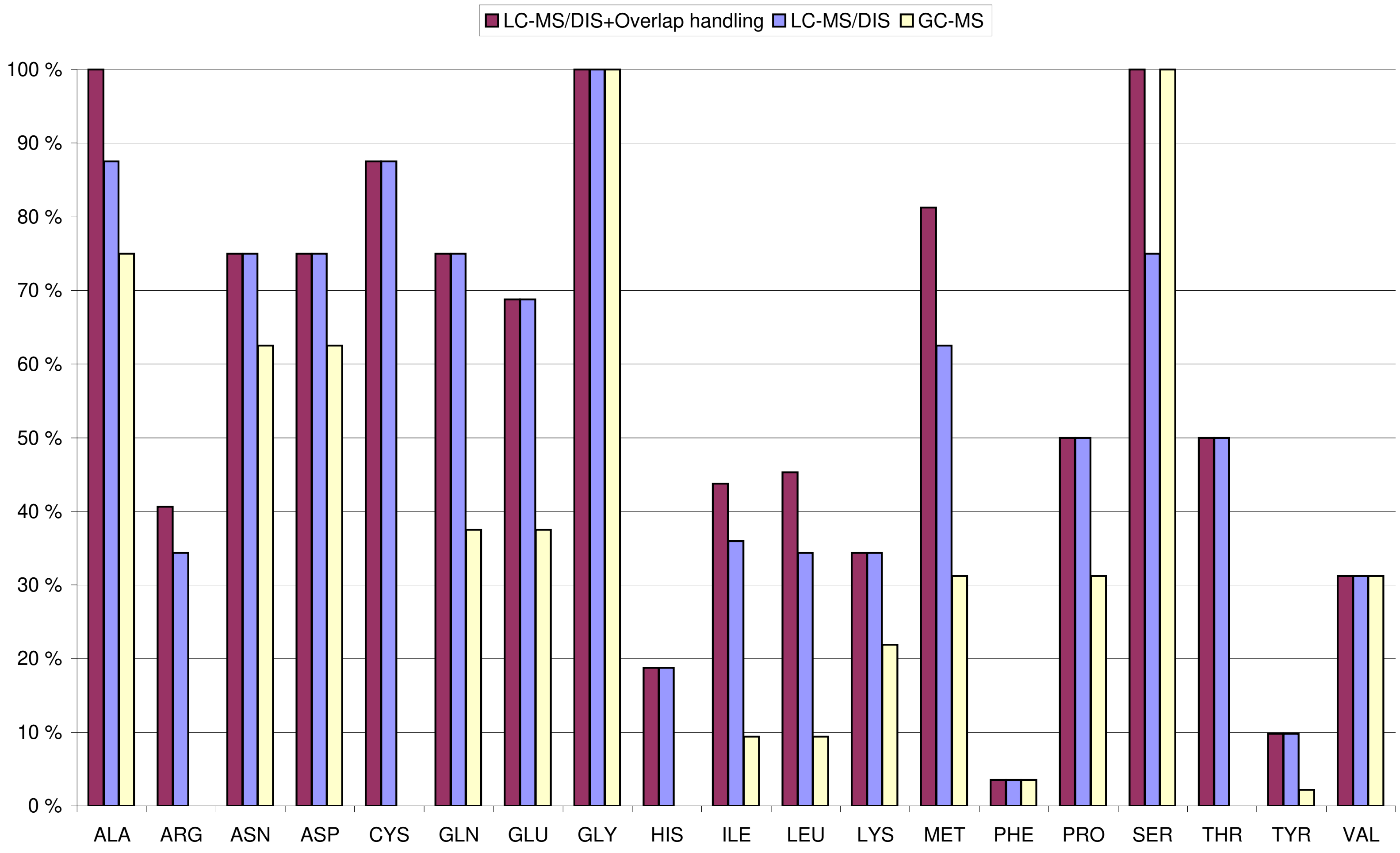
In the daughter spectra all fragments appear at the same time.

If they have similar weight, we may have overlaps (e.g. C_2H_3 (m/z=27) and CH_3N (m/z=29) fragments of Alanine)



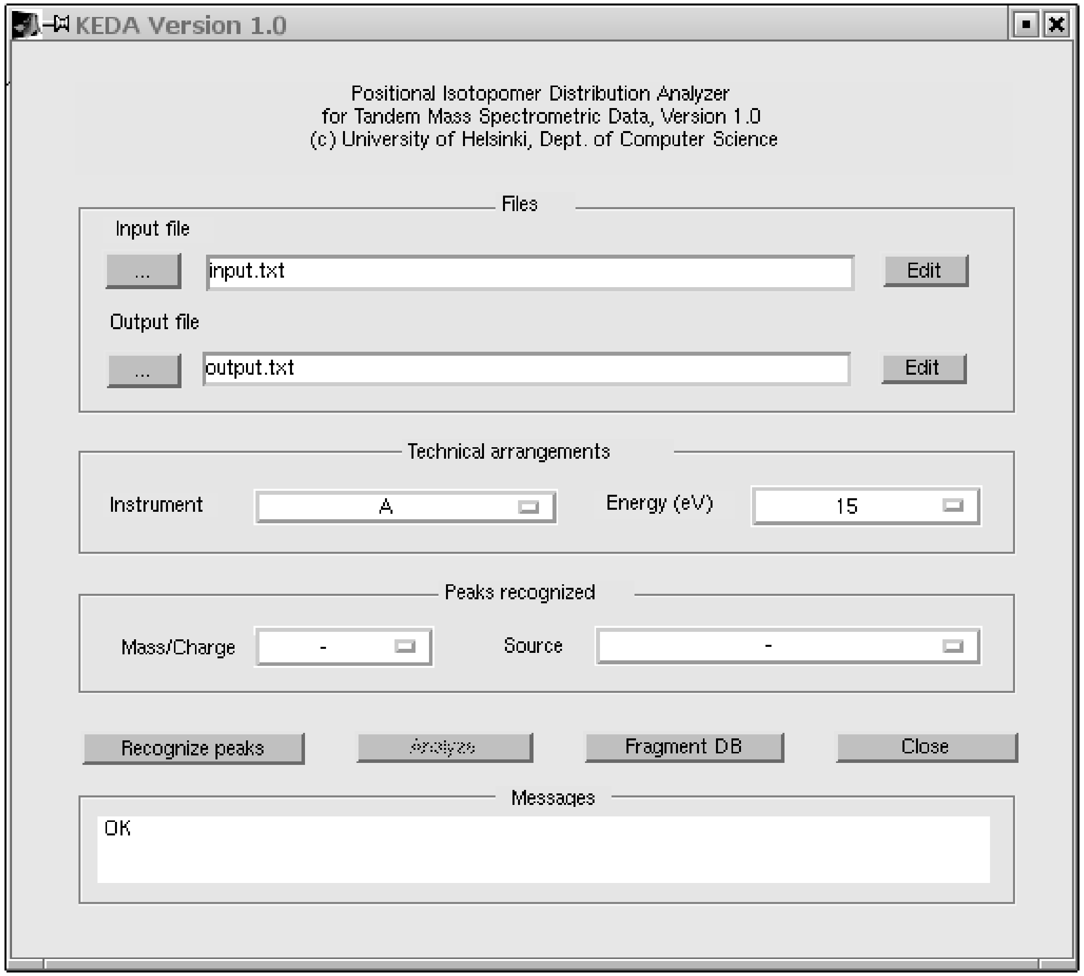
Our method can handle the overlaps assuming that the relative abundance of different fragments (colored area) is known a priori.

Identifiability of isotopomer distributions



PIDC software

- Computes isotopomer distributions from MS and tandem MS data
- Derivatized molecules can be handled as well
- A Matlab applet, runs under both Linux and Windows
- Recognizes peaks automatically based on a database of molecule information.



Flux estimation from incomplete isotopomer data

In practice, we are faced with incomplete isotopomer data:

- MS data: our PIDC software generally groups some isotopomers together so we get data like

$$aP(^{010}ALA) + bP(^{100}ALA) = d$$

- NMR measurements require ^{13}C in a specific position e.g. the middle carbon in alanine

$$\frac{P(^{010}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d.$$

- In addition, some intermediates are present in a too low concentration to be quantified at all.

Our flux estimation method should account for this incompleteness.

Flux estimation from incomplete isotopomer data

Our flux estimation method consists of three main steps

- Propagate the measured isotopomer distributions (using carbon maps) to constrain also some unmeasured isotopomer distributions
- For each metabolite that is produced by two or more reactions (the **junction metabolites**), add corresponding block of isotopomer balance equations to the original stochiometric flux balances determined by the topology of the network
- Solve the resulting equation system to obtain constraints to the fluxes

Isotopomer measurements as linear constraints

Complete isotopomer distribution, NMR and mass spectrometric data, and the **absence** of isotopomer information all can be expressed as a set of linear constraints to the isotopomer distribution.

So we model the measurements as sets of equations

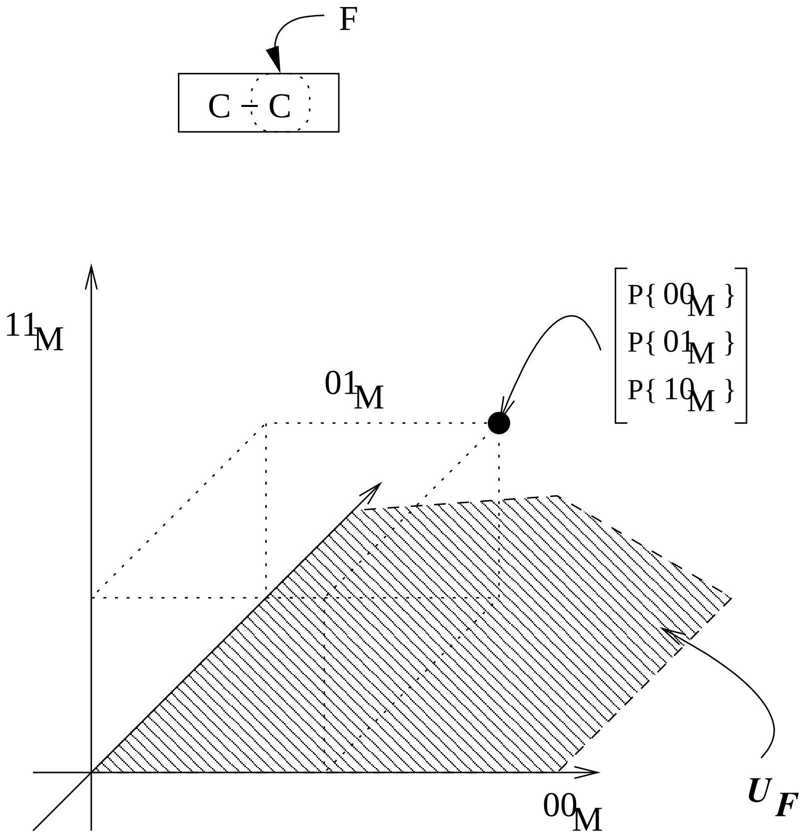
$$\vec{s}^T \mathbb{I}_{ALA} = \sum_{xyz} s_{xyz} \mathbb{P}\{^{xyz}ALA\} = d$$

to the isotopomer distribution $\mathbb{I}_{ALA}$, represented as a matrix system

$$S^T \mathbb{I}_{ALA} = \vec{d}$$

Vector space interpretation of the measurements

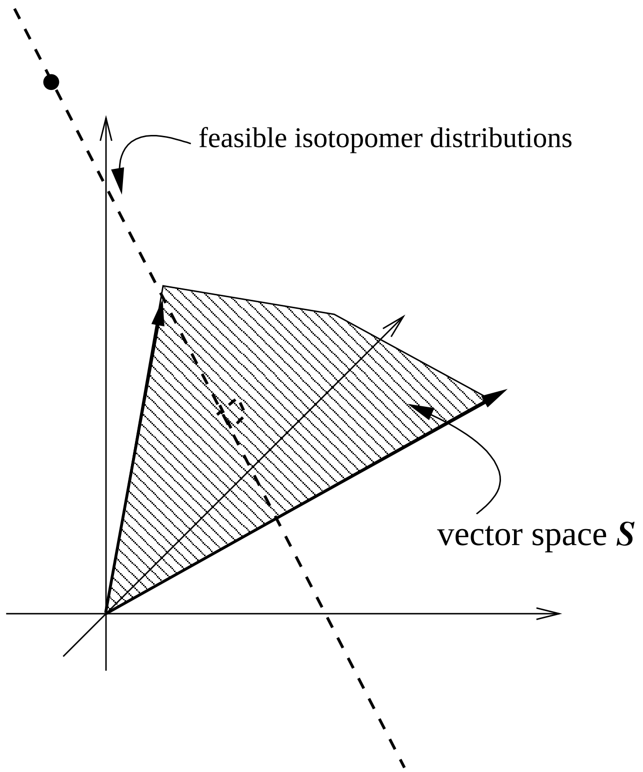
- An n-carbon metabolite M is associated with a 2^n-dimensional **vector space** $\mathcal{I}_M$.
- Isotopomer distributions of M are **points** in $\mathcal{I}_M$.
- Each k-carbon **fragment** F of $\mathcal{I}_M$ is associated with a 2^k-dimensional vector **subspace** $\mathcal{U}_F \subset \mathcal{I}_M$.
- Isotopomer distributions of F are **points** in $\mathcal{U}_F$.



(a 3-dimensional subspace of the 4-dimensional isotopomer space depicted)

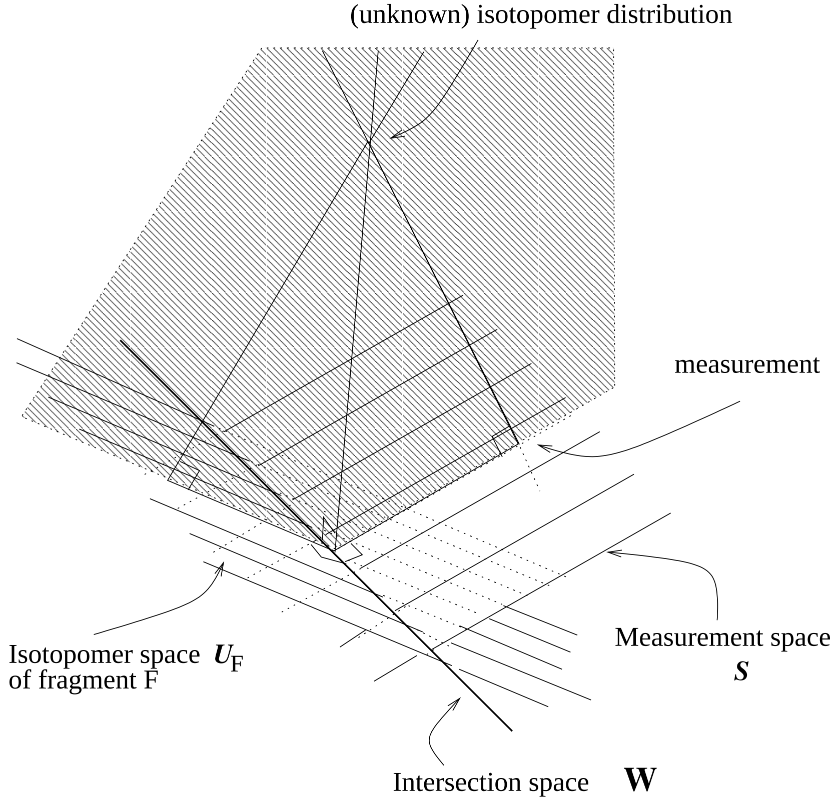
A measurement is an orthogonal projection

- A measurement $S^T \mathbb{I}_M = \vec{d}$ defines a point (the inverse image of $\vec{d}$) in the column space $\mathcal{S} \subset \mathcal{I}_M$ of matrix S.
- Isotopomer distributions $\mathbb{I}_M$ consistent with the measurement have their **orthogonal projections** on that point.
- Any such isotopomer distribution $\mathbb{I}_M$ **must be considered possible**, in the light of the measurement.



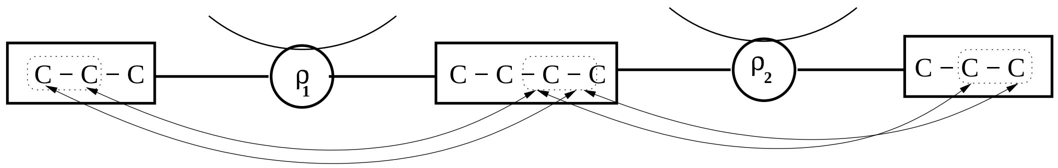
Mapping measurement data to metabolite fragments

- Given isotopomer measurement $S^T \mathbb{I}_M = \vec{d}$ of the metabolite, what is known about the isotopomer distributions $S^T \mathbb{I}_F$ of its fragments F?
- Any linear constraint to $\mathbb{I}_F$ lies in the subspace $\mathcal{U}_F$.
- This calls for computing the intersection space $\mathcal{W} = \mathcal{U}_F \cap S$ and the orthogonal projection of the inverse image of the measurement $\vec{d}$ to $\mathcal{W}$.

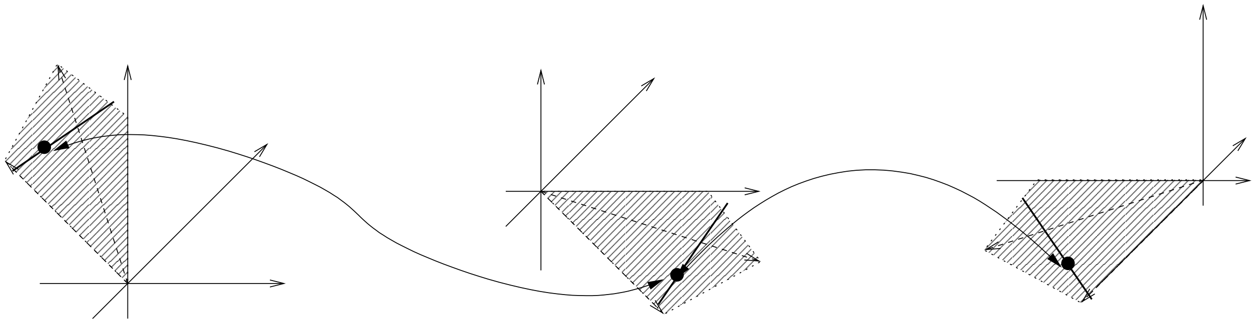


Propagating isotopomer constraints over reactions

Each reaction (and transitively, an unbranched pathway) defines a one-to-one mapping between carbon atom locations ('carbon map')

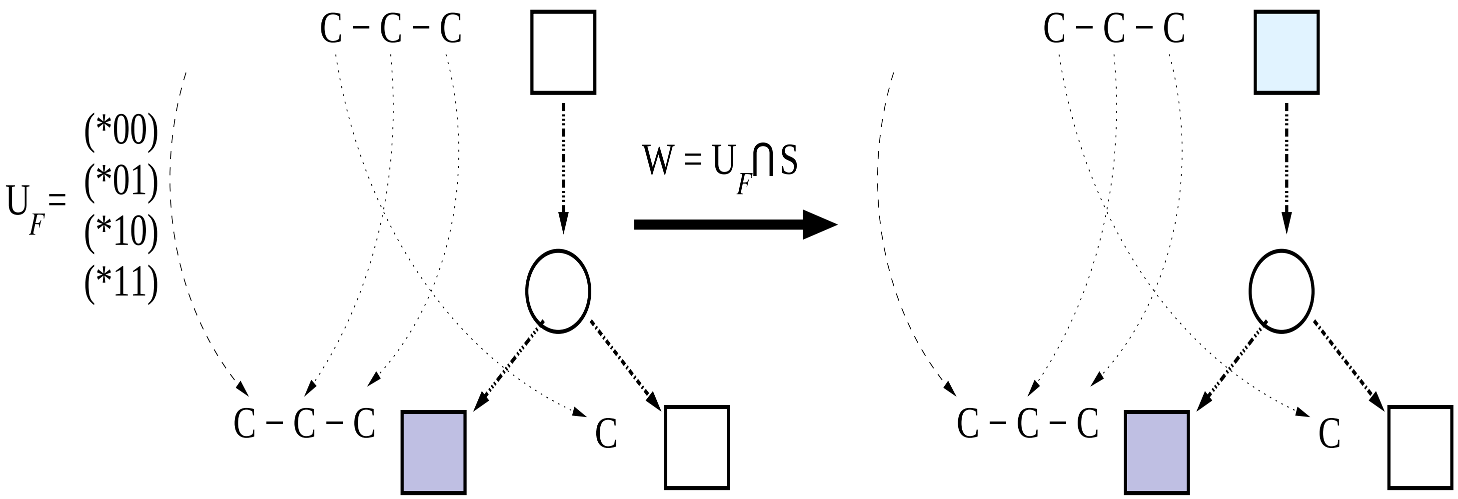


This translates to an mapping between the **subspaces** of the fragments along an unbranched pathway



Mapping isotopomer constraints over reactions

$$\alpha\,(P(*00) + P(*01)) + \beta\,(P(*10) + P(*11)) = f_1$$
$$\chi\,(P(*00) + P(*01)) + \delta\,(P(*10) + P(*11)) = f_2$$



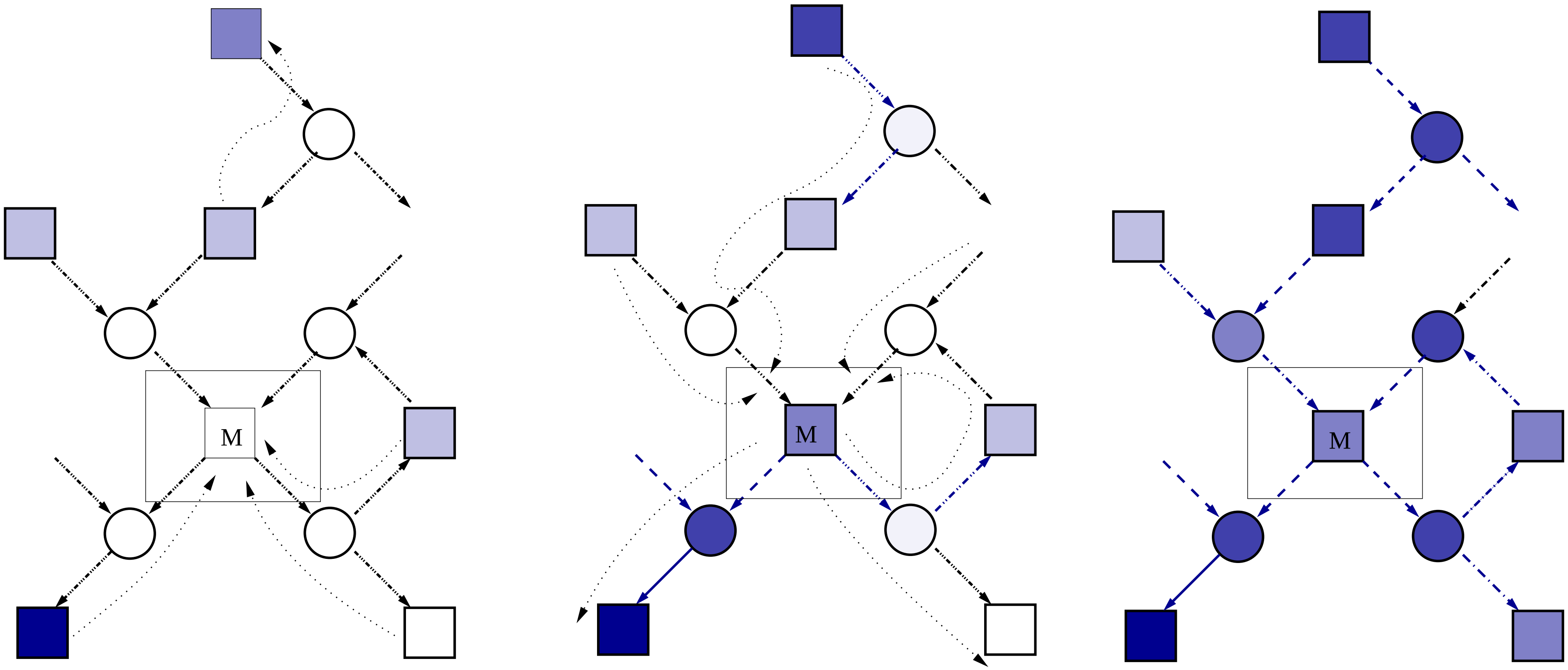
$$S = \begin{array}{ll} P(000) & = d_1 \\ P(001) + P(010) + P(100) & = d_2 \\ P(011) + P(101) + P(110) & = d_3 \\ P(010) + P(110) + P(011) + P(111) & = d_4 \\ P(111) & = d_5 \end{array}$$

Propagation algorithm

1) Backward propagation toward junctions where pathways join

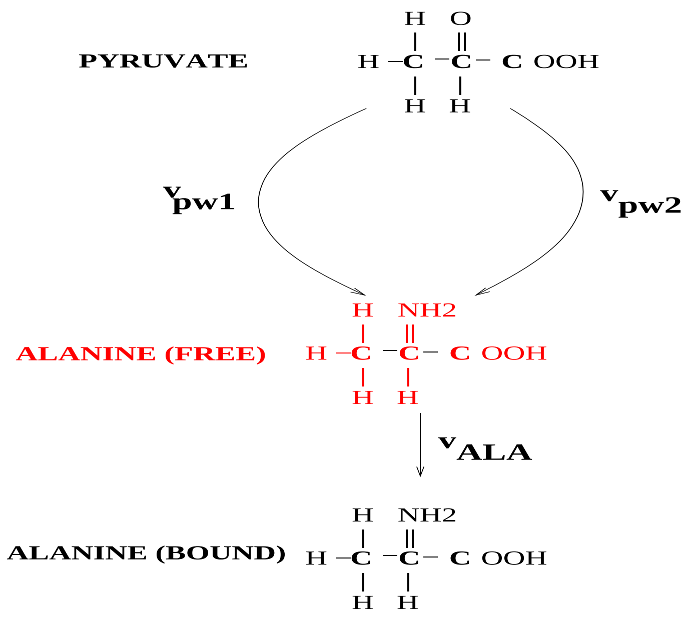
2) Forward propagation toward edges preceding the junctions

3) After propagation, some isotopomer constraint are known for each junction



Generalized isotopomer balances

Assume measurements
$S^T \mathbb{I}_{ALA|pw_1} = \vec{d_1}$,
$S^T \mathbb{I}_{ALA|pw_2} = \vec{d_2}$ and
$S^T \mathbb{I}_{ALA} = \vec{d}$.



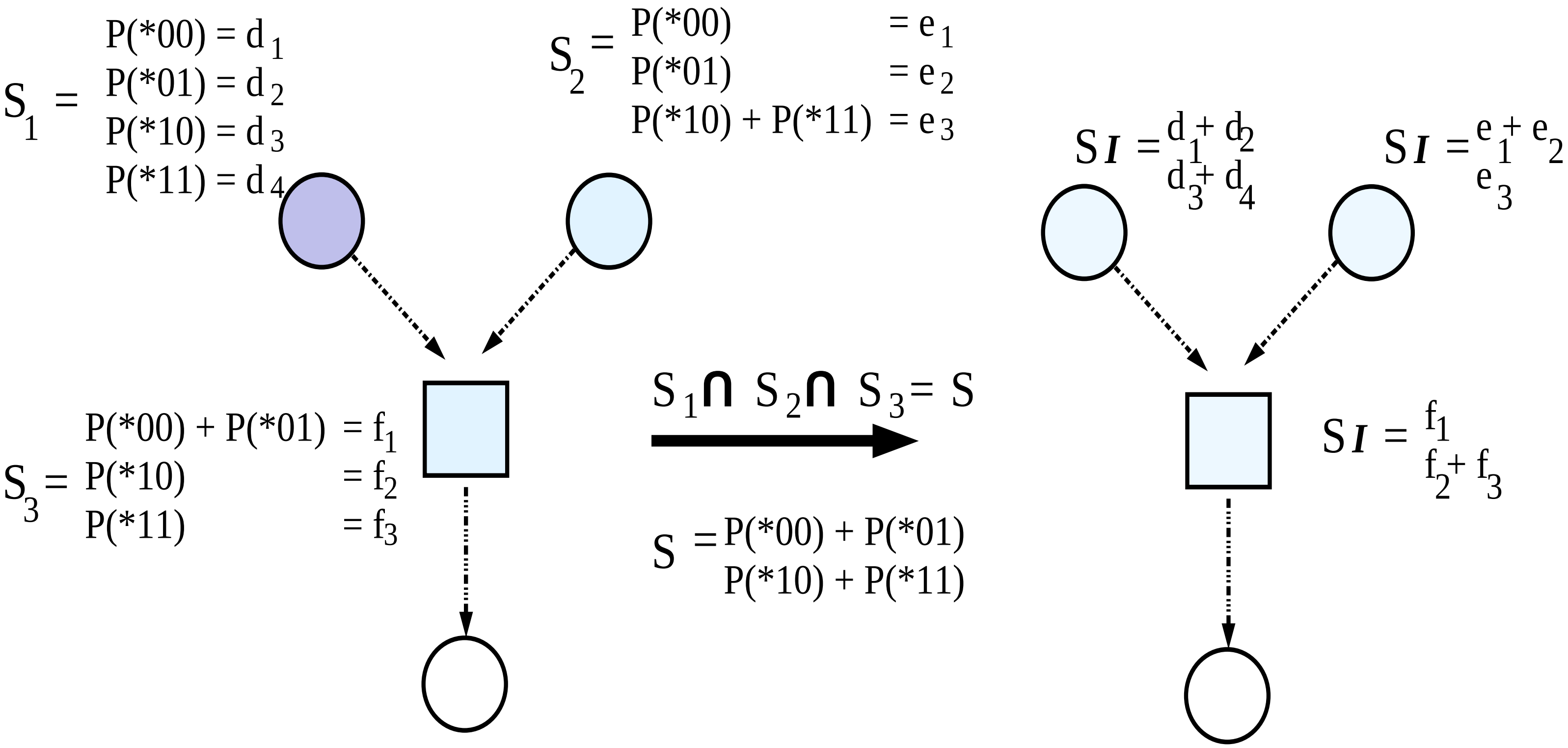
Assuming zero accumulation, we get a generalized isotopomer balance

$$\vec{d_1} \cdot v_{pw_1} + \vec{d_2} \cdot v_{pw_2} = \vec{d} \cdot v_{ALA}$$

The vectors $\vec{d_1}$, $\vec{d_2}$ and $\vec{d}$ are obtained by propagating the measurement information from upstream (e.g. pyruvate) and possibly from downstream (e.g. protein-bound alanine) metabolites towards the metabolite of interest (e.g. free alanine).

Generalized isotopomer balances

In order to get generalized isotopomer balance, the $\vec{d_i}$'s have to be defined in the same subspace. We must compute the intersection of S_i's.

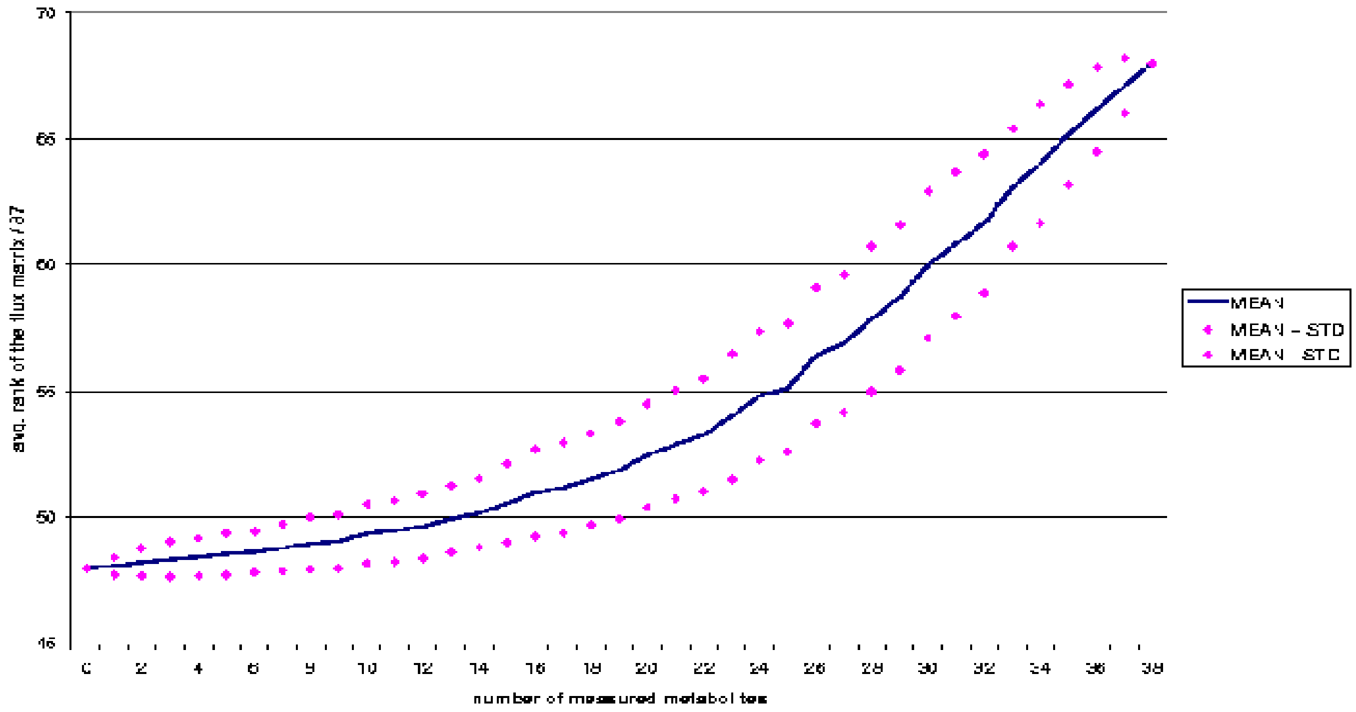


Identifiability of fluxes

We tested our method with a metabolic network containing EMP, PPP and simplified amino acid producing pathways and the TCA cycle. Network had 38 metabolites and 52 reactions, of which 38 were bidirectional. Isotopomer measurements were simulated using 13C Flux software by Wiechert et al (1999).

Identifiability of fluxes

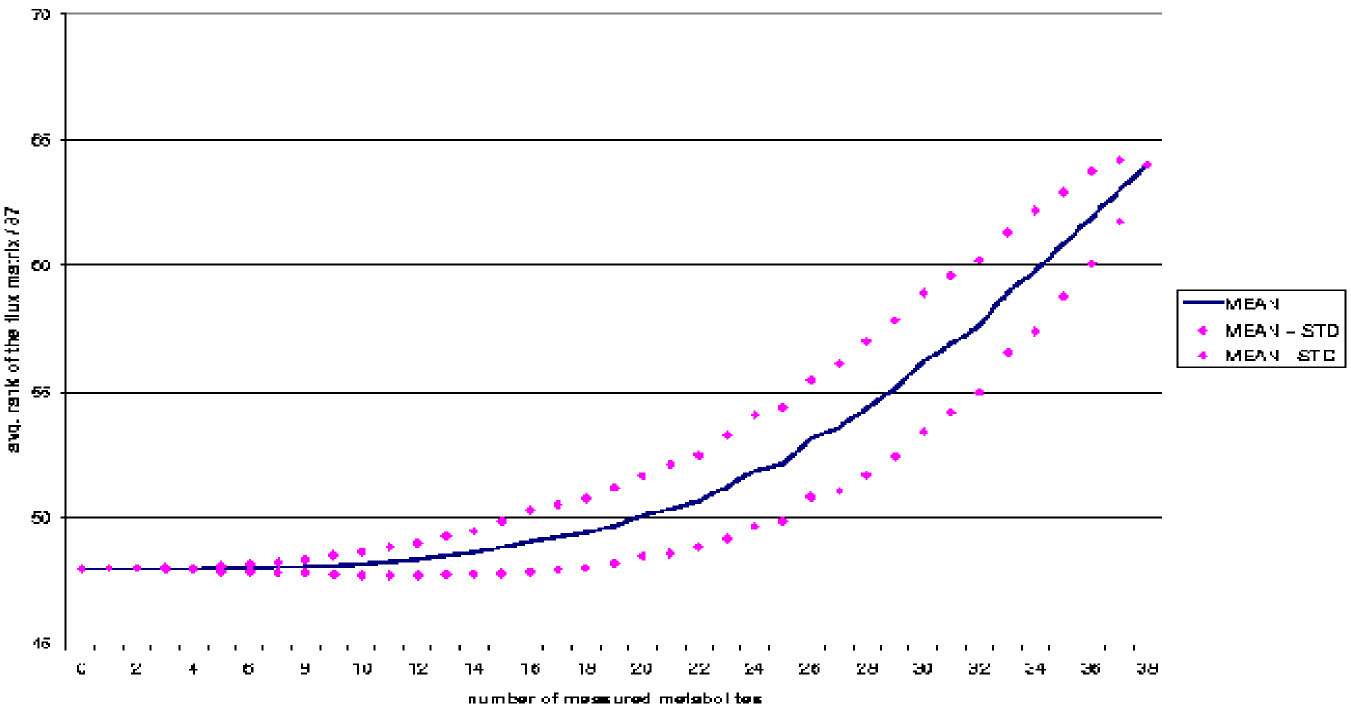
Complete isotopomer information



avg. rank of the flux matrix / 37
number of measured metabolites
MEAN
MEAN - STD
MEAN + STD

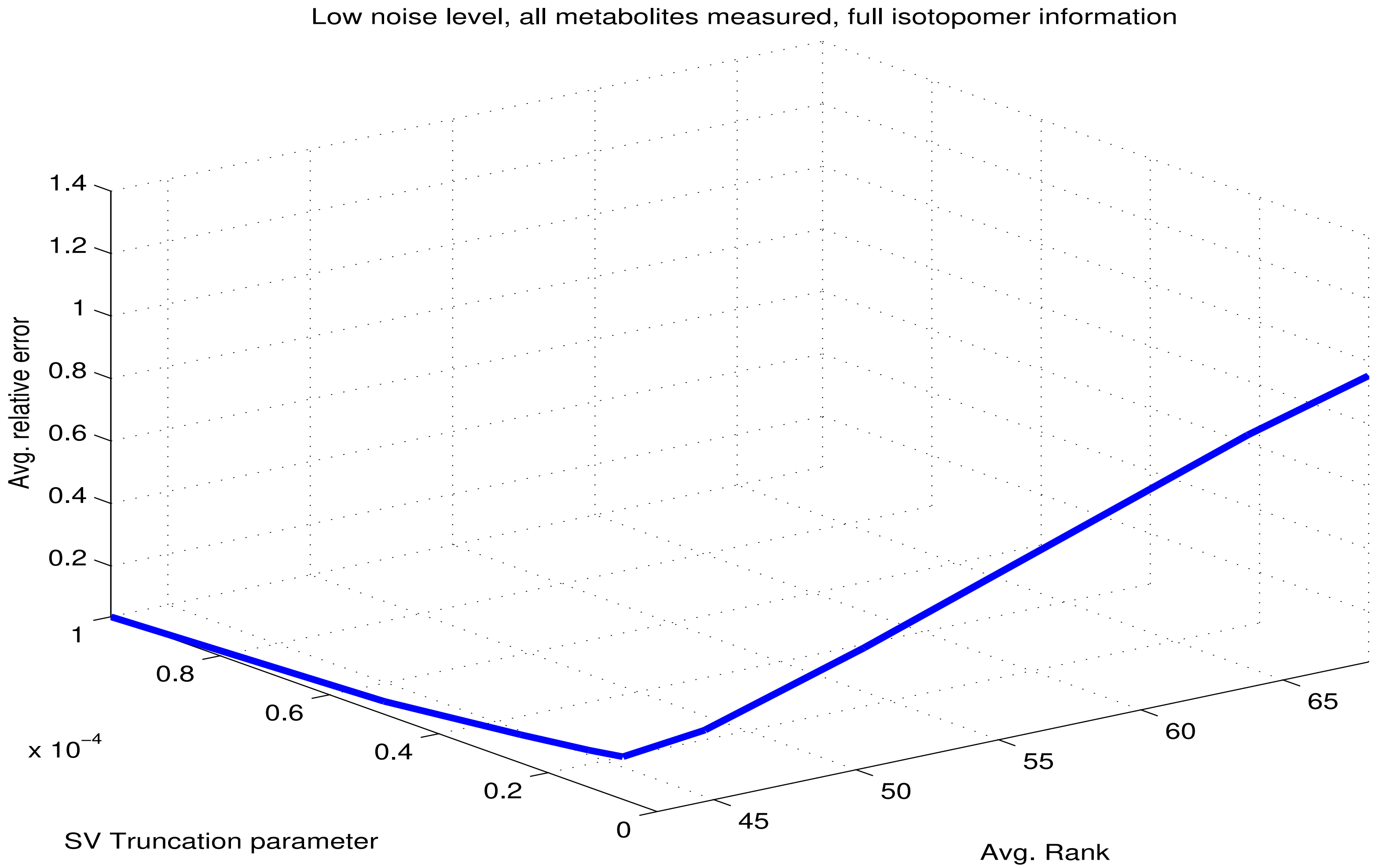
Identifiability of fluxes

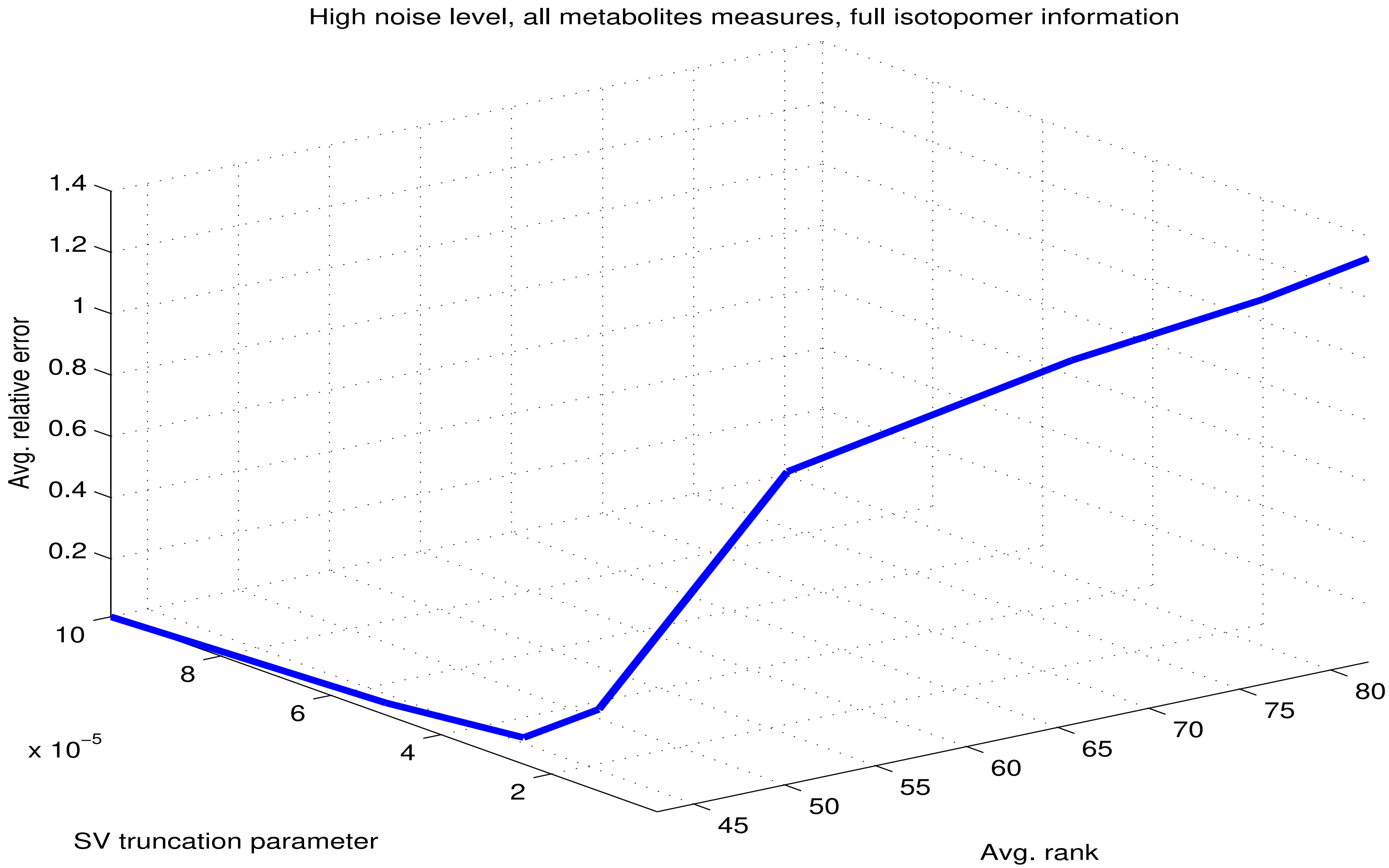
Positional enrichment information


70
65
60
55
50
45
avg. rank of the flux matrix / 87
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
number of measured metabolites
MEAN
MEAN - STD
MEAN STD

Behaviour under noise

- We generated noisy isotopomer distributions based on simulated distributions and sampling from multinomial distribution.
- Distance between the image of the correct flux distribution and the b-vector of the constructed equation system was observed.
- Resulting flux matrices highly ill-conditioned.
- We are currently testing several regularization methods.
  - singular value truncation + Tikhonov regularization
  - number of nonzero singlular values == rank of the equation system
  - "trade off between the rank and the error rate"


Low noise level, all metabolites measured, full isotopomer information
Avg. relative error
1.4
1.2
1
0.8
0.6
0.4
0.2
1
0.8
0.6
0.4
0.2
0
x 10^-4
SV Truncation parameter
45
50
55
60
65
Avg. Rank


High noise level, all metabolites measures, full isotopomer information
Avg. relative error
1.4
1.2
1
0.8
0.6
0.4
0.2
10
8
6
4
2
x 10^−5
SV truncation parameter
45
50
55
60
65
70
75
80
Avg. rank

More powerful propagation?

- Better treatment of gaps in measurement data.
  - Currently propagation stops at each junction, requirements to the measurement technology are too high
  - Isotopomer distributions of some fragments of the junction metabolite are invariate to the values of the fluxes. Propagate these through the junctions.

Conclusions

- New, fast flux estimation methods based on isotopomer tracing experiments was presented
- Only mild assumptions about the isotopomer data
- Room for improvements in the propagation of the measurement data and behaviour under noise (both under construction)