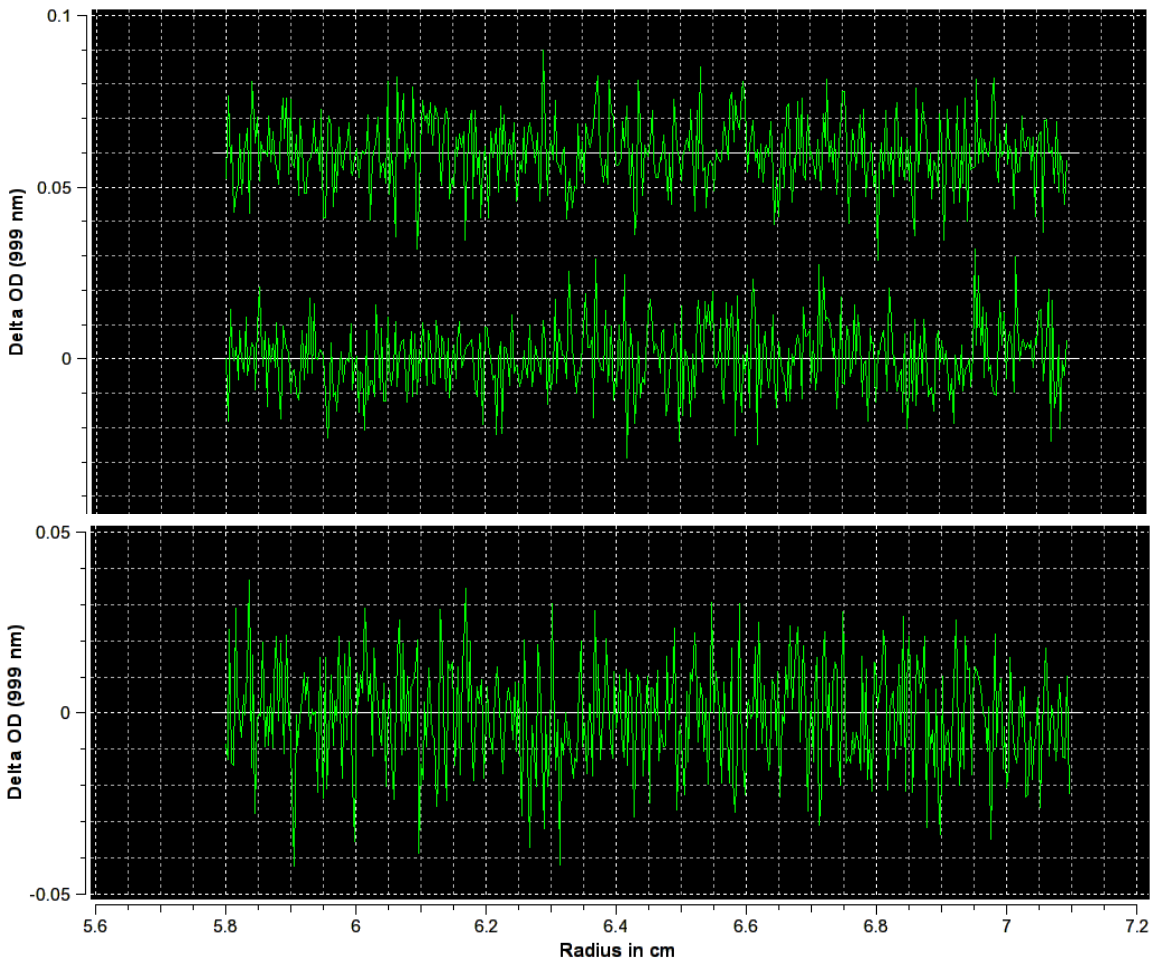
Tale of 2 noisy vectors



Noise of Scan 1

+

Noise of Scan 2

=

Noise * $\sqrt{2}$

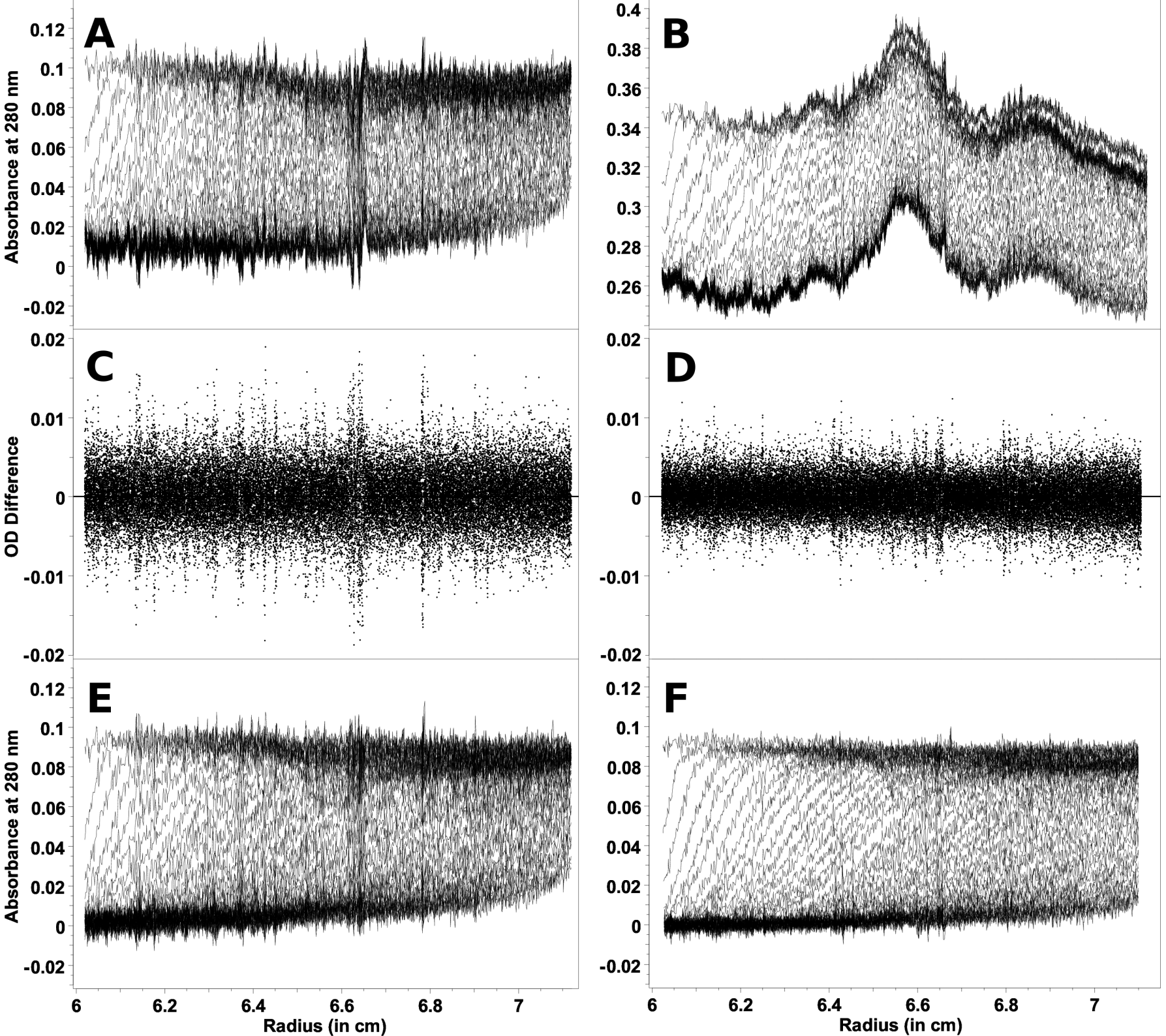
Intensity vs. Absorbance

Optical system considerations

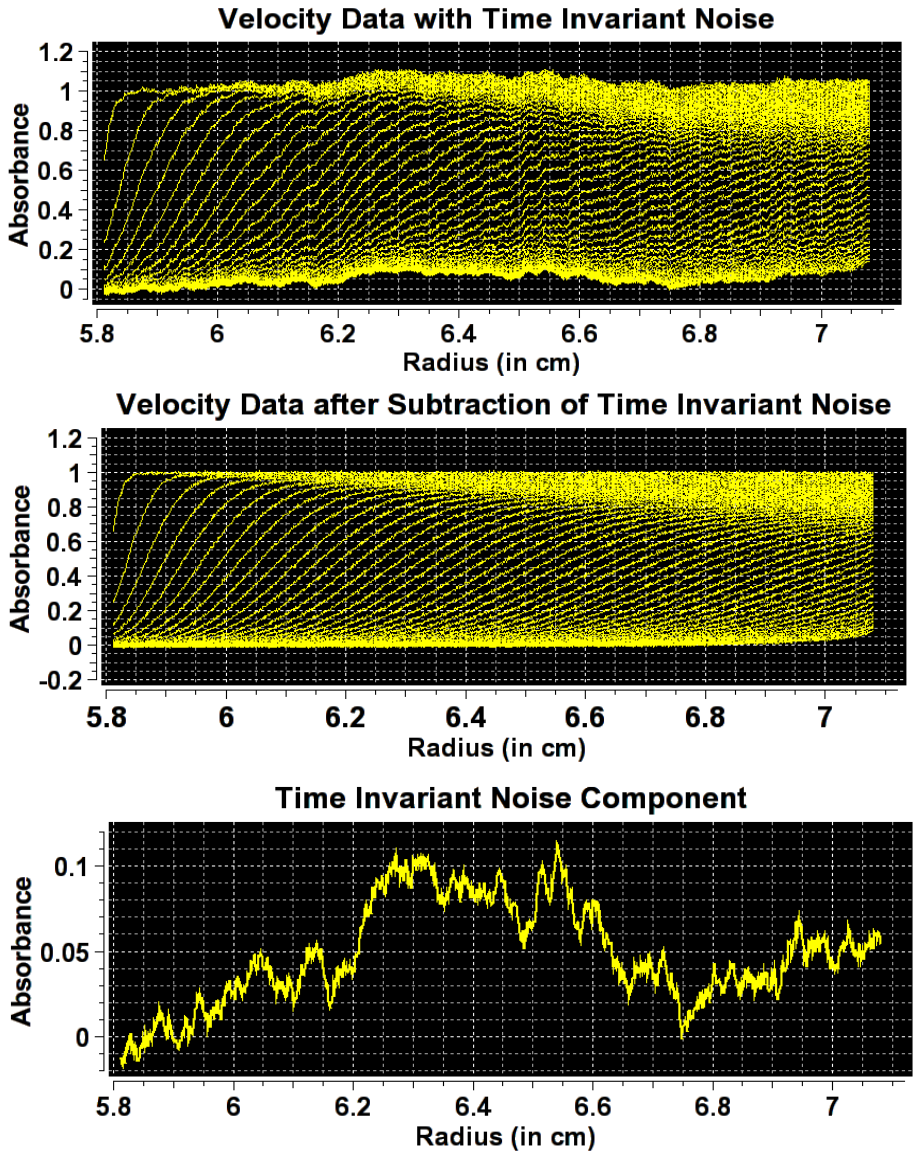
Intensity measurements record the intensity of light passing through one channel. Absorbance measurements record the intensity of light passing through one channel, then record the intensity of the light passing through the reference channel, and subtract it from the first channel. Each channel recording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

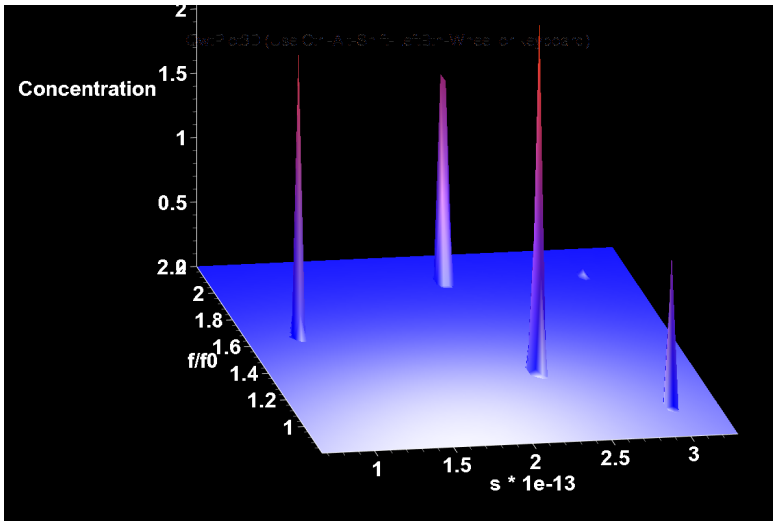
$$scan_1 = signal_1 + N_{s1} + N_{ti}$$
$$scan_2 = signal_2 + N_{s2} + N_{ti}$$
$$scan_1 - scan_2 = signal_1 - signal_2 + N_{s1} + N_{s2}$$

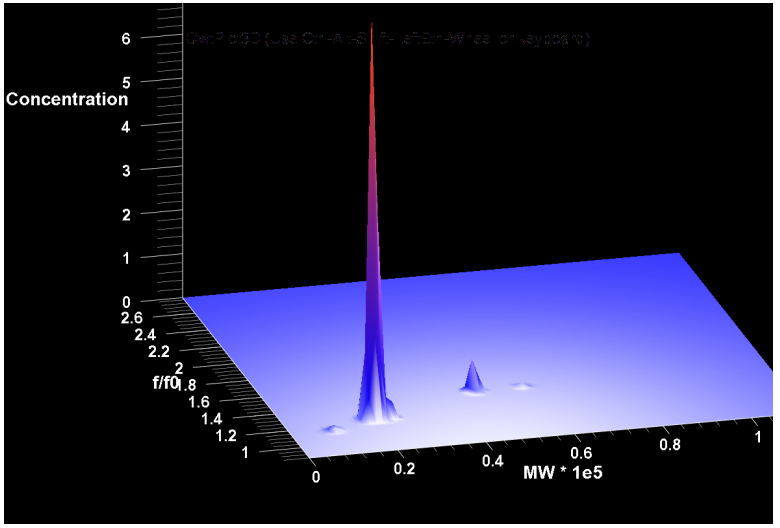
$$N_{s1} + N_{s2} \approx N_{s1}\sqrt{(2)}$$


A
B
C
D
E
F
Absorbance at 280 nm
OD Difference
Absorbance at 280 nm
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
0.4
0.38
0.36
0.34
0.32
0.3
0.28
0.26
0.02
0.01
0
-0.01
-0.02
6
6.2
6.4
6.6
6.8
7
Radius (in cm)
Radius (in cm)

Factors that affect Accuracy – Time-invariant Noise







How can we deal with uncertainty?
A Recipe for Optimal Resolution:

Our solution is affected by random noise, time-invariant noise, the available signal, and its ratio to the random noise.

- Improve signal to noise by:
  - Obtain lots of high quality data
  - Exploit the entire dynamic range of the acquisition system
  - Reduce noise and only use a well-functioning instrument
  - Remove systematic noise from experimental data
- Replace fitting parameters with experimentally determined values from separate experiments
- Explore the parameter space with a grid method, then parsimoniously regularize solution with GA, and use Monte Carlo to explore confidence regions
- Perform global fits for multiple experimental conditions to improve signal

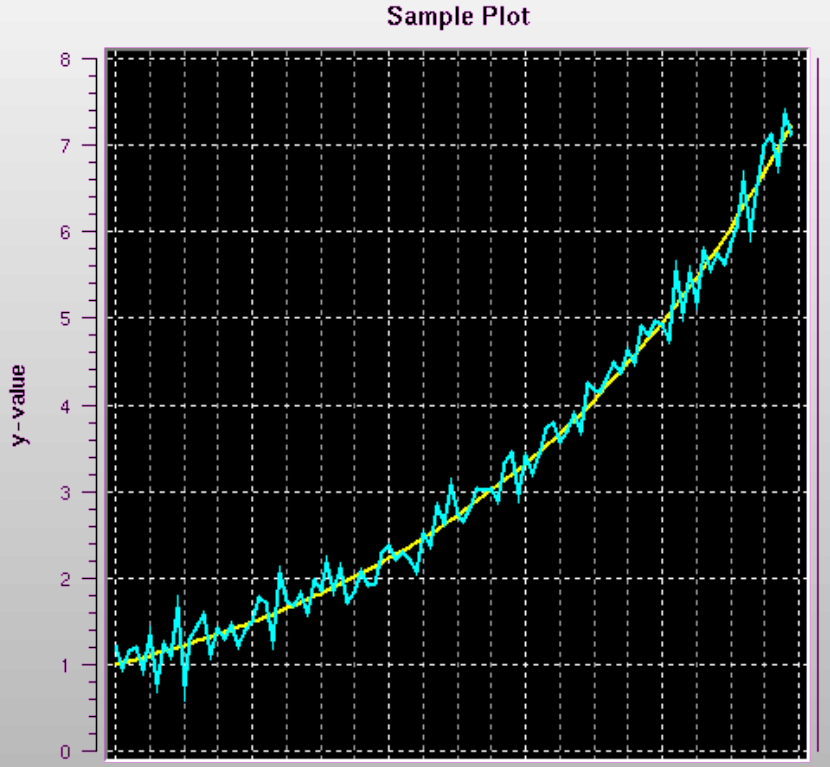
Summary:

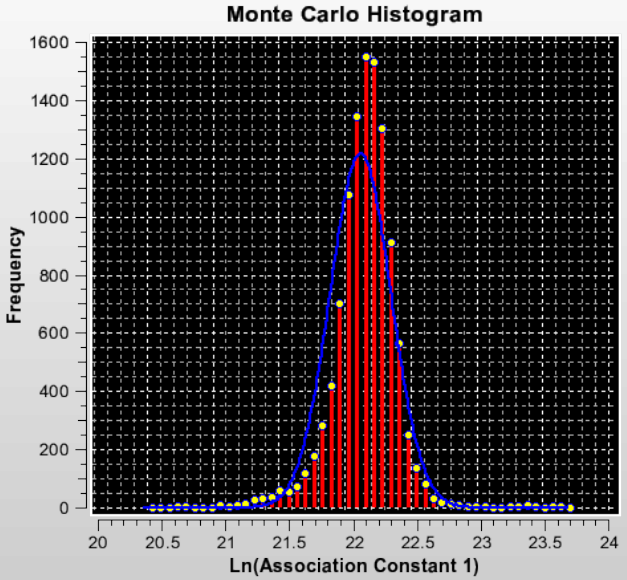
Remember:

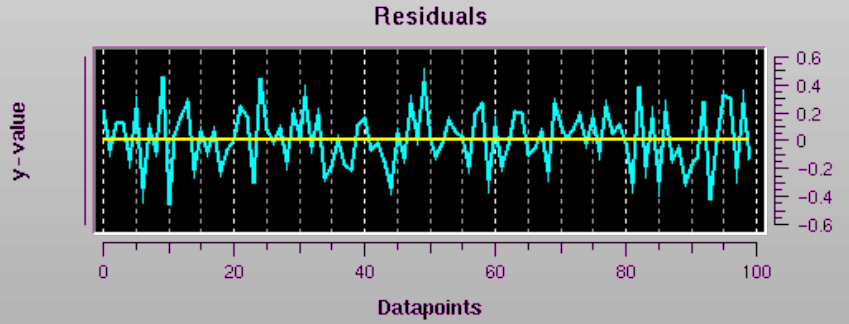
you cannot get reliable answers if you start with low quality input data!

Statistical Verification

Use Monte Carlo to determine the reliability of the fitted data:







Implementation of Monte Carlo Method

The Monte Carlo method is a stochastic approach that can be used to identify the effect noise has on the reliability of determined parameters. With the Monte Carlo approach the statistical confidence limits of each measured parameter can be determined.

Recipe for Monte Carlo:

Obtain a best-fit solution from model function fit and confirm that the residuals are random and without systematic deviation

Generate new synthetic Gaussian noise with the same quality as was observed in the original experiment and add it to the best-fit solution

Re-fit the solution

Repeat (2-4) at least 100 times and collect all parameter values

Calculate statistics from Monte Carlo distribution for each parameter

Implementation of Monte Carlo Method

Generation of synthetic noise:

Method 1 – use Bootstrapping:

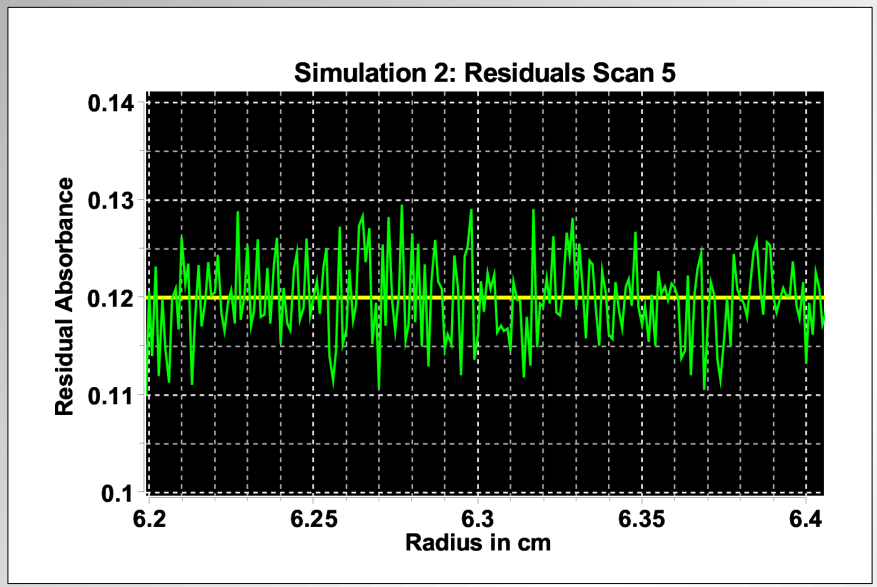
Permute a percentage of residuals. Take any residual, positive or negative, and place it elsewhere in the data. Assumption: The likelihood of a residual's magnitude is the same anywhere in the dataset.

Method 2 – generate Gaussian noise:

Run a 5-7 point Gaussian kernel over the residual vector from the best fit and use the Box-Müller algorithm to generate random new noise with the same quality at that position based on the variance obtained from the 5-point kernel.

Implementation of Monte Carlo Method

Generation of synthetic Gaussian noise with Method 2:



Take the absolute value of the residual values from 5-7 points, and smooth them with a Gaussian kernel (most weight on the center point). The average residual value is fed into the Box – Müller function to generate a new random residual that has a Gaussian probability distribution with a mean of the average residual value.

If there is a lot of noise in some area of the scan, the new data will have noise locally equivalent to the original data.

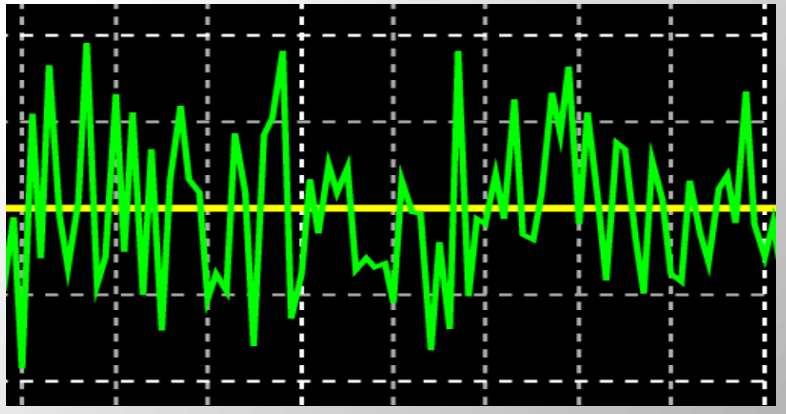
Generation of Gaussian noise is preferred because the quality of the noise varies at different points in the cell and it also varies with absorbance (more absorbance = more noise).

Implementation of Monte Carlo Method

Box – Müller function:

```
float box_muller(float m, float s)        /* normal random variate generator */
{                                         /* mean m, standard deviation s    */
     float x1, x2, w = 2.0;
     while (w >= 1.0)
     {
          x1 = 2.0 * ranf() - 1.0;
          x2 = 2.0 * ranf() - 1.0;
          w = x1 * x1 + x2 * x2;
     }
     w = sqrt((-2.0 * log(w)) / w);
     return( m + x1 * w * s );
}

float ranf()
{
     int N = 1;
     float temp = 0.0;
     temp = (((float) rand() / ((float) RAND_MAX + 1) * N));
     return(temp);
}
```

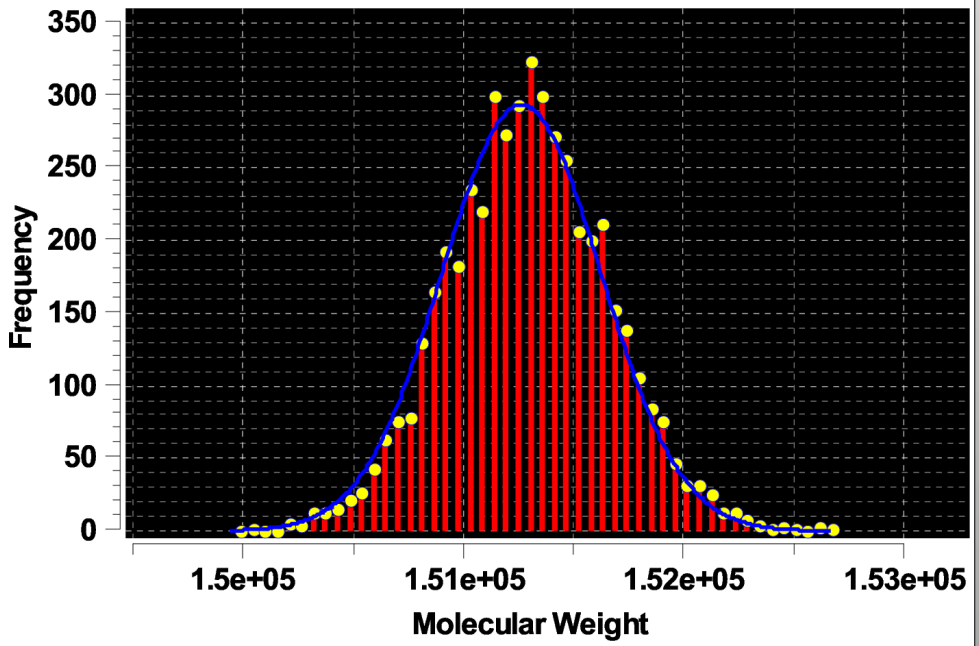


Obtaining Confidence Intervals

Once a Monte Carlo analysis is completed, a frequency distribution of parameter values for each parameter is obtained, and statistics for the distribution can be calculated:

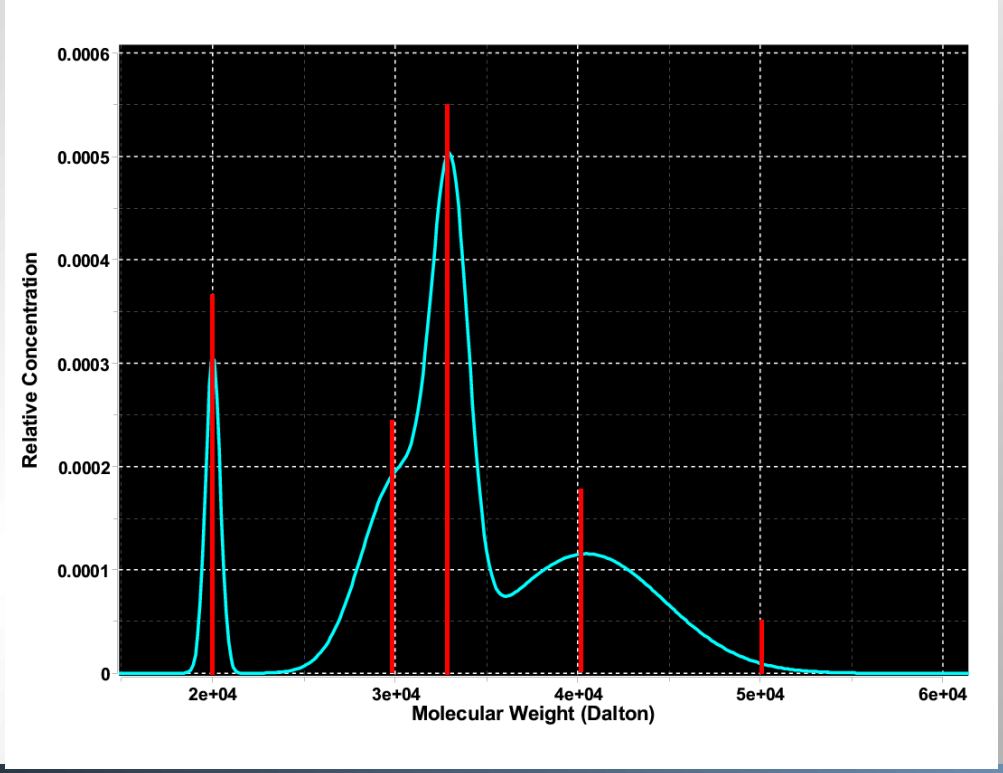
```
Molecular Weight

Maximum Value:                1.526728e+05
Minimum Value:                1.499355e+05
Mean Value:                   1.512681e+05
Median Value:                 1.513042e+05
Skew Value:                   2.345094e-02
Kurtosis Value:               4.796727e-02
Lower Mode Limit:             1.513042e+05
Upper Mode Limit:             1.513589e+05
Mode Center:                  1.513315e+05
95% Confidence Limits:        +6.445781e+02
                              -7.715156e+02
99% Confidence Limits:        +8.671094e+02
                              -9.940469e+02
Standard Deviation:           3.612517e+02
Standard Error:               3.572899e+00
Variance:                     1.305028e+05
Correlation Coefficient:      1.387418e-02
Gaussian Area:                5.592846e+05
95 % Confidence Interval:     1.505600e+05 (low),  1.519761e+05
(high)
99 % Confidence Interval:     1.503375e+05 (low),  1.521986e+05
(high)
```

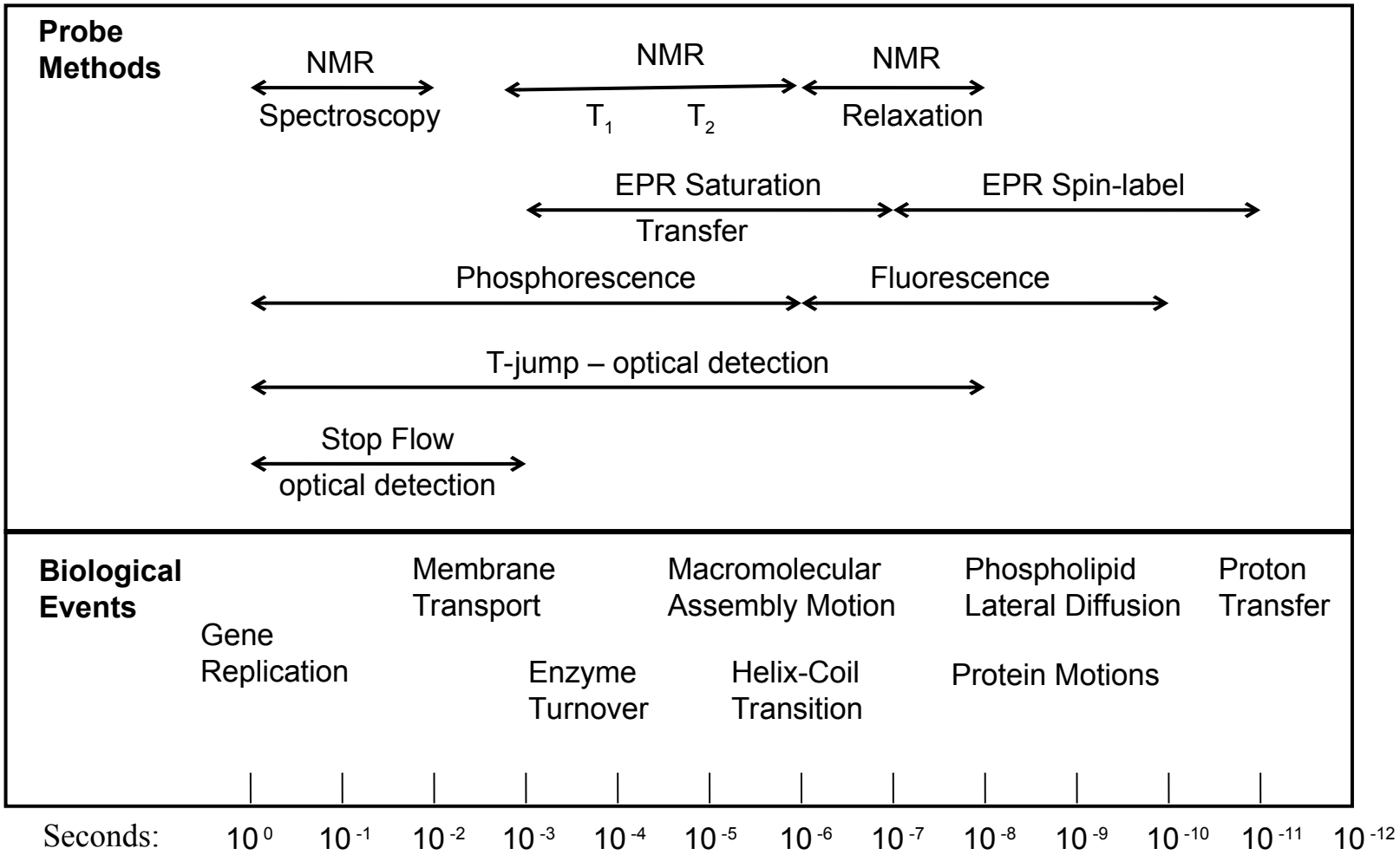


Regularization:

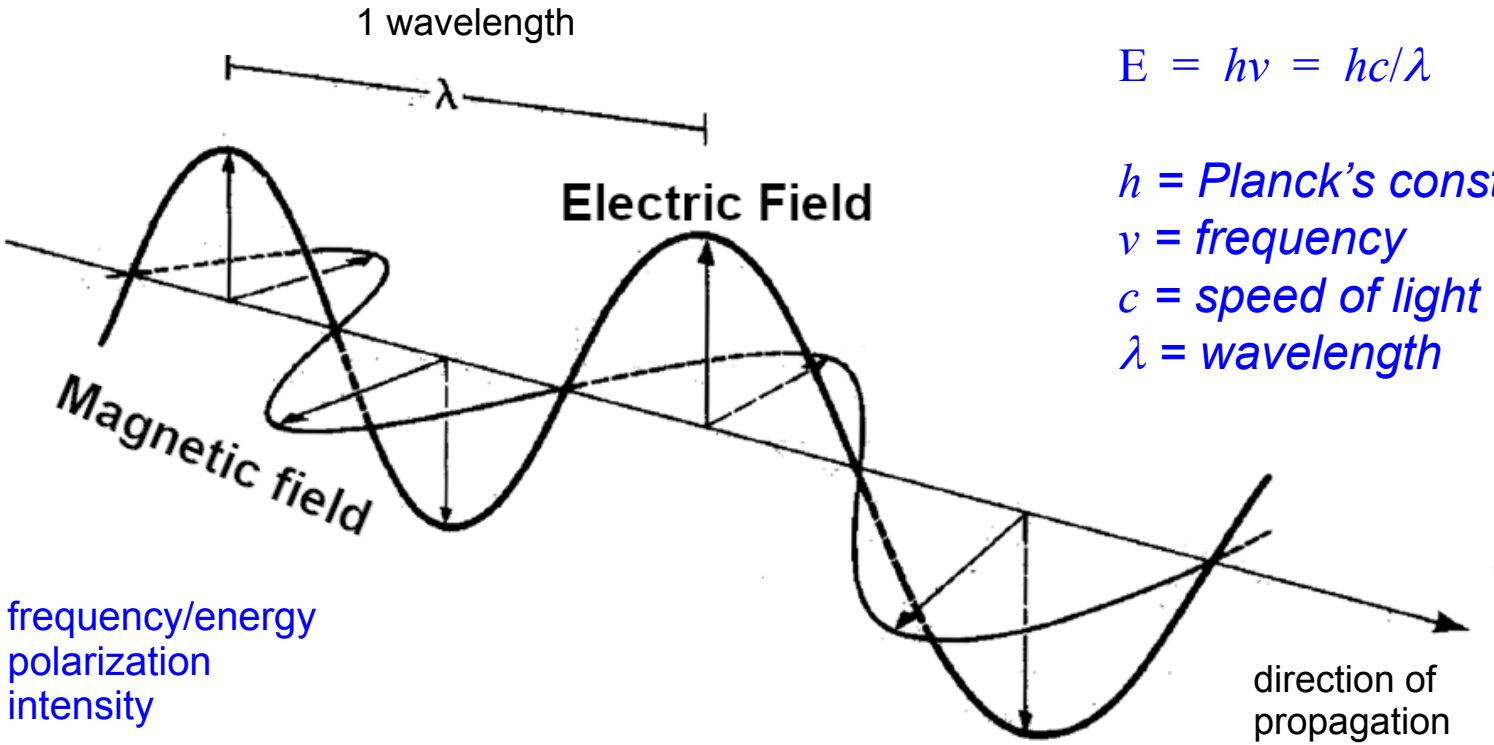
Occam's Razor: The simplest solution describing the data is the preferred solution. Instead of regularizing the solution and introducing infinitely many solutes with different probability, we want to *REDUCE* the solution space to find the solution that has the smallest number of possible solutes. For probabilities of solutes inspect the GA evolution profile which provides MC statistics.



Spectroscopy



Electromagnetic Radiation



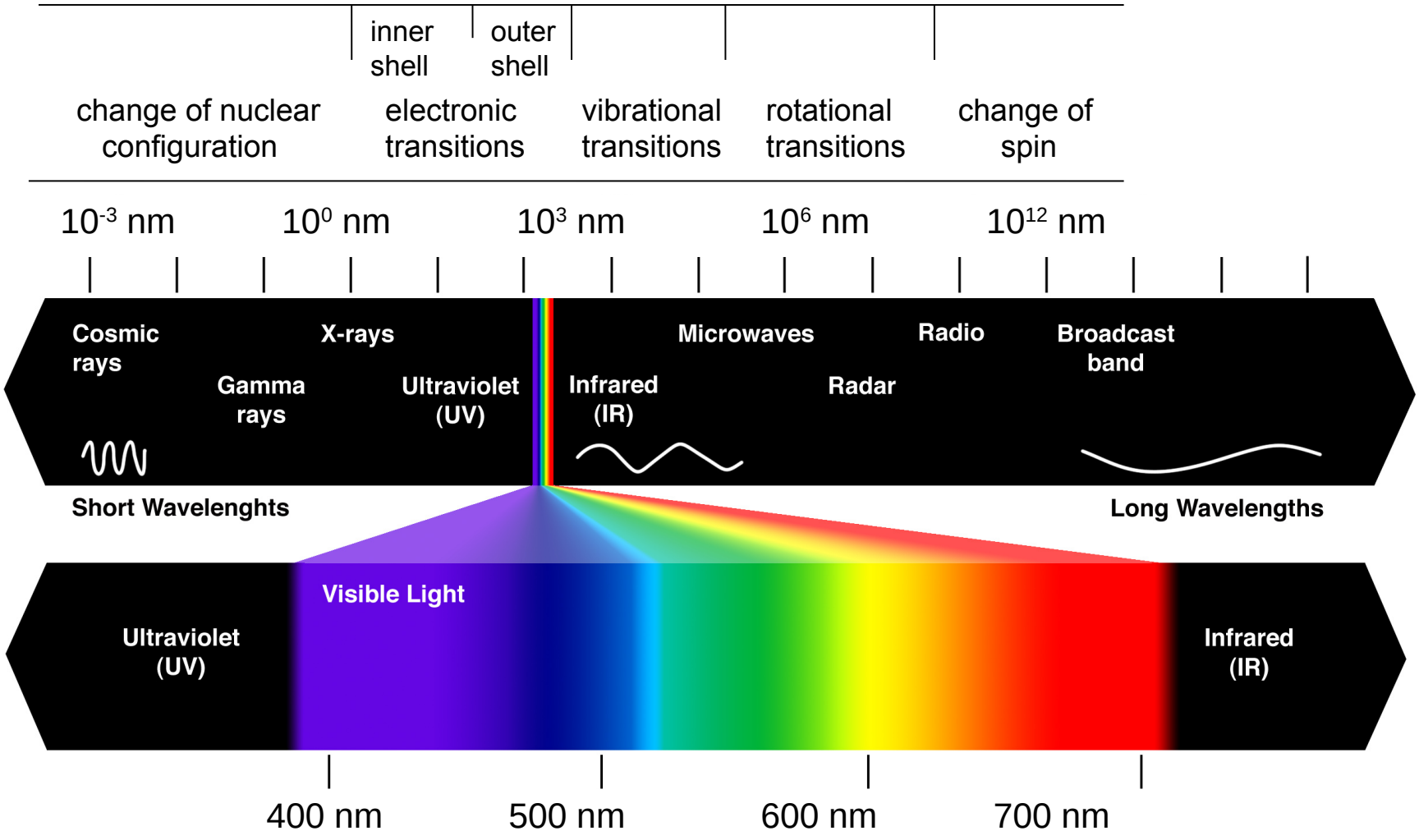
Planck's Law:

$$E = h\nu = hc/\lambda$$

h = Planck's constant
ν = frequency
c = speed of light
λ = wavelength

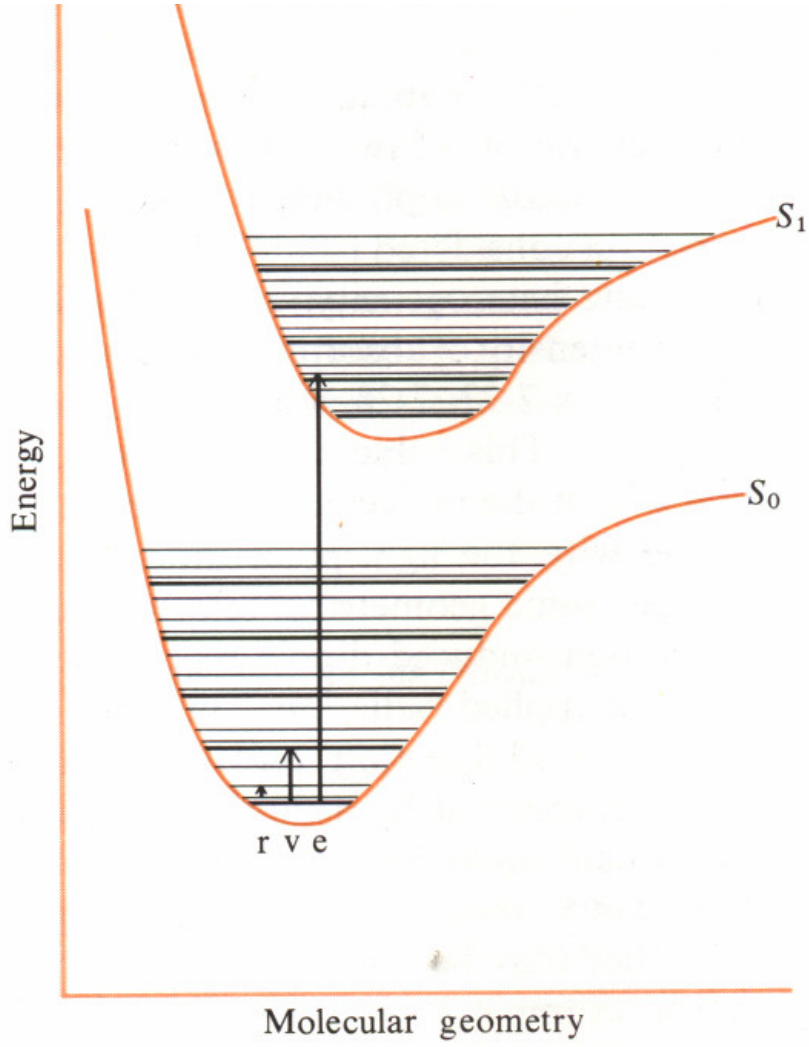
- frequency/energy
- polarization
- intensity

Spectroscopy



Spectroscopy

- Energy levels of a small molecule. Transitions occur between electronic states (e), vibrational (v) and rotational (r) levels. There can be many energy levels making absorbance bands broad.
- Energy spacing between electronic states is around 80 kcal/mol → higher than room temperature, only possible by absorbing light
- Energy spacing between vibrational leves is about 10 kcal/mole, higher than thermal energies, so only the ground states are normally populated
- Energy spacing between rotational levels is about 1 kcal/mol, sufficiently small for multiple levels to be populated at room temperature.
- The energy spacing between vibrational and electronic states is large enough that molecules are mostly in their ground states at room temperature.

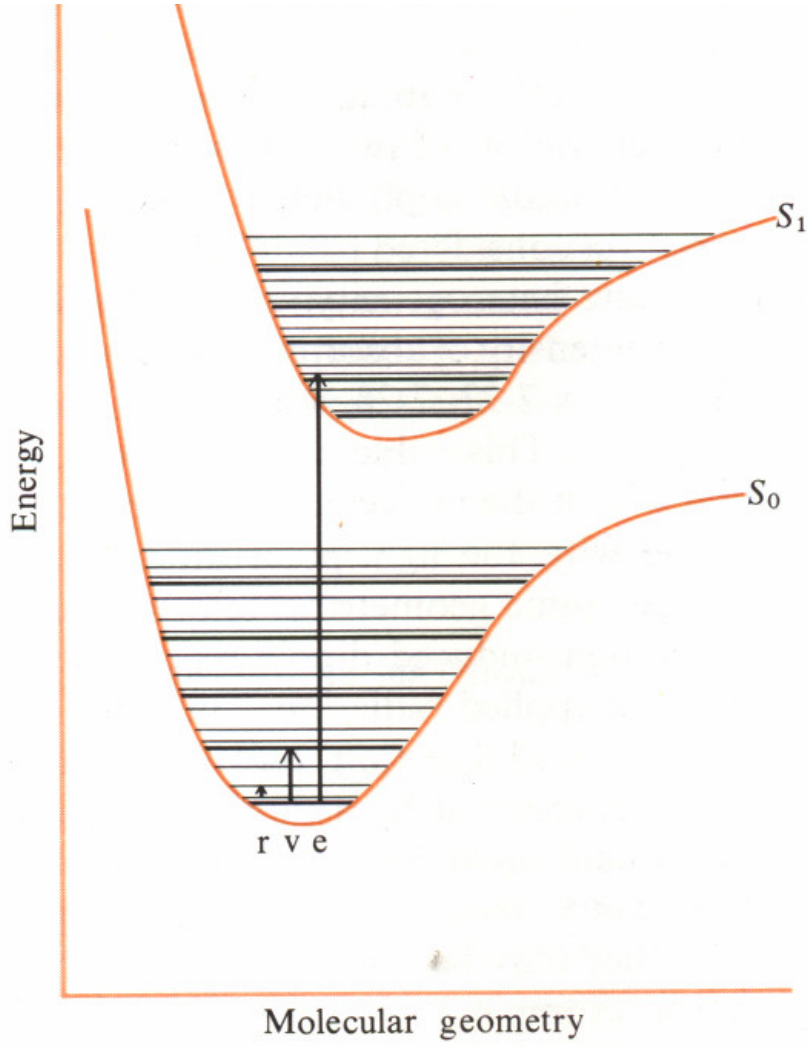


Spectroscopy

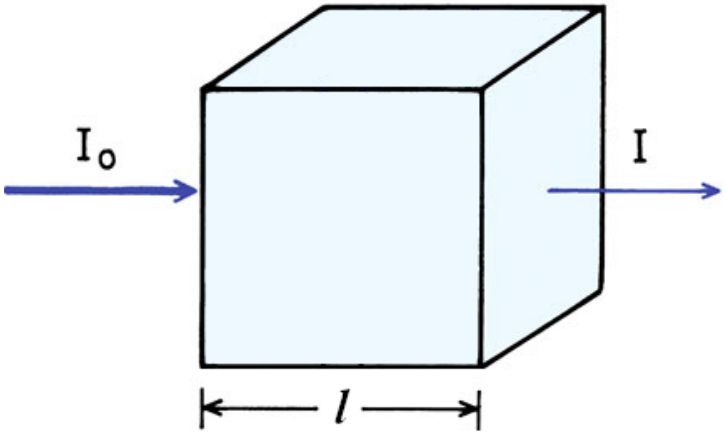
- The Franck-Condon principle states that electronic transitions that involve different vibrational levels have the highest probability of occurring when the vibrational levels overlap in the momentum and nuclear positions, i.e., no displacement of nuclei.

Energies required for various transitions:

- Electronic transitions: UV/visible light
- Vibrational transitions: Infrared light
- Rotational transitions: microwave radiation, not suitable for absorbance spectroscopy, they will be considered with nuclear magnetic resonance.



Beer-Lambert Law (of absorbance)



Intensity: $I_A = I_0 - I$

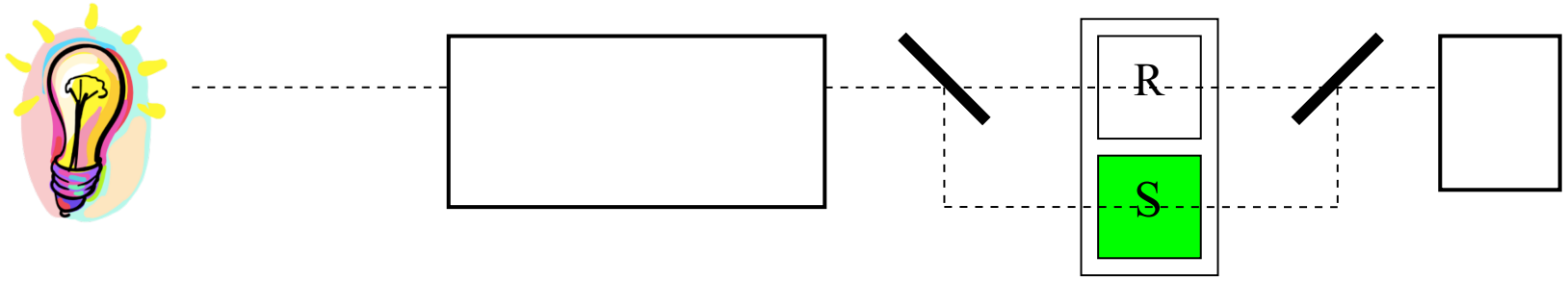
Transmittance: $T = I/I_0$

Absorbance $A = - \log T$

Incoming intensity, I_0, outgoing intensity, I, and absorbed intensity, I_A, in a cell of length l. Consider the light intensity absorbed (I_A) in a volume of thickness l and cross-sectional area *A*. Each molecule, randomly oriented in solution, contributes to the observed absorption of light based on its cross-sectional area, and the probability of absorbing a photon of the energy from the incident light over a given time period.

Absorption

Measured by: single or double beam spectrophotometer



light source (xenon or deuterium)

monochromator

cuvette holder

detector

Absorption Spectroscopy

BEER LAMBERT LAW:

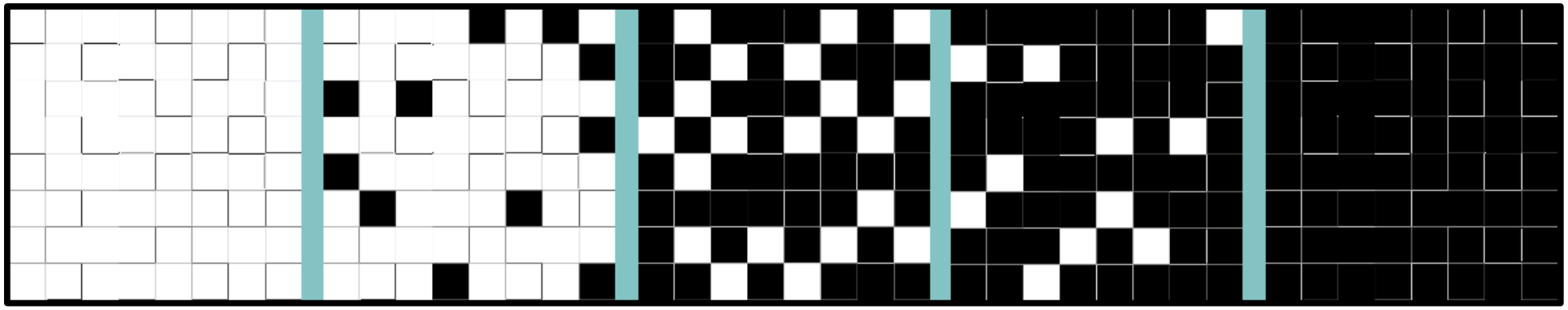
$$\frac{dI}{dl} = I C \epsilon' \qquad \ln\left(\frac{I_0}{I}\right) = C \epsilon' l$$

$$A(\lambda) = log_{10}\left(\frac{I_0}{I}\right) = C \epsilon(\lambda) l$$

A **is the absorbance or optical density (OD) of the sample at a particular wavelength. This measure can be used to determine the concentration of sample**

The extinction coefficient ε is a function of wavelength and independent of concentration, where ε = ε'/2.303. ε is typically expressed in OD/mol for a 1 cm pathlength.

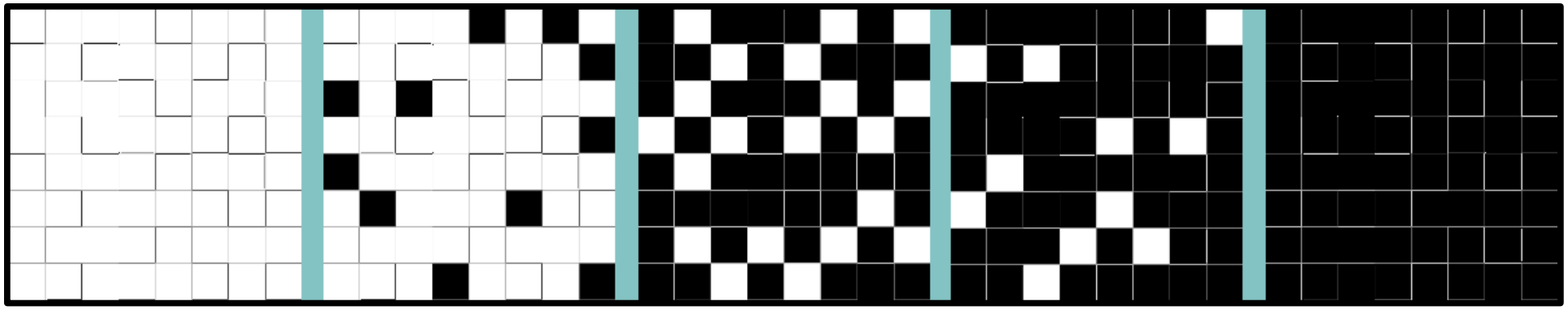
Spectroscopy

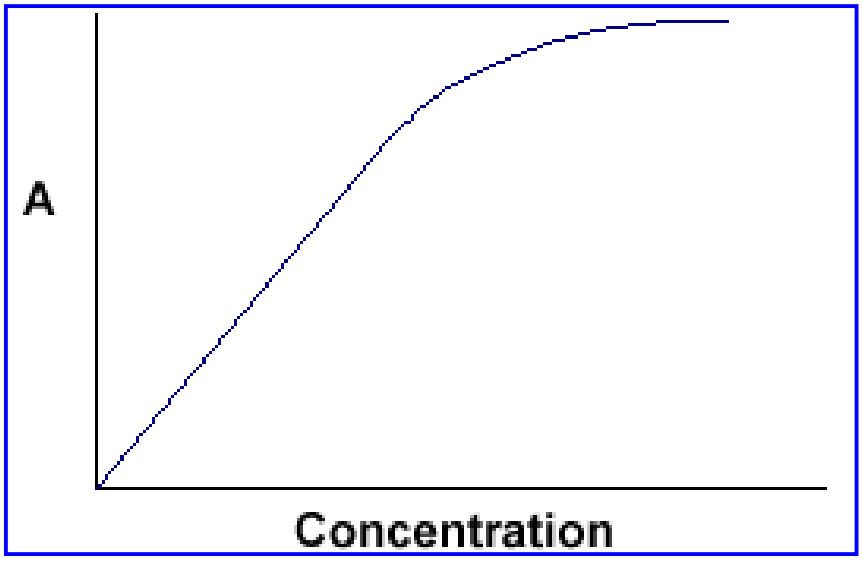


The absorbance is proportional to the number of the molecules (concentration C) and the “darkness” (molar absorptivity ε) of the molecules. The number of the molecules interfering with the light beam increases with the thickness of the volume element l

$$\mathrm{A} = -\log\left(I/I_0\right) = \varepsilon\, C\, l = \mathrm{OD}$$

C is the concentration, [M]; ε is the molar extinction coefficient, [M^{-1} cm^{-1}]; l is the pathlength (in cm), and OD is the optical density, also called absorbance.



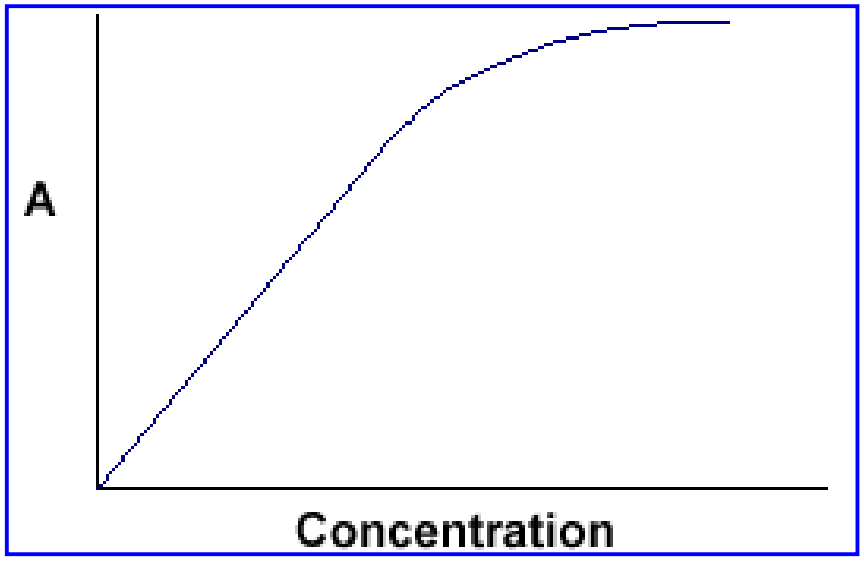


Relationship between Absorbance and Concentration

Why is this plot not linear for the entire range?

At some point the absorbance is so high that not sufficient light passes through to the detector, and linearity is no longer satisfied.

Relationship between Absorbance and Concentration



The absorbance at which UV-visible spectrophotometer becomes non-linear depends on the following factors:

1. Concentration of the analyte
2. Lamp intensity at the measured wavelength
3. Extinction coefficient of the analyte at measured wavelength
4. Sensitivity of detector at the measured wavelength

To measure C accurately, always measure around 0.2-0.8 OD, which corresponds to a transmission of ~63% - 16% T

Deviations from the Beer-Lambert Law – watch out for these:

- **Exceeding the dynamic range of the detector**
- **Chromatic shifts due to macromolecular interaction**
- **Chromatic shifts due to solvent variation**
- **Chromatic shifts due to conformational changes in analyte**
- **Improper blanking of the reference absorbance**
- **Dirty cuvette in the reference beam**
- **Dilutions with an absorbing buffer different from the solvent**
- **Analyte is contaminated with an absorbing molecule**