UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

Basic Error Analysis

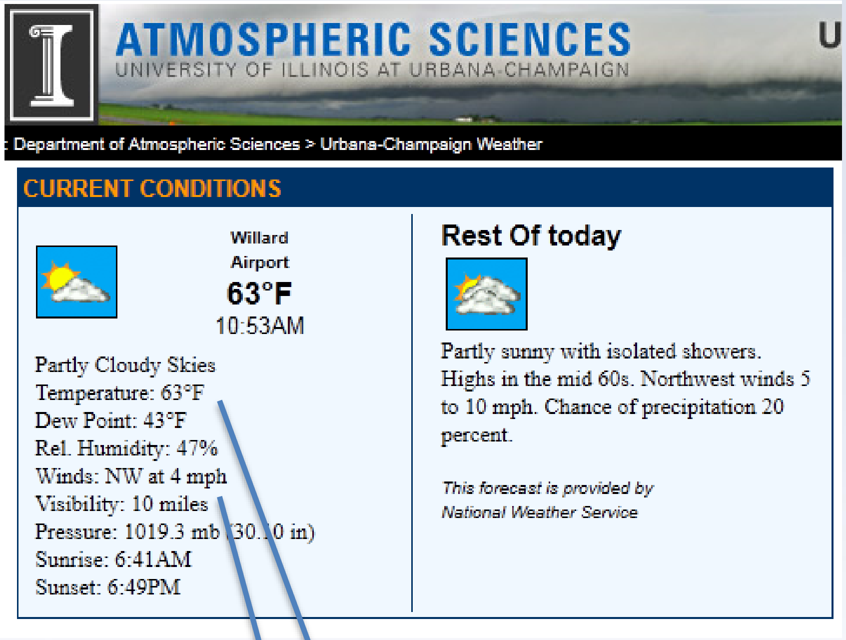
Physics 401
Fall 2016
Eugene V Colla

illinois.edu

Agenda

- Errors and uncertainties
- The Reading Error
- Accuracy and precession
- Systematic and statistical errors
- Fitting errors
- Appendix. Working with oil drop data
    - Nonlinear fitting

What and when we need to know about errors. Everyday life.



ATMOSPHERIC SCIENCES
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Department of Atmospheric Sciences > Urbana-Champaign Weather

CURRENT CONDITIONS

Willard
Airport
63°F
10:53AM

Partly Cloudy Skies
Temperature: 63°F
Dew Point: 43°F
Rel. Humidity: 47%
Winds: NW at 4 mph
Visibility: 10 miles
Pressure: 1019.3 mb (30.10 in)
Sunrise: 6:41AM
Sunset: 6:49PM

Rest Of today

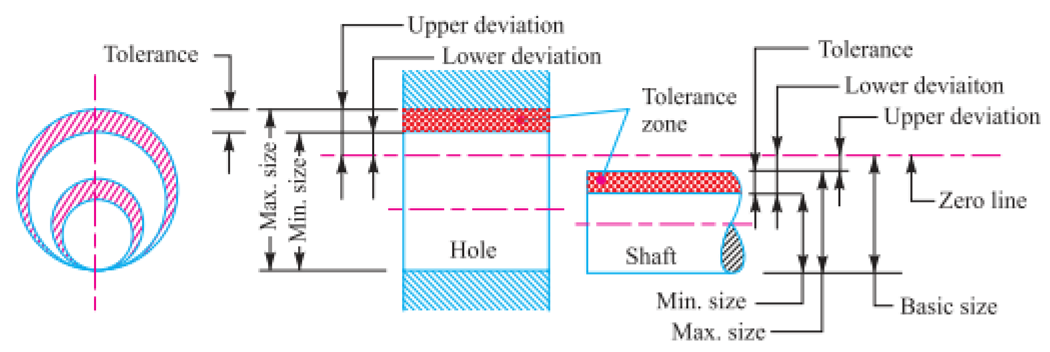
Partly sunny with isolated showers. Highs in the mid 60s. Northwest winds 5 to 10 mph. Chance of precipitation 20 percent.

This forecast is provided by National Weather Service

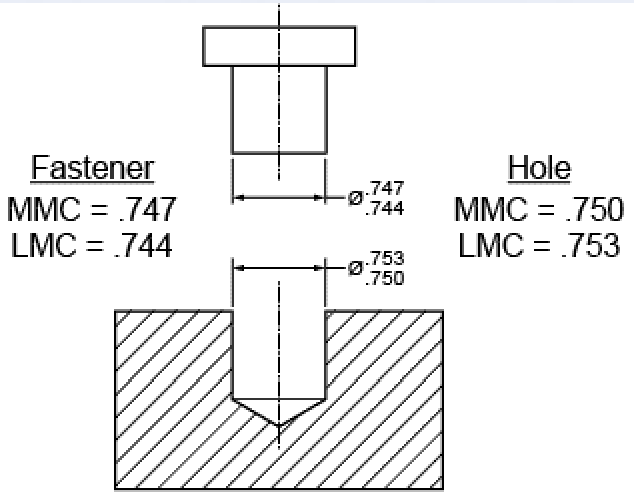
$T = 63°\text{F} \pm ?$ → Best guess $\Delta T \sim 0.5$°F

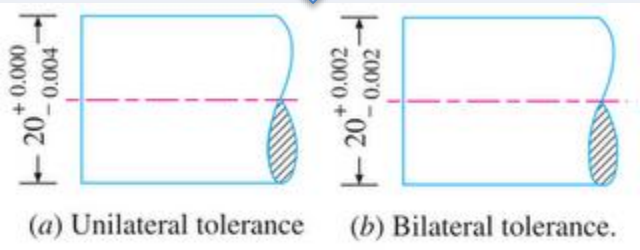
Wind speed 4mph$\pm ?$ → Best guess $\pm 0.5 mph$

What and when we need to know about errors. Industry.



Clearance fit





(*a*) Unilateral tolerance (*b*) Bilateral tolerance.

What and when we need to know about errors. Science.

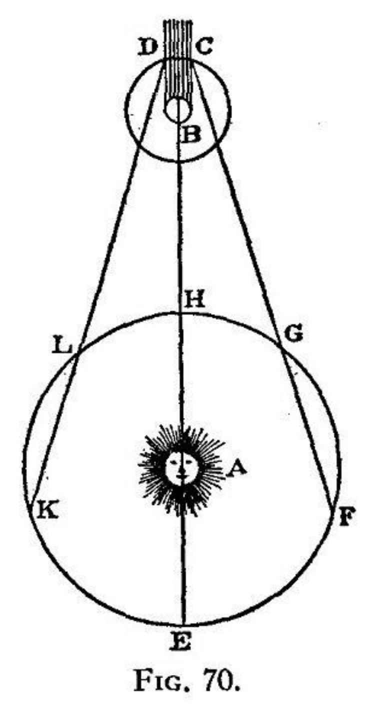


FIG. 70.

Measurement of the speed of the light

1675 Ole Roemer: 220,000 Km/sec

Ole Christensen Rømer
1644-1710

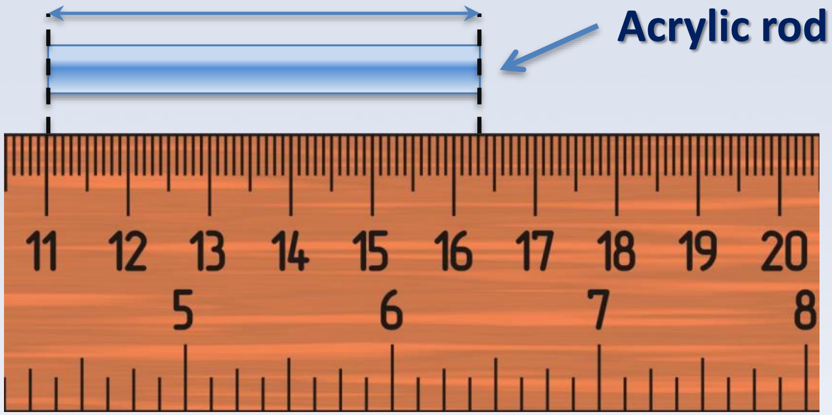
Does it make sense?
What is missing?

NIST Bolder Colorado c = 299,792,456.2±1.1 m/s.

Reading error

$L = 53mm \pm \Delta L(?)$

$\Delta L \cong 0.5mm$



Acrylic rod

$\Delta L \cong 0.03mm$

How far we have to go in reducing the reading error?

We do not care about accuracy better than 1mm

If ruler is not okay, we need to use digital caliper

Probably the natural limit of accuracy can be due to length uncertainty because of temperature expansion. For 53mm $\Delta L \cong 0.012mm/K$

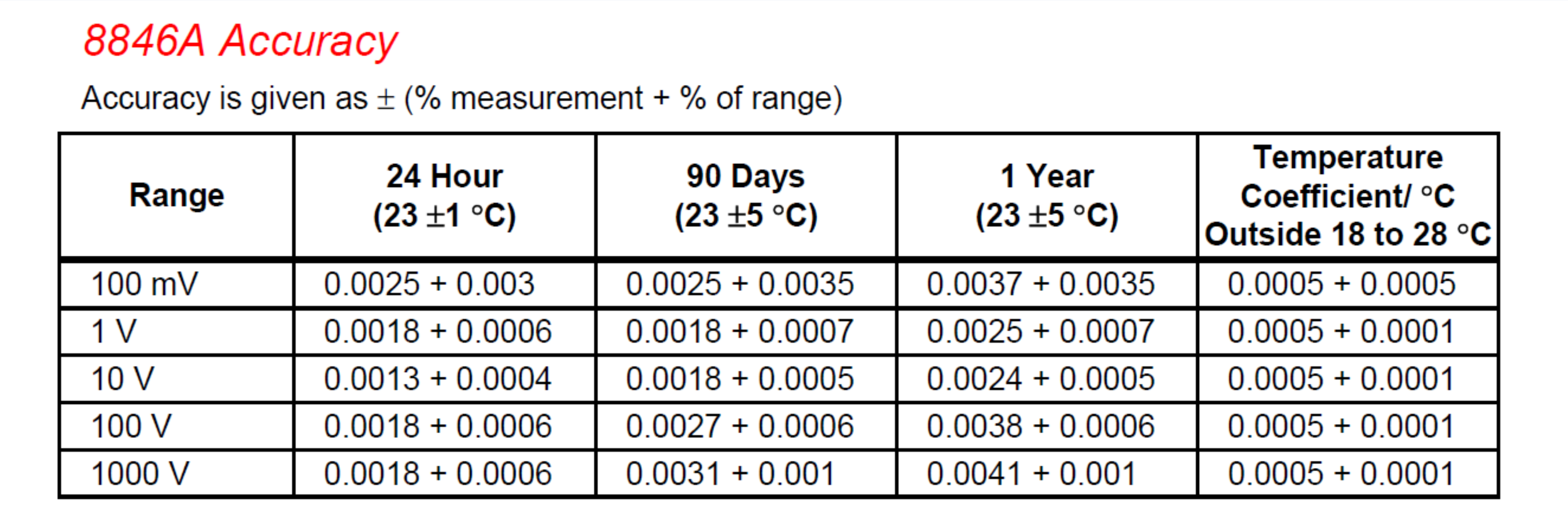
Reading Error = $\pm \frac{1}{2}$ (least count or minimum gradation).

Reading error. Digital meters.

Fluke 8845A multimeter

Example Vdc (reading)=0.85V

$$\Delta V = 0.83 \times \left(1.8 \times 10^{-5}\right)$$
$$+1.0 \times \left(0.7 \times 10^{-5}\right) \cong 2.2 \times 10^{-5}$$
$$= 22\mu V$$



8846A Accuracy

Accuracy is given as ± (% measurement + % of range)

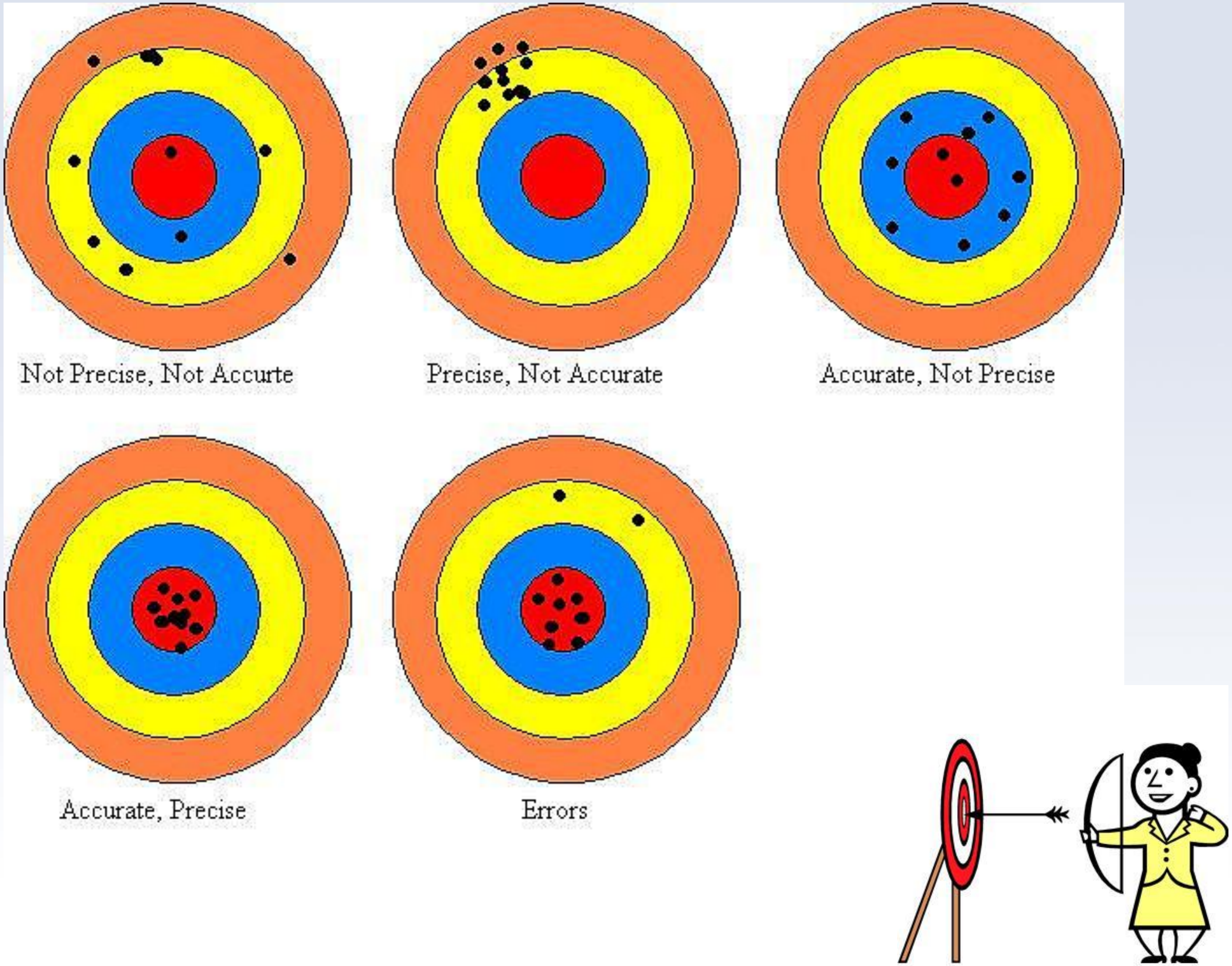
Range	24 Hour (23 ±1 °C)	90 Days (23 ±5 °C)	1 Year (23 ±5 °C)	Temperature Coefficient/ °C Outside 18 to 28 °C
100 mV	0.0025 + 0.003	0.0025 + 0.0035	0.0037 + 0.0035	0.0005 + 0.0005
1 V	0.0018 + 0.0006	0.0018 + 0.0007	0.0025 + 0.0007	0.0005 + 0.0001
10 V	0.0013 + 0.0004	0.0018 + 0.0005	0.0024 + 0.0005	0.0005 + 0.0001
100 V	0.0018 + 0.0006	0.0027 + 0.0006	0.0038 + 0.0006	0.0005 + 0.0001
1000 V	0.0018 + 0.0006	0.0031 + 0.001	0.0041 + 0.001	0.0005 + 0.0001

Accuracy and precession

The accuracy of an experiment is a measure of how close the result of the experiment comes to the true value

Precision refers to how closely individual measurements agree with each other

Accuracy and precession



Not Precise, Not Accurte

Precise, Not Accurate

Accurate, Not Precise

Accurate, Precise

Errors

Systematic and random errors

- ***Systematic Error:*** **reproducible inaccuracy introduced by faulty equipment, calibration or technique.**

- ***Random errors*****: Indefiniteness of results due to finite precision of experiment. Measure of fluctuation in result after repeatable experimentation.**

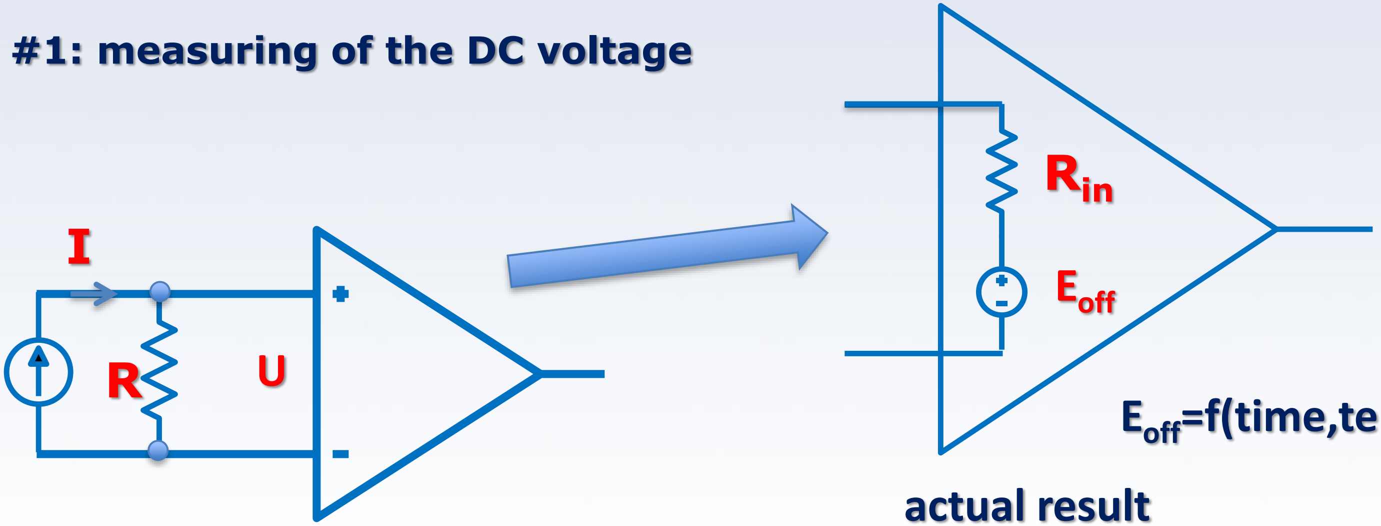
Philip R. Bevington "Data Reduction and Error Analysis for the Physical sciences", McGraw-Hill, 1969

Systematic errors

Sources of systematic errors: **poor calibration of the equipment, changes of environmental conditions, imperfect method of observation, drift and some offset in readings etc.**

Example #1: measuring of the DC voltage

Current source



E_{off}=f(time,temperature)

actual result

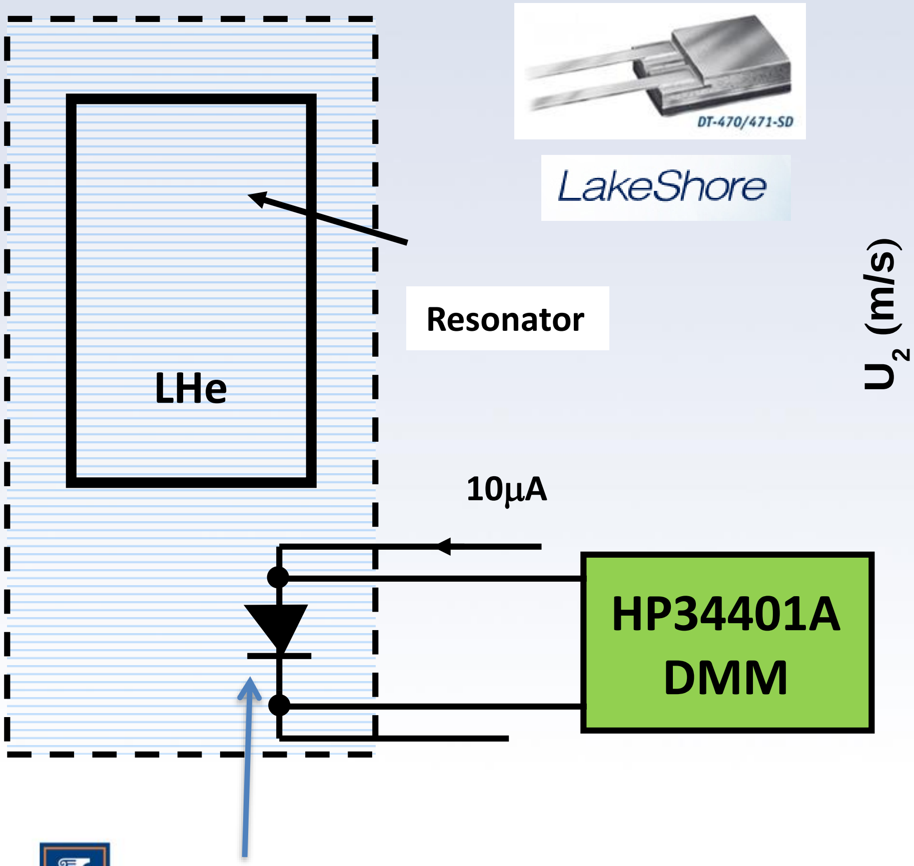
$$U = \frac{R * I - \left(\frac{R}{R_{in}}\right) E_{off}}{\left(1 + \frac{R}{R_{in}}\right)}$$

expectation

U=R*I

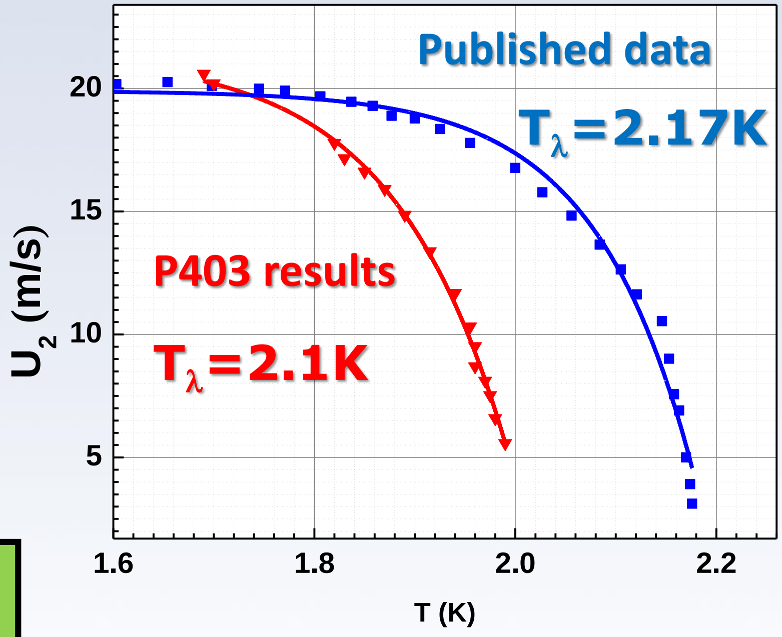
Systematic errors

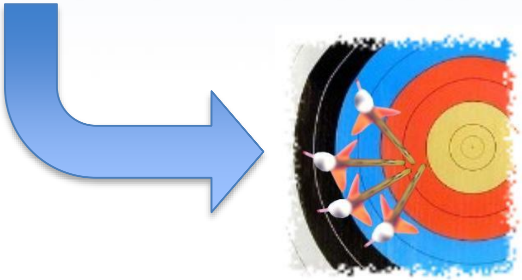
Example #3: poor calibration



Temperature sensor

Measuring of the speed of the second sound in superfluid He4





Random errors

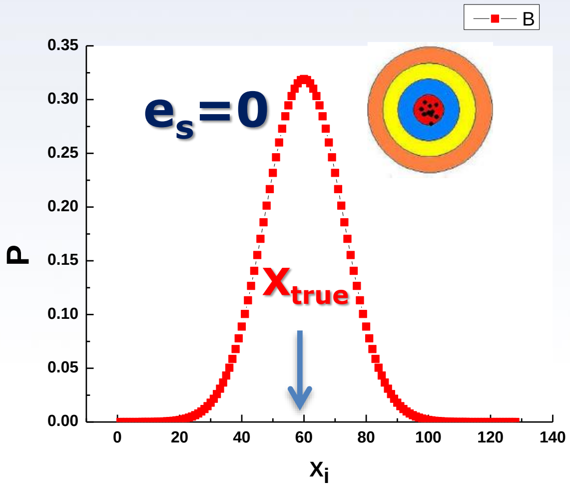
Systematic error

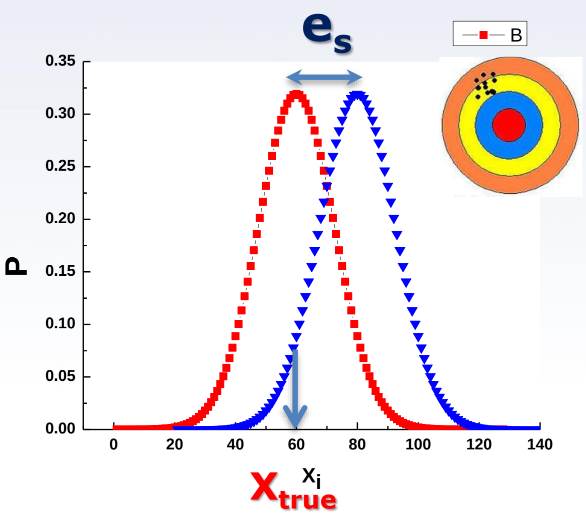
Result of measurement

$$X_{meas} = X_{true} + e_s + e_r$$

Random error

Correct value





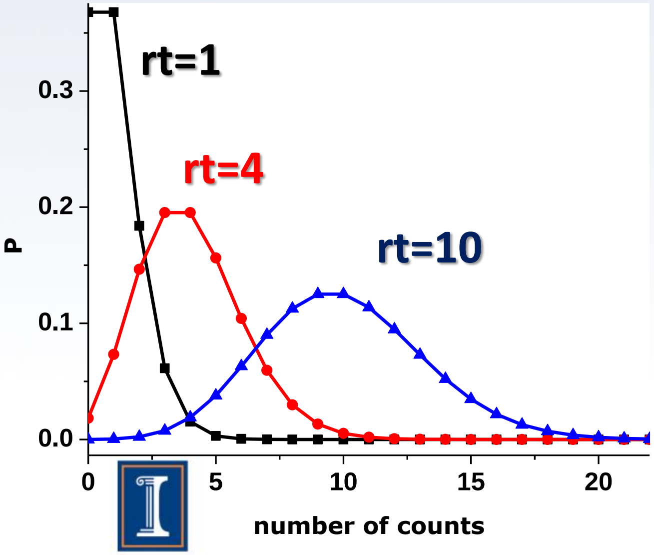
Random errors. Poisson distribution

Siméon Denis Poisson
(1781-1840)

$$P_n(t) = \frac{(rt)^n}{n!} e^{-rt} \quad n = 0,1,2,\dots$$

r: decay rate [counts/s] t: time interval [s]

➔ $P_n(rt)$: Probability to have n decays in time interval t



illinois.edu

A statistical process is described through a Poisson Distribution if:

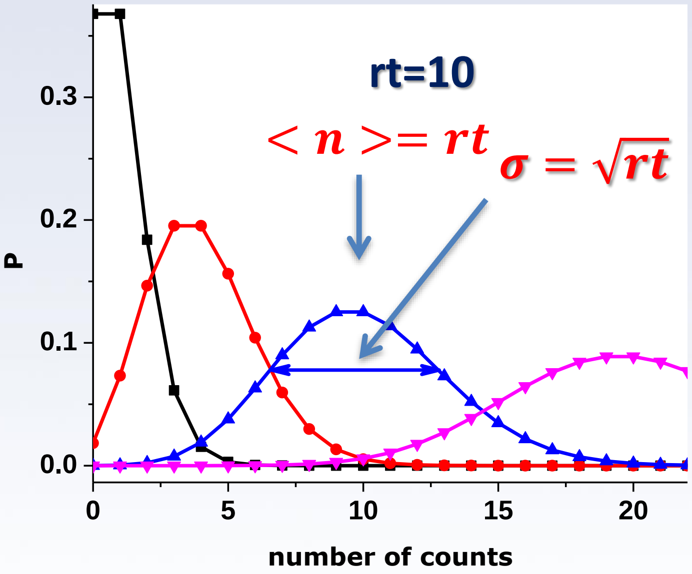
- ○ **random process** → for a given nucleus probability for a decay to occur is the same in each time interval.
- ○ **universal probability** → the probability to decay in a given time interval is same for all nuclei.
- ○ **no correlation between two instances** (the decay of on nucleus does not change the probability for a second nucleus to decay.

Poisson distribution

$$P_n(t) = \frac{(rt)^n}{n!} e^{-rt} \quad n = 0,1,2,\ldots$$

r: decay rate [counts/s] *t*: time interval [s]

➔ $P_n(rt)$: Probability to have n decays in time interval t



Properties of the Poisson distribution:

$$\sum_{n=0}^{\infty} P_n(rt) = 1$$, probabilities sum to 1

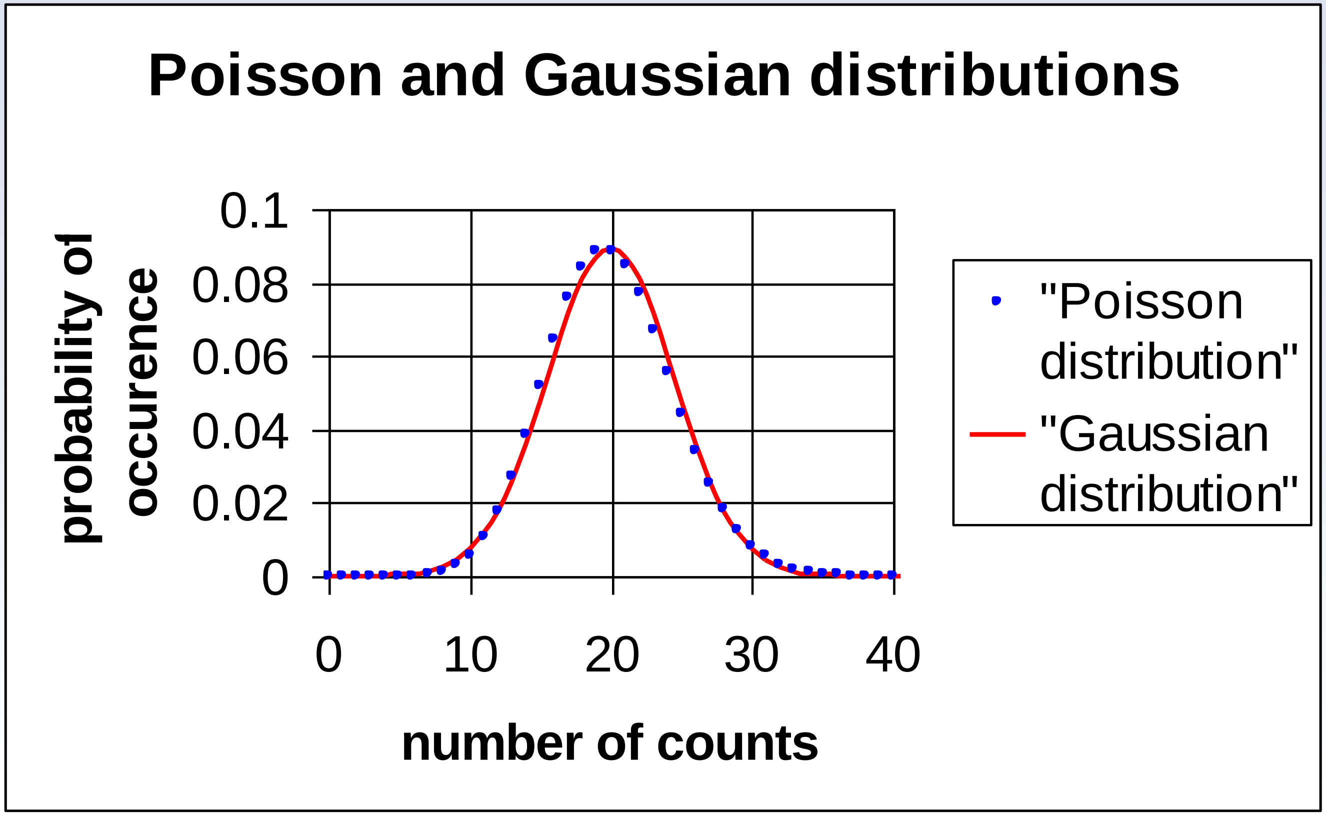
$$< n >= \sum_{n=0}^{\infty} n \cdot P_n(rt) = rt$$, the mean

$$\sigma = \sqrt{\sum_{n=0}^{\infty} (n - < n >)^2 P_n(rt)} = \sqrt{rt}$$,

standard deviation

Poisson distribution at large *rt*

$$P_n\left(t\right) = \frac{\left(rt\right)^n}{n!} e^{-rt} \quad n = 0,1,2,\ldots$$

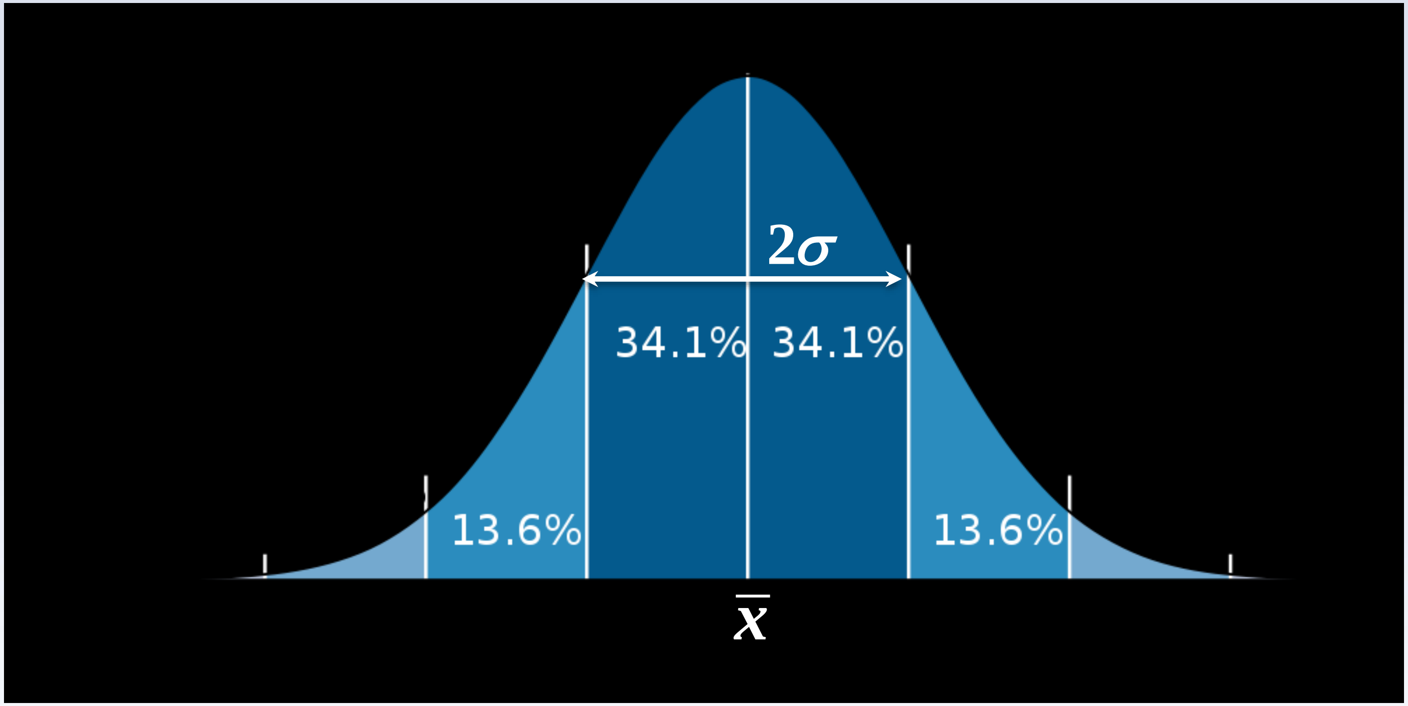


Carl Friedrich Gauss
(1777–1855)

$$\boxed{P_n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}}$$

Gaussian distribution: continuous

Normal (Gaussian) distribution



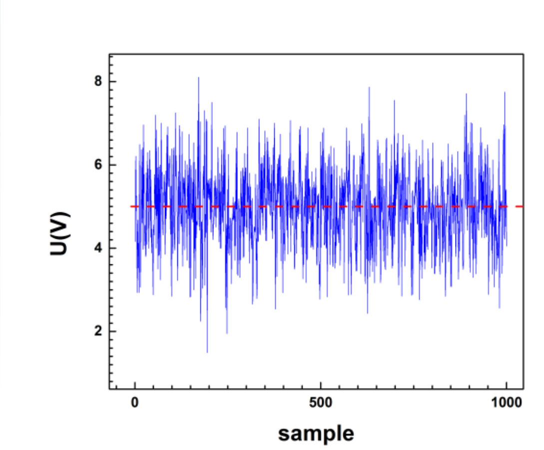
$$\boxed{P_n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}}$$

Error in the mean is given as $\frac{\sigma}{\sqrt{N}}$

illinois.edu

Measurement in presence of noise

Source of noisy signal

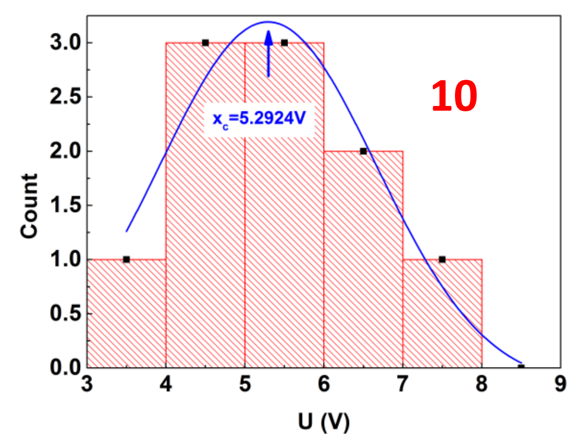


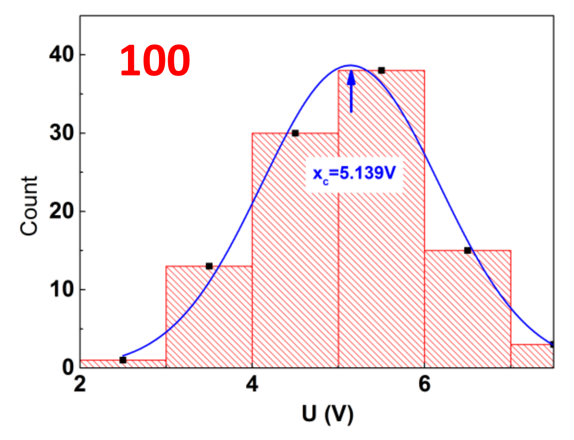
4.89855
5.25111
2.93382
4.31753
4.67903
3.52626
4.12001
2.93411

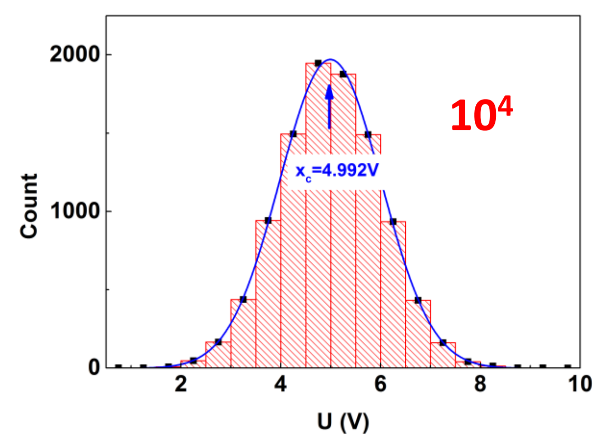
Expected value 5V

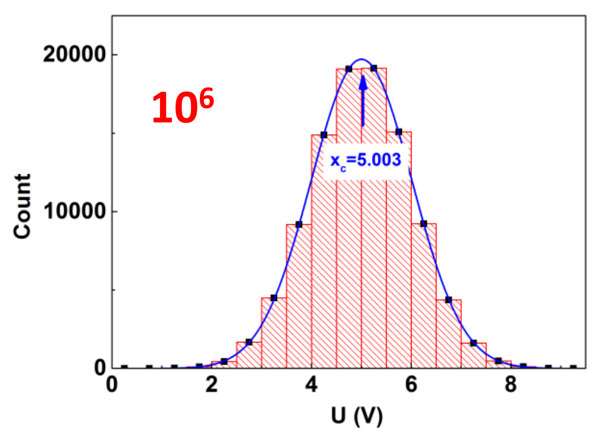
Actual measured values

Measurement in presence of noise

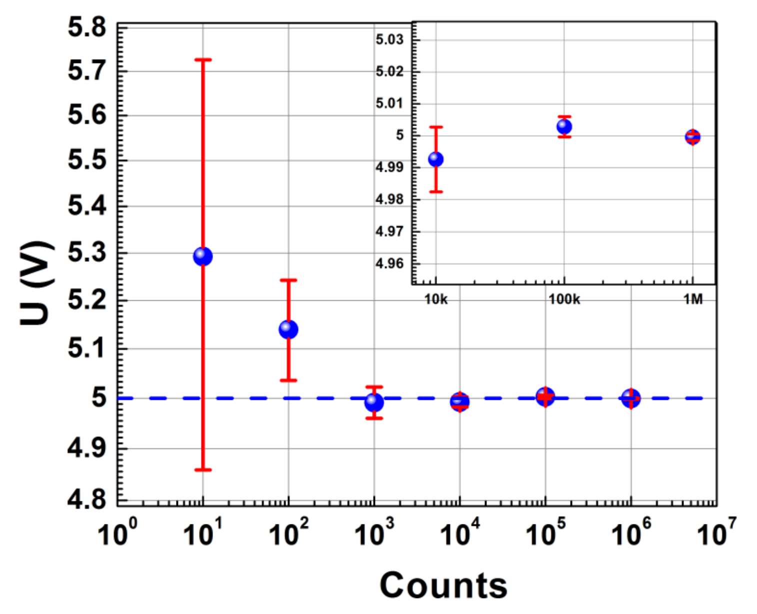


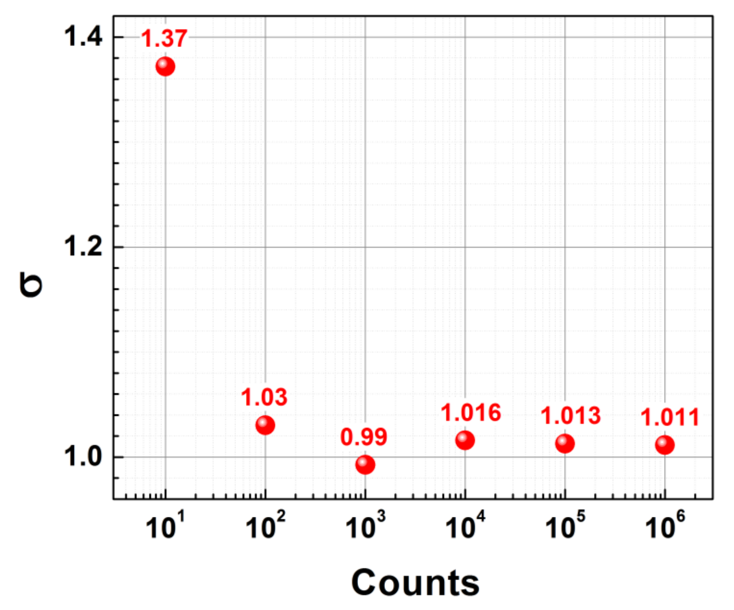






Measurement in presence of noise





Result →

$$U = x_c \pm \frac{\sigma}{\sqrt{N}}$$

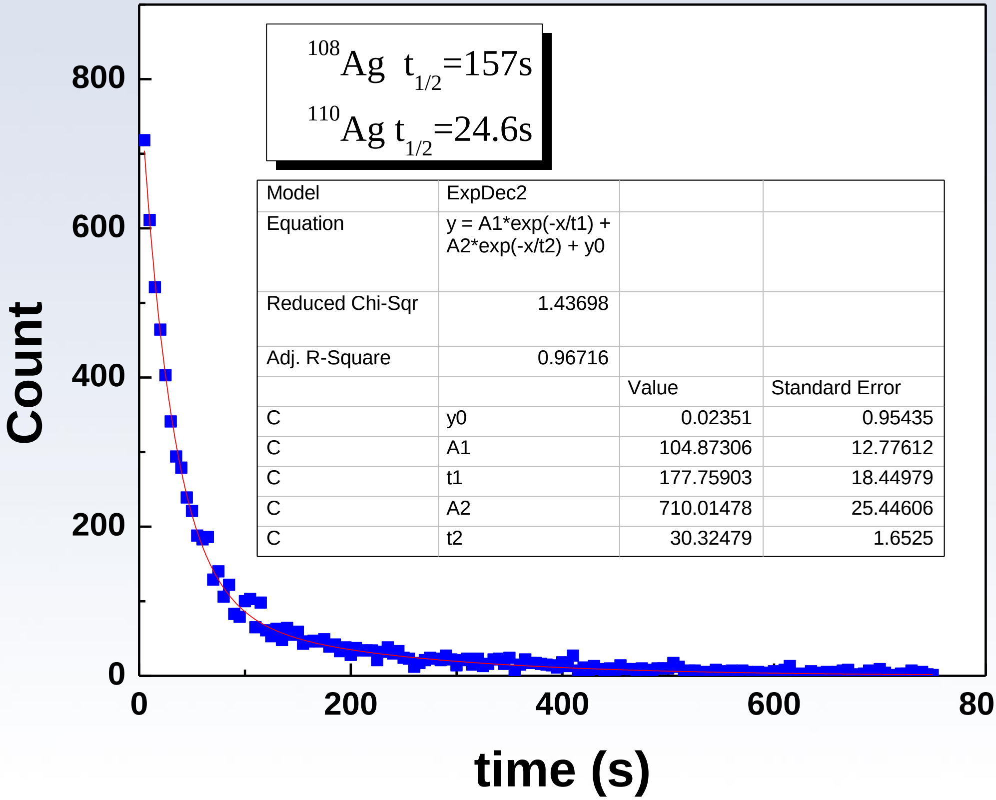
σ - standard deviation
N – number of samples

For N=10^6 U=4.999±0.001 0.02% accuracy

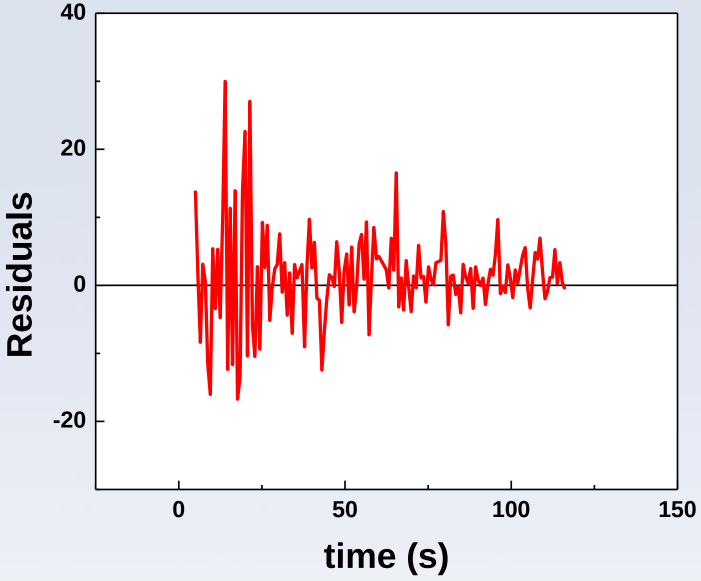
illinois.edu

Fitting errors

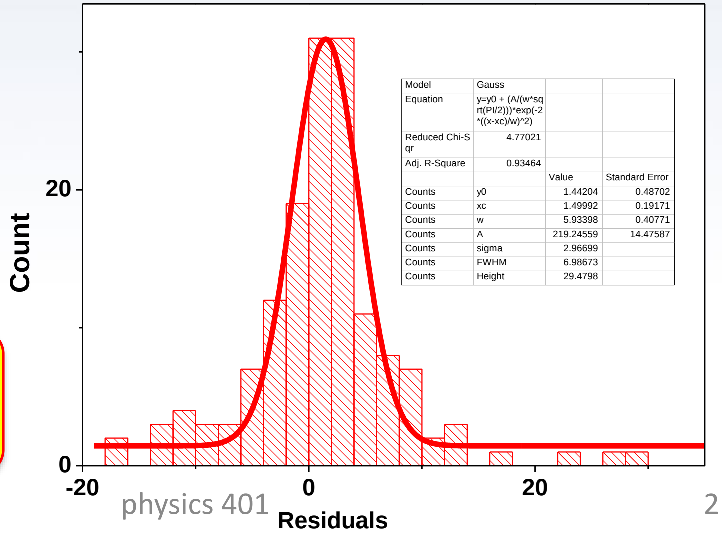
Ag β decay



Model	ExpDec2		
Equation	y = A1*exp(-x/t1) + A2*exp(-x/t2) + y0		
Reduced Chi-Sqr	1.43698		
Adj. R-Square	0.96716		
		Value	Standard Error
C	y0	0.02351	0.95435
C	A1	104.87306	12.77612
C	t1	177.75903	18.44979
C	A2	710.01478	25.44606
C	t2	30.32479	1.6525



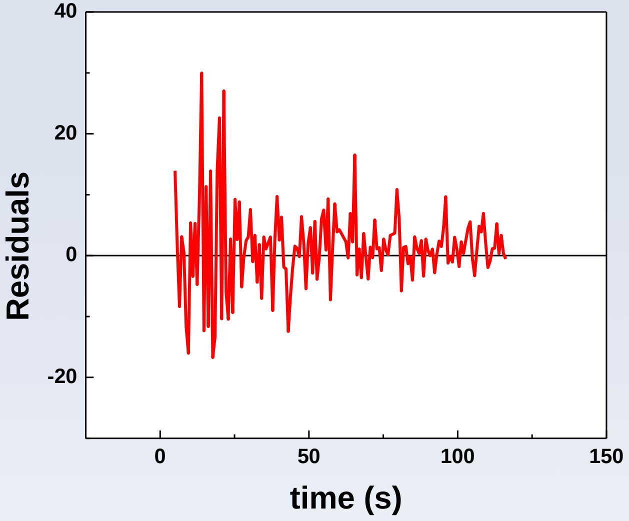
$$y = A1 \bullet \exp\left(\frac{-t}{t_1}\right) + A2 \bullet \exp\left(\frac{-t}{t_2}\right) + y_0$$

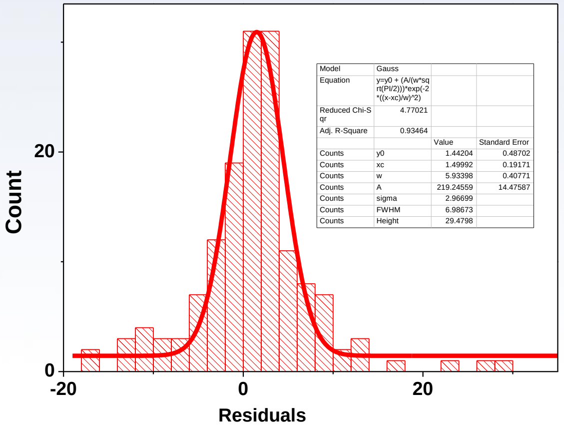


Model	Gauss		
Equation	y=y0 + (A/(w*sqrt(PI/2)))*exp(-2*((x-xc)/w)^2)		
Reduced Chi-Sqr	4.77021		
Adj. R-Square	0.93464		
		Value	Standard Error
Counts	y0	1.44204	0.48702
Counts	xc	1.49992	0.19171
Counts	w	5.93398	0.40771
Counts	A	219.24559	14.47587
Counts	sigma	2.96699	
Counts	FWHM	6.98673	
Counts	Height	29.4798	

Fitting. Analysis of the residuals

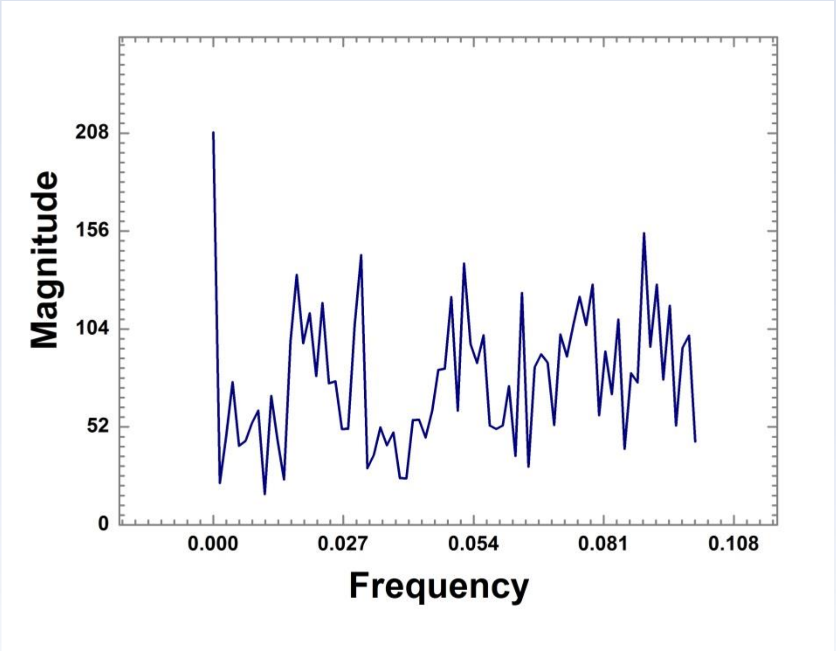
Ag β decay





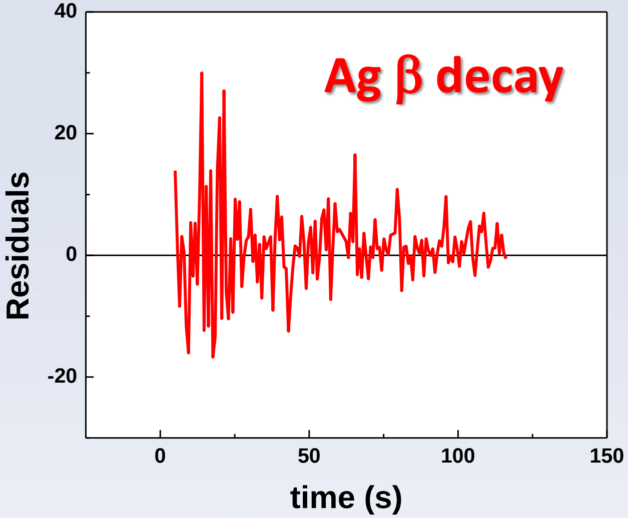
Model	Gauss		
Equation	y=y0 + (A/(w*sqrt(PI/2)))*exp(-2*((x-xc)/w)^2)		
Reduced Chi-Sqr	4.77021		
Adj. R-Square	0.93464		
		Value	Standard Error
Counts	y0	1.44204	0.48702
Counts	xc	1.49992	0.19171
Counts	w	5.93398	0.40771
Counts	A	219.24559	14.47587
Counts	sigma	2.96699	
Counts	FWHM	6.98673	
Counts	Height	29.4798	

Test 1. Fourier analysis

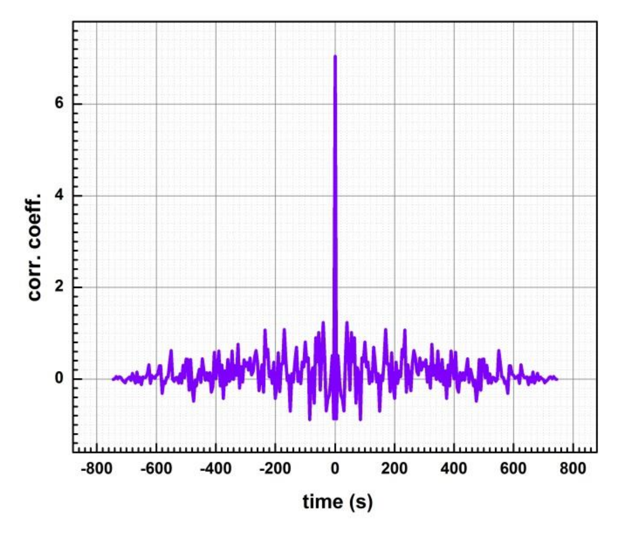


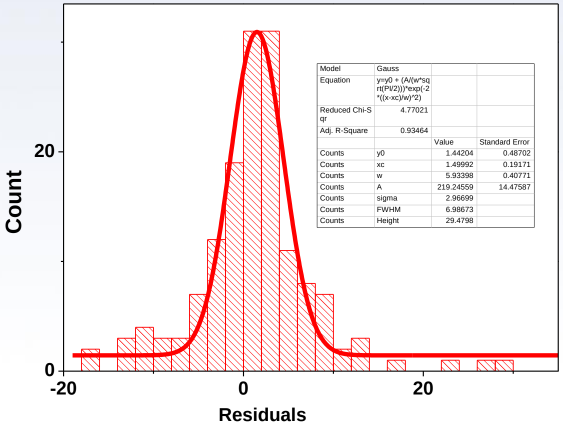
No pronounced frequencies found

Fitting. Analysis of the residuals



Test 1. Autocorrelation function





Model	Gauss		
Equation	y=y0 + (A/(w*sqrt(PI/2)))*exp(-2*((x-xc)/w)^2)		
Reduced Chi-Sqr	4.77021		
Adj. R-Square	0.93464		
		Value	Standard Error
Counts	y0	1.44204	0.48702
Counts	xc	1.49992	0.19171
Counts	w	5.93398	0.40771
Counts	A	219.24559	14.47587
Counts	sigma	2.96699	
Counts	FWHM	6.98673	
Counts	Height	29.4798	

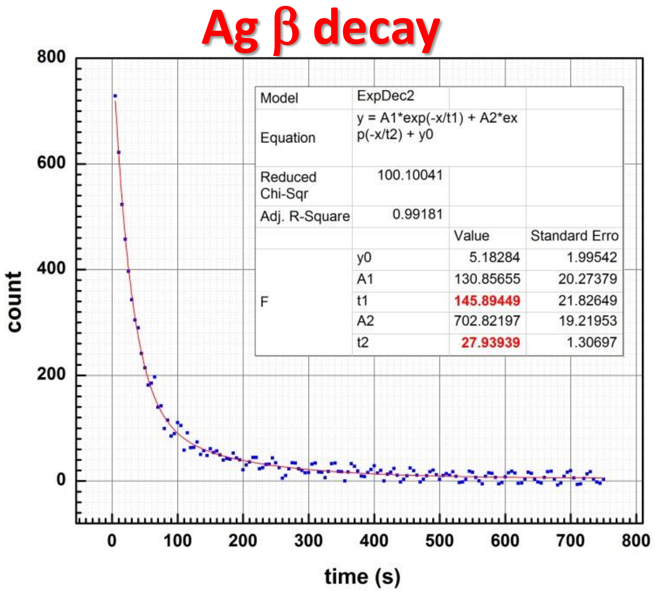
Correlation function

$$y(m)=\sum_{n=0} f(n)g(n-m)$$

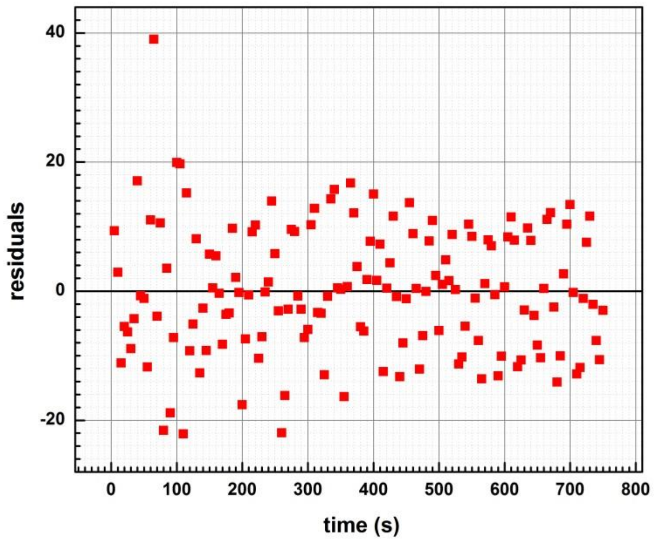
autocorrelation function

$$y(m)=\sum_{n=0}^{M-1} f(n)f(n-m)$$

Fitting. Analysis of the residuals. Non "ideal" case

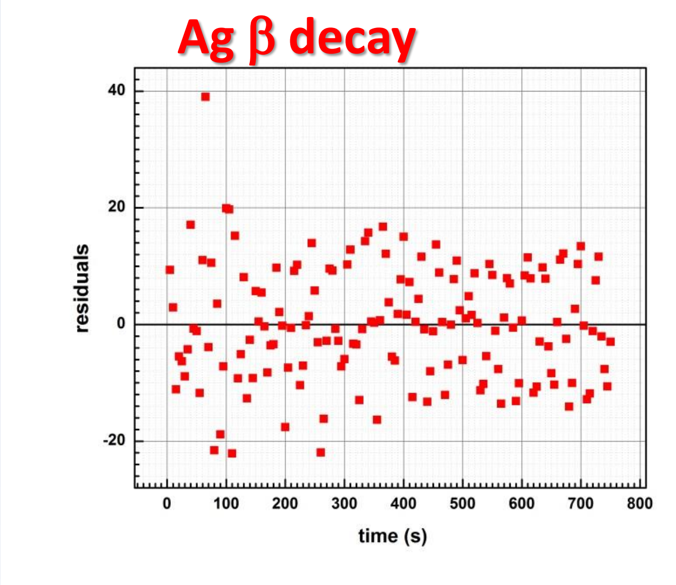


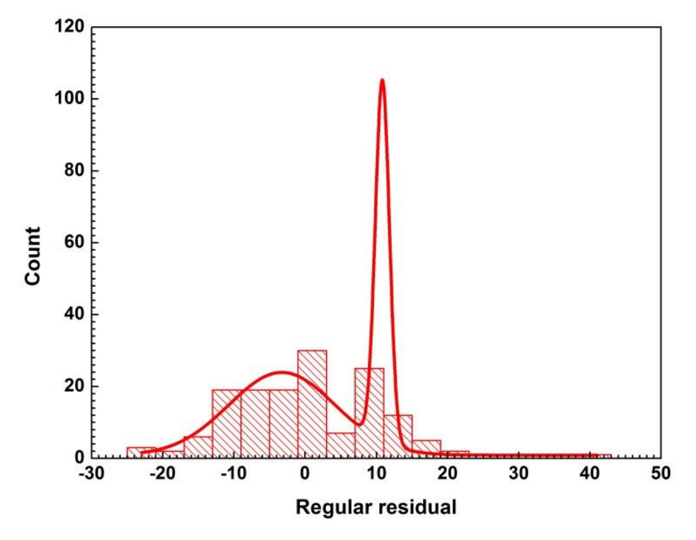
Model	ExpDec2		
Equation	y = A1*exp(-x/t1) + A2*exp(-x/t2) + y0		
Reduced Chi-Sqr	100.10041		
Adj. R-Square	0.99181		
		Value	Standard Erro
F	y0	5.18284	1.99542
	A1	130.85655	20.27379
	t1	**145.89449**	21.82649
	A2	702.82197	19.21953
	t2	**27.93939**	1.30697

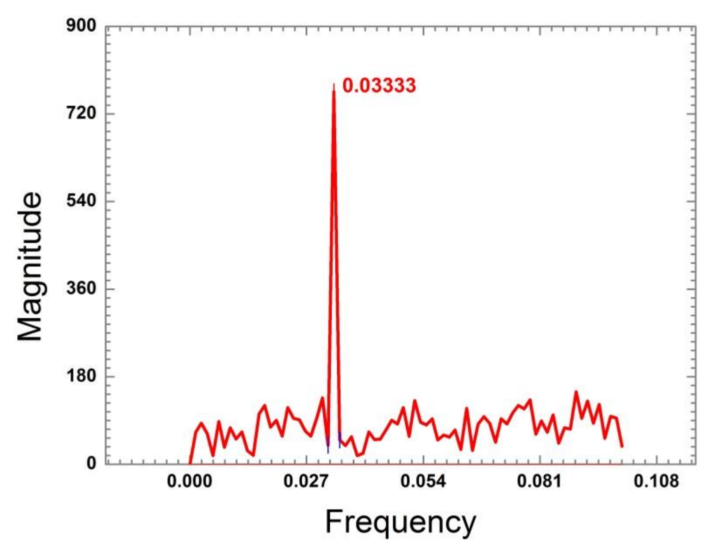


	Clear experiment	Data + "noise"
t_1(s)	177.76	145.89
t_2(s)	30.32	27.94

Fitting. Analysis of the residuals. Non "ideal" case

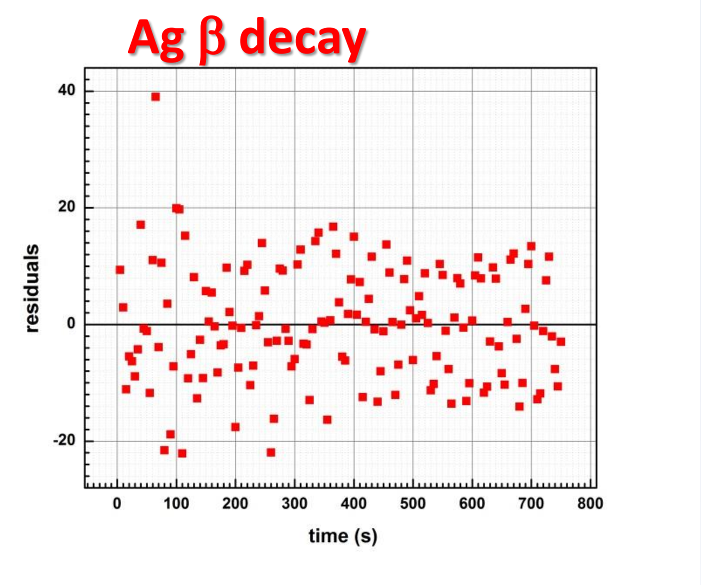


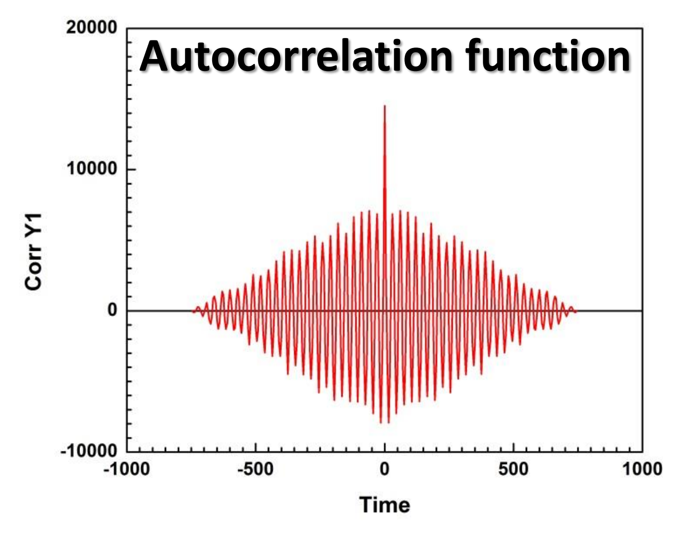




Histogram does not follow the normal distribution and there is frequency of 0.333 is present in spectrum

Fitting. Analysis of the residuals. Non "ideal" case

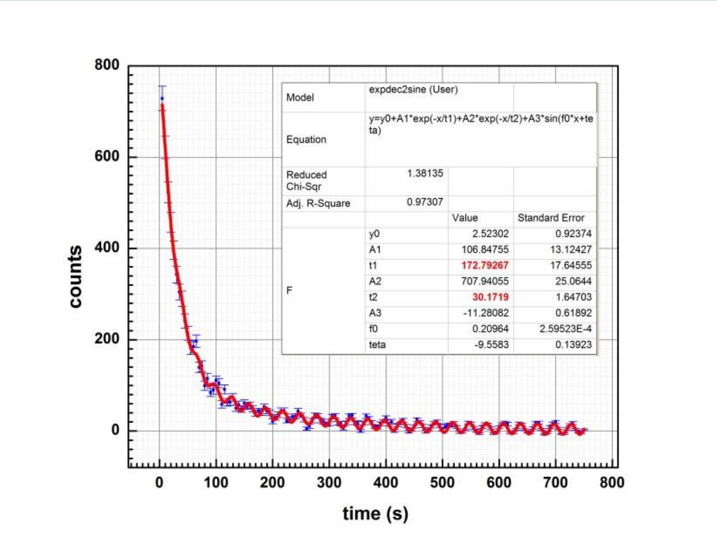




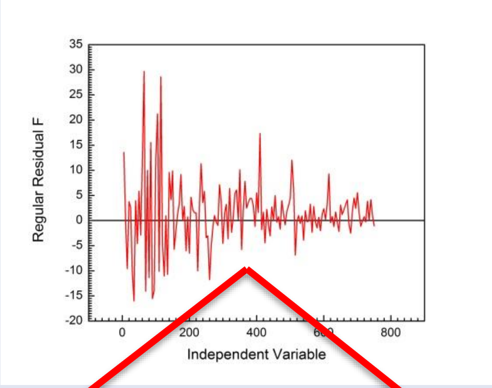
Conclusion: fitting function should be modified by adding an additional term:

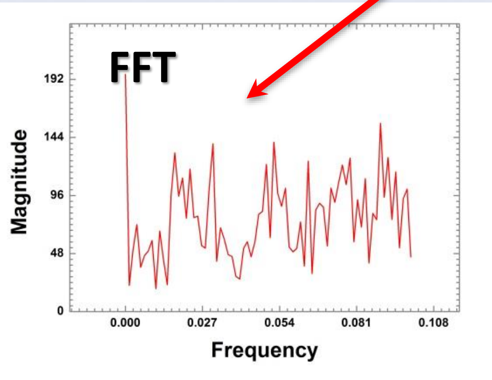
$$y(t) = y_0 + A_1 \exp\left(\frac{-t}{t_1}\right) + A_2 \exp\left(\frac{-t}{t_2}\right) + A_3 \sin(\omega t + \theta)$$

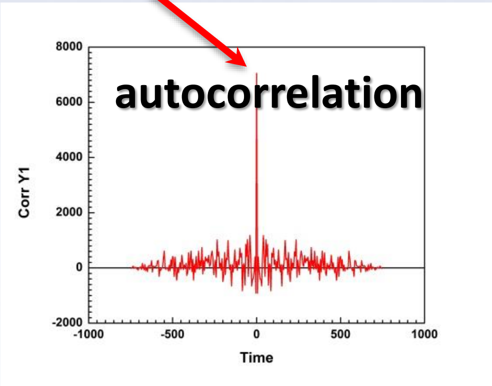
Fitting. Analysis of the residuals. Non "ideal" case



Model	expdec2sine (User)		
Equation	y=y0+A1*exp(-x/t1)+A2*exp(-x/t2)+A3*sin(f0*x+teta)		
Reduced Chi-Sqr	1.38135		
Adj. R-Square	0.97307		
		Value	Standard Error
F	y0	2.52302	0.92374
	A1	106.84755	13.12427
	t1	172.79267	17.64555
	A2	707.94055	25.0644
	t2	30.1719	1.64703
	A3	-11.28082	0.61892
	f0	0.20964	2.59523E-4
	teta	-9.5583	0.13923







	Clear experiment	Data + noise	Modified fitting
t_1(s)	177.76	145.89	172.79
t_2(s)	30.32	27.94	30.17

Error Analysis. Millikan oil drop experiment.

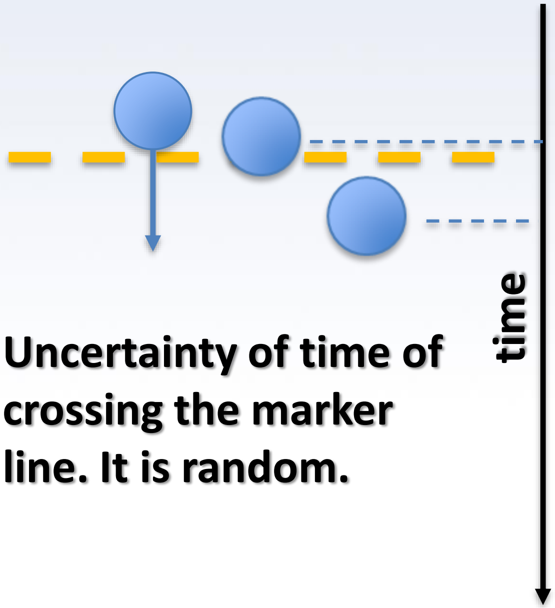
In general we could expect both components of errors

$$Q_{meas} = Q_{true} + e_s + e_r$$

e_s - systematic error comes from uncertainties of plates separation distance, applied DC voltage, ambient temperature etc.

$V = V_{DC} \pm \Delta V, \ d = d_0 \pm \Delta d \ \ldots$

e_r - random errors are related to uncertainty of the knowledge of the actual t_g and t_{rise}.



Uncertainty of time of crossing the marker line. It is random.

Systematic component. Error propagation. Millikan oil drop experiment.

$$\boxed{Q = F \bullet S \bullet T = \left(\frac{1}{f_c^{3/2}}\right)\frac{9\pi d}{V}\sqrt{\frac{2\eta^3 x^3}{g\rho}}\frac{1}{\sqrt{t_g}}\left(\frac{1}{t_g}+\frac{1}{t_{rise}}\right)}$$

$$\Delta Q = \sqrt{\left(S \bullet T\right)^2 \Delta F^2 + \left(F \bullet T\right)^2 \Delta S^2 + \left(F \bullet S\right)^2 \Delta T^2}$$

$$\boxed{T = \frac{1}{\sqrt{t_g}}\left(\frac{1}{t_g}+\frac{1}{t_{rise}}\right)}$$

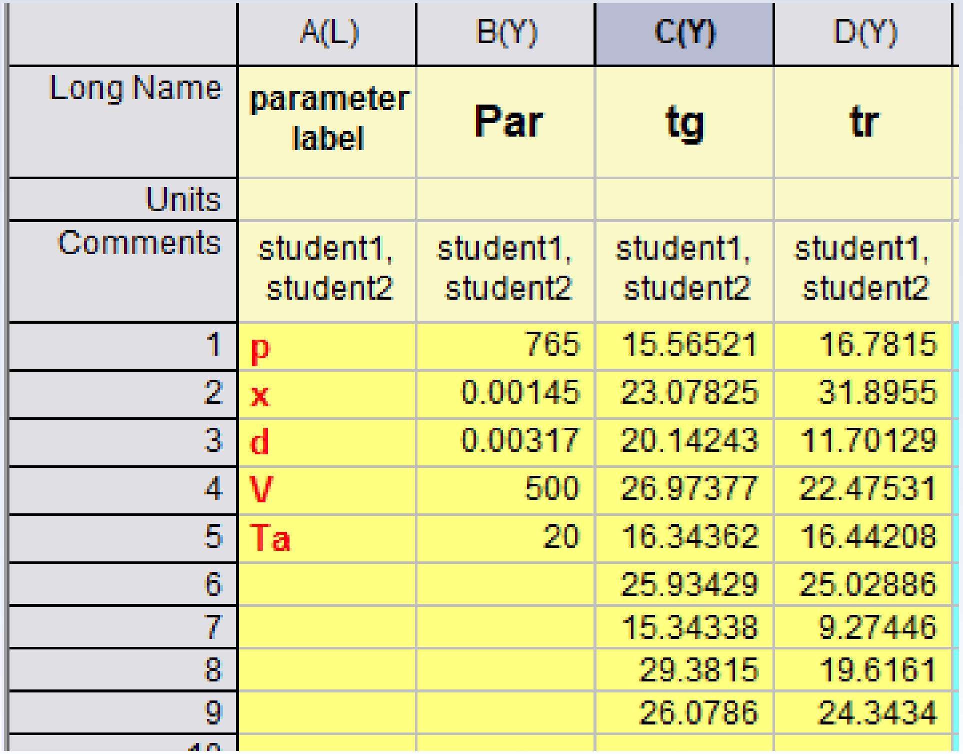
$$\boxed{\Delta T = \sqrt{\left(\frac{3/2}{t_g^{5/2}}+\frac{1/2}{t_g^{3/2}}\frac{1}{t_{rise}}\right)^2 \Delta t_g^2 + \left(\frac{1}{t_g^{1/2}}\frac{1}{t_{rise}^2}\right)^2 \Delta t_{rise}^2}}$$

Appendix #1. Analyzing of the statistical data.

Step 1. Collect your data + parameters of the experiment in:

\\engr-file-03\PHYINST\APL Courses\PHYCS401\Students\1. Millikan Oil Drop experiment\Section L1m.opj

Use different columns for each student or team. This Origin project is for data collecting only but not for data analysis. For data analysis you have to copy these data and experiment parameters obtained by different students/team and paste it in one in your personal Origin project.



	A(L)	B(Y)	C(Y)	D(Y)
Long Name	parameter label	Par	tg	tr
Units				
Comments	student1, student2	student1, student2	student1, student2	student1, student2
1	p	765	15.56521	16.7815
2	x	0.00145	23.07825	31.8955
3	d	0.00317	20.14243	11.70129
4	V	500	26.97377	22.47531
5	Ta	20	16.34362	16.44208
6			25.93429	25.02886
7			15.34338	9.27446
8			29.3815	19.6161
9			26.0786	24.3434

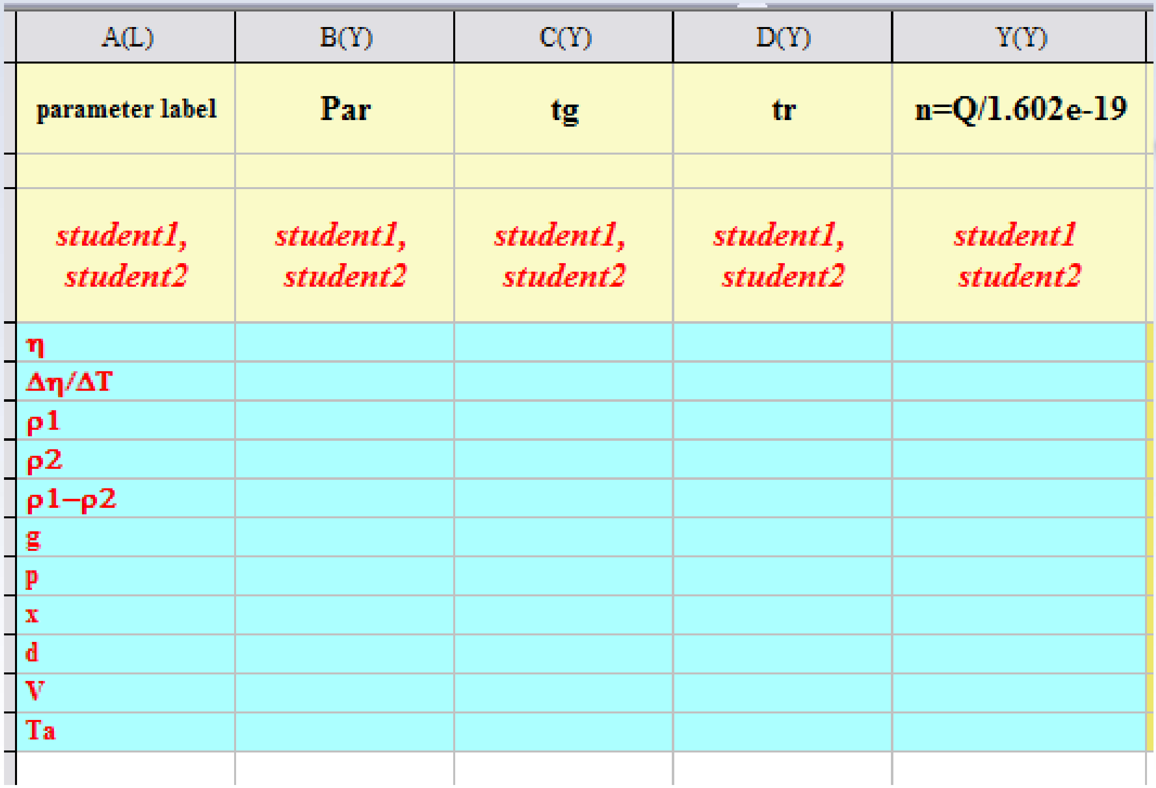
Setup and environmental parameters

Raw data

Appendix #1. Analyzing of the statistical data.

Step 1. Slightly Modified Origin Prpjeckt For Data Collection:

\\engr-file-03\PHYINST\APL Courses\PHYCS401\Students\1. Millikan Oil Drop experiment\Section L1m.opj



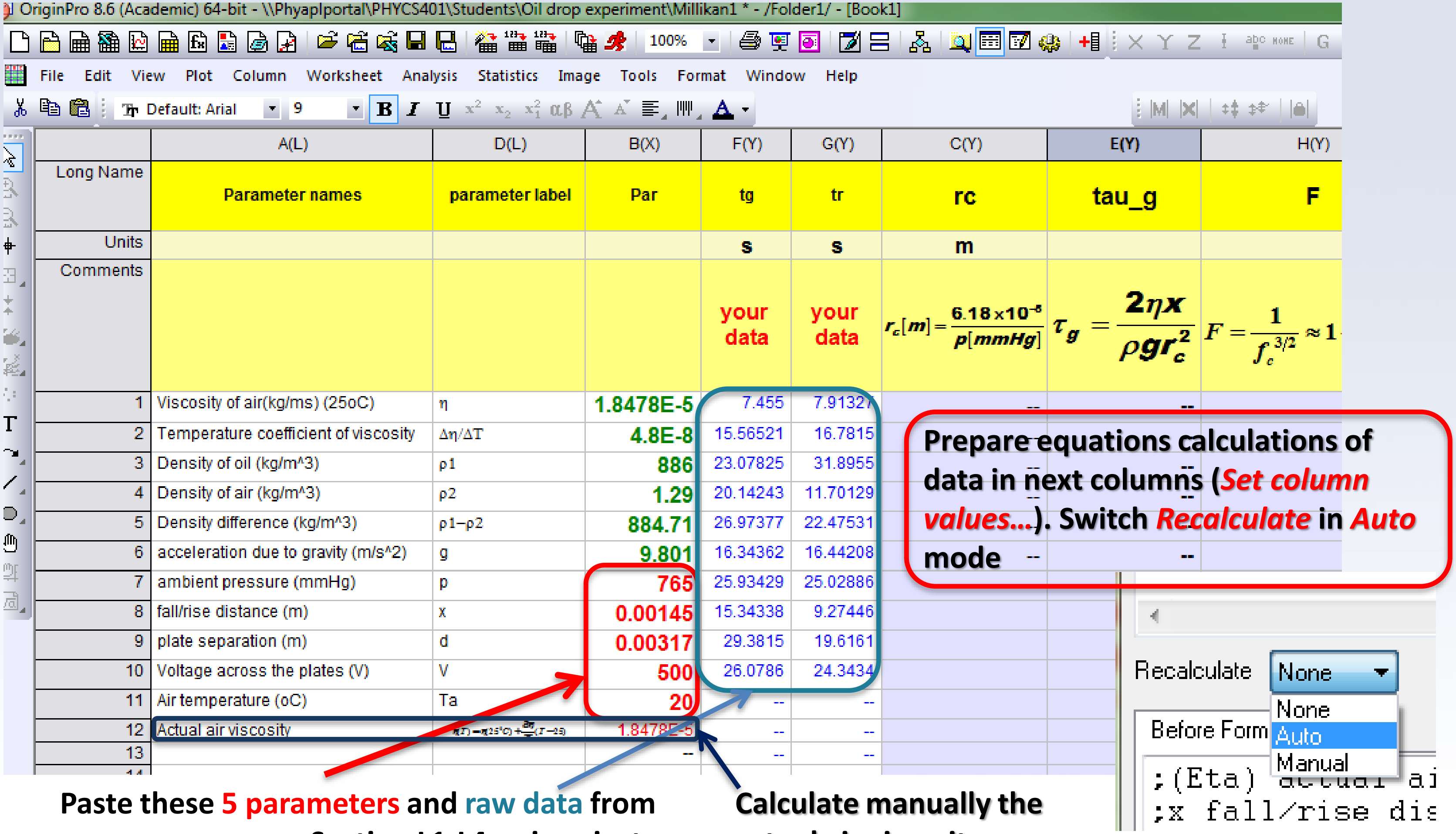
A(L)	B(Y)	C(Y)	D(Y)	Y(Y)
parameter label	Par	tg	tr	n=Q/1.602e-19
student1, student2	*student1, student2*	*student1, student2*	*student1, student2*	*student1 student2*
η				
Δη/ΔT				
ρ1				
ρ2				
ρ1–ρ2				
g				
p				
x				
d				
V				
Ta				

Extra column for number of elementary charges

Appendix #1. Analyzing of the statistical data.

Step 2. Working on personal Origin project

Make a copy of the Millikan1 project to your personal folder and open it



OriginPro 8.6 (Academic) 64-bit - \\Phyaplportal\PHYCS401\Students\Oil drop experiment\Millikan1 * - /Folder1/ - [Book1]

	A(L)	D(L)	B(X)	F(Y)	G(Y)	C(Y)	E(Y)	H(Y)
Long Name	Parameter names	parameter label	Par	tg	tr	rc	tau_g	F
Units				s	s	m		
Comments				your data	your data	$r_c[m]=\frac{6.18\times10^{-6}}{p[mmHg]}$	$\tau_g=\frac{2\eta x}{\rho g r_c^2}$	$F=\frac{1}{f_c^{3/2}}\approx 1$
1	Viscosity of air(kg/ms) (25oC)	η	1.8478E-5	7.455	7.91327	--	--	
2	Temperature coefficient of viscosity	Δη/ΔT	4.8E-8	15.56521	16.7815	--	--	
3	Density of oil (kg/m^3)	ρ1	886	23.07825	31.8955	--	--	
4	Density of air (kg/m^3)	ρ2	1.29	20.14243	11.70129	--	--	
5	Density difference (kg/m^3)	ρ1−ρ2	884.71	26.97377	22.47531	--	--	
6	acceleration due to gravity (m/s^2)	g	9.801	16.34362	16.44208	--	--	
7	ambient pressure (mmHg)	p	765	25.93429	25.02886			
8	fall/rise distance (m)	x	0.00145	15.34338	9.27446			
9	plate separation (m)	d	0.00317	29.3815	19.6161			
10	Voltage across the plates (V)	V	500	26.0786	24.3434			
11	Air temperature (oC)	Ta	20	--	--			
12	Actual air viscosity	$\eta(T)=\eta(25^{\circ}C)+\frac{\partial\eta}{\partial T}(T-25)$	1.8478E-5	--	--			
13			--	--	--			

Recalculate: None

```
;(Eta) actual ai
;x fall/rise dis
```

Prepare equations calculations of data in next columns (*Set column values…*). Switch *Recalculate* in *Auto* mode

Paste these 5 parameters and raw data from Section L1-L4.opj projects

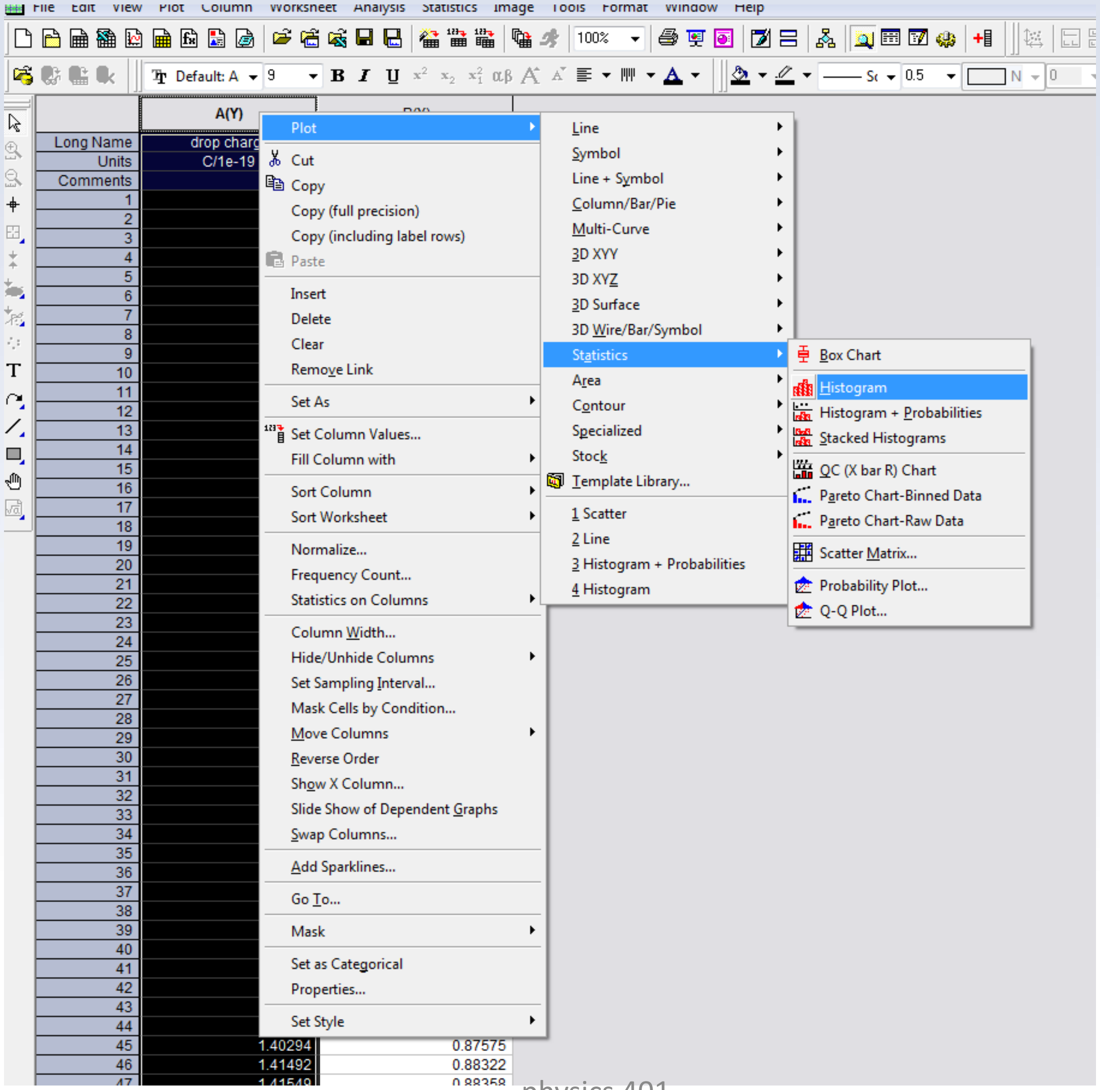
Calculate manually the actual air viscosity

Appendix #1. Analyzing of the statistical data.

Millikan oil drop experiment

Step 3. Histogram graph

First use the data from the column with drop charges and plot the histogram



illinois.edu

Appendix. Analyzing of the statistical data.

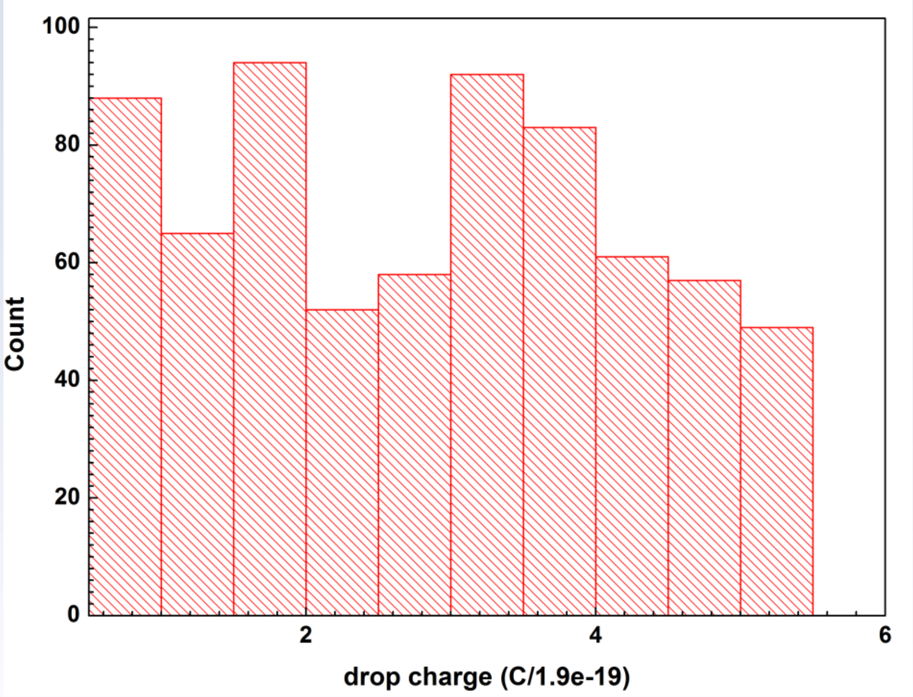
Millikan oil drop experiment

Step 4. Histogram. Bin size

Origin will automatically but not optimally adjust the bin size h. In tis page figure **h=0.5**. There are several theoretical approaches how to find the optimal bin size.

$$h = \frac{3.5\sigma}{n^{1/3}}$$

σ Is the sample standard deviation and **n** is total number of observation. For presented in Fig.1 results good value of **h ~0.1**



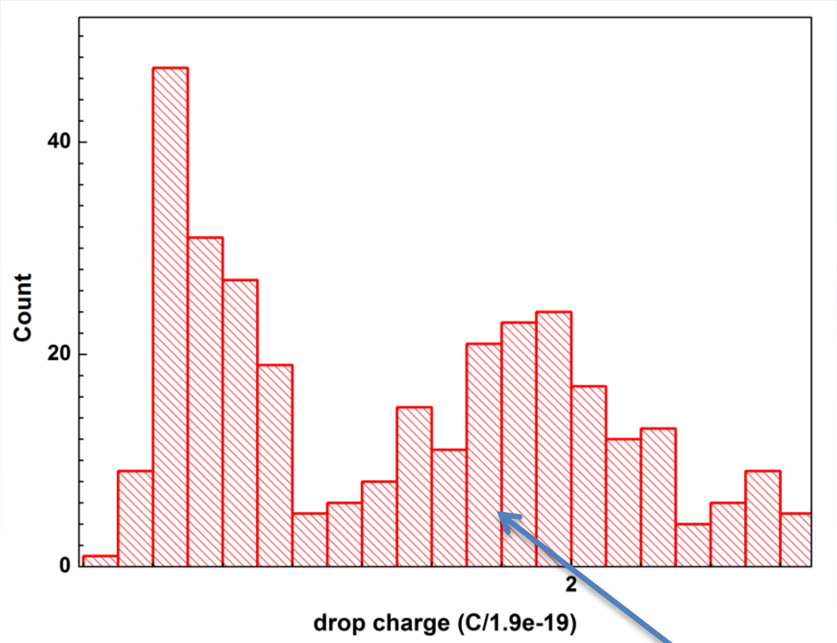
illinois.edu

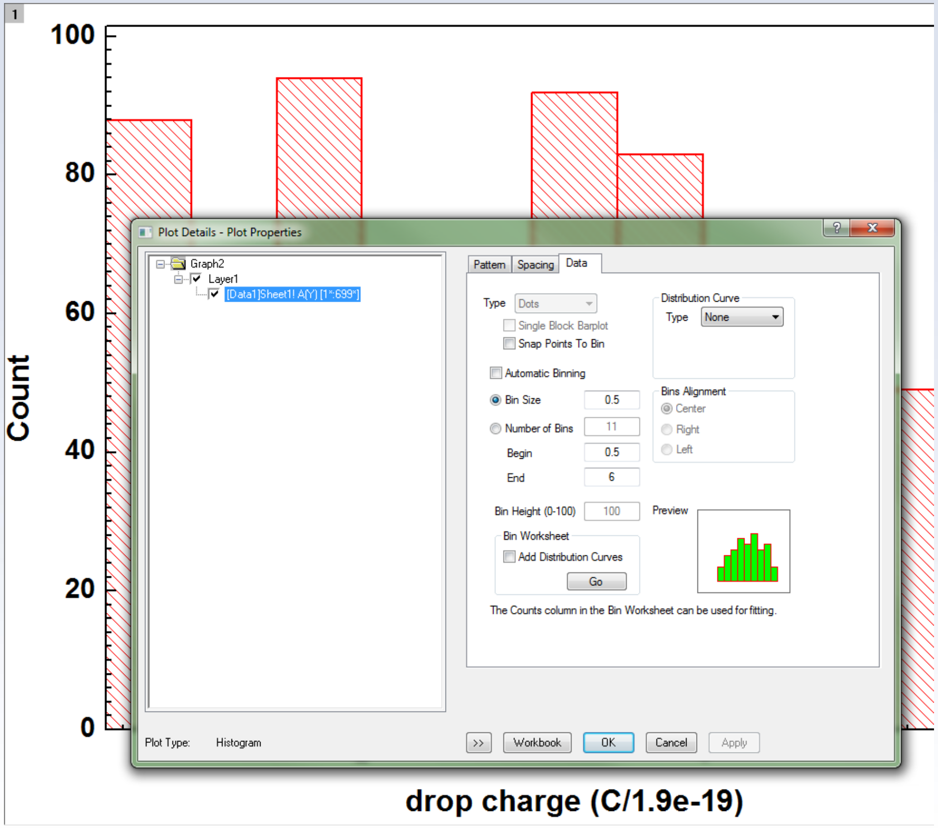
Appendix #1. Analyzing of the statistical data.

Millikan oil drop experiment

Step 4. Histogram. Bin size

To change the bin size click on graph and unplug the "*Automatic Binning*" option





Type: Dots

☐ Single Block Barplot

☐ Snap Points To Bin

☐ Automatic Binning

● Bin Size 0.5

○ Number of Bins 11

Begin 0.5

End 6

Bin Height (0-100) 100

Distribution Curve Type: None

Bins Alignment: ● Center ○ Right ○ Left

☐ Add Distribution Curves

The Counts column in the Bin Worksheet can be used for fitting.

Plot Type: Histogram

Bin size in this histogram is 0.1

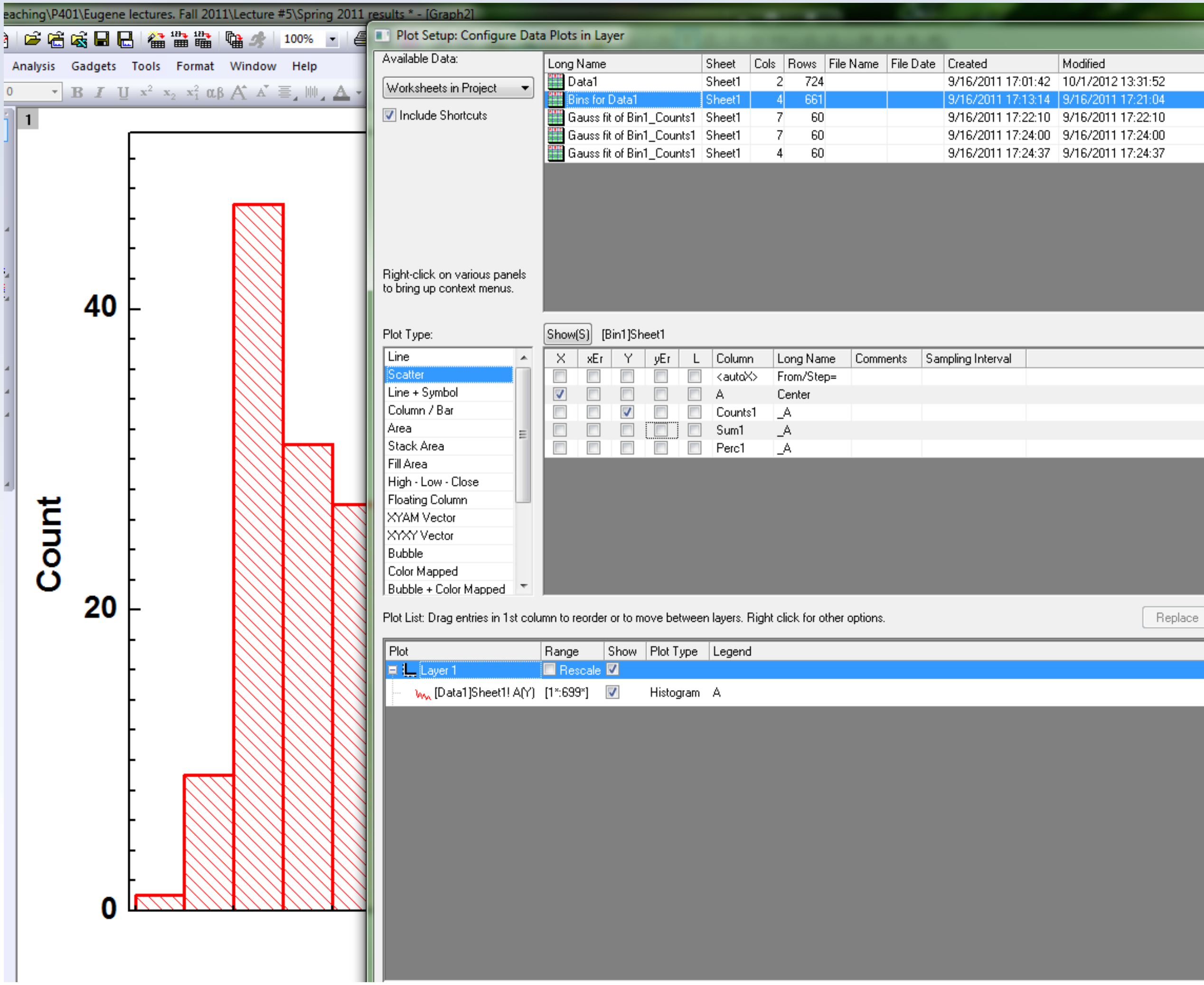
Appendix #1. Analyzing of the statistical data.

Step 4. Multipeak Gaussian fitting

Millikan oil drop experiment

To do this you have to add an extra plot to the graph

Counts vs. Bin Center



Available Data:

Worksheets in Project ▾

☑ Include Shortcuts

Long Name	Sheet	Cols	Rows	File Name	File Date	Created	Modified
Data1	Sheet1	2	724			9/16/2011 17:01:42	10/1/2012 13:31:52
Bins for Data1	Sheet1	4	661			9/16/2011 17:13:14	9/16/2011 17:21:04
Gauss fit of Bin1_Counts1	Sheet1	7	60			9/16/2011 17:22:10	9/16/2011 17:22:10
Gauss fit of Bin1_Counts1	Sheet1	7	60			9/16/2011 17:24:00	9/16/2011 17:24:00
Gauss fit of Bin1_Counts1	Sheet1	4	60			9/16/2011 17:24:37	9/16/2011 17:24:37

Right-click on various panels to bring up context menus.

Plot Type:

- Line
- Scatter
- Line + Symbol
- Column / Bar
- Area
- Stack Area
- Fill Area
- High - Low - Close
- Floating Column
- XYAM Vector
- XYXY Vector
- Bubble
- Color Mapped
- Bubble + Color Mapped

Show(S) [Bin1]Sheet1

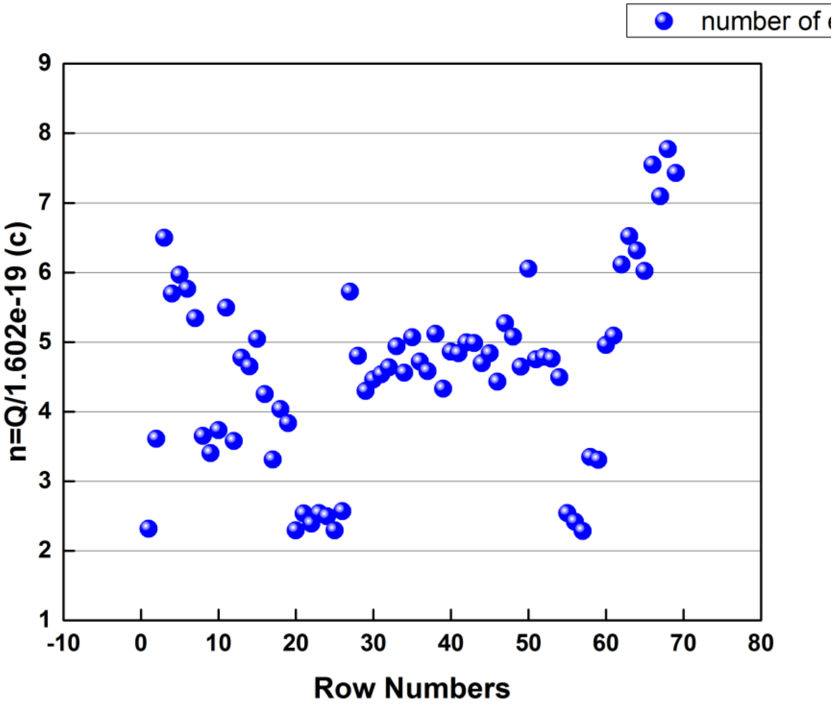
X	xEr	Y	yEr	L	Column	Long Name	Comments	Sampling Interval
☐	☐	☐	☐	☐	<autoX>	From/Step=		
☑	☐	☐	☐	☐	A	Center		
☐	☐	☑	☐	☐	Counts1	_A		
☐	☐	☐	☐	☐	Sum1	_A		
☐	☐	☐	☐	☐	Perc1	_A		

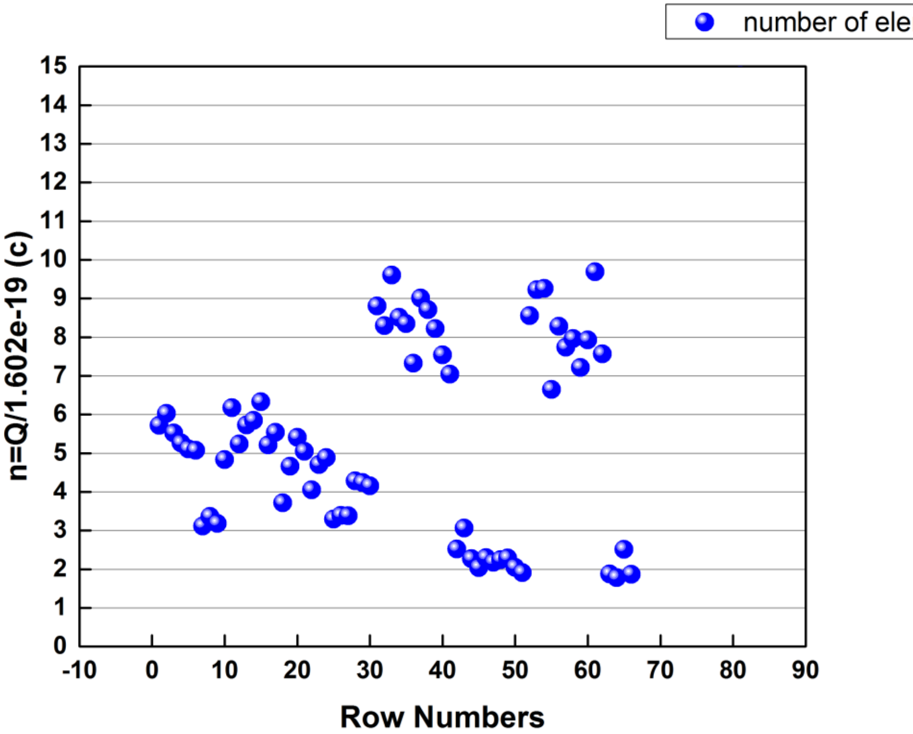
Plot List: Drag entries in 1st column to reorder or to move between layers. Right click for other options.

Plot	Range	Show	Plot Type	Legend
Layer 1	☐ Rescale	☑		
[Data1]Sheet1! A(Y)	[1*:699*]	☑	Histogram	A

Appendix #1. oil Drop Data Issue.

Be careful with data selection obtained by different teams!





Appendix #1. oil Drop Data Issue.

Write-up, page 7. mistype in some copies

Wrong!

quantity	Value
Viscosity of air	η = 1.8478 10^{-5} kg/m·s (25 ºC) [illegible])
Density of oil	ρ_{oil} = 886 kg/m^3
Density of air	ρ_{air} = 1.29 kg/m^3
Acceleration due to gravity	g = 9.801 m/s^3

Correct number

quantity	Value
Viscosity of air	η = 1.8478 10^{-5} kg/m·s (25 ºC)) $\frac{d\eta}{dT}$=**4.8 10^{-8} kg/m·s/ºC**
Density of oil	ρ_{oil} = 886 kg/m^3
Density of air	ρ_{air} = 1.29 kg/m^3
Acceleration due to gravity	g = 9.801 m/s^3

$$\Delta\eta \Big/ \Delta T = 4.8 \times 10^{-8}\ kg / m \cdot s / {}^{\circ}C$$

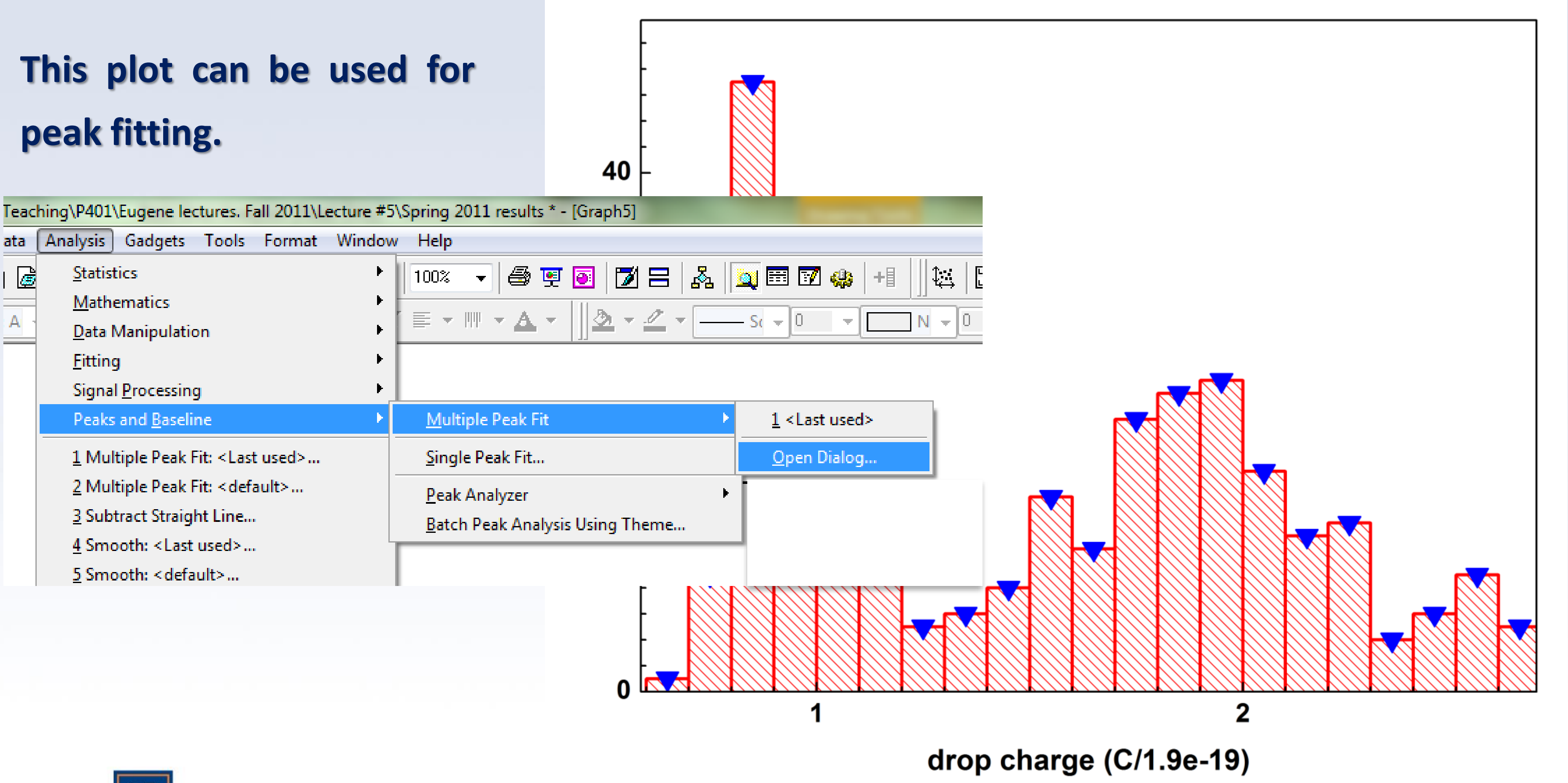
$$\eta(T) = \eta(25C^{o}) + 4.8 \times 10^{-8} \times (T - 25)(\mathrm{kg/\ m * s/\ }{}^{o}C)$$

Appendix #1. Analyzing of the statistical data.

Millikan oil drop experiment

Step 4. Multipeak Gaussian fitting

This plot can be used for peak fitting.



Appendix #1. Analyzing of the statistical data.

Millikan oil drop experiment

Step 4. Multipeak Gaussian fitting

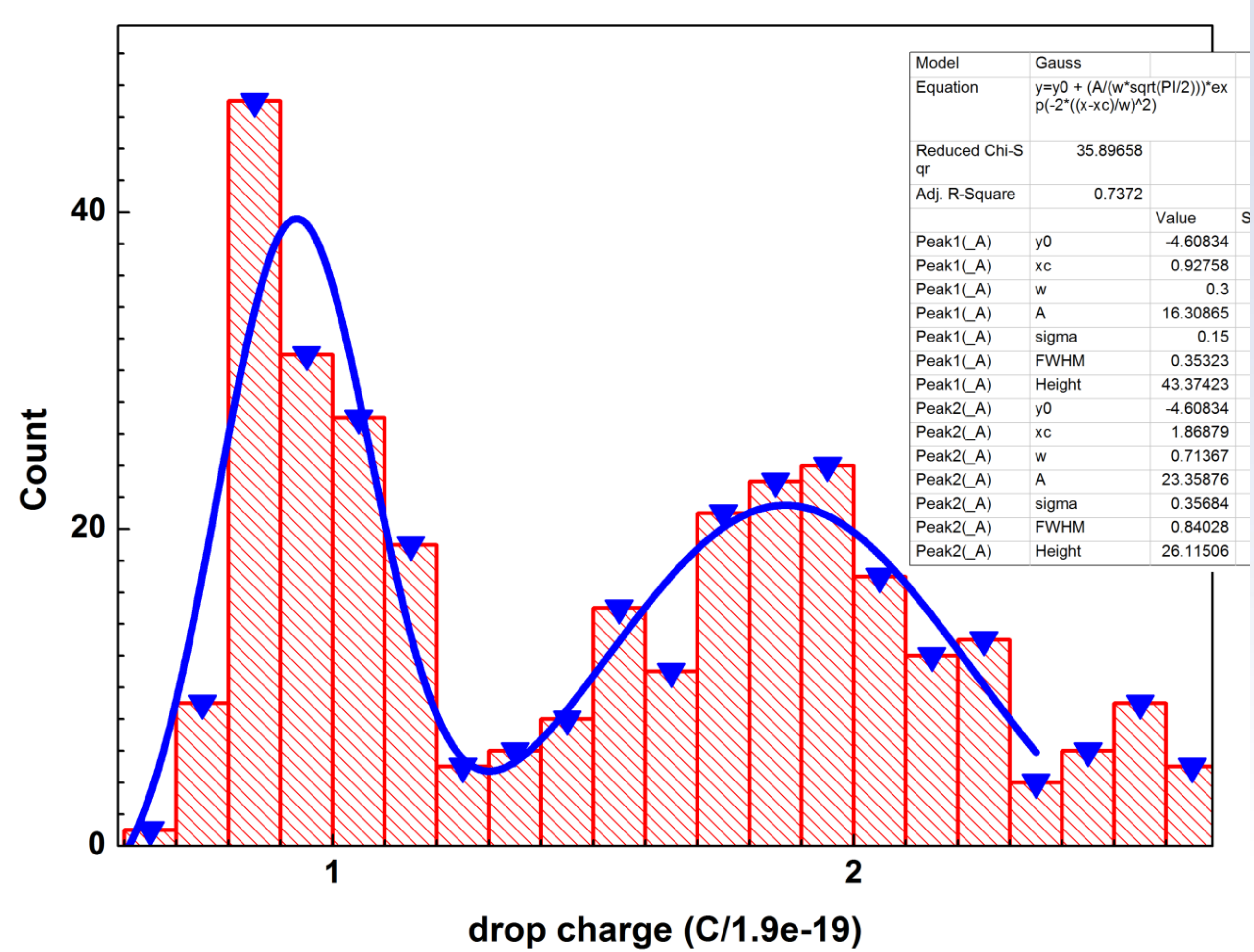
This plot can be used for peak fitting.

Final result for first two peaks:

Q/e=0.93±0.01

Q/e=1.87±0.07

This pretty close to e and 2e



Model	Gauss		
Equation	y=y0 + (A/(w*sqrt(PI/2)))*exp(-2*((x-xc)/w)^2)		
Reduced Chi-Sqr	35.89658		
Adj. R-Square	0.7372		
		Value	S
Peak1(_A)	y0	-4.60834	
Peak1(_A)	xc	0.92758	
Peak1(_A)	w	0.3	
Peak1(_A)	A	16.30865	
Peak1(_A)	sigma	0.15	
Peak1(_A)	FWHM	0.35323	
Peak1(_A)	Height	43.37423	
Peak2(_A)	y0	-4.60834	
Peak2(_A)	xc	1.86879	
Peak2(_A)	w	0.71367	
Peak2(_A)	A	23.35876	
Peak2(_A)	sigma	0.35684	
Peak2(_A)	FWHM	0.84028	
Peak2(_A)	Height	26.11506	

Here $w = 2\sigma$ *and error of the mean* $= \frac{\sigma}{\sqrt{N}} \sim \frac{\boldsymbol{sigma}}{\sqrt{\boldsymbol{A}}}$

Appendix #1. Fitting. Main Idea.

(x_i,y_i) is an experimental data array. **x_i** is an independent variable and **y_i** - dependent

$f(x,\beta)$ is a model function and β is the vector of fitting (adjustable) parameters

The goal of the fitting procedure is to find the set of parameters which will generate the function ***f*** closest to the experimental points.

To reach this goal we will try to minimize the sum of squared deviation function (**S**):

$$S(\beta)=\sum_{i=1}^{m}[y_i-f(x_i,\beta)]^2$$

illinois.edu

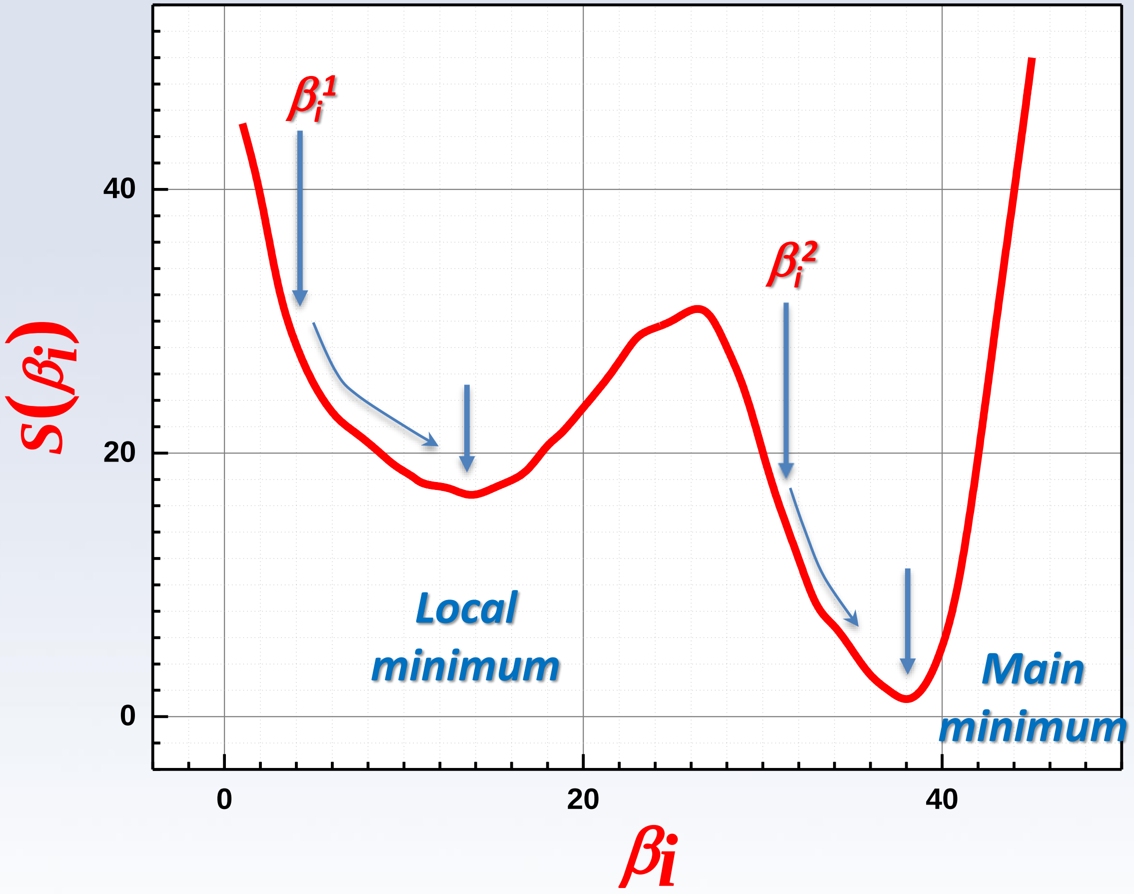
Appendix #1. Fitting. The Choice of Parameters.

The goal of fitting is not only to find the curve best matching the experimental data but also to find the corresponding parameters which in majority cases are the important physical parameters

There are several known mathematical algorithms for optimizing these parameters but in general the fitting procedure could have not only unique solution and the choice of initial parameters is very important issue

$$S(\beta) = \sum_{i=1}^{m} \left[y_i - f(x_i, \beta) \right]^2$$

Appendix #1. Fitting. The Choice of Parameters.



Let we have the ***S*** function dependent on parameter β_i as shown on this graph