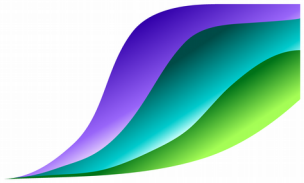




Borries Demeler, Ph.D.  
Dept. of Chemistry and Biochemistry  
Canada 150 Research Chair  
in Biophysics



NORTHWEST  
BIOPHYSICS  
CONSORTIUM

## Finite Element Modeling in UltraScan



# What is UltraScan and what does it offer?

## UltraScan is a Comprehensive Hydrodynamic Analysis Platform



### Some Statistics:

- Approximately 6.5 M lines of code
- \$3.5 M NIH/NSF/HHMI funding
- 30 years of development
- Model-E, XLA, XLI, XLF, Optima support
- Well over 500 publications supported by UltraScan analysis

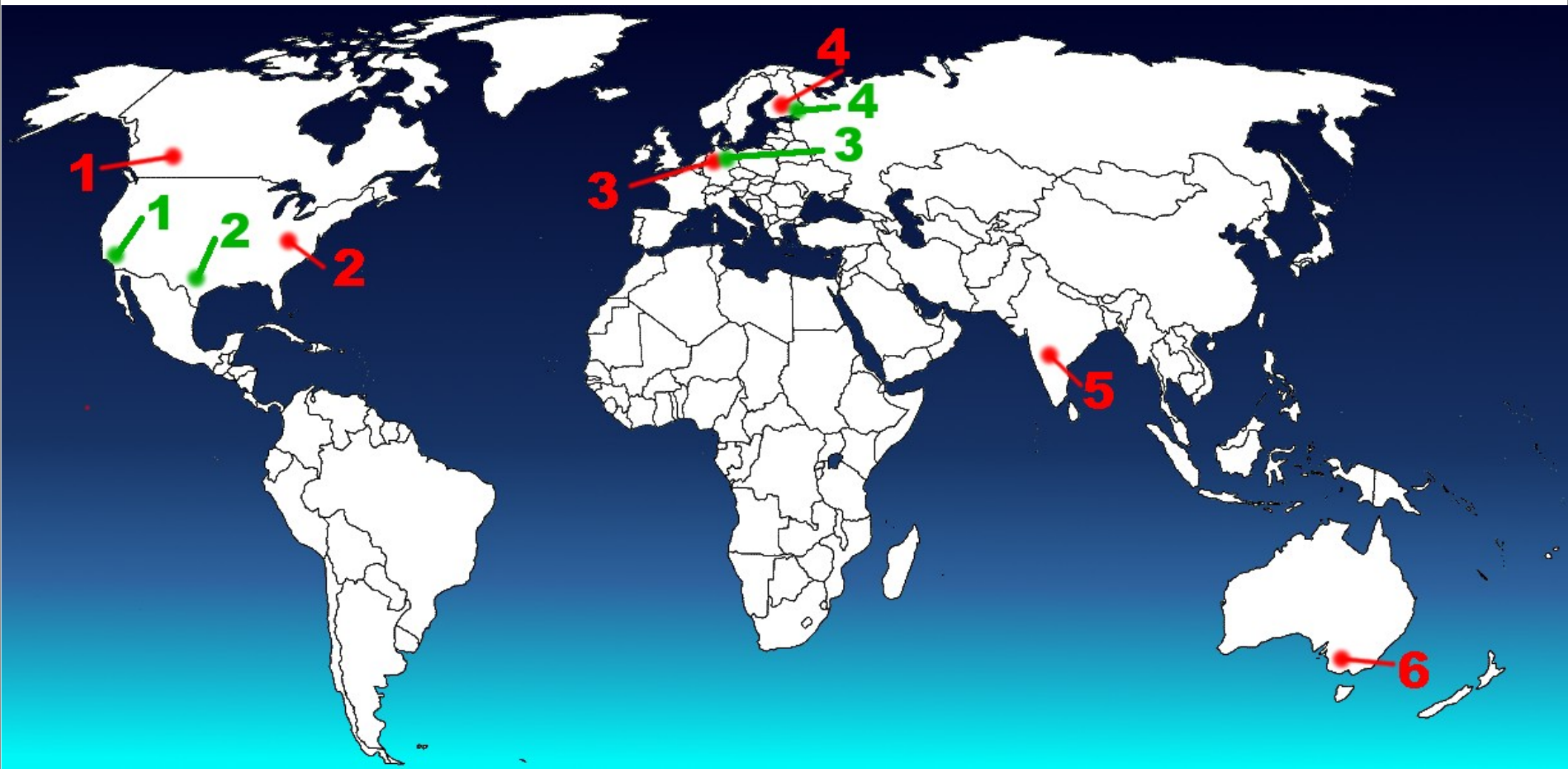
- Complete Optima Integration/Data Acqui.
- Supports high-performance computing
- Adheres to the OpenAUC data standard
- Provides highest resolution analysis
- Provides best fit/lowest RMSD solutions
- 2DSA, PCSA, GA, MC, DMGA, CG, vHW ...
- Highest accuracy ASTFEM solutions
- Advanced simulation support
- Integrated LIMS system, SSL encryption
- Relational database support
- Provides best fit/lowest RMSD solutions
- Hydrodynamic bead modeling (US-SOMO)
- Global fitting support for unlimited data
- Distribution analysis and statistics
- SAXS/SANS analysis
- Multi-platform and UltraScan-in-a-Box
- Multi-wavelength/Multi-Speed Analysis
- And now GMP...

## Worldwide UltraScan LIMS and HPC Server Installations

The UltraScan Science Gateway is a worldwide initiative facilitating AUC analysis on various high performance computing installations (Development funded by NSF/XSEDE)

**Red:** LIMS servers in Canada, USA, Germany, Finland, India and Australia

**Green:** HPC installations at SDSC (1), TACC (2), FZ-Jülich SCC (3), and U Helsinki (4)



[Welcome!](#)[Request New LIMS](#)[US3 Integration](#)[Partners](#)[Contacts](#)[Webmaster](#)

## Welcome to the UltraScan III LIMS Portal...

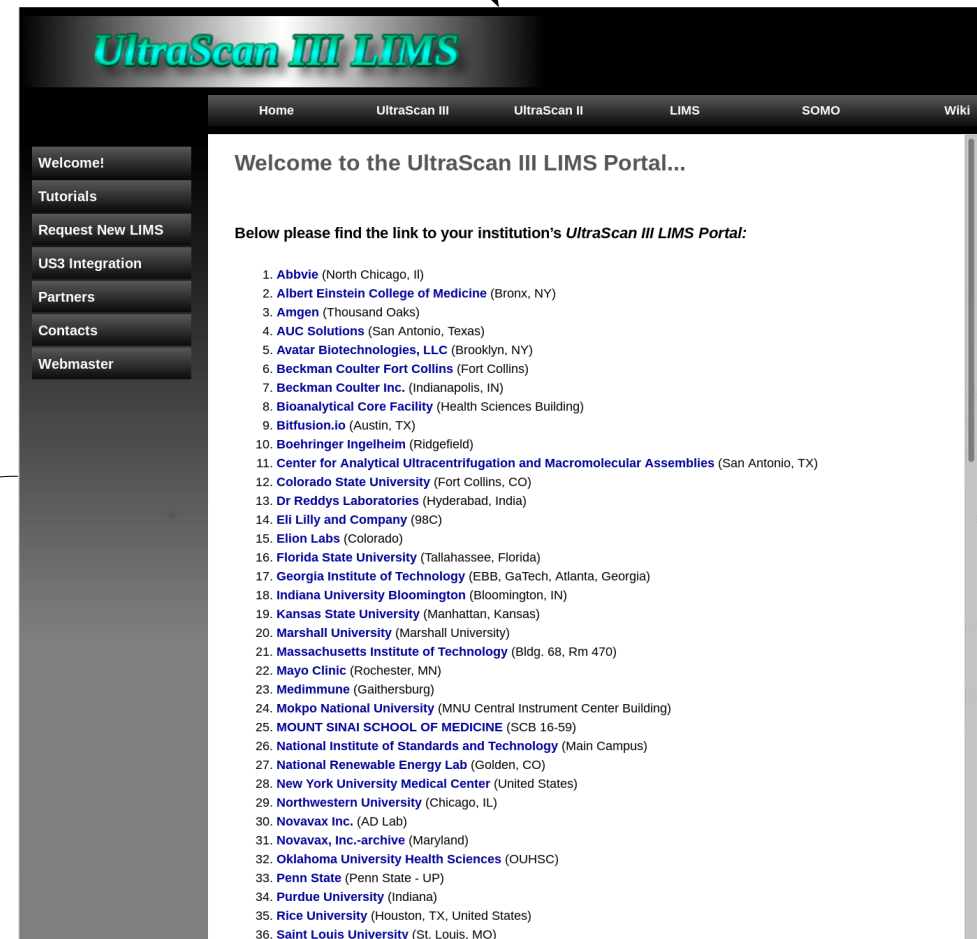
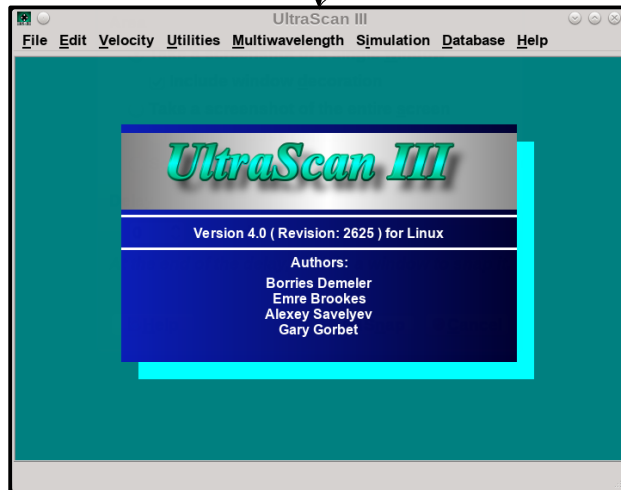
Below please find the link to your institution's *UltraScan III LIMS Portal*:

1. [AESKU.KIPP Institute](#) (Germany, RLP)
2. [Avatar Biotechnologies, LLC](#) (Brooklyn, NY)
3. [Beckman Coulter Inc.](#) (Indianapolis, IN)
4. [Bioanalytical Core Facility](#) (Health Sciences Building)
5. [CAUMA3](#) (San Antonio, TX)
6. [cauma3d](#) (San Antonio, TX)
7. [Center for Analytical Ultracentrifugation and Macromolecular Assemblies](#) (San Antonio, TX)
8. [Colorado State University](#) (Fort Collins, CO)
9. [Ecole Polytechnique Federale de Lausanne](#) (Lausanne, France)
10. [Florida State University](#) (Tallahassee, Florida)
11. [Heinrich-Heine-University Duesseldorf](#) (Duesseldorf, Germany)
12. [IGBMC ISB](#) (France)
13. [INDIAN INSTITUTE OF SCIENCE](#) (IISC, BANGALORE)
14. [Indiana University Bloomington](#) (Bloomington, IN)
15. [Institute of Microbial Technology](#) (Chandigarh, India)
16. [King Abdullah University of Science and Technology](#) (Thuwal, Saudi Arabia)
17. [Lund University](#) (Lund, Sweden)
18. [Marshall University](#) (Marshall University)
19. [Max Planck Institute for Colloids and Interfaces](#) (Potsdam, Germany)
20. [Max-Planck-Institute of Biochemistry](#) (Munich, GERMANY)
21. [MOUNT SINAI SCHOOL OF MEDICINE](#) (SCB 16-59)
22. [National Renewable Energy Lab](#) (Golden, CO)
23. [Northwestern University](#) (Chicago, IL)
24. [Purdue University](#) (Indiana)
25. [Rice University](#) (Houston, TX, United States)
26. [Science gateways group](#) (Bloomington)
27. [Texas A & M University](#) (College Station, TX)
28. [The University of Melbourne](#) (The University of Melbourne, Australia)
29. [Thomas Jefferson University, Kimmel Cancer Center](#) (KCC X-ray Crystallography and Molecular Interaction)
30. [Tokvo Institute of Technoloav](#) (Room 1003)

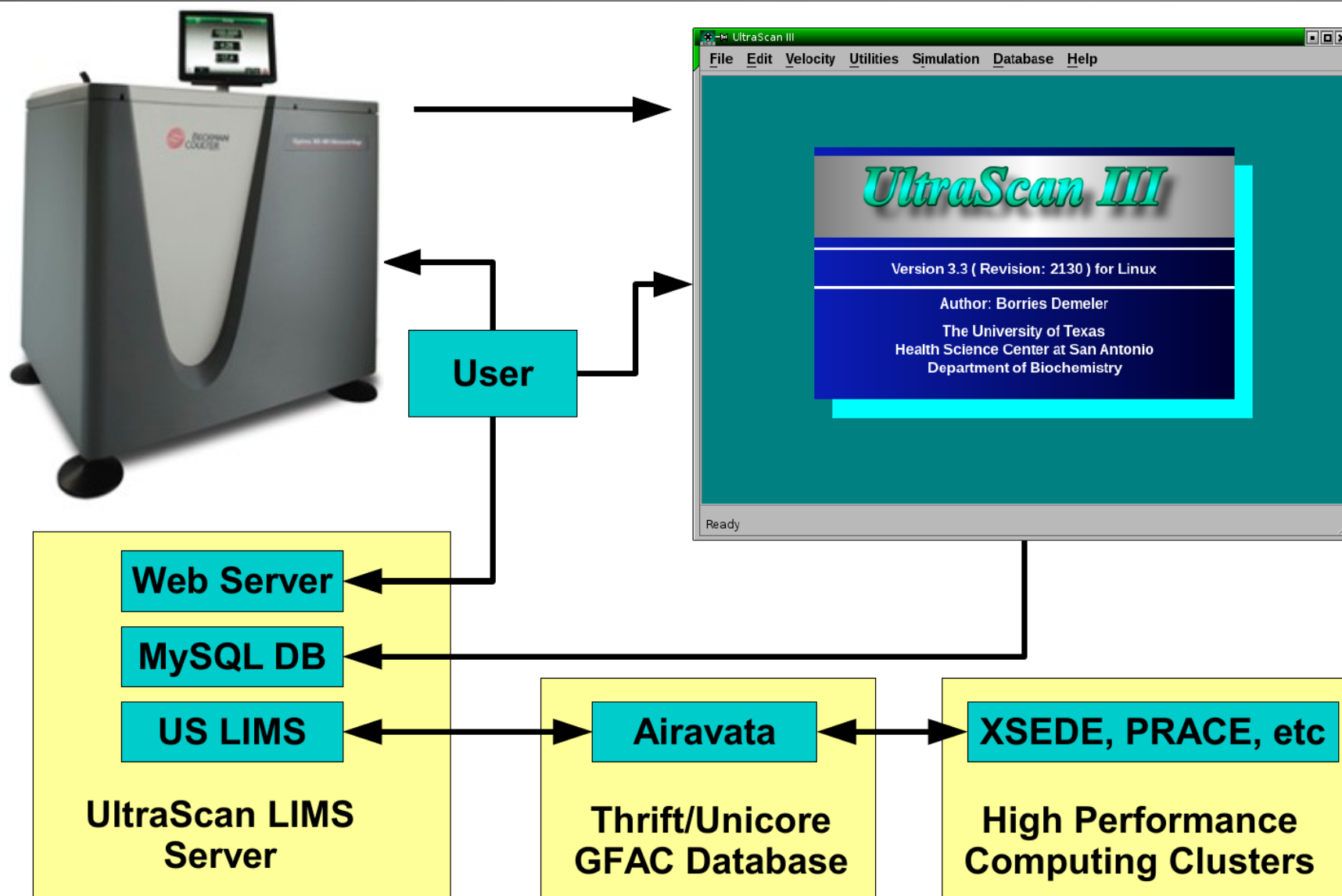


# Application of MWL-AUC to complex systems

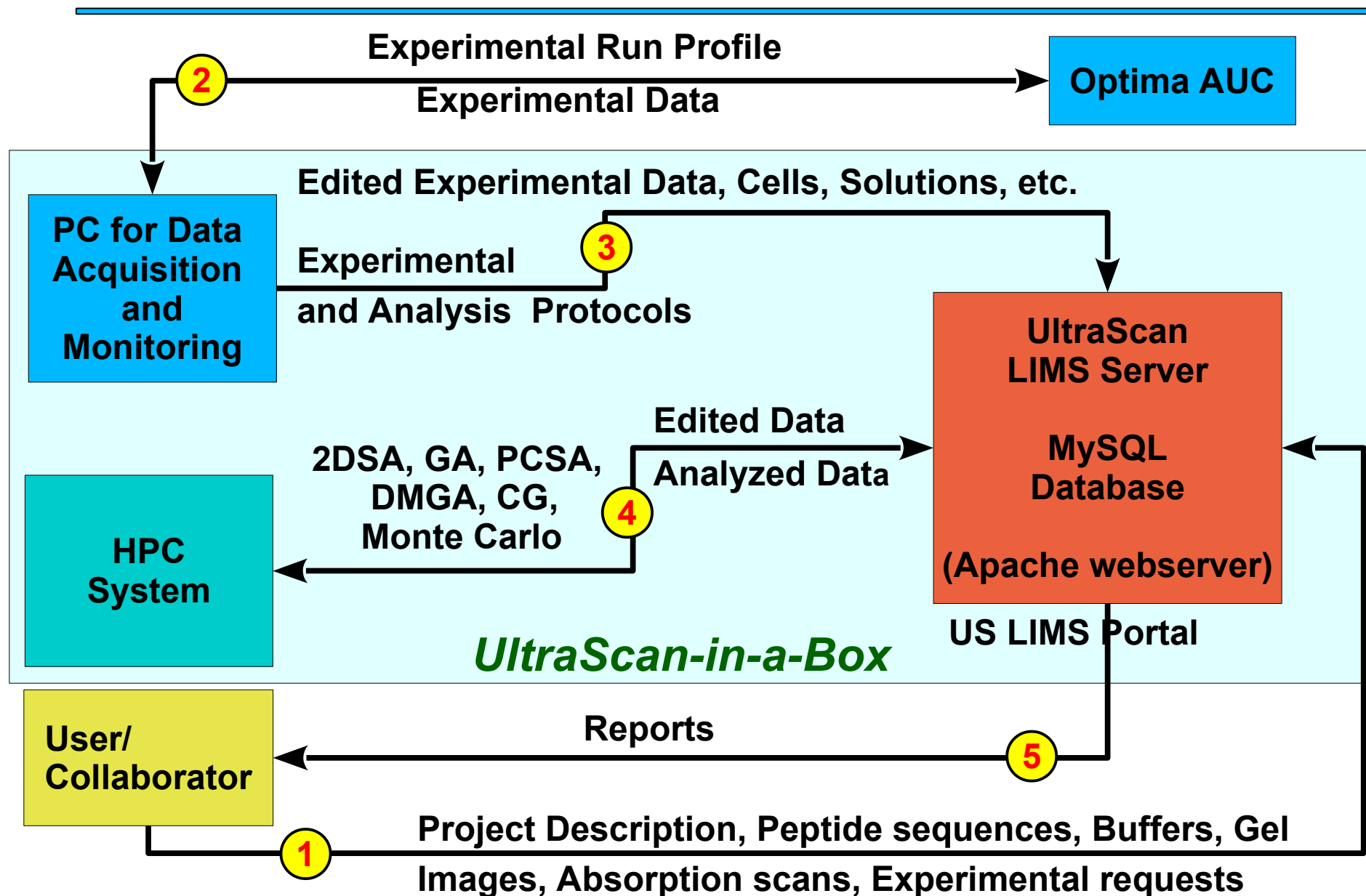
## UltraScan AUC Analysis Workflow



# *UltraScan Layout (multi-platform, open source)*

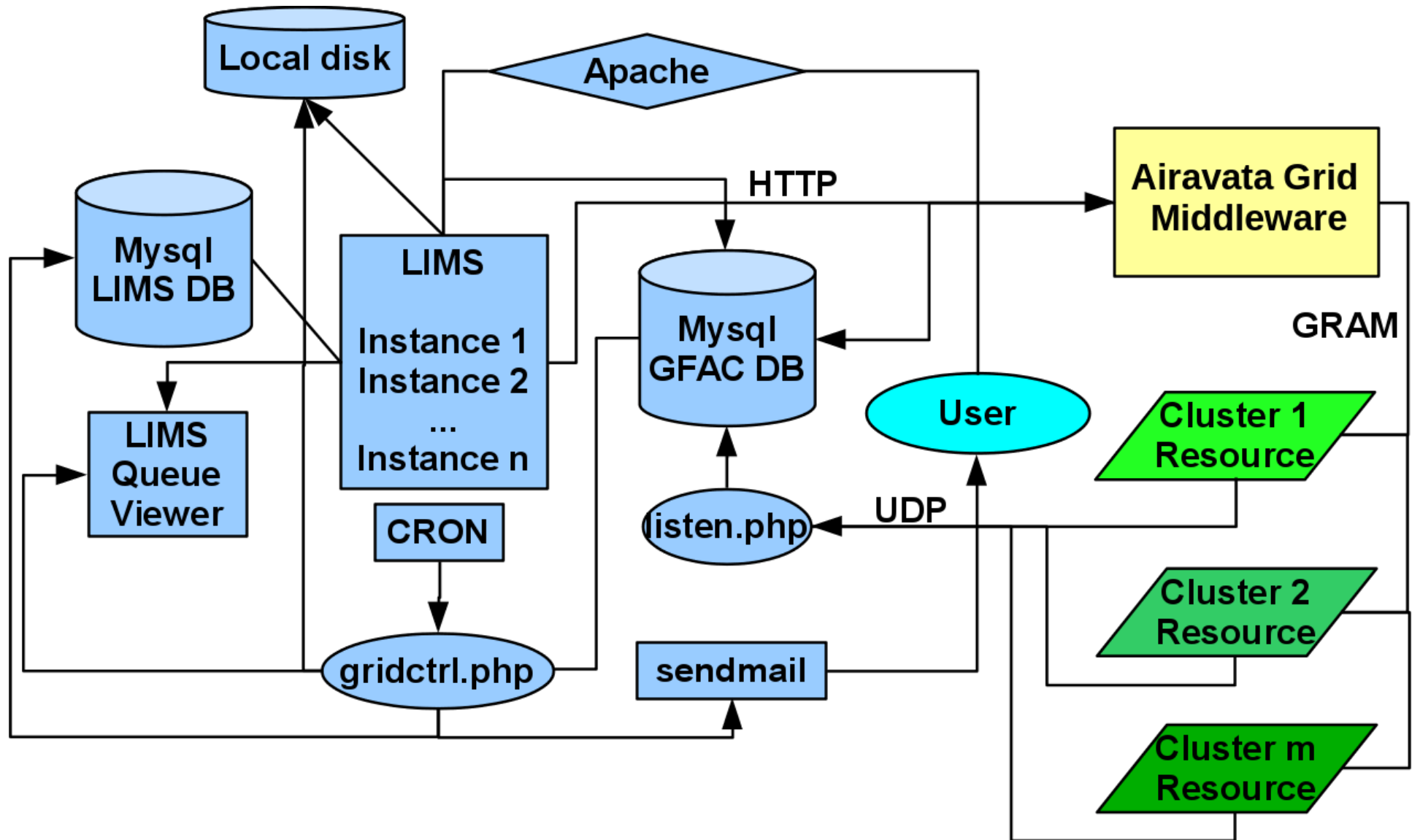


## Future UltraScan Data Flow



More Information: <http://www.aucsolutions.com>

# UltraScan HPC Submission Flowchart





# ***Analytical Ultracentrifugation Background***

## **What can be learned from AUC?**

- Excellent method for characterizing any molecule or molecular interaction in the solution environment – small sample requirement, up to 14 samples can be analyzed. Analysis is based on first principles - No standards are required
- Molecules can be studied in a physiological environment – solution conditions can be adjusted (concentration dependency, effect of pH, ionic strength, buffer type, ligands, oxidation state, temperature, etc.)
- Very large size range ( $10^2$  –  $10^8$  Dalton)
- Dynamics - measure oligomerization states of reversible self- or hetero-associations, ligand binding, slow kinetics and  $K_d$
- Composition analysis – number of components, their partial concentration, molecular weight, and anisotropy
- Conformational analysis - folding/melting studies of biopolymers, conformational changes based on changes in solution properties

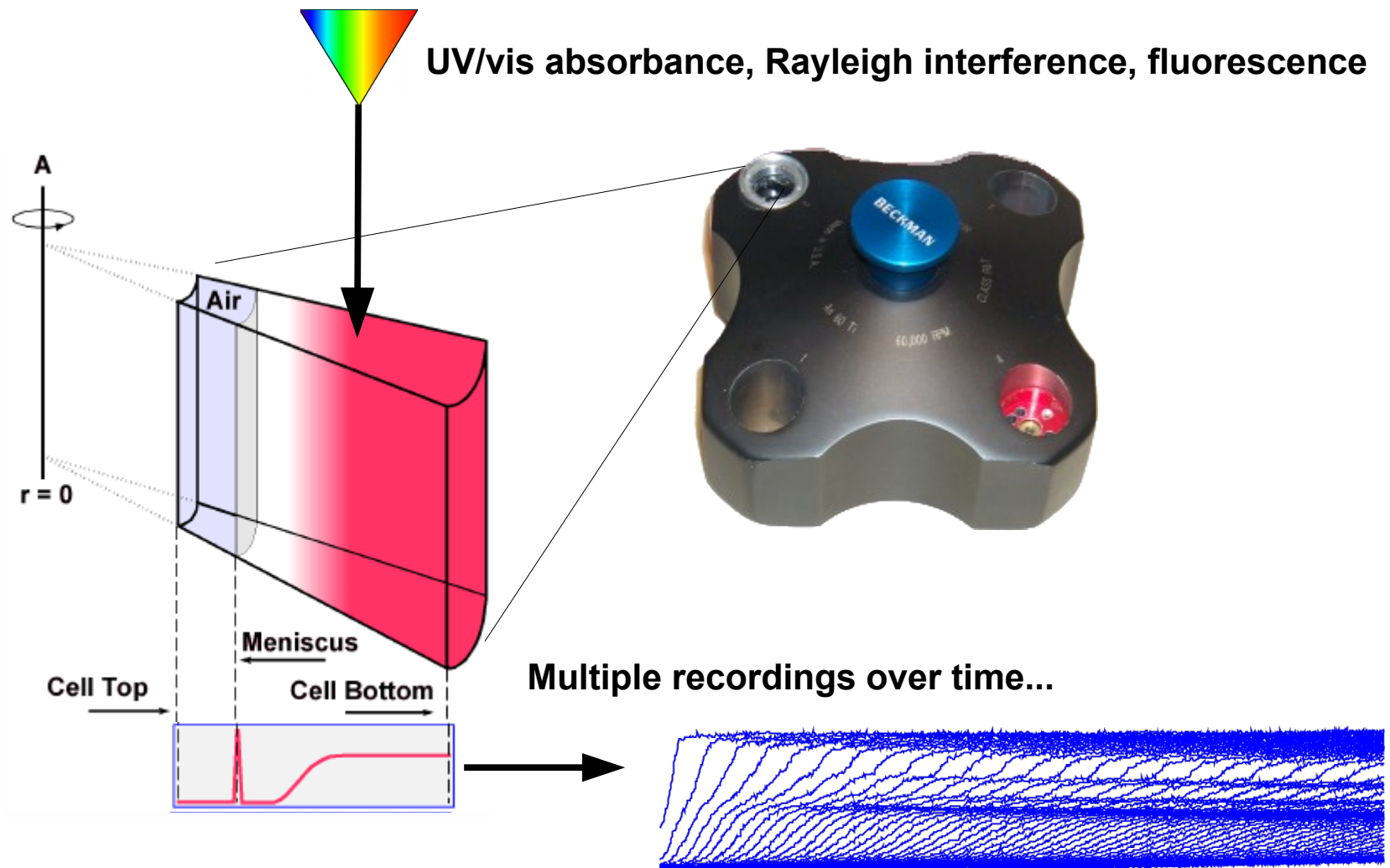
# ***Analytical Ultracentrifugation Background***

## **Available Optical Systems:**

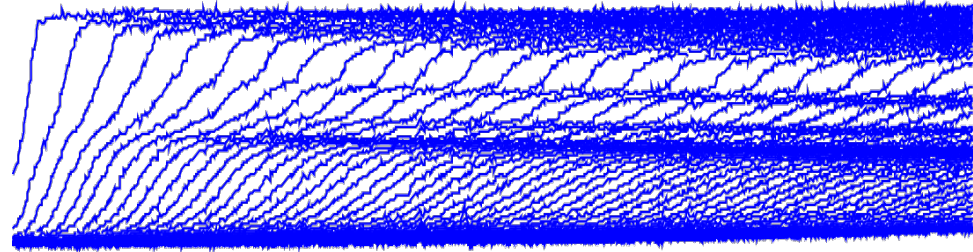
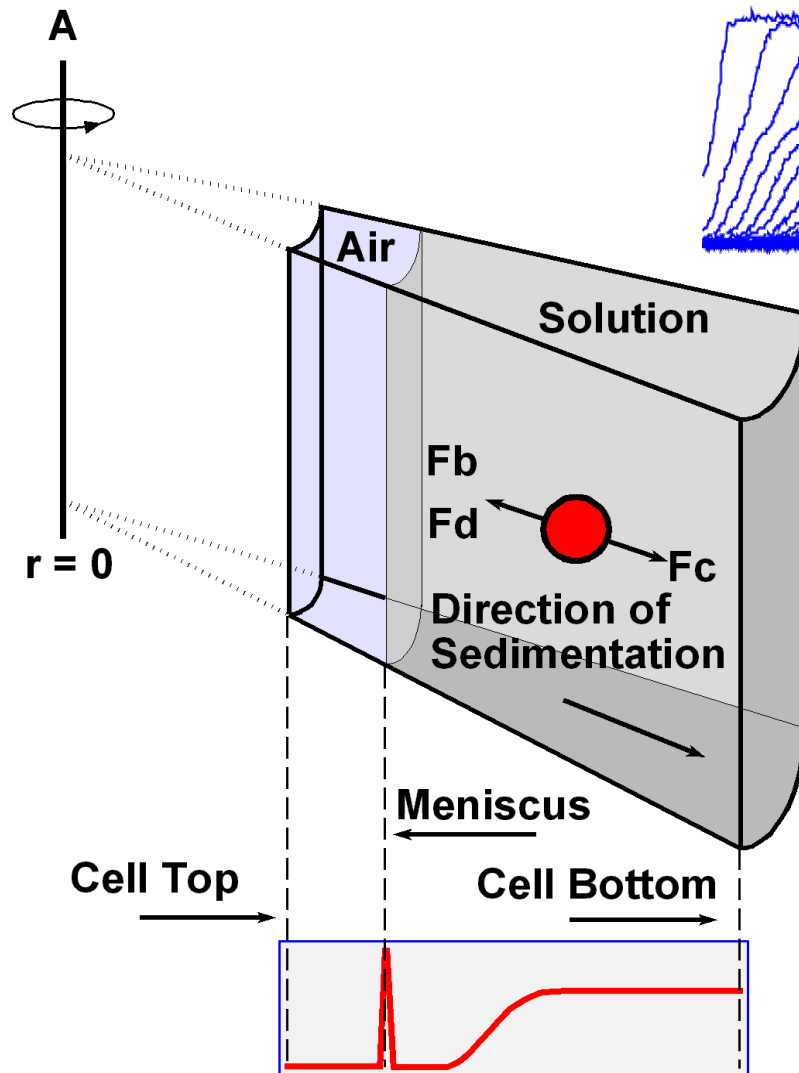
**Multiple detectors extend the range of AUC applications:**

- **UV/vis absorbance (single wavelength):** all-purpose detector for biopolymers and materials absorbing in the visible. Slower detection, good for low protein concentration – measure intensity of transmitted light.
- **Fluorescence:** Study molecules with intrinsic fluorophores or eGFP fusions in impure cell extracts, exquisite selectivity for binding experiments, add fluorescently labeled antibodies to an impure cell extract, study the order of assembly in a multi-domain protein complex – measure intensity of fluorescence emission with a confocal microscope setup (488 nm excitation)
- **Interference:** fast data acquisition and measurement of non-absorbing molecules, carbohydrates, high-concentration studies – measure refractive index differences at 675 nm
- **Multi-wavelength UV/vis:** obtain additional spectral dimension in addition to hydrodynamic separation for independent characterization of molecular properties – measures intensity of transmission of multiple wavelengths

## AUC Instrumentation



# Sedimentation Review



## Sedimentation:

### Forces at Equilibrium:

$$F_c - F_b - F_d = 0$$

$$F_b \text{ (buoyancy)} = \omega^2 r m_s$$

$$F_d \text{ (viscous drag)} = f v$$

$$F_c \text{ (centrifugal force)} = \omega^2 r m$$

### Explanation:

$F_b$  is the buoyancy force - the force required to displace the buffer surrounding the solute, and  $m_s$  is the mass of the displaced solvent.



<b>F<sub>b</sub> (buoyancy)</b>	<b>= <math>\omega^2 r m_s</math></b>
<b>F<sub>d</sub> (viscous drag)</b>	<b>= <math>f v</math></b>
<b>F<sub>c</sub> (centrifugal force)</b>	<b>= <math>\omega^2 r m</math></b>

Substitute the mass of the solvent,  $m_s$ ,  
with the mass of the solute,  $m$

$$m_s = m \bar{v} \rho, \quad F_b = \omega^2 r m \bar{v} \rho$$

Rearrange the force equation:  
 $F_c - F_b - F_d = 0$  and substitute

$$\omega^2 r m - \omega^2 r m \bar{v} \rho = f v$$

Place molecular parameter on one side  
and experimental parameters on the other

$$\frac{m (1 - \bar{v} \rho)}{f} = \frac{v}{\omega^2 r}$$

Put into molar units by multiplying with  
Avogadro's number,  $N$

$$\frac{M (1 - \bar{v} \rho)}{N f} = \frac{v}{\omega^2 r} = s$$

## Partial Specific Volume and Buoyancy

**Sedimentation Forces:  $F_c - F_b - F_d = 0$**

$$F_b \text{ (buoyancy)} = \omega^2 r m_o$$

$$F_d \text{ (viscous drag)} = fv$$

$$F_c \text{ (centrifugal force)} = \omega^2 r m$$

$$\frac{M(1 - \bar{v}\rho)}{Nf} = \frac{v}{\omega^2 r} = s$$

**We need:**

**Partial specific volume ( $\bar{v}$ ) of solute (we can estimate this from the protein sequence or measure in a densitometer – UltraScan will estimate from protein sequence automatically)**

**Viscosity of solvent (we can obtain this measure from the known composition of the buffer or measure in a viscometer – UltraScan will estimate this quantity from buffer composition automatically)**

**Density of solvent (we can obtain this measure from the known composition of the buffer or measure it in a densitometer – UltraScan will estimate this quantity from buffer composition automatically)**

**Temperature of experiment (recorded by instrument)**

**Rotor speed (recorded by instrument)**

$$\frac{M(1 - \bar{v}\rho)}{Nf} = \frac{v}{\omega^2 r} = s$$

***Definition:***

The sedimentation velocity,  $v$ , divided by the centrifugal field strength,  $\omega^2 r$ , is equal to the sedimentation coefficient,  $s$

***Take-home message:***

The sedimentation coefficient is proportional to  $M$  and inversely proportional to  $f$

## Viscosity and Density Corrections

The density and also the viscosity of the solvent affect the sedimentation and diffusion of the particle in solution, so the measured values need to be corrected to standard conditions. Moreover, temperature and buffer composition affect the solvent density and viscosity, so they need to be considered. To correct for density and viscosity, use these formulas:

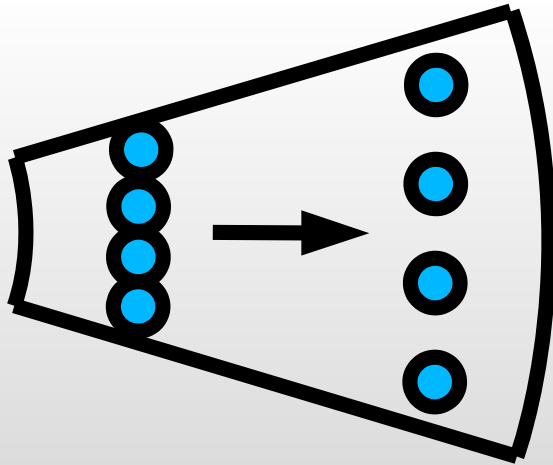
$$s_{20,W} = s_{T,B} \frac{(1 - \bar{v} \rho)_{20,W} \eta_{T,B}}{(1 - \bar{v} \rho)_{T,B} \eta_{20,W}}$$

$$D_{20,W} = D_{T,B} \frac{293.15 \eta_{T,B}}{T \eta_{20,W}}$$

**UltraScan will automatically apply these corrections  
for aqueous buffers!**



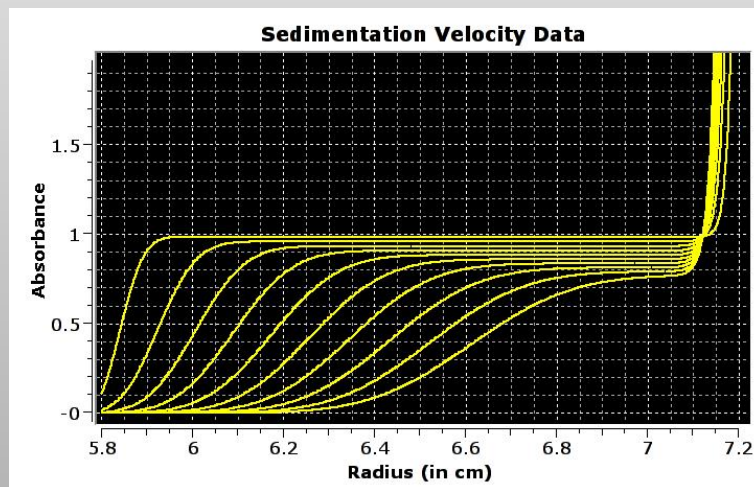
# Radial Dilution



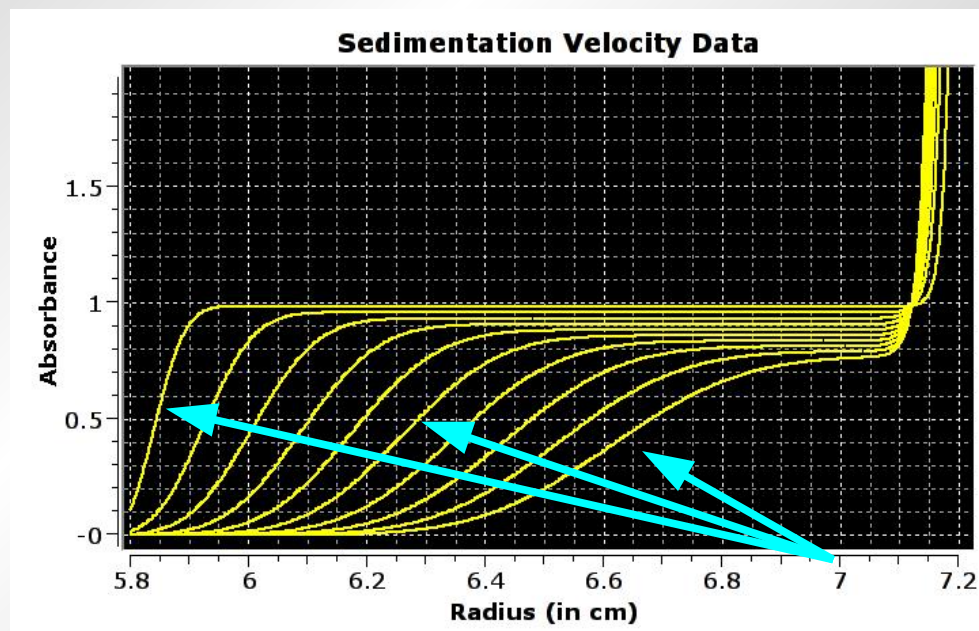
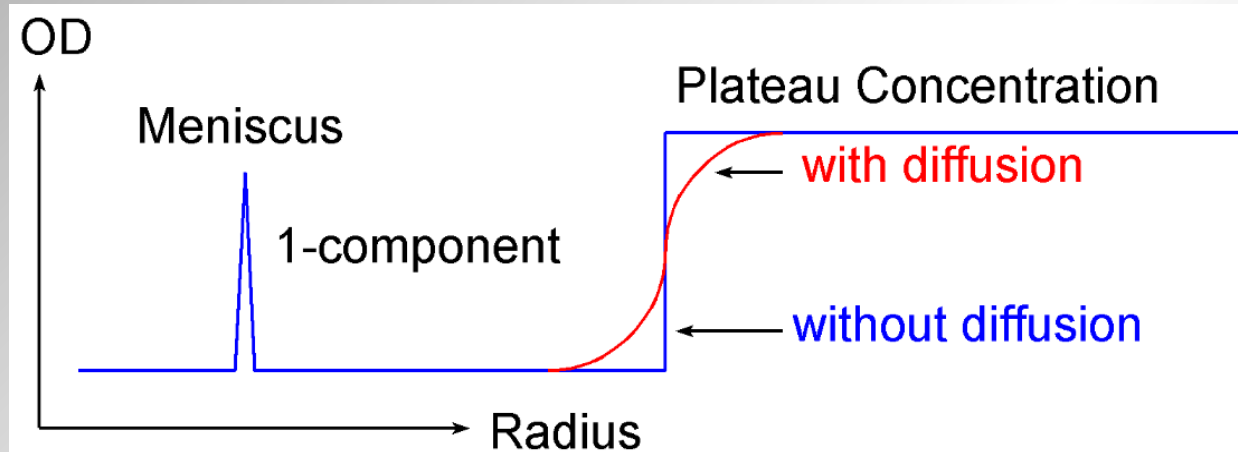
Radial Dilution occurs because of the cell's sector shape. Molecules sedimenting towards the outside of the cell will dilute as they sediment.

All molecules - no matter at what position they are - will dilute the same amount over a given distance, the speed of dilution is proportional to their sedimentation speed. This causes a reduction in the observed optical density.

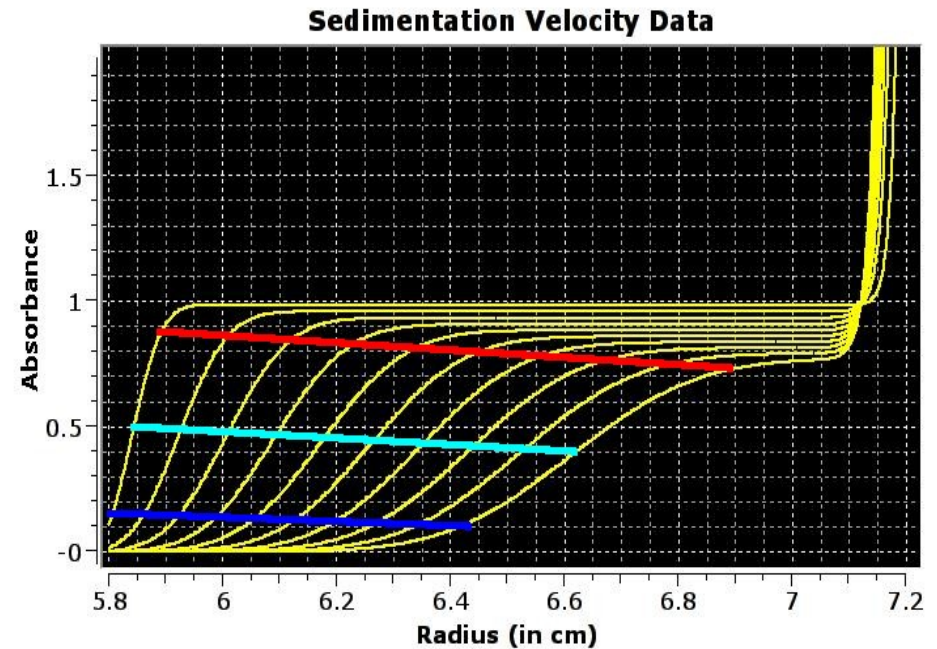
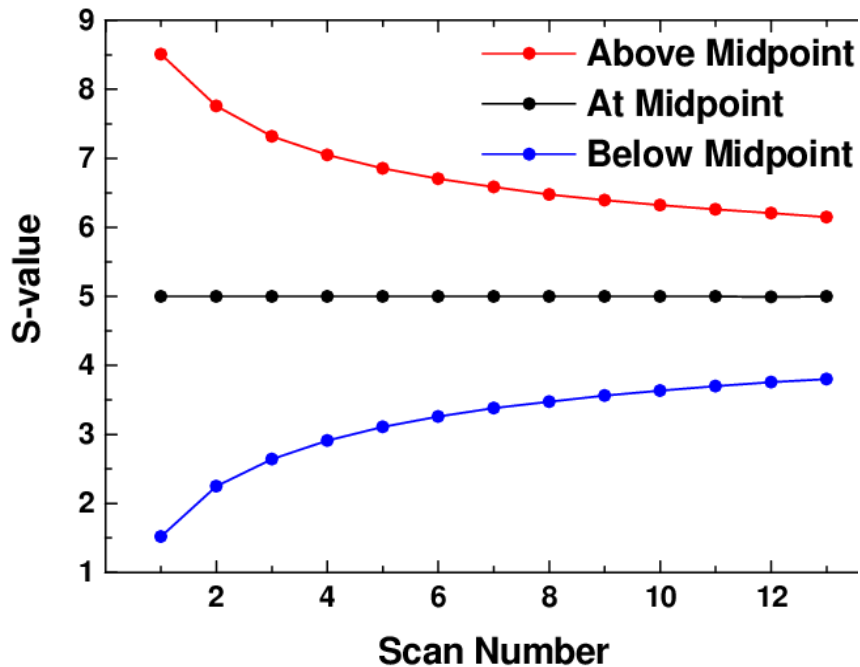
Radial Dilution can be observed through a reduction in the plateau absorbance in successive scans



## Effect of Diffusion on the Sedimentation Boundary



## Calculation of the Sedimentation Coefficient:



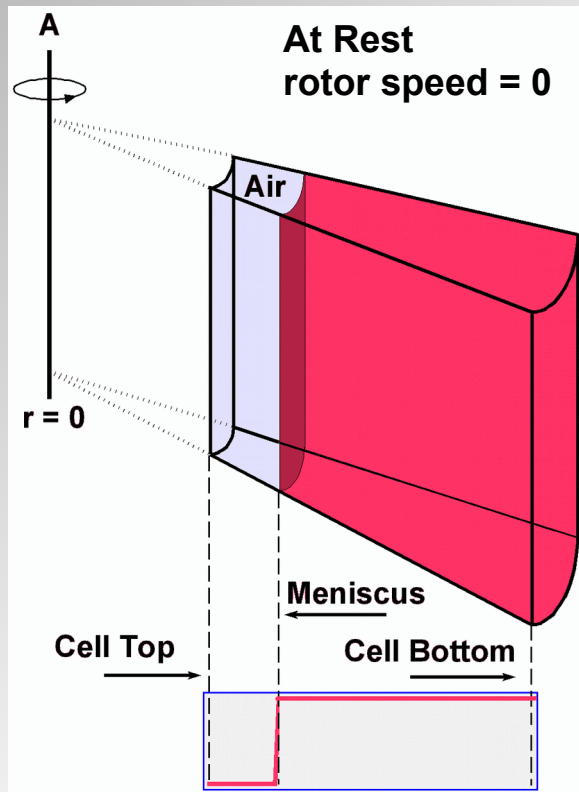
$$\frac{M(1 - \bar{v}\rho)}{Nf} = \frac{v}{\omega^2 r} = s$$

$$v = \frac{dr}{dt} \quad \frac{dr}{r} = \omega^2 s dt$$

$$\int_{r=m}^{r=b} \frac{1}{r} dr = \int_{t=0}^{t=scan} s \omega^2 dt$$

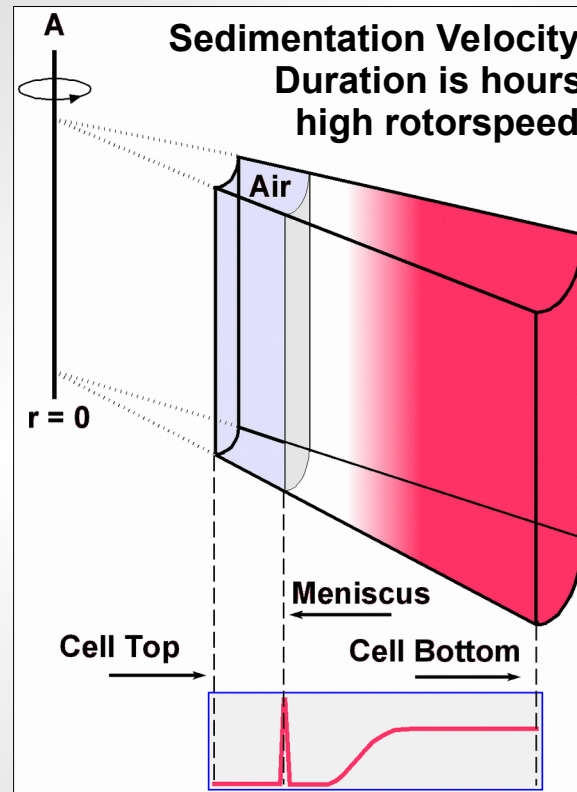
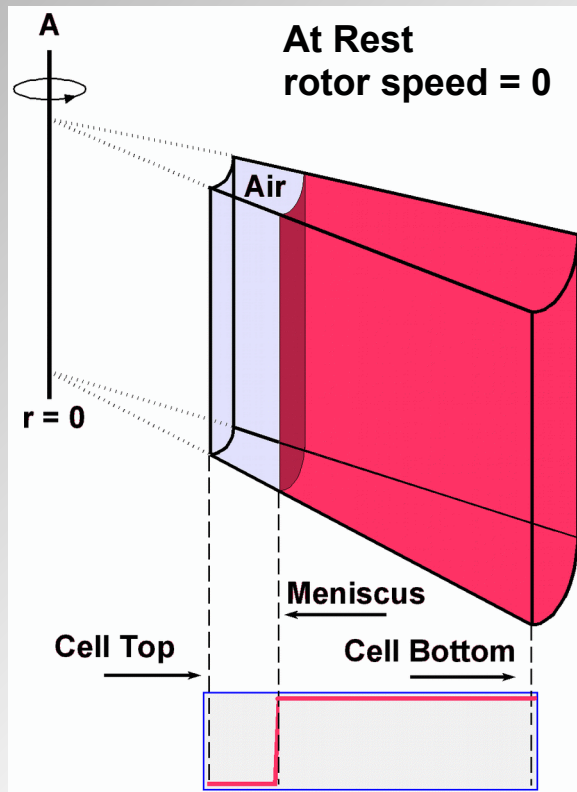
$$\hat{s}_b = \ln \left( \frac{r_b(t)}{r_m(t_0)} \right) (\omega^2 (t - t_0))^{-1}$$

## Transport Processes – Sedimentation and Diffusion



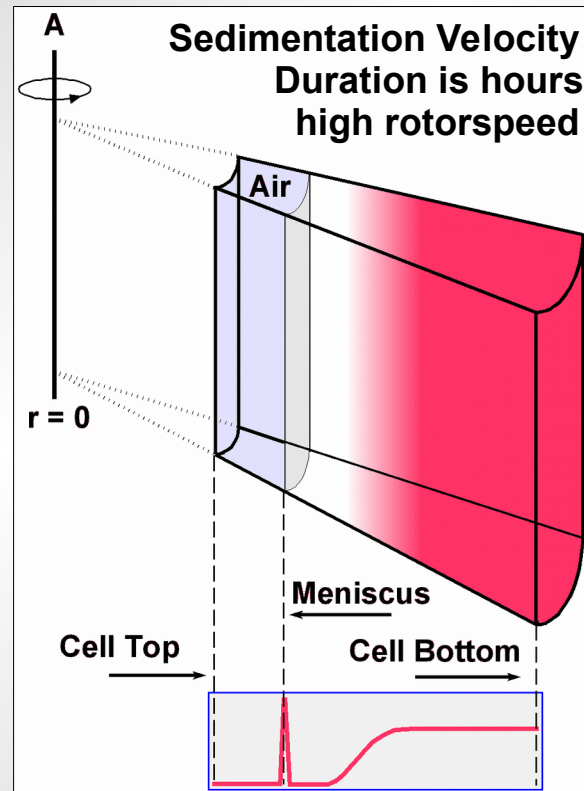
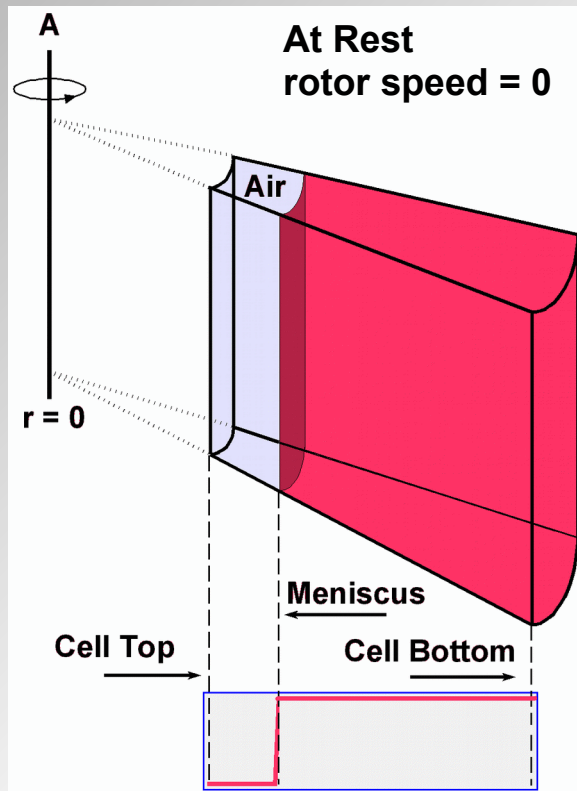


# Transport Processes – Sedimentation and Diffusion



## Transport Processes – Sedimentation and Diffusion

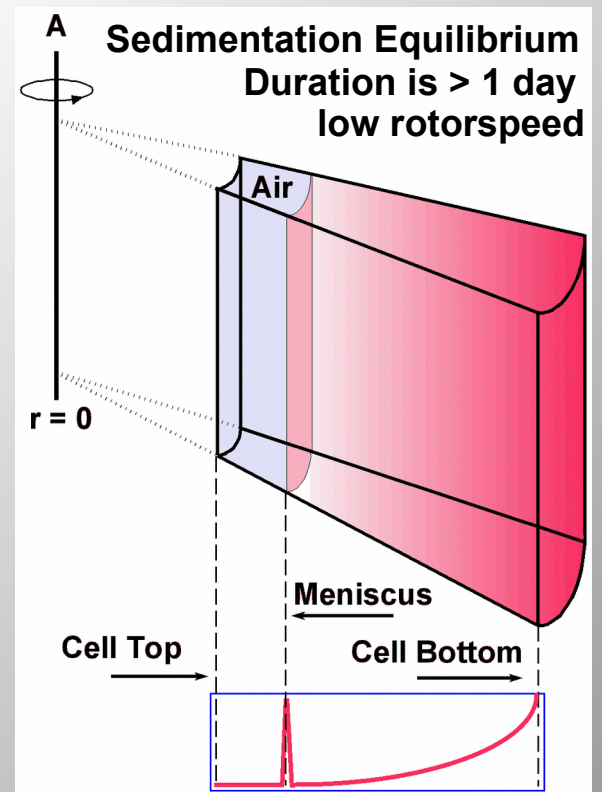
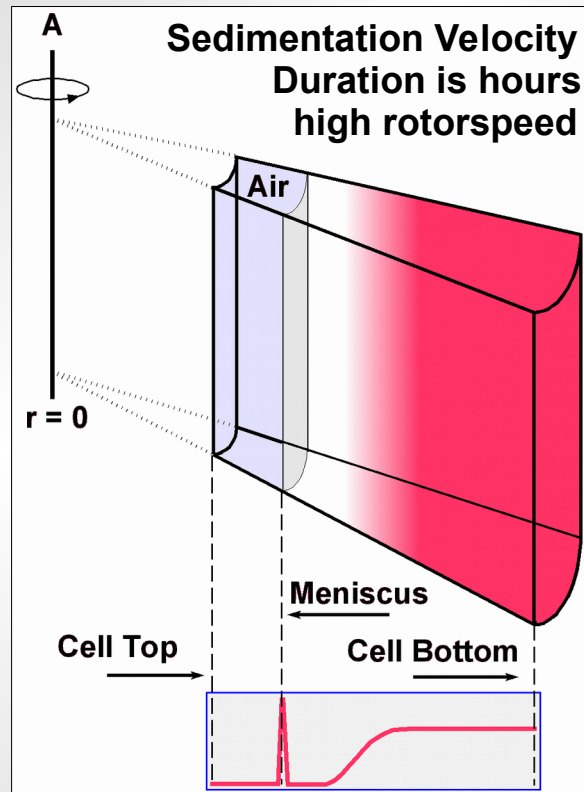
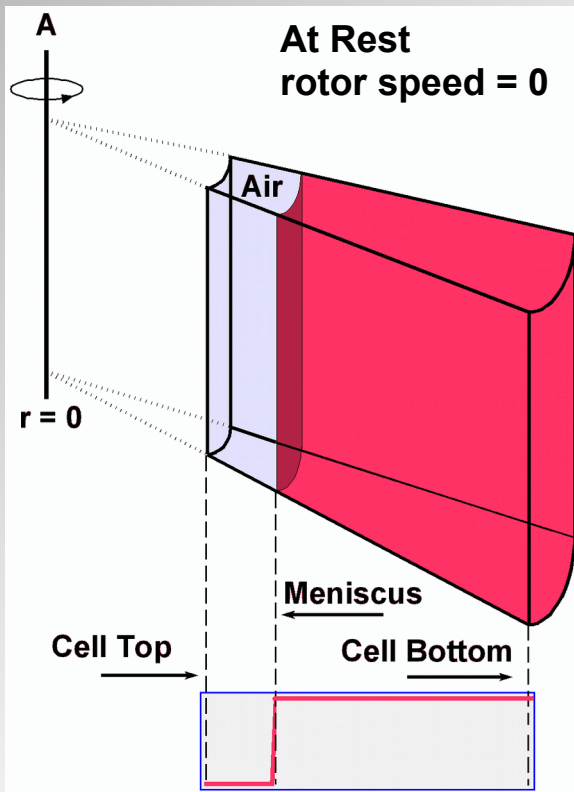
$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$



# Transport Processes – Sedimentation and Diffusion

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

$$J = s \omega^2 r C - D \frac{\partial C}{\partial r} = 0$$

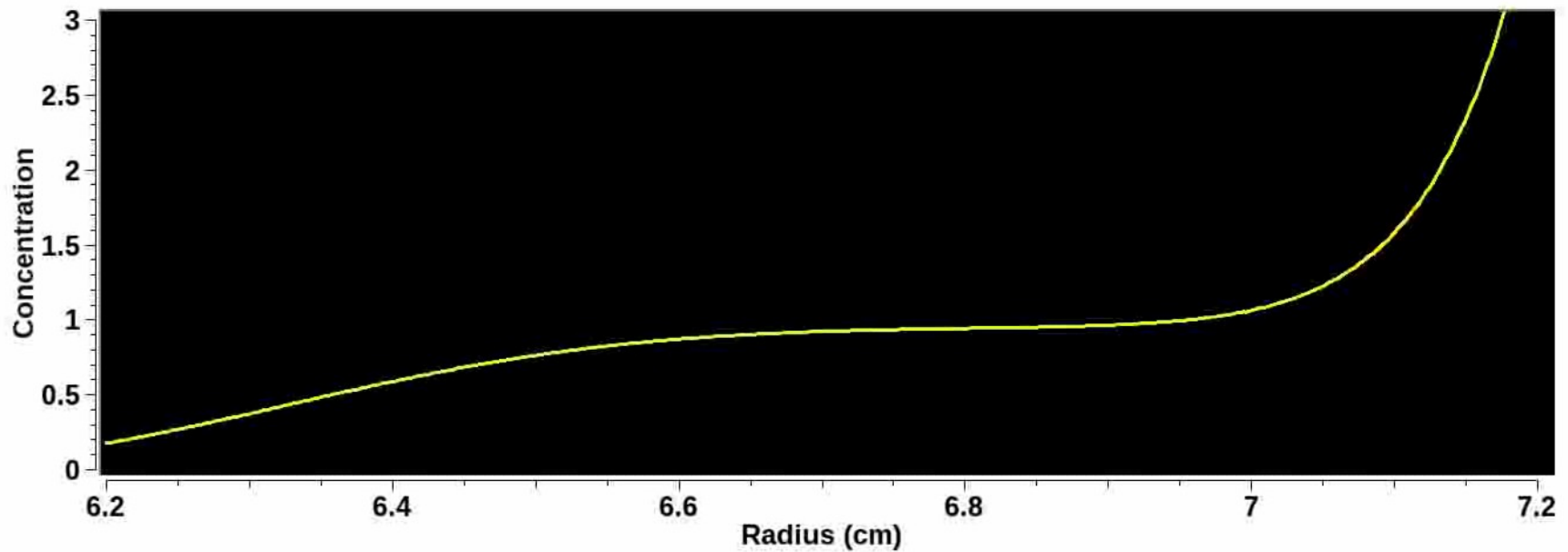


## Background

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

$$J = s \omega^2 r C - D \frac{\partial C}{\partial r} = 0$$

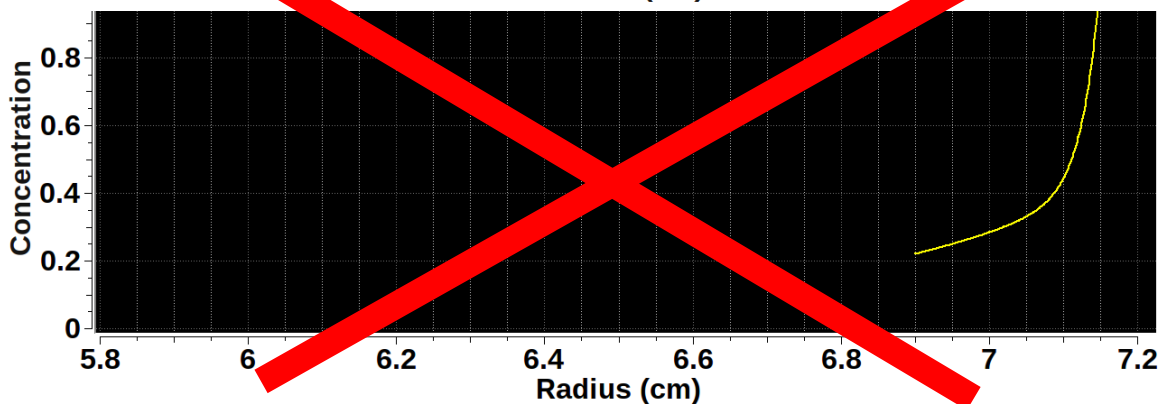
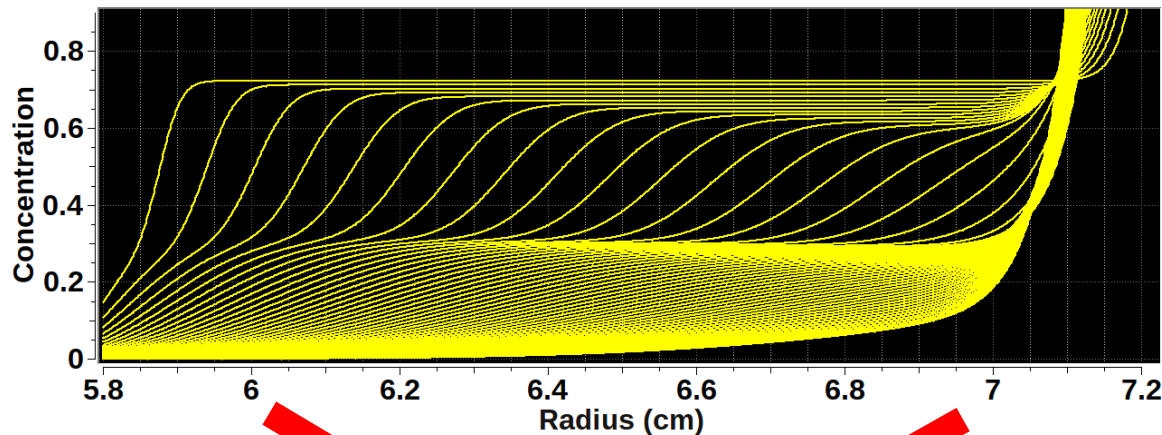
$$C = C_0 e^{\left[\sigma(r^2 - r_m^2)\right]}$$



## Background

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

$$J = s \omega^2 r C - D \frac{\partial C}{\partial r} = 0$$



$$C = C_0 e^{\left[\sigma(r^2 - r_m^2)\right]}$$

**Sedimentation  
Velocity**

or

**Sedimentation  
Equilibrium?**

## Relationship between $M$ , $f$ , $\varphi$ , $s$ , and $D$

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ \underbrace{s \omega^2 r^2 C}_{\text{Sedimentation}} - \underbrace{D r \frac{\partial C}{\partial r}}_{\text{Diffusion}} \right]_t$$

Concentration
Sedimentation
Diffusion

$$f = \frac{RT}{N D}$$

$$M = \frac{s N f}{1 - \bar{v} \rho}$$

$$V = \frac{M \bar{v}}{N}$$

$$r_0 = \left( \frac{3 V}{4 \pi} \right)^{1/3}$$

$$f_0 = 6 \pi \eta r_0$$

$$\varphi = \frac{f}{f_0}$$



# Relationship between $M$ , $f$ , $\varphi$ , $s$ , and $D$

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

Concentration

Sedimentation

Diffusion

Shape  
Dependence

$$\begin{aligned} f &= \frac{RT}{N D} \\ M &= \frac{s N f}{1 - \bar{v} \rho} \\ V &= \frac{M \bar{v}}{N} \\ r_0 &= \left( \frac{3 V}{4 \pi} \right)^{1/3} \\ f_0 &= 6 \pi \eta r_0 \\ \varphi &= \frac{f}{f_0} \end{aligned}$$

# Relationship between $M$ , $f$ , $\phi$ , $s$ , and $D$

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

Concentration

Sedimentation

Diffusion

Size

$$\begin{aligned}
 f &= \frac{RT}{N D} \\
 M &= \frac{s N f}{1 - \bar{v} \rho} \\
 V &= \frac{M \bar{v}}{N} \\
 r_0 &= \left( \frac{3 V}{4 \pi} \right)^{1/3} \\
 f_0 &= 6 \pi \eta r_0 \\
 \phi &= \frac{f}{f_0}
 \end{aligned}$$

## Relationship between $M$ , $f$ , $\varphi$ , $s$ , and $D$

$$\left( \frac{\partial C}{\partial t} \right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

Concentration

Sedimentation

Diffusion

$$f = \frac{RT}{N D}$$

$$M = \frac{s N f}{1 - \bar{v} \rho}$$

$$V = \frac{M \bar{v}}{N}$$

$$r_0 = \left( \frac{3 V}{4 \pi} \right)^{1/3}$$

$$f_0 = 6 \pi \eta r_0$$

Anisotropy —  $\varphi = \frac{f}{f_0}$

## Relationship between $M$ , $f$ , $\varphi$ , $s$ , and $D$

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ \underbrace{s \omega^2 r^2 C}_{\text{Sedimentation}} - \underbrace{D r \frac{\partial C}{\partial r}}_{\text{Diffusion}} \right]_t$$

Concentration
Sedimentation
Diffusion

$$f = \frac{RT}{N D}$$

$$M = \frac{s N f}{1 - \bar{v} \rho}$$

$$V = \frac{M \bar{v}}{N}$$

$$r_0 = \left( \frac{3 V}{4 \pi} \right)^{1/3}$$

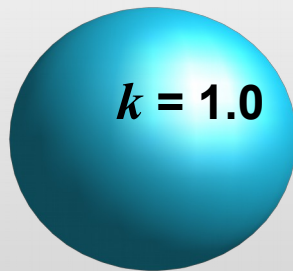
$$f_0 = 6 \pi \eta r_0$$

$$\varphi = \frac{f}{f_0}$$

Partial specific  
volume and  
density

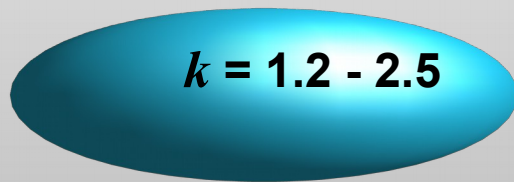
## ***Parametrization of Anisotropy (Shape)***

The frictional ratio  $f/f_0 = k$  is a convenient way to parameterize the diffusion coefficient and the shape of a molecule .



$$k = 1.0$$

The frictional ratio  $k$  is 1.0 for a sphere since  $f = f_0$  and hence  $k$  has a convenient lower limit



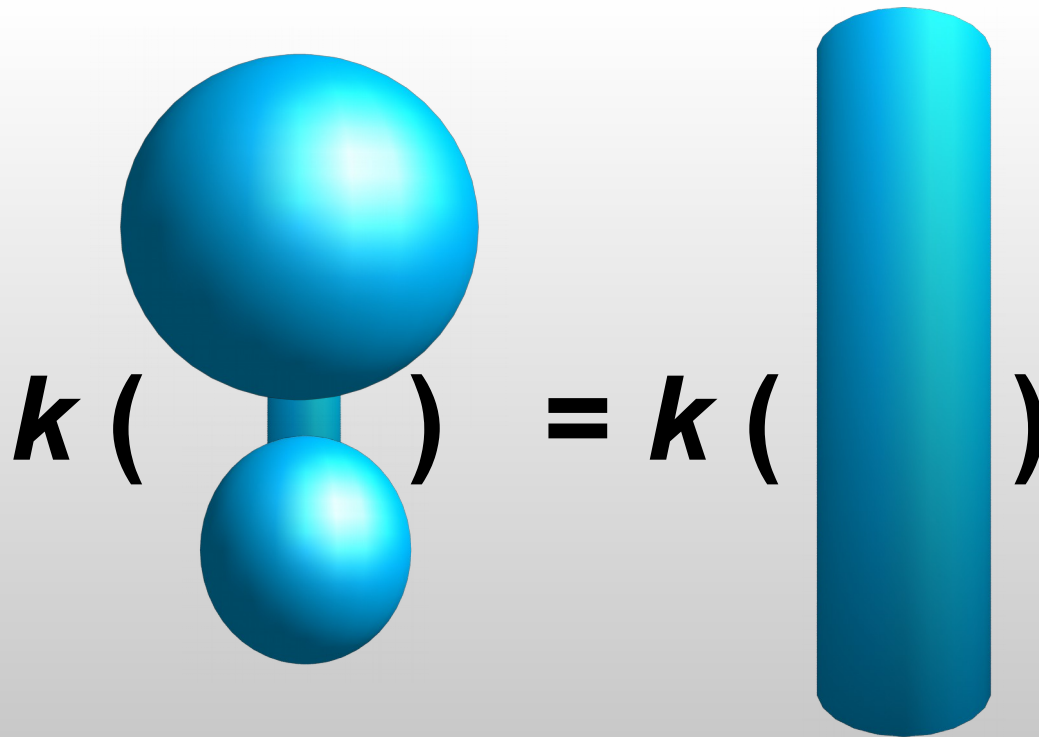
$$k = 1.2 - 2.5$$

$1.0 \leq k \leq 4.0$  for most proteins, higher for rod-shaped, disordered and unfolded proteins, DNA, fibrils and aggregates or linear molecules



$$k > 3$$

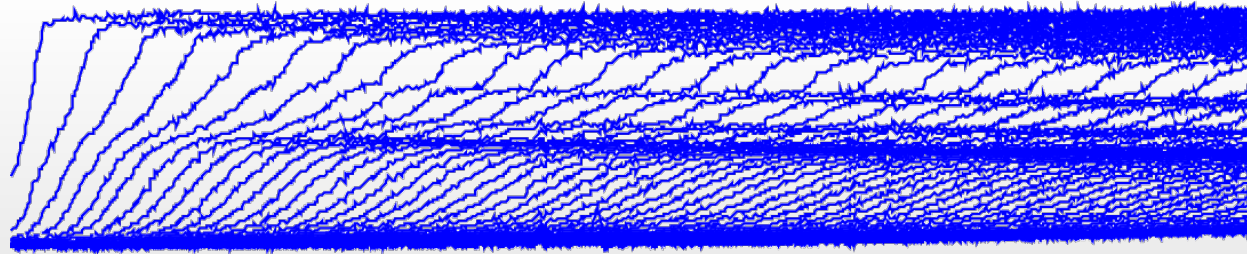
## *Degeneracy of Shape Determination*



***$k$  may be the same for different shapes, we cannot distinguish them by AUC, we can only measure the anisotropy!***

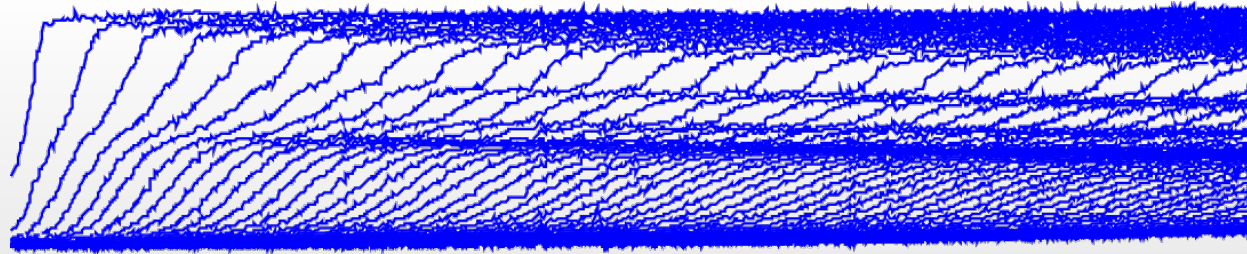


# ***Sedimentation Velocity***



**Sedimentation velocity profile of a mixture of  
macromolecules over time**

# *Sedimentation Velocity*

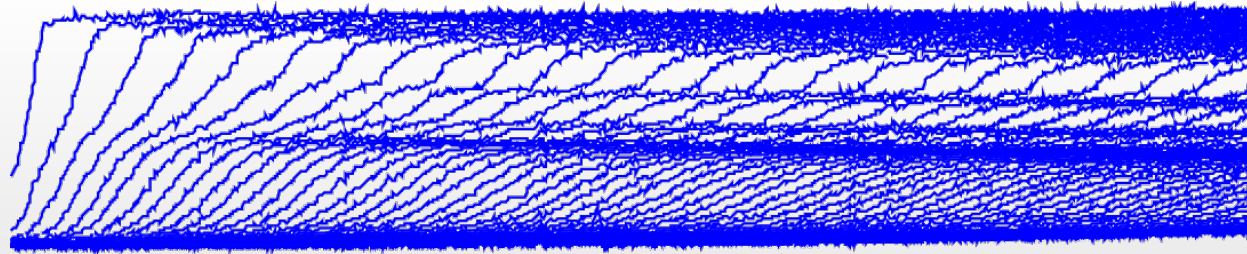


## **Composition Analysis**

**We can answer these questions:**

**How many components?**

# *Sedimentation Velocity*



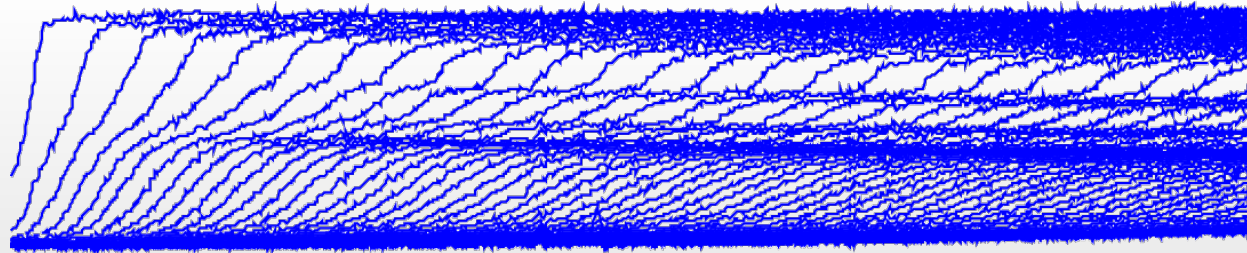
## **Composition Analysis**

**We can answer these questions:**

**How many components?**

**What are their sizes and molecular weights?**

# *Sedimentation Velocity*



## **Composition Analysis**

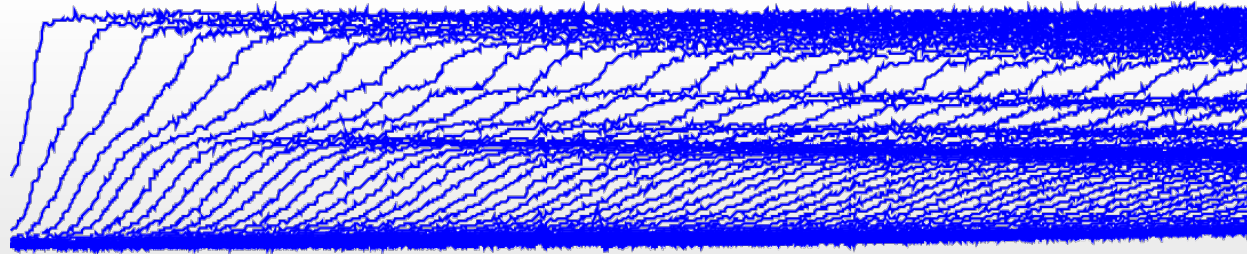
**We can answer these questions:**

**How many components?**

**What are their sizes and molecular weights?**

**What are their anisotropies?**

# *Sedimentation Velocity*



## **Composition Analysis**

**We can answer these questions:**

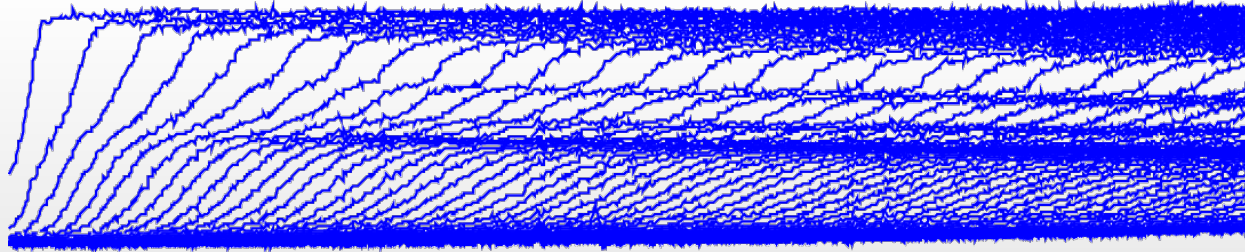
**How many components?**

**What are their sizes and molecular weights?**

**What are their anisotropies?**

**What is the partial concentration of each component?**

# *Sedimentation Velocity*



## **Composition Analysis**

**We can answer these questions:**

**How many components?**

**What are their sizes and molecular weights?**

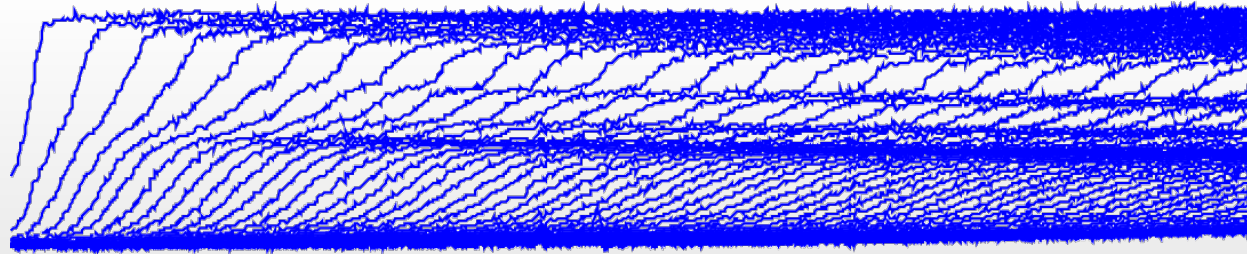
**What are their anisotropies?**

**What is the partial concentration of each component?**

**Do the components interact (how fast, strong)?**



# ***Sedimentation Velocity***



## **Composition Analysis**

**We can answer these questions:**

**How many components?**

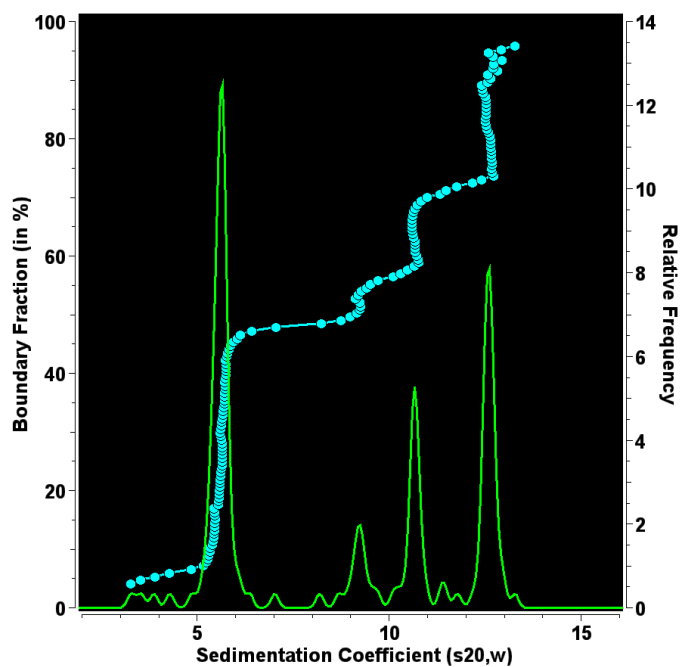
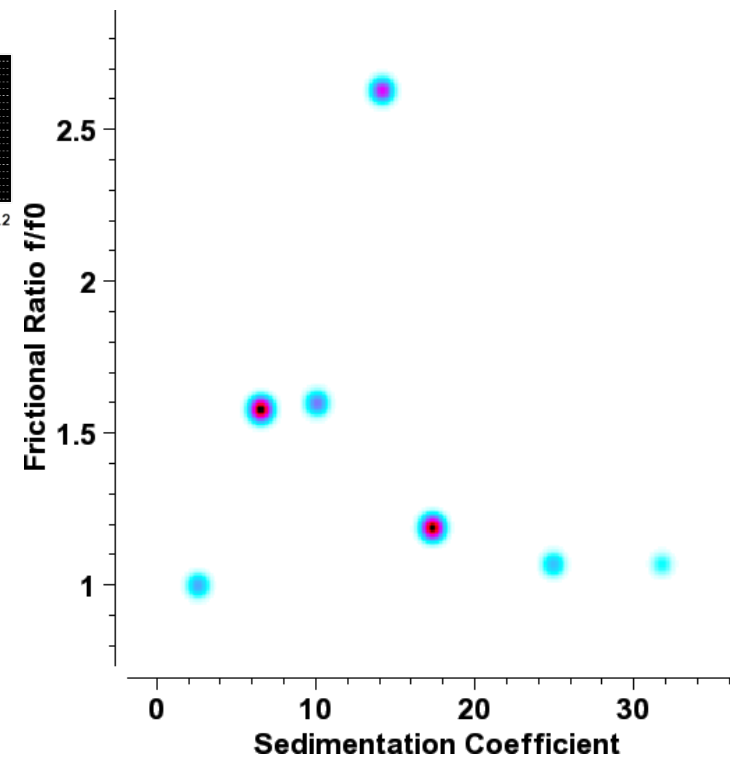
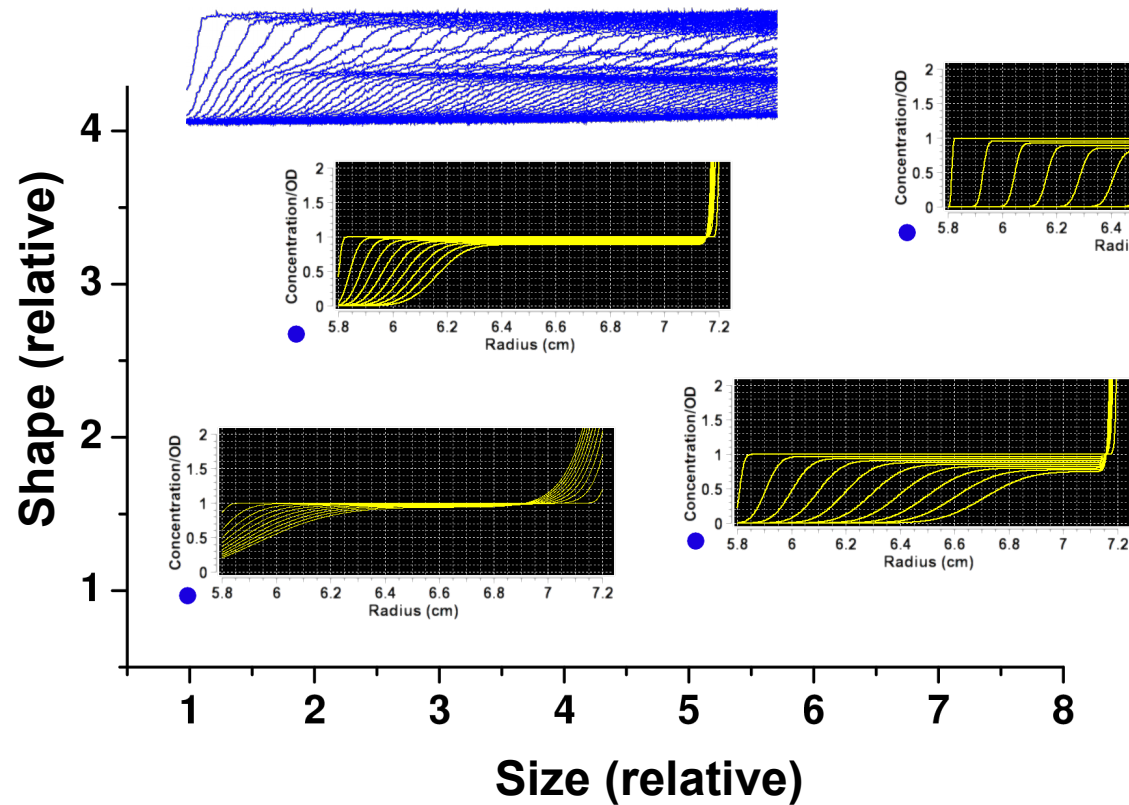
**What are their sizes and molecular weights?**

**What are their anisotropies?**

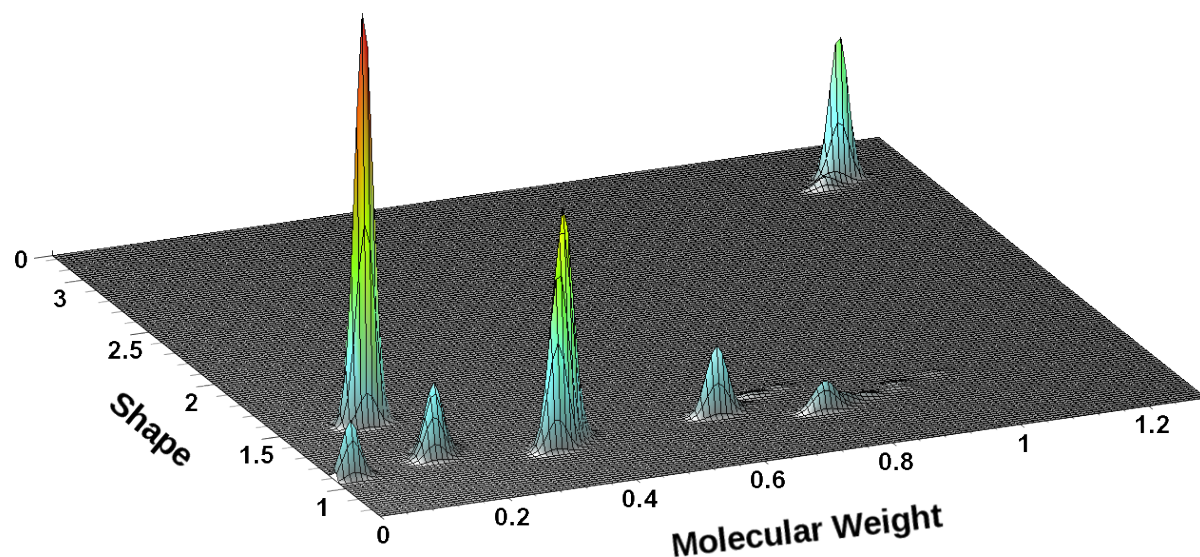
**What is the partial concentration of each component?**

**Do the components interact (how fast, strong)?**

**What is the reliability of our measurement?**



Find: partial concentration,  $s$ ,  $D$ ,  $MW$ ,  $K_d$ ,  $k_{off}$  and  $f/f_0$



## Modeling of Data

A *first principles* approach allows us to model the experimental data by fitting it to a mathematical model. The model represents the physics of the experiment, and contains parameters of interest to the experimentalist. We need to find the values of these parameters by adjusting the model so it matches the data. This is a hard problem called an “*inverse problem*” that requires optimization (fitting) algorithms which aid us in adjusting the parameters so the model fits the data.

In UltraScan this is accomplished by a least squares fitting approach that compares each data point from the model with the corresponding point in the experimental data:

$$\text{Minimize } \sum_{i=1}^N (Data_i - Model_i)^2 \quad (i \text{ over radius and time})$$

Optimally, the difference is zero, but because of experimental noise this never happens, since the model is noise free.

## Noise & Data Modeling Considerations

Fitting of noisy data prevents unique solutions – multiple solutions are possible.

**We need to *minimize noise* when modeling data.**

There are three ways to reduce or eliminate noise:

1. fit the noise
2. maintain an exceptionally well tuned instrument
3. design your experiment to optimize the quality of the data

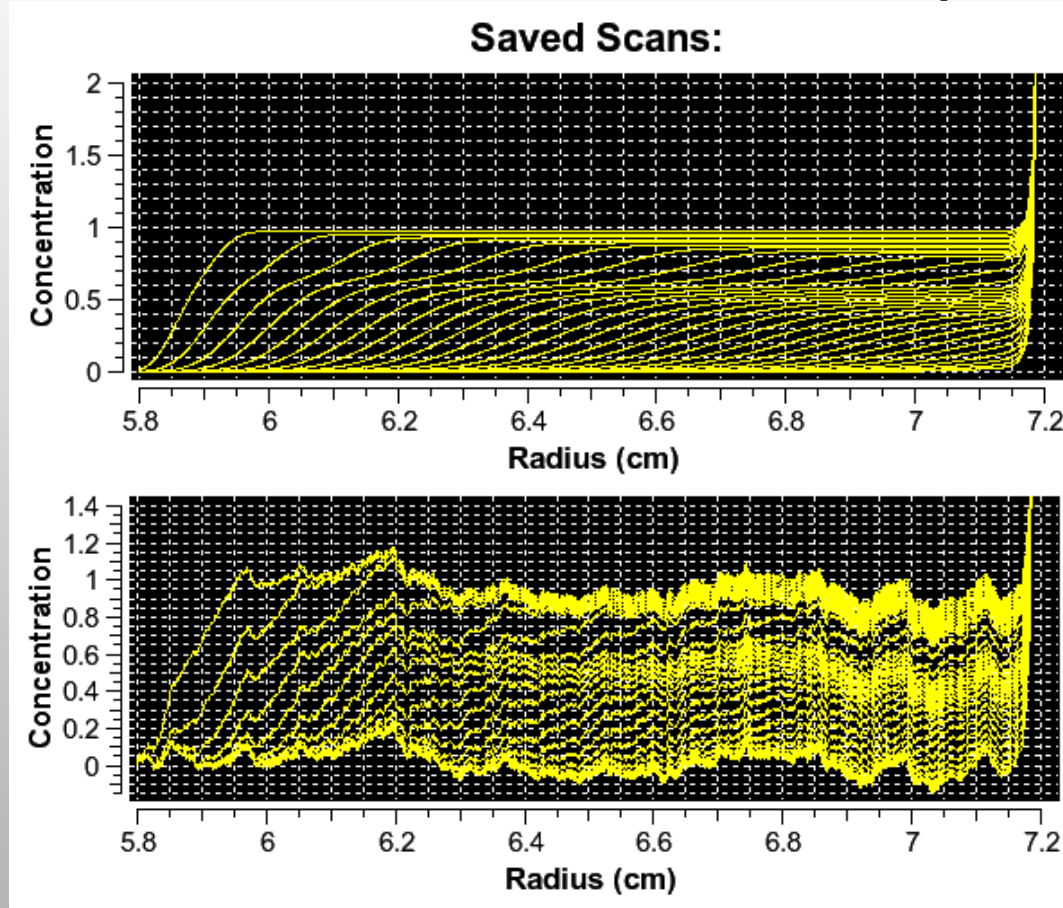
There are three noise types:

1. **Time Invariant noise:** Noise is different for each radial position, but the same offset for each scan, and hence time independent (finger prints).
2. **Radially Invariant noise:** Noise is different for each scan, but each radial position is offset by the same amount throughout the scan (baseline variation)
3. **Stochastic (random noise):** Noise is different for each radial and time point and it is (hopefully) Gaussian in distribution

(1) and (2) can be fitted by UltraScan and removed from the data

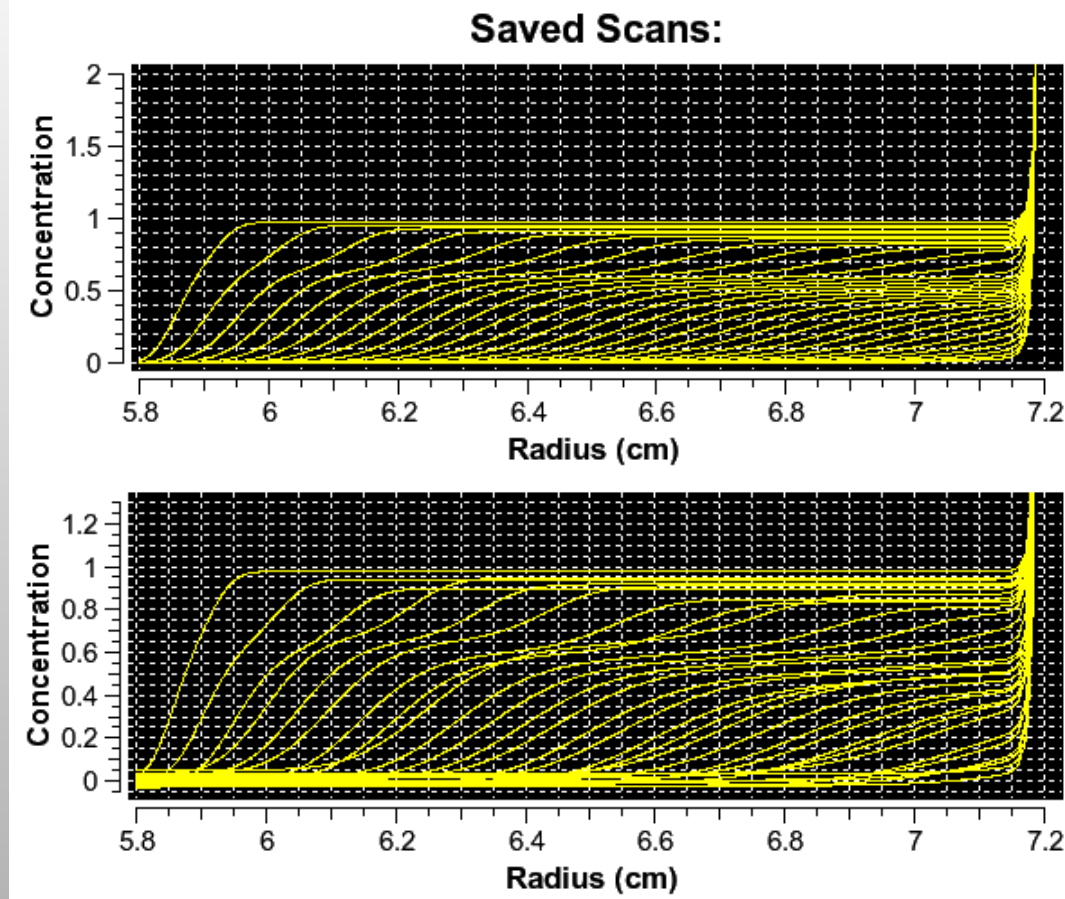
## *Different Types of Noise*

**Time Invariant noise:** Noise is different for each radial position, but the same offset for each scan, and hence time independent.



## *Different Types of Noise*

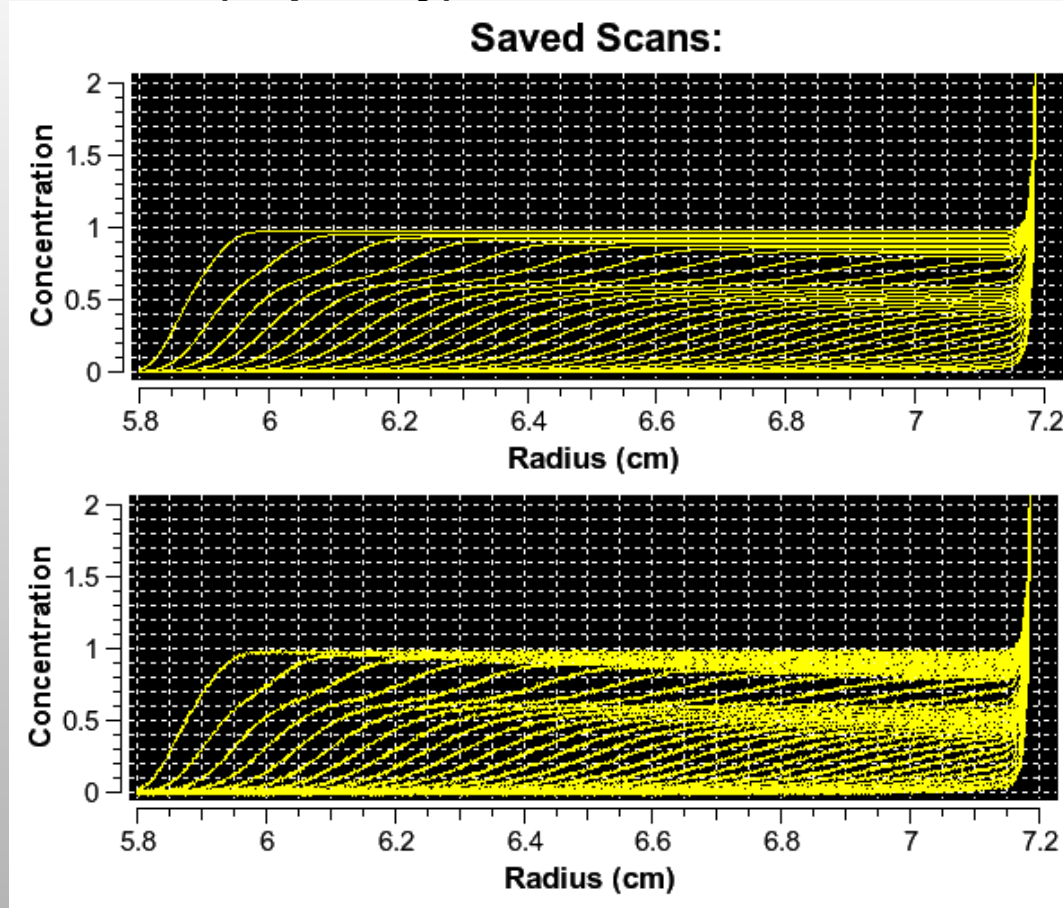
**Radially Invariant noise:** Noise is different for each scan, but each radial position is offset by the same amount throughout the scan



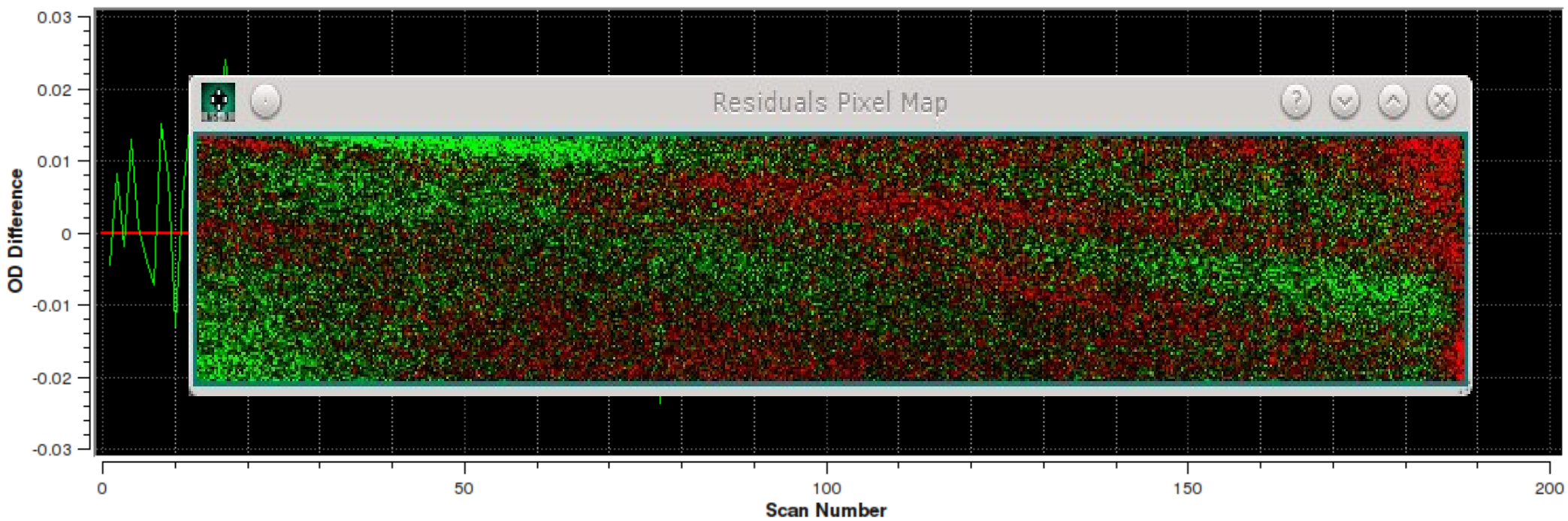
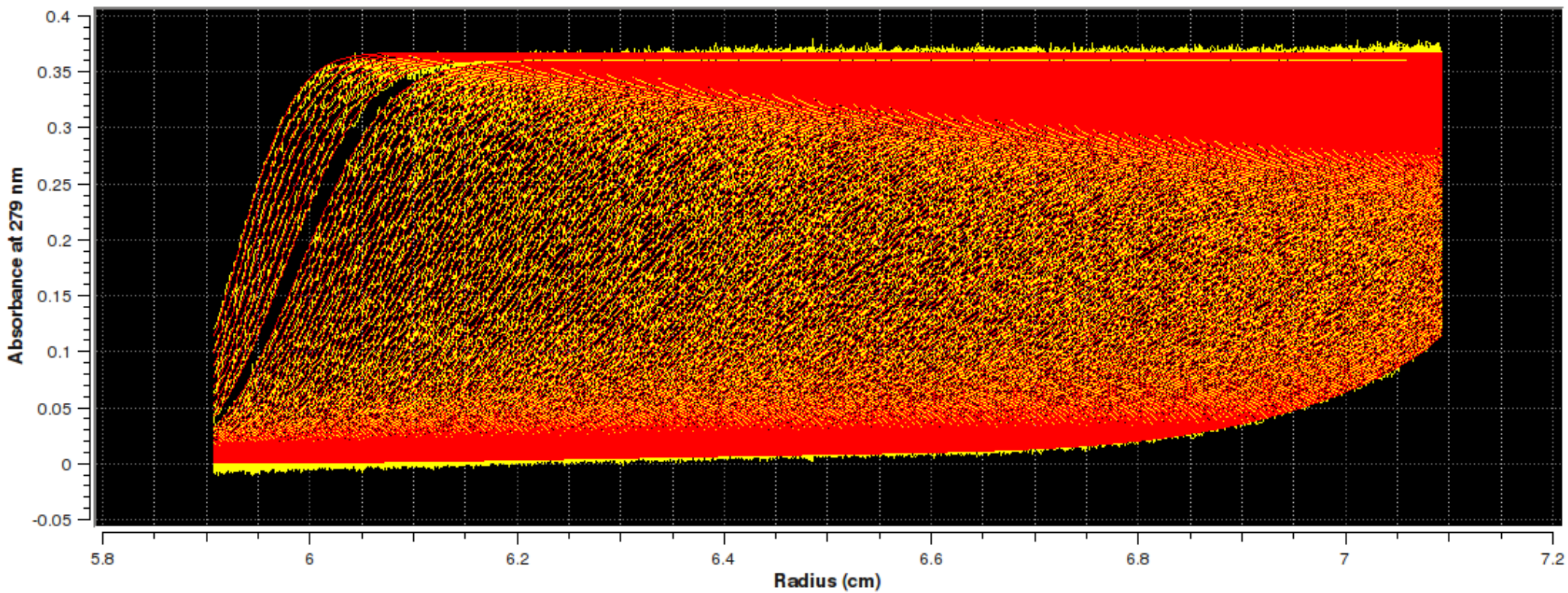


## *Different Types of Noise*

**Stochastic (random noise):** Noise is different for each radial and time point and it is (hopefully) Gaussian in distribution:



# Experimental and Simulated Data



# Intensity vs. Absorbance

## Time invariant noise sources:

### Reference Channel:

Lamp window  
Monochromator optics  
Cell window (reference channel)  
Slit assembly  
Photomultiplier tube optics

### Sample Channel:

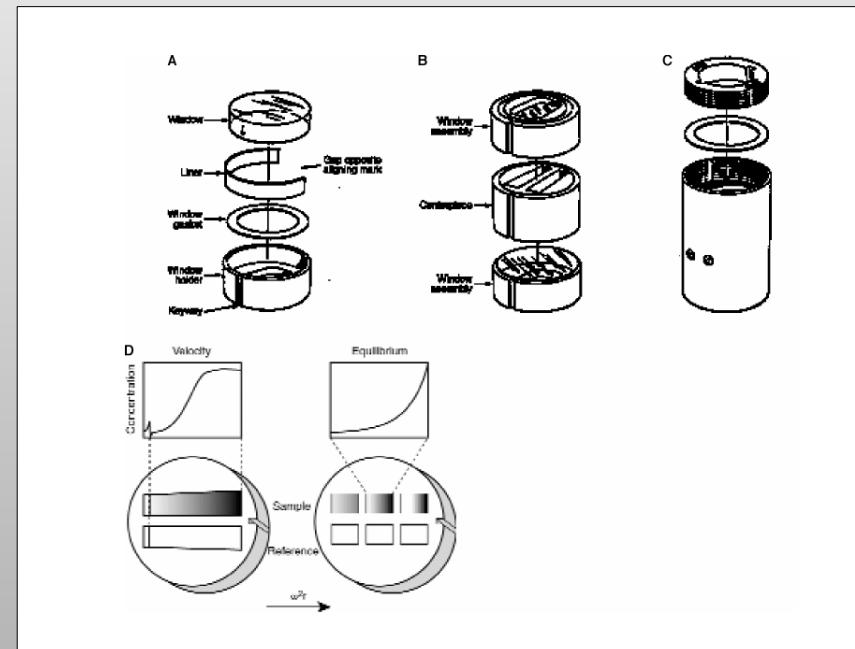
Lamp window  
Monochromator optics  
Cell window (sample channel)  
Slit assembly  
Photomultiplier tube optics

## Radially invariant noise sources:

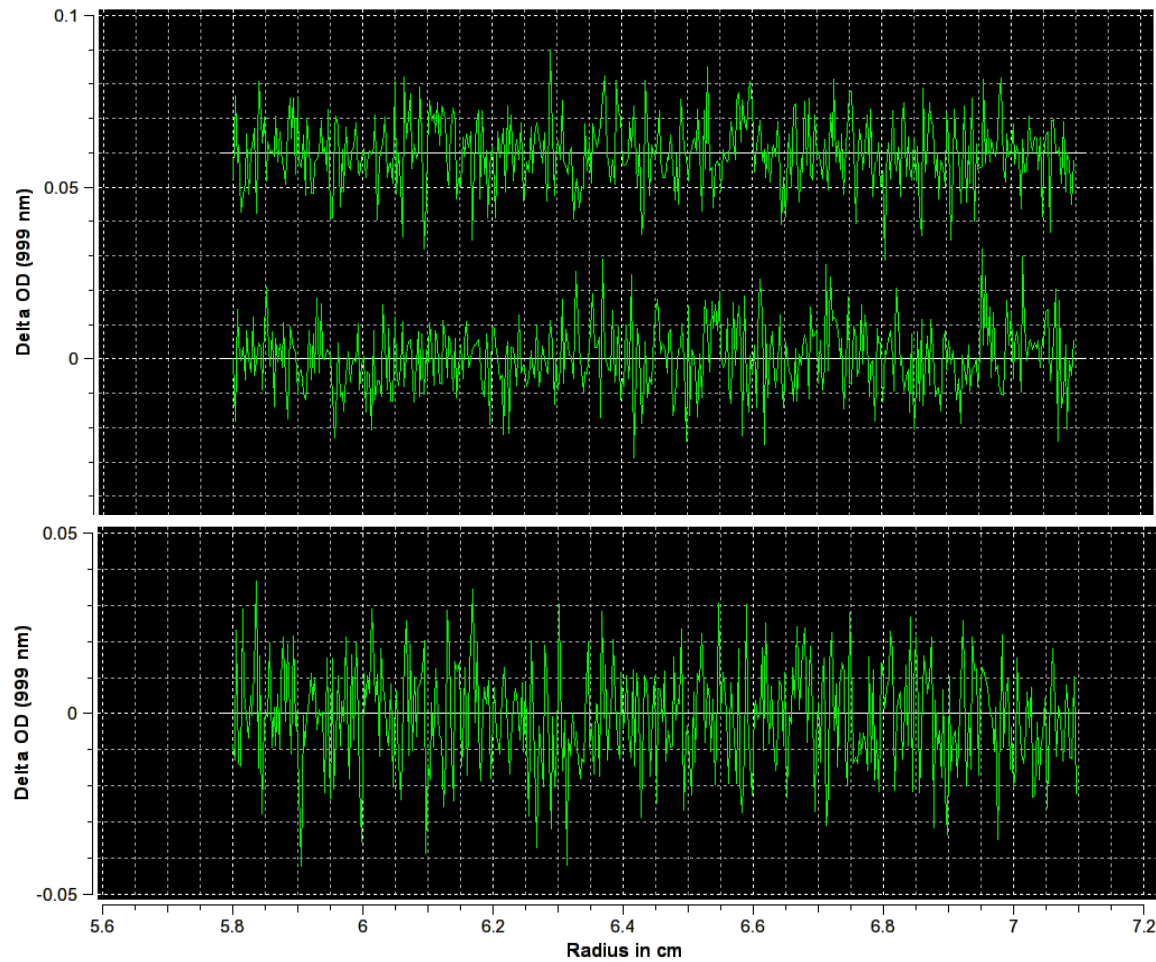
Optical path length changes  
Low frequency intensity changes

## Stochastic noise sources:

Electronics  
Photomultiplier tube  
Lamp flashes



## *Tale of 2 noisy vectors*



Noise of Scan 1

+

Noise of Scan 2

=

Noise \*  $\sqrt{2}$

## ***Intensity vs. Absorbance***

### **Optical system considerations**

**Intensity measurements record the intensity of light passing through one channel. Absorbance measurements record the intensity of light passing through one channel, then record the intensity of the light passing through the reference channel, and subtract it from the first channel. Each channel recording contains the (nearly) same amount of time invariant noise, but different amount of stochastic noise. Subtraction of time invariant noise will eliminate it:**

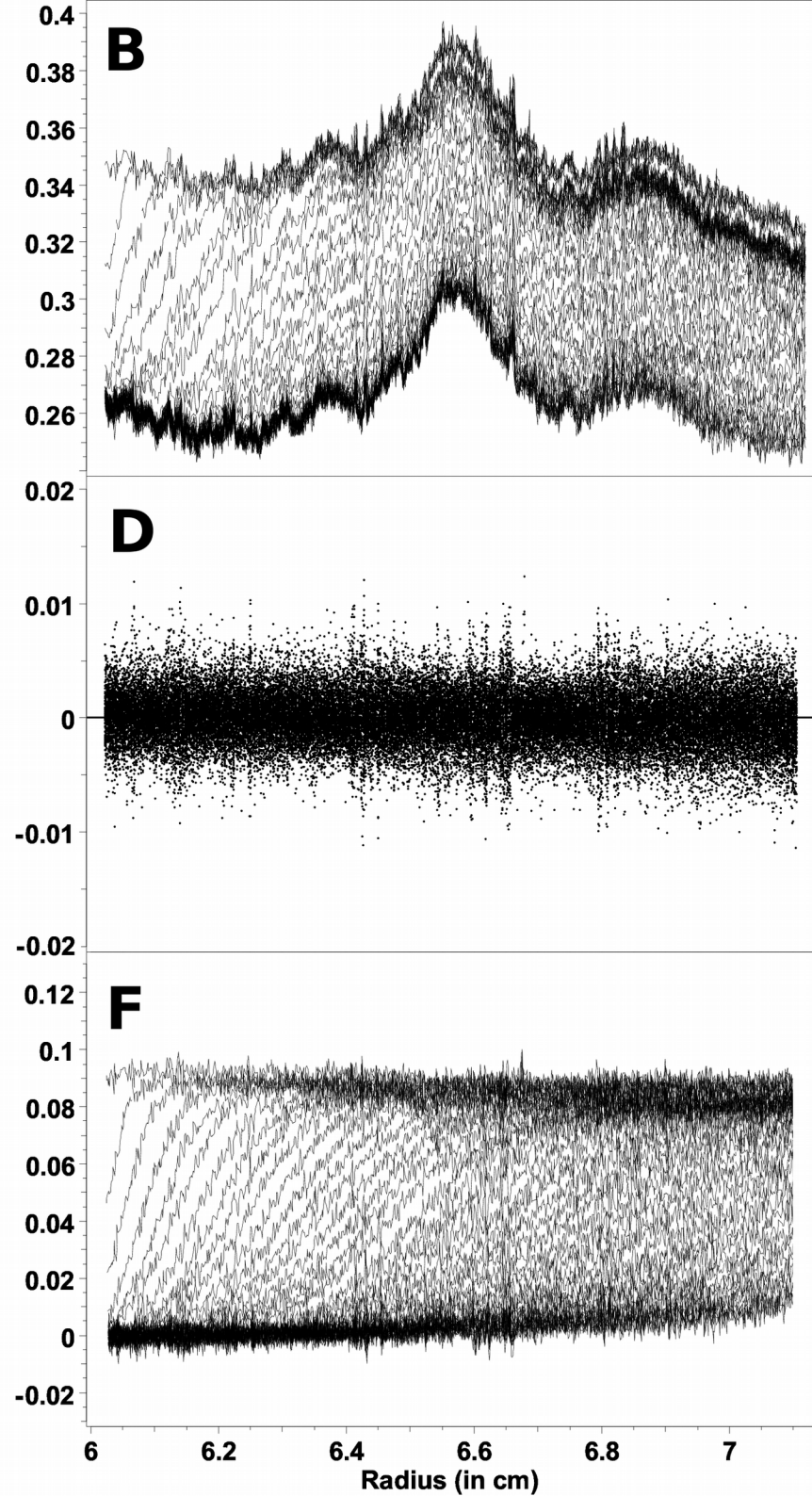
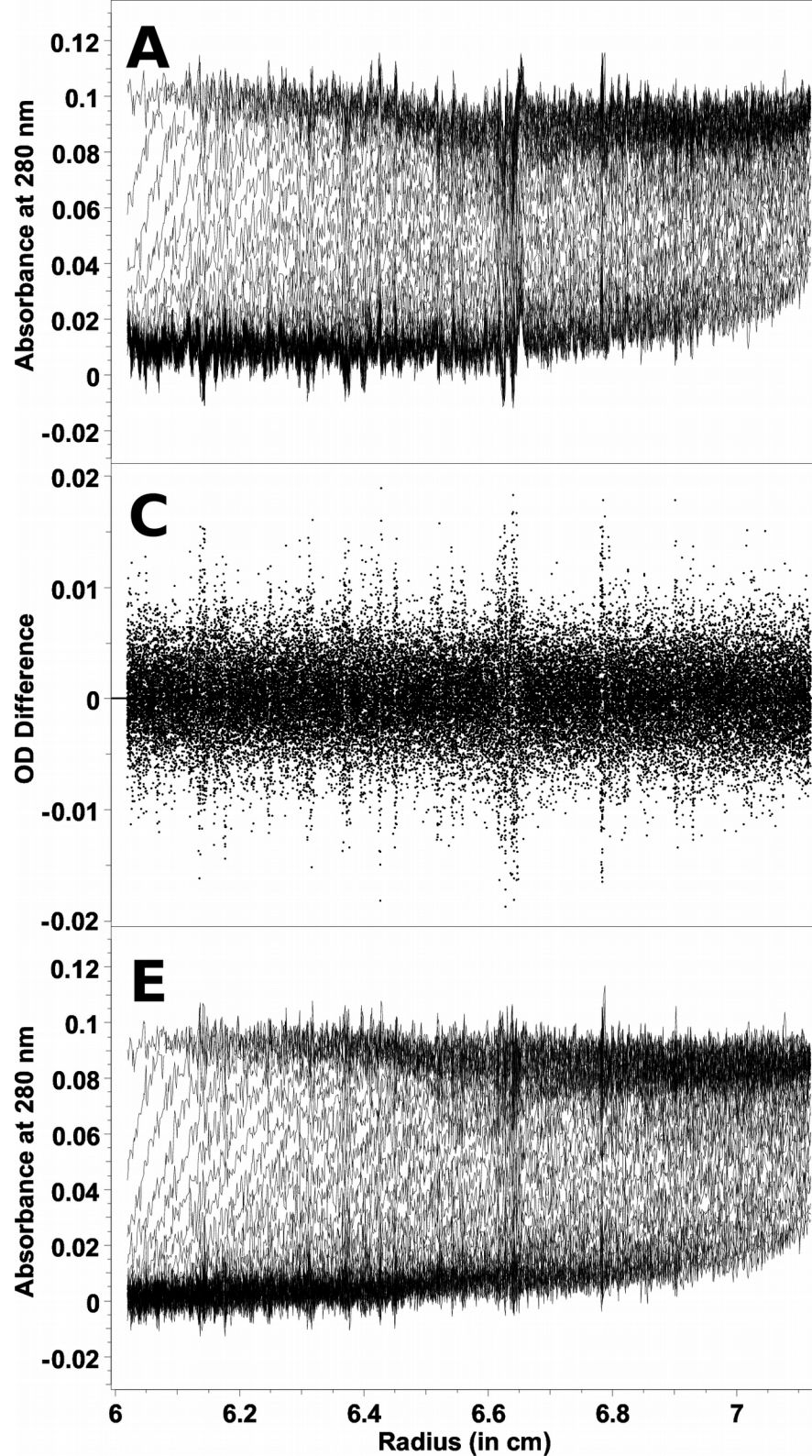
$$scan_1 = signal_1 + N_{s1} + N_{ti}$$

$$scan_2 = signal_2 + N_{s2} + N_{ti}$$

$$scan_1 - scan_2 = signal_1 - signal_2 + N_{s1} + N_{s2}$$

$$N_{s1} + N_{s2} \approx N_{s1} \sqrt{(2)}$$







## *Advantages of Intensity Data*

Measurement of intensity data when measuring **velocity experiments** has multiple advantages:

Stochastic noise is reduced by approximately a factor of 1.4, providing much better quality data and improving the RMSD of finite element solution data fits.

The capacity of the instrument is increased because the reference channel can be filled with sample, although the total OD in the reference cell should remain below 0.5 (for the reference cell). The sample OD can be higher, but just as in absorbance optics, it should not exceed the dynamic range of the UV-vis detector.

Due to the lower stochastic noise the sample concentration can be reduced while maintaining the same signal-to-noise ratio as observed in the absorbance measurement.

## ***Advantages of Intensity Data***

**Additional considerations when measuring intensity data:**

**Intensity data can only be used for velocity experiments, since considerable time invariant noise and also increased radially invariant noise (from variations in the lamp intensity) is present in such data. Equilibrium data by definition are time-invariant and hence correlate 100% with the noise removal routine which is essential for using time-invariant noise.**

**A Reference Channel is no longer required, the air-to-air region above the meniscus can be used as a reference.**

**Baseline absorbance must be added to sample absorbance and the total must be lower than 0.5 OD to avoid gain setting changes of the photomultiplier tube.**

## ***Summary:***

**Time- and radially- invariant noise can be fitted by UltraScan and removed from the data to improve the results.**

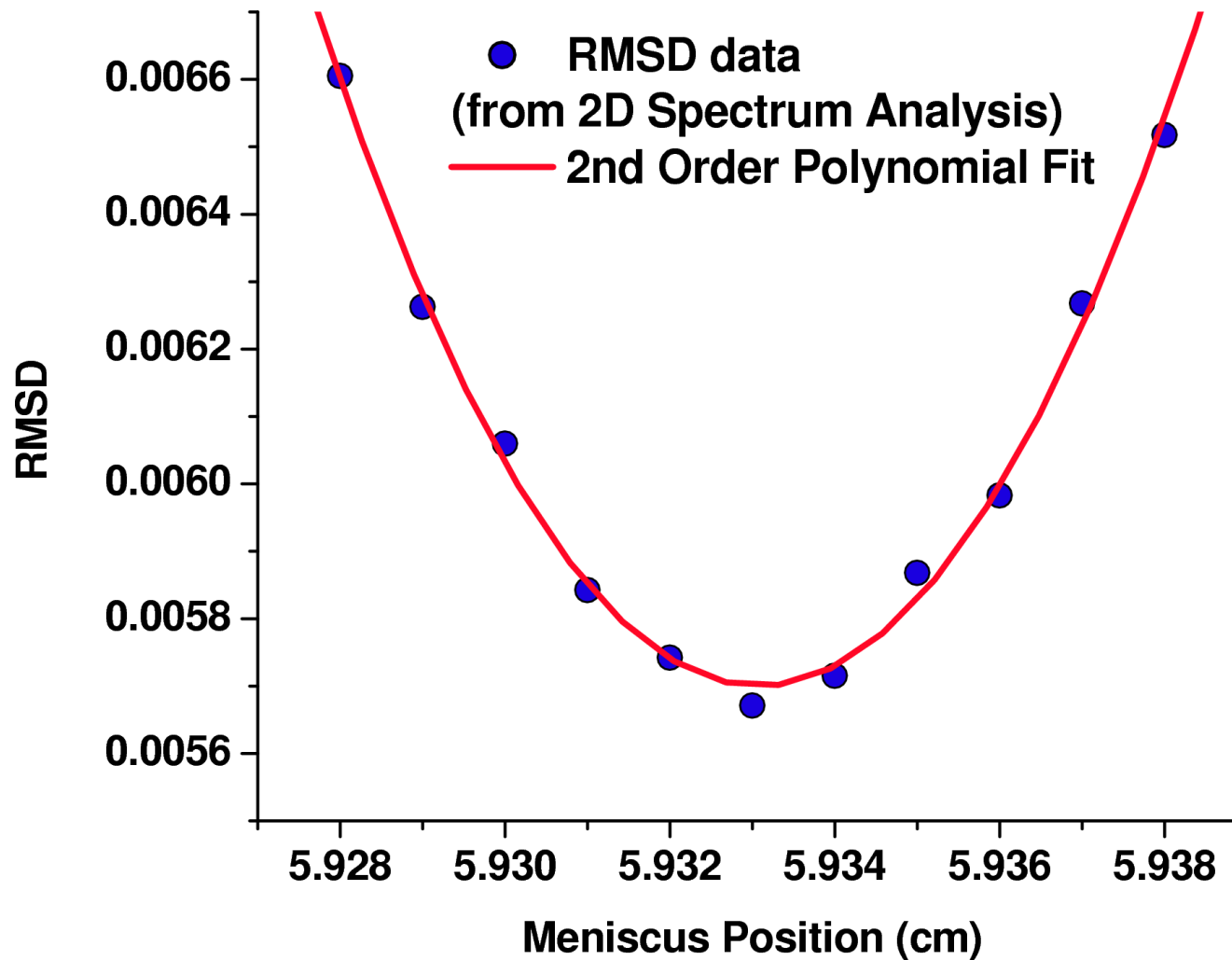
**Stochastic noise cannot be removed and should be minimized by maintaining a well calibrated instrument and performing a well-designed experiment!**

**Data subtraction in absorbance mode  $\sqrt{2}$  convolutes two stochastic vectors and leads to an increase in stochastic noise by  $\sim$**

**Remember:**

**you cannot get reliable answers if you start  
with low quality input data!**

## *Factors that affect Accuracy - Meniscus*



## Modeling Flow with the Lamm Equation

$$\left( \frac{\partial C}{\partial t} \right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ s \omega^2 r^2 C - D r \frac{\partial C}{\partial r} \right]_t$$

**Concentration**                      **Sedimentation**    **Diffusion**

The Lamm Equation describes the flow of a single solute in the sector- shaped analytical ultracentrifugation cell over time and radius. This allows us to simulate an entire experiment from start to finish.

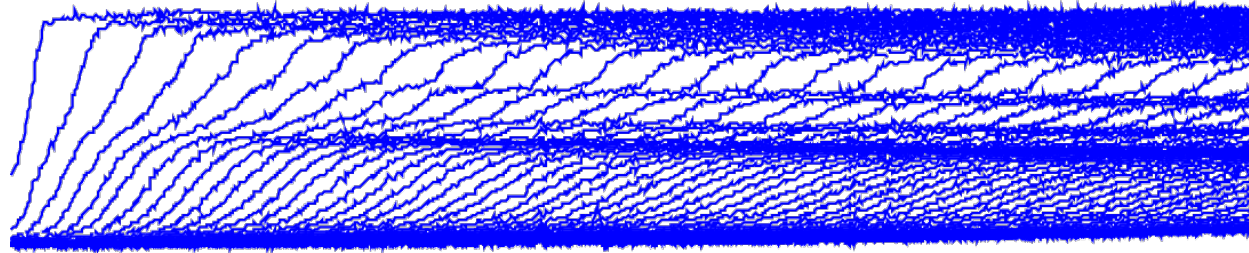
To solve this equation we use the finite element method. This method discretizes the two independent variables, the radius and the time.

This way we can calculate the concentration of the solute during the experiment for each radial point at each time point (scan).

Multiple non-interacting solutes are modeled by summing the results from two independent simulations.

*Cao W., Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. (2005) Biophys J. 89(3):1589-602.*

# Lamm Equation for Non-interacting Systems:



Lamm equation  
 $L(s, D, C)$  for a single  
 ideal solute:

$$\left( \frac{\partial C}{\partial t} \right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ \underbrace{s \omega^2 r^2 C}_{\text{Sedimentation}} - \underbrace{D r \frac{\partial C}{\partial r}}_{\text{Diffusion}} \right]_t$$

Concentration

Lamm equation for a  
 mixture of non-  
 interacting solutes:

$$C = \sum_{i=1}^n c_i L(s_i, D_i)$$



# Lamm Equation for Interacting Systems

Lamm equation  
 $L(s, D, C)$  for a single  
 ideal solute:

$$\left( \frac{\partial C}{\partial t} \right)_r = \frac{-1}{r} \frac{\partial}{\partial r} \left[ \underbrace{s \omega^2 r^2 C}_{\text{Sedimentation}} - \underbrace{D r \frac{\partial C}{\partial r}}_{\text{Diffusion}} \right]_t$$

Concentration

Lamm equation for an  
 interacting system  
 (e.g., monomer-dimer,  
 mass action applies):

$$M + M \rightleftharpoons D \quad K_a = \frac{[D]}{[M]^2}$$

$$C = [L(\bar{s}, \bar{D})]_{r,t} \quad \bar{s} = \frac{\sum_{j=1}^m s_j C_j}{C_T} \quad \bar{D} = \frac{\sum_{j=1}^m D_j (\partial C_j / \partial r)}{\sum_{j=1}^m (\partial C_j / \partial r)}$$

# *Optimization and Analysis Methods for Sedimentation Velocity*

**2-dimensional Spectrum Analysis (2DSA):** High-resolution, general and model-independent solution for size and anisotropy distributions of non-interacting systems

**Parametrically Constrained Spectrum Analysis (PCSA):** Identifies size/anisotropy relationships for polymerizing systems and provides a constrained fit over the 2-dimensional sedimentation/diffusion space.

**Custom Grid Analysis (CG):** Takes advantage of prior knowledge to parameterize the 2DSA grid in terms of alternate hydrodynamic variables.

**Genetic Algorithms (GA):** Robust non-linear least squares optimization method that provides parsimonious regularization of 2DSA spectra. Also used for fitting of discrete, non-linear models (reversible association, non-ideality, co-sedimenting solvents).

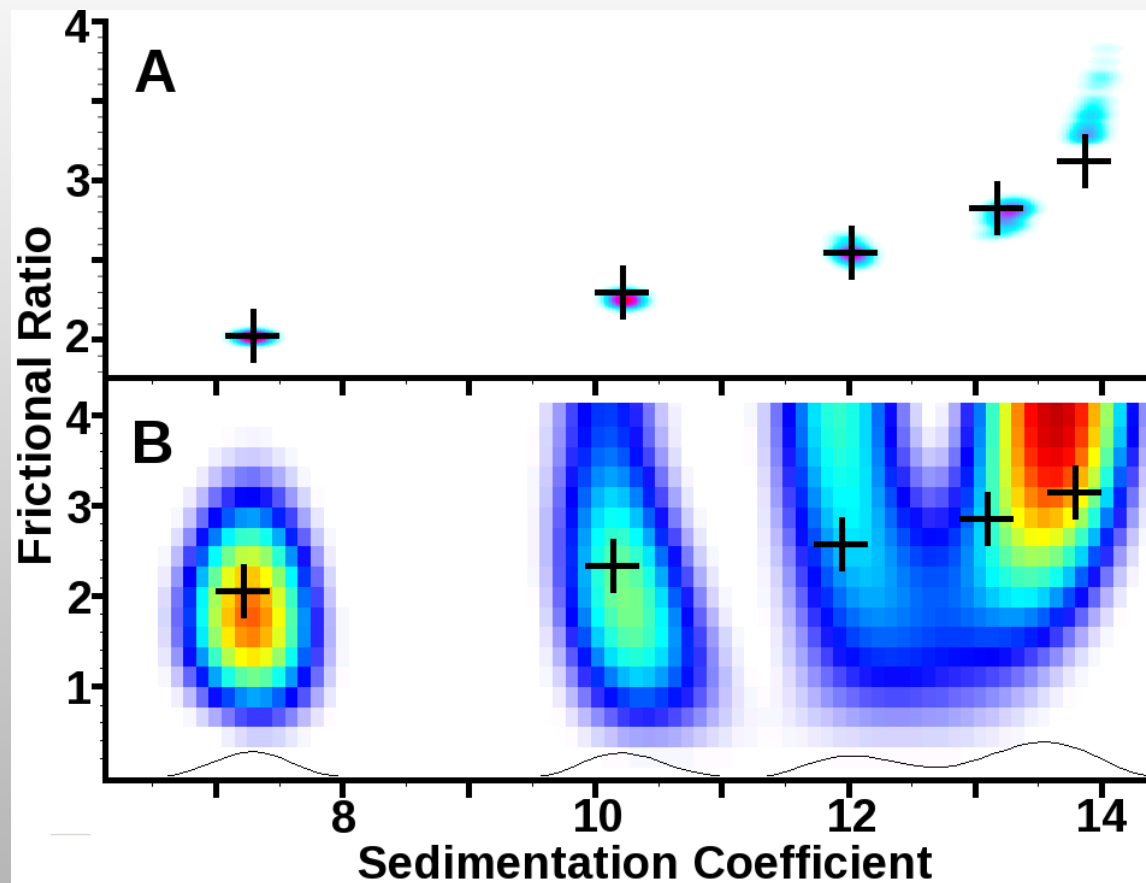
**Monte Carlo Analysis (MC):** Used to measure the effect of noise on the fitted parameters, yields parameter distribution statistics

**van Holde – Weischet Method (vHW):** Used to generate diffusion-corrected sedimentation profiles which provide finely detailed comparisons between multiple samples.

**C(s), C(s, f/f<sub>0</sub>), C(s, M):** Low resolution methods - not used in UltraScan.

# Optimization and Analysis Methods for Sedimentation Velocity

$C(s)$ ,  $C(s, f/f_0)$ ,  $C(s, M)$ : Low resolution methods - not used in UltraScan.



# ***Nonlinear Least Squares Finite Element Fitting***

**Direct Boundary fitting uses a nonlinear least squares minimization approach to fit a model function (a sum of Lamm equations)  $Y^*$  to an experimental dataset  $Y$ :**

**Our Model:** 
$$Y^* = \sum_{k=1}^n c_k L(s_k, D_k) + b$$

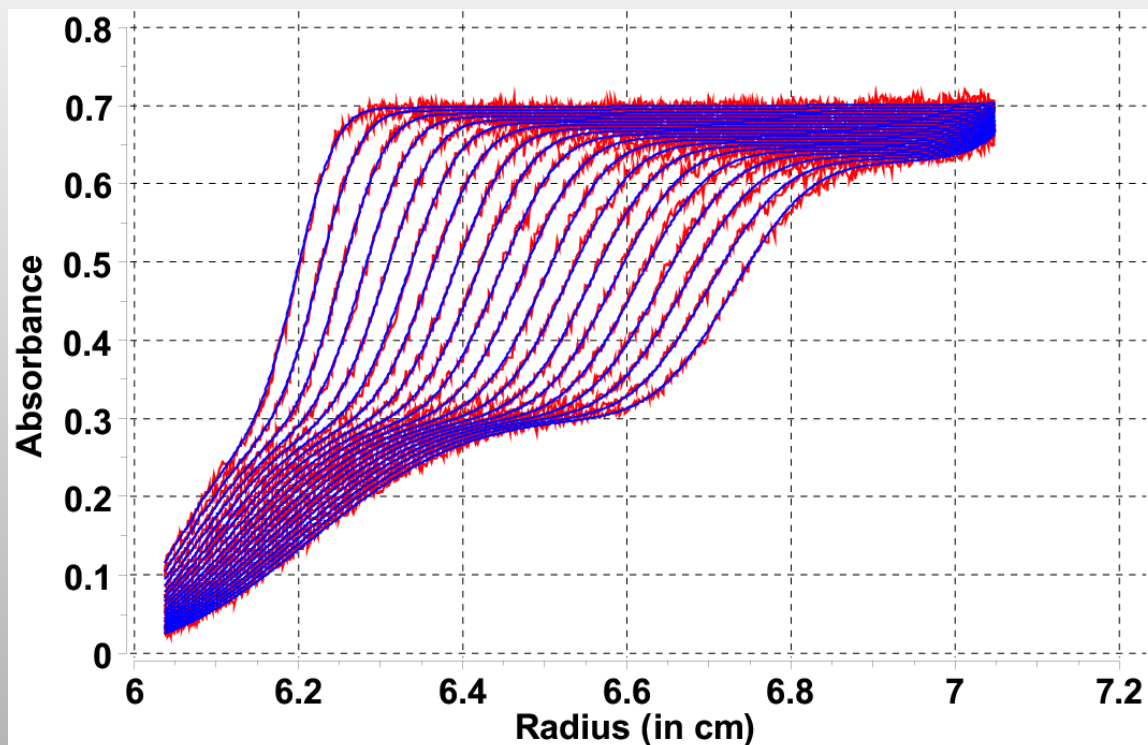
**The model is compared to the experimental data in the least squares sense for each data point in the experiment (over time and radius)**

$$\text{Min} \sum_{i=1}^r \sum_{j=1}^t \left[ Y_{ij}^* - Y_{ij} \right]^2$$

**here,  $c$ ,  $b$ ,  $s$  and  $D$  are nonlinear parameters, and are adjusted independently in an iterative fit (Svedberg, SedAnal, Lamm) .**

# Nonlinear Least Squares Finite Element Fitting

Finite Element - Nonlinear Least Squares (RMSD:  $4.61 \times 10^{-3}$ )  
Monte Carlo is needed to define statistical confidence of fitted parameters.



$M_1$ : 128.8 kD (135.7 kD)

$f/f_0$ : 3.10

$s_1$ :  $5.43 \times 10^{-13}$

$D_1$ :  $2.28 \times 10^{-7}$

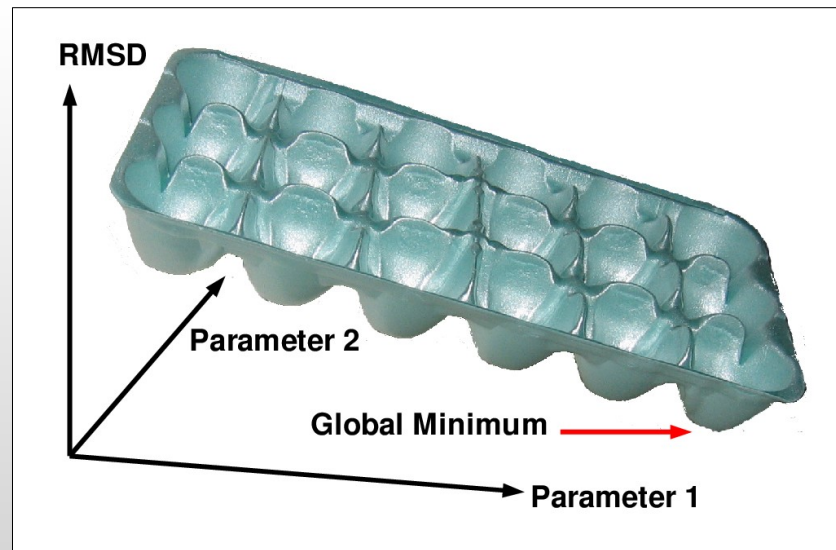
$M_2$ : 14.6 kD (14.3 kD)

$f/f_0$ : 1.29

$s_2$ :  $1.71 \times 10^{-13}$

$D_2$ :  $1.02 \times 10^{-7}$

## ***The Optimization Challenge:***



### **Problem with nonlinear least squares optimization:**

For multi-component systems, the nonlinear least squares fitting algorithm gets easily stuck in local minima and the solution depends on the starting points. Problem gets worse with more parameters (i.e., multiple components).



## ***The Optimization Challenge:***

- 1. For complicated problems, nonlinear optimization will fail and the fitting algorithm will not converge to the global optimum.**
- 2. In addition, due to noise the solution will not be unique and there will be an infinite number of equally likely solutions with the same  $\chi^2$**

**How do we get around these problems?**

**Problem 1 can be alleviated by *linearizing* the problem**

**Problem 2 is intractable. The best we can do is to perform a statistical error analysis and use Monte Carlo methods.**

## ***C(s)/C(M) Method (P. Schuck)***

### **Linearization Approach 1 – keeping a constant $f/f_0$ value:**

Decomposition of the concentration function into a linear combination of orthogonal basis functions (Lamm equations) distributed over a partitioned s-value range and a constant frictional ratio  $\Phi = f/f_0$ :

$$C = \underbrace{c_1 L(s_1, D(s_1, \Phi))}_{\text{Component 1}} + \underbrace{c_2 L(s_2, D(s_2, \Phi))}_{\text{Component 2}} + \dots$$

Fit only the amplitudes ( $c_j$ ) of those components that make a non-zero contribution by doing a non-negatively constrained *linear* least squares fit over all components.

## ***C(s)/C(MW) Method (P. Schuck)***

### **Parameterization Approach:**

Instead of using nonlinear fitting parameters  $s$  and  $D$  (which are required for the solution of the Lamm equation), we treat these parameters as constants. The  $s$ -value is partitioned over a range from  $s_{min} < s < s_{max}$  in equi-distant intervals. Using the Stokes-Einstein relationship, the diffusion coefficient can be expressed as a function of the sedimentation coefficient and a constant frictional ratio  $\Phi = f/f_0$

$$D = \frac{RT}{18\pi N(\Phi \eta)^{2/3}} \sqrt{\frac{2(1 - \bar{v} \rho)}{s \bar{v}}}$$

This way, given an  $s$ -value and a fixed shape, a corresponding diffusion coefficient can be calculated for each  $s$ -value and the Lamm equation term for each species can be calculated. Then the only question remaining is the amplitude of each term, which is a linear fit, and the best match for  $k$ . The frictional ratio can be adjusted for a best fit average using a line search.

*Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. 78(3):1606-19, 2000*

## ***C(s)/C(MW) Method (P. Schuck)***

Perform a *linear* fit using the NNLS method\* and only fit the amplitudes  $c_j$  subject to the constraint  $c_j \geq 0$

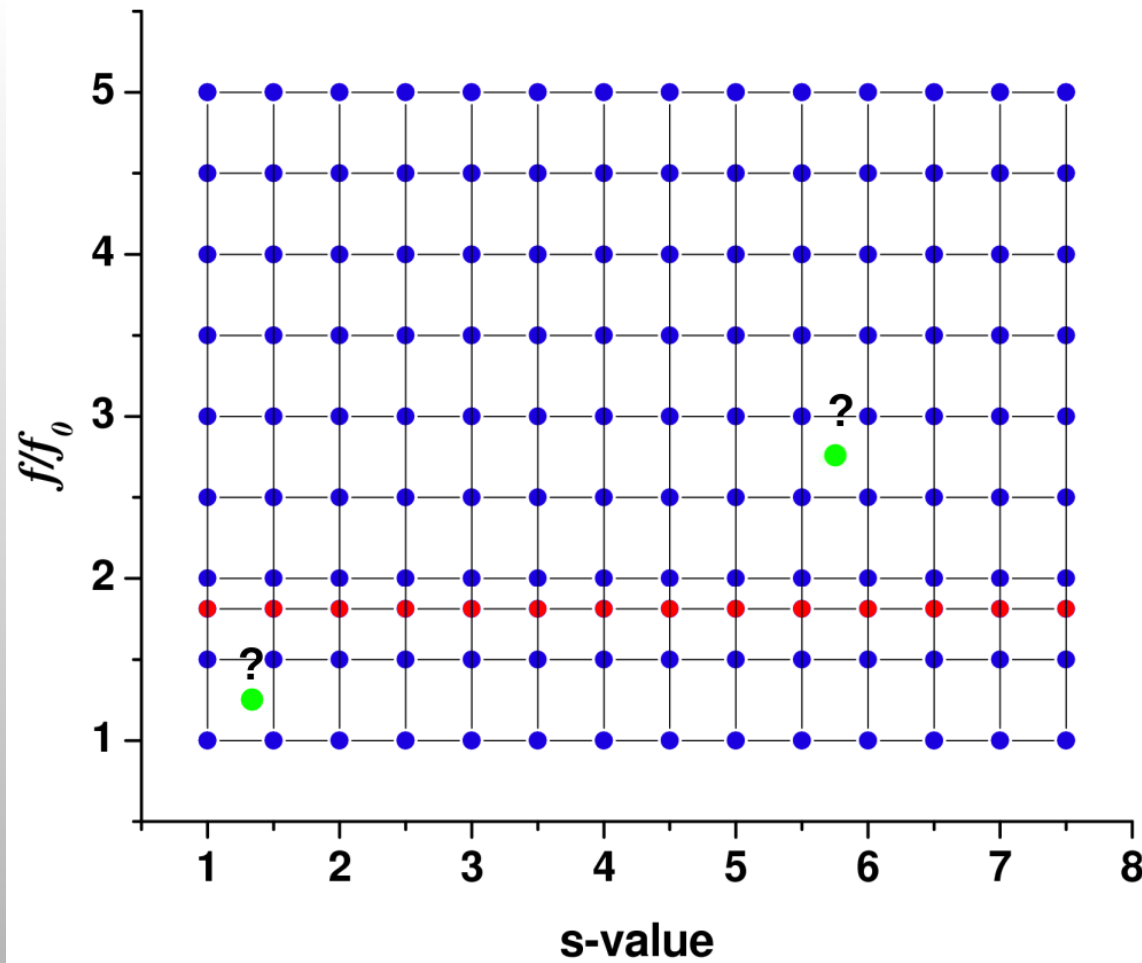
$$\text{Min} \sum_{i=1}^m \sum_{j=1}^n \left[ c_j L(s_j, D(s_j)) - Y_i \right]^2$$

**Note:** This will generate Lamm equations that have a fixed frictional ratio and a diffusion coefficient that is linked to the sedimentation coefficient.

**ALL PARAMETERS EXCEPT THE AMPLITUDES ARE CONSTANT!**

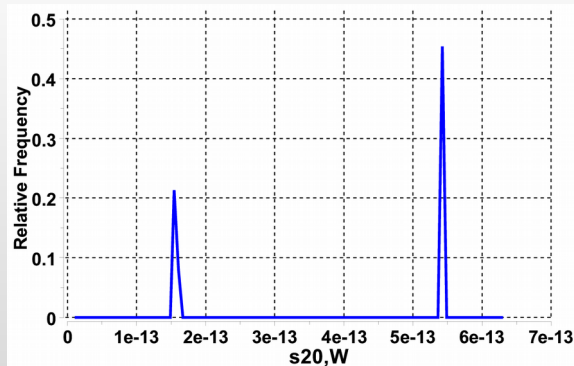
*Lawson, C. L. and Hanson, R. J. 1974. Solving Least Squares Problems. Prentice-Hall, Inc. Englewood Cliffs, New Jersey*

## ***C(s)/C(MW) Method (P. Schuck)***

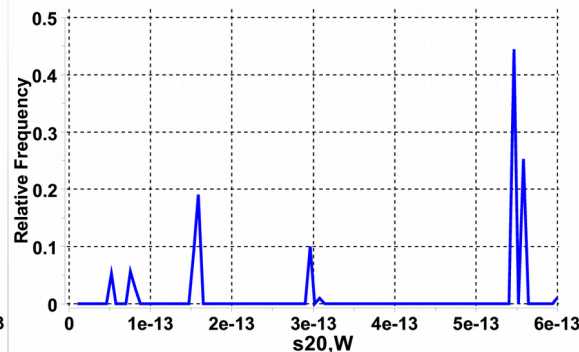


## *C(s)/C(MW) Method (P. Schuck)*

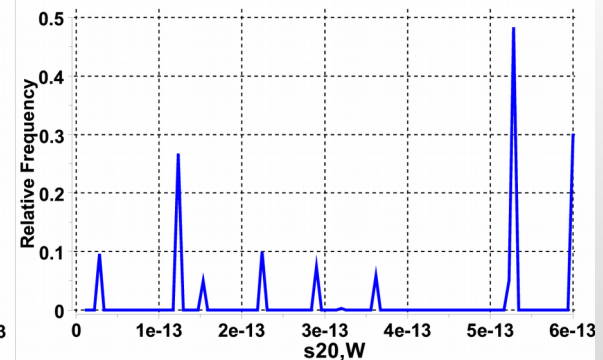
**C(s) method (lowest  $f/f_0$ ):**  $s_1$ :  $5.43 \times 10^{-13}$  (61 %),  $s_2$ :  $1.56 \times 10^{-13}$  (39 %)



$f/f_0 = 1.29$  (Lysozyme)



$f/f_0 = 2.297$  (fitted)



$f/f_0 = 3.10$  (DNA)

RMSD for C(s) fit:  $6.0 \times 10^{-3}$ , RMSD for FE fit:  $4.61 \times 10^{-3}$

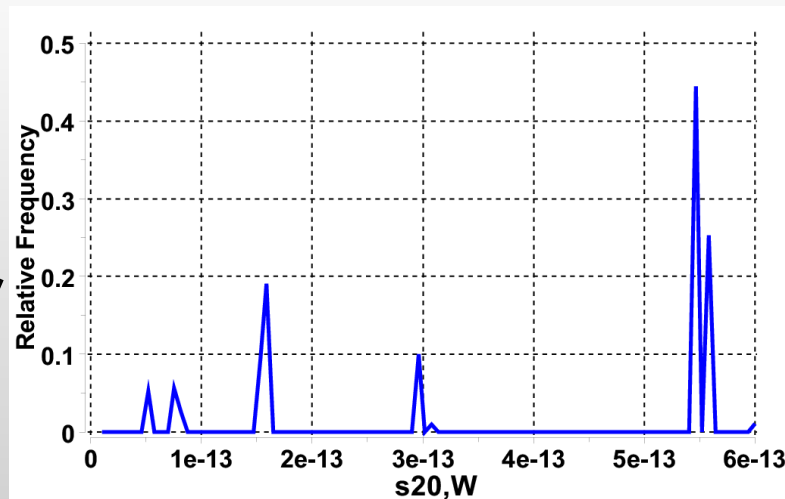
**With increasing  $f/f_0$ , the number of artifactual peaks increases**  
(regularization hides this problem)

Fitted  $f/f_0$  values provide an **average** of all components



## ***C(s)/C(MW) Method (P. Schuck)***

**C(s) method (lowest  $f/f_0$ ):  $s_1$ :  $5.43 \times 10^{-13}$  (61 %),  $s_2$ :  $1.56 \times 10^{-13}$  (39 %)**



**$f/f_0 = 2.297$  (fitted)**

**Lysozyme:**

**Molecular Weight = 30.1 kD  
too high**

**DNA:**

**Molecular Weight = 120.6 kD  
too low**

## ***Motivation: Wish List for an Optimal Method:***

**We need a method that satisfies the following criteria:**

Generality – works accurately and reliably for ***any*** system

High resolution/high information content (s, D, partial conc., Kds)

Model independent – it needs to be able to find it's own model

Suitable for global fitting – can integrate other experiments

Always converges to the global minimum (overcomes the egg carton problem!)

Computationally efficient

## 2-Dimensional Spectrum Analysis

**Solution:** Allow for variation in  $f/f_0$  as well.

This is now a very large problem, but one that can fortunately be calculated in a single iteration, with one Lamm equation for each coordinate point in the grid:

$$Y^* = \sum_{s=s_{min}}^{s_{max}} \sum_{k=1}^{k_{max}} c_{s,k} L[s, D(s, k)] + b \quad \text{Min} \sum_{i=1}^r \sum_{j=1}^t [Y_{ij}^* - Y_{ij}]^2$$

$$Ax = b \quad Lc = Y$$

Using **NNLS** for this problem guarantees  $c_{s,k} > 0$

m = # of radial points \* # of time points = 1000 \* 100 = 100,000

n = # of sedimentation value grid points (~30 - 50)

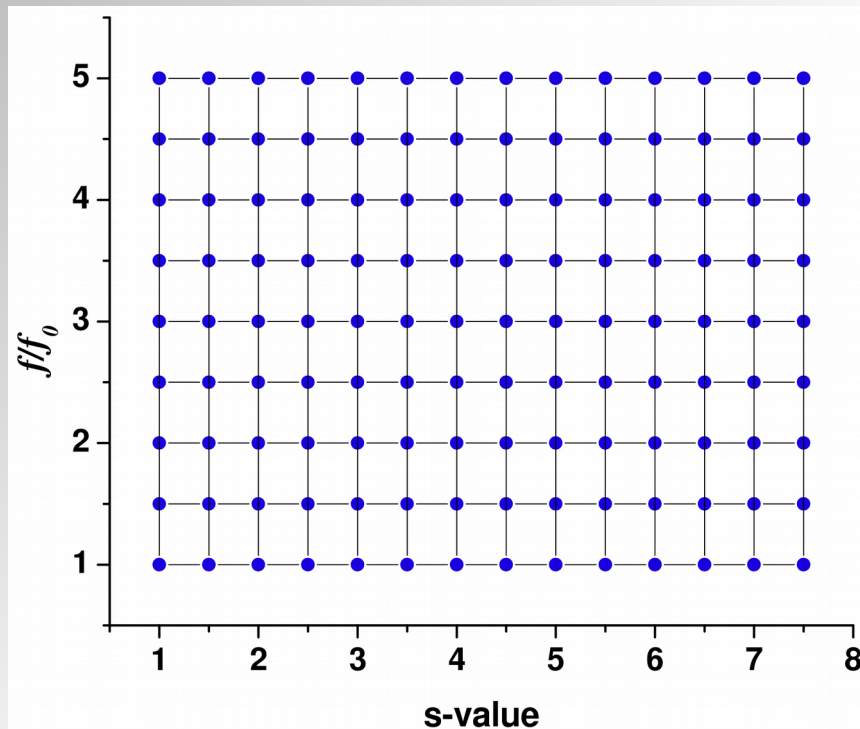
f = # of  $f/f_0$  value grid points (~30-50)

Total size: 250 million \* 4 bytes/value + workspace, altogether > 1 GB

*Brookes, E, Cao, W, Demeler, B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010 39(3):405-14.*

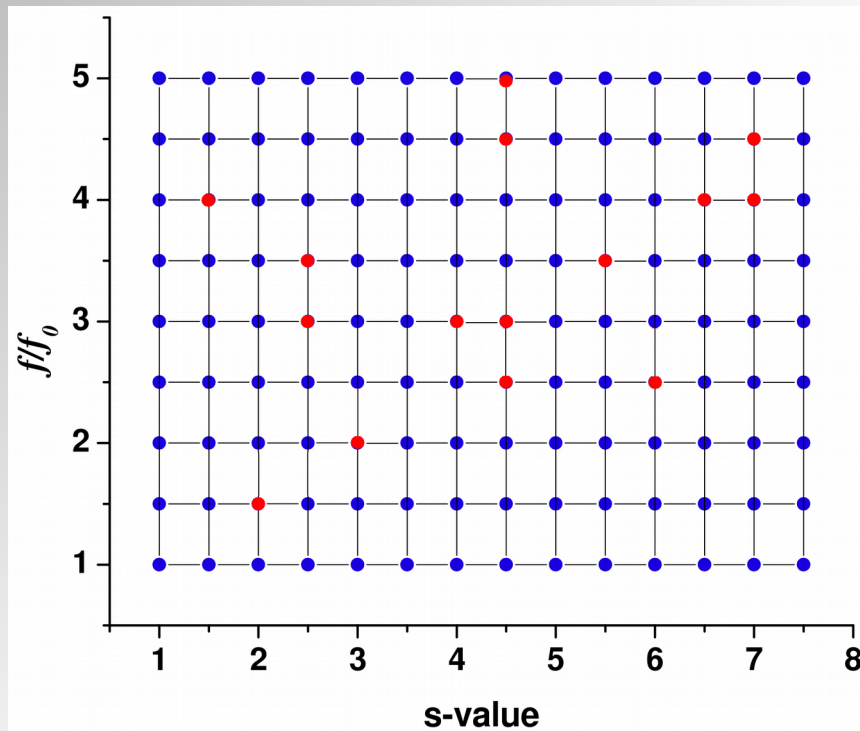
## 2-D Spectrum Analysis - Refinement:

**Step 1: Start with original grid definition:**



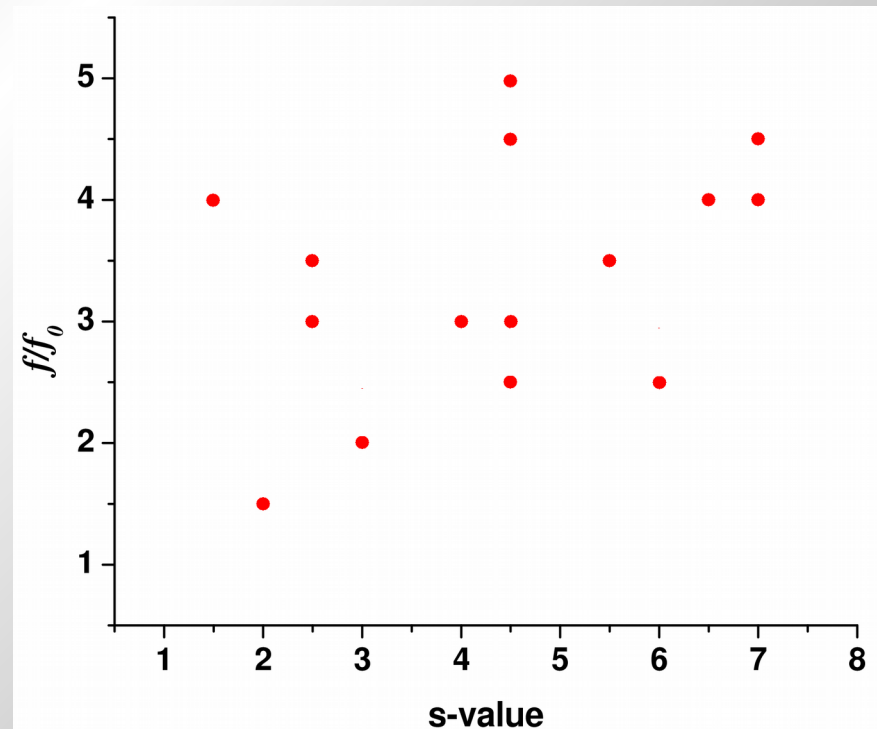
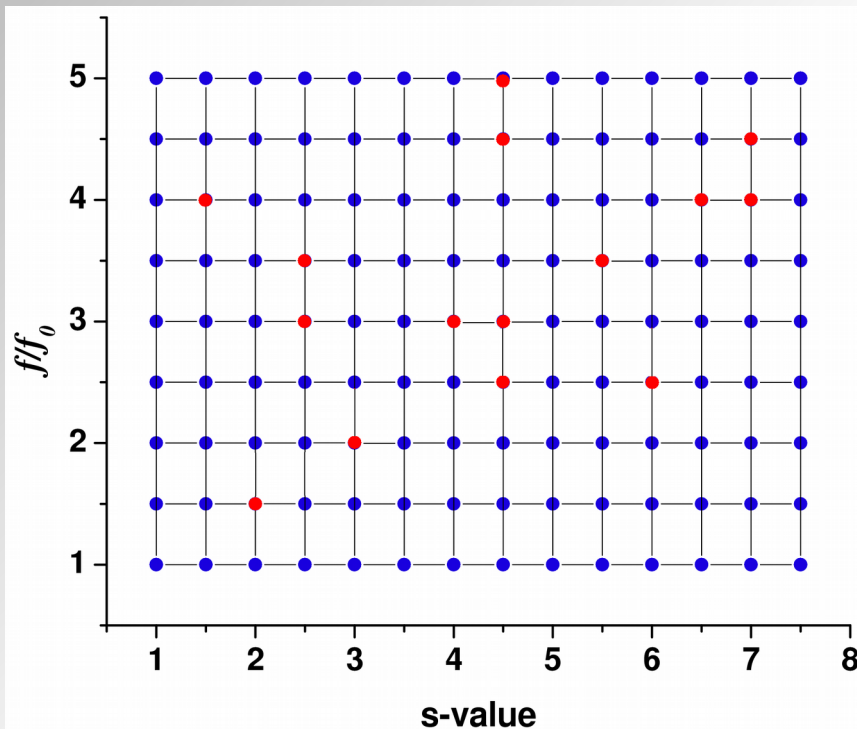
## 2-D Spectrum Analysis - Refinement:

## Step 2: Perform NNLS



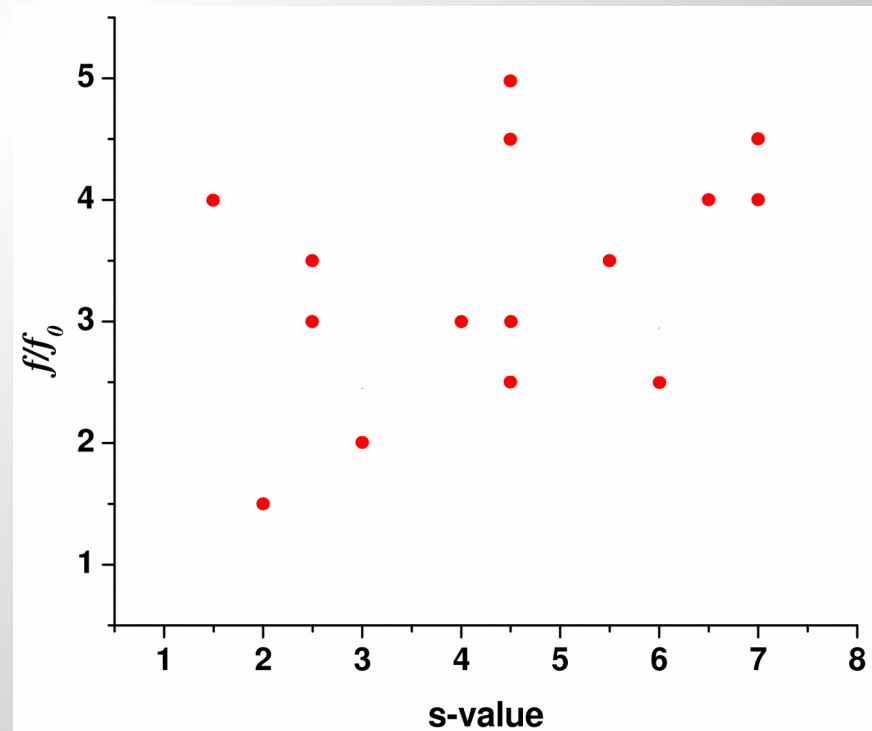
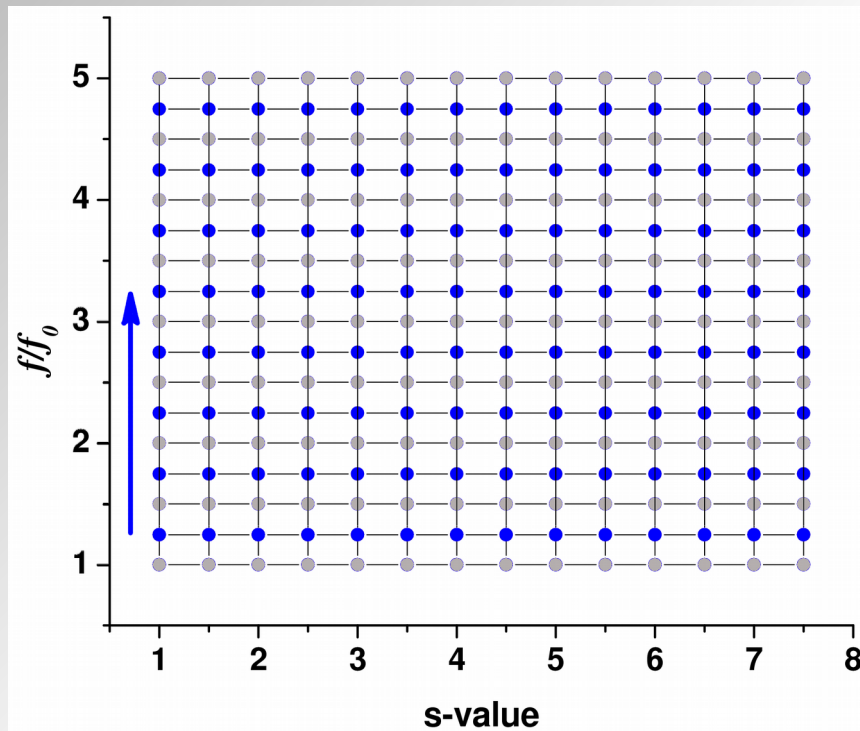
## 2-D Spectrum Analysis - Refinement:

Step 3: Save non-zero elements into a separate array



## 2-D Spectrum Analysis - Refinement:

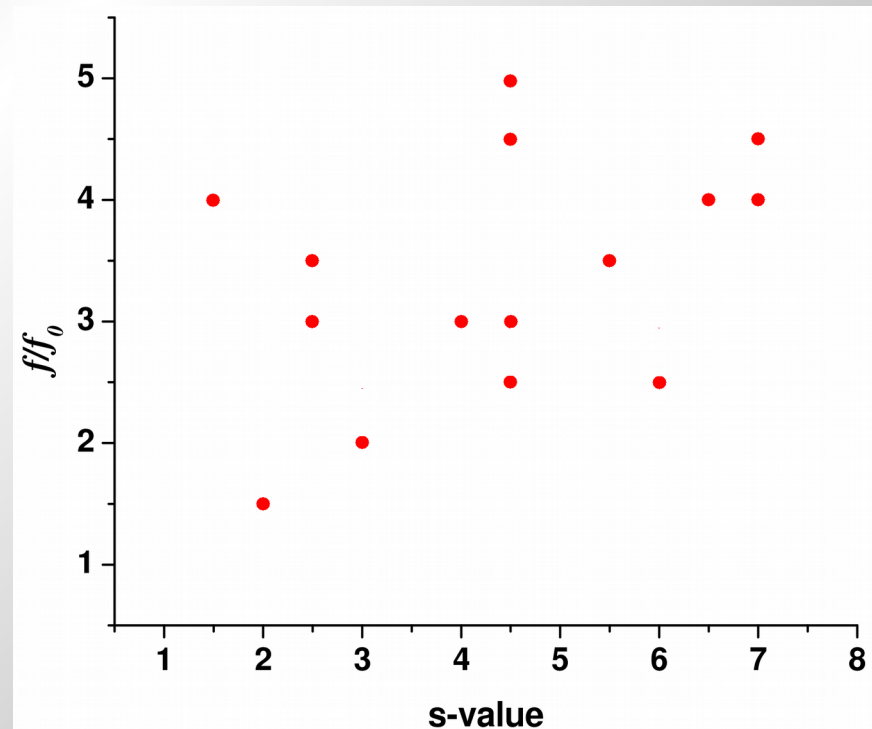
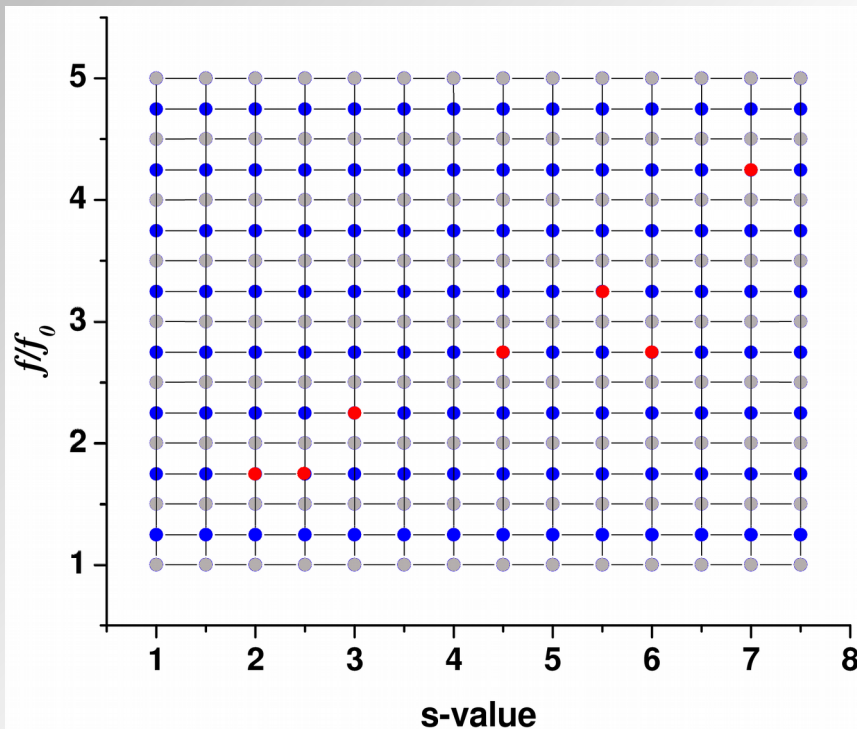
### Step 4: Shift grid into Y-direction





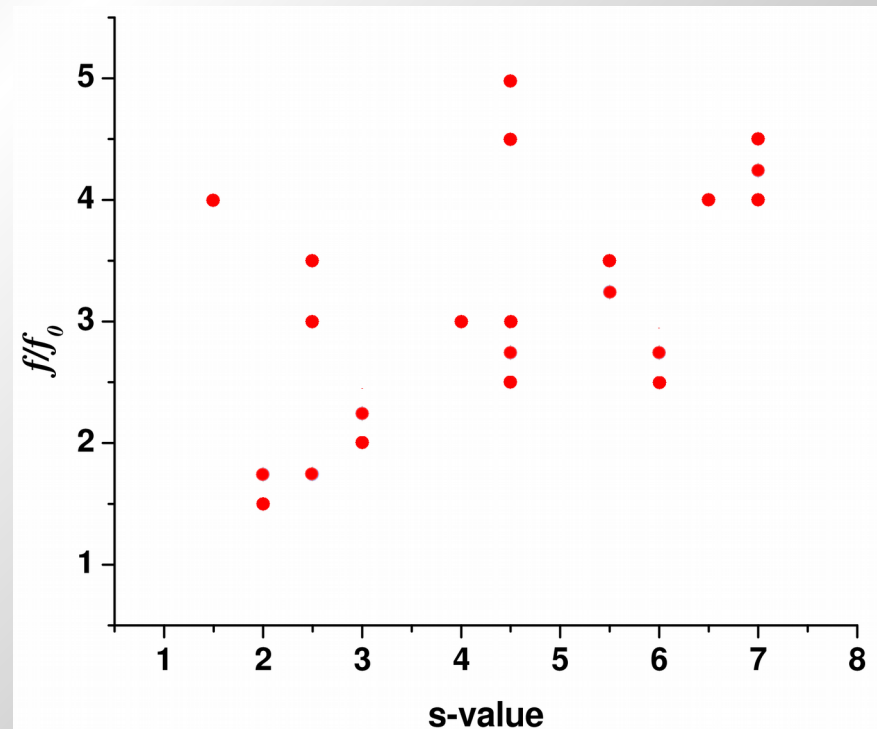
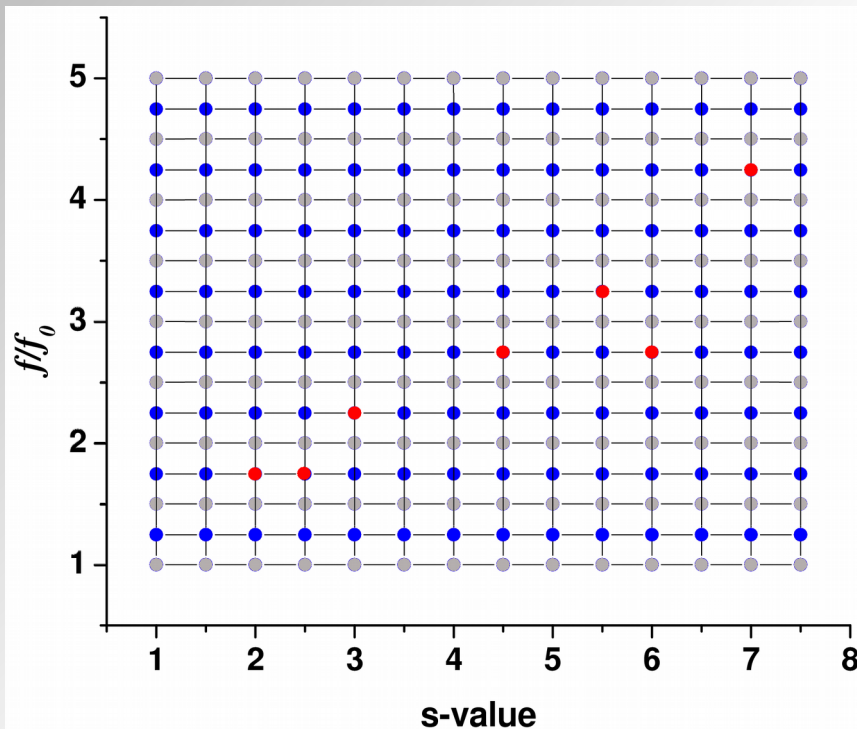
## 2-D Spectrum Analysis - Refinement:

Step 5: Perform NNLS again, but only on the shifted grid (blue)



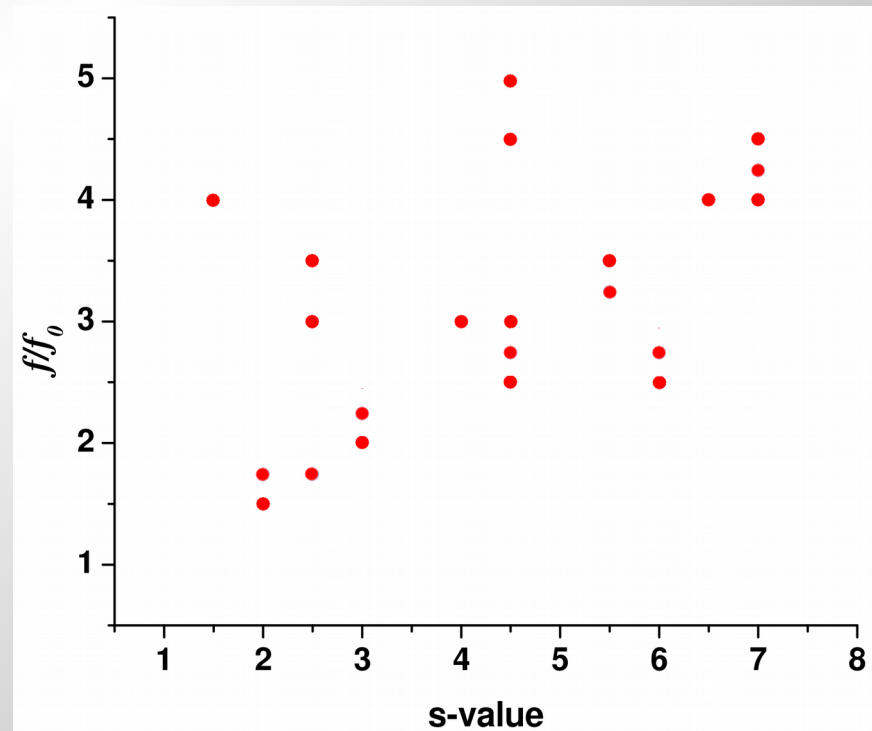
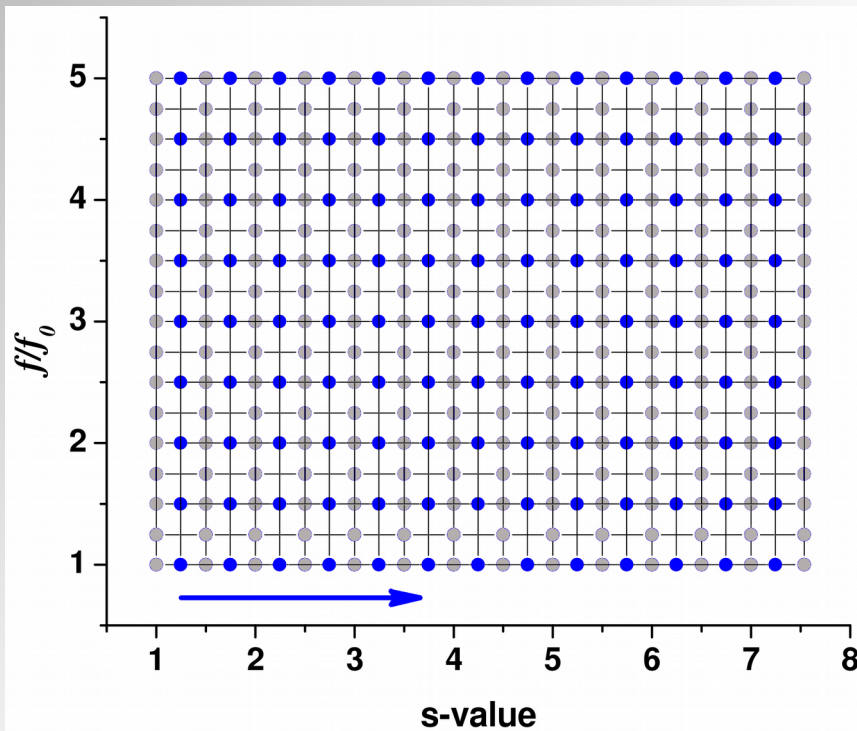
## 2-D Spectrum Analysis - Refinement:

**Step 6: Add the newly found non-zero elements to the stored array**



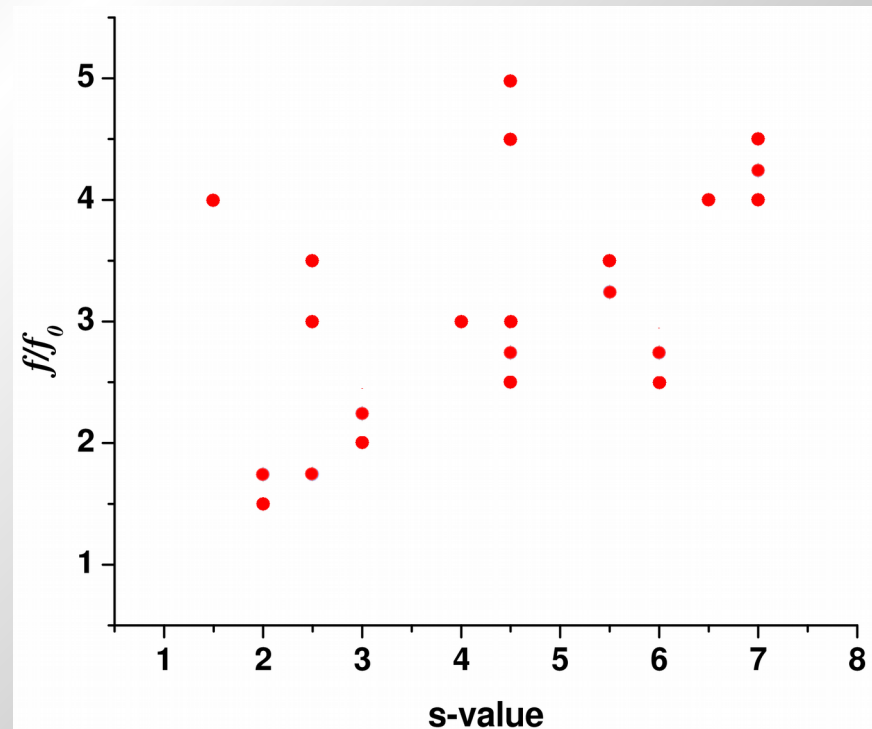
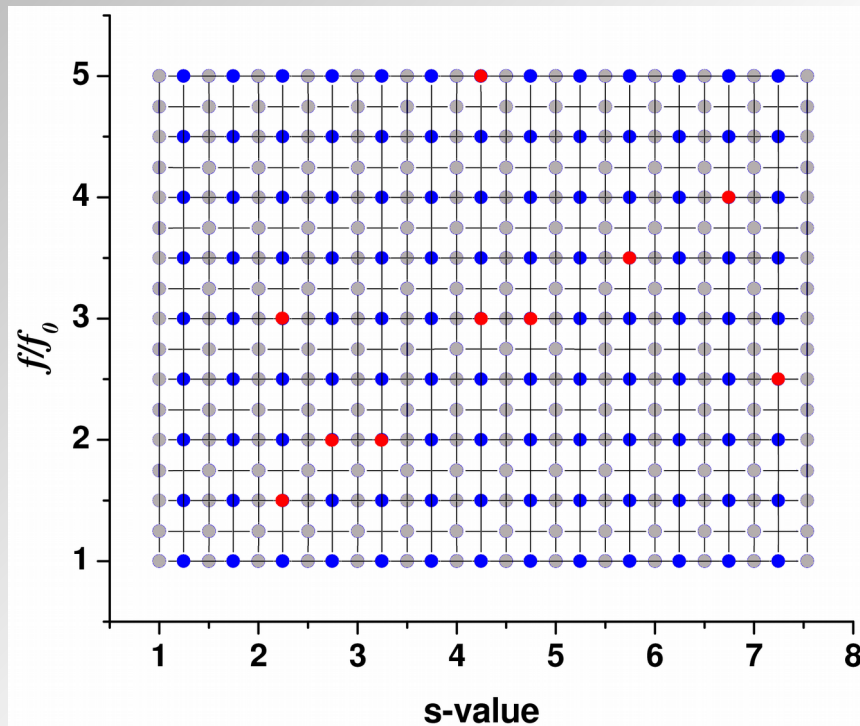
## 2-D Spectrum Analysis - Refinement:

Step 7: Now shift the grid into the X-direction



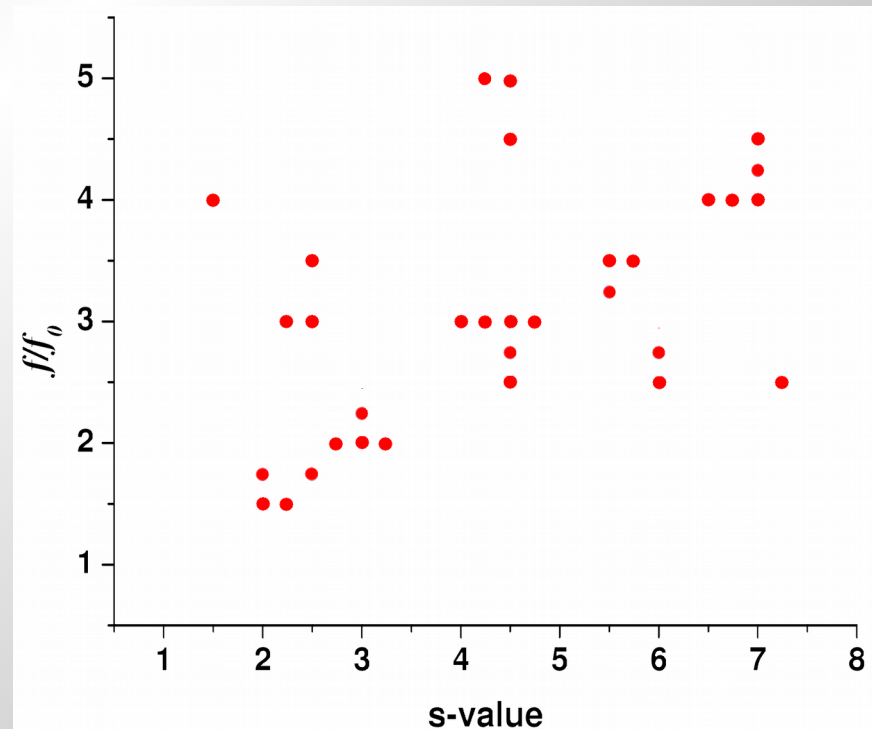
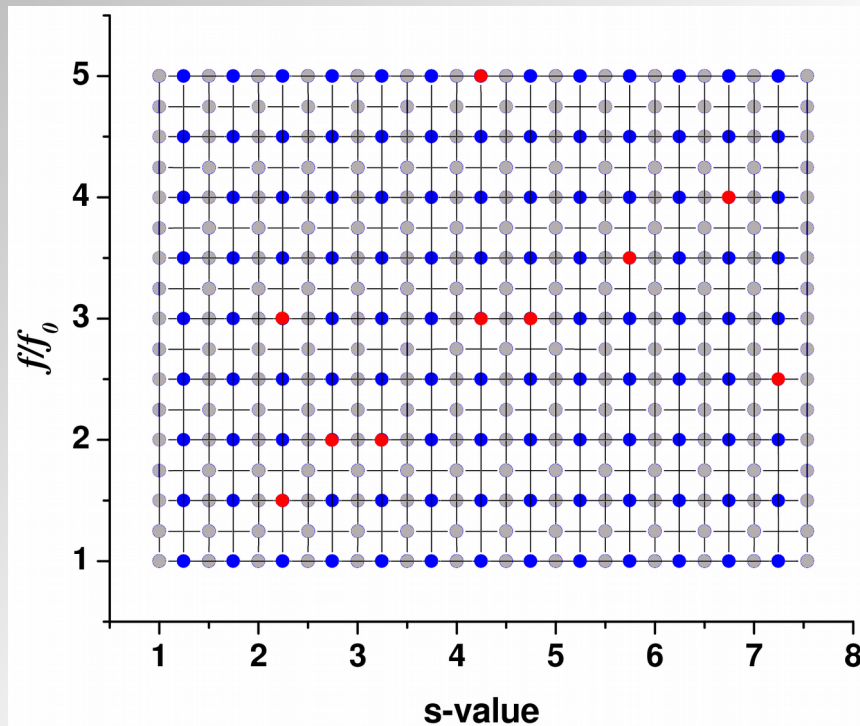
## 2-D Spectrum Analysis - Refinement:

Step 8: Perform NNLS on the shifted grid again



## 2-D Spectrum Analysis - Refinement:

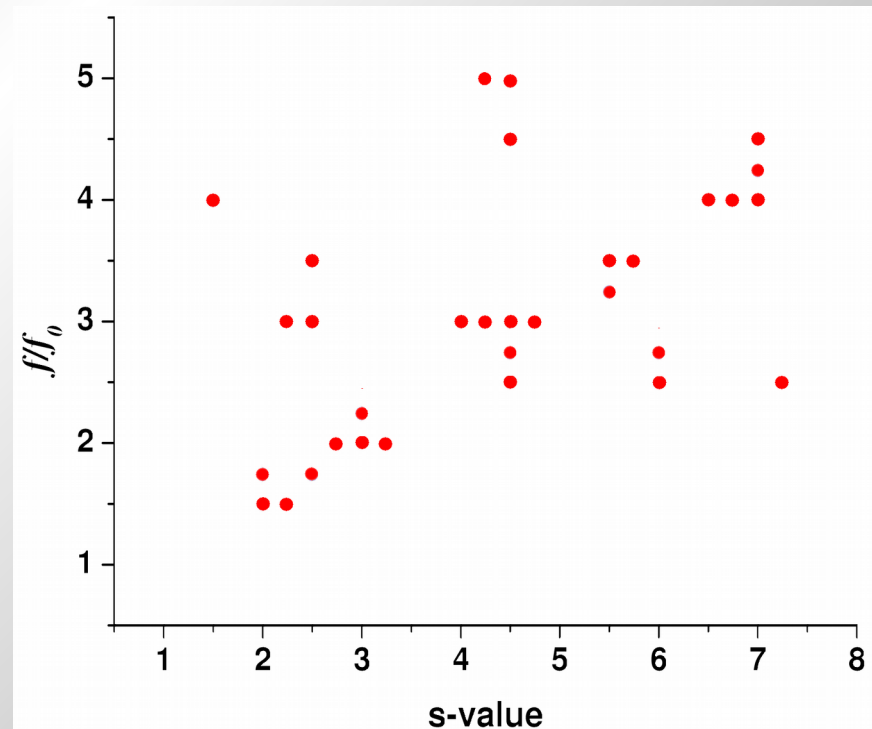
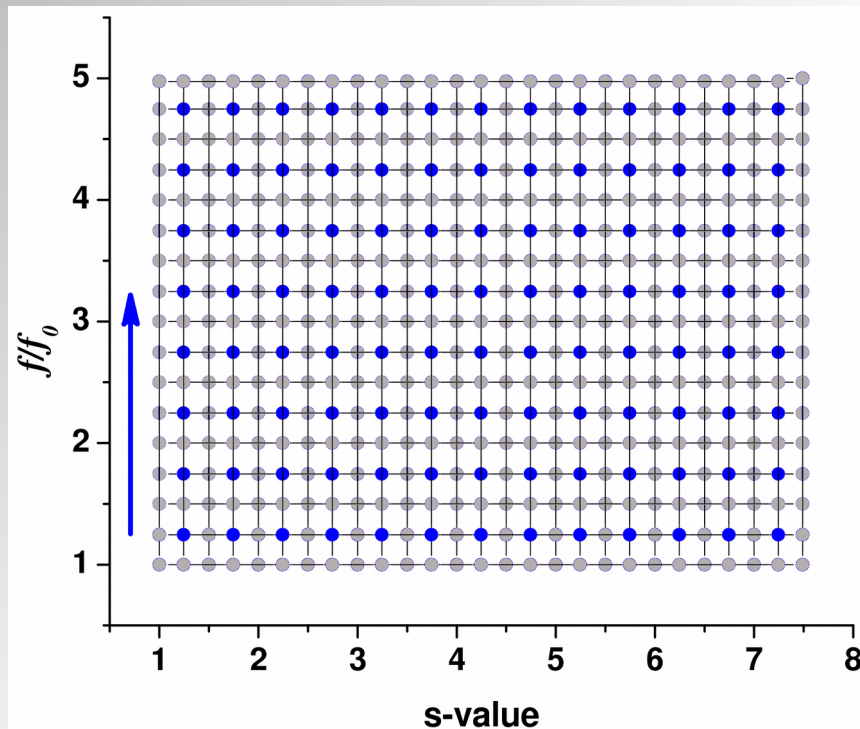
Step 9: Add the new non-zero elements to the stored array





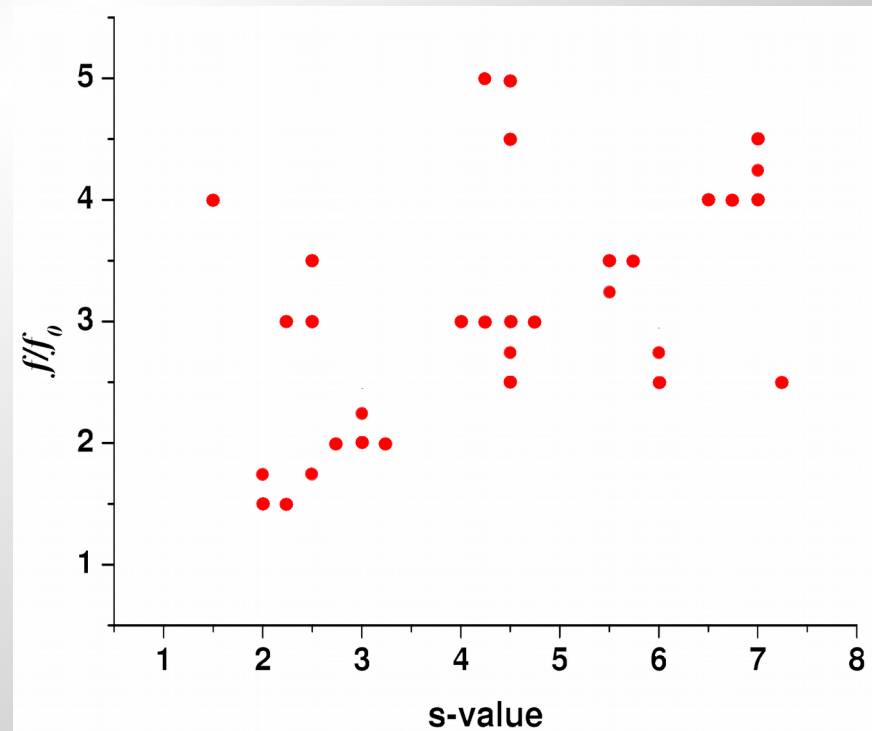
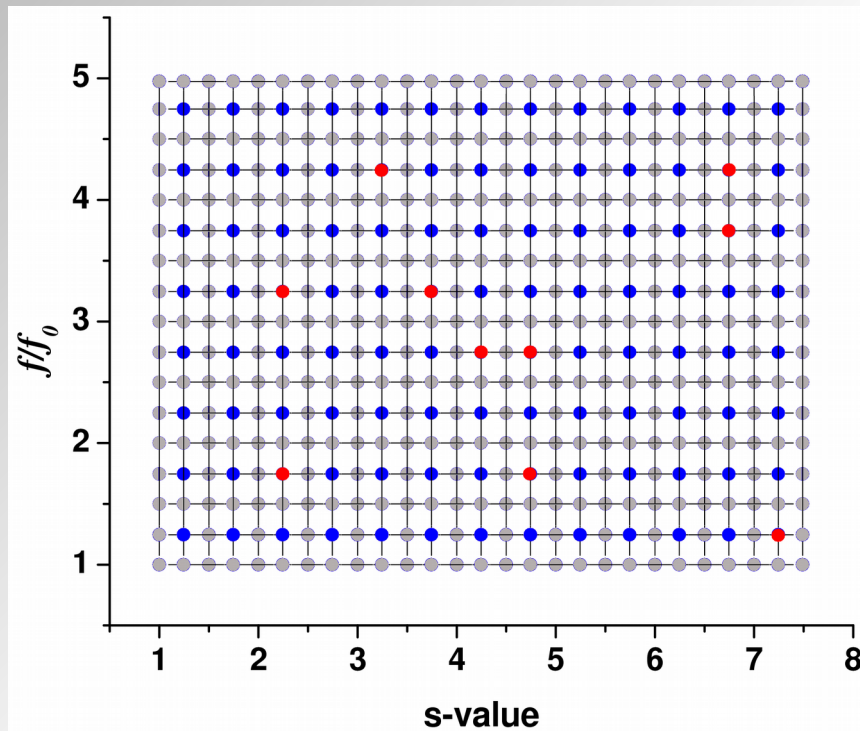
## 2-D Spectrum Analysis - Refinement:

**Step 10: Complete the square and shift the grid once more in the Y-direction**



## 2-D Spectrum Analysis - Refinement:

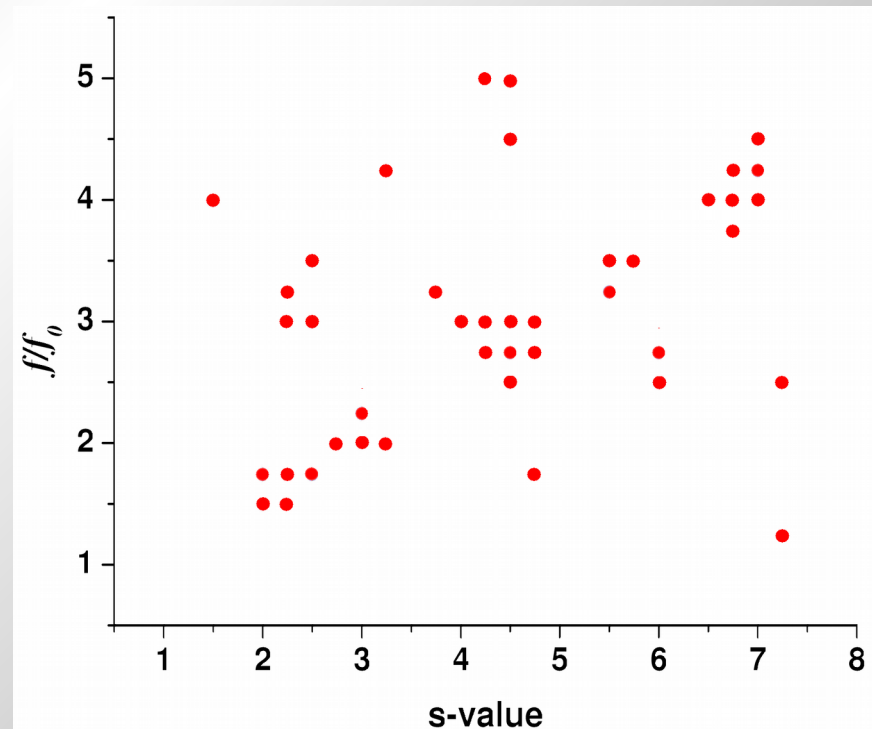
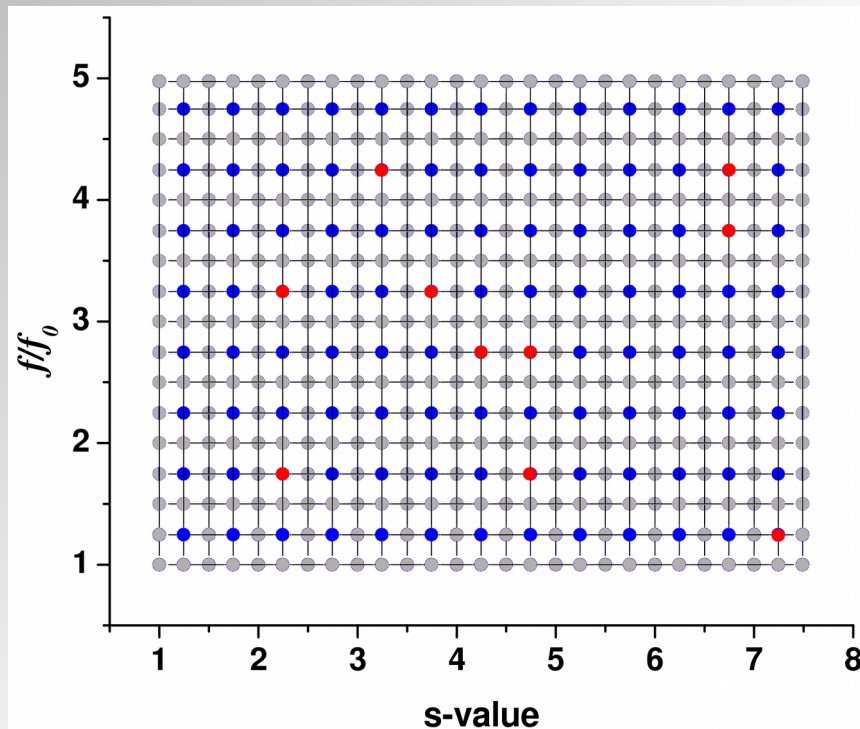
Step 11: Perform NNLS on the new grid





## 2-D Spectrum Analysis - Refinement:

Step 12: ... and add the non-zero points to the storage array



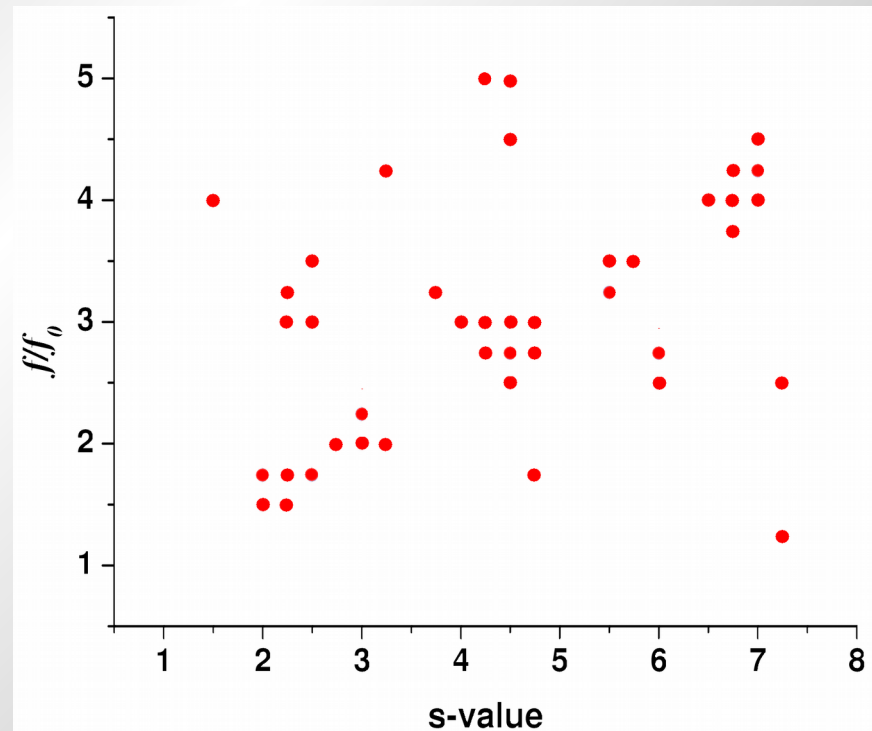
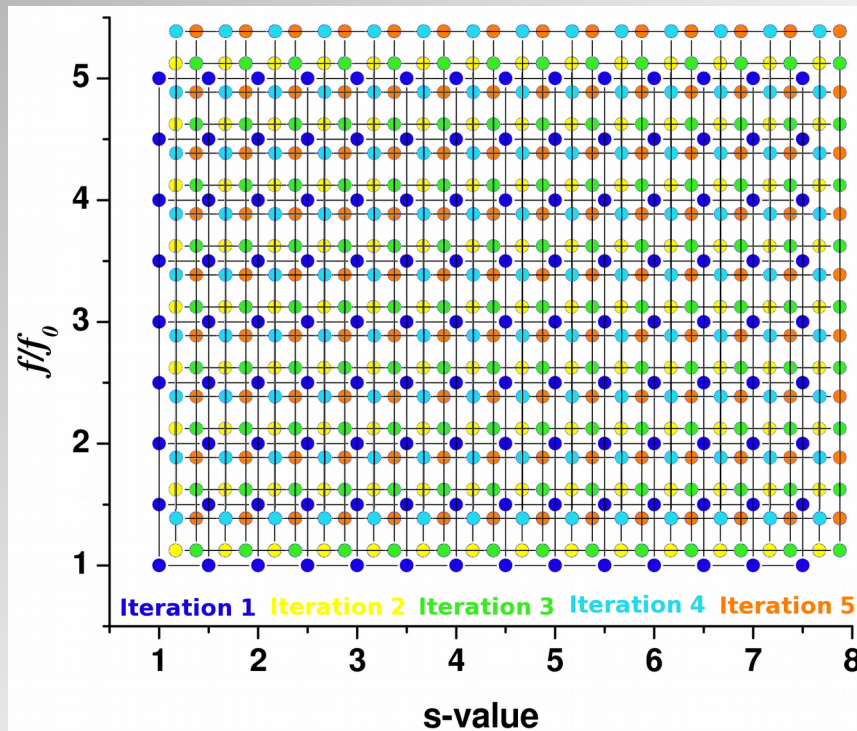
## ***2-D Spectrum Analysis - Refinement:***

---

**Repeat this process  
until the desired grid  
size has been reached**

## 2-D Spectrum Analysis - Refinement:

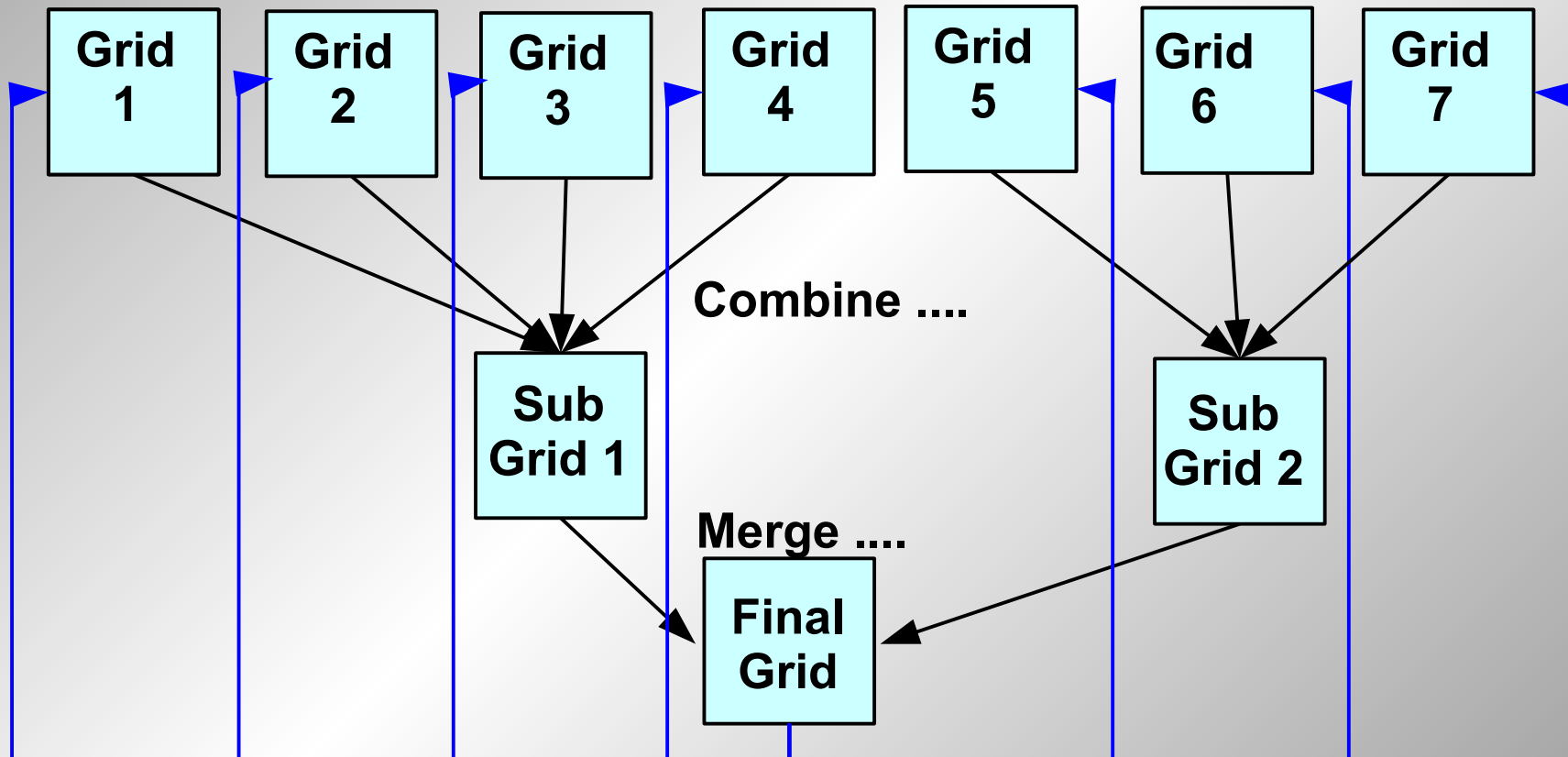
Divide and Conquer approach – evaluate multiple grids slightly off-set against each other, and accumulate results:



Final result is fairly sparse, but it is also degenerate, includes false positives and needs further refinement. It can be used to identify regions that contain signal.

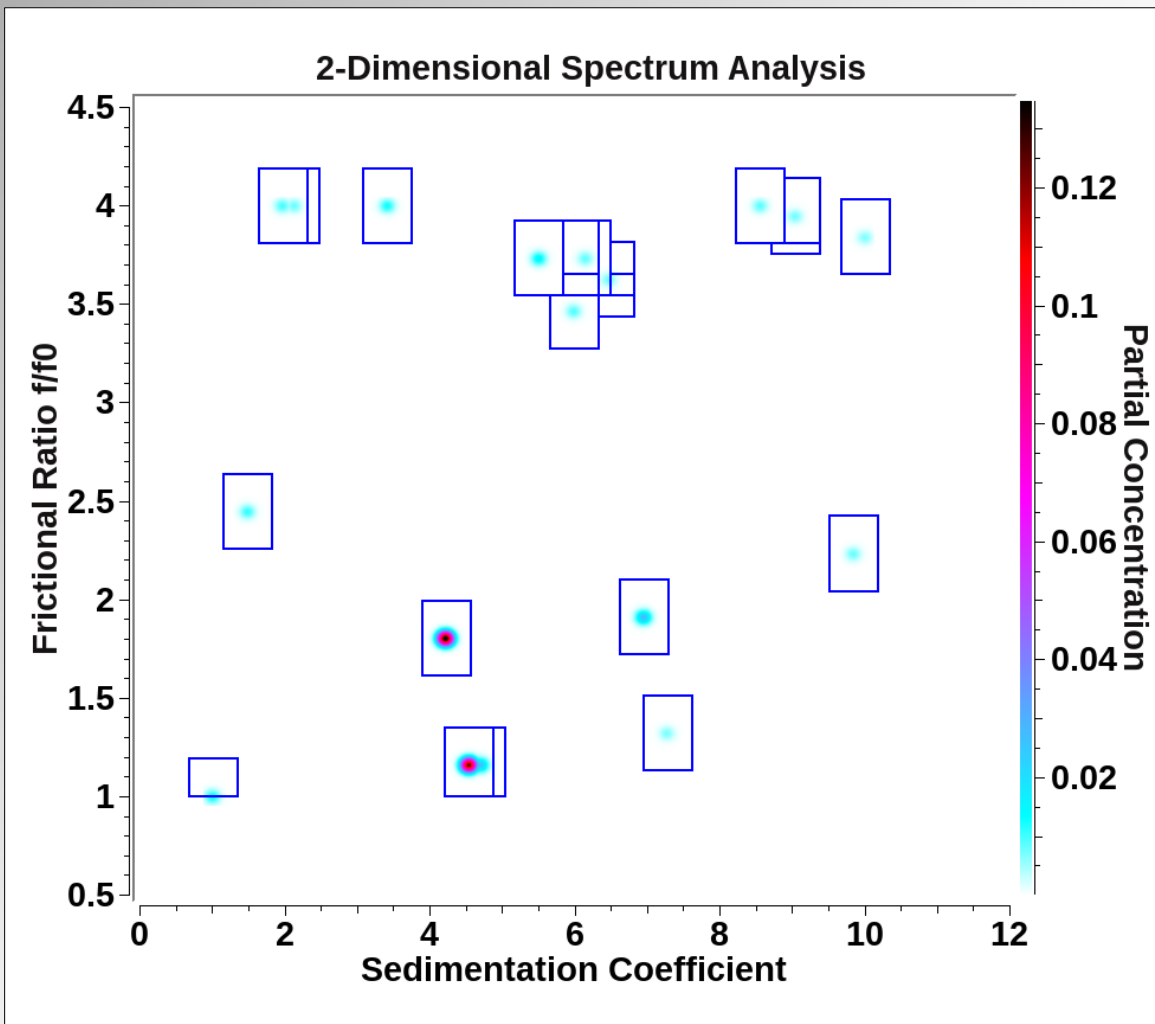
## ***Moving Grid Approach – parallel HPC implementation***

**Calculate each individual grid in parallel ....**



**Evaluate each grid on a different processor, and communicate by MPI  
Iterate until there is no more change ....**

## ***2DSA Result is used to initialize Genetic Algorithms***



The 2DSA finds regions with signal in the parameter space.

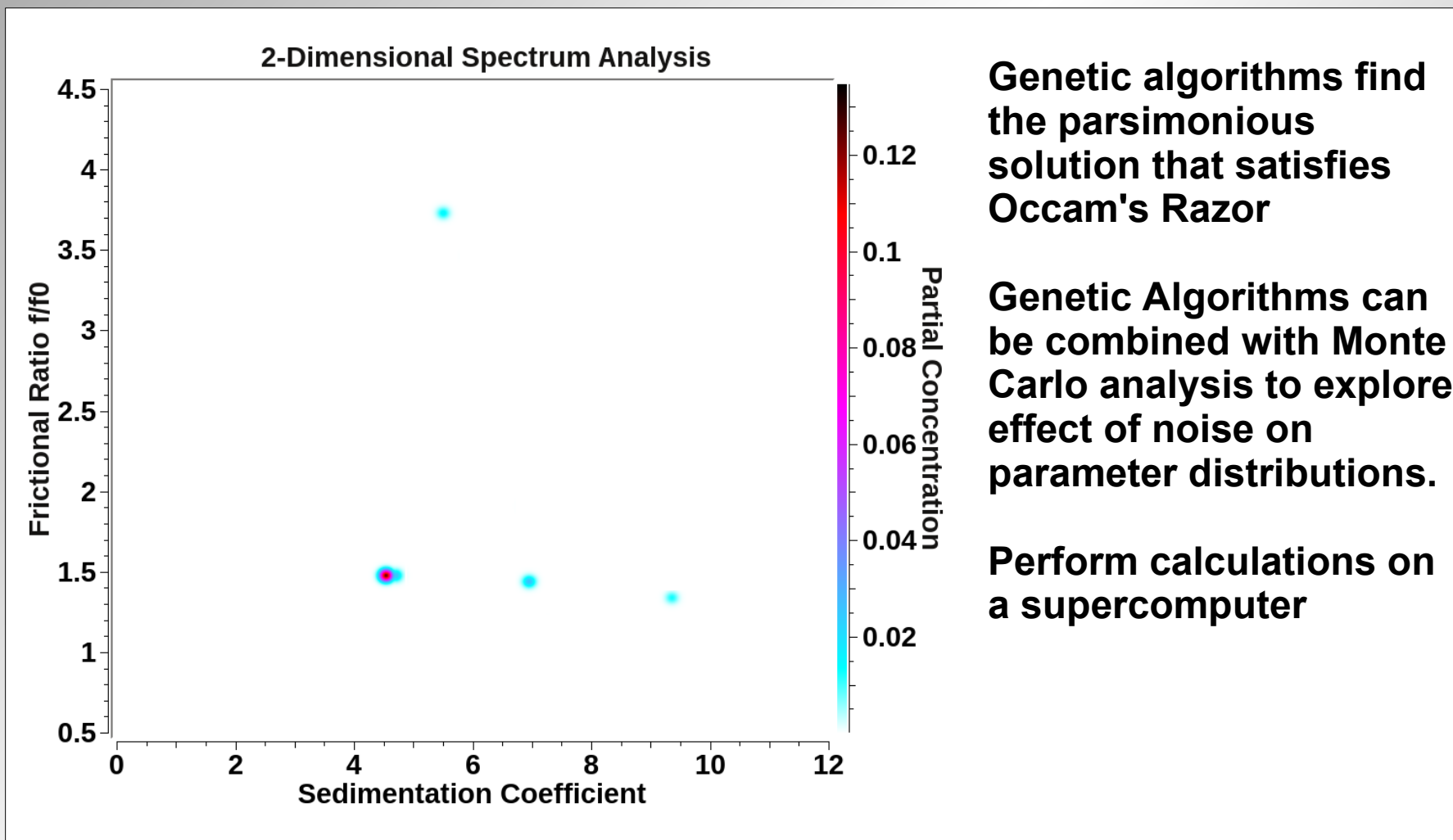
Genetic Algorithms are used to refine 2DSA solutions and to remove false positives through

*“Parsimonious Regularization”*

Initialize GA with 2DSA results

Perform calculations on a supercomputer

## ***2DSA Result is used to initialize Genetic Algorithms***



# **Genetic Algorithms (GA)**

**Genetic Algorithms (also called evolutionary programming)  
provide a stochastic optimization method**

*John H Holland, Adaption in Natural and Artificial Systems, 1975, U. of Michigan Press*

*John R Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, 1992, MIT Press*

**Based on nature – evolutionary paradigm**

**Mutation, recombination, deletion, insertion, crossover operators**

**Multiple populations (“demes”) are allowed to compete, limited migration rates between demes are allowed.**

**Random number generators are used to manipulate operators**

**Generational Model – survival of the fittest (...fitting function)**

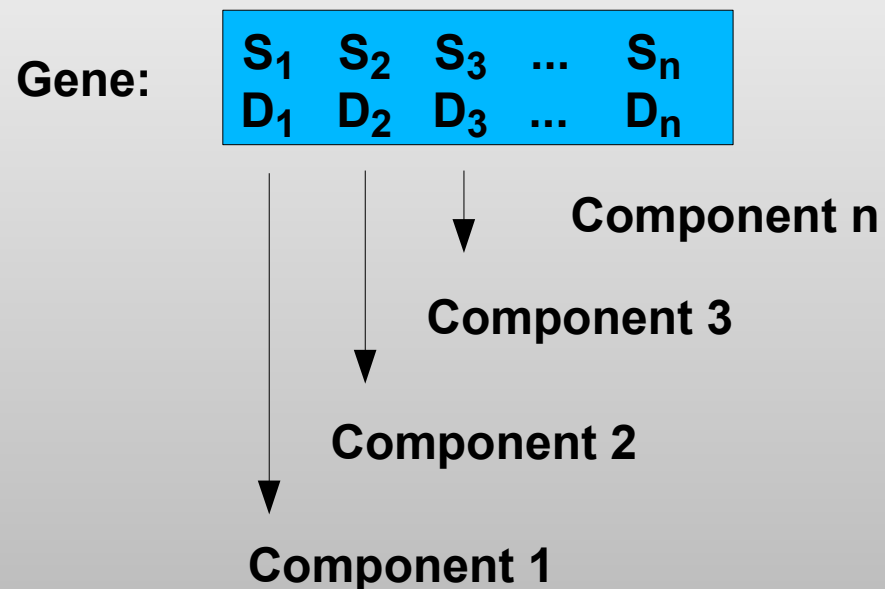
**Generation → iterations, genes → parameter strings, bases → s, D**

**Each solute is simulated with the Lamm equation, solutes are summed**

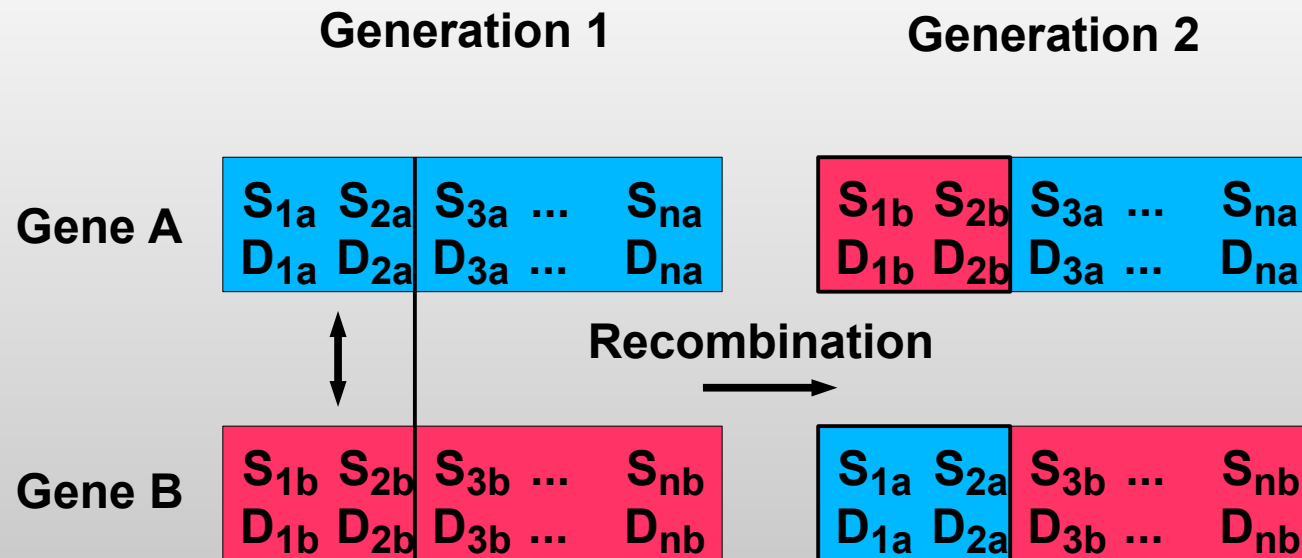


## *GA genes:*

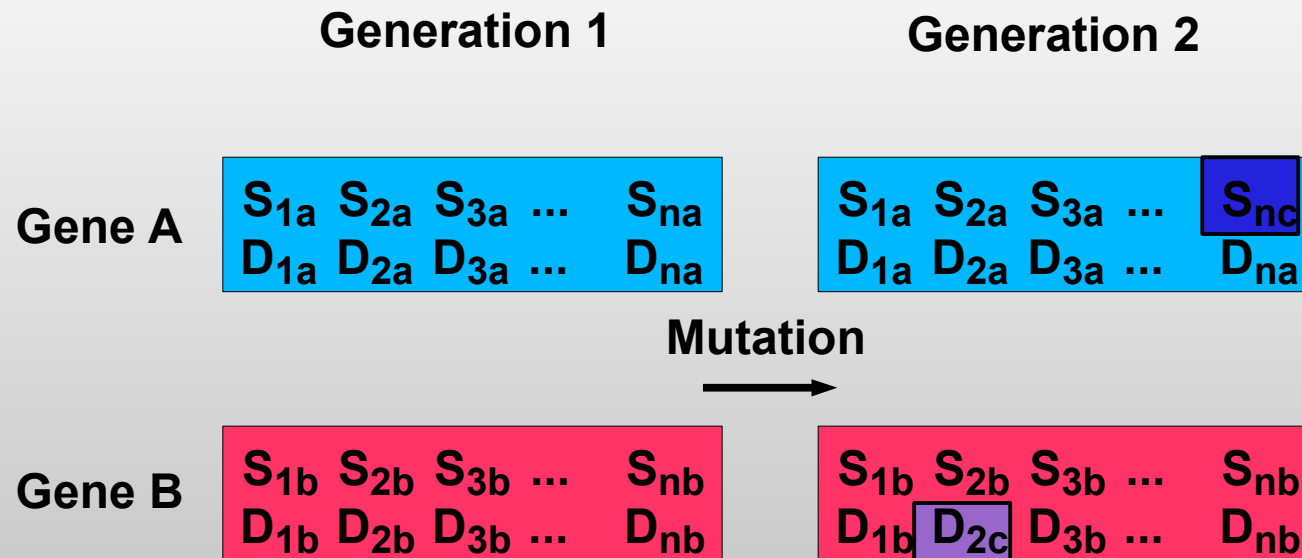
Genes are strings of parameters, each gene consists of a pair of corresponding sedimentation and diffusion coefficients.



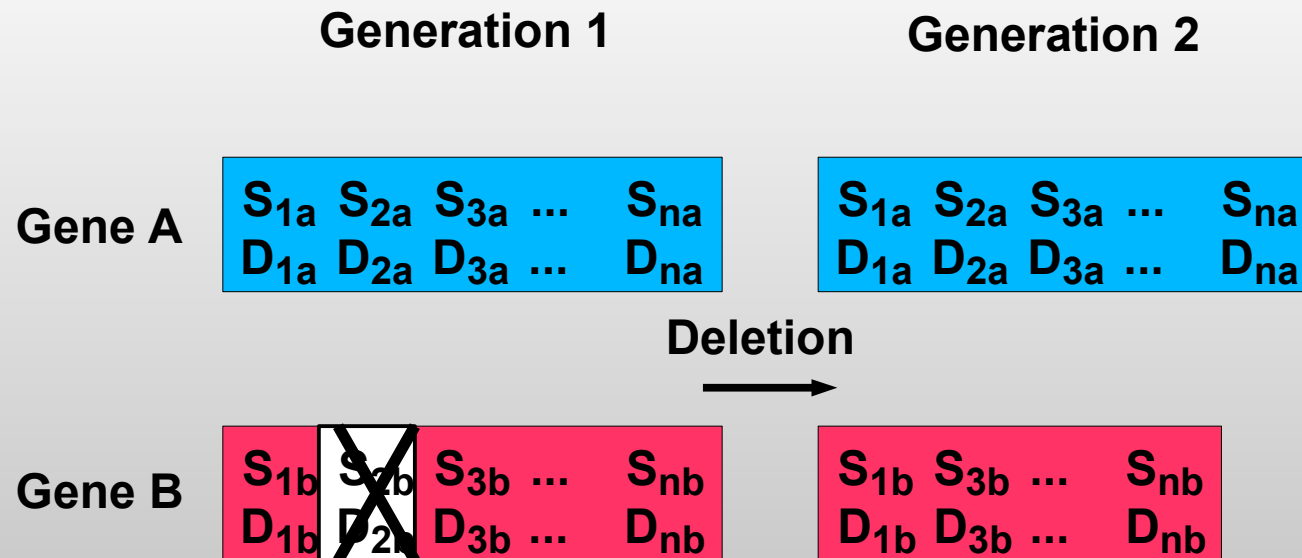
# Crossover/Recombination



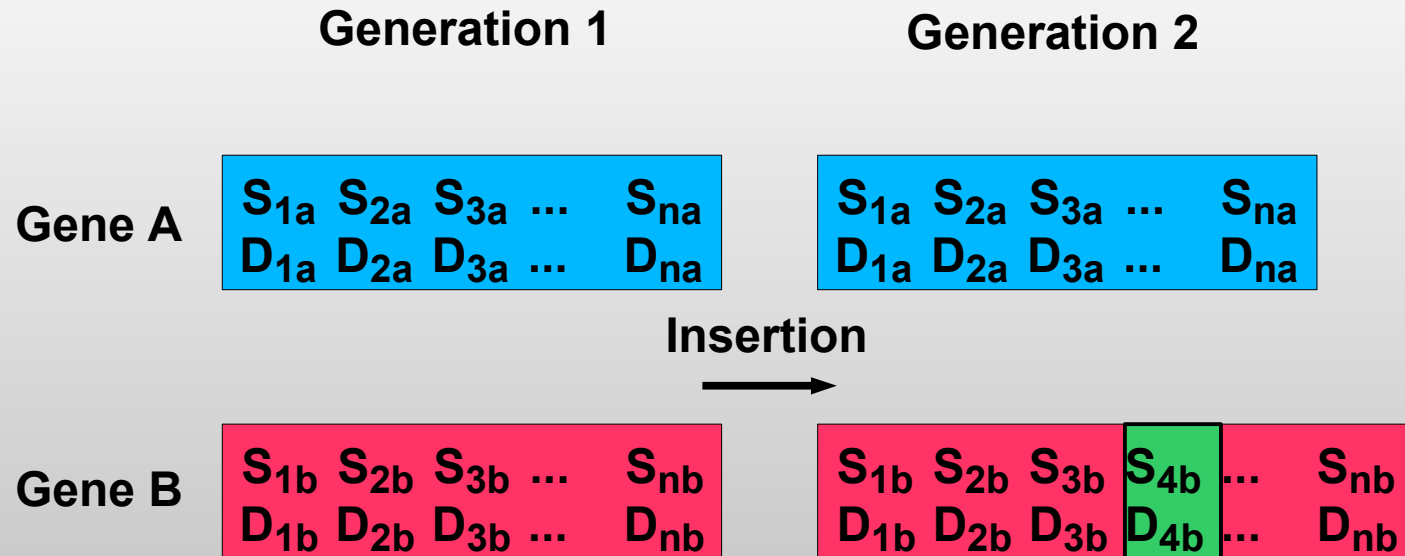
# Mutation



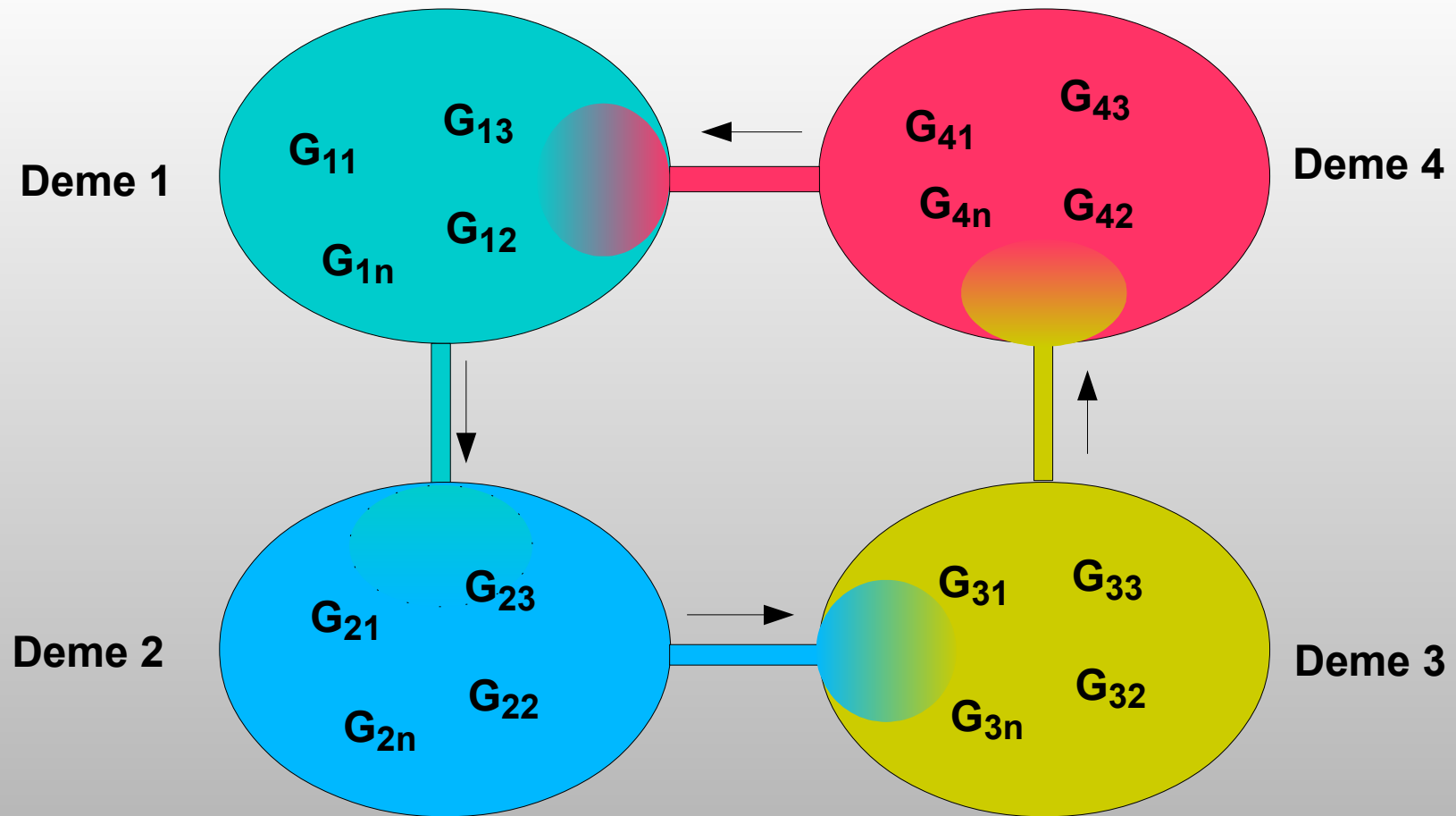
# Deletion



# *Insertion*



# Deme Topology



## ***Initialization of Genetic Algorithms***

**Parameters from all populations are initialized with reasonable starting guesses to create “genes”.**

s-values are initialized using the model independent van Holde – Weischet analysis\*. It provides a good way to assess the limits and possible number of components.

Corresponding diffusion coefficients are randomly assigned based on a reasonable range for  $k=f/f_0$  values between given limits (i.e. 1-4):

$$D = \frac{RT}{18 \pi N (k \eta)^{2/3}} \sqrt{\frac{2(1 - \bar{v} \rho)}{s \bar{v}}}$$

*\*Demeler, B. and K. E. van Holde. Sedimentation velocity analysis of highly heterogeneous systems. (2004). Anal. Biochem. Vol 335(2):279-288*



## ***Approach and Implementation - Initialization***

Concentration values are determined with NNLS\*, components with values below a threshold are eliminated.

Demes are initially kept isolated

Mutation/Crossover/Recombination operators are applied

Progeny is calculated and this process is iterated

After some iterations, migration rates are applied and nonlinear optimization (Quasi-Newton/Inverse Hessian) is applied for a few iterations.

*\* Lawson, C. L. and Hanson, R. J. 1974. Solving Least Squares Problems. Prentice-Hall, Inc. Englewood Cliffs, New Jersey*

# Parametrically Constrained Spectrum Analysis

Biophysical Journal Volume 106 April 2014 1741–1750

1741

## A Parametrically Constrained Optimization Method for Fitting Sedimentation Velocity Experiments

Gary Gorbet,<sup>†</sup> Taylor Devlin,<sup>†</sup> Blanca I. Hernandez Uribe,<sup>†</sup> Aysha K. Demeler,<sup>†</sup> Zachary L. Lindsey,<sup>‡</sup> Suma Ganji,<sup>†</sup> Sabrah Breton,<sup>†</sup> Laura Weise-Cross,<sup>§</sup> Eileen M. Lafer,<sup>†</sup> Emre H. Brookes,<sup>†</sup> and Borries Demeler<sup>†\*</sup>

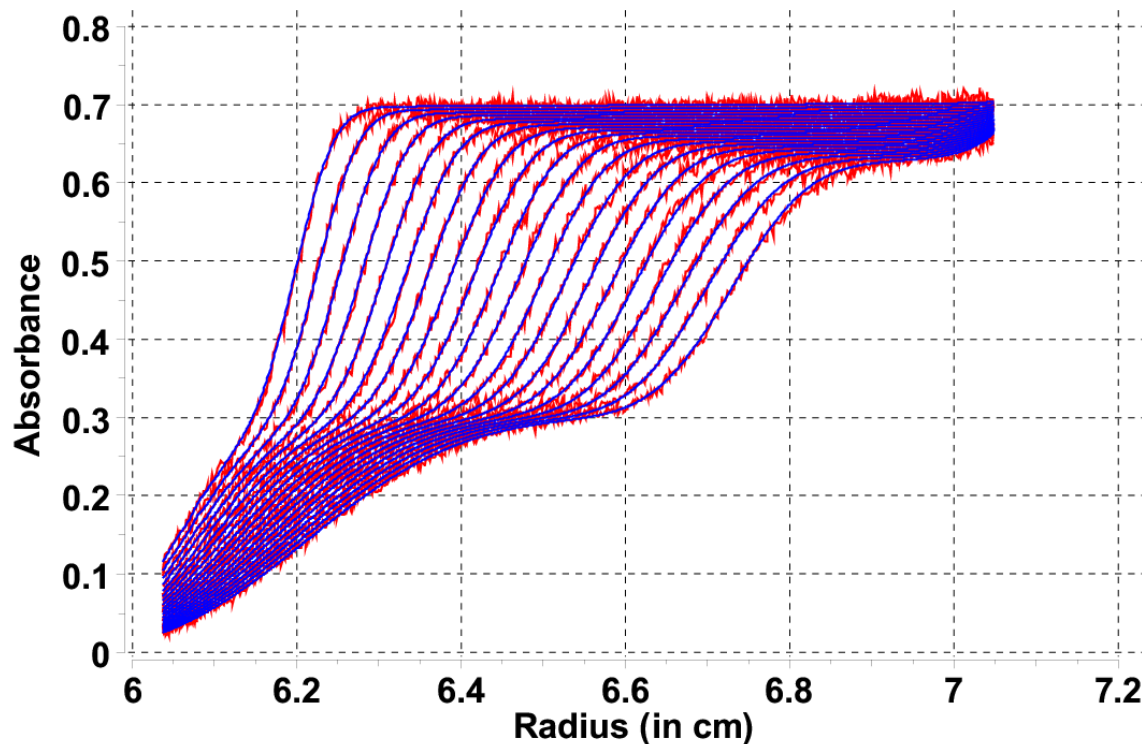
<sup>†</sup>The University of Texas Health Science Center at San Antonio, Department of Biochemistry, San Antonio, Texas; <sup>‡</sup>Texas A&M University, Department of Mechanical Engineering, College Station, Texas; and <sup>§</sup>University of North Carolina at Chapel Hill, Department of Pathology and Laboratory Medicine, Chapel Hill, North Carolina

### Motivation:

**We want a method that can model polymerizing systems that follow a systematic size/shape growth function (for example, end-to-end polymerization) where the anisotropy for each size changes in a predictable fashion**

# Parametrically Constrained Spectrum Analysis

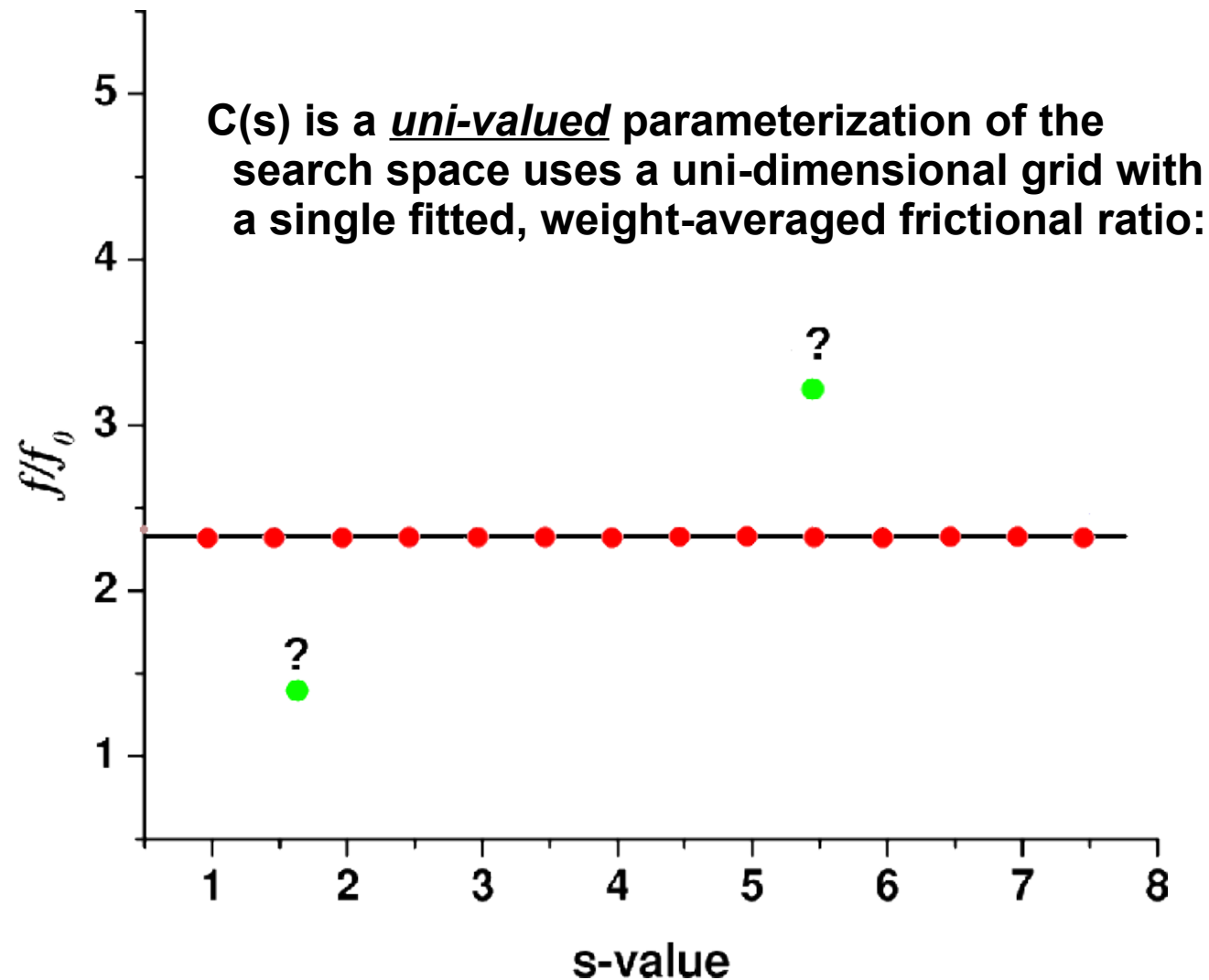
Simple two component system where both components have different anisotropy, fitted with a nonlinear method:



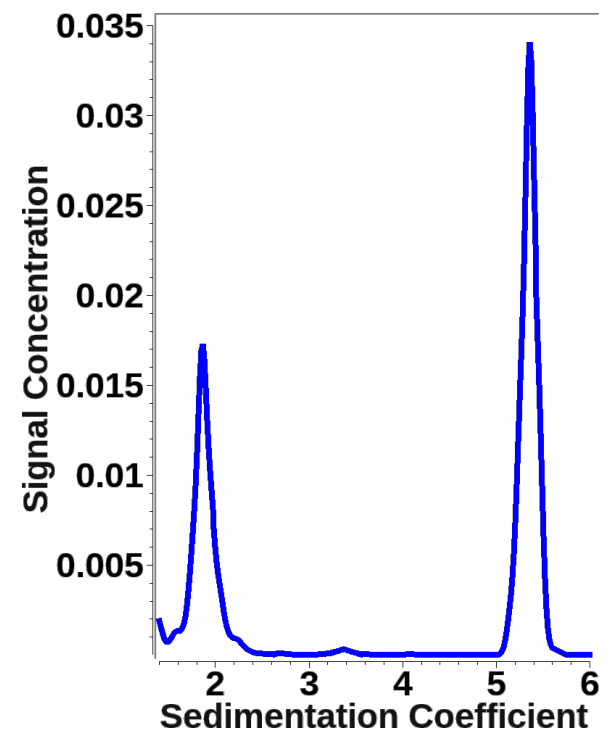
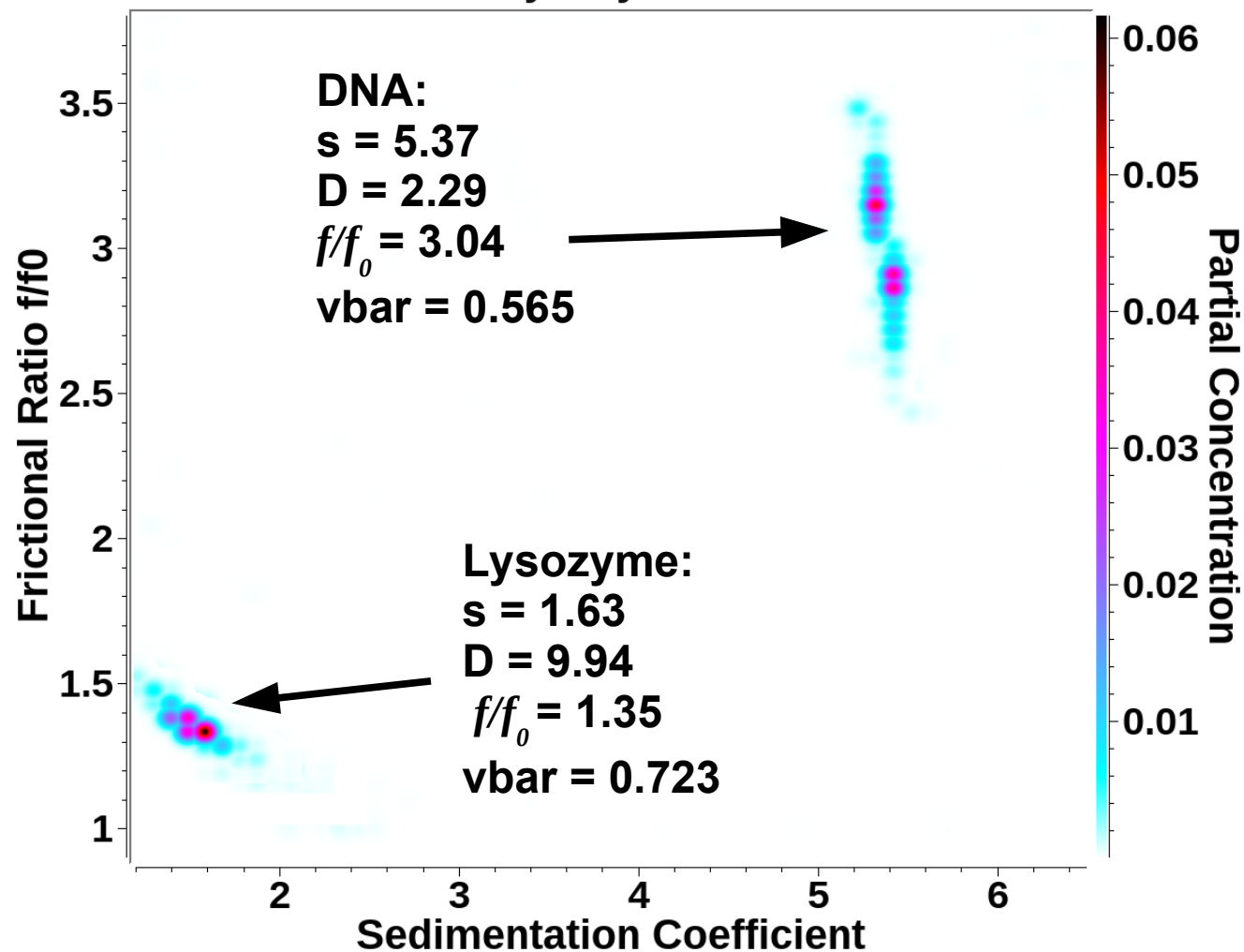
**Lysozyme: 14.3 kDa**  
(globular protein)

**208 bp DNA: 131.0 kDa**  
(extended linear dsDNA)

# Parametrically Constrained Spectrum Analysis

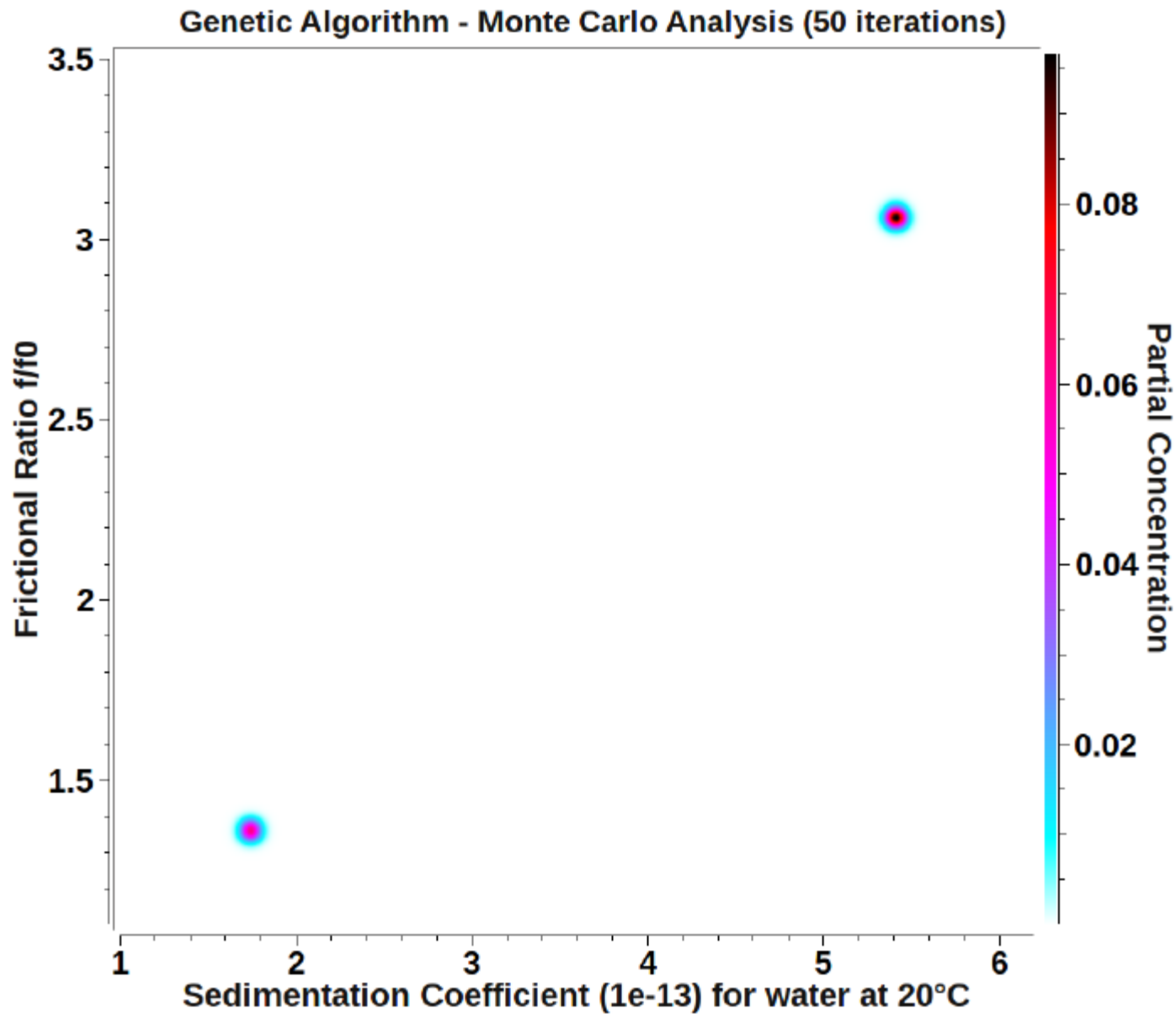


## 2-Dimensional Spectrum Analysis DNA/Lysozyme Mixture



### Goal:

Identify a uni-valued parameterization for the 2-dimensional size and shape domain that models polymer growth as function of its intrinsic shape changes. **Constrain** molecular weight to a single anisotropy.



**Genetic algorithms give the right answer,  
but computationally expensive**

# Parametrically Constrained Spectrum Analysis

Biophysical Journal Volume 106 April 2014 1741–1750

1741

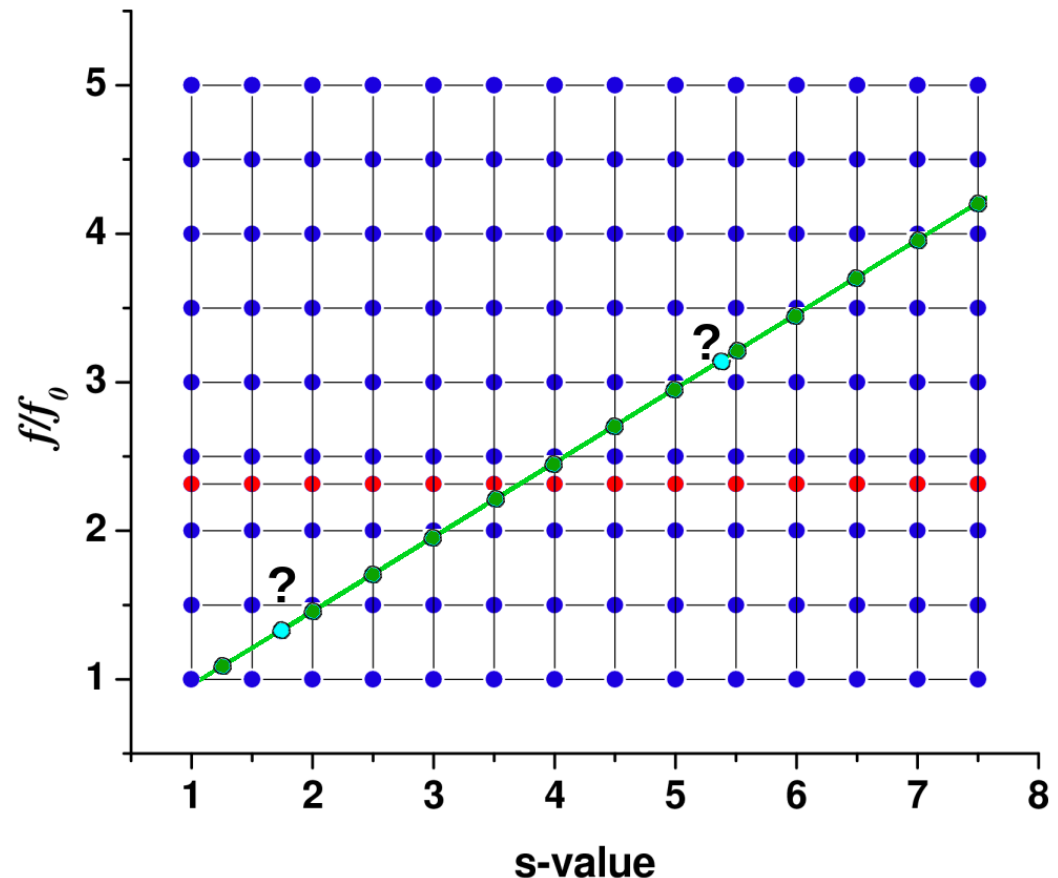
## A Parametrically Constrained Optimization Method for Fitting Sedimentation Velocity Experiments

Gary Gorbet,<sup>†</sup> Taylor Devlin,<sup>†</sup> Blanca I. Hernandez Uribe,<sup>†</sup> Aysha K. Demeler,<sup>†</sup> Zachary L. Lindsey,<sup>†</sup> Suma Ganji,<sup>†</sup> Sabrah Breton,<sup>†</sup> Laura Weise-Cross,<sup>§</sup> Eileen M. Lafer,<sup>†</sup> Emre H. Brookes,<sup>†</sup> and Borries Demeler<sup>†\*</sup>

<sup>†</sup>The University of Texas Health Science Center at San Antonio, Department of Biochemistry, San Antonio, Texas; <sup>†</sup>Texas A&M University, Department of Mechanical Engineering, College Station, Texas; and <sup>§</sup>University of North Carolina at Chapel Hill, Department of Pathology and Laboratory Medicine, Chapel Hill, North Carolina

### Motivation:

We want a **general** method that can model polymerizing systems that follow a systematic size-anisotropy growth function (e.g., end-to-end polymerization) where the anisotropy for each size changes in a predictable fashion, using a **uni-valued** relationship that maps one size to one anisotropy value.

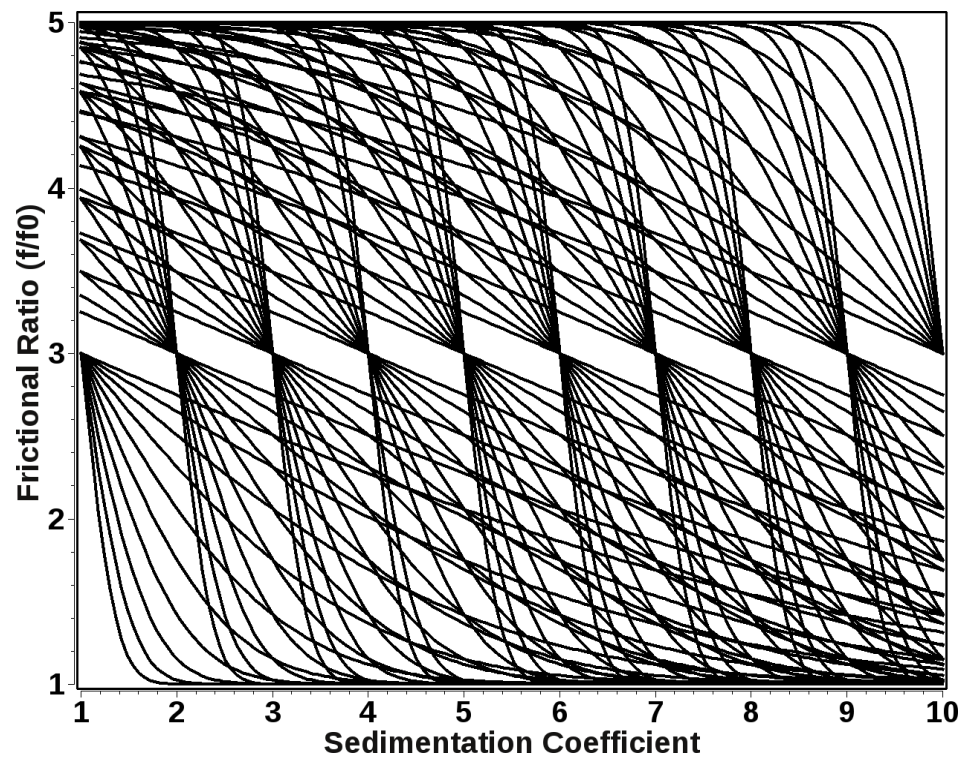
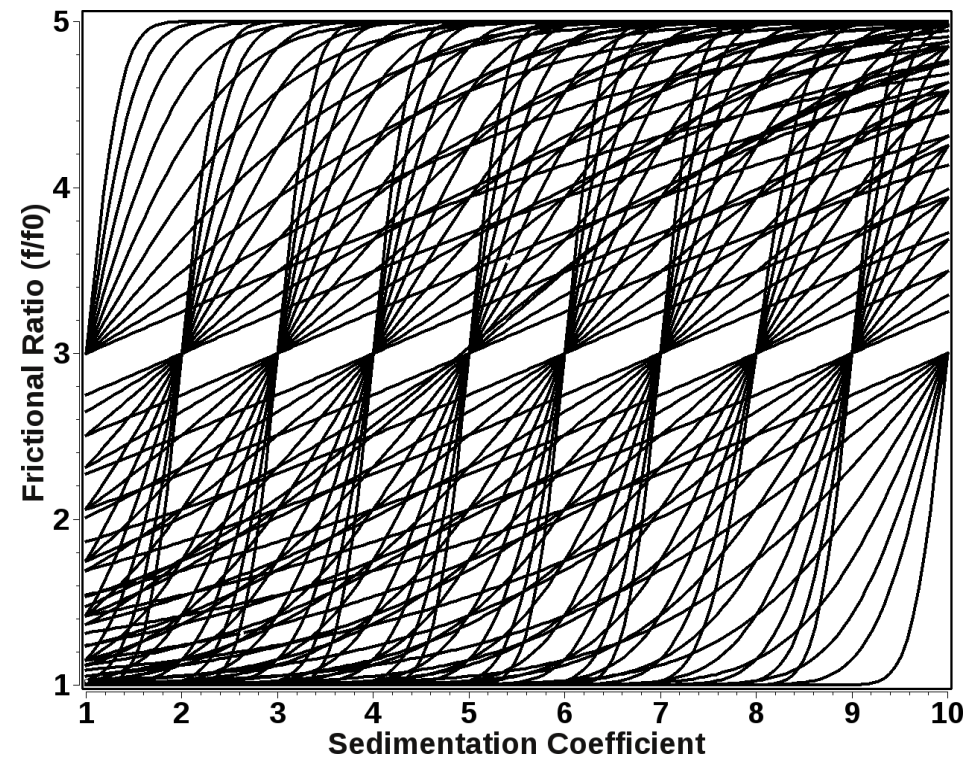
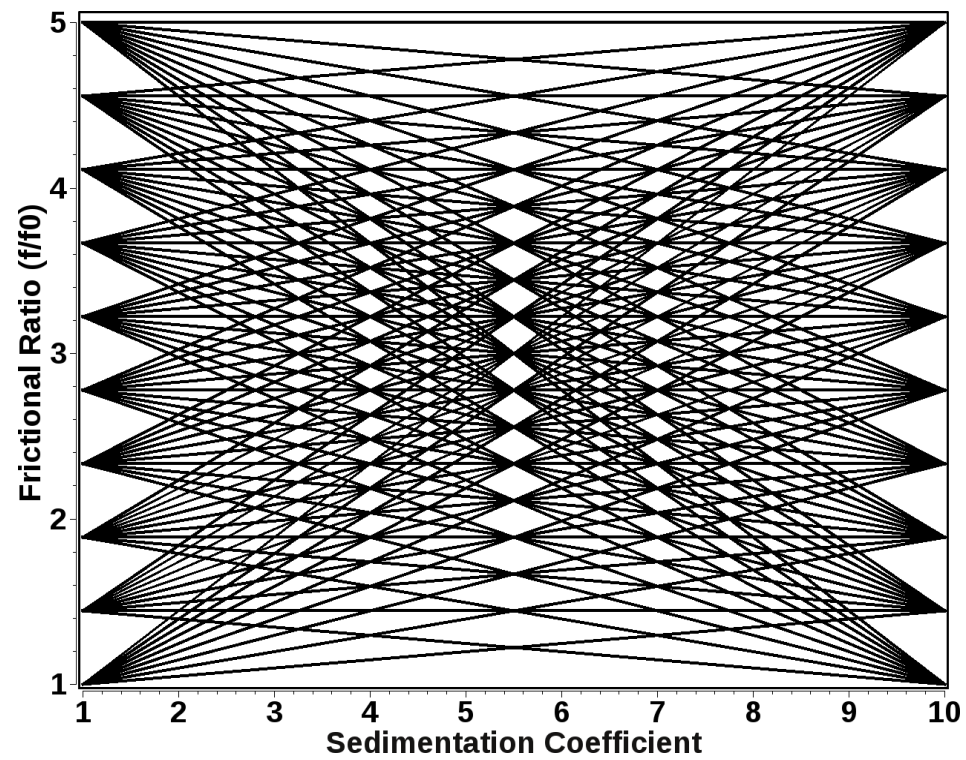




# ***Parametrically Constrained Spectrum Analysis***

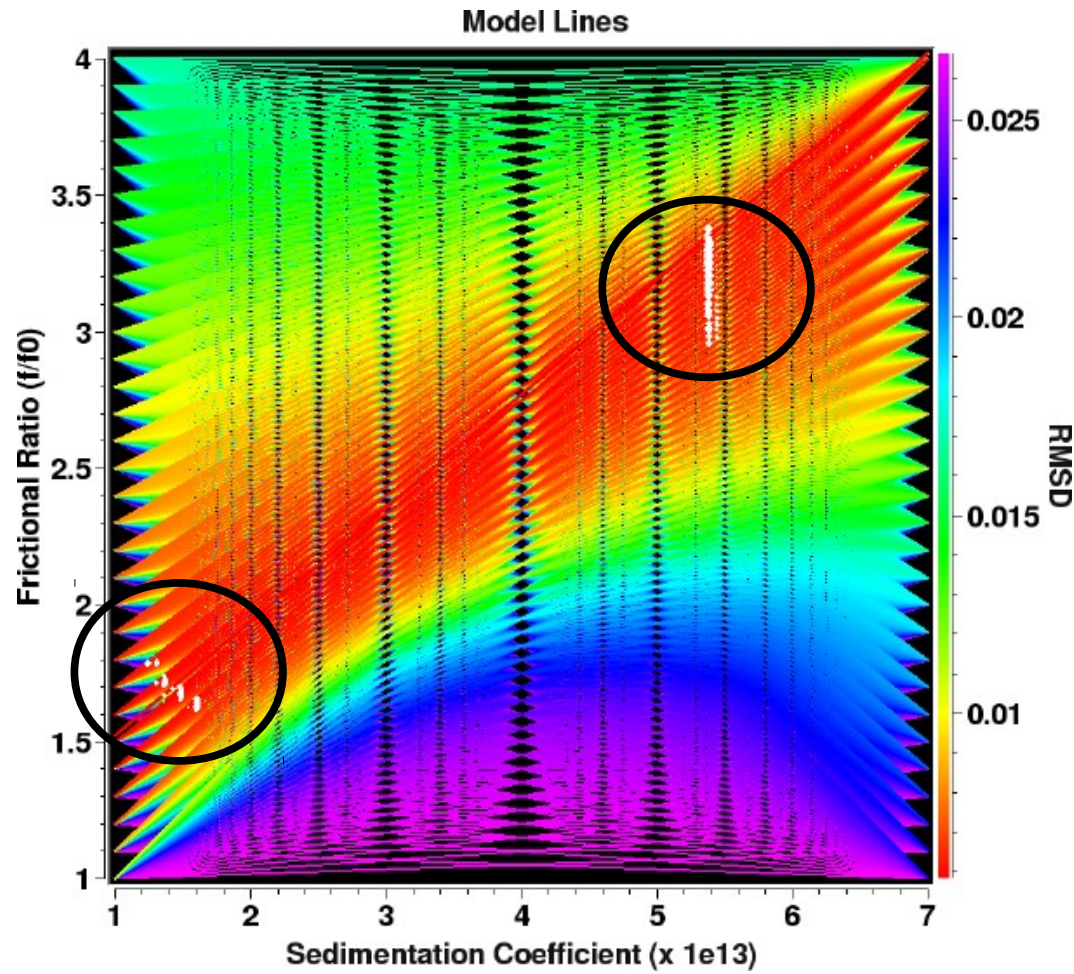
## **PCSA Approach:**

- **Select any single-valued function (straight line, hyperbolic functions, increasing/decreasing sigmoid, exponential growth/decay, etc.)**
- **Generate a discrete grid of functions by varying the function's parameters to achieve a good coverage between the user-selected limits for the 2-dimensional range  $\langle f/f_{0,min}, f/f_{0,max} \rangle$ ,  $\langle S_{min}, S_{max} \rangle$ .**
- **Discretize each function over the 2-dimensional parameter space and solve with finite element and NNLS.**

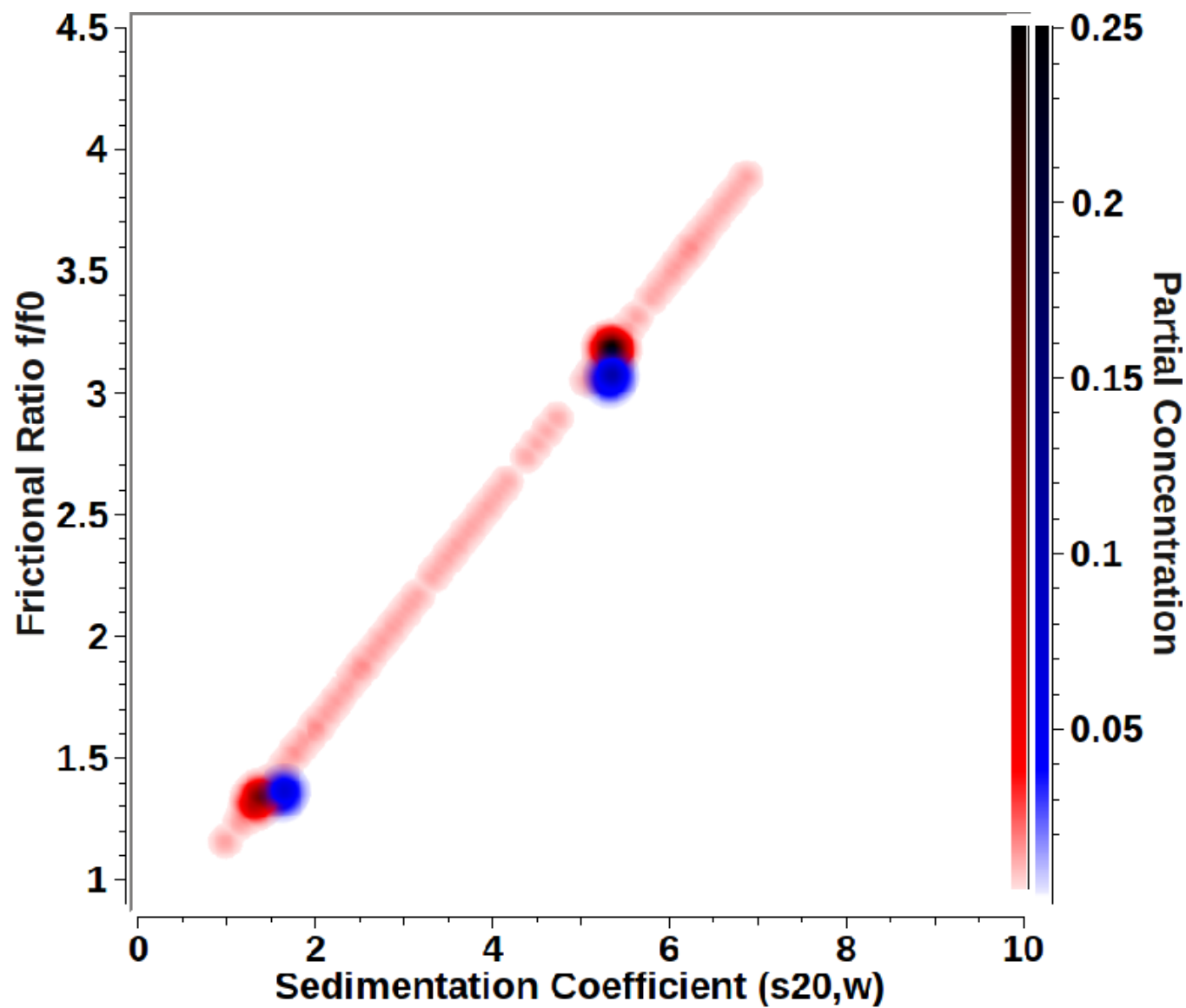


# Parametrically Constrained Spectrum Analysis

Select the NNLS fit with the lowest RMSD and perform a Levenberg-Marquardt fit of the function's parameters to find the best model.

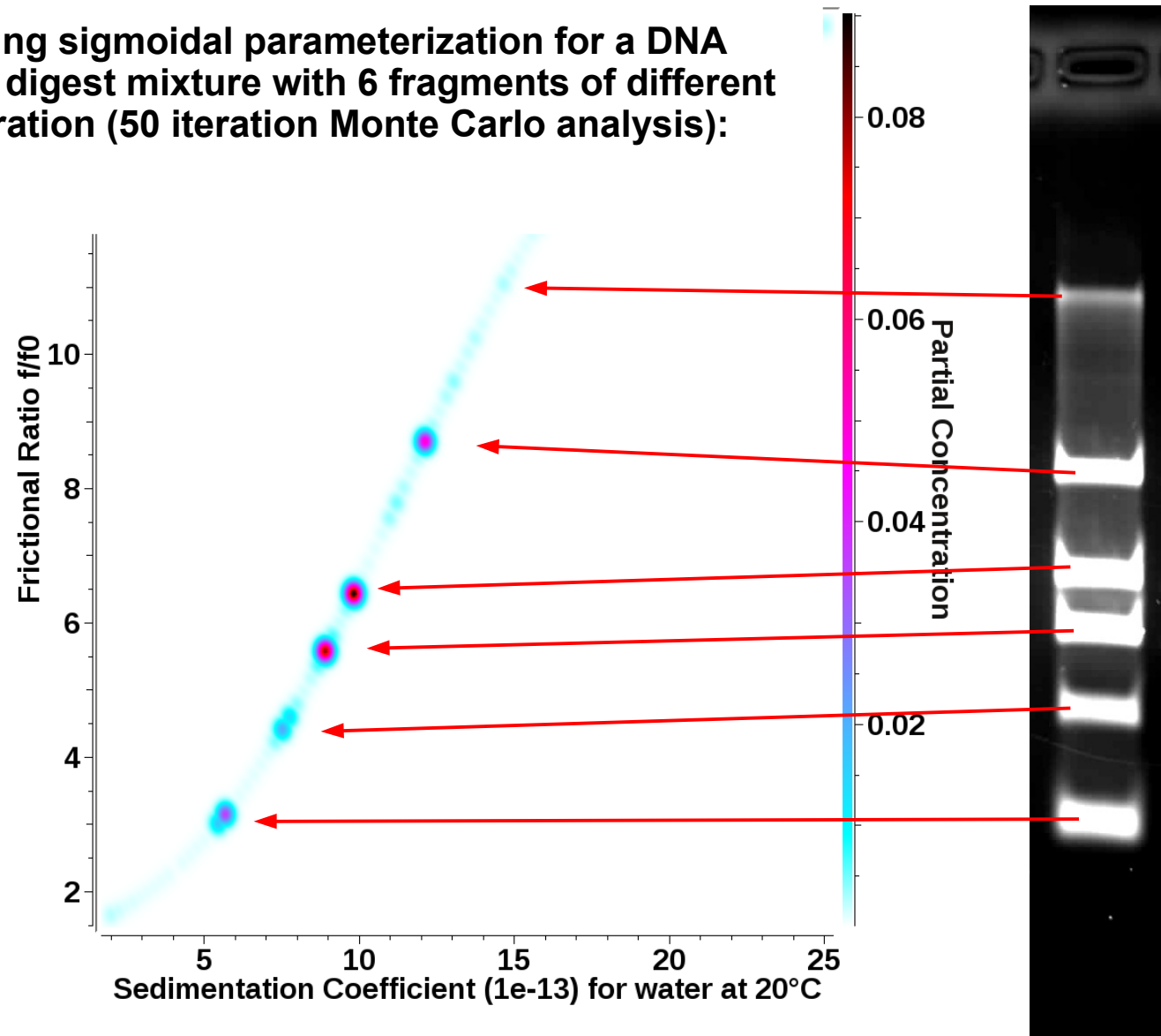


## Overlay plots for PCSA (red) with Genetic Algorithm - Monte Carlo (blue)



# Parametrically Constrained Spectrum Analysis

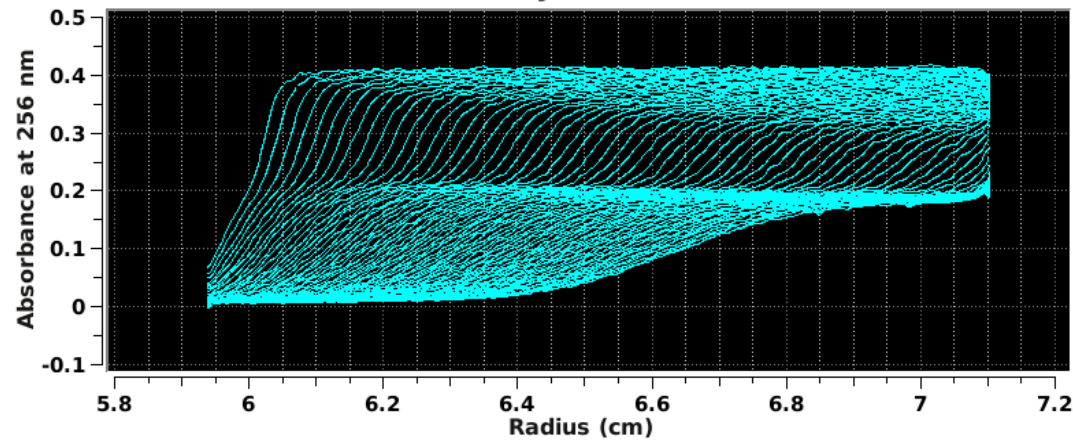
Increasing sigmoidal parameterization for a DNA restriction digest mixture with 6 fragments of different concentration (50 iteration Monte Carlo analysis):



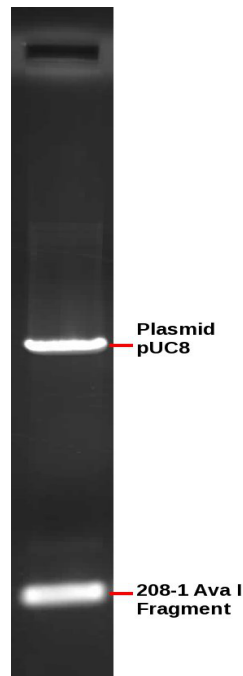


# Parametrically Constrained Spectrum Analysis

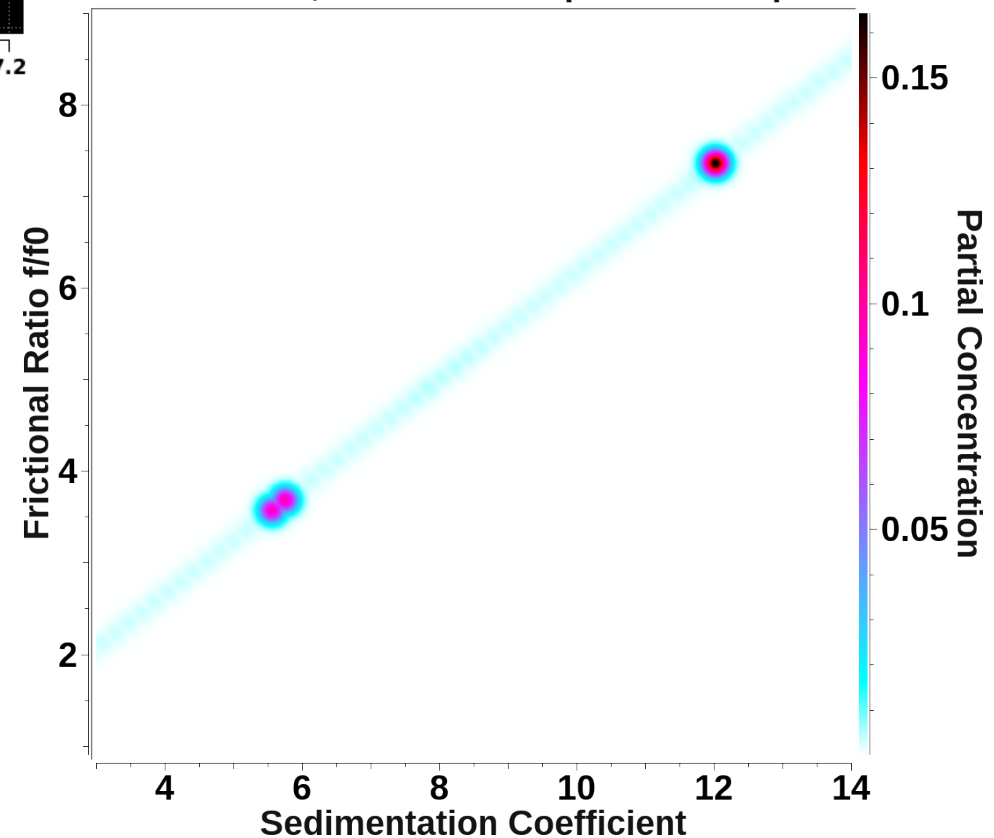
Sedimentation Velocity Data for 100 mM NaCl DNA



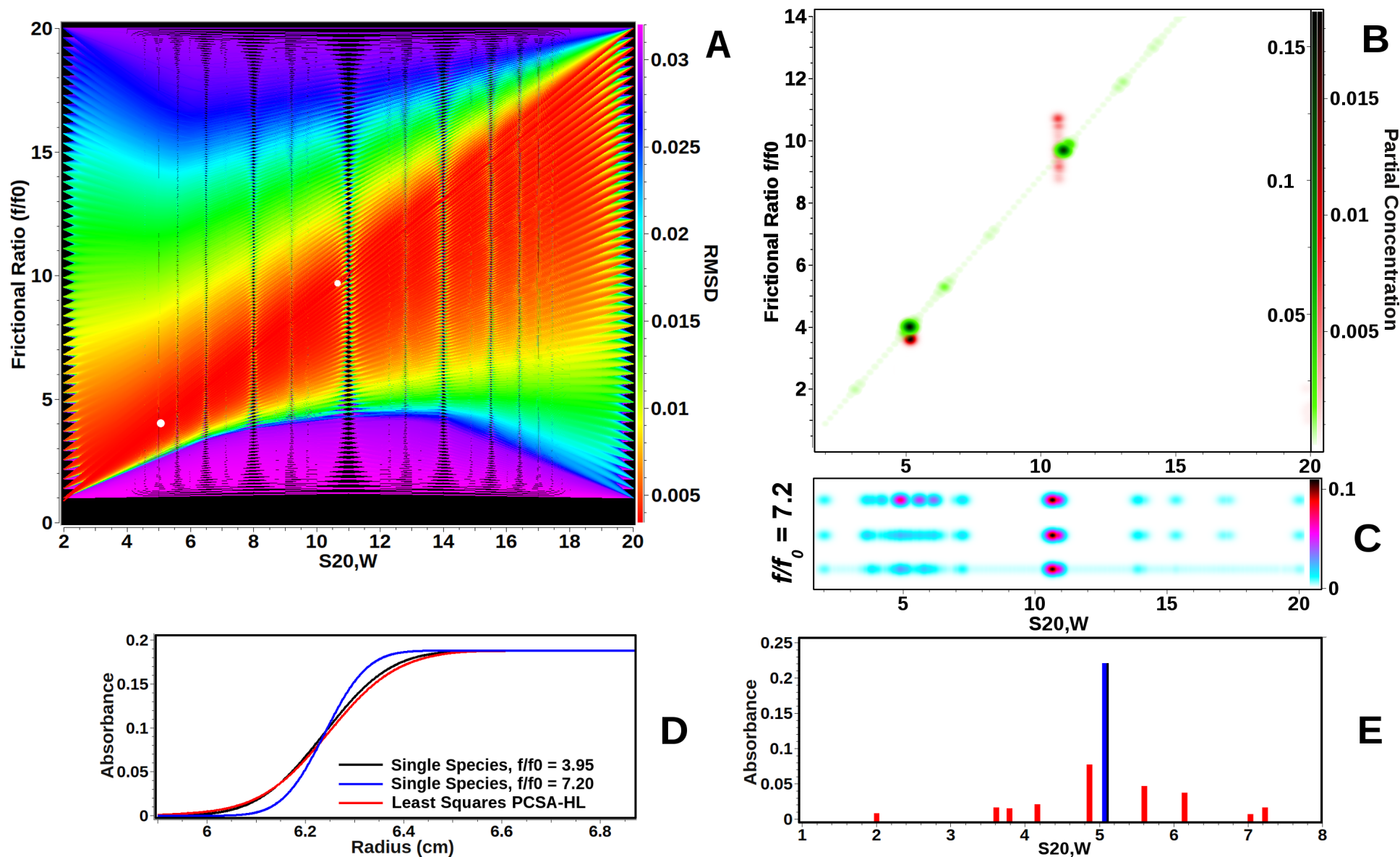
Straight line PCSA  
Monte Carlo results for  
two DNA fragments in  
150 mM NaCl



PCSA, dsDNA - 208 bp and 2811 bp

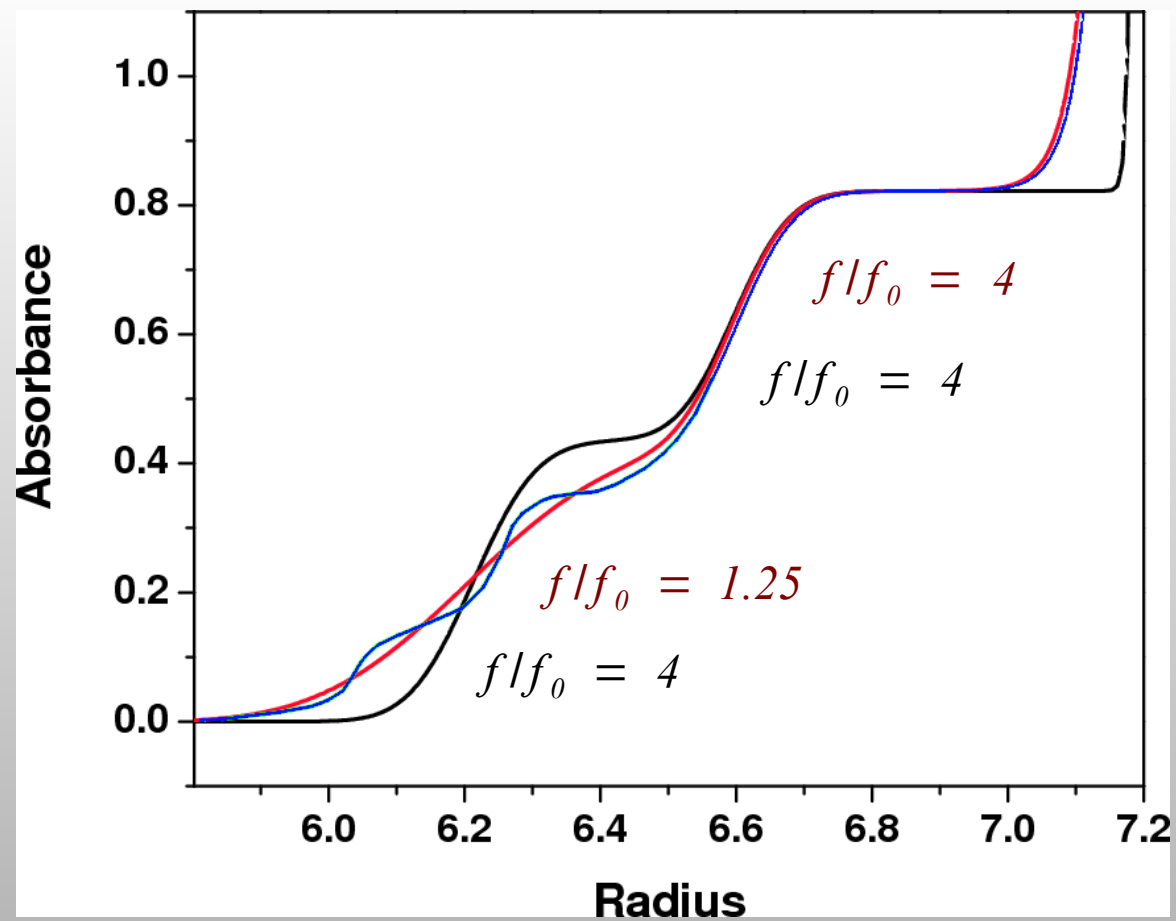


# Parametrically Constrained Spectrum Analysis

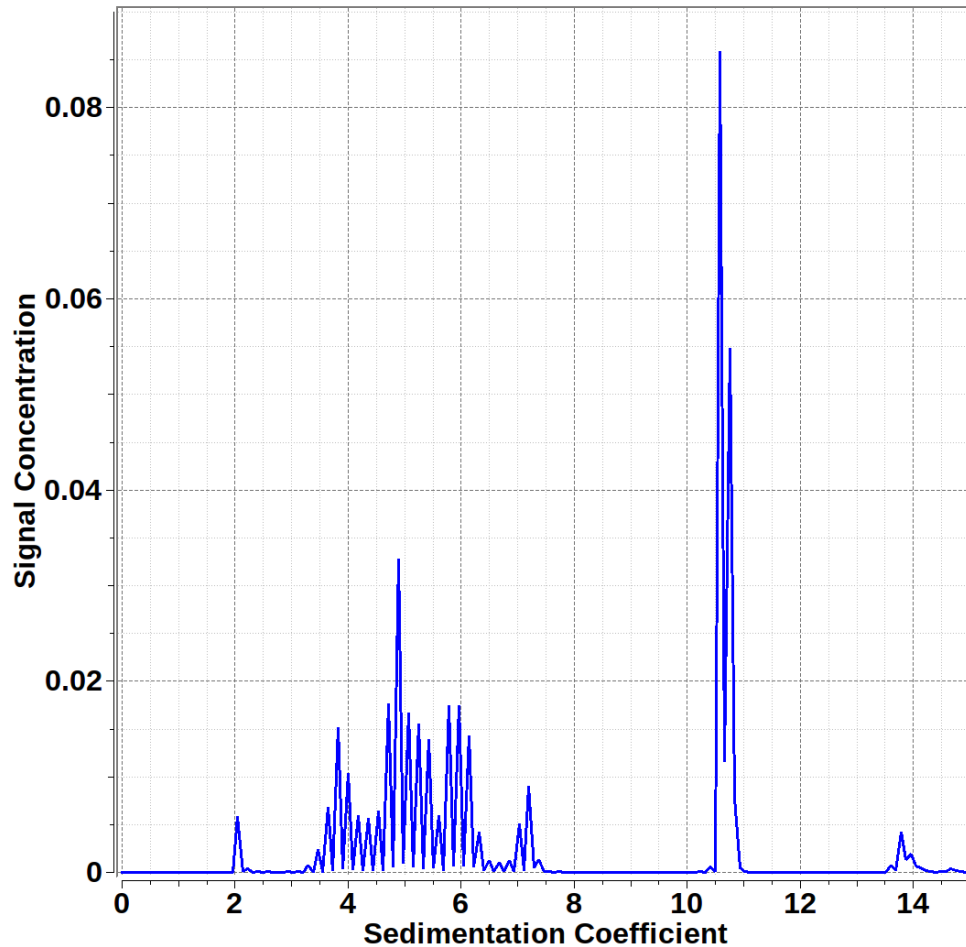




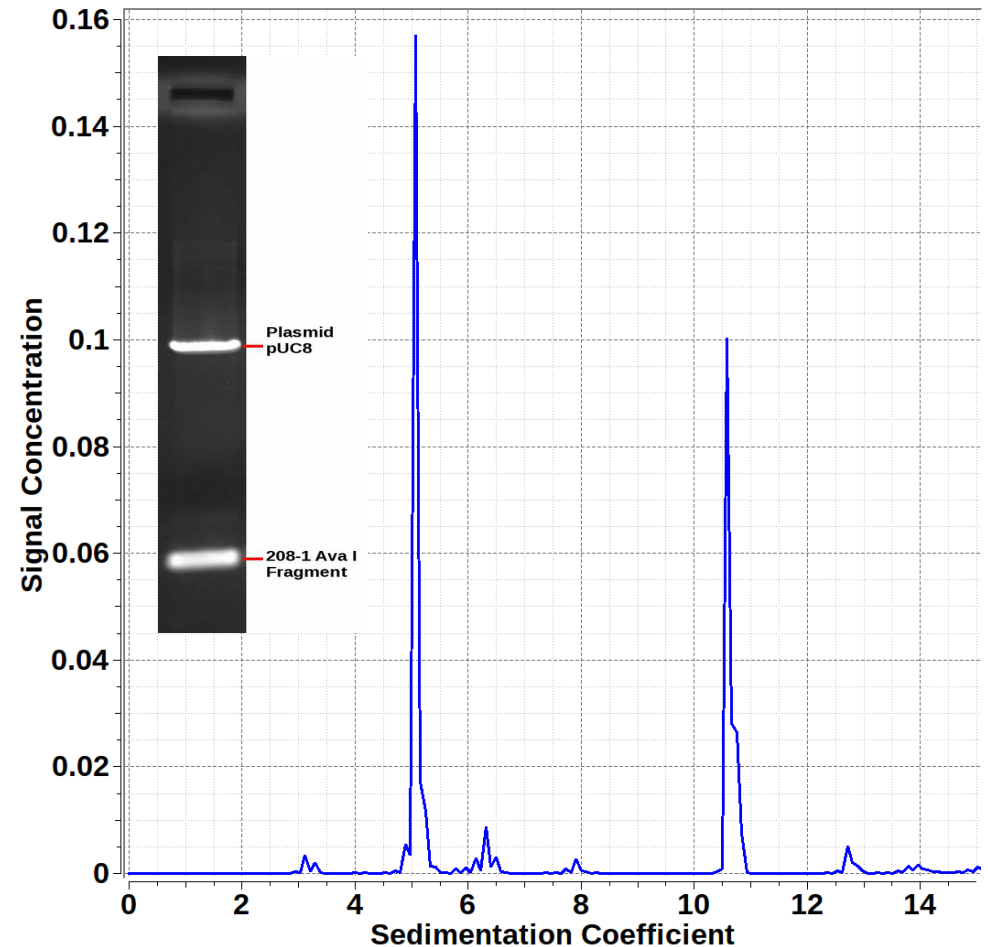
## ***C(s)/C(MW) Method (P. Schuck)***



# Parametrically Constrained Spectrum Analysis

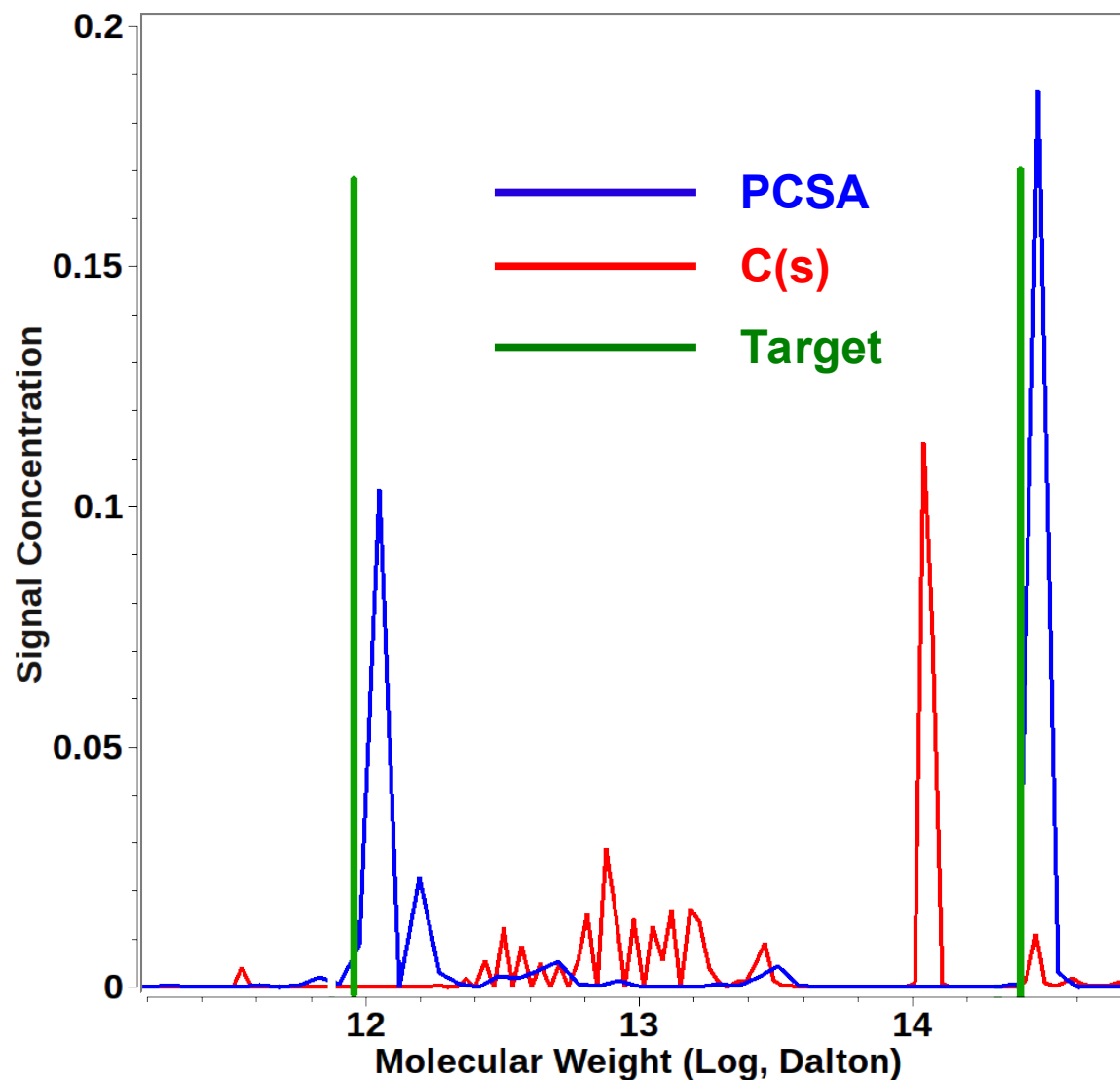


**C(s) analysis,  
high RMSD**



**PCSA analysis,  
low RMSD**

# Parametrically Constrained Spectrum Analysis



**C(s) is unreliable for fitting any velocity data except when anisotropy is constant. The PCSA method produces accurate distributions and mol. weight**