

Chemical equilibria from a statistical perspective

1 Introduction

The science of chemistry as typically taught in high school develops the subject by building upon the experimental results from scientists primarily working in the 18th-19th centuries. Although these scientists made remarkably accurate conclusions about chemistry and nature through experimentation, we have since developed more rigorous physical theories and mathematical tools which paint a complementary picture to the original results reported by Avogadro, Lavoisier, Dalton, etc. As a result, we strive to rationalize those foundational topics from introductory chemistry, such as the laws of thermodynamics, LeChatelier's principle, and equilibrium constants through an approach guided by probability and statistics. I hope that the presented approach serves a pedagogical purpose in enlightening a different way of thinking about the introductory chemistry curriculum and cultivating new understanding.

The discussion presented here has two general sections. First, we introduce fundamental concepts in statistical mechanics that will be crucial to the second section, where we apply those concepts to model and discuss the behavior of chemical reactions at the molecular level. The first section (parts 1-5) is **not** an attempt to comprehensively teach statistical mechanics. Instead, I try my best to present a crystallized summary of my mental picture of specific topics that I have learned through coursework. Specifically, I would like to acknowledge the material presented by Professors Senthil Todadri and Julien Tailleur in 8.333, MIT's graduate intro to statistical mechanics course, for both building fundamental knowledge and sparking insights into connections with chemistry. The course follows *Statistical Physics of Particles* by Mehran Kardar, and another text I find very interesting is *Entropy, Order Parameters, and Complexity* by James Sethna. The first section is simply a distillation of content from the course presented to make the second section (part 6+) intelligible, and the above professors and authors deserve the credit for their initial presentation. To be explicit, the discussion on AP chemistry topics in the second section is not anything novel either - I must imagine that this type of thinking is crucial to any research work in classically modelling molecular and reaction dynamics. Nevertheless, I hope this discussion brings these ideas closer to the realm of high school or undergraduate chemistry education.

2 Chemical systems as random variables

The root from which our discussion stems from is a formal description of the chemical system as a random variable. We seek to formulate the results presented in introductory chemistry courses through physically guided interpretations of the random variable associated with the system. We introduce the formalism of the chemical system as a random variable through the example of an ideal gas and continue to add complexity as our model becomes more physical.

What is an ideal gas?

An ideal gas is a collection of N particles where each particle i has some position \mathbf{x}_i and momentum \mathbf{p}_i .¹ This model is typically used as a simplified description of a real gas, where the atomic makeup and quantum properties

¹These variables are used to describe each particle of the ideal gas because of the hidden assumption that the system is governed by Newton's laws of physics, which uses a second-order linear differential equation to describe the trajectory of each particle. As a result, we require an initial condition for x (position) and \dot{x} (momentum) of each particle to guarantee a unique description of system behavior per existence and uniqueness theorems for the solution of such equations.

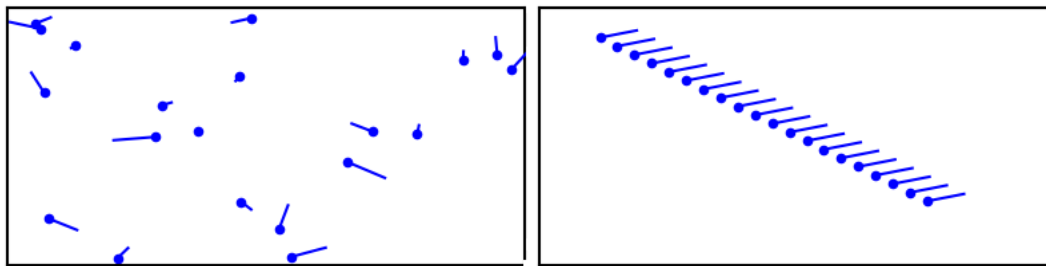


Figure 1: Two different valid configurations of positions and momenta for $N = 20$ ideal gas particles, where the position of each dot corresponds to each \mathbf{x}_i and its respective tail corresponds to \mathbf{p}_i . As we will see, we seek to construct a distribution over all such configurations $\{\mathbf{x}_i, \mathbf{p}_i\}$, irregardless of how "disordered" (L) or "ordered" (R) atoms appear in the colloquial sense.

of each gas molecule have been binned away into the description of a generic particle with no size. Additionally, there is no appreciable interaction between particles that we should account for. Furthermore, a real gas occupies a physical container with volume V constraining the gas volume; in the ideal model, we impose a volume $V \subset \mathbb{R}^3$ such that each $\mathbf{x}_i \in V$.

How can we describe an ideal gas as a random variable?

Random variables are typically specified by a set of outcomes occurring with some known probability or probability density. For example, for the random variable of a die throw, the outcomes $\{1, 2, 3, 4, 5, 6\}$ each occur with probability $\frac{1}{6}$, while for a normally distributed random variable, the outcome can be any real number with probability density given by a bell curve centered at the mean.

For the case of an ideal gas, we want to consider the outcome space of all possible positions and momenta of each particle in the gas - that is, each outcome associated with the ideal gas random variable is specified by N positions $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in V$ and N momenta $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N \in \mathbb{R}^3$. For example, two outcomes are depicted in Fig. 1. Describing how we assign a probability density function over the $6N$ -dimensional outcome space is very difficult - in fact, we devote all of section 3 to discussing how to construct such a probability density to model gas behavior in a physically realistic manner. For now, we will refer to that distribution as $p(\{\mathbf{x}_i, \mathbf{p}_i\})$ but make no assumptions about its properties.

How does the random variable picture connect with experimental chemistry?

At this point, it is natural to question the connections between this model and experiments on physical systems. After all, our distribution is supported on the positions and momenta of each atom of the gas, but no chemistry experiments in reality seek to keep track of the behavior of every single atom; at least, certainly not when chemistry is done on the molar $N \sim 10^{23}$ scale and certainly not in the era when the foundational experiments described in introductory chemistry were performed. Instead, statements that scientists make as a result of chemistry experiments (ex. the reaction to form product had 90% yield, the product has volume 10 mL, etc.) are statements about global system properties on the human scale.

Luckily, there is a clean mathematical description that bridges the human scale that experiments are performed on to the atomic scale that our probability distribution $p(\{\mathbf{x}_i, \mathbf{p}_i\})$ exists on. This is because human scale, experimentally measurable quantities are often simple functions of the microscopic quantities. As a result, each element of the outcome space of the ideal gas random variable associated with the microscopic $\{\mathbf{x}_i, \mathbf{p}_i\}$, can also be associated with a macroscopic, measurable quantity. For example, the total kinetic energy of the system,

a quantity that is strongly correlated with the measurable system temperature as we will see later, is simply the sum of the microscopic kinetic energy of each particle:

$$E(\{\mathbf{x}_i, \mathbf{p}_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} \quad (1)$$

where each particle is assumed to have mass m . As a result, the system energy that is seen by humans in an experiment follows the distribution

$$p(E = E_0) = \int p(\{\mathbf{x}_i, \mathbf{p}_i\}) \delta(E(\{\mathbf{x}_i, \mathbf{p}_i\}) - E_0) d^3\mathbf{x}_1 d^3\mathbf{p}_1 \dots d^3\mathbf{x}_N d^3\mathbf{p}_N \quad (2)$$

that is, the probability distribution of the energy E at any fixed E_0 can be calculated by finding the total contribution of the microscopic probability distribution $p(\{\mathbf{x}_i, \mathbf{p}_i\})$ for all values of $\{\mathbf{x}_i, \mathbf{p}_i\}$ which have energy E_0 , ie. solutions to the equation $\sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} = E_0$.² We approach modelling experimental results related to the energy of a gas by associating its energy with the distribution given by $p(E)$. In the following sections, we will see that the unique properties of such a distribution for a real system at equilibrium can be leveraged to construct intuitive statements about real chemistry.

3 Equilibrium probability distribution of the ideal gas

We return to the discussion of how the probability density $p(\{\mathbf{x}_i, \mathbf{p}_i\})$ for the ideal gas actually behaves. Here, we use the notation μ_j to denote some arbitrary fixed value of the $6N$ parameters $\{\mathbf{x}_i, \mathbf{p}_i\}$ and refer to each distinct value of μ_j as a microstate. For ease of the intuitive description in this section, we consider a finite number of microstates $\mu_1, \mu_2, \dots, \mu_\Omega$ which we can think of as arising from some artificial discretization of position/momentum space. Thus, the distributions p discussed here associate discrete microstates μ_j with finite probabilities $p(\mu_j)$ such that $\sum_j p(\mu_j) = 1$. The result that we arrive at lends itself very well to the continuous limit where the discretization size goes to zero.

To understand the nature of $p(\mu_j)$, we begin by posing a fundamental postulate from which we can derive mathematically sensible results:

(Postulate) Given a physical system associated with a random variable with probability distribution $p(\mu_j)$, there exists an function $S(p)$ representing our ignorance of the system with the following properties:

- After not interacting with or observing the system, $S(p)$
- The system is said to be in equilibrium if it is well described by a probability distribution p where S has reached a maximum.

The general concept that "ignorance in a system will not decrease if we do not do anything to try to reduce our ignorance" seems almost too obvious to be explicitly written. However, a *proof* of such a statement regarding why and how systems evolve from order to disorder as time progresses is an important open question about the nature of our universe. Although there exist proved theorems with similar implications,³ I believe that taking the above postulate to be true provides a good starting point for the meat of the discussion.⁴

²We note for any E_0 , the solution set represents a $3N$ -dimensional sphere with radius $\sqrt{2mE_0}$ in momentum space.

³<https://en.wikipedia.org/wiki/H-theorem>, https://en.wikipedia.org/wiki/Arrow_of_time

⁴This is no doubt reflective of the dilemma in science education and communication that I alluded to in the introduction - like the choices that educators are posed with in designing chemistry curricula, I have not presented further rigorous discussion of the postulate. However, I feel comfortable that the amount of additional insight that stems from the postulate constructively builds upon that presented in a generic introductory chemistry course.

This concept of the ignorance function $S(p)$ arises from information theory, from which we borrow the concept that **the ignorance present in a distribution over system microstates is proportional to the number of additional bits of information needed to describe the exact system microstate**. This should feel reasonably intuitive since this number of bits reflects the amount of knowledge that we do not have but is required to pin the system down to one of the μ_j 's. For example, if we had perfect knowledge (ie. no ignorance) that our system is always in some microstate μ_1 , we trivially do not need any additional information to say that the system is in microstate μ_1 , corresponding to the value of the function $S(p_1 = 1) = 0$. If we lose some knowledge and now only know that the system could be in microstates μ_1 and μ_2 with equal probability, we can correspond the bitstrings "0" and "1" to the event that the system is in either of the two microstates; then, receiving a single bit would be sufficient for us to know the system state. Therefore, $S(p_1 = \frac{1}{2}, p_2 = \frac{1}{2}) = 1$.

We can assign a value of S for any distribution over the Ω outcomes μ_j in a result originally due to Shannon. Let us consider an ensemble of M i.i.d. copies of our system. If under p , our system attains state μ_j with probability p_j , then, the state μ_j shows up in the ensemble Mp_j times in expectation. Then, the total number of distinct permutations of the M observed outcomes is

$$\frac{M!}{\prod_{j=1}^{\Omega} (Mp_j)!} \quad (3)$$

Therefore, if we want associate each of these permutations to a bitstring, which corresponds to the amount of additional information needed to fully disambiguate the different permutations, we would need

$$\log_2 \frac{M!}{\prod_{j=1}^{\Omega} (Mp_j)!} = \log_2 M! - \sum_{j=1}^{\Omega} \log_2 (Mp_j)! \quad (4)$$

bits. By applying Stirling's approximation $\log N! = N \log N - N + O(\log N)$ and dropping terms subdominant to N , we see that the number of bits required is

$$\begin{aligned} (M \log_2 M - M) - \left(\sum_{j=1}^{\Omega} Mp_j \log_2 (Mp_j) - Mp_j \right) &= \left(\sum_{j=1}^{\Omega} Mp_j \log_2 M \right) - \left(\sum_{j=1}^{\Omega} Mp_j \log_2 (Mp_j) \right) \\ &\propto -Mk \sum_{j=1}^{\Omega} p_j \log p_j \end{aligned} \quad (5)$$

We note that scaling by constants is unimportant since we can impose that the ignorance cost of a single bit is any arbitrary $k > 0$ units of ignorance as long as each bit is worth the same k , which also allows us to absorb logarithm base-changes into k . Thus, the number of ignorance units per system copy is

$$S = -k \sum_{j=1}^{\Omega} p_j \log p_j \quad (6)$$

which we define as the **entropy** S . In fact, following the approach presented by Sethna in his text in problem 5.17, it can be proven that Eqn. 6 is the unique function up to scaling that represents an intuitive formulation of ignorance from a probabilistic standpoint. Circling back to our postulate, our system tends towards distributions over microstates that maximize S , which we claim occurs when the distribution is uniform over the Ω outcomes. The key observation is that $f(x) = x \log x$ is a convex function of x on $(0, 1)$. Therefore, Jensen's inequality implies that the average value of $x \log x$ over inputs $\{p_j\}$ is bounded by the value of $x \log x$ at the average input $\frac{1}{\Omega} \sum p_j$:

$$\frac{1}{\Omega} \sum_j p_j \log p_j \geq \left(\frac{1}{\Omega} \sum_j p_j \right) \log \left(\frac{1}{\Omega} \sum_j p_j \right) \quad (7)$$

Since we must have $\sum p_j = 1$, it follows that the entropy is bounded above by $-k \log \frac{1}{\Omega} = k \log \Omega$, which we can confirm is achieved when $p_1 = \dots = p_\Omega = \frac{1}{\Omega}$. We note that this formula for the entropy

$$S = k \log \Omega \quad (8)$$

matches the infamous result found on Boltzmann's gravestone. We can also arrive at this result through the method of Lagrange multipliers, where if we impose that the gradient of S with respect to $\{p_j\}$ is parallel to the gradient of the constraint $(\sum p_j) - 1 = 0$, we obtain that uniform distribution over all p_j is a unique extrema of $S(p)$.

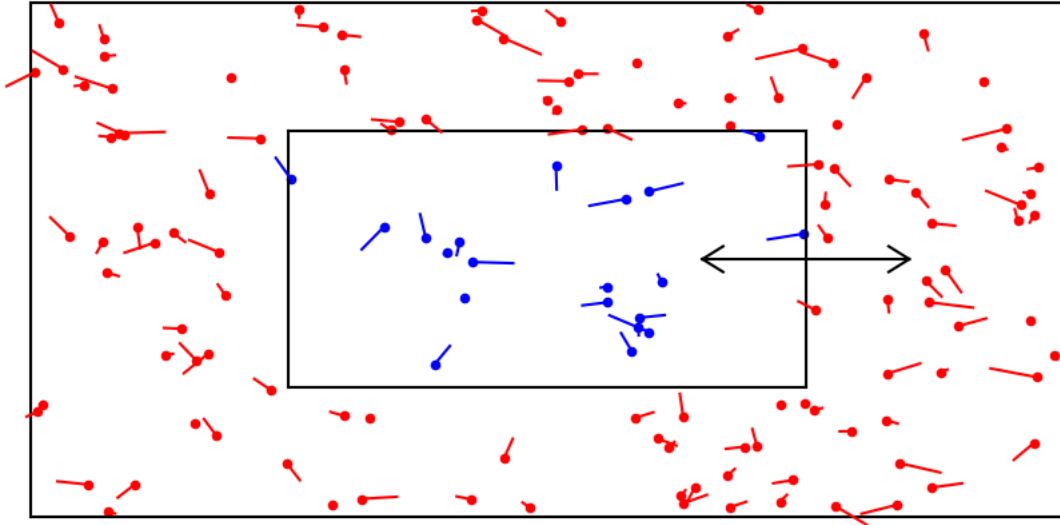
As a result, our postulate allows us to narrow down the uniform distribution as the single entropy-maximizing distribution among all possible distributions over microstates. Given how many possible distributions there are over the $6N$ -dimensional parameter space of $\{\mathbf{x}_i, \mathbf{p}_i\}$, the fact that there is this unique maximum is quite a remarkable result. We will exploit this uniform structure in the unique entropy-maximizing distribution to draw conclusions about how chemical systems behave at equilibrium.

It's also worth noting that our postulate *doesn't* discuss how the system achieves the equilibrium distribution over outcomes, only that such an equilibrium distribution exists and is characterized via entropy maximization. The study of the processes by which systems reach equilibrium, typically falling under the realm of kinetics, is outside of the scope of this discussion. There are many chemically and mathematically interesting phenomena in the field of kinetics, including the tools used to prove theorems closely linked to our aforementioned postulate. We reserve that discussion for another time and choose to focus on the aspects of interest of chemical systems that have already reached equilibrium and are thus characterized by the entropy-maximizing distribution. In the language of a practical experiment, if we perform a chemical reaction between colorless A and B to form blue-colored C, we aim to describe the concentration of blue-colored molecules in the solution after the solution color has stabilized,⁵ but not the rate at which the blue color appears in the reaction mixture.

4 Constant-temperature chemical systems

Due to our statistical result that all systems converge to a equilibrium state that is uniform over all microstates, it may naively come as a surprise that there are any emergent properties on the macroscopic scale that can be detected via experiment. However, our lived experiences and intuition clearly run against this: running multiple instances of a chemical reaction under identical conditions yields an identical composition of the final reaction mixture across all instances, not a uniform distribution over possible final compositions. There are two intuitive descriptions for why these seemingly deterministic processes emerge from entropy-maximizing uniform distributions. First, we can refer back to the two examples of $\{\mathbf{x}_i, \mathbf{p}_i\}$ configurations presented in Fig. 1. If one simply randomly generated pictures of position and momentum coordinates for the atoms in the box, many more pictures would qualitatively look like the left panel compared to the right panel. Second, real chemical systems are not described by the idealized picture of a box floating in space, but rather are constantly exchanging energy with the environment. As a result, the uniform equilibrium distribution that we derived in the previous section is uniform over all system and environment microstates, which is generally not equivalent to the uniform distribution over purely system microstates. Through this section, we will describe the aforementioned intuition with more mathematical rigor.

⁵ie. via a UV-vis experiment after stirring overnight after the onset of a stable blue color has been observed



Energy exchange for thermalization

Figure 2: Ideal gas with $N = 20$ particles (blue) enclosed in an environment (red) that maintains the system temperature at some fixed $T_1 = T$. For example, a container of gas (the system) placed on a lab bench in a room (the environment) will eventually reach the room temperature in its equilibrium state, independently of the temperature that the container begins at, without significantly changing the temperature of the room.

Consider a system considering the aforementioned ideal gas in a box of volume V (denoted 1) connected to a large thermal bath at fixed temperature T (denoted 2) which behaves as the environment. We treat the composite system-environment construction as an isolated system that achieves a uniform distribution over microstates at equilibrium according to our postulate. The environment fixes the system temperature at T because if the system temperature reaches some $T_1 \neq T$, it will fall out of thermal equilibrium with the environment. The environment being "large" implies that there are sufficiently many particles to provide or absorb the energy difference to bring the system temperature back to $T_1 = T$ without changing $T_2 = T$.

What does it mean for a system to have temperature?

Although we have imposed that the temperature of the system and environment must be equal based on our intuitive understanding of temperature, we have not yet identified the temperature with any quantities from our statistical model. We attempt to do this by identifying a statistical quantity that also must be equal among system and environment.

We begin by noting that the number of microstates of the system-environment composite system can be calculated by summing over all ways the total energy E can be partitioned among the system and environment:⁶

$$\Omega(E) = \int dE' \Omega_1(E') \Omega_2(E - E') = \int dE' \exp \left(\frac{1}{k} (S_1(E') + S_2(E - E')) \right) \quad (9)$$

where we have used the Boltzmann entropy formula (Eqn. 8) to simplify the integrand. This may seem like an intractable integral, especially as we have no assumptions or modelling for the behavior of the environment.

⁶Recall E is equivalent to the microscopic energy defined in Eqn. 1. Although microscopic energy can be exchanged through the wall separating system and environment, there is nowhere for the energy of the total system to dissipate and thus remains constant.

However, we can use a technique known as **saddle point integration** to gain some intuition about how this integral behaves. Consider the integers from 1 – 20, over which $1 + 2 + \dots + 20 = 210$, of which the the largest integer 20 contributes $\frac{20}{210} \approx 9.5\%$ to the sum. In contrast, e^{20} contributes 63% to the sum $e^1 + e^2 + \dots + e^{20}$. This is due to how much more rapidly the exponential function e^x grows relative to all other functions - although 20 is of a similar order of magnitude as the rest of the integers 1, ..., 19, e^{20} is at least ~ 10 times more than the rest of the exponentials except e^{19} . This effect becomes even stronger for functions e^{Nx} for integer N , ie. for $N = 10$, e^{200} is $e^{10} \approx 20000$ times larger than e^{190} . More rigorously, for any set of numbers X , we have that $\sum_{x \in X} e^{Nx} \rightarrow \exp(N \max_{x \in X} x)$ as $N \rightarrow \infty$. Extending to the integral of the exponential of a function $\phi(s)$, $\lim_{N \rightarrow \infty} \int_S ds e^{N\phi(s)} = \max_{s \in S} e^{N\phi(s)}$, where the set X in the previous discrete case is analogous to the continuous set $\{\phi(s) \mid s \in S\}$.

For a system with $N \sim 10^{23}$ particles, each particle contributes to the total ignorance S ; as a result, almost all integrals of relevant exponential functions in our chemical system can be evaluated by the saddle point method. Therefore, we have that $\Omega(E) = \Omega_1(E^*)\Omega_2(E - E^*)$, where E^* maximizes the argument of the exponential $\frac{1}{k}(S_1(E') + S_2(E - E'))$ over all E' . Therefore, although the entropy maximizing distribution is uniform over all composite system microstates, there are **so many more** microstates with energy E^* in the system and $E - E^*$ in the environment than any other configuration due to the exponential dependence. As a result, the system energy behaves deterministically despite the uniform distribution, foreshadowing the behavior of many other experimental observables that we will consider.

Additionally, since the entropy-maximizing energy E^* satisfies

$$\frac{\partial}{\partial E'}(S_1(E') + S_2(E - E')) = 0 \rightarrow \frac{\partial S_1(E)}{\partial E} = \frac{\partial S_2(E)}{\partial E} \quad (10)$$

after applying the requisite chain rule to convert to parameterizing each subsystem with its own energy E , we have found our statistical quantity $\frac{dS}{dE}$ that must be equal between a system and its environment for equilibrium to be established. Note that this result is even system independent - we never used the assumption that the system/environment was composed of ideal gases. However, we must tread lightly in relating this quantity to an intuitive measure of the temperature T , since our assumption of equal T among system and environment only implies that $\frac{\partial S}{\partial E}$ is some function $f(T)$. We do not try to nail down the exact function here; instead, we discuss some general properties. We note that if $\frac{\partial S_1(E)}{\partial E} > \frac{\partial S_2(E)}{\partial E}$, we can increase the total composite system entropy by increasing E . Therefore, the system is not in equilibrium with the environment, implying that one has a higher temperature than another. As energy flows in the direction of higher to lower temperature according to our everyday understanding, the system, assumed here to have larger $\frac{\partial S}{\partial E}$, must have a lower empirical temperature than the environment, with the same argument holding in the opposite direction when the system has smaller $\frac{\partial S}{\partial E}$ than the environment. Additionally, adding energy to a system of ideal gas particles increases the average momentum (Eqn. 1), increasing the amount of momentum space enclosed by our spherical manifold of states, which in turn increases the number of possible microstates and thus S . As a result, $\frac{\partial S}{\partial E}$ is generically positive.

In fact, the exact relation, which follows more easily from a classical thermodynamic treatment, is⁷

$$\frac{1}{T} \equiv \frac{\partial S}{\partial E} \quad (11)$$

which is consistent with our previous intuition that $\frac{\partial S}{\partial E}$ was somehow inversely related to temperature and always positive. In order to satisfy historical conventions on the definition of the Kelvin and measures of thermodynamic heat, the cost we impose on a bit in the statistical description is given by Boltzmann's constant

⁷Here is one of the only instances here where no statistical treatment I could arrive at was helpful in my opinion at describing the background behind the chemistry. There is a reasonable description of why the statistical definition of S is equivalent to thermodynamic one at <https://arxiv.org/abs/1903.02121>; then, the thermodynamic description of entropy as a heat per unit temperature gives a much more direct route to the expression as the heat contribution to the system energy comes as $dE = T dS$.

$k = k_B = 1.38 \cdot 10^{-23} \text{ J/K}$. Note the units of k_B are also suggestive of the inverse proportionality in Eqn. 10, but there is nothing fundamental in the information-theoretic view of entropy that points us towards the units of k .

How are states distributed in a system equilibrated to its environment at temperature T ?

We previously alluded to the fact that the uniform distribution over all system-bath microstates is not uniform when we focus on only the system, that is, when we integrate out states associated with the environment. To characterize the probability distribution p_1 over system microstates only, we consider some system-bath composite system with energy E and system in microstate μ_1 with energy E_1 . Since the equilibrium distribution over all system-bath microstates is uniform, we have

$$p_1(\mu_1) \propto \frac{1}{\Omega(E)} \Omega_2(E - E_1) \quad (12)$$

We immediately see the origin of non-uniformity in p_1 : since different system microstates can have different energies, they require a different amount of energy to be borrowed from the environment. As a result, the environment is left with a different number of allowed microstates $\Omega_2(E - E_1)$, which equals

$$\exp\left(\frac{1}{k} S_2(E - E_1)\right) = \exp\left(\frac{1}{k} (S_2(E) - \frac{\partial S_2}{\partial E} E_1)\right) \propto e^{-\frac{E_1}{kT}} \quad (13)$$

Therefore, for any microstate μ_j of the system only, its probability

$$p_S(\mu_j) \propto e^{-\frac{E_j}{kT}} \quad (14)$$

as all other prefactors are constant. We turn the proportionality into a direct equality by renormalizing the probability distribution to 1

$$p_S(\mu_j) = \frac{1}{Z} e^{-\frac{E_j}{kT}} \quad (15)$$

where constants have been binned into the partition function Z :

$$Z = \sum_{\mu_j} \exp\left(-\frac{E_j}{kT}\right) \quad (16)$$

Therefore, the famed Boltzmann distribution has a simple intuitive origin: it is nothing but the entropy-maximizing distribution of microstates in a system with overall temperature equilibrated by the environment. Again, it is important to reiterate that although the distribution over all system-environment microstates is uniform, the distribution over system states is not since the environment would have needed to provide a different amount of energy to the system to thermally equilibrate system microstates with different energy. As chemical experiments only focus on the behavior of the system (the environment is typically just the atmosphere surrounding a chemical reaction), chemical experiments on equilibrium systems behave according to the emergent non-uniform Boltzmann-distributed behavior for $p(S)$.

What general properties are observed for a Boltzmann-distributed system?

To understand the general behavior for systems obeying the Boltzmann distribution, it is important to understand what the distribution of energies $p_S(E)$ over Boltzmann-distributed microstates looks like, where E now represents the system energy.⁸

As the Boltzmann distribution is still uniform when we consider any subspace of microstates within the whole state space with the same energy, we simply have

$$p_S(E) = \Omega(E) \cdot \frac{1}{Z} \exp\left(-\frac{E}{kT}\right) \quad (17)$$

where $\Omega(E)$ is the number of microstates with energy E . Again replacing $\Omega(E)$ with the Boltzmann entropy $S(E) = k \log \Omega(E)$ gives

$$p_S(E) = \frac{1}{Z} \exp\left(-\frac{E - TS(E)}{kT}\right) \quad (18)$$

Again, we can avoid direct evaluation of integrals as in Eqn. 2 by using the saddle point method. Since each particle independently contributes its characteristic energy to the total energy E and its characteristic degrees of freedom in position and momentum to any given microstate, we have $E \propto N$ and $\Omega \propto e^N \rightarrow S \propto N$ for an N particle system. Therefore, for $N \sim 10^{23}$, particles, the distribution of $p(E)$ will be sharply peaked around the energy that maximizes $-E + TS(E)$ - again, the system exhibits pseudo-deterministic behavior in the distribution of a macroscopically observable quantity.

To make matters more convenient, we define the free energy of a system

$$F = E - TS(E) \quad (19)$$

The free energy gives us a convenient way to extend the equilibrium condition of maximum S over the system and bath to functions of E, T, S **of the system only**. This is an easy point of confusion, as our postulate seeks maximum S over system and environment. However, the S that enters the formula for F is the system entropy only. As a result, F provides a quantity that is both system-exclusive and minimized if and only if the system is at equilibrium by the previous saddle point argument. This is of particular utility since the entropy of the bath is poorly-defined and generically not interrogable with a basic experiment, while macroscopic system observables are much more closely associated with experimental results.

We let F^* be the minimum free energy obtained by the system (ie. the equilibrium free energy). Since the equilibrium energy distribution is sharply peaked around F^* , the probability of observing the system in its free-energy-minimizing state implies the condition $p(F = F^*) = 1$ at equilibrium, which dictates that

$$1 = \frac{1}{Z} \exp\left(-\frac{F^*}{kT}\right) \rightarrow F^* = -kT \log Z \quad (20)$$

It turns out that for most systems, the easiest way to compute the equilibrium free energy is through the partition function Z via some sort of summation which can often be made easier by exploiting some nice property of the system. For example, let us evaluate Z for our ideal gas.

The total system energy is $\sum_j \frac{\mathbf{p}_j^2}{2m}$, so the partition function is given by

$$Z = \sum_{\mu_j} \exp\left(\sum_j -\frac{\mathbf{p}_j^2}{2mkT}\right) \quad (21)$$

⁸As we have narrowed down to describing only the system with the distribution $p_S(E)$, subscripts of all quantities are dropped and assumed to refer to the system only unless otherwise noted.

We take the problem to the continuum limit by converting the partition function to an integral over all allowed states parameterized by position and momentum $\{\mathbf{x}_j, \mathbf{p}_j\}$:

$$Z = \frac{1}{N!} \int d^{3N}x d^{3N}p \exp\left(\sum_j -\frac{\mathbf{p}_j^2}{2mkT}\right) \quad (22)$$

where the integral is performed over all positions in the allowed volume V and all momenta in \mathbb{R}^3 . The factor of $\frac{1}{N!}$ accounts for the combinatorial overcounting associated with the indistinguishable of permutations of ideal gas particles. Then,

$$\exp\left(\sum_j -\frac{\mathbf{p}_j^2}{2mkT}\right) = \prod_j \exp\left(-\frac{\mathbf{p}_j^2}{2mkT}\right) \rightarrow Z = \frac{1}{N!} \left(V \int d^3p \exp\left(-\frac{\mathbf{p}^2}{2mkT}\right)\right)^N \quad (23)$$

Furthermore, $\int d^3p \exp\left(-\frac{\mathbf{p}^2}{2mkT}\right) = \int 4\pi p^2 dp \exp\left(-\frac{p^2}{2mkT}\right)$, which can be efficiently evaluated through the equivalence to the variance of a Gaussian random variable⁹. Thus,

$$Z = \frac{1}{N!} V^N \cdot 2^N \cdot (2\pi mkT)^{3N/2} \rightarrow F^* = -kT \log Z = -NkT(-\log N + 1 + \log V + \log 2 + \frac{3}{2} \log 2\pi mkT) \quad (24)$$

As the free energy represents an environment-independent quantity that is extremized at equilibrium, we will extensively use the result from Eqn. 24 to model the equilibrium behavior of real chemical systems in the upcoming sections. Although F^* is already the minimum free energy with respect to the current system configuration, we will see that real chemical reactions typically define a constrained parameter space over system variables (eg. N and V in this case) that we have taken to be arbitrary fixed numbers. Since the saddle point argument still holds when we extend the integral dimensionality, free-energy minimization for finding equilibrium states of chemical systems will be a key approach moving forward.

Additionally, we note that $F^* = E - TS$ represents a competition between two qualitative phenomena: at higher values of E , we generally have higher $S(E)$ increasing the system disorder, but higher values of E also require more energy from the bath and thus we have compensate for generally lower bath entropies as well. This also speaks to why this quantity is known as the *free* energy - although we have borrowed energy E from the bath to store in the system, we have a penalty $TS(E)$ associated with our ability to access $\Omega(E)$ microstates that is inherent to system of the structure. As a result, only $E - TS$ of the total energy E present in the system is "free" to perform tasks such as mechanical work.

We conclude our discussion of the constant temperature and volume ideal gas by mentioning that only some gas-phase chemical reactions occur in fixed-volume containers. Depending on experimental setup, gases can also exist at ambient pressure and expand and contract depending on the number of gas molecules yielded or consumed in a reaction, as we will cover in the next section. Interestingly, the constant temperature and volume ideal gas is in fact also a good model for the solute in a solution phase. Although this does not sound very intuitive at first; the basic assumptions still hold, as the microstates associated with each solvated solid particle can still be parameterized by their positions and momenta. The position of any solvated particle can be at any location in the solvent, which has a constant volume, and any solvated particle can roughly have any momentum as it moves through the solvent. The only complication with this is that since solvated particles have to move through the solvent, there is an upper bound on their momentum set by the rate at which solute particles scatter with solvent particles in a process similar to diffusion. Conveniently for us, the microstate occupation probability is $p \propto \exp(-\mathbf{p}^2/(2mkT))$ and the corresponding upper bound on the kinetic energy is very high relative to kT at relevant temperatures (eg. 373 K for water boiling at standard temperature). As a result, we can approximate by integrating over all momenta like we do in the ideal gas case, as the microstates with particle momenta higher than this upper bound are very unlikely to be occupied.

⁹For random variable X with $\langle X \rangle = 0$, its variance is simply $\langle X^2 \rangle = \int x^2 p_X(x) dx$, but the variance of a Gaussian RV can just be read off from argument in the exponential. This gives a simple way to evaluate the integral with care taken to get the correct normalization factor.

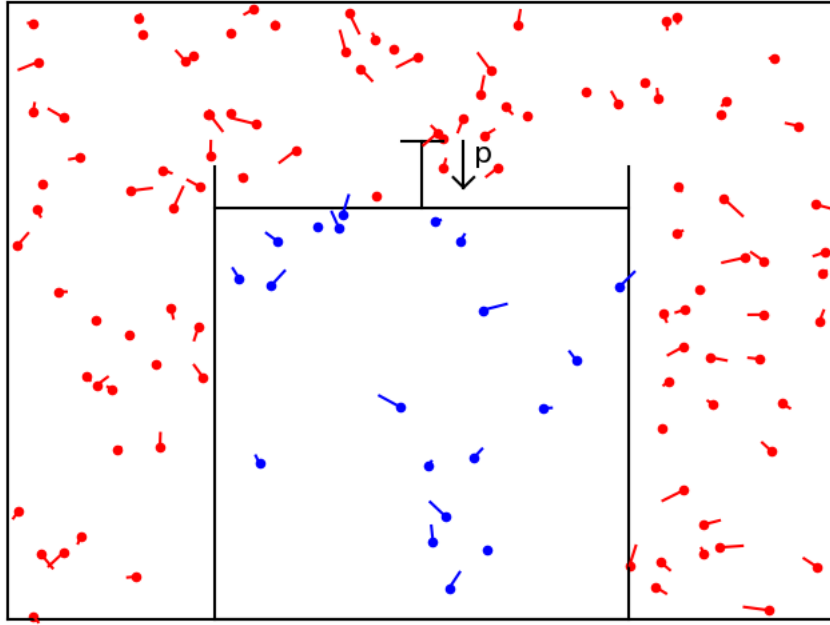


Figure 3: System equilibrated by its environment via both energy and volume exchange between system and environment. Not only is the temperature equal between the system and environment at equilibrium, the pressure must also be equal - if not, the depicted piston height shifts to counteract the side with larger pressure, effectively performing a volume exchange between system and environment.

5 Constant-pressure chemical systems

As we alluded to, gas phase chemical reactions generically occur in a system with variable volume equilibrated against the fixed external atmospheric pressure, not a container with fixed volume. As a result, we need to include the exchange of volume as well as energy between system and environment (Fig. 3). We see that we can derive a probability distribution and free energy expression analogous to the fixed volume case, which use to model chemical systems of real gases.

First, we perform a similar analysis as the one we did at the beginning of Sec. 4 in order to relate the quantity we intuitively understand as pressure to quantities in our statistical model. We assume that the system achieves some equilibrium, entropy-minimizing configuration at volume V . As a result, at V , the total derivative

$$\frac{dS}{dV} = 0 \quad (25)$$

However, S depends on V both explicitly, as increasing V increases the number of allowed positions gas molecules can have which increases Ω , as well as through the total energy E , since expanding the system by dV against the external pressure p leads to energy change $dE = -p dV$ via expansion work. Therefore,

$$\frac{dS}{dV} = \frac{\partial S}{\partial V} + \frac{\partial S}{\partial E} \frac{\partial E}{\partial V} = \frac{\partial S}{\partial V} + \frac{1}{T} \cdot -p = 0 \rightarrow \frac{\partial S}{\partial V} = \frac{p}{T} \quad (26)$$

Thus, much like how T is related to the statistical model through $\frac{\partial S}{\partial E}$, p is related to the statistical model through $\frac{\partial S}{\partial V}$. As a result, the probability distribution of a constant-pressure ideal gas can be derived similarly as in Eqn. 12, except now both energy and volume are partitioned between system and environment. Again, we wish to characterize the equilibrium probability distribution over microstates when we focus on the system only. Associating subscript 1 with the system and 2 with the environment, for microstate μ_1 characterized by energy E_1 and volume V_1 , we have that

$$p(\mu_1) = \frac{1}{\Omega(E, V)} \Omega_2(E - E_1, V - V_1) \quad (27)$$

Similarly expanding $\Omega_2 = \exp\left(\frac{1}{k_B}S(E - E_1, V - V_1)\right)$ gives that

$$p(\mu_1) \propto \exp\left(\frac{1}{k_B}S(E, V) + \frac{\partial S}{\partial E} \cdot -E_1 + \frac{\partial S}{\partial V} \cdot -V_1\right) \propto \exp\left(-\frac{E_1 + pV_1}{kT}\right) \quad (28)$$

where we use the relation $\frac{\partial S}{\partial V} = \frac{-p}{T}$ with opposite sign as previously derived since here, p represents external pressure compressing the system, so an increase in the system volume V corresponds to additional energy being borrowed from the reservoir. Again, we define the total partition function as

$$Z = \sum_{\mu_j} \exp\left(-\frac{1}{kT}(E_i + pV_i)\right) \quad (29)$$

To make evaluating this sum a bit easier, we first group together microstates with the same volume V :

$$Z = \sum_V \sum_{\mu_j \text{ with volume } V} \exp\left(-\frac{1}{kT}(E + pV)\right) = \sum_V e^{-\frac{pV}{kT}} \sum_{\mu_j} \exp\left(-\frac{E}{kT}\right) \quad (30)$$

Now, the inner sum is nothing more than the previously computed partition function at fixed V , which we will call Z_V . Then, converting the sum to an integral over the continuous space of possible system volumes, we arrive at

$$Z = \int_0^\infty dV e^{-\frac{pV}{kT}} Z_V \quad (31)$$

which converges as Z_V experiences an exponential suppression in the integrand. Then,

$$Z = \int_0^\infty dV e^{-\frac{pV}{kT}} \frac{1}{N!} V^N \cdot 2^N \cdot (2\pi m kT)^{3N/2} = C \int_0^\infty e^{-\frac{pV}{kT}} V^N dV \quad (32)$$

for a temperature-independent constant C , which is just a gamma function integral! Even more nicely, we end up with $\Gamma\left(\frac{pV}{kT}\right) = \Gamma(N) \approx N!$ by the ideal gas law, which cancels out the $\frac{1}{N!}$ prefactor. Thus, the overall partition function is

$$Z = \left(\frac{2}{p}\right)^N (2\pi m)^{3N/2} (kT)^{5N/2} \quad (33)$$

The analogous quantity to the free energy F (19) is the Gibbs free energy G , which is given a separate name for historical reasons but behaves similarly. The probability of obtaining any macrostate (E, V) is $\Omega(E, V) \exp\left(-\frac{1}{kT}(E + pV)\right) = \exp\left(-\frac{1}{kT}(E + pV - TS)\right)$. By the same logic associated with the saddle point method, we have $p((E, V) = (E^*, V^*)) = 1$ for (E^*, V^*) minimizing $E + pV - TS$, and $p((E, V) = 0$ otherwise. Then, for G to be the system quantity that must be minimized in the equilibrium state, we define

$$G = E + pV - TS \quad (34)$$

such that at equilibrium, the Gibbs free energy G^* satisfies

$$G^* = -kT \log Z \quad (35)$$

where Z now sums over microstates with different energy and volume. In particular for our ideal gas,

$$G^* = NkT(\log p - \frac{5}{2} \log T + G_g) \quad (36)$$

for some p, T independent constant G_g .

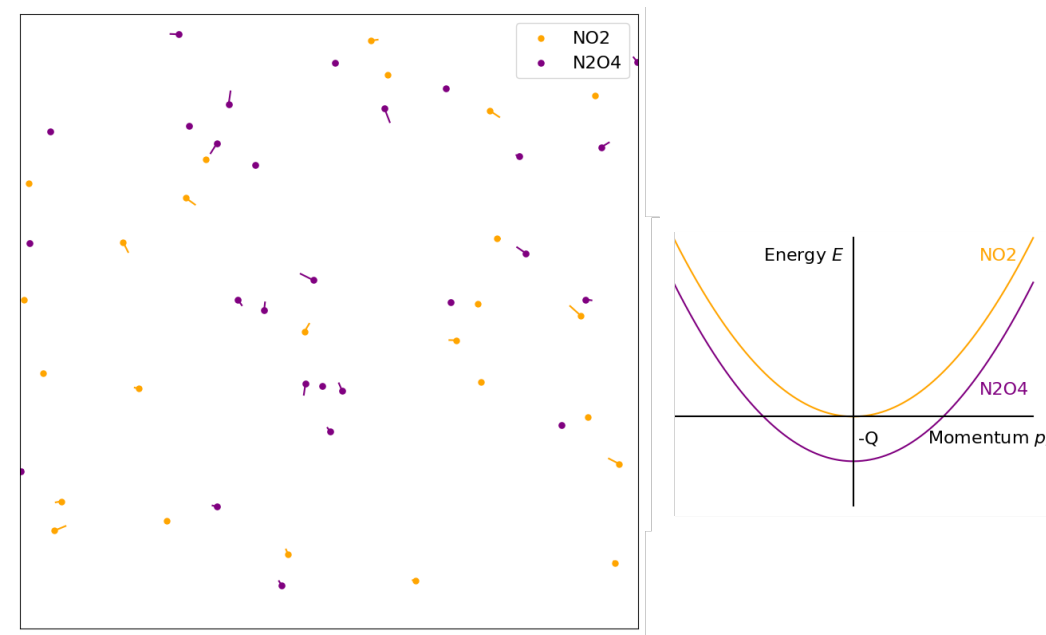


Figure 4: To model a chemical reaction containing two species A , B as products and reactants, we consider the ensemble over all possible microstates of N_A molecules of A and N_B molecules of B . To account for the physical reality that some energy is stored in the bond formed when two B molecules join together to form one A molecule, we introduce the $-Q$ offset such that for molecules of A and B with identical kinetic energy, the molecule of A has an extra $-Q$ contribution towards the total energy.

6 An example constant volume gas-phase reaction

At this point, we have developed the tools we need to describe the measurable outcomes of chemical reactions. In particular, we have demonstrated how the uniform distribution over all system-bath microstates leads to an emergent deterministic free energy on the macroscopic scale that governs the behavior of chemical equilibrium. Now, we can take the analysis in many directions - for example, as you may have noted already, the equilibrium free energy $G^* = -kT \log Z$ is intimately tied to reaction standard free energy differences $\Delta G^o = -RT \log K$ and the $\frac{3}{2}$ and $\frac{5}{2}$ coefficients on the $\log T$ term in the constant volume and constant pressure free energy is similarly related to the ideal gas heat capacity in both conditions. Here, we choose to use the statistical framework to evaluate the behavior of a specific dissociation reaction



at constant volume and point out aspects the derivation that are generalizable for all gas-phase chemical reactions.

To treat this system with the framework of the constant volume ideal gas that we previously developed, we model N_2O_4 as ideal gas A and NO_2 as ideal gas B . TO account for the difference in chemical bonding between N_2O_4 and NO_2 , we impose that ideal gas A has a $-Q$ energy offset per molecule from ideal gas B . This offset arises due to the N–N covalent bond holding together the two individual NO_2 units in the N_2O_4 molecule. Due to the bond, the electrons associated with the N–N bond in N_2O_4 have a lower energy than the resulting electrons localized on the central N atom when the bond dissociates to form NO_2 (Fig. 4). Although this modelling is very simplistic for the real gases involved in the reaction, as we neglect intermolecular forces, quantum structure of electrons and nuclei, etc., we proceed to show that it is typically sufficient to extract all the details needed in many introductory-level chemistry problems with small modifications.

In order to understand the behavior equilibrium state in the dissociation reaction $A(\text{g}) \longleftrightarrow 2 B(\text{g})$, we first consider the free energy of the system under an arbitrary composition of N_A molecules of A and N_B molecules

of B . Then, we examine which ones lead to a minimum in G , corresponding to microstates that are occupied with nontrivial probability at equilibrium. Finally, we try to determine what, if any, measurable macroscopic properties these microstates have in common. For any system microstate, specified by $\{\mathbf{x}, \mathbf{p}\}$ of all A and B molecules, its contribution to the partition function is

$$\exp \frac{1}{kT} \left(N_A Q + \sum_{i_A=1}^{N_A} -\frac{\mathbf{p}_{i_A}^2}{2m} + \sum_{i_B=1}^{N_B} -\frac{\mathbf{p}_{i_B}^2}{2m} \right) \quad (38)$$

where we have added the offset of $-N_A Q$ to the energy of all microstates with N_A molecules of A . Then, summing over all possible positions and momenta of A and B molecules that characterize a system microstate $\mu_j = \{\{\mathbf{x}_{i_A}, \mathbf{p}_{i_A}\}, \{\mathbf{x}_{i_B}, \mathbf{p}_{i_B}\}\}$, we note that the partition function can actually be factored into a product of sums over microstates of a system only with gas A and gas B respectively:

$$Z_{\text{total}} = \sum_{\mu_j} \exp \frac{1}{kT} \left(N_A Q + \sum_{j_A=1}^{N_A} -\frac{\mathbf{p}_{j_A}^2}{2m} + \sum_{j_B=1}^{N_B} -\frac{\mathbf{p}_{j_B}^2}{2m} \right) = \\ \left(e^{-N_A Q/kT} \sum_{\{\mathbf{x}_{j_A}, \mathbf{p}_{j_A}\}} \exp \frac{1}{kT} \sum_{j_A=1}^{N_A} -\frac{\mathbf{p}_{j_A}^2}{2m} \right) \left(\sum_{\{\mathbf{x}_{j_B}, \mathbf{p}_{j_B}\}} \exp \frac{1}{kT} \sum_{j_B=1}^{N_B} -\frac{\mathbf{p}_{j_B}^2}{2m} \right) \quad (39)$$

This is because each unique microstate of the whole system can be described by a unique microstate of gas A and a unique microstate of gas B , and all pairs of (A microstate, B microstate) describe one unique microstate of the whole system. As a result, the equilibrium free energy $F^* = -kT \log Z_{\text{total}}$ is simply the sum of the free energies of the two ideal gases, plus a term due to the energy offset:

$$F^* = -N_A Q + -N_A kT(-\log N_A + 1 + \log V + \log 2 + \frac{3}{2} \log 2\pi m kT) + \\ -N_B kT(-\log N_B + 1 + \log V + \log 2 + \frac{3}{2} \log 2\pi m kT) \quad (40)$$

Although this is the equilibrium free energy for a system with arbitrary composition of N_A and N_B for non-interacting A and B , in reality the chemical reaction between A and B implies that only N_A and N_B such that $N_A + \frac{1}{2}N_B = N$ for some fixed constant N are allowed due to matter conservation. As a result, F^* should be thought of as a function of N_A , which changes as the reaction proceeds: N_B varies along with N_A due to the matter balance constraint and generic system compositions yield different equilibrium free energies $F^*(N_A, N_B = 2(N - N_A))$. Thus, the true equilibrium free energy in a system where A and B react must also extremize F with respect to N_A by a similar saddle point argument to before - not only does the free energy behave deterministically upon averaging over all system-bath microstates for two isolated ideal gases A and B , but when they are brought together to react, the deterministic equilibrium state is arrived at by minimizing F over all variables that become free upon reaction. **This strategy of writing down the total free energy of all components in the system as if they were non-interacting, then solving the optimization problem over the constraints due to the interaction, is generally successful for analyzing chemical systems.**

Following this approach, we drop the $*$ notation on the free energy and solve for the overall minimum F by imposing the condition

$$dF = 0 = \frac{\partial F}{\partial N_A} dN_A + \frac{\partial F}{\partial N_B} dN_B \quad (41)$$

combined with the mass balance condition

$$N_A + \frac{1}{2}N_B = N \rightarrow -2 dN_A = dN_B \quad (42)$$

implies

$$\frac{\partial F}{\partial N_A} = 2 \frac{\partial F}{\partial N_B} \quad (43)$$

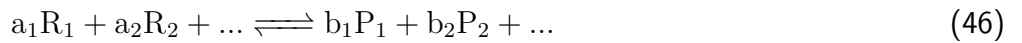
Thus,

$$-kT[-\log N_A + \log V + \frac{3}{2} \log 2\pi m_A kT + \log 2] - Q = -2kT[-\log N_B + \log V + \frac{3}{2} \log 2\pi m_B kT + \log 2] \quad (44)$$

Letting $n_A = \frac{N_A}{V}$ and same for B , rearranging gives

$$\log \frac{n_B^2}{n_A} = \alpha(T), \text{ where } \alpha(T) = \frac{-Q}{kT} + \log 2 - \frac{3}{2} \log 2\pi kT \frac{m_B^2}{m_A} \text{ is a temperature-dependent constant} \quad (45)$$

Therefore, all equilibrium microstates of the composite A-B system have $\frac{n_B^2}{n_A} = e^{\alpha(T)}$. If we define $K = e^{\alpha(T)}$, this is nothing more than the equilibrium constant formulation of the equilibrium position of a chemical reaction! In particular, our statistical model provides us with mathematical justification for why the form of the equilibrium constant is the way that it is - the deterministic quantity characterizing a chemical reaction at equilibrium for any reaction



will always be

$$K = \frac{\prod n_{P_i}^{b_i}}{\prod n_{R_i}^{a_i}} \quad (47)$$

not other expressions such as $(\sum n_{P_i})/((\sum n_{R_i}) + \sum n_{P_i})$ (constant yield), $(\sum b_i n_{P_i}) - (\sum a_i n_{R_i})$ (constant stoichiometric difference), or any of the other literally infinite possibilities of functions on the concentrations and coefficients of products and reactants.

To understand how characterizing the equilibrium state with the deterministic relation $K = \frac{\prod n_{P_i}^{b_i}}{\prod n_{R_i}^{a_i}}$, we substitute $N_A = N - N_B$ to solve for the amount of B at equilibrium:

$$N_{B,eq}(T, V, N) = \frac{\sqrt{(KV)^2 + 4NKV} - KV}{2} \quad (48)$$

LeChatelier's principles

From here, we can characterize how $N_{B,eq}$ changes with the system parameters V, N , and T . In high school chemistry, these rules are commonly presented as LeChatelier's principles but presented without much justification.

The first LeChatelier's principle we will consider is the dependence of equilibrium composition on the volume V of the reaction vessel. Students are taught that increasing the volume of the reaction vessel shifts the equilibrium composition to favor the reaction direction that produces more gas molecules and the associated intuition is often along the lines that there is more "room" in the vessel to fit more molecules of gas. We can confirm this by calculating

$$\frac{\partial N_{B,eq}}{\partial V} = \frac{1}{2} \left(-K + K \frac{KV + 2N}{\sqrt{(KV)^2 + 4NKV}} \right) > \frac{1}{2} \left(-K + K \frac{KV + 2N}{\sqrt{(KV)^2 + 4NKV + 4N^2}} \right) = 0 \quad (49)$$

for a box with finite volume V . Thus, we indeed have that in the general case, increasing V shifts the distribution of products at equilibrium to prefer N_B . More rigorously, the previous few pages of derivation tell us that the reason for this is because when we increase V , the $N \log V$ contribution to the free energy corresponding to accessing more position microstates increases more than the free energy decrease from the $-N \log N$ penalty due to the indistinguishability of B molecules.

The second LeChatelier's principle we consider is the dependence of the equilibrium composition on initial reactant concentration. As previously noted, our statistical interpretation of the equilibrium state says nothing about the approach to equilibrium from initial nonequilibrium states. Therefore, we can just treat an increase in initial reactant concentration as an increase in N and letting the thermalization process that we have black-boxed away take care of differentiating the new molecules into N_A or N_B . Then, a similar calculation tells us

$$\frac{\partial N_{B,eq}}{\partial N} = \frac{KV}{\sqrt{(KV)^2 + 4NKV}} \quad (50)$$

which is clearly always nonnegative. Again, we confirm the notion that adding reactant shifts the reaction equilibrium to the right.

Finally, we consider the equilibrium dependence on temperature. Since the reaction is endothermic, LeChatelier's principles suggest that increasing the temperature shifts the equilibrium position to the right. Again, we can confirm this by checking

$$\frac{\partial N_{B,eq}}{\partial K} = \frac{1}{2} \left(\frac{KV^2 + 2NV}{\sqrt{(KV)^2 + 4NKV}} - V \right) > 0 \quad (51)$$

and

$$\frac{\partial \log K}{\partial T} = \frac{Q}{kT^2} + \frac{3}{2T} > 0 \quad (52)$$

at all temperatures T . Therefore, $N_{B,eq}$ increases with temperature. However, the common intuition presented with this rule - that "heat should be treated as a reactant and increasing temperature provides heat, reducing the treatment to the second case" - is shown to be incorrect using the statistical view. Instead, the role of temperature is primarily through the dependence of $\log K(T) \sim \frac{-C}{kT}$; ie., the free energy offset between the reactants and the products. Since higher temperatures make reactant microstates, which initially lie at a lower free energy due to the $-Q$ offset, more uniformly distributed in energy,¹⁰ reactant molecules become more likely to have a free energy matching typical product free energies at higher temperature. Thus, it is more accurate to say that at lower temperature, the equilibrium position favors lower energy states, which is dominated by offsets like $-Q$ here between reactants and products, while at higher temperature, the equilibrium position is associated with a more intermediate position between reactants and products which maximizes the number of total microstates.¹¹

7 Putting it all together

At this point, we've introduced all the mathematical machinery to analyze reactions at an introductory level. To conclude, let's discuss the extended example of a chemical equilibrium between a liquid and a gas with the wrinkle that there is some material solvated in the liquid (Figure 5). Not only is this equilibrium relevant in real life, such as in making food with a pressure cooker, it is also a chemical scenario that ties together all of our previous discussion.

Now, we have three chemical species we need to keep track of in our system: the gas phase and liquid phase that are in equilibrium, as well as the solvated phase in the liquid. We treat the gas phase as an ideal gas under constant pressure. The solvated solid phase is treated as a constant volume ideal gas as previously discussed, with volume $V = \rho N_l$ for N_l solvent atoms and "density" ρ . Furthermore, we model the liquid as having a

¹⁰For states separated in energy by ΔE , their ratio of the probability density according to the Boltzmann distribution is $\exp(-\Delta E/kT)$. Thus in the temperature regime $kT \gg \Delta E$, the states are associated with roughly equal probability density

¹¹Note that there are many more microstates in the intermediate regime of equilibrium position due to the permutation of distinguishable reactants and products which we do not have if the equilibrium position lies far to the side of reactants or products.

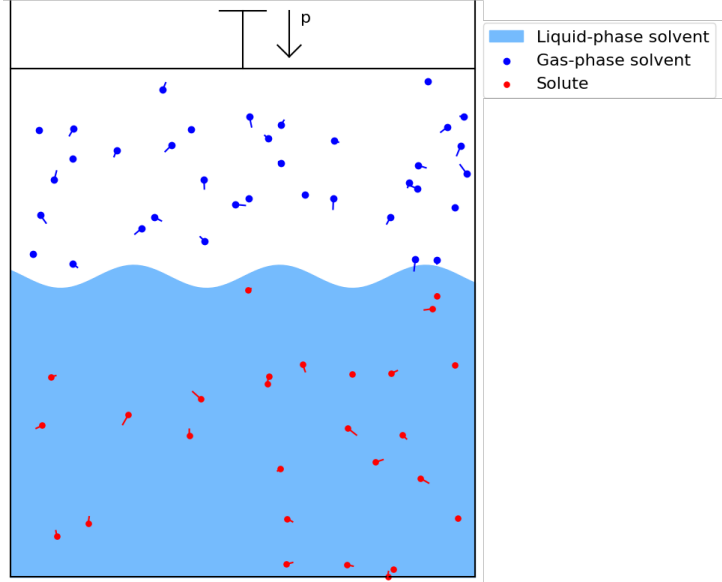


Figure 5: We consider the equilibrium between the liquid and gas phases of a solvent such as water with the restriction that there is a solute in the liquid phase. If the liquid phase evaporates, there is less volume for solute molecules to occupy and thus a lower number of solute microstates and lower entropy. As a result, we show that the liquid-gas equilibrium shifts to favor the presence of more liquid at any given pressure and temperature.

fixed, temperature independent free energy per particle F_l . This is because atoms in a liquid flow much more collectively than in the gas phase due to strong forces between particles. As a result, atoms in a liquid have similar momenta across some finite length scale. Thus, if we try to perform a similar integral as in (Eqn. 24) over all liquid microstates, we integrate over a very small portion of the overall momentum space compared to the gas phase in which particles can have any momentum. For example, if the length scale enforced that the 100 nearest atoms to each liquid particle had the same momentum,¹² the number of free variables in the integral and thus the free energy is reduced by a factor of $\frac{1}{100}$. As a result, $\Omega_{\text{gas}} \gg \Omega_{\text{liquid}}$, implying that the gas phase has a much stronger temperature-dependent free energy than the liquid phase, justifying the approximation that the liquid phase has a temperature-independent free energy per particle F_l . The origin of such nonzero F_l energy typically arises from intermolecular attractions ($F_l < 0$) preserving the collective nature of the liquid, which are not present in the gas phase.

Therefore, using the expressions for the free energy of an ideal gas at constant volume and constant pressure respectively, the total system free energy is¹³

$$F_{\text{total}} = N_s kT (\log N_s - \log \rho N_l + F_s) + N_l F_l + N_g kT (\log p - \frac{5}{2} \log T + F_g) \quad (53)$$

Since we consider the liquid-gas equilibrium of our system, we constrain that the system has a total of $N = N_g + N_l$ particles partitioned between the two phases. Again applying the principle that deterministic properties of the equilibrium state are arrived through minimization of the free energy along the constraint introduced by physical process, we write that

$$\frac{dF_{\text{total}}}{dN_g} = \frac{\partial F_{\text{total}}}{\partial N_g} + \frac{\partial F_{\text{total}}}{\partial N_l} \frac{\partial N_l}{\partial N_g} = 0 \quad (54)$$

¹²a very short length scale in practice

¹³We have abstracted away irrelevant constant terms into F_s and F_g in preparation of what is to come

at the equilibrium state, and $\frac{\partial N_l}{\partial N_g} = -1$. Therefore, we have that

$$kT \log p - \frac{5}{2}kT \log T = C - \frac{N_s kT}{N_l} \quad (55)$$

for a temperature independent constant C , or

$$\frac{5}{2} \log T - \frac{C}{kT} = \log p + \frac{N_s}{N_l} \quad (56)$$

First, let us consider what happens when there is no solute, where the condition to have a liquid-gas equilibrium reduces to

$$\frac{5}{2} \log T - \frac{C}{kT} = \log p \quad (57)$$

Therefore, at any given value of the external pressure p , there is a characteristic temperature T at which the liquid and gas phases exist in equilibrium in our pressure-regulated box. In the language associated with the study of phase transitions, we call this characteristic temperature the phase transition temperature T_c . Although the study of phase transitions and condensation in real gases is very rich and merits at least its own separate discussion,¹⁴ we choose to present the extremely simplified analysis here that the free energy associated with the gas phase decreases as $c_1 T + c_2 T \log T$ (Eqn. 57) for c_1, c_2 while the free energy associated with the liquid phase remains roughly constant (Figure 6). Thus, we see that the conversion of liquid to gas has positive free energy change for $T < T_c$ and negative free energy change when $T > T_c$. This describes the qualitative behavior of the system well: at low temperatures, the equilibrium state is established by the competitive balance between molecules incurring the energy benefit of being condensed liquid phase and the entropic benefit of being in the gas phase. However, at sufficiently high temperature, the fixed free energy benefit per particle to be in the liquid phase F_l is too small such that any particle would contribute towards a lower free energy by moving to the gas phase. At $T = T_c$ where the free energy of the two phases are equal, implying that the thermodynamically preferred equilibrium distribution have microstates with a liquid-gas composition with gas pressures that match the external pressure p .

Conversely, we can look at (57) with the mindset that T is fixed. Then, the equation suggests that at any temperature T , there exists a characteristic external pressure $p \propto \exp\{-c/T\}$ such that the internal pressure of the gas phase at the equilibrium liquid-gas composition is also p . This characteristic value of p is known as the **vapor pressure** and is commonly used as a proxy for solvent volatility. For example, the vapor pressure of water at 25°C is about 0.03 atmospheres, while the vapor pressure of isopropanol at the same temperature is almost double that. This explains why the smell when opening a bottle of rubbing alcohol is so pungent - there are many more alcohol molecules in the gas phase compared to water molecules in the headspace above the liquid.

Now, adding back in a small amount of solute $N_s > 0$, we add a $-T \log N_s/N_l$ contribution to the free energy of the liquid phase, which shifts the transition temperature from T_c to $T_c + \Delta T$ (Figure 7). Considering our original expression to the first order Taylor expansion around T_c , we have

$$\frac{5\Delta T}{2T_c} + \frac{c\Delta T}{T_c^2} = \frac{N_s}{N_l} \quad (58)$$

Treating the $c_1 T + c_2 T \log T$ expression as simply a linear term $c_3 T$ in this temperature regime gives

$$\Delta T = \frac{T_c^2}{c_3} \frac{N_s}{N_l} \quad (59)$$

is the **boiling point elevation** of the solvent, proportional to the concentration of solute particles.

¹⁴or even a whole scientific career

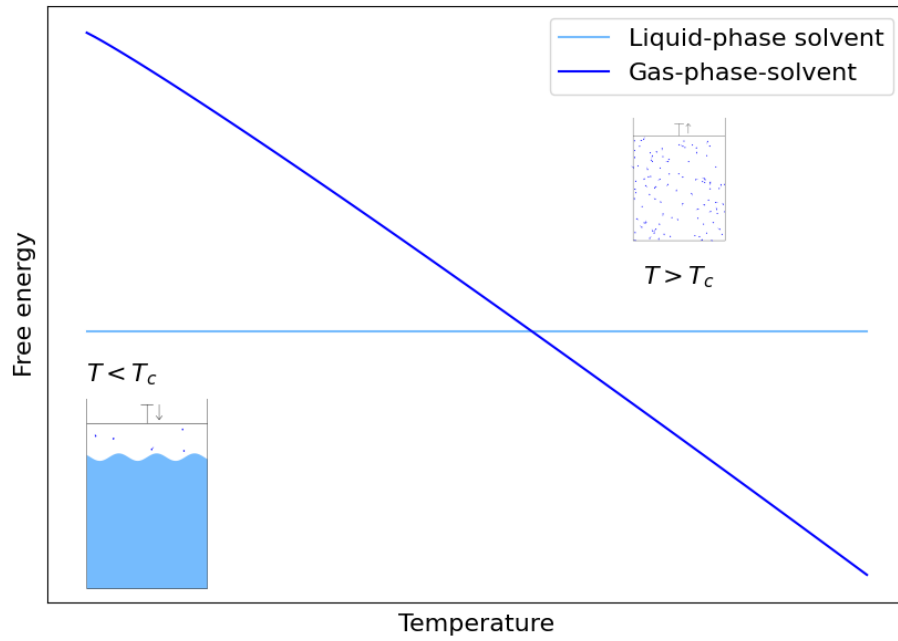


Figure 6: Free energy of the solvent in the liquid phase and gas phase. In the regime associated with most real solvents, the combination of $|c_1| > |c_2|$ and temperatures $T \sim 300$ K imply that the gas free energy can be treated as decreasing linearly as depicted in the graph. However, it's still important to know that under the hood, there is both a T and $T \log T$ contribution to the free energy due to how the ideal gas entropy scales with temperature.

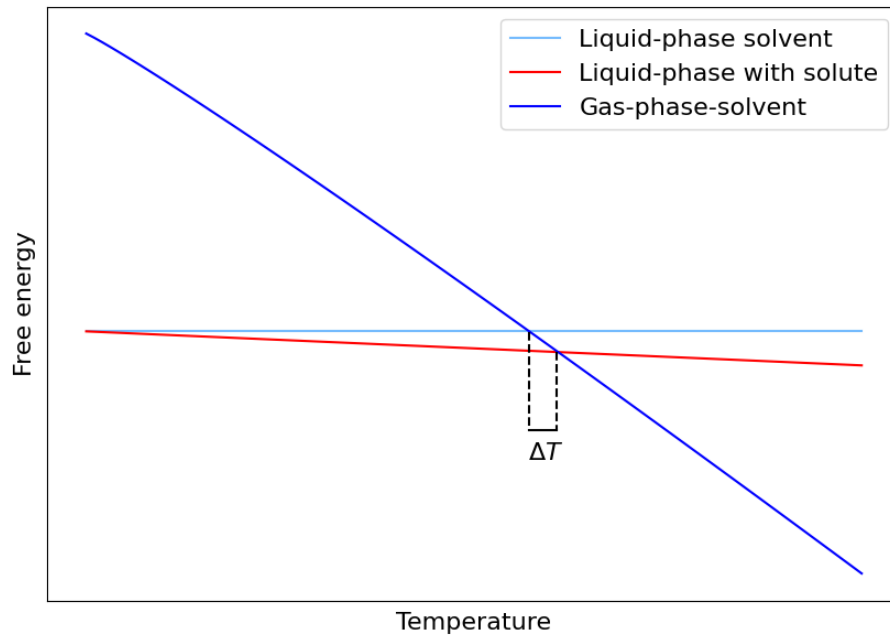


Figure 7: Upon decrease in the free energy of the solution phase due to the introducing microstates parameterized by solute positions and momenta, there is a characteristic increase ΔT in the phase transition temperature between liquid and gas.

This is a classic example of a **colligative property** which is introduced in high school chemistry classes, which is commonly defined as a property of a solution that only depends on the solute concentration and not the solute identity. However, we note that there is one complication with this definition that is revealed through the statistical interpretation of thermodynamics. In our previous assumption, we assumed that all solute particles were **indistinguishable** ideal gas molecules. This seems fine at face value assuming you buy that molecules moving through a solvent behave similar to an ideal gas, until we note many solutes, in particular ionic solutes, do not maintain a single-molecule structure as they are dissolved in a solvent. For example, N molecules of sucrose dissolve to form N solvated sucrose molecules in water, but $\frac{N}{2}$ units of table salt NaCl dissolve to form $\frac{N}{2}$ Na^+ and $\frac{N}{2}$ Cl^- ions. Although our analysis would treat both cases as N ideal gas molecules in solvent, we incur an $N!$ factor due to overcounting equivalent microstates in the first case, while we incur an overall factor of $(N/2)!^2$ in the second case due to the distinguishability of sodium and chloride ions. Thus, the free energy $F \propto -\log Z$ in the second case incurs a $2 \cdot N/2 \log N/2$ term compared to $N \log N$ in the first case per Stirling's approximation. Of course, the difference of $N \log 2$ is subdominant to $N \log N$ in the $N \sim 10^{23}$ regime that characterizes physically-sized systems, which is why this complication is ignored when colligative properties are introduced in a classroom setting. However, it is important to note that such a discrepancy does indeed exist and would be observable given a sufficiently small system.

8 Conclusions

In this journey through viewing equilibrium chemistry through a statistical approach, we began with the fundamental principle of viewing equilibrium chemical systems a probability distribution. We used our postulate of entropy maximization over system and environment to establish equilibrium distributions for the constant volume and constant pressure ideal gases. In the second half of our discussion, we focused on extended examples in modelling the equilibrium state of two chemically relevant systems to extract additional understanding from topics typically introduced in high school chemistry. Throughout, we emphasized that observable thermodynamic equilibria are emergent from shockingly simple probability distributions over microstates. In particular, many thermodynamic quantities that we observe in chemistry experiments actually turn out to be deterministic due to the thermodynamic limit as exemplified by the saddle point integral.

Using the formalism described in this chapter, one can show why a fixed amount of solid can be dissolved in solution and no matter how much additional solid is added past saturation none will be dissolved; or why universal pH indicators are often made by simply mixing solutions of individual pH indicators. Indeed, this approach for analyzing chemical systems can be extended to nearly all equilibrium chemical phenomena with quite different experimentally observable outcomes. It is quite remarkable to me that they can all be quantitatively described by the free energy approach to the entropy maximization postulate of the system and environment. More extended discussion on these phenomena would certainly be insightful, however they are not included for the sake of ending the chapter somewhere. I encourage the interested reader to work out these examples independently with this framework.

Please feel to reach out via email (listed on the main site) with any suggestions or thoughts for discussion, I would love to hear from anyone who has any thoughts/feedback/advice about the content here. In particular, I am most interested in honest and direct criticism on points to improve the writing and strengthen the argument.