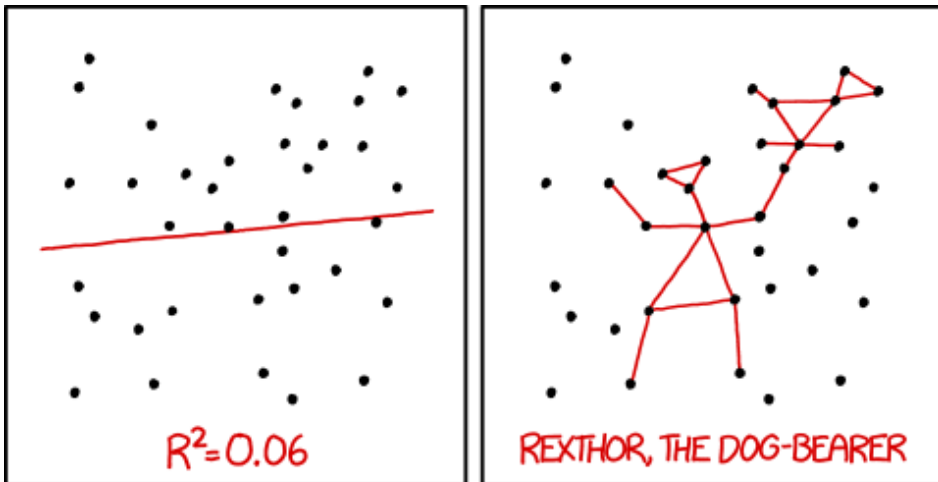


Basic Error Analysis



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

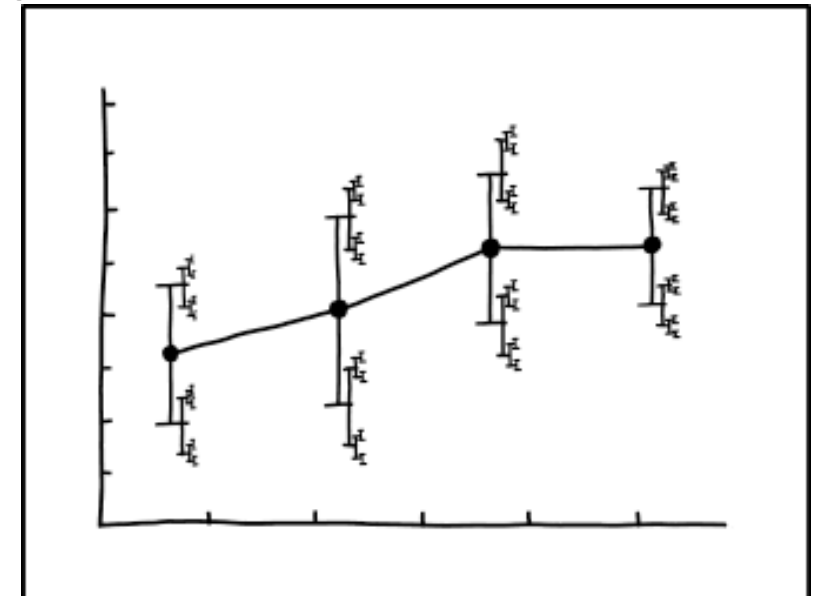
[XKCD #1752](#)

Professor Jeff Filippini

Physics 401

Spring 2020

I ILLINOIS



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

[XKCD #2110](#)

Today's Topics

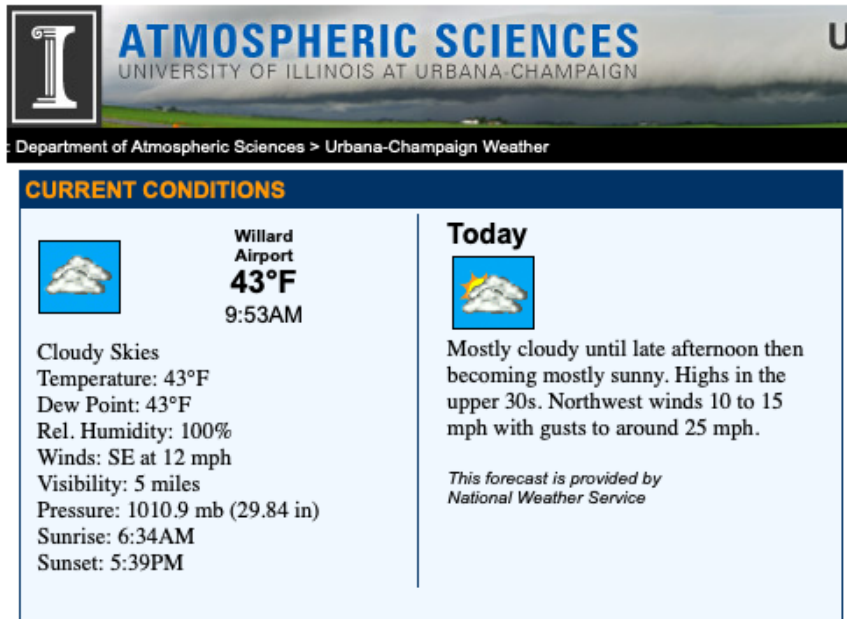
1. Errors and Uncertainties
2. Instrument Errors
3. Accuracy and Precision
4. Systematic and Statistical Errors
5. Evaluating Fits and Parameter Errors

Appendix: More on oil drop data analysis, fitting

Redeeming some “Bad Words”

- **Errors** are not mistakes!
- **Statistics** is, in a deep sense, what science is about!
 - “The discipline that concerns the collection, organization, analysis, interpretation and presentation of data” ([Wikipedia](#))
 - What conclusions can you draw about the universe from your observations?
- Careful **error analysis** is critical to learning from data

Errors: Everyday Life



ATMOSPHERIC SCIENCES
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Department of Atmospheric Sciences > Urbana-Champaign Weather

CURRENT CONDITIONS

Willard Airport
43°F
9:53AM

Today

Mostly cloudy until late afternoon then becoming mostly sunny. Highs in the upper 30s. Northwest winds 10 to 15 mph with gusts to around 25 mph.

This forecast is provided by National Weather Service

Cloudy Skies
Temperature: 43°F
Dew Point: 43°F
Rel. Humidity: 100%
Winds: SE at 12 mph
Visibility: 5 miles
Pressure: 1010.9 mb (29.84 in)
Sunrise: 6:34AM
Sunset: 5:39PM

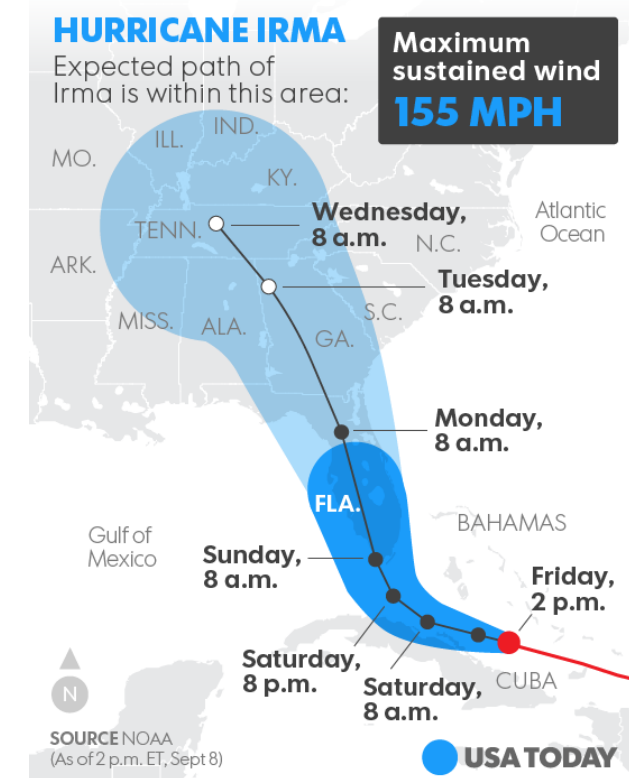


$$T = 43^{\circ}\text{F} \pm ??$$

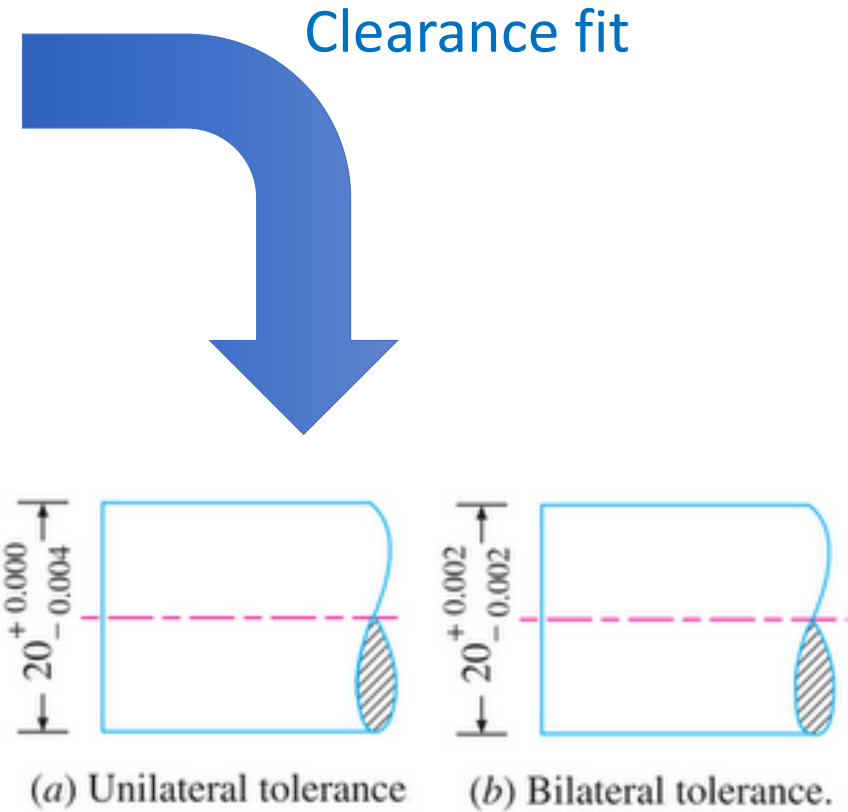
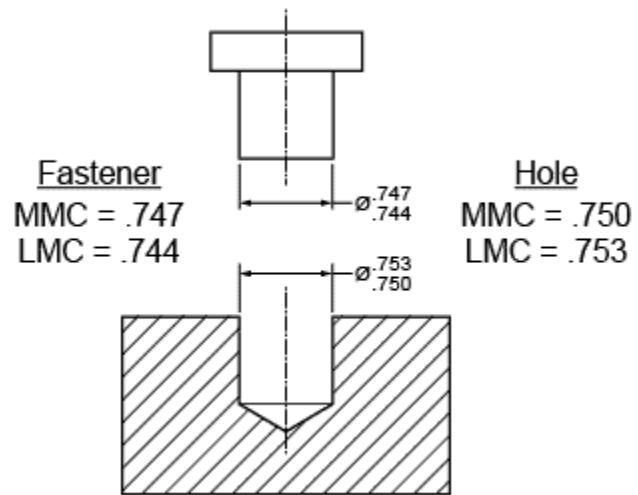
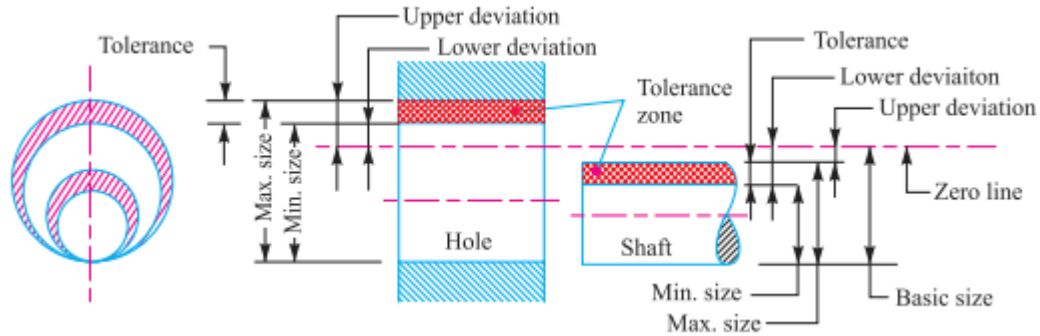
Best guess: $\Delta T \sim 0.5^{\circ}\text{F}$

$$\text{Wind speed } 12 \text{ mph} \pm ??$$

Best guess: $\Delta v \sim 0.5 \text{ mph}$



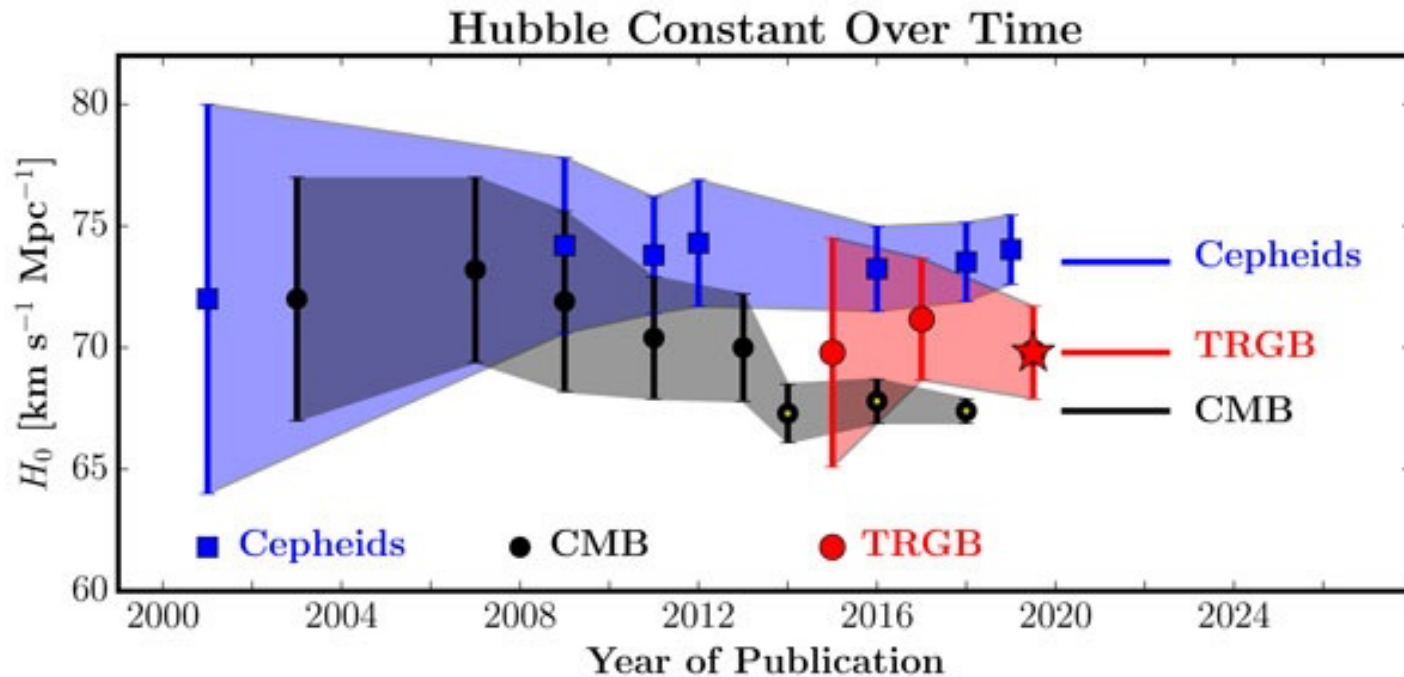
Errors: Industry



Tight tolerances cost \$\$\$\$!!!

Errors: Science

How fast is the universe expanding? The **Hubble Constant**



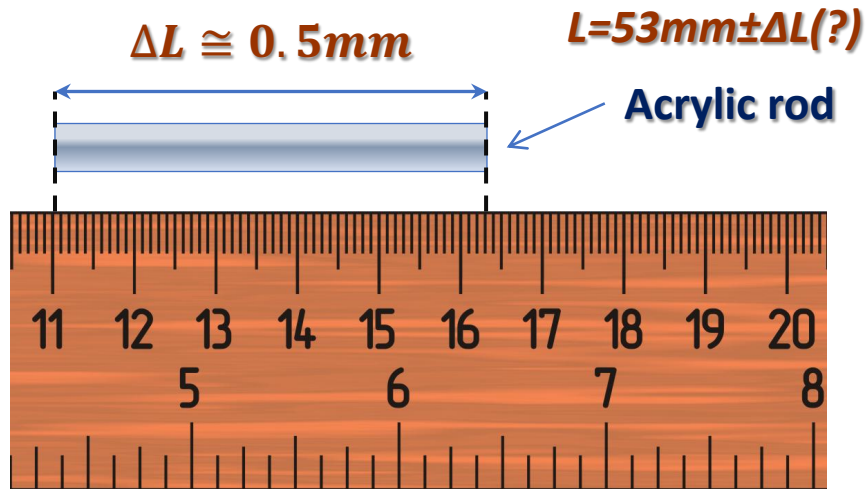
We seek [consilience](#) between different techniques
Do we need a new model, or are we missing sources of error??

Multiple methods:

- **Standard candles:**
How bright do objects of known properties appear?
 - Peculiar variable stars
 - Largest possible red giant stars
- **Standard rulers:**
How big do ripples in the cosmic microwave background look?

Measurements: Reading Errors

Typical reading error: $\pm\frac{1}{2}$ of minimum gradation or count



How far should we go to minimize reading error?

What do we Need?

In some cases we don't care about accuracy better than 1mm

Tools

If a ruler doesn't do the job, use a digital caliper or Vernier caliper

Intrinsic Limits

E.g. thermal expansion
CTE $\Delta L/L = 7.5 \times 10^{-5} \text{K}^{-1}$
 $\Delta L \sim 0.004 \text{ mm/K}$

Measurements: Reading Errors



Fluke 8846A multimeter

Example Vdc (reading)=0.85V on 1V range

$$\Delta V = 0.85 \times (1.8 \times 10^{-5})$$

$$+ 1.0 \times (0.6 \times 10^{-5}) \cong 2.2 \times 10^{-5} = 22 \mu V$$

8846A Accuracy

Accuracy is given as \pm (% measurement + % of range)

Range	24 Hour (23 \pm 1 $^{\circ}$ C)	90 Days (23 \pm 5 $^{\circ}$ C)	1 Year (23 \pm 5 $^{\circ}$ C)	Temperature Coefficient/ $^{\circ}$ C Outside 18 to 28 $^{\circ}$ 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 Precision: Conceptual



Accuracy

How close does the measurement come to the **true value**?

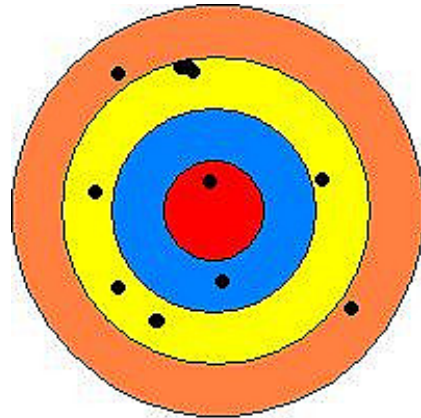


Precision

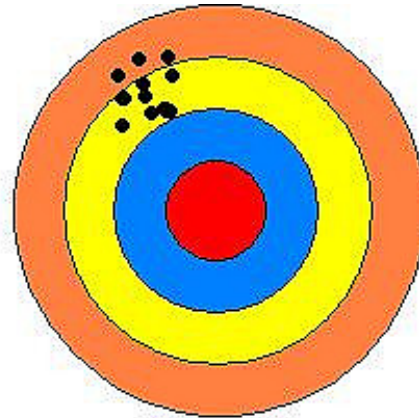
How close do individual measurements come to **each other**?

Repeatability, consistency

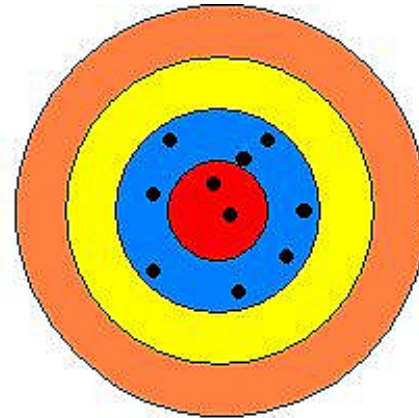
Accuracy and Precision: Conceptual



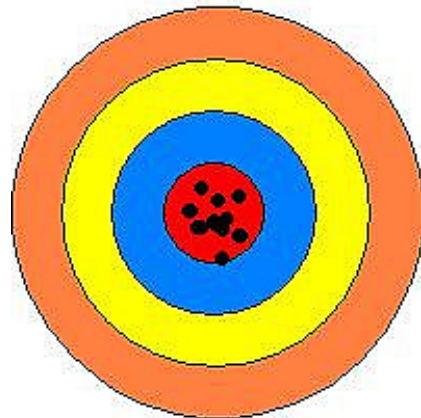
Not Precise, Not Accurate



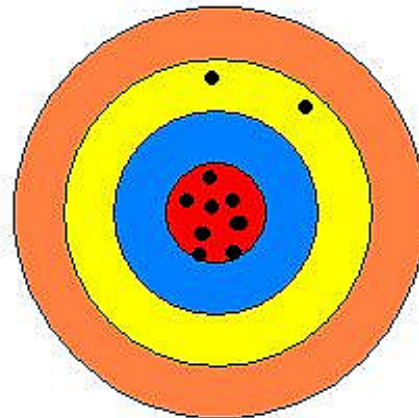
Precise, Not Accurate



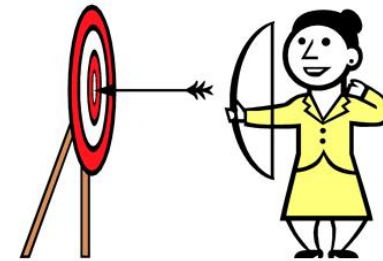
Accurate, Not Precise



Accurate, Precise

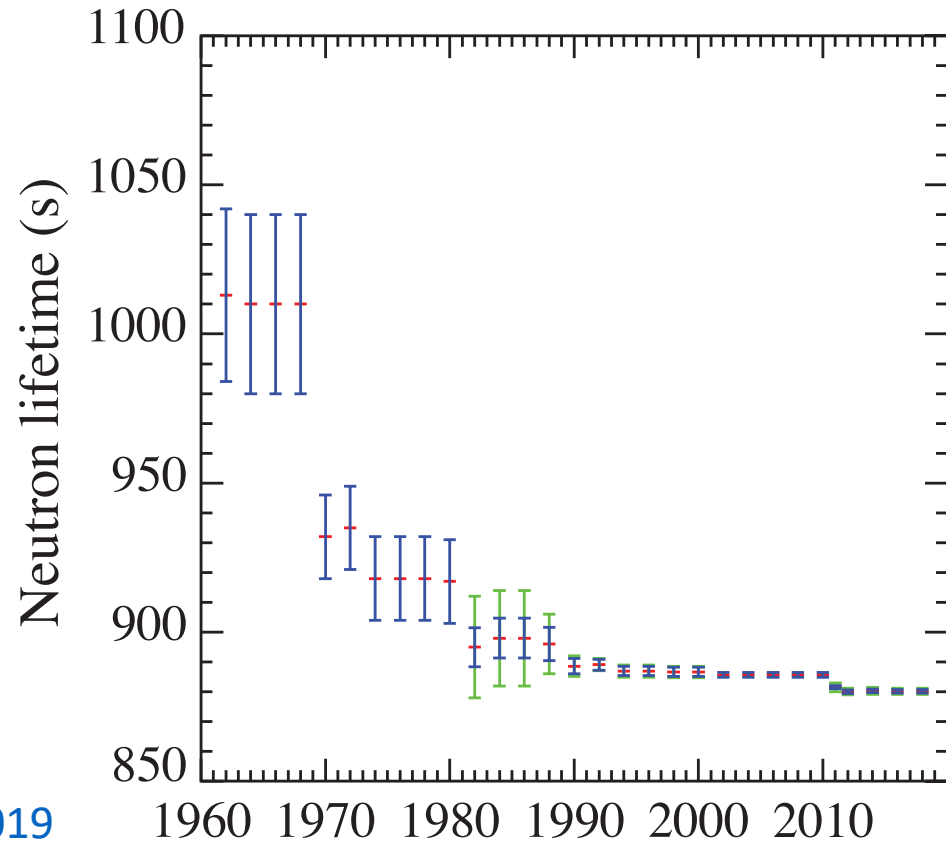


Errors



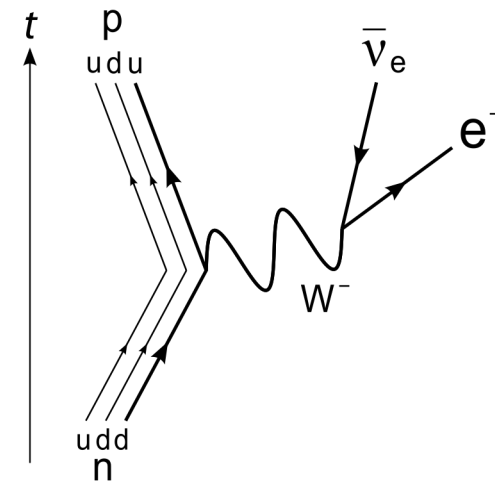
Accuracy and Precision: Reality

... but in experimental science we don't *know* the true value *a priori*!



[PDG 2019](#)

How long does a **free neutron** live before decaying to a proton, an electron, and a neutrino?



[Wikipedia: Free neutron decay](#)

Two Classes of Error

- **Systematic error**: **Reproducible** inaccuracy introduced by imperfect equipment, calibration or measurement technique.
- **Random error**: Indefiniteness of results due to finite precision of experiment. Measure of **fluctuation** in result after repeatable experimentation.

“Bias”

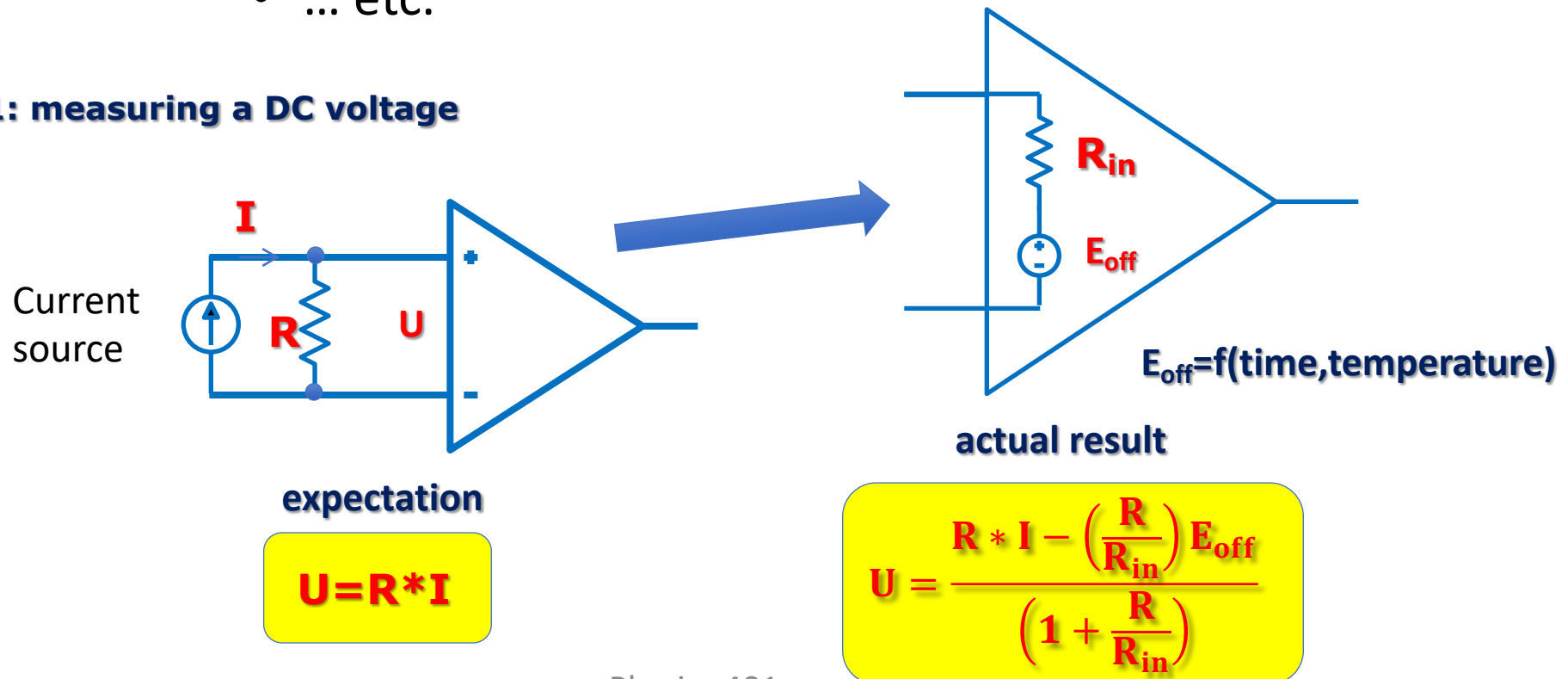
“Noise”

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

Sources of Systematic Error

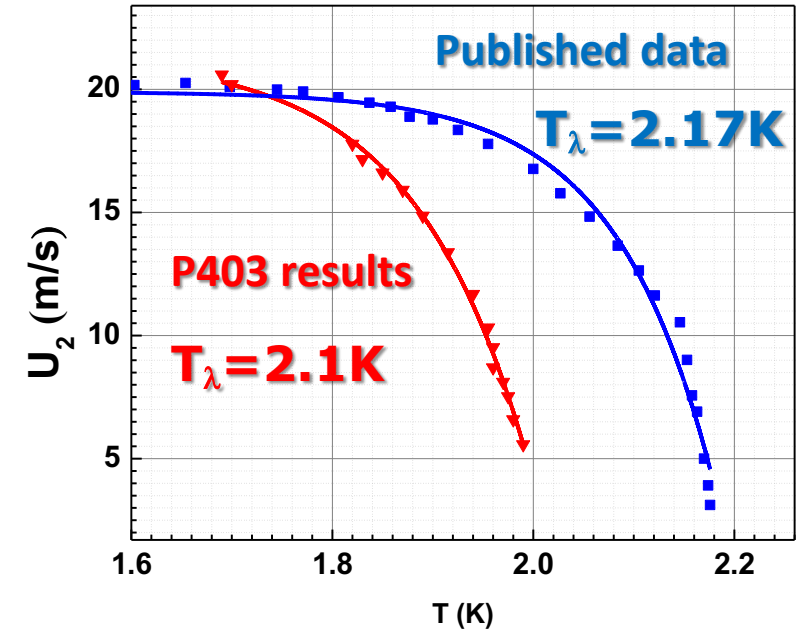
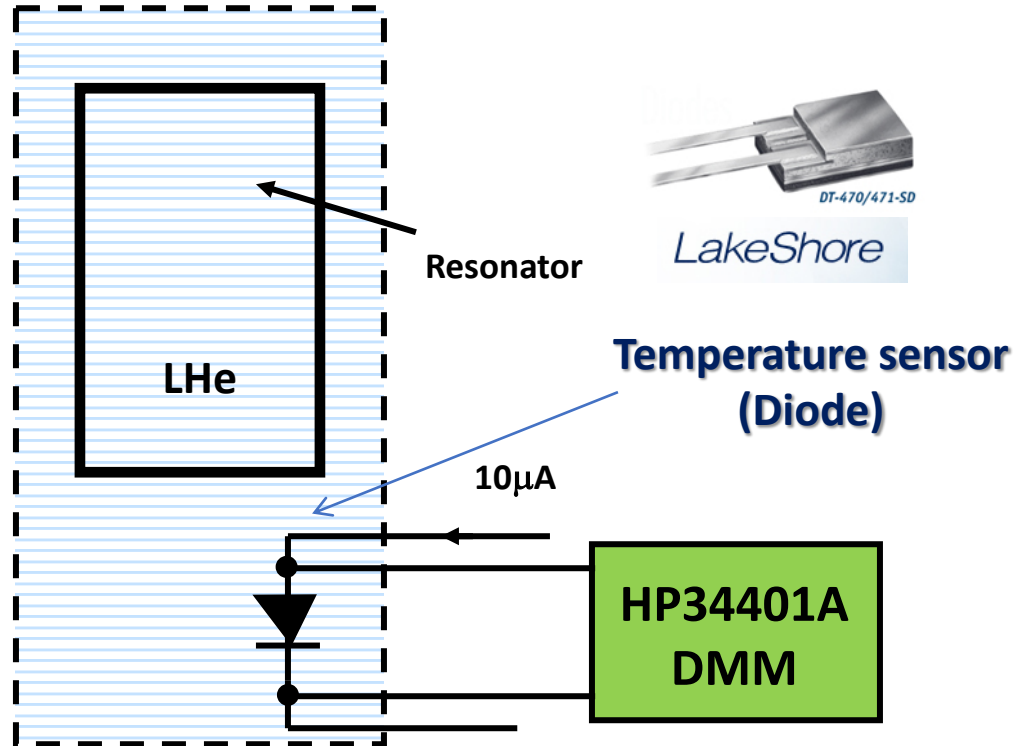
- Poor calibration of equipment
- Changes in environmental conditions
- Imperfect methods of measurement
- Offsets and drifts in instruments
- ... etc.

Example #1: measuring a DC voltage



Sources of Systematic Error

Example #2: Measuring the speed of "second sound" in superfluid 4He



Explanation is **imperfect calibration**
Common issue with diodes at low T , where dV/dT is small

Random Errors

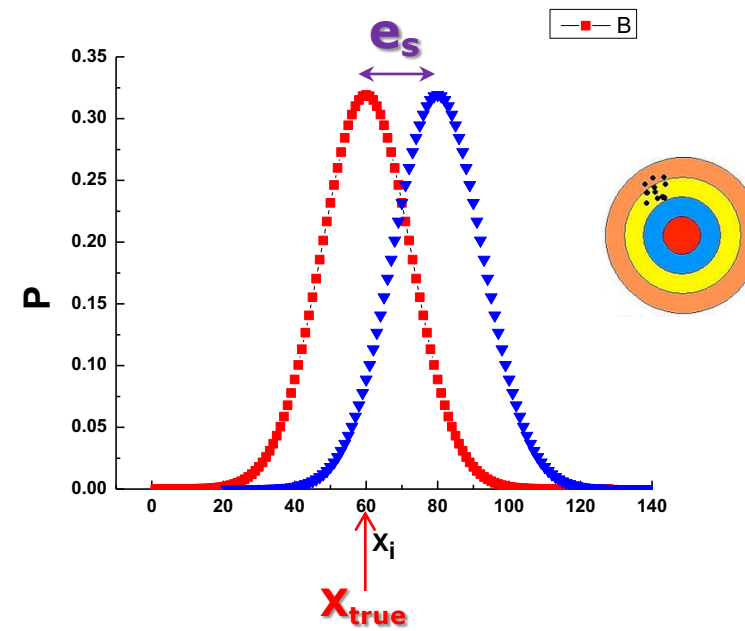
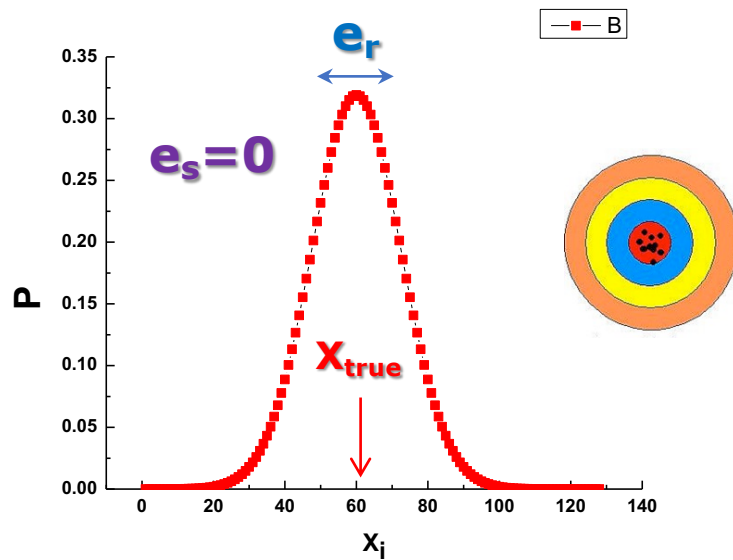
Measured value

$$x_{meas} = x_{true} + \boxed{e_s} + \boxed{e_r}$$

Correct value

Systematic error

Random error

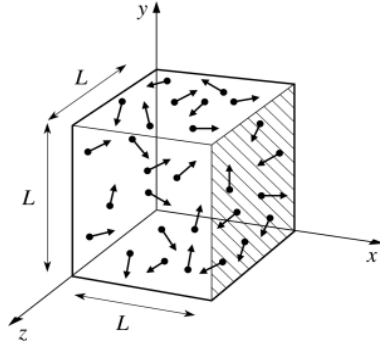


Where Does Randomness Come From?

Where have you found randomness in your other physics classes?

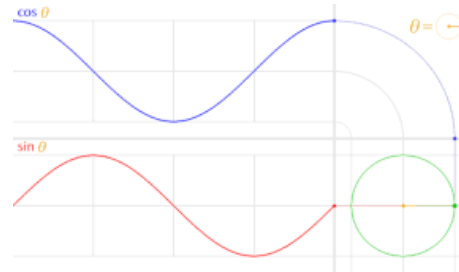
Ignorance

Coupling to large ensembles



Thermal fluctuations

Coupling through complicated functions



*Chaos
Pseudo-random #s*

Quantum mechanics



Measurement

Random Errors: Poisson Distribution



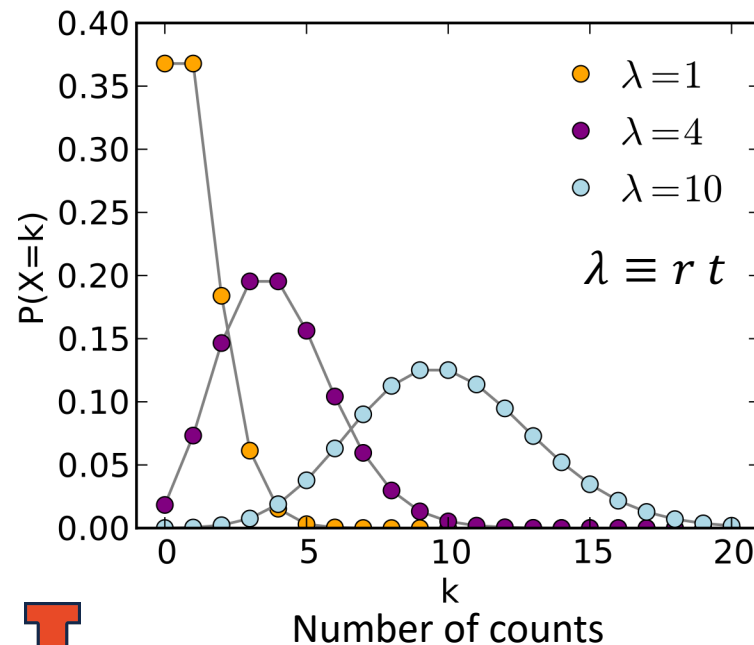
Siméon Denis Poisson
(1781-1840)

[Wikipedia: Poisson distribution](#)

$$P_n(t) = \frac{(r t)^n}{n!} e^{-rt}$$

Probability to observe n events in a time interval of length t .

Sole parameter is **rate** r (average events/second)



A statistical process (e.g. radioactive decay) is described through a Poisson distribution if it is/has:

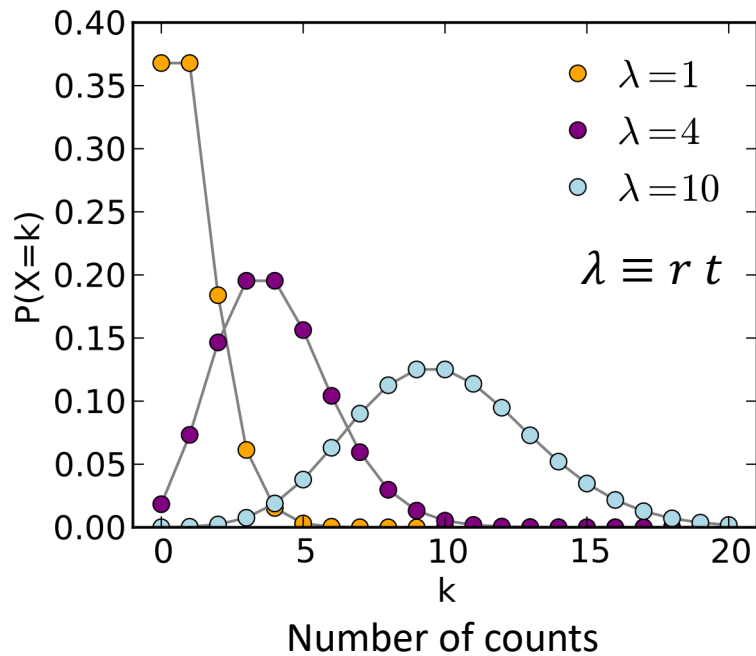
- **Discrete**: something decays or it doesn't ($n=0,1,2,\dots$)
- **No memory**: Probability for a decay is the same in any time interval, no matter how early or late
- **Universal probability**: the probability to decay in a given time interval is the same for all nuclei
- **Independence**: The decay of one nucleus does not affect the probability of a second nucleus decaying

Random Errors: Poisson Distribution

$$P_n(t) = \frac{(r t)^n}{n!} e^{-rt}$$

Probability to observe n events ($n=0, 1, 2, \dots$) in a time interval of length t .

Sole parameter is rate r (average events/second)



Properties of the Poisson distribution

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \text{Probabilities sum to 1}$$

$$\langle n \rangle = \sum_{n=0}^{\infty} n P_n(t) = r t \quad \text{Mean}$$

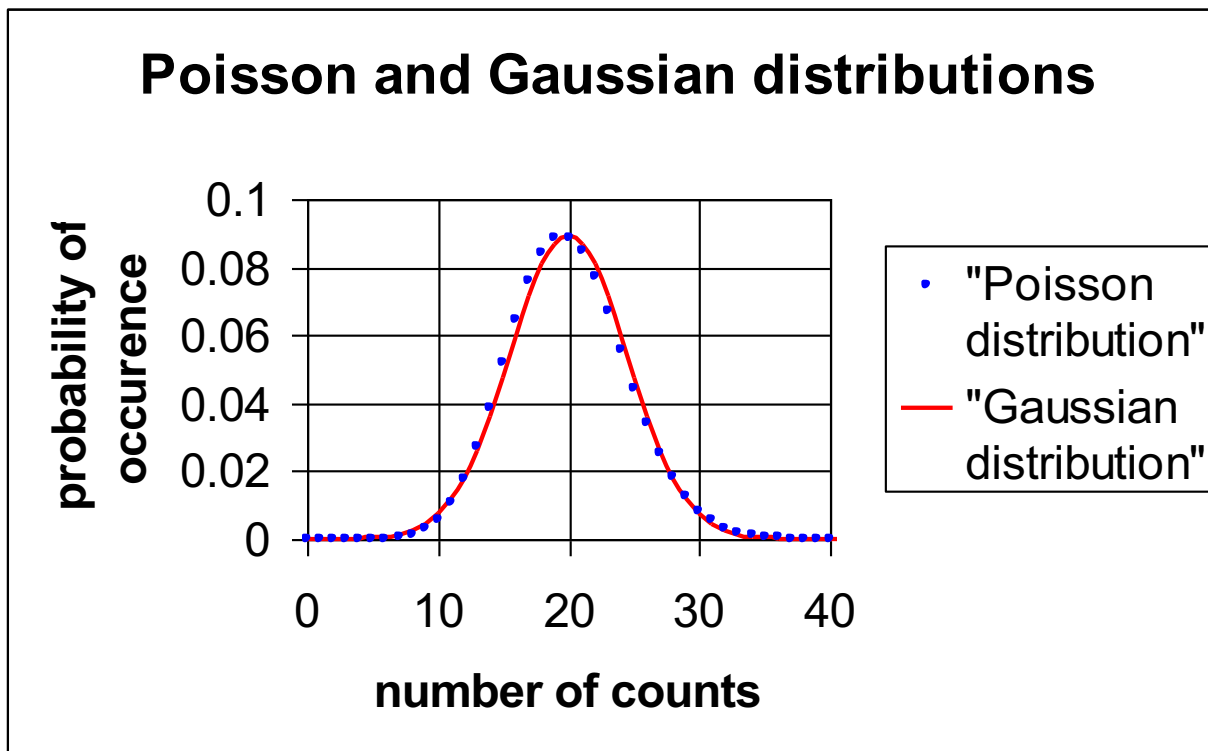
$$\sigma^2 = \sum_{n=0}^{\infty} (n - \langle n \rangle)^2 P_n(t) = r t \quad \text{Variance}$$

Poisson Distribution at large rt

$$P_n(t) = \frac{(rt)^n}{n!} e^{-rt}$$



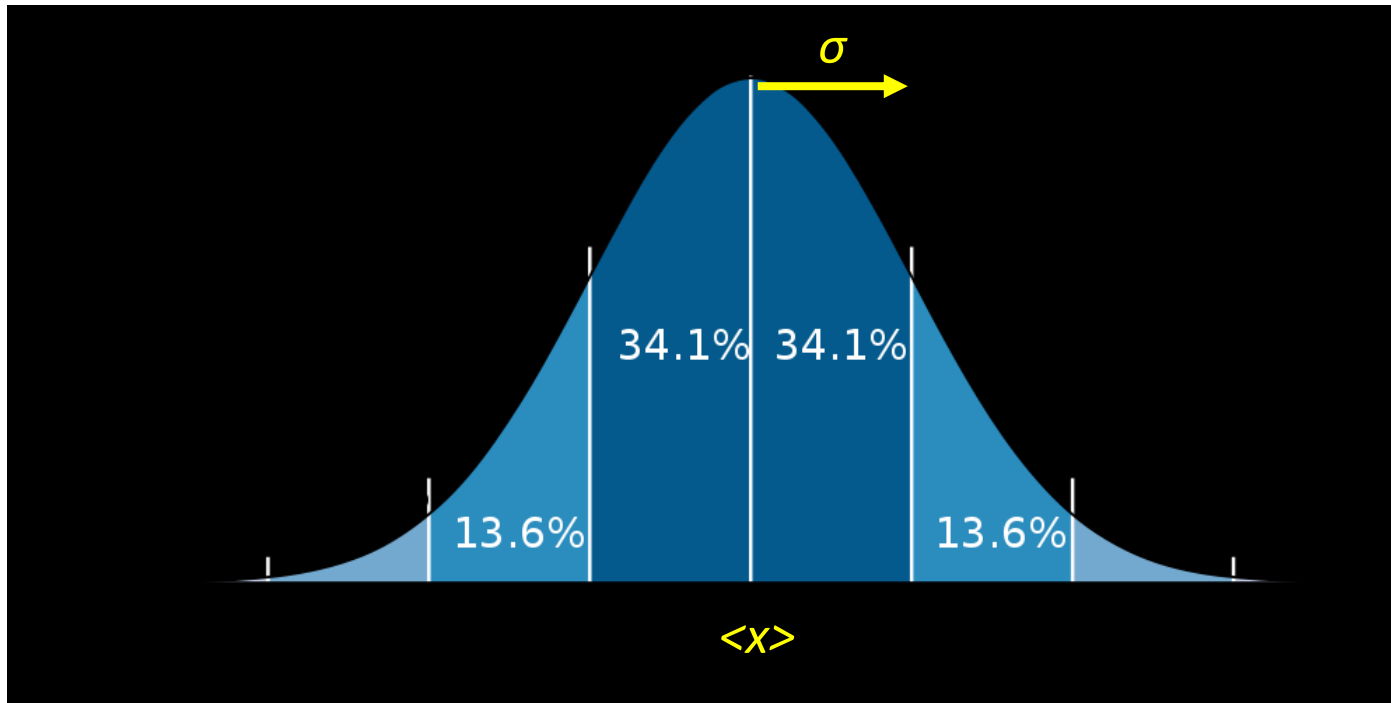
**Carl Friedrich Gauss
(1777–1855)**



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

Gaussian (continuous) distribution

Gaussian (Normal) Distribution



Rules of thumb

- $\pm 1\sigma$: 68.3%
- $\pm 2\sigma$: 95.5%
- $\pm 3\sigma$: 99.7%

- $\pm 4\sigma$: 1 in 16k
- $\pm 5\sigma$: 1 in 1.75M
- $\pm 6\sigma$: 1 in 500M

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

If I measure N events, the “best” estimate of the mean is their arithmetic mean.

The error on this estimate is $\frac{\sigma}{\sqrt{N}}$

Measurements in the Presence of Noise

Source of noisy signal

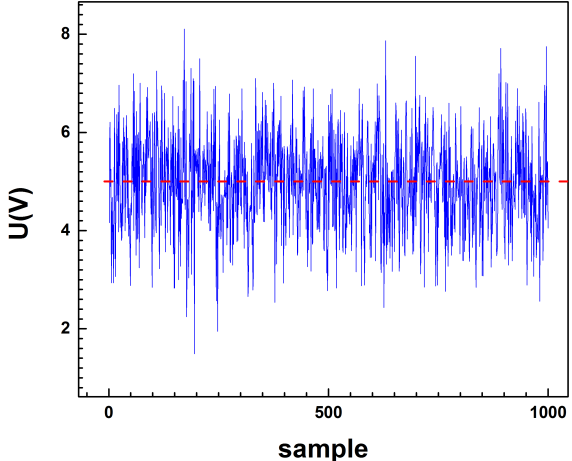


Expected value 5V

- 4.89855
- 5.25111
- 2.93382
- 4.31753
- 4.67903
- 3.52626
- 4.12001
- 2.93411

Noise is often (*not always!*) coupling between your test setup and some thermal bath

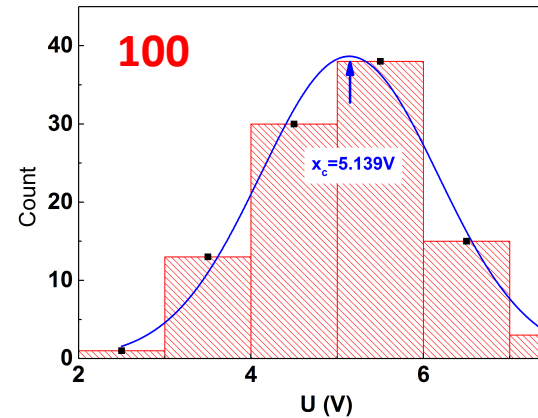
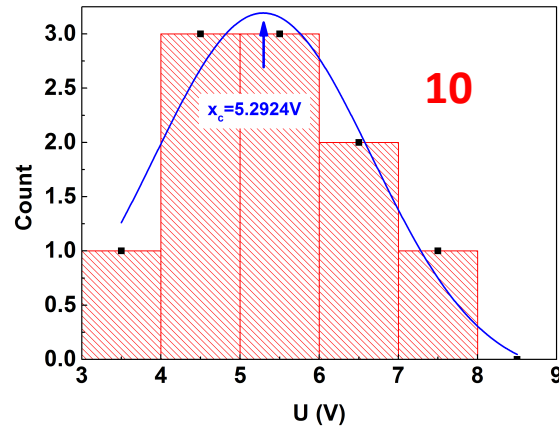
Hence the use of *cryogenic detectors!*



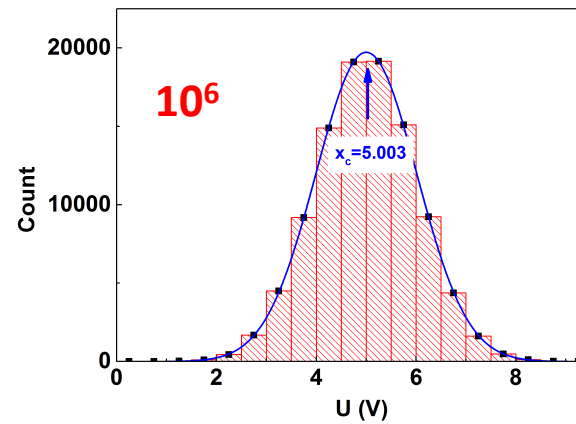
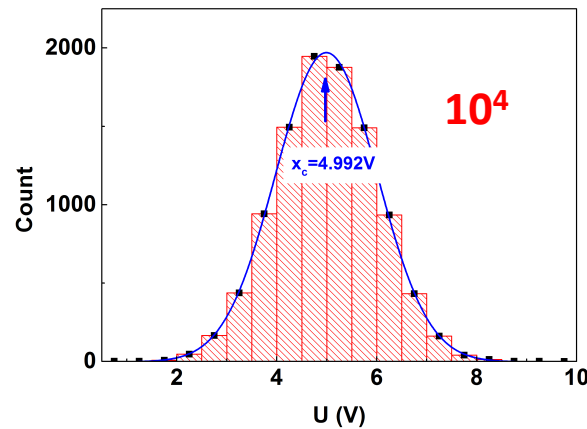
Actual measured values



Measurements in the Presence of Noise



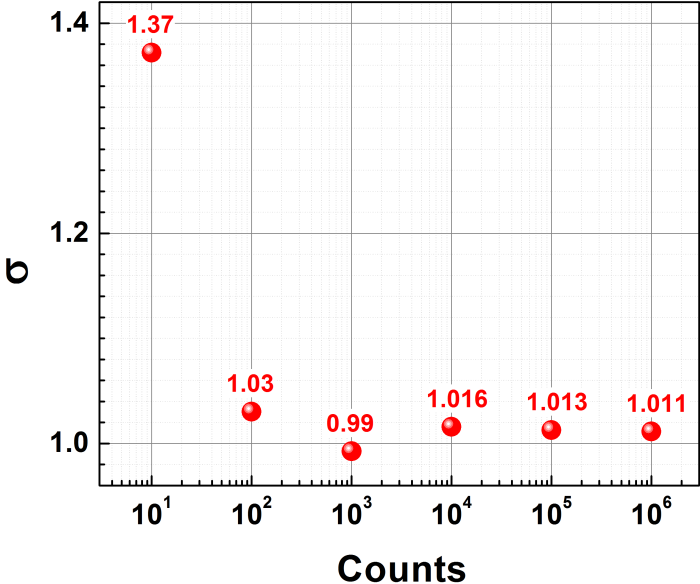
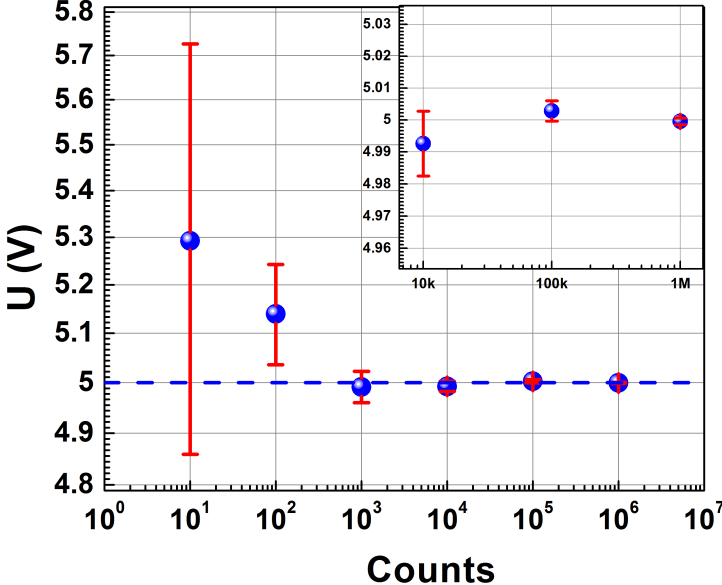
Note that the distribution isn't getting *narrower* at large N .



Instead, you'd have to be more and more unlucky to have enough fluctuations in one direction to move the arithmetic mean significantly



Measurements in the Presence of Noise



Result



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

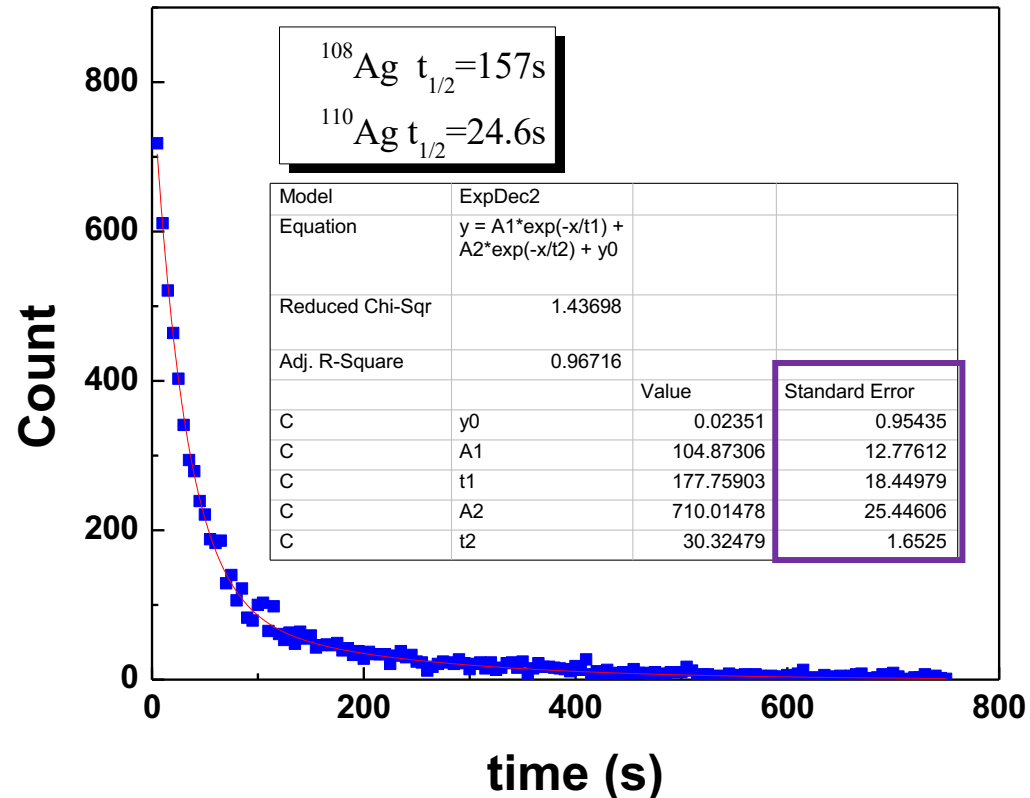
σ - standard deviation
 N - number of samples

For $N=10^6$ $U=4.999 \pm 0.001$ 0.02% accuracy



Fitting Errors

Beta decay of neutron-activated silver foil



$$y = A_1 e^{-t/t_1} + A_2 e^{-t/t_2} + y_0$$

Fitting a **parameterized model** to data means finding the set of parameter values that minimizes the **residuals** – distances between data points and model curve.

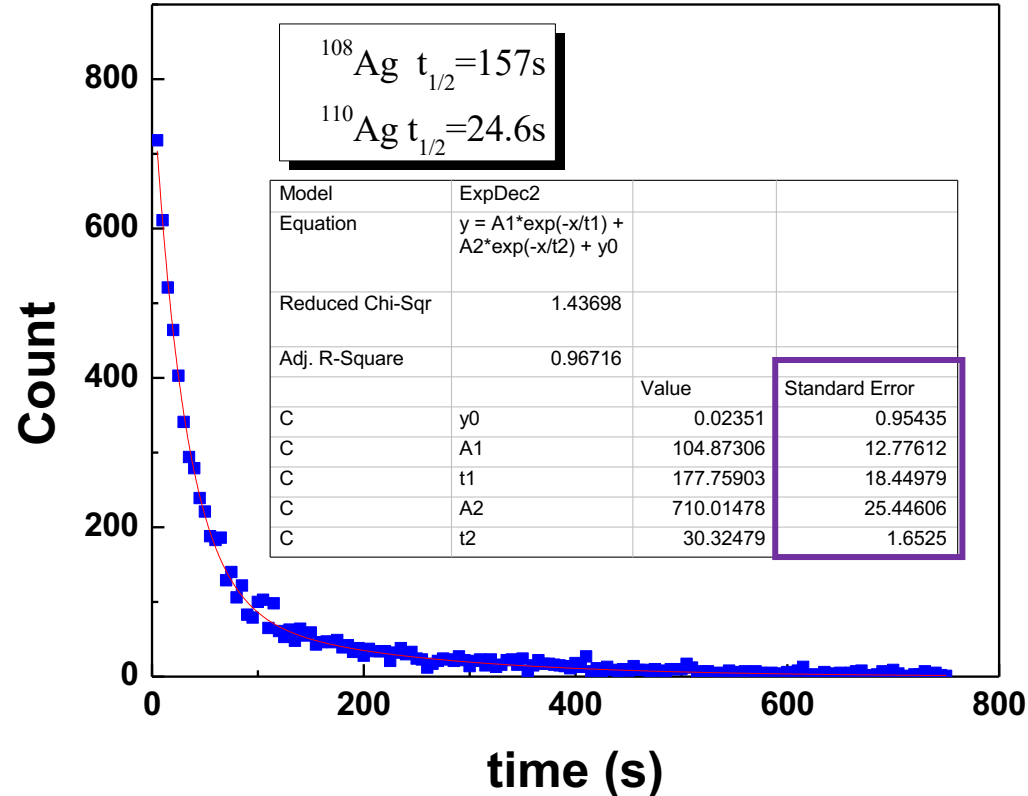
Further details in the appendix slides: minimize what exactly, and how?

Error estimates: how far can we wiggle the parameters before the residuals increase “significantly”?

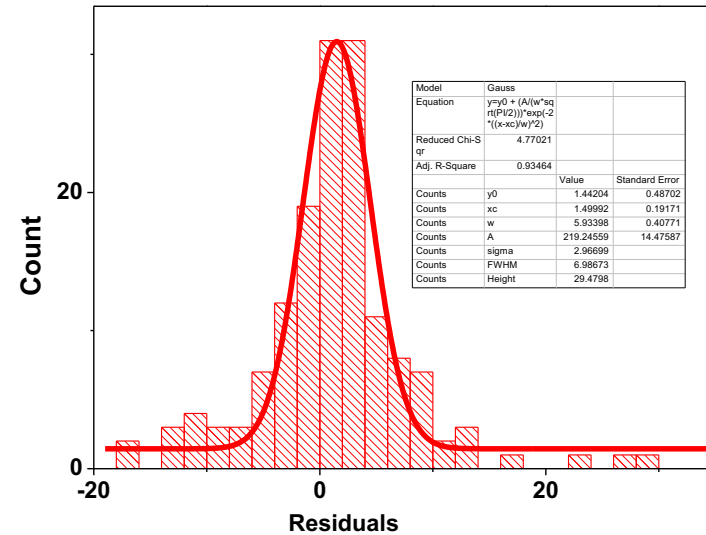
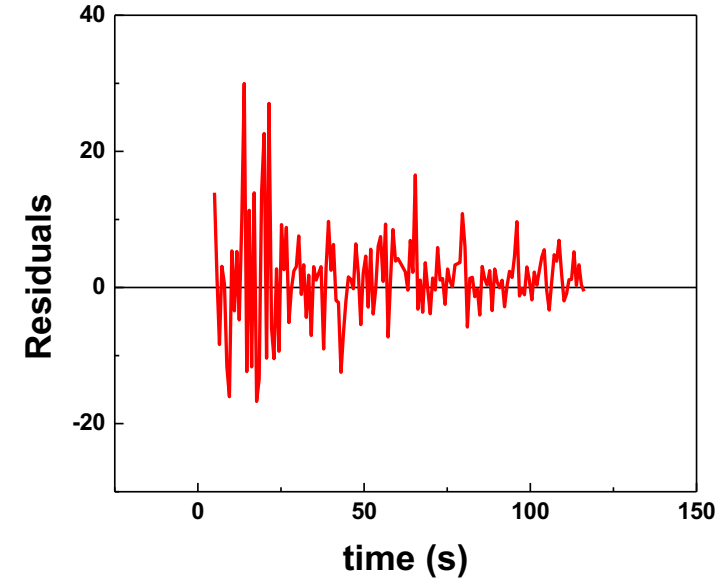
*Error estimates assume **model is correct** and **residuals are gaussian noise**. Errors not trustworthy if these aren't roughly true!*

Fitting Errors

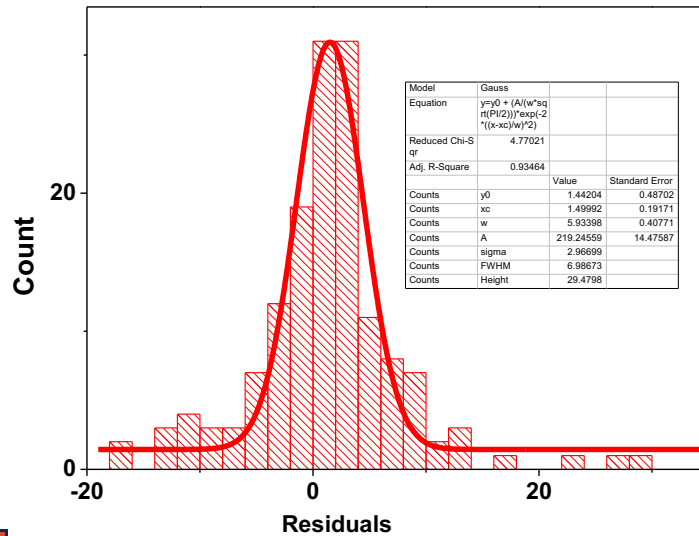
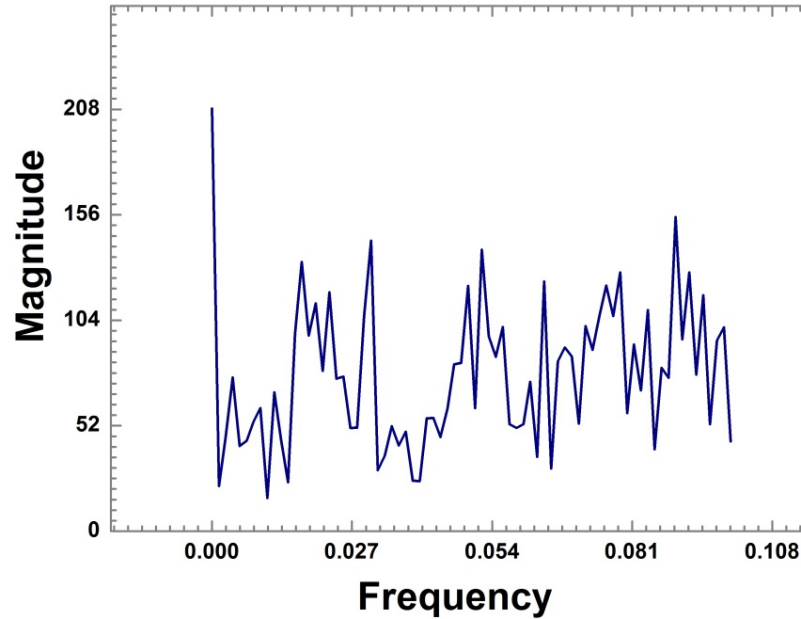
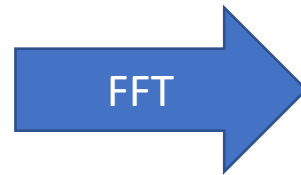
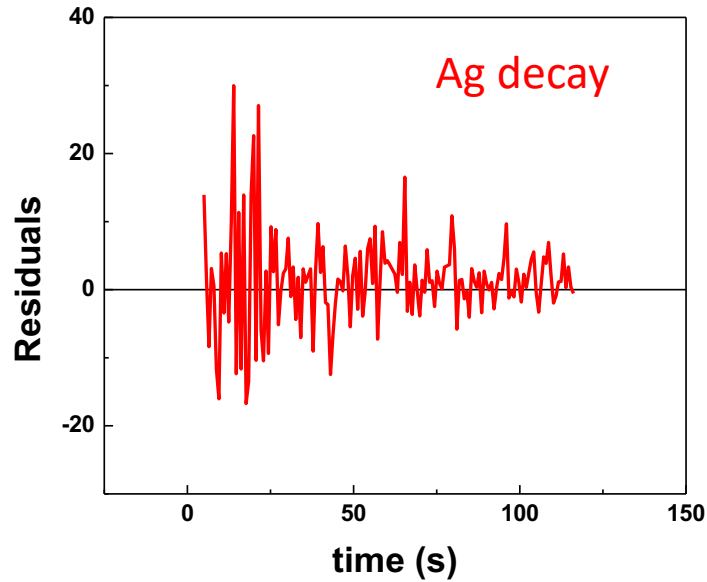
Beta decay of neutron-activated silver foil



$$y = A_1 e^{-t/t_1} + A_2 e^{-t/t_2} + y_0$$



Fitting Errors: Examining the Residuals

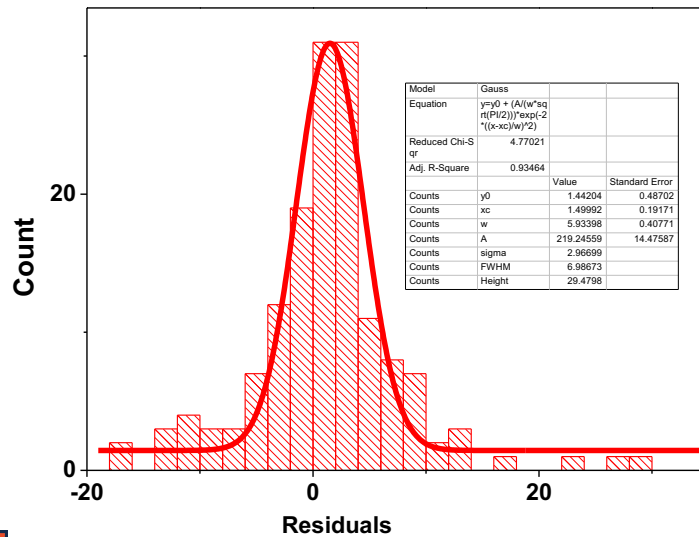
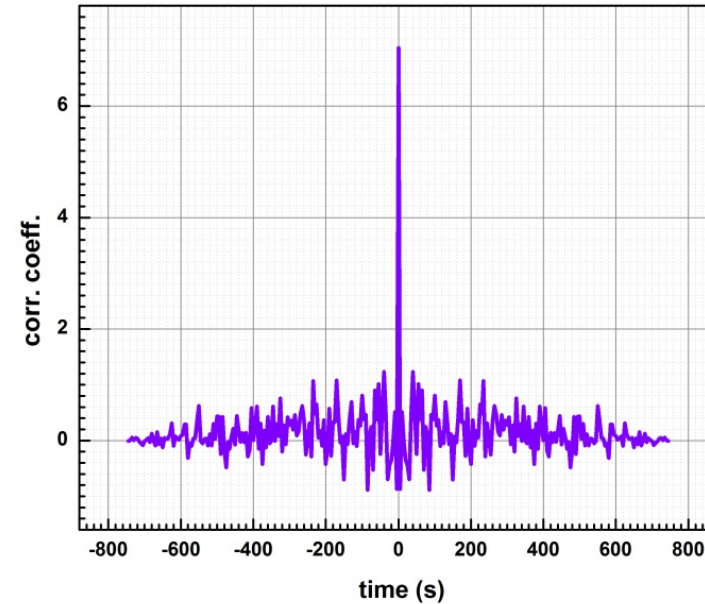
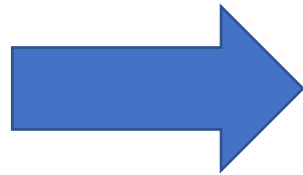
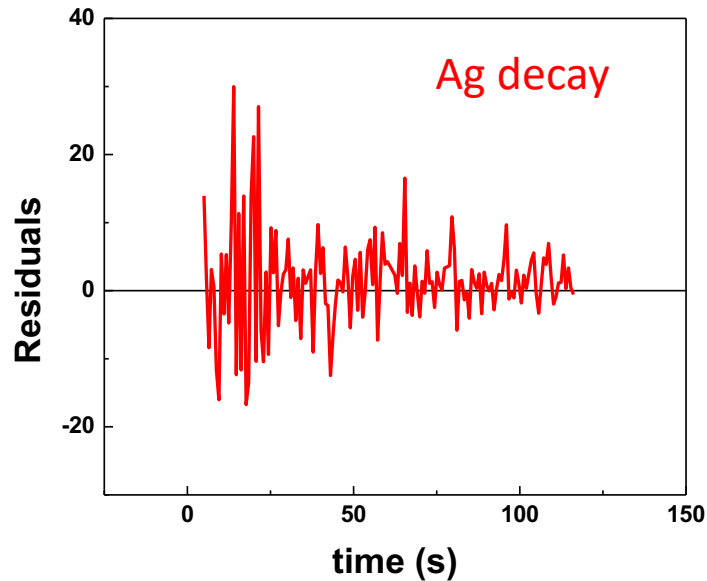


Test #2: No pronounced frequencies visible

Test #1: generally Gaussian/symmetric residuals



Fitting Errors: Examining the Residuals



Test #3: Autocorrelation shows no prominent correlation time scales

Correlation

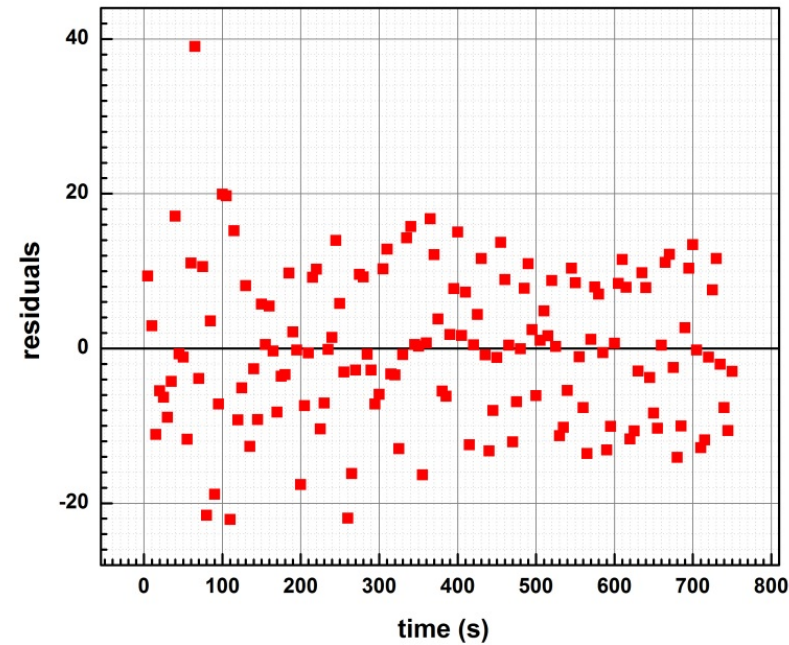
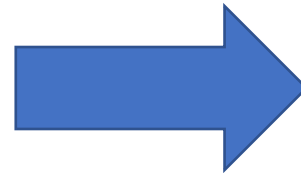
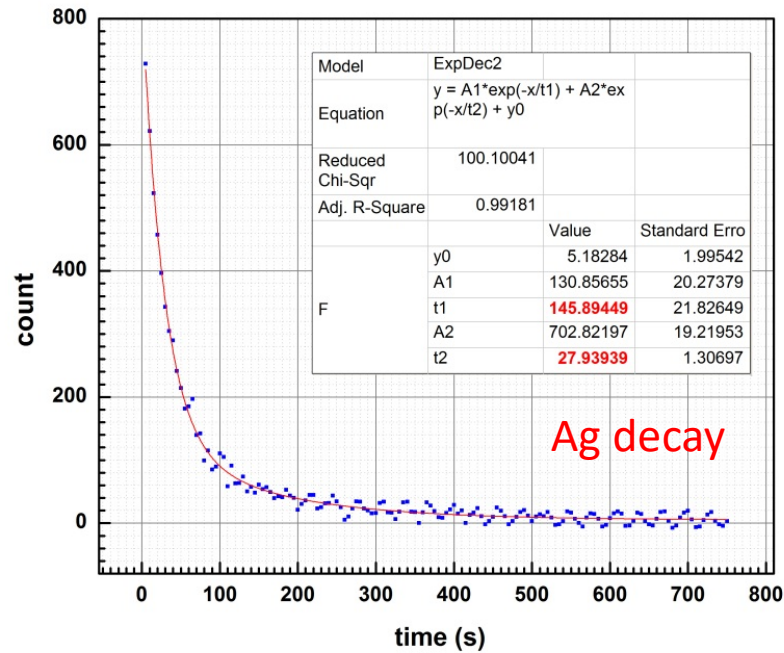
$$y_{fg}(m) = \sum_{n=0}^{N-1} f(n)g(n-m)$$

Auto-correlation

$$y_{ff}(m) = \sum_{n=0}^{N-1} f(n)f(n-m)$$

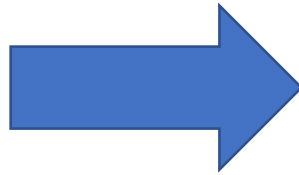
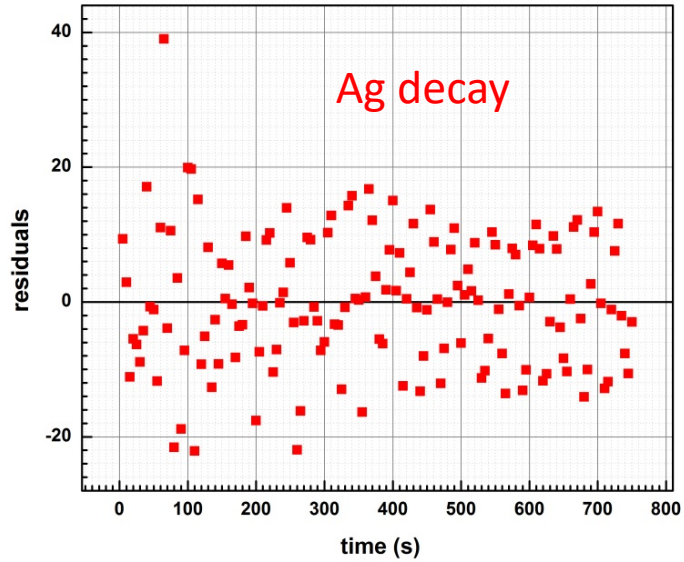


Examining Non-Ideal Fit Residuals

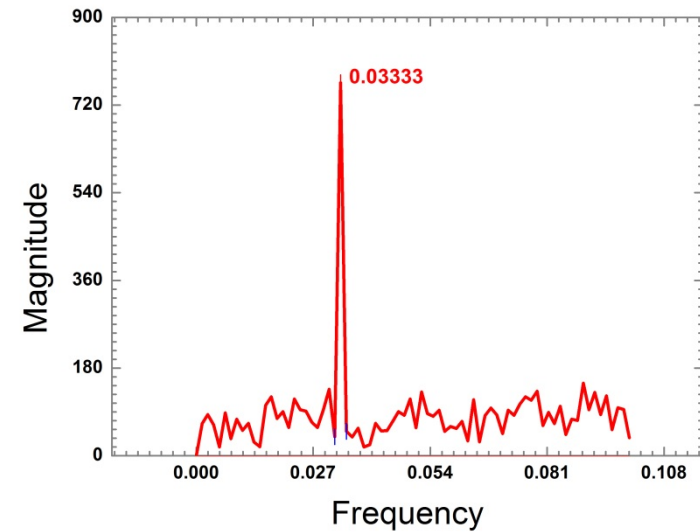
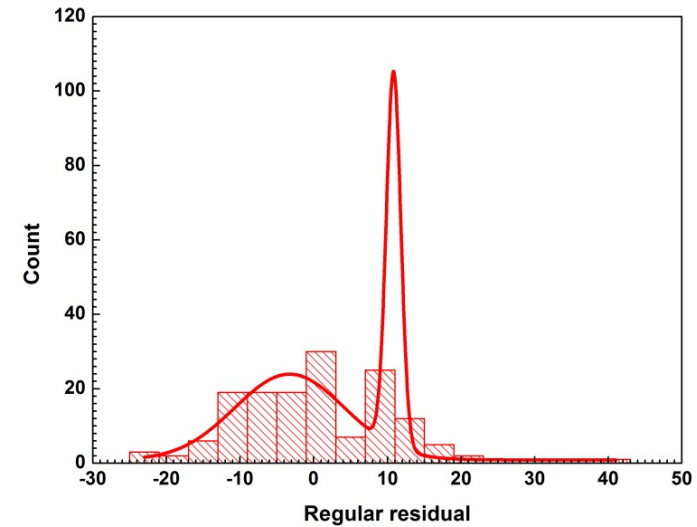


	Clean experiment	Data + "noise"
$t_1(s)$	177.76	145.89
$t_2(s)$	30.32	27.94

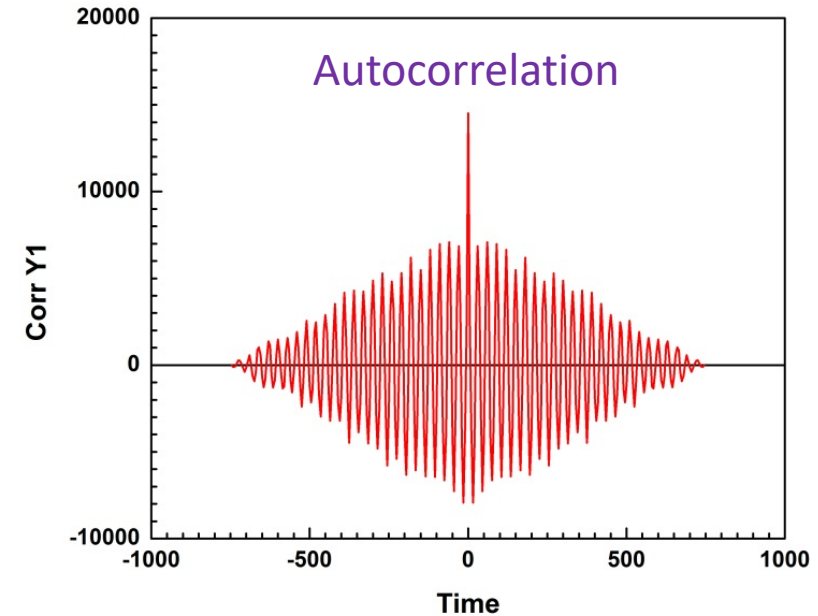
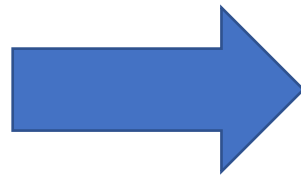
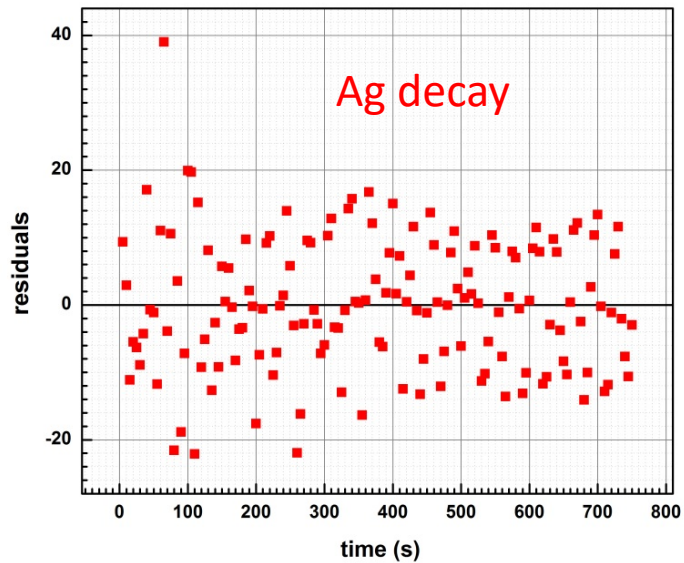
Examining Non-Ideal Fit Residuals



1. Histogram does *not* resemble a normal distribution
2. Prominent component at frequency 0.03333



Examining Non-Ideal Fit Residuals

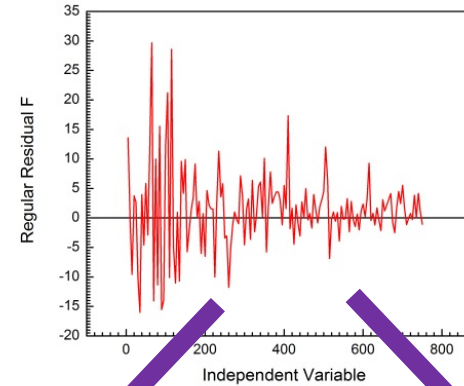
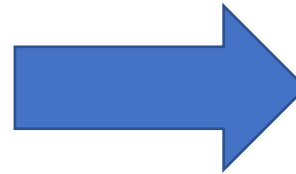
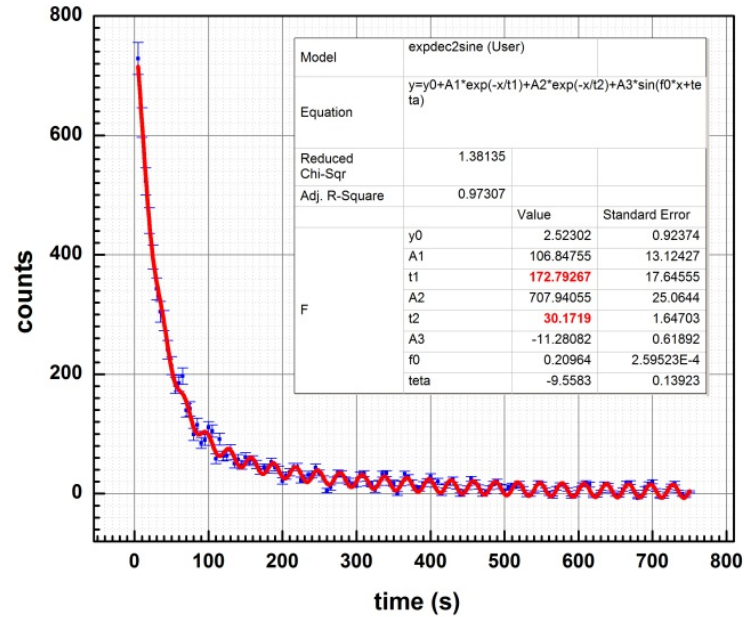


Conclusion: Fitting function needs to be modified to incorporate an additional term

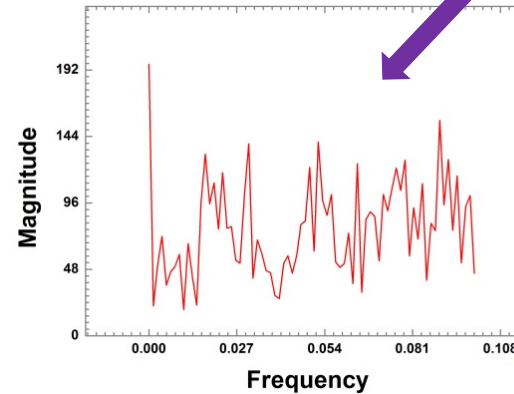
$$y(t) = y_0 + A_0 e^{-t/t_1} + A_1 e^{-t/t_2} + A_3 \sin(\omega t + \theta)$$

Caution: “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk”
John von Neumann, attributed by Enrico Fermi

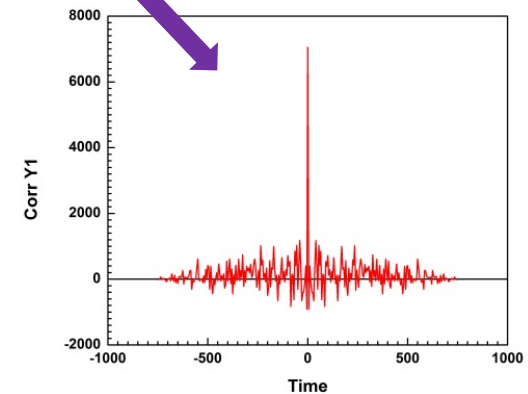
Examining Non-Ideal Fit Residuals



FFT



Autocorrelation



	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

We expect both error components in each measurement

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

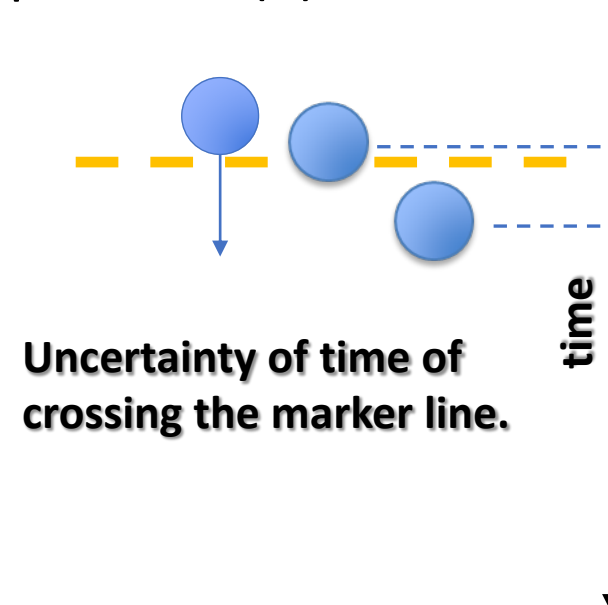
... and there are various sources for each

e_s

Systematic error could include uncertainties in plate separation (d), applied DC voltage (V), ambient temperature (T), etc.

e_r

Random errors include uncertainties in measurement of t_{rise} , t_g , etc.



Systematic Error Analysis: Millikan Oil Drop Experiment

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

$$Q = ne = \mathbf{F} \mathbf{S} \mathbf{T} = \frac{\mathbf{1}}{f_c^{2/3}} \frac{9\pi d}{V} \sqrt{\frac{2\eta^3 x^3}{g\rho}} \sqrt{\frac{\mathbf{1}}{t_g} \left[\frac{\mathbf{1}}{t_g} + \frac{\mathbf{1}}{t_{rise}} \right]}$$

$$\frac{1}{f_c^{2/3}} \approx 1 - \left(\frac{t_g}{\tau_g}\right)^{\frac{1}{2}}$$

$$\Delta Q = \sqrt{\left(\frac{dQ}{dF}\right)^2 (\Delta F)^2 + \left(\frac{dQ}{dS}\right)^2 (\Delta S)^2 + \left(\frac{dQ}{dT}\right)^2 (\Delta T)^2} \approx \sqrt{\left(\frac{dQ}{dS}\right)^2 (\Delta S)^2 + \left(\frac{dQ}{dT}\right)^2 (\Delta T)^2}$$

Generally negligible

$$= \sqrt{(FT)^2 (\Delta S)^2 + (FS)^2 (\Delta T)^2} = Q \sqrt{\left(\frac{\Delta S}{S}\right)^2 + \left(\frac{\Delta T}{T}\right)^2}$$

Systematic Error Analysis: Millikan Oil Drop Experiment

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

$$Q = ne = FST = \frac{1}{f_c^{2/3}} \frac{9\pi d}{V} \sqrt{\frac{2\eta^3 x^3}{g\rho}} \sqrt{\frac{1}{t_g} \left[\frac{1}{t_g} + \frac{1}{t_{rise}} \right]}$$

$$\Delta Q \approx Q \sqrt{\left(\frac{\Delta S}{S}\right)^2 + \left(\frac{\Delta T}{T}\right)^2}$$







$$\frac{\Delta S}{S} = \sqrt{\left(\frac{\Delta d}{d}\right)^2 + \left(\frac{\Delta V}{V}\right)^2 + \left(\frac{3}{2} \frac{\Delta x}{x}\right)^2 + \left(\frac{3}{2} \frac{\Delta \eta}{\eta}\right)^2 + \left(\frac{1}{2} \frac{\Delta \rho}{\rho}\right)^2 + \left(\frac{1}{2} \frac{\Delta g}{g}\right)^2} \approx \sqrt{\left(\frac{\Delta d}{d}\right)^2 + \left(\frac{3}{2} \frac{\Delta x}{x}\right)^2}$$

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

Appendix #1: Analyzing the Charge Data

Step 1. Origin Project For Raw Data :

\\engr-file-03\PHYINST\APL Courses\PHYCS401\Students\2. Millikan Raw Data

 Data Analysis for Millikan Oil Drop Experi...	2/22/2008 9:36 AM	Adobe Acrobat D...	59 KB
 Millikan_raw data.opj	10/5/2017 4:50 PM	OPJ File	15 KB
 Millikan_raw data1.opj	9/25/2018 1:38 PM	OPJ File	14 KB
 Millikan1_calc.opj	9/26/2018 1:36 PM	OPJ File	95 KB
 Millikan1_no_calc.opj	9/25/2017 2:03 PM	OPJ File	66 KB
 T measurement.opj	9/26/2018 1:36 PM	OPJ File	316 KB

All project files with raw data should be stored in:

\\engr-file-03\PHYINST\APL
Courses\PHYCS401\Students\2. Millikan Raw Data

Only files with *raw data* should be stored in this *common* folder. Please remove files that are unrelated to the experimental results!

All *other files*, including those used for calculations, should be saved in your personal folder

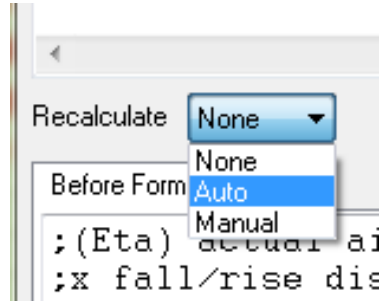
Appendix #1: Analyzing the Charge Data

Step 2. Working on your personal Origin project

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

	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	$r_c[m] = \frac{6.18 \times 10^{-5}}{\rho[mmHg]}$	$\tau_g = \frac{2\eta x}{\rho g r_c^2}$	$F = \frac{1}{f_c^{3/2}} \approx 1$
1	Viscosity of air(kg/ms) (25oC)	η	1.8478E-5	7.455	7.91327	--	--	--
2	Temperature coefficient of viscosity	$\Delta\eta/\Delta T$	4.8E-8	15.56521	16.7815	--	--	--
3	Density of oil (kg/m^3)	ρ_1	886	23.07825	31.8955	--	--	--
4	Density of air (kg/m^3)	ρ_2	1.29	20.14243	11.70129	--	--	--
5	Density difference (kg/m^3)	$\rho_1 - \rho_2$	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		1.8478E-5	--	--	--	--	--

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

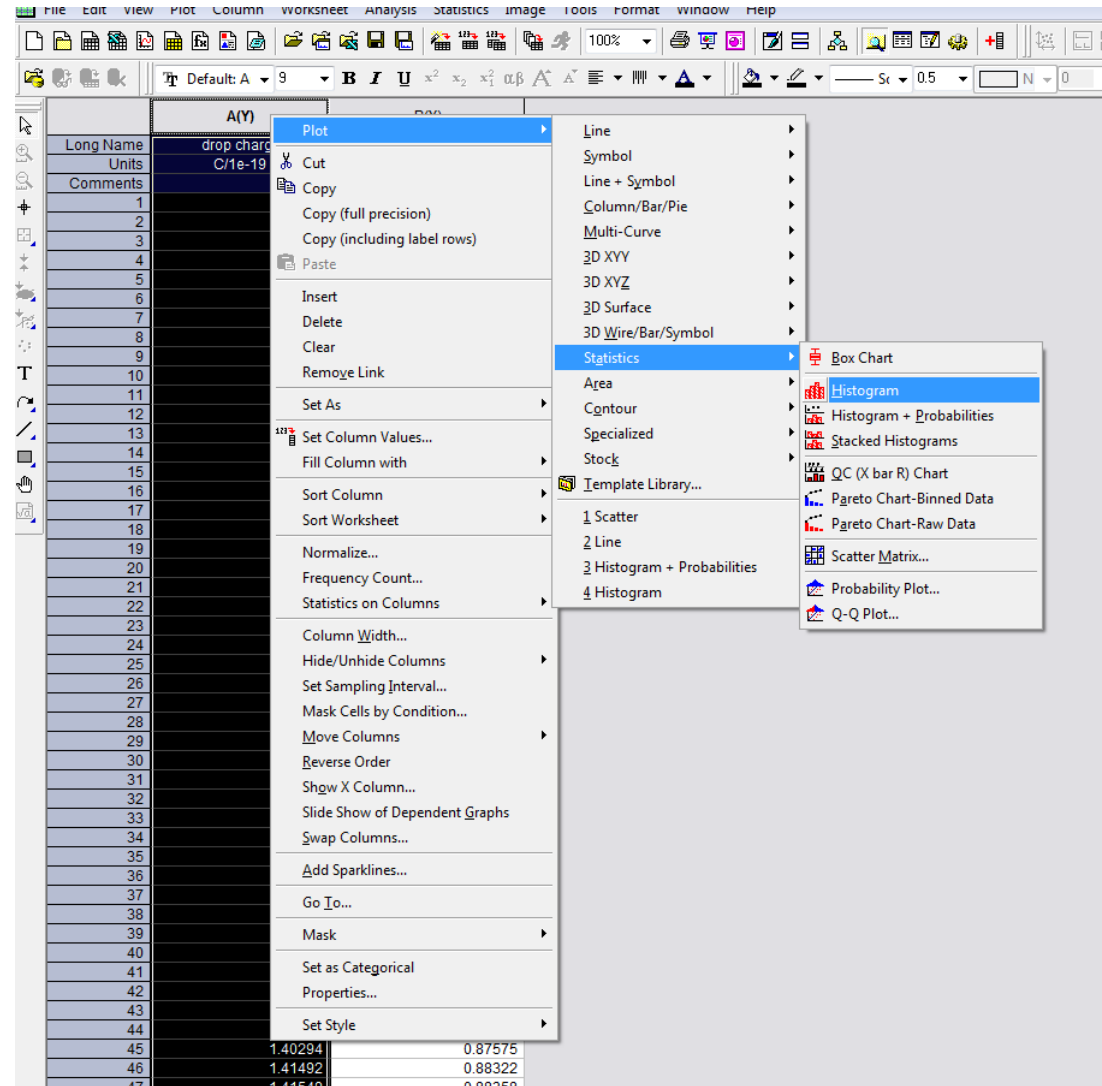
Manually calculate the actual air viscosity



Appendix #1: Analyzing the Charge Data

Step 3. Make a histogram

Plot a histogram from the data in the column of drop charges



Appendix #1: Analyzing the Charge Data

Step 4. Histogram – set bin size

Origin will choose the bin size h *automatically*, but not *optimally*. In the plot at right, $h=0.5$.

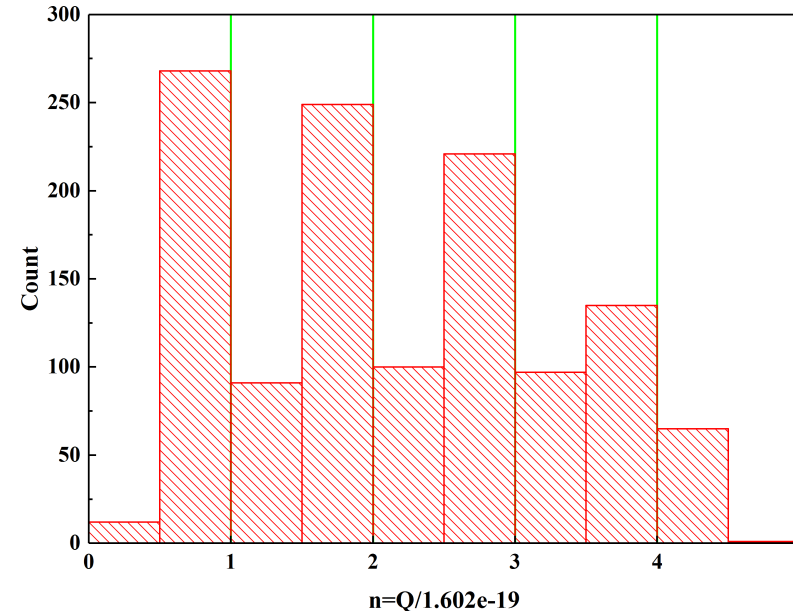
We want bins *wide* enough to control bin-to-bin fluctuations that obscure the pattern, but *narrow* enough not to wash out the pattern.

There are several theoretical approaches to finding the optimal bin size. One example:

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

... where σ is the sample standard deviation, n is the total number of observations.

For the data at right, this yields $h \sim 0.1$

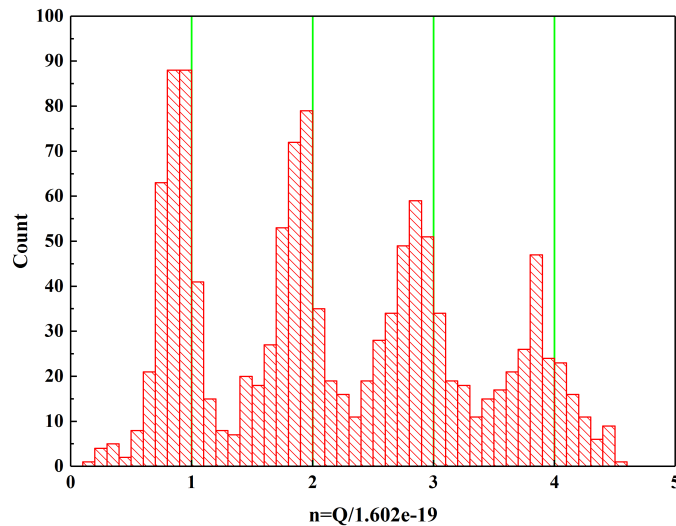


Bin size above is 0.5
Peaks vaguely visible, but can't
determine width

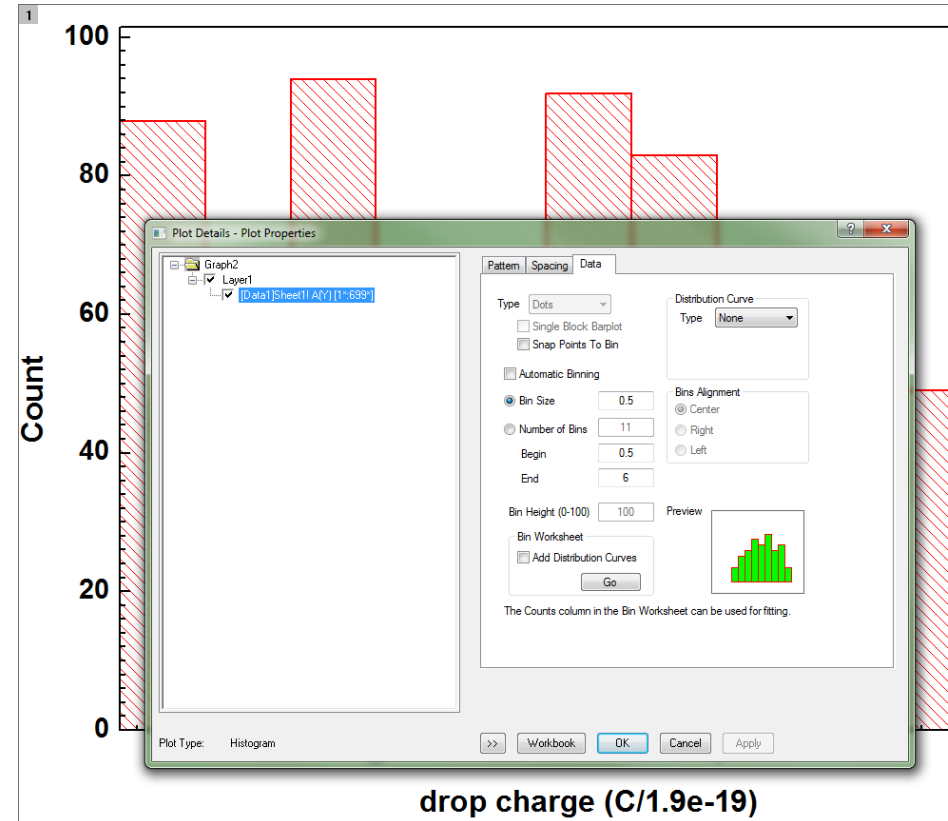
Appendix #1: Analyzing the Charge Data

Step 4. Histogram – set bin size

To change the bin size, click on the graph the uncheck the “*Automatic Binning*” option



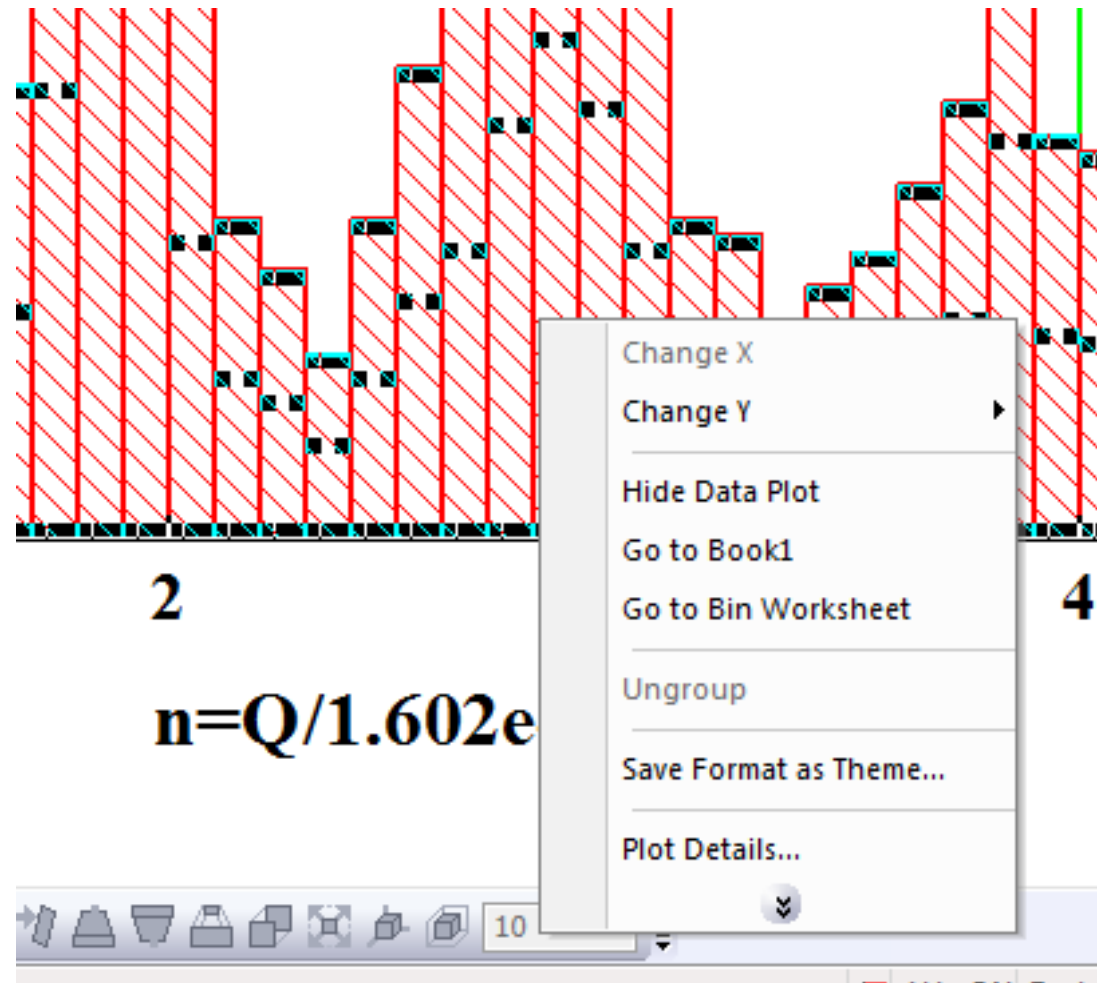
Bin size is now 0.1, pattern is visible!



Appendix #1: Analyzing the Charge Data

Step 4. Histogram - find the bin worksheet

Right-click on the histogram and choose
“Go to bin worksheet”



Appendix #1: Analyzing the Charge Data

Step 5. Add counts-vs-bin plot

Plot Setup: Configure Data Plots in Layer

Available Data:

Long Name	Sheet	Cols	Rows	File Name	File Date	Created	Modified
Book1	Sheet1	1	1260			10/6/2008 12:43:53	10/2/2017 14:...
Book1	Book1_B Bins	4	60			10/6/2008 12:43:53	10/2/2017 14:...

Plot Type:

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

Show(S) [Book1]"Book1_B Bins"

X	xEr	Y	yEr	L	Column	Long Name	Comments	Sampling Interval
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<autoX>	From/Step=		
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	A	Bin Centers	Bins	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B	Counts	Bins	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C	Cumulative Sum	Bins	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	D	Cumulative Percent	Bins	

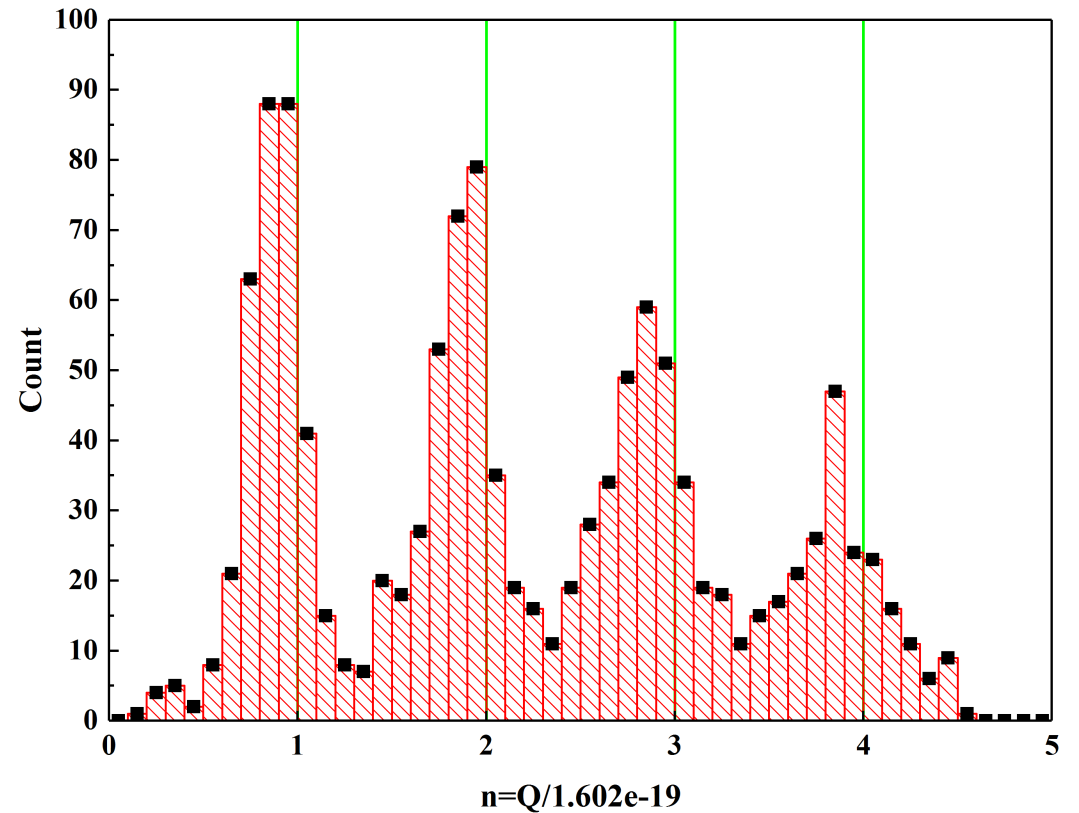
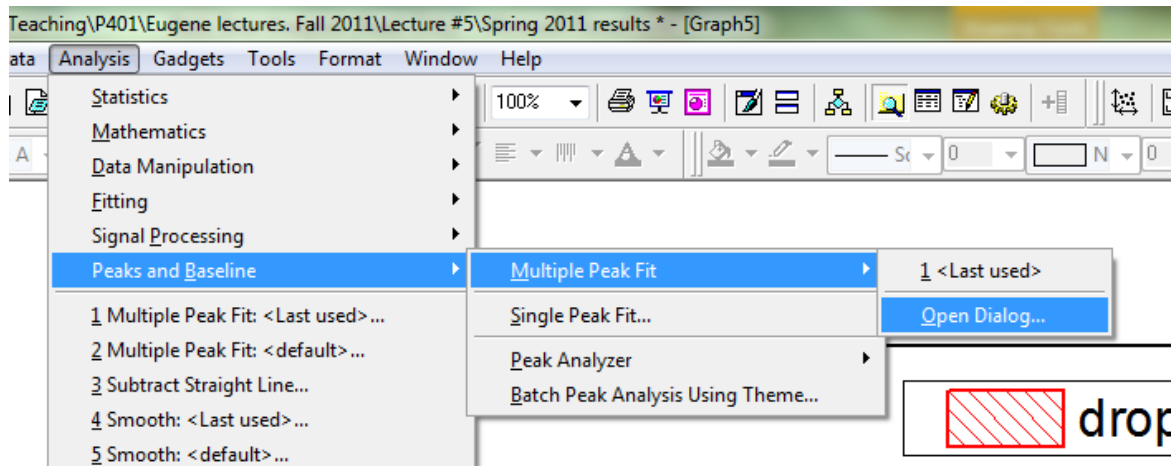
Plot List: Drag entries in 1st column to reorder or to move between layers. Right click for other options.

Plot	Range	Show	Plot Type	Legend
Layer 1		<input checked="" type="checkbox"/> Rescale		
[Book1]Sheet1! B(Y)	[1*:1241*]	<input checked="" type="checkbox"/>	Histogram	B
[Book1]Book1_B Bins! "Bin Centers"(X), "Counts"(Y)	[1*:50*] 0.05 < X < 4.95, 0 < Y < 88	<input checked="" type="checkbox"/>	Scatter	Bins



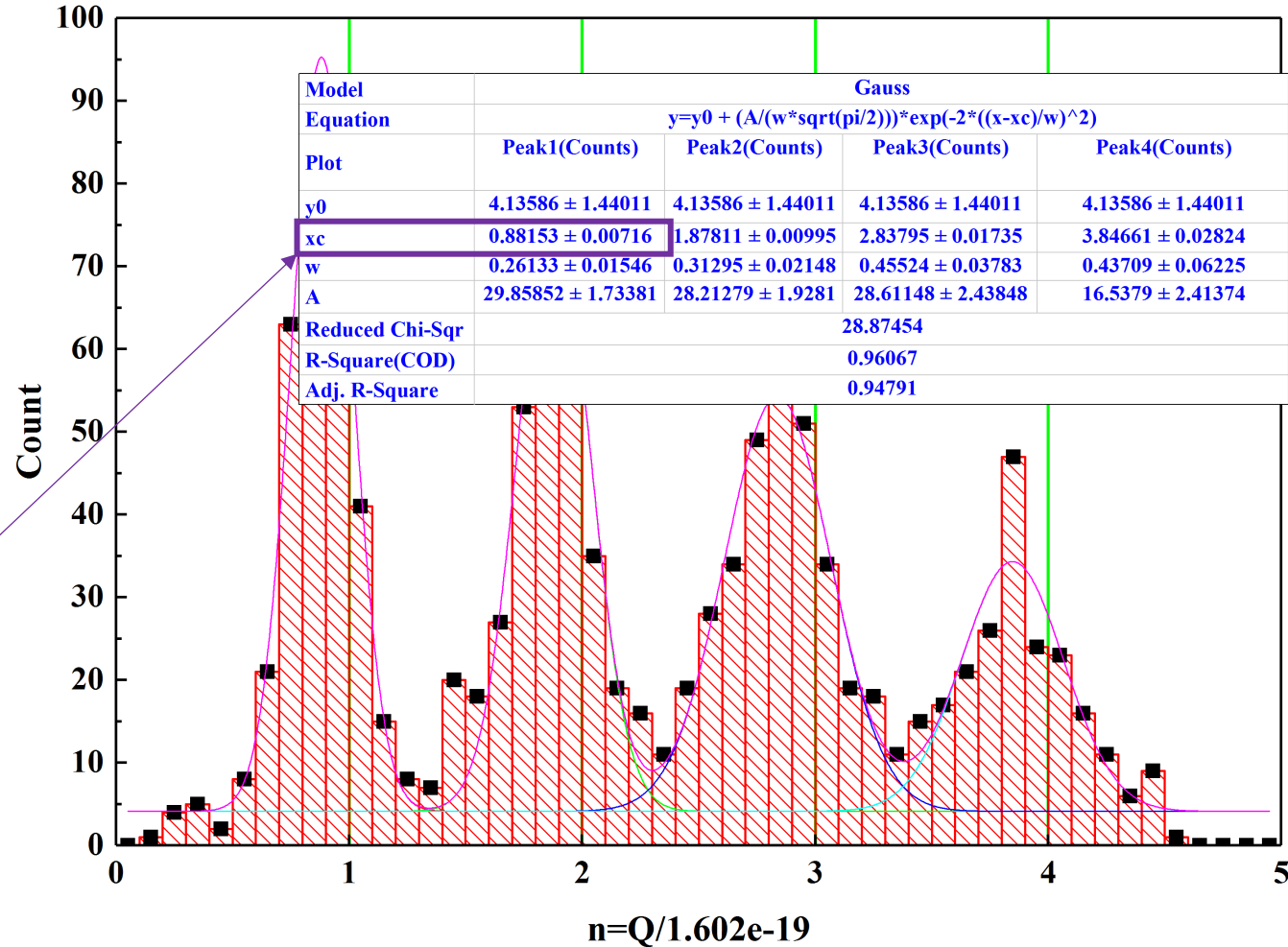
Appendix #1: Analyzing the Charge Data

Step 5. Multi-peak gaussian fit



Appendix #1: Analyzing the Charge Data

Step 5. Multi-peak gaussian fit

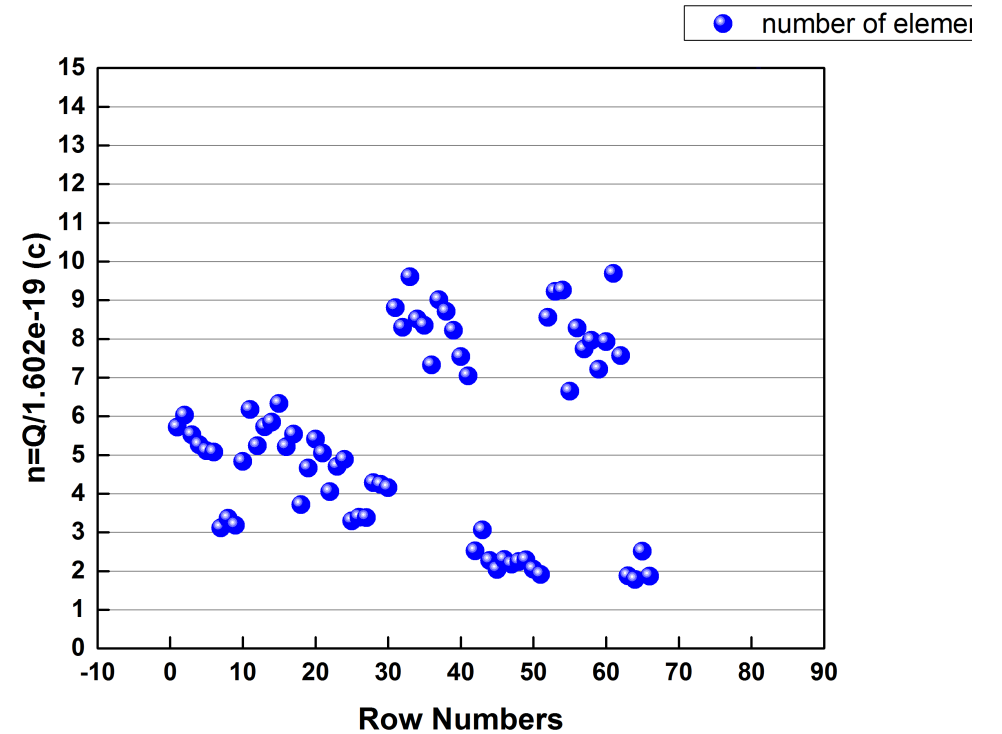
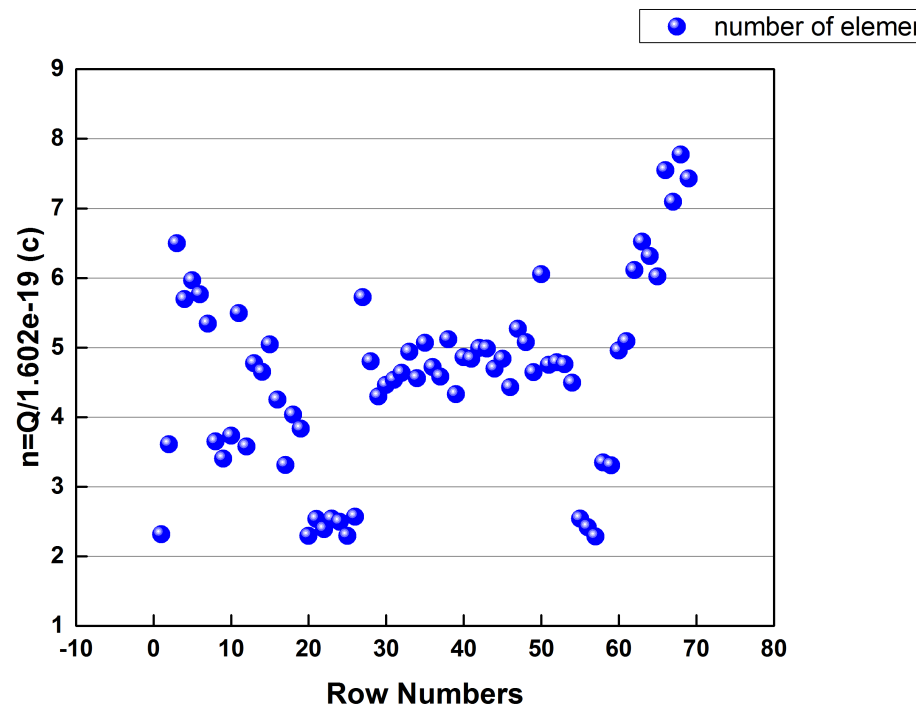


First peak: mean 0.882 ± 0.007



Appendix #1: Issues with Oil Drop Data

Be careful with data selections obtained by different teams!



For more details how to create the histogram plot and do the analysis, see [“Working with Histogram Graph. Millikan Oil Drop Experiment” \(E.V. Colla\)](#)

Appendix #1: Increasing Experimental Accuracy

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

e_s **Systematic error** could be improved with more precise knowledge of the experimental parameters: plate separation (d), applied DC voltage (V), ambient temperature (T), etc.
Usually these are limited by existing measurement equipment

e_r **Random or statistical errors** can be reduced only by increasing the number of data points (*no limit!*)

Appendix #2: Fitting Models to Data

Fitting a model to data means finding the set of *parameter values* that gets the *model function* “closest” to the data

- **Data:** ordered pairs (x_i, y_i) , often in the form of an $N \times 2$ matrix
 - Independent variable x_i , e.g. frequency, time, etc.
 - Dependent variable y_i , e.g. signal magnitude
- **Model function:** $y = f(x; \beta)$, which takes some set of parameters β

What exactly does “closest” mean? Minimizing some *cost function* – most often the sum of squared deviations:

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

Appendix #2: Fitting Models to Data

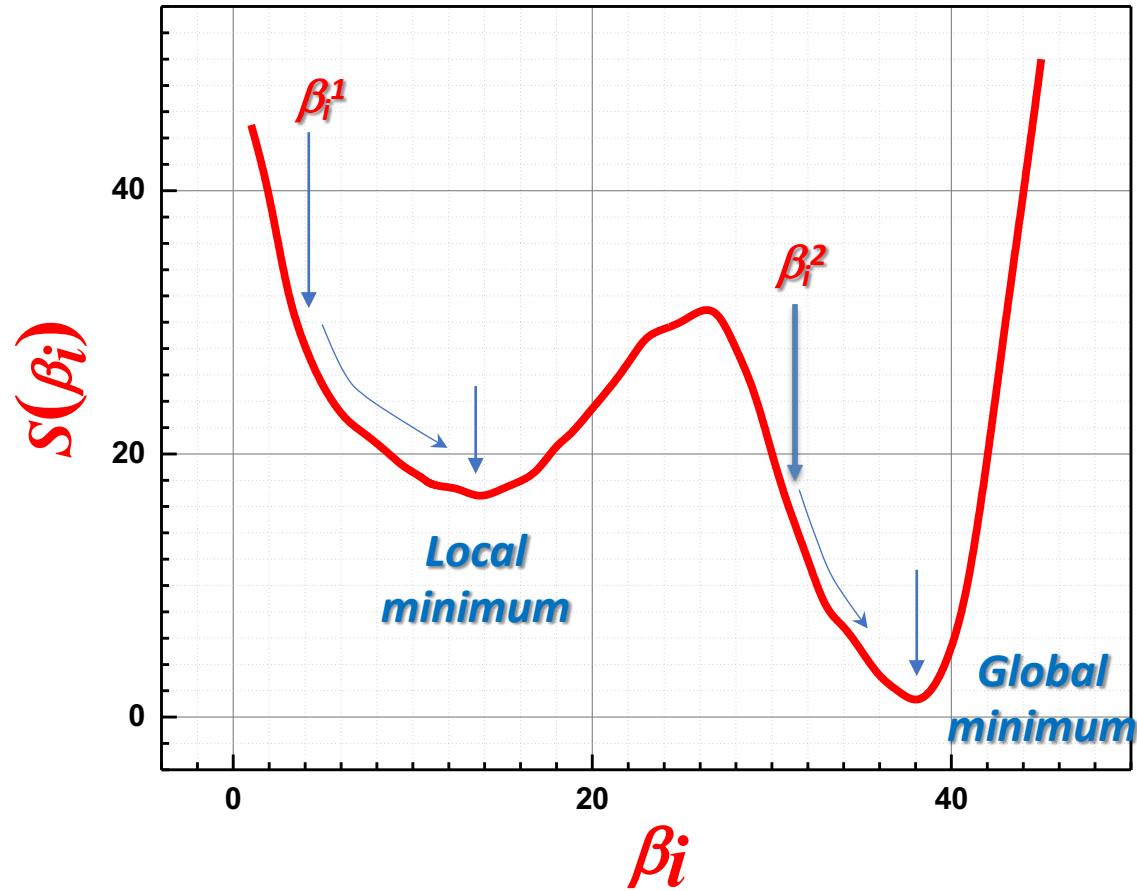
$$S(\beta) = \sum_{i=1}^N [f(x_i; \beta) - y_i]^2$$

There are several minimization algorithms for solving such problems. Origins uses the [Levenberg-Marquardt algorithm](#) for nonlinear fitting, which is optimized for quadratic cost functions like this one.

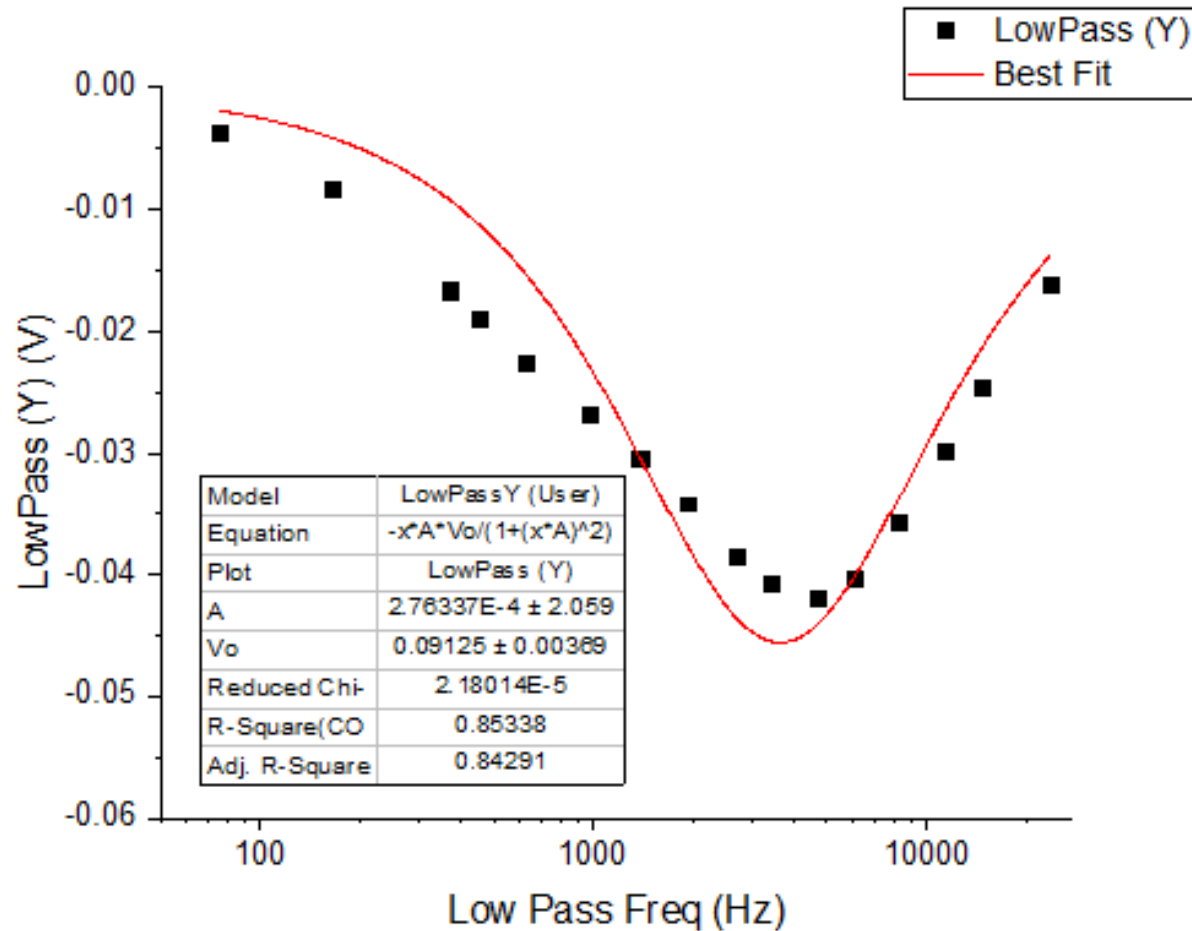
These algorithms proceed from a [starting guess](#) and traverse the cost function “landscape” (hopefully efficiently!) to find a [global minimum](#).

How well they do depends on the quality of the starting guess, and it’s easy to get trapped in [local minima](#).

Appendix #2: Fitting Models to Data



Appendix #2: Fitting Models to Data



An example of poor results from the fitting procedure



Appendix #3: Some Reminders

1. Reports should be uploaded *only* to the **proper folder** for your **activity** and section
 - For example, folder **Frequency domain analysis_L1** should only be used by students from section L1
 - Submit only one copy (no need to submit e.g. both Word and PDF)
 - Show your name, and only your name, in the list of authors on the title page of the report
 - I recommend the following file name style:

L1_lab5_LastName