

## Notes O - Overview and tools

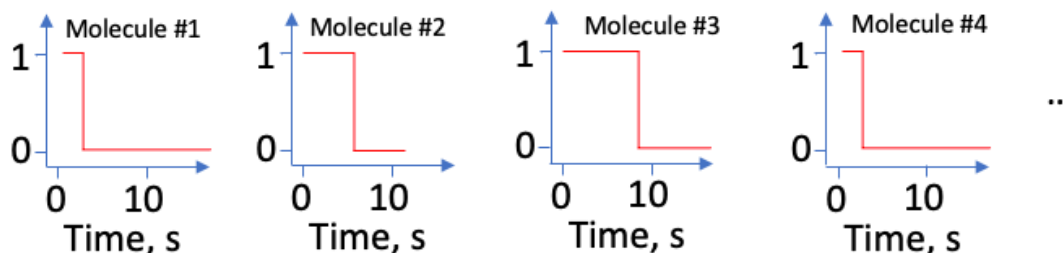
### Chapter 1. The goal of physical chemistry

The goal of physical chemistry is to build models to predict how molecules behave, using the smallest possible number of assumptions while keeping reasonable accuracy.

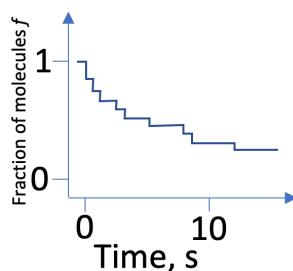
Our goal in this class will be to see how physical chemists derive properties of molecules from fundamental laws of nature, such as ‘ $H\Psi=E\Psi$ ,’ and how one might come up with new and (hopefully!) useful equations based on those ‘laws.’ Unlike a multi-semester class, we will not try to prove every detail and study 500 pages worth of a textbook. Rather, I want you to get a feel for what the most important quantitative concepts in quantum mechanics, statistical mechanics & thermodynamics, and kinetics & transport are, using just a few examples ranging from the simple, like  $H_2$  forming a bond, to the complex, like a protein folding. Therefore, these lecture notes will be all your reading, ca. 6-8 pages per week or 2-3 pages per lecture. It’s a short read each week, but it is important that you do it before class, so you know what is puzzling you in the notes and can ask questions when I lecture in class. Interspersed in the notes will be in-class exercises, thought problems, and homework assignments that are just as important as the notes and our discussion. The exam questions will be open book and similar to these exercises and problems.

To do calculations with models, there is a variety of tricks that we need to use. Most of them are not too hard to use, but they do require practice to build good intuition, just like any skill. Some important examples and ‘tricks of the trade’:

**1. Averaging and average quantities.** Often we don’t want to know what every single molecule does, but rather what many molecules together do. This simplifies life because instead of worrying about  $10^7$  positions of  $10^7$  molecules, we only need to worry about the volume  $V$  they occupy, for example. For example, if you observed several molecules undergoing a chemical reaction and ‘disappearing’ by turning into a new molecule, your observations might be summarized in a series of plots, such as



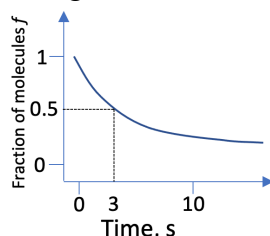
“1” means the molecule is still there, “0” means it has reacted. Each molecule reacts at a specific time, but you might observe that more of them react quickly. If we added these ‘single molecule’ traces together for 10 molecules, and divided by 10 to get the average, you might get a trace that looks as follows:



It looks like about half the molecules are gone after 3 seconds, so you might conclude from this plot that the reaction has a ‘half life’ of  $t_{1/2} \approx 3$  s.

How do we calculate averages? For a discrete number of points, like  $y=1,3$  and  $5$ , it’s just  $\bar{y} = (1+3+5)/3=3$ . This assumes that each value of  $y$ ,  $1$ ,  $3$ , and  $5$ , has the same probability of occurring, namely  $1/3$ . Of course if  $y=1$  is twice as likely as the others, then the answer is  $\bar{y}=(1+1+3+5)/4=2.5$ . Or in terms of probabilities,  $1$  occurs  $1/2$  the time,  $3$  occurs  $1/4$  the time and  $5$  occurs  $1/4$  the time, the average is  $1 \cdot (1/2)+3 \cdot (1/4)+5 \cdot (1/4)=2.5$ . For a continuous function  $y(x)$ , if we know with what probability  $P(x)$  the values of  $x$  are distributed, this sum just becomes an integral  $\bar{y} = \int dx y(x)P(x)$ .

**2. Derivatives or how things change with time.** As we all know, the derivative  $\partial f/\partial t$  tells us how quickly ‘ $f$ ’ changes with time. Looking at the above plot, you might wonder that if you had looked at 10,000 molecules instead of 10, and normalized to the fraction of surviving molecules  $f(t)$ , the plot would have invisibly tiny steps and might look like



You might wonder what that function might be. Perhaps it is universal for all chemical reactions? Or at least certain reactions? Let’s make an intuitive assumption:

“If you double the number of molecules, you get twice as many reactions”

or, more generally

“The rate of molecules reacting (disappearing) is proportional to the number of molecules”.

The number of molecules still there is given by  $N_0 f$ , where  $N_0$  is the initial number of molecules (e.g. 10) and  $f$  is the fraction surviving. And the rate of disappearance is the derivative  $-\partial N_0 f/\partial t$ . If they are proportional, we have

$$\frac{\partial N_0 f}{\partial t} \sim N_0 f.$$

Putting our “Gedankenexperiment” in the form of an equation like the one above, instead of just saying it in words, is often the hard part of physical chemistry that requires practice! Canceling out the constant  $N_0$  on each side (it can be pulled in front of the derivative since it is constant), and calling the unknown proportionality constant “ $-k$ ” (the - sign is not necessary in the definition of the proportionality constant, but simplifies things later), we can write

$$\frac{\partial f}{\partial t} = -k f.$$

This is a differential equation. If our intuition ‘twice as many molecules = twice as many reactions’ is right, it should describe the observation. It is easy to see by inserting

$$f(t) = e^{-kt}$$

into the above equation, that  $f$  satisfies the equation. Note that when  $f(t)$  has decayed to 0.5, we have reached the half life  $t_{1/2}$ . We can easily solve for  $t_{1/2}$  in terms of  $k$  since  $0.5=e^{-k t_{1/2}}$ , or  $t_{1/2} = \ln(2)/k$ .

Generally in this class, we are not going to bother solving differential equations in all the detail, but often check solutions by inserting a good guess. Of course solving differential equations either numerically (on a computer) or analytically (by hand) is important, but it is part of the ‘technology side’ of physical chemistry. The conceptual side is more important: how to go from expressing a

model in words, to writing down an equation (algebraic or differential) that makes those words quantitative, is the essence of building models in physical chemistry.

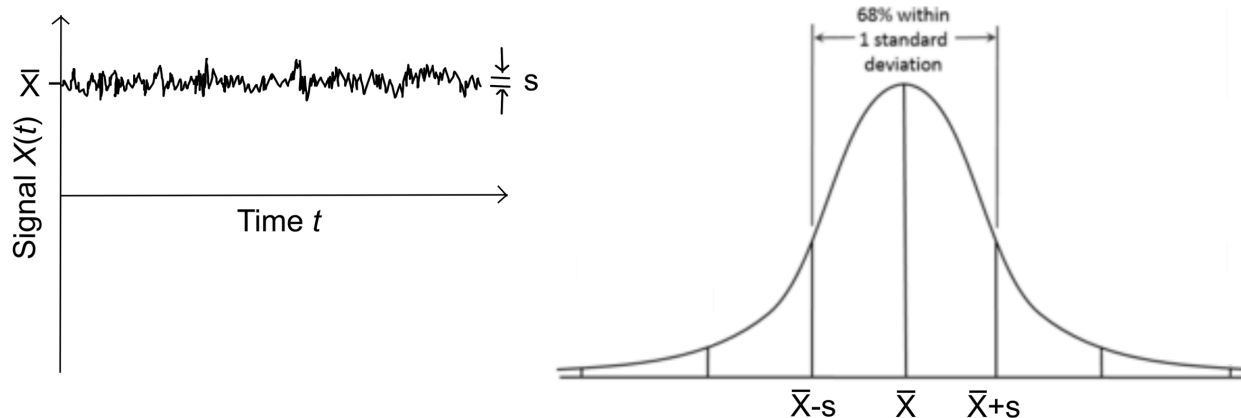
So from our “Gedankenexperiment” in our armchair, we come to the conclusion that chemical reactions should obey the **very general law we derived above**. Indeed, when you make measurements, many reactions obey this equation very well, but not all reactions do: some don't decay to 0 as  $t$  goes to  $\infty$ ; others seem to fit different equations, such as  $f \sim 1/[k+t]$ , and so forth. But this does not mean there's something ‘wrong’ with our equation. It's just that our assumption

“If you double the number of molecules, you get twice as many reactions” may not apply to all reactions.

**Thought experiment:** If you had two molecules “A” react to make “A<sub>2</sub>,” how would you put in words how the rate depends on concentration? Is that easier to see when you think of A+B→AB, so the two molecules are different? Can you put it in terms of a differential equation?

This is what physical chemistry is all about: look at phenomena surrounding us that involve molecules experimentally, take a guess at what the underlying reasons for the experimental answer may be, formulate that guess in terms of mathematics, and see how well it matches the data. When it does not, the guess may need refinement or replacement. The payoff is that the equations derived often apply to many related cases, so once you ‘know one, you know them all.’

**Homework Problem 01.1:** For the function  $y=t$ , calculate the average  $\bar{y}$ , if  $P(t)=ke^{-kt}$  per unit time over the range  $t=0$  to  $\infty$ . Since  $t$  is time, this gives the average time for a molecule with rate coefficient  $k$  to decay. How does that compare with the half-life of the molecule? [Hint:  $y=t$  is a continuous function, so you need to use the integral formula, not the summation formula.]



**3. Randomness and the Central Limit Theorem.** When one is dealing with many molecules, the behavior of the individual particles tends to smooth out to an average, but unless we are averaging over  $\infty$  many molecules, there will always be some random fluctuations left, like the little bumps in the figure on the previous page. Mathematicians have proved a very handy theorem, the Central Limit Theorem (see also Figure above):

When a random variable “ $X$ ” is added up from many random variables “ $X_i$ ” that have averages  $\bar{X}_i$  and variances  $V_i$ , then, no matter what the random distribution of the  $X_i$  looks like,  $X$  obeys a normal (Gaussian) distribution with mean (or average)  $\bar{X} = \bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \dots$  and variance  $V = V_1 + V_2 + V_3 + \dots$ . This explains a somewhat mysterious fact, that when you measure just about any experimental data, the noise in the data is always Gaussian (or normal) distributed. The individual

noise sources  $X_i$  could be elevators shaking the building, air current in the room, or electrical noise from the power plug, but when it all adds up, it looks like random Gaussian noise. In the picture above, as always, the standard deviation is the square root of the variance, given by  $s = \sqrt{V}$ , and the variance is the average of the squared deviations of  $X$  from its mean value, e.g.  $V_i = \Sigma (X_i - \bar{X}_i)^2$  for variable  $X_i$ .

**Homework Problem 01.2:** Go to the applet

[http://195.134.76.37/applets/AppletCentralLimit/Applet\\_CentralLimit2.html](http://195.134.76.37/applets/AppletCentralLimit/Applet_CentralLimit2.html), which shows you how starting with 1 random variable  $X_1$  with distribution  $P(X_1)$ , then adding a second, and so forth, eventually produces a Gaussian. Click on “8” = exponential decay, Population size ‘Medium’, and draw for sample size 1. Then Increase sample size to 2 (two random variables, both exponentially distributed added up), 5 and 10. Does it begin to look like a Gaussian for 10 random exponential variables added up?

**4. Bayesian inference:**

Consider a coin toss experiment:



$P(T)=P(H)=0.5$  means there’s a 50% chance of getting “H” = heads or “T” = tails. Here your best bet is just to take a random guess each time and you should be right half of the time. If you were to toss several =  $N$  coins and heads is worth 0 and tails is worth 1, then by the Central Limit Theorem (CLT), the sum of the each toss will have average value  $N/2$  and will obey a random Gaussian distribution.

But consider the following coin toss experiment:



Still 50:50, but I bet that if you look at the pattern, and saw that two “tails” had just been tossed, you would guess “heads”, not randomly! Looking at the sequence, the tosses are rigged, that is, there is a pattern that is not totally random. One can also hear this well in a sonification of tosses, such as at

Video Sonification of a simple coin-toss that is either fair, biased toward Heads, or biased toward Tails ([https://www.youtube.com/watch?v=zGcPIZy\\_mnI](https://www.youtube.com/watch?v=zGcPIZy_mnI))

Experiment in class: In the following Video Coin-toss with an unknown bias, can you hear the bias - is it heads or tails or fair? (<https://www.youtube.com/watch?v=e8yeLEX14lk>)

What about the following coin toss sequence?



This is where Bayesian statistics can help you make the best possible decision by using *incomplete prior knowledge together with statistics to make the best possible next guess*. For instance, if you

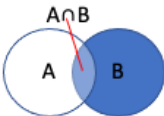
think that after a single “T”, another “T” is very likely above (they do seem to come in pairs), you’ll guess “T” with good certainty. On the other hand, after seeing a single “H”, you can’t be too sure whether another “H” or “T” will follow. Bayesian statistics allows you to crunch the math to take that best guess.

**Example:** If you had designed an expensive experiment that you want to do only when T = true, in case 3 you might measure until you get T, measure once more, then skip one measurement before trying again to save yourself some time and effort, even if  $P(T) = \frac{1}{2}$ . After all, it looks like an “H” comes at least once after every pair of “T.” This is an example of Bayesian thinking.

When analyzing statistics in terms of beliefs about future outcomes, one constantly deals with conditional probabilities: **What is the probability P that A will happen, given I know that B is true, or  $P(A|B)$ . (Will “T” happen again, given I know B = “I just got T”?)**

Often  $P(A|B)$  is not very intuitive, even though  $P(B|A)$  is. For example, you can think of the first as “What is the chance that A will happen, given that B is true” and the second “What is the chance B will happen given that A is true”?

Bayes theorem relates these complementary ideas: is a theory true given an observation? How likely will an observation happen based on a given theory?

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$


$P(A|B) = \frac{P(A \cap B)}{P(B)}$ 
 This reasoning is symmetrical, so
  $P(B|A) = \frac{P(A \cap B)}{P(A)}$ 
 prior probability

**In class exercise:** I used to follow Formula 1 racing’s Austrian star Nikki Lauda as a child in the 1970s, and found the following ‘cute’ example at <https://www.analyticsvidhya.com/blog/>:

Suppose Nikki Lauda won 3 out of 4 times against James Hunt. Who would you bet on in race #5, and what odds?

**Answer:** Nikki Lauda, with 3:1 odds for your betting money;  $P(B) = 1/4$  is the prior probability that Hunt wins. So,  $P(A|B) = P(\text{Lauda wins next} | \text{Hunt won 25\% of time before}) = 75\%$ .

Now suppose I gave you some extra info: Lauda sucks at driving in the rain; the one time Hunt actually won it rained hard. All the times Lauda won the weather was nice. At race #5, the forecast is hard rain. Who would you bet on in race #5 NOW?

**Answer:** Independently of the racing, let  $P(A)$  be the prior probability that it rains. This is  $P(A) = 0.25$  since we know it rained once out of 4 races previously.

Similarly  $P(A|B) = 1$  is the probability that it rained given that Hunt won. (It rained every time Hunt won, which was one rainy day, one win for Hunt).

So here  $P(A)$ ,  $P(B)$  and  $P(A|B)$  are pretty easy to figure out from the data. We want to know “What is the estimated probability Hunt wins, given that it will rain?” This is the less obvious quantity. But luckily, it’s  $P(B|A)$  !

$$P(B|A) = P(A|B) * P(B) / P(A) = 1 * 0.25 / 0.25 = 1$$

So, you better bet on Hunter if you know it rains, given that he’s always won against Lauda in the rain.

Note that this does not have to hold up with additional measurements; Lauda might rarely win in the rain when Hunter has a bad day. If Hunter lost race #5, then the re-calculated  $P(B|A)=0.5$  based on 5 races is lower than 1, but still not as low as 0.2. You can keep updating your probabilistic expectations, and bet accordingly as more data comes in. So Bayesian statistics is predictive, whereas conventional statistics is purely retrospective.

**Homework problem 01.3:** Consider the molecular DNA sequence CCAGCAC\_GGCAGCAT, with a base missing. What would be a good guess for that base? To start, what is the probability  $P(A)$  that adenine occurs in that known sequence? The probability  $P(C)$ ? The probability that whenever there is an A, there's a C before it? (that is  $P(C, \text{ given } A \text{ is in the next position})$ ). Therefore what is  $P(A|C) = P(A, \text{ given } C \text{ is in the previous position})$ ? So what odds do you give an A being in the blank space? **Note: this homework is not in green; no need to turn it in, but make sure to do it, it could be on the exams, like all homework, exercises and thought experiments.**

**5. Logarithms: when adding is easier than multiplying.** In many areas of science, quantities are multiplicative. For example, two unrelated probabilities are multiplicative. If 1% (0.01) of cars have an accident every year, and 10% (0.1) of cars are tan-colored, then  $0.1\% = 0.01 \times 0.1$  of cars will have accidents *and* be tan. If the fraction were smaller or bigger, then we would know that the two probabilities are not independent; for example, tan cars might be involved in accidents more often than the average if they are more difficult to see. In chemistry we also deal with probabilities, such as what is the probability that all molecules of gas are on one side of a box. These probabilities can get astronomically large or small, in which case multiplying the numbers is a hassle. The log comes to the rescue:

- $\ln(x)$  compresses the dynamic range of really small and really large numbers
- if  $z=xy$ , then  $\ln(z)=\ln(x)+\ln(y)$ , turning multiplication into addition.

Additivity is also a very useful concept in chemistry. For example, if we have a box full of molecules, with energy  $E$ , and we stick next to it an identical box of molecules, we expect the total energy to be  $E_{total}=E_1+E_2$ . We call such quantities 'extensive', as opposed to an intensive quantity. For example, the density of each box is  $\rho=M/V$ , where  $V$  is the volume of the box and  $M$  the total mass of molecules in the box, and the density of both identical boxes together is  $2M/2V=M/V$ , which is still equal to  $\rho$ .

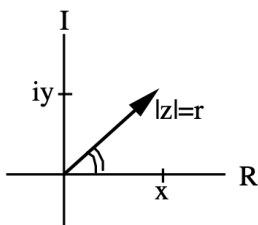


Figure: A complex number  $z=x+iy$  is really two real numbers  $x$  and  $y$ , like components of a 2-dimensional vector, with a special rule for multiplication so  $i$  time  $i = 1$ . For example,  $z^2=(x+iy)(x+iy) = x^2+2iy-y^2 = (x^2-y^2)+2iy$  if we group by the real and imaginary part at the end.

**6. Complex numbers - not so complex.** Many times in human history, people have realized that they were missing some kind of number. Consider negative numbers. They arose from the need to understand things like debt: "I lent you 5, you gave me 3 back so far, you still owe me 2". The concept "owe-me-2" is written in shorthand as "-2". Later, people realized they were missing irrational numbers.  $22/7=3.142857142857\dots$  is just not equal to  $\pi=c/d$ , where  $c$  is the circumference and  $d$  the diameter of a circle. Eventually people came to grips with the idea that some numbers have digits that never repeat and are not ratios - hence 'ir-ratio-nal.'

And finally in the 1600s, people started worrying about roots because they are needed in many area-to-length conversion problems, and the problem of  $\sqrt[4]{-1}=?$  came up. The solution was to invent numbers that are like 2-D vectors and have two parts, 'real', and 'imaginary', such that  $i=\sqrt{-1}$  and, for example,  $a=2+3i$  is a complex number with a 'real' part = 2 and an 'imaginary'

part = 3i. The addition rule is just like vectors, so if  $b=1+1i$ , then  $a+b=3+4i$ . Soon someone realized that since the Taylor expansion for  $\cos(\varphi)$  is  $1-(1/2!)\varphi^2+\dots$  and for  $\sin(\varphi)$  it is  $\varphi - (1/3!)\varphi^3+\dots$  and for  $e^z$  is  $1+z+(1/2!)z^2+(1/3!)z^3+\dots$ , therefore  $e^{i\varphi} = \cos(\varphi)+i \sin(\varphi)$ . This was a real breakthrough in mathematics: until then, people thought that the exponential function and trigonometric functions are two separate functions, but they are actually the same function, just with real or imaginary numbers plugged in. So, complex numbers can simplify our understanding of things, not just make things more messy!

Multiplying complex numbers together uses the usual associativity rule, for example  $a \cdot b = (2+3i)(1+1i) = 2 \cdot 1 + 2 \cdot 1i + 3i \cdot 1 + 3i \cdot 1i = 2 + 5i - 3 = -1 + 5i$ . The trigonometric notation allows a nice interpretation of multiplication: Any complex number can now be written as  $z=x+iy=re^{i\varphi}$ , where  $r=\sqrt{x^2+y^2}$  is its length, and  $\varphi = \tan^{-1}(y/x)$  is its ‘angle.’ For example  $a=2+3i=\sqrt{(2^2+3^2)} e^{0.983i}$  and  $b=1+1i=\sqrt{2} e^{i\pi/4}$ , since  $\tan^{-1}(1/1)=\pi/4$ . Multiplying them together we get  $a \cdot b \approx 3.60 \cdot 1.41 \cdot e^{1.768i} \approx 5.09 e^{1.768i} \approx -1+5i$  again (rounding error!), where we used the multiplication rule for exponentials (see logarithms above!), namely  $e^a \cdot e^b = e^{a+b}$ . Thus, when multiplying two complex numbers together, you multiply their magnitudes together and add their angles together.

**Homework problem 01.4:** Write the number  $i$  in exponential notation (what is its angle  $\varphi$ ?). Now take the natural logarithm  $\ln(i)$ . What do you get? Can you devise a simple procedure that extracts the magnitude ‘ $r$ ’ and angle ‘ $\varphi$ ’ of any complex number by distributing them into a real and imaginary piece?

Why use complex numbers in Chemistry? It turns out that you can talk about subjects like sound waveforms or quantum mechanics without complex numbers, by just talking about a ‘wavefunction’ (we’ll see what that is later) as being two functions arranged as a 2-D vector, but it is often easier to think of the ‘wavefunction’ or any other wave as a single function with a real and an imaginary part. Luckily, many wavefunctions of interest in Chemistry happen to be real anyway, i.e. their imaginary part is equal to zero.