

# EVRI-thing You Need to Know About Shift Invariant Tri-linearity for Spectral Unmixing in LC/GC-MS

Neal B. Gallagher<sup>1</sup>, Paul-Albert Schneide<sup>2</sup>, Rasmus Bro<sup>2</sup>

<sup>1</sup>Eigenvector Research, Inc., Manson, WA USA

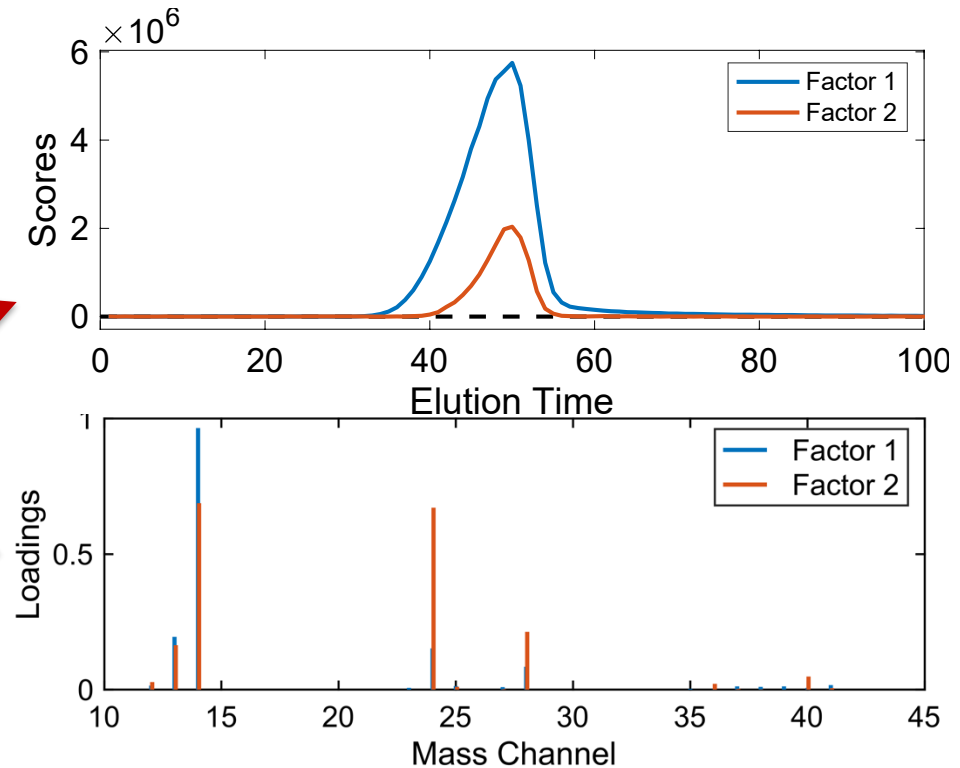
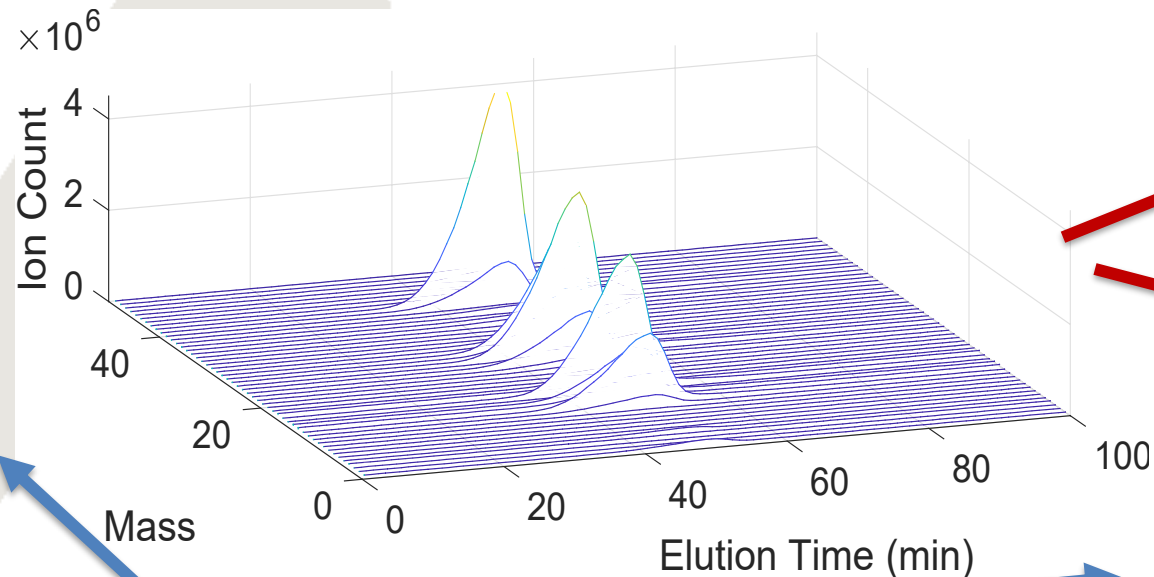
<sup>2</sup>Univeristy of Copenhagen, Copenhagen, Denmark

# SIT Webinar Outline

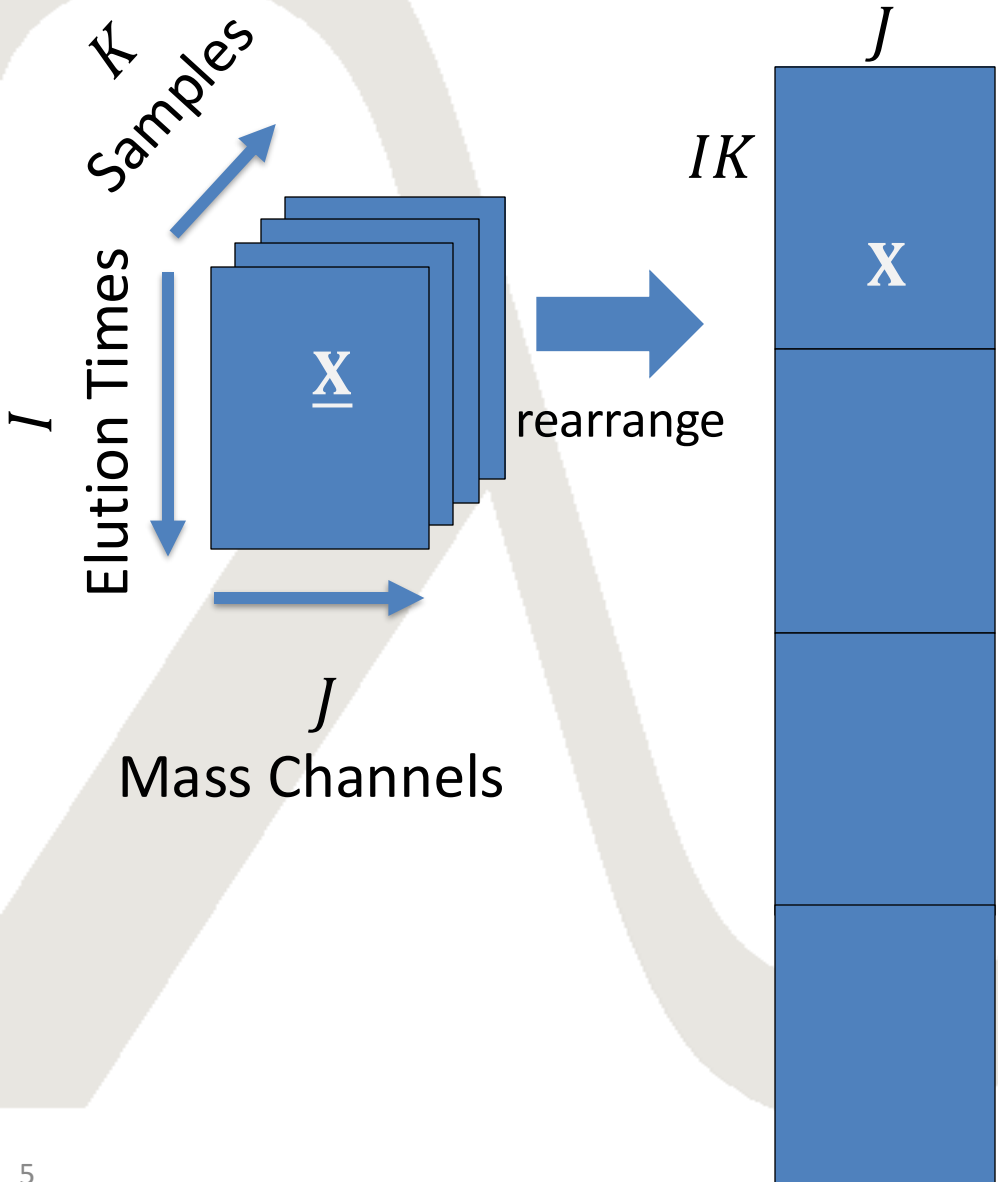
- Data analysis objective
- What is shift invariant tri-linearity (SIT) and how does it work?
- Command line examples from PLS\_Toolbox
- What is the future of SIT?
- Questions

# Blind source separation for hyphenated chromatography

- the objective is to estimate the underlying signal sources from measurements on multiple samples to obtain pure component elution profiles and spectra



# Data Arrangement



- $\underline{\mathbf{X}}$  is  $I \times J \times K$  (tri-linear)
  - (elution time) by (spectra) by (samples) and, ideally, can be modeled by parallel factor analysis (PARAFAC)
  - shifts in elution profiles violate tri-linearity
- $\mathbf{X}$  is  $IK \times J$  (bi-linear)
  - (elution time\*samples) by (spectra)
  - Principal components analysis (PCA)
  - $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$
  - Multivariate curve resolution (MCR)
  - $\mathbf{X} = \mathbf{CS}^T + \mathbf{E}$

# MCR/ALS

## Alternating constrained least squares

Start with an estimate of  $\mathbf{S}_0$ , set  $i = 0$

→ While convergence criteria not met

$$i = i + 1$$

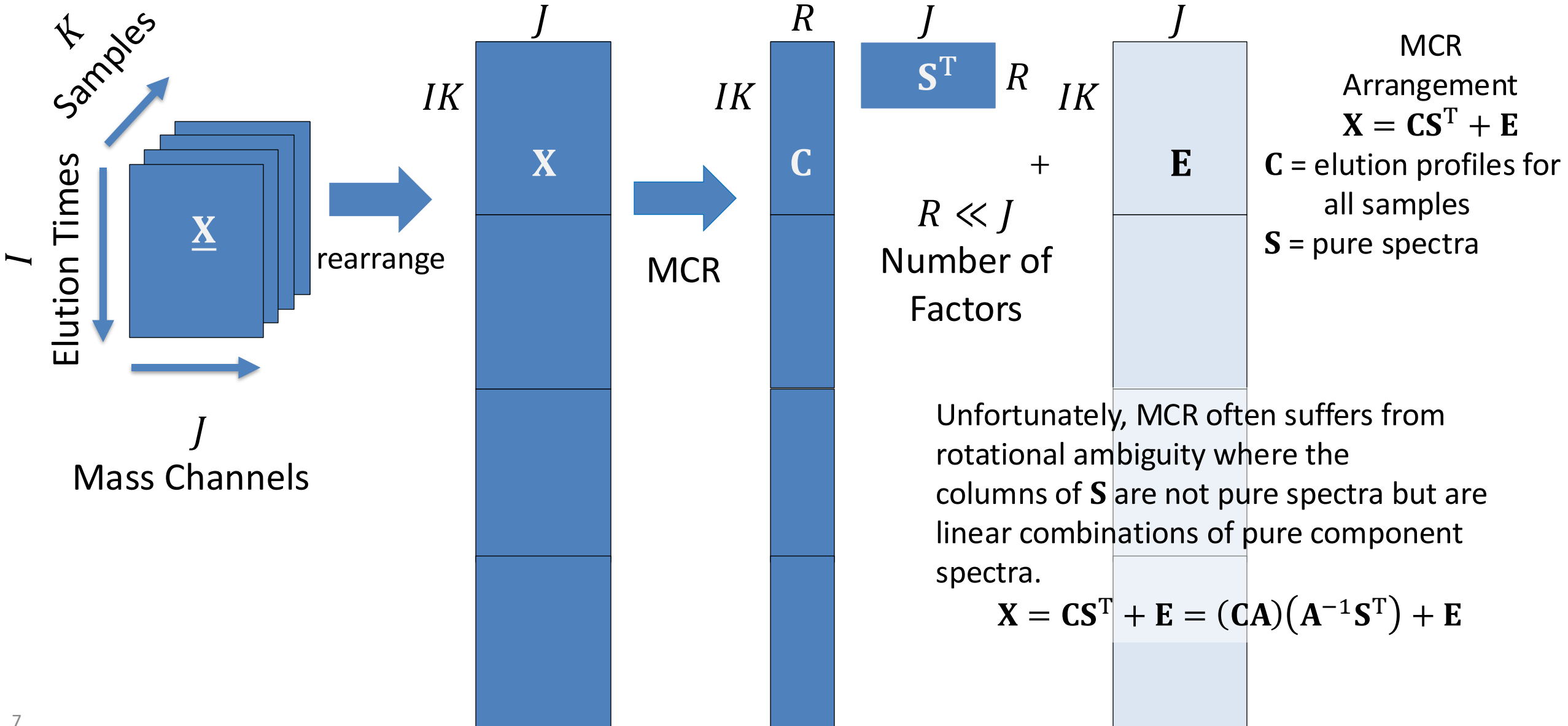
Calculate  $\mathbf{C}_i$ , apply non-negativity

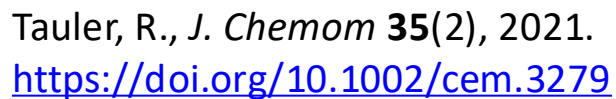
Calculate  $\mathbf{S}_i$ , apply non-negativity

Normalize, set  $\mathbf{s}_k \rightarrow \mathbf{s}_k / \|\mathbf{s}_k\|$  for  $k = 1, \dots, K$

continue until convergence

# MCR

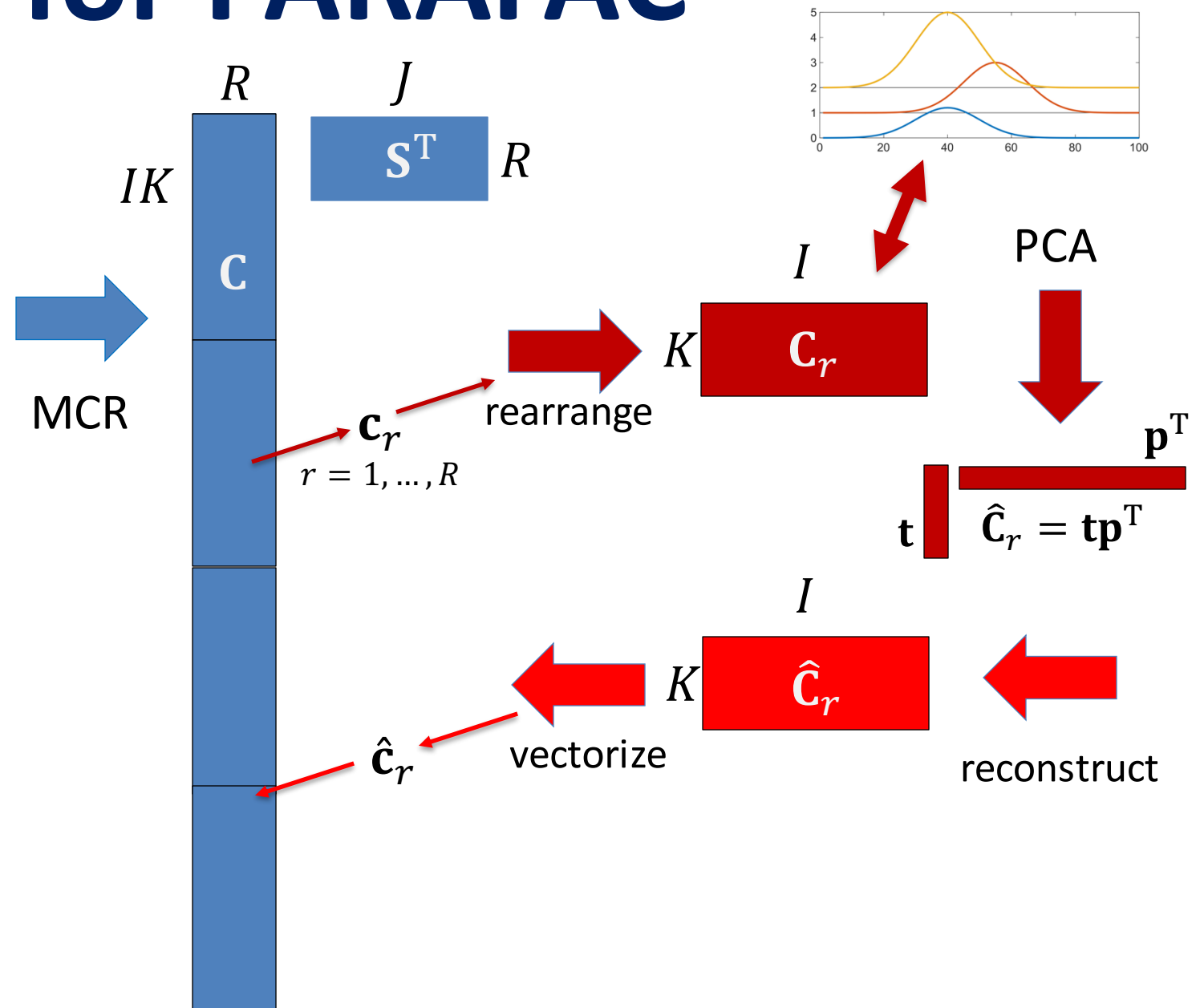




# MCR for PARAFAC

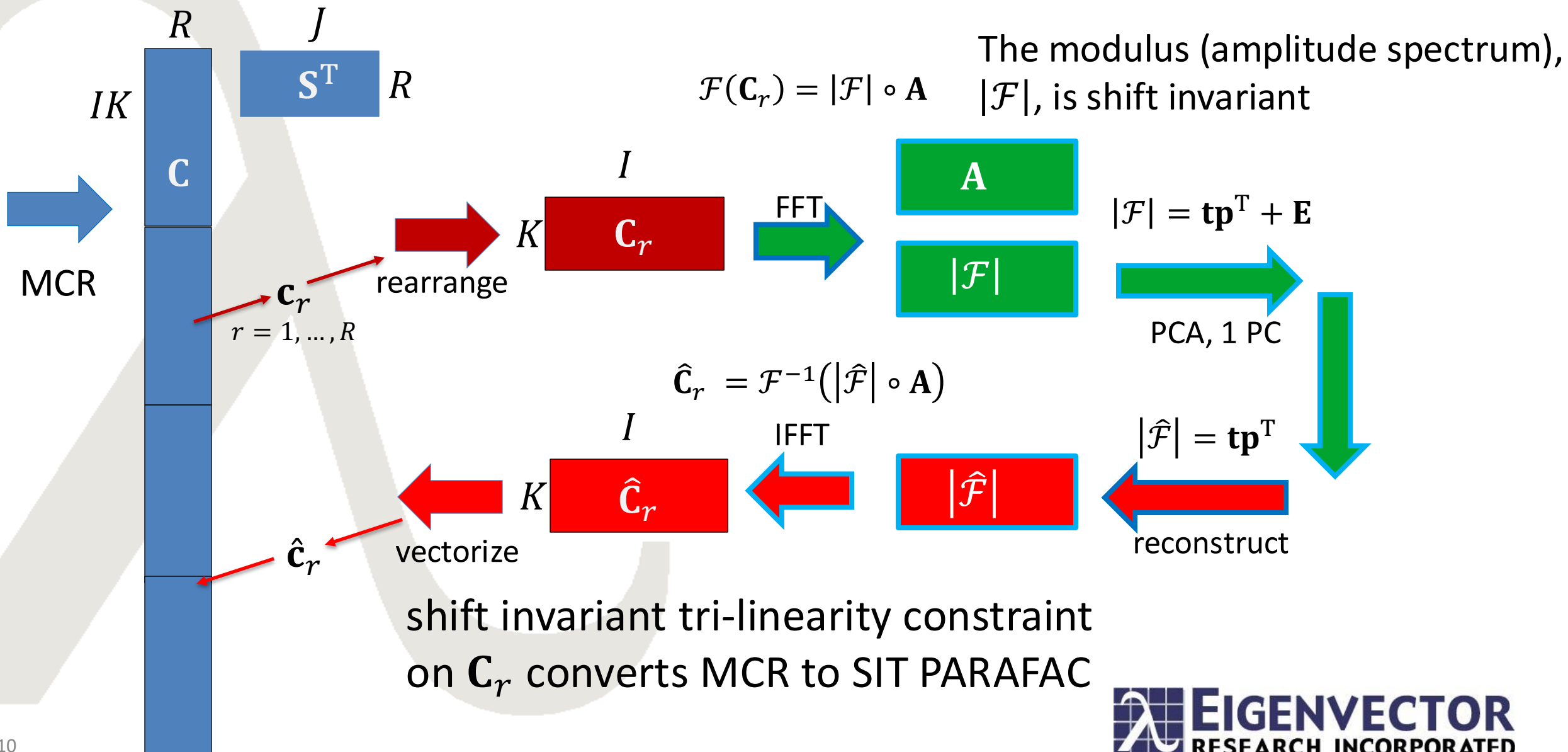
Problem:  
If elution profiles in  $\mathbf{C}_r$  shift,  
then a one-PC PCA model does  
not appropriately represent the  
elution signal, and  
 $\mathbf{C}_r$  will have rank  $\geq 1$ .

This is the same problem for  
PARAFAC – the shifting causes  
non-tri-linear data.





# Shift-Invariant Tri-linearity



# State of the Art Models

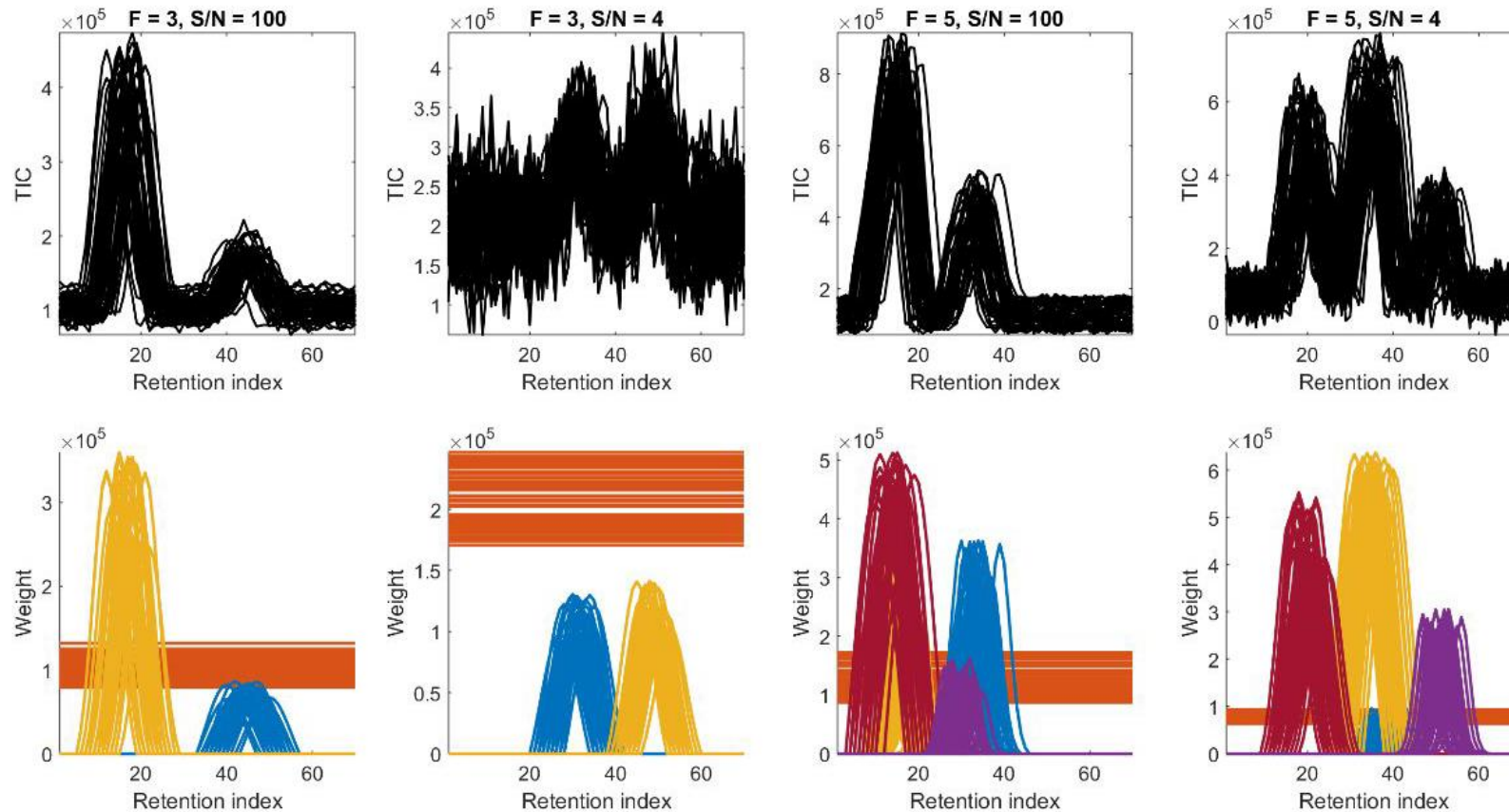
- MCR (Multivariate Curve Resolution)
  - very flexible and can be used if elution profiles shift
  - often suffers from rotational ambiguity: **non-unique**
    - can work for MS due to selectivity
- PARAFAC (PARAllel FACtor Analysis)
  - appropriate if the **data follow tri-linearity**
  - PARAFAC is not useful if elution profiles shift
  - PARAFAC2 (slow) is designed for shifts (plus) but tends to ~suffer from poor accuracy
  - PARAFAC2 with flexible coupling

# Shift-Invariant Tri-linearity

- MCR implementation of PARAFAC
  - **limited to non-shifting profiles**
  - flexible, can be relaxed on non-trilinear factors
  - can use >1 PC to account for elution peak shape changes
- SIT implementation of PARAFAC<sup>1</sup>
  - **allows for shifting profiles**
  - flexible, can be relaxed on non-trilinear factor
  - can use >1 PC to account for elution peak shape changes (SISoftT)<sup>2</sup>
  - **allows for profiles to shift relative to each other**
  - extended to GCxGCxTOFMS<sup>3</sup>
    - shifts observed in multiple modes

<sup>1-3</sup> references on last slide of the deck

# Simulated Data



simulated profiles

Number

of Factors

S/N

3

100

3

4

5

100

5

4

true profiles

# Explained Fit cosine with true profile

S/N = 100

S/N = 4

S/N = 100

S/N = 4

Algorithm	Components	EFP <sub>max</sub> [%]	Computation	Tucker congruence		
			time [s]	A	B	C
PF2	3	90.0542	24.0	1.000	1.000	1.000
PFFC	3	90.0542	20.0	0.979	0.989	1.000
SIT	3	90.0542	0.7	1.000	1.000	1.000
PF2	3	80.8803	14.0	0.976	0.989	0.999
PFFC	3	80.8803	11.0	0.973	0.984	1.000
SIT	3	80.8803	0.5	1.000	0.997	1.000
PF2	5	99.0607	42.0	0.999	0.999	1.000
PFFC	5	99.0607	84.0	0.966	0.987	1.000
SIT	5	99.0607	2.2	1.000	1.000	1.000
PF2	5	80.7828	47.0	0.892	0.945	0.999
PFFC	5	80.7828	32.0	0.946	0.978	1.000
SIT	5	80.7828	1.4	0.997	0.991	1.000

PF2, PARAFAC2–ALS

PFFC, PARAFAC2-flexible coupling

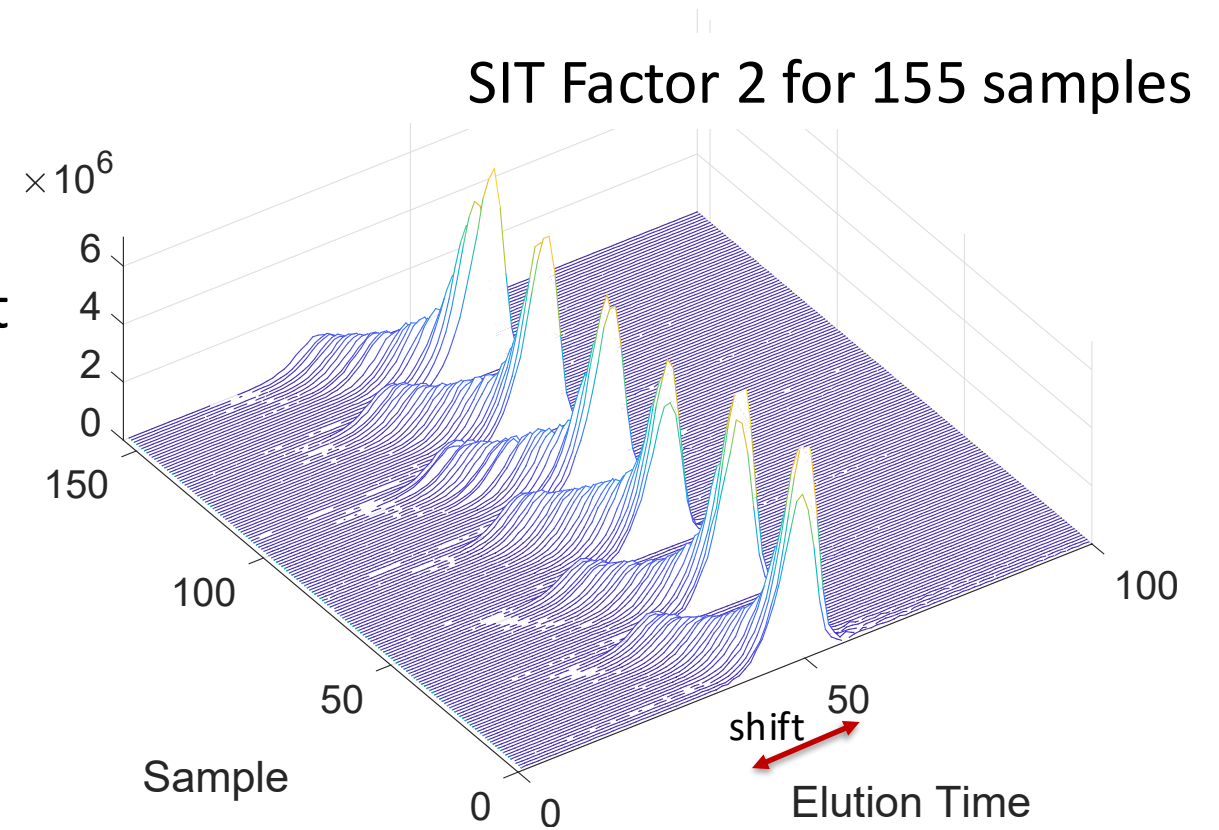
SIT, shift-invariant tri-linearity.

0 ≤ Tucker congruence ≤ 1  
(worst) (best)

SIT ~20x faster than PF2

# Apple Wine (Section 18)

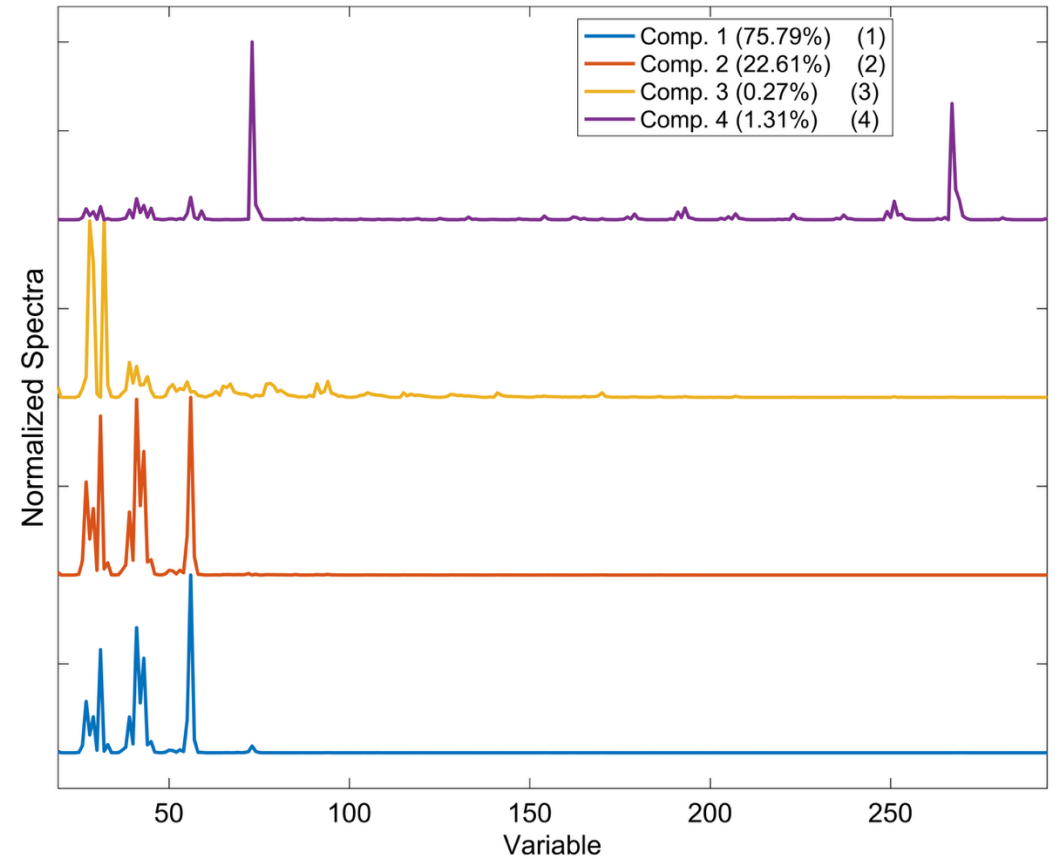
- The Apple Wine data example
  - (122 elution time points) x (286 masses) x (155 samples)
    - elution profiles shift and have slight shape changes
  - <https://web.tresorit.com/l/ADl7Q#X3jPgWg4lbt8C8TmMDD5dA>



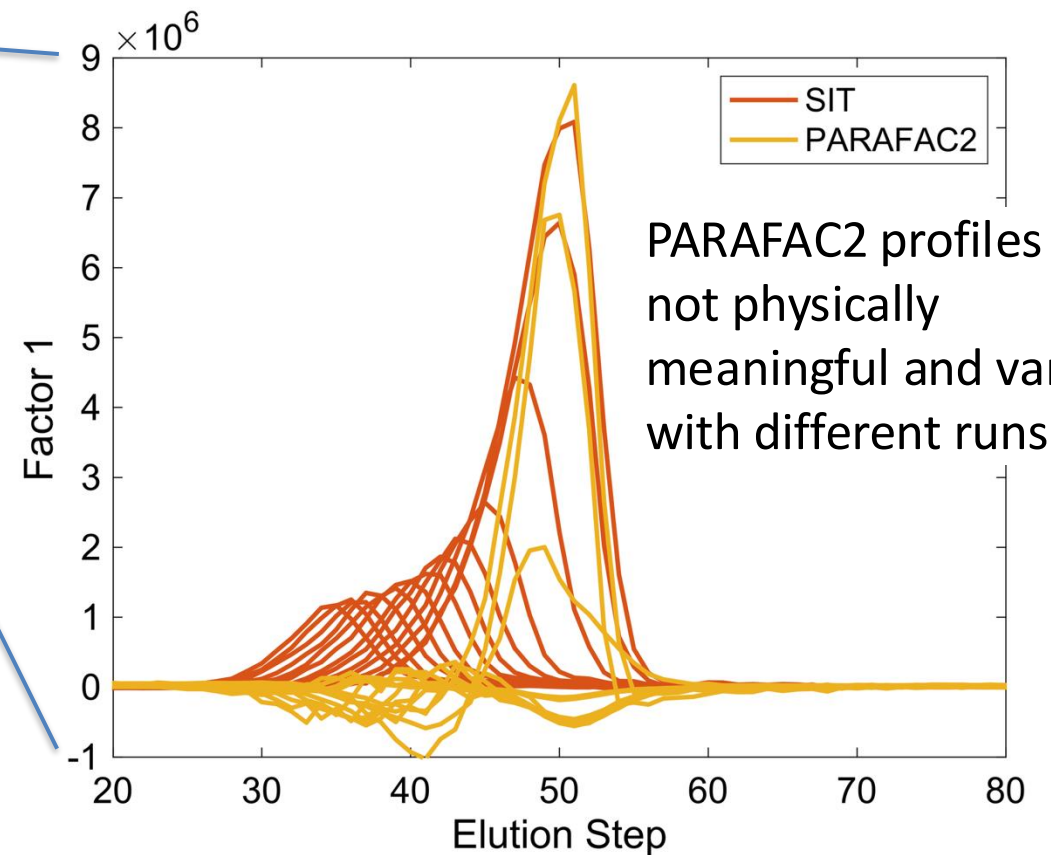
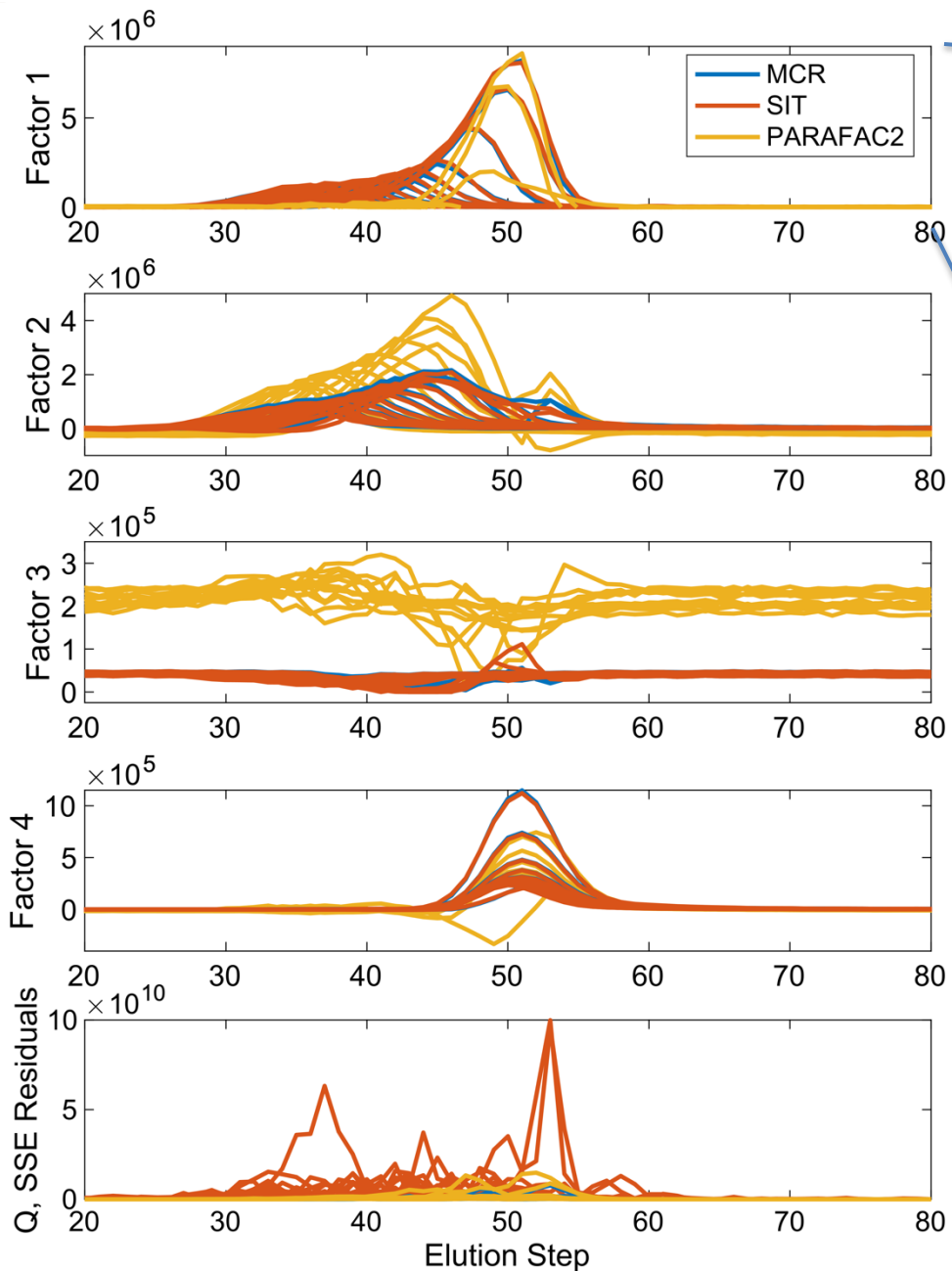
# Model Parameters

- Fit four factors
- MCR used non-negativity
- SIT used non-negativity and relaxed tri-linearity on Factor 3
  - baseline
- PARAFAC2 cannot be constrained non-negative on the elution profiles

SIT Estimated Mass Spec







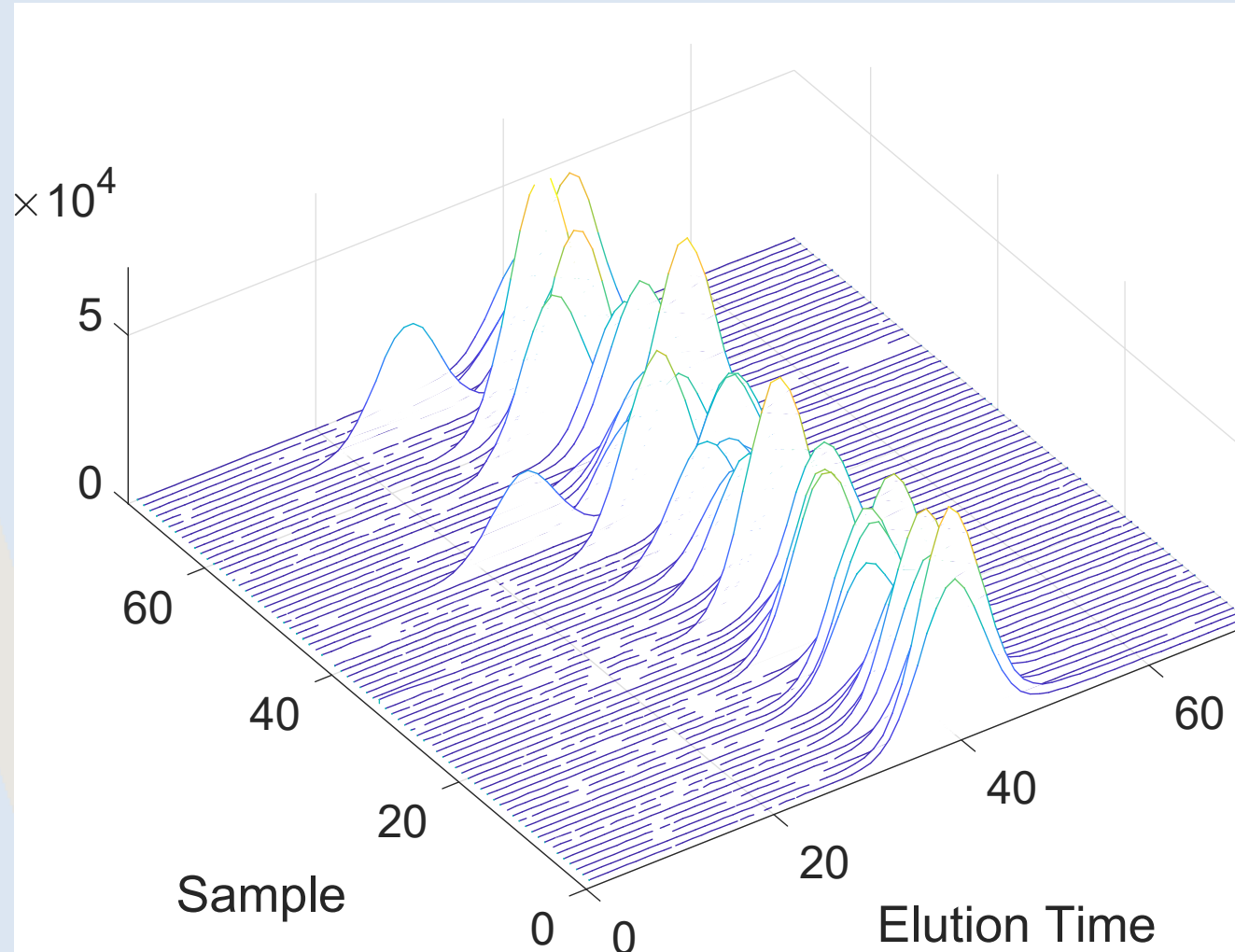
Factors 1 and 2 shift  
 Factor 4 does not shift  
 Factor 3 is ~baseline  
 SIT has lowest fit – it's not as flexible as MCR



# GC Wine Example

- GC-MS on Wine Samples: 71x199x72
  - Dynamic headspace GCMS data of red wines from different regions.
    - J. M. Amigo, T. Skov, J. Coello, S. MasPOCH, R. Bro. Solving GC-MS problems with PARAFAC2. *Trac-Trends in Analytical Chemistry* **27** (8):714-725, 2008.
    - A part of a GC-MS chromatogram of red wine.  
This area contains two analytes: an acetic-acid hexyl ester and 3-hydroxy-2-butanone.
  - 71 time points in each elution profile
  - 199 mass channels
  - 72 samples

# Example of Elution Profile Shifting



Factor 3 from ALS\_SIT

# Example Live Script

- Als\_Sitdemo.mlx
  - [https://www.dropbox.com/scl/fi/ev6n4bt52erc8syqwo5nq/Als\\_Sitdemo.mlx?rlkey=yf67cjbd1a4rapq6muhkj2n3g&dl=0](https://www.dropbox.com/scl/fi/ev6n4bt52erc8syqwo5nq/Als_Sitdemo.mlx?rlkey=yf67cjbd1a4rapq6muhkj2n3g&dl=0)
  - DOI: [10.13140/RG.2.2.34408.38402](https://doi.org/10.13140/RG.2.2.34408.38402)

# Recent and future developments

- Applications of SIST to HPLC-DAD data:
  - Shifts in LC are often more severe (even peak inversion can be observed)
  - Peak shape changes are more common (broadening, skewing)
  - DAD detector is less selective, therefore rotational ambiguity becomes a bigger problem for MCR<sup>4</sup>
- Incorporating noise filtering strategies<sup>3</sup> to make the model more robust in low SNR situations
- Incorporating essential variable selection<sup>5</sup> to make SIT even faster!

# Conclusions

- Shift Invariant Tri-linearity is a fast, flexible and accurate alternative to PARAFAC2
  - ~20x faster for cases studied
  - tri-linearity can be relaxed on factors that do not exhibit tri-linearity
  - non-negativity can be employed (have observed  $<0$  after iFFT)
    - PARAFAC2 cannot implement non-negativity in the shifting mode but alternatives are available

- P.-A. Schneide, R. Bro, N.B. Gallagher, “Shift-Invariant Tri-linearity - A new model for resolving untargeted GC-MS data”, *J. Chemom.* **37**(8) (2023); e3501. doi: [10.1002/cem.3501](https://doi.org/10.1002/cem.3501)
  - **2025**, *J. Chemom.* Bruce Kowalski Award for best paper
  - **2023**, top viewed paper in *J. Chemom.*
  - **2023**, top cited paper in *J. Chemom.*
- P.-A. Schneide, R. Bro, N.B. Gallagher, “Shift invariant soft tri-linearity: Modelling shifts and shape changes in gas-chromatography coupled mass spectrometry,” *Chemometr. Intell. Lab.*, **251**, 105155 (2024). doi: [10.1016/j.chemolab.2024.105155](https://doi.org/10.1016/j.chemolab.2024.105155).
- Schneide, P.-A., Armstrong, M., Bro, R., Gallagher, N.B., “Unlocking new capabilities in the analysis of GC×GC-TOFMS data with shift-invariant multi-linearity,” *J. Chemom.*, **39**(1) / e3623 (2025). doi: [10.1002/cem.3623](https://doi.org/10.1002/cem.3623). [arxiv.org/abs/2412.12114](https://arxiv.org/abs/2412.12114).
- P.-A. Schneide, R. Bro, R. Tauler, N.B. Gallagher, J.L. Hinrich, “A unifying framework for modelling non-negative bi-linear, tri-linear and “in-between” data in chemometrics. Part I: Theoretical framework and concepts,” (in progress, 2025)
- R. Vitale & A. Azizi & M. Ghaffari & N. Omidikia & C. Ruckebusch. “Three-Way Data Reduction Based on Essential Information” *J. Chemom.*, **38** / e3617 (2024). doi: [10.1002/cem.3617](https://doi.org/10.1002/cem.3617)