Notes on Statistical Physics, PY 541

Anatoli Polkovnikov, Boston University

(Dated: December 9, 2008)

These notes are partially based on notes by Gil Refael (CalTech), Sid Redner (BU), Steven Girvin (Yale), as well as on texts by M. Kardar, Statistical Mechanics of Particles and L. Landau and L. Lifshitz, Statistical Physics vol. V.

Contents

I. LIMITATIONS OF MICROSCOPIC APPROACHES.

- We deal with very large amounts of atoms in everyday scale 10^{30}. It is impossible to solve these many equations of motion, it is even harder to define the initial conditions.
- Most of dynamics we are dealing with is chaotic. Chance plays a very important role everywhere around us. This means that the outcome of a certain process is very susceptible to slightest changes in the initial conditions in external perturbations etc. Chaos: information which is not significant at present determines the future. Usually chaos is opposite to dissipation associated with the loss of information. Connections of statistical physics and information is coming!
- Quantum mechanics is even more complex. The Hilbert space is exponentially large. Take e.g. 100 particles in 200 lattice sites. The size of the Hilbert space for (spinless) fermions is $200!/(100!)^2$ 10^{50} and for bosons $300!/(200!100!)$ 10^{80} for bosons. The amount of particles in the Universe is 10^{50}. It is fundamentally impossible to simulate dynamics microscopically for even this small system. Also the typical level spacing is exponentially small: impossible to resolve levels during life time of universe. No Schrödinger cat state is possible!!! Also any tiny perturbation will can mix these levels - quantum chaos.
- Yet we have compelling evidence that often phenomenological macroscopic approaches work very well: fluid and gas dynamics - airplanes and, newton's equations for collective degrees of freedom and phenomenological friction - cars, trains, etc, phenomenological Ohm's laws (based on kinetic equations) - electronics, many predictions based on statistics and probability in biology and chemistry and so on.

The main subject of statistical physics is understanding the connection between microscopics and behavior of large macroscopic systems. Many things are already understood but many still remain a mystery even now. Despite statistical physics is so old, it is now an area of very active research by many people.

II. PROBABILITY AND STATISTICS

Good book Feller. Suppose we have a set $\mathcal{S}$, discrete or continuous. Probability $p(s \in \mathcal{S})$ is a map to real axis, which satisfies the following properties:

- $p(s) \geq 0$ - probability can not be negative.
- $p(\mathcal{S}) = 1$ - we will get some outcome withe the probability one
- If $s_1 \bigcap s_2 = 0$ then $p(s_1 | s_2) = p(s_1) + p(s_2)$.

One has to differentiate between probability and statistics. The latter is the analysis of events which already happened. Example toss a coin. Before the process the probability of "head" or "tail" is 50%. However after the coin is tossed it is definite outcome. Physics definition: probability is a preassigned chance of a certain outcome, statistics is what we measure.

Intuitive definition of probability: do many (N) identical experiments. Suppose that the outcome s happened N_s times. Then probability of N_s is

$$p(s) = \lim_{N \to \infty} \frac{N_s}{N}. \tag{1}$$

A. Discrete Sets.

Assume now that our set $\mathcal{S}$ is a discrete set of numbers. Then the normalization requires that

$$\sum_s p(s) = 1. \tag{2}$$

Next we introduce some important concepts. Expectation value of s or mean or average is

$$\overline{s} = \sum_s s p_s. \tag{3}$$

This average satisfies simple equalities:

$$\overline{s_1 + s_2} = \overline{s}_1 + \overline{s}_2, \tag{4}$$

If a is some constant then

$$\overline{as} = a\overline{s}. \tag{5}$$

Variance is defined as

$$\delta s^2 = \overline{(s-\overline{s})^2} = \overline{s^2} - (\overline{s})^2 \tag{6}$$

Very important result to remember. Variance tells us about the width of distribution. Note that variance is in general not additive:

$$\overline{(s_1+s_2)^2} - (\overline{s_1+s_2})^2 = \delta s_1^2 + \delta s_2^2 + 2\overline{s_1 s_2} - 2\overline{s_1}\,\overline{s_2}. \tag{7}$$

The two events are called independent if $\overline{s_1 s_2} = \overline{s_1}\,\overline{s_2}$. Then the variance is additive. Example what is the probability that we will get a sequence a "head" and a "tail". The answer is obviously $1/4 = 1/2 * 1/2$. These two event are independent. What is the probability that in a sequence of three tosses we will get (i) 2 out of 3 heads and (ii) the last toss is the head. By simple counting we will get that $p(i-ii) = 1/4$ while $p(i) = 3/8$ and $p(ii) = 1/2$. Note $p(i-ii) > p(i) * p(ii)$ because these events are not independent. In fact one can prove that $p(i-ii) = p(i) * p(ii|i) = 3/8 * 2/3 = 1/4$. The quantity $p(ii|i)$ is called a conditional probability: what is the probability of ii given i happened.

Higher moments of s are defined as expectation values of corresponding powers of s. The best (unbiased) statistical estimate for the variance:

$$\delta s^2 \sim \frac{1}{M-1} \sum_{i=1}^{M-1} \left(s_i - \frac{1}{M} \sum_j s_j \right)^2. \tag{8}$$

Indeed if we compute expectation value of δs^2 we find:

$$\overline{\delta s^2} = \frac{1}{M-1} \sum_i \overline{s_i^2} - \frac{1}{M(M-1)} \sum_{i,j} \overline{s_i s_j} = \overline{s^2} - (\overline{s})^2. \tag{9}$$

Examples of discrete distributions.

Binomial distribution. Imagine that we have a process (a coin toss) with the probability of success p and probability of failure q such that $p + q = 1$. Note that for the coin toss $p = q = 1/2$. Suppose we perform the process N times. What is the probability of finding n successes and $N-n$ failures? First let us solve a simpler problem. What is the probability of having first n successes and next $N-n$ failures? The answer is obviously $p^n q^{N-n}$. Now let is find the number of independent configurations of n successes and $N-n$ failures, each having the same probability. First let us choose the positions for "successes". The number of ways we can distribute them is $N(N-1)\dots(N-n+1)$. But there is over-counting so we need to divide by $n!$ because all permutations within successes are equivalent so the net result is

$$P(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}. \tag{10}$$

Note that $\sum P(n) = (p+q)^N$ so the name is binomial distribution. The coefficients are the binomial coefficients. Let us compute expectation value and variance:

$$\overline{n} = \sum_n n \frac{N!}{n!(N-n)!} p^n q^{N-n} = Np \sum_n \frac{(N-1)!}{(n-1)!(N-1-(n-1))!} p^{n-1} q^{(N-1)-(n-1)} = pN \tag{11}$$

as it should be.

$$\overline{n}^2 = \sum_n n^2 \frac{N!}{n!(N-n)!} p^n q^{N-n} = \sum_n (n-1+1) \frac{N!}{(n-1)!(N-n)!} p^n q^{N-n} = N(N-1)p^2 + Np \tag{12}$$

Therefore

$$\delta n^2 = Np(1-p) \tag{13}$$

Note that $\delta n/\overline{n} = \sqrt{\frac{1-p}{p}} \frac{1}{\sqrt{N}} \to 0$. This justifies intuitive definition of probability. If $N \to 0$ the distribution is highly peaked near the average. This is a cartoon version of the central limit theorem which plays the key role in statistical physics.

Poisson distribution. This is another very important distribution in statistics. It describes variety of phenomena from radioactivity to statistics of defective items. It can be obtained from Poisson distribution when $p \to 0$, $N \to \infty$ such that $pN = \text{const}$. Derivation: assume that the probability of decay of a nucleus per unit time is Γ. What is the probability that we will detect n decayed nucleus during time t. Solution:let us split the interval t into subintervals δt and let $M = t/\delta t$ to be the number of such intervals. The probability that there will be decay during a particular interval is $\delta p = \Gamma \delta t$. Correspondingly the probability there is no decay is $1 - \Gamma \delta t$. First let us evaluate the probability that the the nucleus does not decay at all:

$$P_0(t) = (1 - \Gamma \delta t)^M = (1 - \Gamma t/M)^M \to \exp[-\Gamma t] \tag{14}$$

This is of course the formula for the radio active decay. Now let us find the probability that n nucleus decayed, we assume that n is small compared to the total number of nucleous

$$P_n(t) = (\Gamma \delta t)^n (1 - \Gamma \delta^t)^{M-n} \frac{M!}{n!(M-n)!} \to \frac{(\Gamma t)^n}{n!} \exp[-\Gamma t] \tag{15}$$

Let us denote $\Gamma t = \lambda$. Then

$$P_n = \frac{\lambda^n}{n!} \exp[-\lambda]. \tag{16}$$

Homework. *Prove that this distribution is normalized. Find the mean and the variance as a function of* λ*. Discuss how the Poisson distribution looks like for large and small values of* λ*.* Example from Rutherford paper of 1920, flying bomb hits London, statistics of permutations in chromosomes after X-ray radiation and many more: pages 160-161 from Feller.

Homework. Problems from Feller.

1. How many random digits one needs to get in order that the probability to have at least one "7" there is 90%?

2. What is the probability that six random people have birthdays within the same two months and the other 10 months are birthday free?

3. A book of 500 pages contains 100 missprints. Estimate the probability that at least one page contains 5 missprints.

B. Continuous distributions.

Now assume that the set $\mathcal{S}$ represents a continuous set of numbers, e.g. rational or complex numbers. We will use x to denote them. Then the probability satisfies

$$\int_{\mathcal{S}} p(x)dx = 1. \tag{17}$$

For concreteness let us use x to be a 1D set of real numbers. The quantity $p(x)$ is called the probability density function. One also defines the cumulative probability function

$$P(x) = \int_{-\infty}^{x} p(x')dx' \tag{18}$$

This function has following properties: $P(x \to -\infty) = 0$, $P(x \to \infty) = 1$, $P(x)$ is a nondecreasing function of x. Expectation value of any function is defined (as in the discrete case)

$$\langle F(x) \rangle = \int_{-\infty}^{\infty} dx p(x) F(x) \tag{19}$$

Note that if y is a monotonic function of x we can change variables from y to x. The conservation of probability requires that $p(x)dx = p(y)dy$. Therefore

$$p(y) = p(x)\left|\frac{dx}{dy}\right|. \tag{20}$$

In general (for arbitrary nonmonotonic function) one can define $p(y)$ as

$$p(y) = \int_{-\infty}^{\infty} p(x)\delta(y(x) - y)dx. \tag{21}$$

Example: let us take a Gaussian distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{22}$$

Let us change the variables to $y = x^2$. Then according to Eq. (21) we have

$$\begin{aligned} p(y) &= 0 \text{ for } y <= 0 \\ p(y) &= \frac{2}{\sqrt{2\pi y}}\exp[-y/2] \text{ for } y > 0. \end{aligned} \tag{23}$$

A very useful concept is the characteristic function:

$$\tilde{p}(k) = \langle \exp[-ikx]\rangle = \int dx \exp[-ikx]p(x). \tag{24}$$

The probability density is the inverse Fourier transform of the characteristic function. One can use also other transforms, e.g. Bessel function instead of exponential, or any other complete set of functions. Note that the characteristic function is the generator of the moments of the distribution:

$$\tilde{p}(k) = \sum_n \frac{(-ik)^n}{n!}\langle x^n\rangle. \tag{25}$$

Similarly

$$\exp[ikx_0]\tilde{p}(k) = \sum_n \frac{(-ik)^n}{n!}\langle (x - x_0)^n\rangle. \tag{26}$$

The logarithm of the characteristic function generates cumulant expansion of the distribution:

$$\ln \tilde{p}(k) = \sum_n \frac{(-ik)^n}{n!}\langle x^n\rangle_c \tag{27}$$

The first four cumulants have special names: mean, variance, skewness, kurtosis and they are defined in the following way

$$\langle x\rangle_c = \langle x\rangle, \tag{28}$$

$$\langle x^2\rangle_c = \langle x^2\rangle - \langle x\rangle^2, \tag{29}$$

$$\langle x^3\rangle_c = \langle x^3\rangle - 3\langle x^2\rangle\langle x\rangle + 2\langle x\rangle^3, \tag{30}$$

$$\langle x^4\rangle_c = \langle x^4\rangle - 4\langle x^3\rangle\langle x\rangle + 3\langle x^2\rangle^2 + 12\langle x^2\rangle\langle x\rangle^2 - 6\langle x\rangle^4. \tag{31}$$

$$\tag{32}$$

Skewness characterizes the degree of asymmetry of the distribution. Dimensionless number is $\xi = \langle x^3\rangle_c/(\langle x^2\rangle_c)^{3/2}$. These cumulants are similarly defined for discrete distributions. **Homework:** *1. Work out ξ for the binomial distribution as a function of N. What happens to ξ for large N. 2. Repeat for Poisson distribution, look carefully at the limit when the mean becomes large. 3. Find first three cumulants for the exponential distribution: $p(x) = \lambda\exp[-\lambda x]$, $x \geq 0$ and $p(x) = 0$ otherwise.*

Examples of continuous distributions Exponential distribution:

$$p(x) = \lambda\exp[-\lambda x] \tag{33}$$

describes the probability density of a nuclei to decay at moment x (if x is interpreted as a time). The parameter λ plays the role of the decay rate. Normal distribution - probably most important in statistics:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right]. \tag{34}$$

The parameters x_0 and σ^2 are the mean and the variance respectively. It has all other cumulants higher than second equal to zero. Characteristic function

$$\tilde{p}(k) = \exp\left[-ikx_0 - \frac{k^2x^2}{2}\right] \tag{35}$$

so that indeed $\ln\tilde{p}(k) = -ikx_0 - k^2x^2/2$.

One can also define joint probability distribution of multiple variables: $p(x_1, x_2, ...)$. The probability density factorizes if and only if the random variables are independent.

Central limit theorem. Suppose that we are dealing with independent random variables x_i having the same distribution with all moments being finite. Then in the limit when number of these variables is large the distribution of their sum $X = \sum_{i=1}^N x_i$ approaches gaussian distribution with mean Nx_0 and variance $N\sigma^2$, i.e.

$$p(X) \to \frac{1}{\sqrt{2\pi N}\sigma}\exp\left[-\frac{(X-Nx_0)^2}{2N\sigma^2}\right]. \tag{36}$$

The theorem is also true if the variables are not equally distributed then the mean and the variance are the sum of individual means and variances. The theorem is also true for weakly dependent variables, where quantities like $\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle$ decay sufficiently fast with $|i-j|$, i.e. the variables, which are far from each other are almost independent. In this case, however, the variance will not be the sum of the variances.

Sketch of proof. We showed that the mean and the variance are additive for independent variables. In order to show that the distribution approaches gaussian it is sufficient to prove that higher order cumulants of X vanish. Let us define a new variable

$$y = \frac{X - Nx_0}{\sigma\sqrt{N}}. \tag{37}$$

Then clearly $\langle y\rangle_c = 0$, $\langle y^2\rangle_c = 1$. It is straightforward to verify that higher order cumulants satisfy:

$$\langle y^n\rangle_c = \langle x^n\rangle_c \frac{N}{N^{n/2}} \tag{38}$$

They vanish in the limit $N \to \infty$ for all $n > 2$.

Example: Take Poisson distribution

$$p(n) = \frac{\lambda^n}{n!}\exp[-\lambda] \tag{39}$$

Recall that this distribution is obtained from binomial distribution with $\lambda = pN$. We are interested in the limit $\lambda \to \infty$. Use saddle point approximation and Stirling's formula and treat n as a continuous variable:

$$p(n) \approx \frac{1}{\sqrt{2\pi n}}\exp[n\log\lambda - n\log n + n - \lambda]. \tag{40}$$

Now expand near $n = \lambda$ up to second order. Then we get

$$p(n) \approx \frac{1}{\sqrt{2\pi\lambda}}\exp\left[-\frac{(n-\lambda)^2}{2\lambda}\right]. \tag{41}$$

QED.

Application: a random walk. Assume that we have a random walker who randomly makes a step to the left of two the right. Find the probability distribution of a position of a random walker after large number of steps. Solution: Each step changes the position of a walker either by $\Delta x = +1$ or by $\Delta x = -1$ so the mean displacement per step is $x_0 = 0$ and the variance is $\sigma^2 = 1/2 + 1/2 = 1$. By the central limit theorem after $N = t$ (interpret N as time) steps the distribution is Gaussian with the probability

$$p(x) \approx \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{x^2}{2t}\right]. \tag{42}$$

This is the famous formula for diffusion. It generalizes to arbitrary number of dimensions.

Remarks. The central limit theorem is valid only if all variables have finite moments. If this is not the case the sum can still converge but to a non-gaussian Levi distribution. The central limit theorem is useless when we are interested by rare events, i.e.by the tails of the distribution (like maxima). Instead we get completely different extreme value statistics distributions.

III. MICROSCOPIC ENSEMBLES, ERGODICITY

As we discussed it is virtually impossible to describe large macroscopic systems deterministically. Due to many various factors we can use only probabilistic description, i.e. there is a function $f(\mathbf{x}, \mathbf{p}, t)$ of all coordinates and momenta of all particles, which describes the probability of a system to occupy a certain microscopic state. Thus the average of any observable is:

$$\langle \Omega(t) \rangle = \int d\mathbf{x} d\mathbf{p} \Omega(\mathbf{x}, \mathbf{p}, t) f(\mathbf{x}, \mathbf{p}, t) d\mathbf{x} d\mathbf{p}. \tag{43}$$

Note by $\mathbf{x}$ and $\mathbf{p}$ we understand all phase space coordinates, discrete (like e.g. magnetization) or continuous. In quantum statistical mechanics one have to distinguish between two averages (often a source of confusion!): quantum mechanical average and statistical (probabilistic) average. The first one is related to fundamental uncertainty of QM which states that even if we have a complete knowledge about the wave-function of the system - we have intrinsically probabilistic description about the system: e.g. we can not possibly measure coordinates and momenta of a particle. Statistical average is related to the fact that we do not know the wave function of the system (or this wave function might not exist at all due to mixing with the environment). Then we have statistical uncertainty about the wave-function itself. General description of mixed states (states not described by a single wave function) is given by a density matrix. There are many ways to introduce it. Let us use the one which emphasizes its statistical nature. Assume that the system is described by some wave function with statistically random coefficients

$$|\Psi\rangle = \sum_m a_m |\psi_m\rangle, \tag{44}$$

where $|m\rangle$ is some basis. Then expectation value of any observable is

$$\overline{\Omega} = \sum_{m,n} a_n^\star a_m \Omega_{n,m}. \tag{45}$$

Note that in order to measure the expectation value we have to perform many experiments. But for each experiment a_n and a_m are randomly chosen according to some statistical probability distribution. So in practice we have to do both quantum and statistical average and it is very hard to distinguish between them (though sometimes it is possible - HBT effect - examples later!). Then

$$\langle\overline{\Omega}\rangle \equiv \langle\Omega\rangle = \sum \rho_{m,n}\Omega_{n,m} = \mathrm{Tr}(\rho\Omega), \tag{46}$$

where $\rho_{nm} = \langle c_n^\star c_m\rangle$. Note that for a pure state we have $(\rho\rho)_{nm} = \sum_{n,p,m} c_n^\star c_p c_p^\star c_m = c_n^\star c_m = \rho_{nm}$. For non-pure states this does not hold.

Density matrix in quantum statistical physics plays the same role as the distribution function $f(\mathbf{x},\mathbf{p})$ in classical statistical physics. Instead of integral over phase space we are dealing with sum over Hilbert space. Diagonal elements of the density matrix play the role of probability of occupying certain microscopic state: $\rho_{nn} = \langle c_n^\star c_n\rangle$.

Homework. *Consider a system of two spins. Find its density matrix in two situations: (i) both spins are in the state* $(|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$ *and (ii) each spin is in either* $|\uparrow\rangle$ *or* $|\downarrow\rangle$ *state with equal probability. Find* ρ^2 *in both situations. What do you get for pure state (i), for mixed state (ii)?*

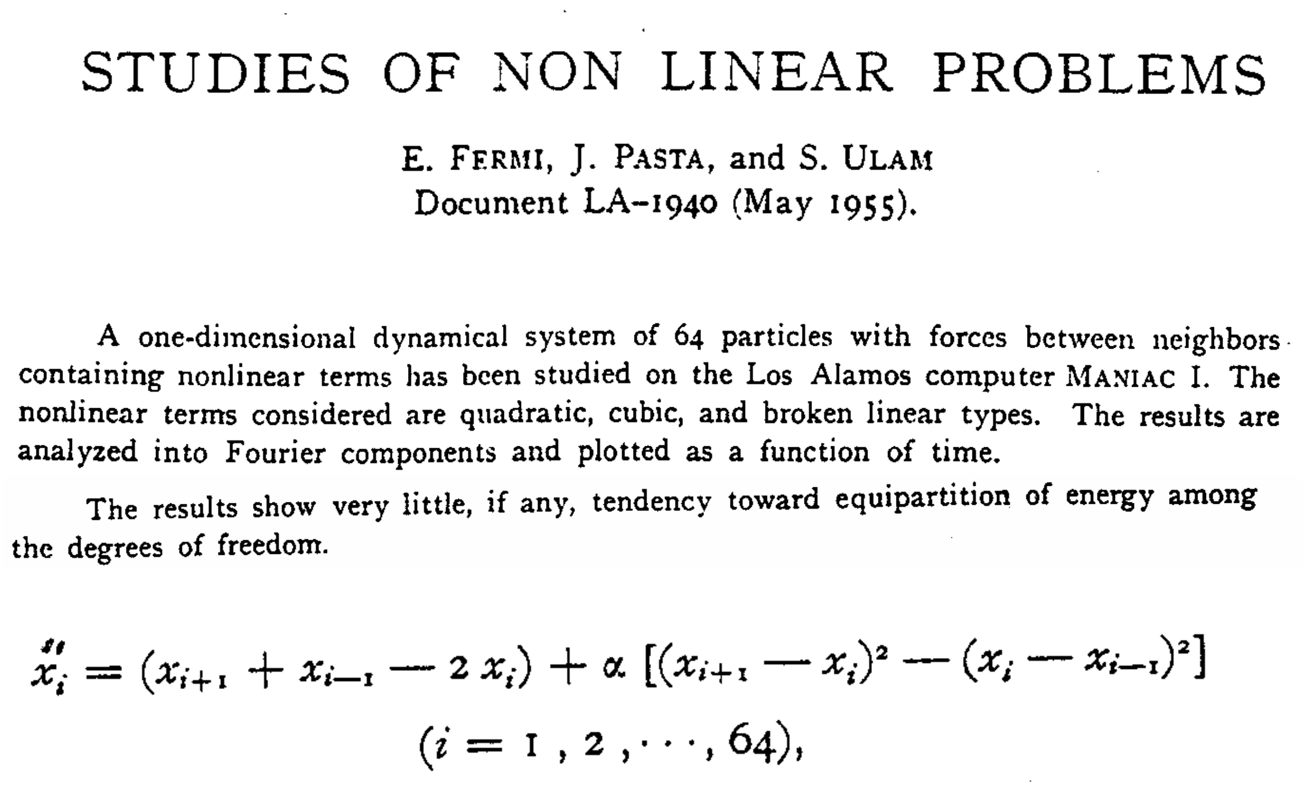
Unless I mention explicitly that we are dealing *only* with classical or quantum systems averaging over classical statistical distribution immediately translates to averaging over density matrix in quantum statistical physics and vice versa.

Ergodic hypothesis: in macroscopic systems average over equilibrium ensemble describing equilibrium is equivalent to time average, i.e.

$$\langle\Omega\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^t \Omega(t)dt \tag{47}$$

The RHS of this equation is statistical average over many realizations of the experiment. The LHS is the time average of a single realization. Ergodic hypothesis is the conjecture. There is no microscopic proof (as far as I know). There is a subtlety in quantum case, because measurement itself has a back-action: it introduces additional statistical uncertainty into the system by randomly projecting the system in the one of the eigenstates of the observable. Usually measurements destroy pure states and make them mixed states.

Fermi Pasta Ulam Problem. Ergodic hypothesis is a very natural concept. We are implicitly dealing with it all time. E.g., if we measure equilibrium fluctuations of voltage we do not care if we average over time or we start over voltmeter again and again. However, this hypothesis is very hard to prove. The first numerical attempt to do this was by Fermi Pasta and Ulam at Los Alamos National lab using one of the first computers. They considered a 1D system of coupled oscillators with a weak nonlinearity (see Fig. 1). They initially excited the lowest mode with $k = 2\pi/L$ and expected that after some time because of nonlinearity the energy will be equipartitioned between different modes (as we will learn later in the course). However, what they found was quite opposite: after a sufficiently long time the energy almost completely returned to the first mode. This is like in movies you destroy some object and then magically it reorganizes back to the same shape. Typically you need some



STUDIES OF NON LINEAR PROBLEMS

E. FERMI, J. PASTA, and S. ULAM
Document LA-1940 (May 1955).

A one-dimensional dynamical system of 64 particles with forces between neighbors containing nonlinear terms has been studied on the Los Alamos computer MANIAC I. The nonlinear terms considered are quadratic, cubic, and broken linear types. The results are analyzed into Fourier components and plotted as a function of time.

The results show very little, if any, tendency toward equipartition of energy among the degrees of freedom.

$$\ddot{x}_i = (x_{i+1} + x_{i-1} - 2x_i) + \alpha\,[(x_{i+1} - x_i)^2 - (x_i - x_{i-1})^2]$$
$$(i = 1, 2, \cdots, 64),$$

FIG. 1 FPU report: the abstract and the main conclusion

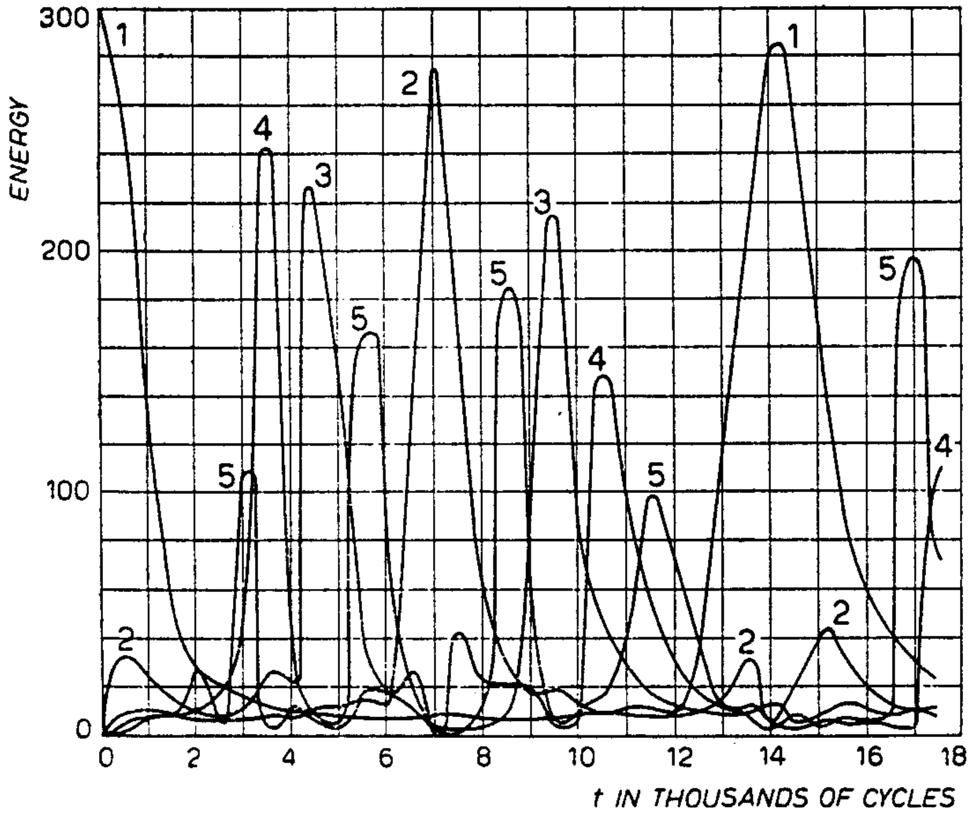


FIG. 2 FPU report: time dependence of the energy of first several modes.

divine form like Fairy to do this. However, here it happened without such intrusion. People later understood why this happened: 1D nonlinear systems can be almost exactly described by nonlinear excitations (solitons), which in many respects behave as noninteracting objects. Thus they do not cause thermalization. This example was understood. But even now we know very little about which systems are ergodic and which are not. This is a subject of active research.

Statistical independence. Imagine that we split a large system into two subsystems. Each subsystem is not independent. However, it knows about another subsystem only because of surface effects. Those effects should be small if the subsystems are small too. Therefore we expect that the two subsystems approximately independent from each other. This means that the distribution function should approximately factorize: $f(\mathbf{x},\mathbf{p}) \approx f_1(\mathbf{x}_1,\mathbf{p}_1) f_2(\mathbf{x}_2,\mathbf{p}_2)$

(the same should be true about density matrix - the latter should be roughly block diagonal). This in turn implies that logarithm of the distribution function is roughly additive: $\ln f(\mathbf{x},\mathbf{p}) \approx \ln f_1(\mathbf{x}_1,\mathbf{p}_1) + \ln f_2(\mathbf{x}_2,\mathbf{p}_2)$. So the log of the distribution function should know only about additive quantities. As we know from the previous chapter statistical independence means that fluctuations of all additive thermodynamic quantities (like energy) should become small in large systems.

Liouville's theorem. In a closed system the distribution function is conserved along the trajectories. Let us consider an element of the phase space $d\mathbf{x}d\mathbf{p}$. Let us now consider an infinitesimal time step dt. Then this volume becomes $d\mathbf{x}'d\mathbf{p}'$. Note that $x_a' = x_a + \dot{x}_a dt$ and similarly $p_a' = p_a + \dot{p}_a dt$. This means that $dx_a' = dx_a + \frac{\partial \dot{x}_a}{\partial x_a} dx_a dt$. From this we find that

$$dx_a' dp_a' = dx_a dp_a \left[1 + \left(\frac{\partial \dot{x}_a}{\partial x_a} + \frac{\partial \dot{p}_a}{\partial p_a}\right) dt\right] = dx_a dp_a \tag{48}$$

by Hamilton equations of motion: $\dot{x}_a = \partial H/\partial p_a$ and $\dot{p}_a = -\partial H/\partial x_a$. The Liouville's theorem implies that all the pure states are transformed from the point $(\mathbf{x},\mathbf{p})$ to the point $(\mathbf{x}',\mathbf{p}')$ and the phase volume does not change. In practice this phase space volume of course impossible to measure, but one can draw several formal consequences from this theorem. In particular, because $fd\Gamma = f'd\Gamma'$ (conservation of probability) we find that $df/dt = 0$. Note that this is a full derivative: $df/dt = \partial f/\partial t + \sum_a \frac{\partial f}{\partial x_a}\dot{x}_a + \frac{\partial f}{\partial p_a}\dot{p}_a$.

Consequences.

- Consider a statistical average of some observable Ω, which explicitly does not depend on time. Then

  $$\begin{aligned} \frac{d\langle\Omega}{dt} &= \int d\Gamma \frac{\partial f(\mathbf{x},\mathbf{p},t)}{\partial t}\Omega(\mathbf{x},\mathbf{p}) = -\sum_a \int d\Gamma\Omega(,\mathbf{p})\left(\frac{\partial f}{\partial x_a}\dot{x}_a + \frac{\partial f}{\partial p_a}\dot{p}_a\right) \\ &= \sum_a \int d\Gamma f\left(\frac{\partial\Omega}{\partial x_a}\frac{\partial H}{\partial p_a} - \frac{\partial\Omega}{\partial p_a}\frac{\partial\Omega}{\partial q_a}\right) = \langle\{\Omega, H\}\rangle. \end{aligned} \tag{49}$$

  If the Poisson brackets of the observable Ω with the Hamiltonian vanish (recall this is true for all conserved quantities) then the corresponding average is a constant of motion. This is a statement,, which is intuitively clear in any case.

- The equilibrium distribution function should satisfy $\{f_{\text{eq}}, H\} = 0$. Clearly any function which depends only on H satisfies this requirement: $\{f(H), H\} = 0$. If we use this choice then we imply that within the energy shall all states are equally probable. Note if there are additional conserved quantities than the stationary distribution function can depend on all these quantities and still be stationary. This implies equipartition between all possible phase space points satisfying the constraints.

Quantum Liouville's theorem. Recall that $\rho_{nm}(t) = \langle c_n^\star(t) c_m(t)\rangle$. Let us compute time derivative of ρ:

$$\dot\rho_{nm}(t) = \langle \dot c_n^\star c_m + c_n^\star \dot c_m\rangle = i(E_n - E_m)\rho_{nm}(t) = iH_{np}\rho_{pm} - i\rho_{np}H_{pm} = i[H,\rho]. \quad (50)$$

So

$$i\frac{d\rho}{dt} = [\rho, H]. \quad (51)$$

Let us check the observables corresponding to stationary operators commuting with the Hamiltonian are conserved in time.

$$d\langle\Omega\rangle/dt = \mathrm{Tr}\,[\dot\rho\Omega] = -i\mathrm{Tr}[\rho H\Omega - H\rho\Omega] = -i\langle[H,\Omega]\rangle = 0. \quad (52)$$

The consequences of the quantum Liouville's theorem are basically identical to those of the classical Liouville's theorem: (i) Stationary density matrix should commute with the Hamiltonian. In turn this implies that it should be diagonalizable simultaneously with the Hamiltonian and (ii) The density matrix, which is some functional of the Hamiltonian and other conserved quantities automatically satisfies (i) and thus is conserved in time.

IV. ENTROPY, TEMPERATURE, LAWS OF THERMODYNAMICS.

A. Statistical Entropy

One of the main postulates of the statistical physics is that the system tends to equally populate all available states within the constraints of total available energy and possibly other conserved quantities. In these situations one can introduce the entropy which is the measure of available phase space Γ. Formally $S = -\ln\Gamma$. (We will later give a more general definition in the case when the probabilities to occupy different states are not equal).

Example: consider a system of spin 1/2 systems with total magnetization $M_z = \gamma\sum_i S_i^z$. In isolated systems with no external magnetic field in the xy-plane then M_z is a conserved quantity.

Next let us assume that the system is in the external magnetic field so the energy of the system is

$$U = -HM_z = -H\gamma \sum_i \hat{S}_i^z. \tag{53}$$

Let us first find how the entropy is connected with magnetization (and thus) energy. We need to find the phase space available for a given magnetization. We know the answer: it comes from the binomial distribution:

$$\Gamma = \frac{N!}{N_\uparrow! N_\downarrow!} = \frac{N!}{(N/+L/2)!(N/2-L/2)!}, \tag{54}$$

where $L = N_\uparrow - N_\downarrow = M_z/\gamma$. This expression is quite cumbersome. Let us study its logarithm, i.e. entropy. For simplicity we assume that L is small

$$S(L) = \ln\Gamma(L) \approx N\ln N - \frac{N+L}{2}\ln\left(\frac{N+L}{2}\right) - \frac{N-L}{2}\ln\left(\frac{N-L}{2}\right) + \mathcal{O}(\ln N) \tag{55}$$

Expanding the expression into the Taylor series in L we find

$$S(L) \approx N\ln 2 - \frac{N+L}{2}\frac{L}{N} + \frac{N-L}{2}\frac{L}{N} = N\ln 2 - \frac{L^2}{2N} = N\ln 2 - \frac{1}{N}\left(\frac{U}{\gamma H}\right)^2. \tag{56}$$

This function is of course strongly peaked around $U = 0$.

Now let us imagine that we split the system into two subsystems and let us ask in how many ways we can distribute magnetization (energy) in a particular way $U = U_1 + U_2$ or $M^z = M_1^z + M_2^z$. Because systems are noninteracting clearly we have

$$\tilde{\Omega}(U_1, U_2) = \Omega_1(U_1)\Omega_2(U_2) \tag{57}$$

and

$$\Omega(U) = \sum_{U_1} \tilde{\Omega}(U_1, U - U_1). \tag{58}$$

Assume that the first block contains N_1 spins and the second block contains N_2 spins. Then

$$S(U_1, U - U_1) \approx N\ln 2 - \frac{1}{N_1}\left(\frac{U_1}{\gamma H}\right)^2 - \frac{1}{N_2}\left(\frac{U - U_1}{\gamma H}\right)^2. \tag{59}$$

Since S is extensive and total number of configurations (=probability) is exponential in S the distribution is highly peaked near the maximum of the entropy (maximum number of available configurations). To find the maximum we need to differentiate S with respect to U_1:

$$\frac{U_1^\star}{N_1} = \frac{U_2^\star}{N_2} \tag{60}$$

thus $U_1^\star = UN_1/N$ Thus we find that with the highest probability the magnetization is uniformly distributed in the system. This result is of course what we anticipate from the central limit theorem. But here it comes from the principle of maximum of entropy.

Note that the entropy is additive. Within the saddle point approximation

$$S(U) = N\ln 2 - \frac{1}{N}\left(\frac{U}{\gamma H}\right)^2 = \tilde{S}(U_1^\star, U - U_1^\star). \tag{61}$$

Lessons about entropy: entropy is additive. Within the postulate that all available states are equally probable maximum of the entropy corresponds to the maximally probable configuration.

B. Temperature

Now assume that we have two arbitrary systems in a contact with each other. Then quite generally we can write

$$\Omega(E) = \sum_{E_1} \Omega_1 E_1 \Omega_2(E - E_1) \tag{62}$$

In equilibrium we require that sum is dominated by the maximum so

$$\Omega(E) \approx \Omega_1(E_1)\Omega_2(E - E_1) \tag{63}$$

where

$$\frac{\partial \Omega_1}{\partial E_1}\Omega_2(E_2) = \Omega_1(E_1)\frac{\partial \Omega_2(E_2)}{\partial E_2} \quad \Leftrightarrow \quad \frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2}. \tag{64}$$

This derivative we call inverse temperature: $\partial S(E)/\partial E = 1/T$. Thus the equilibrium between the two subsystems immediately results in the requirement of constant temperature: $T_1 = T_2$.

Let us return to our example. From Eq. (55) we find

$$\frac{1}{T} = \frac{1}{2\gamma H}\ln\left[\frac{N\gamma H - U}{N\gamma H + U}\right]. \tag{65}$$

Note that the minimum of the energy occurs at $U = -N\gamma H$ when all the spins are polarized along the magnetic field. In this case we clearly have $T = 0$ (as well as $S = 0$) - there is only one configuration corresponding to the minimal energy (well known statement from quantum mechanics). Correspondingly this configuration corresponds to zero temperature. Now assume that $U = -N\gamma H + \delta U$. Then we have

$$T \approx 2\gamma H \frac{1}{\ln\left[\frac{2N\gamma H}{\delta U}\right]} \tag{66}$$

or in a more familiar form

$$\delta U \approx N 2\gamma H \exp\left[-\frac{2\gamma H}{T}\right]. \tag{67}$$

This is of course the familiar Boltzmann's distribution. We can interpret Eq. (67) in the following way. The energy required to excite a particular spin is $2\gamma H$. The probability of this excitation is $\exp[-2\gamma H/T]$ (later we will return to a more general justification of this postulate). The average energy (above the ground state) is then the probability of exciting a particular state times the energy of each excitation times the total number of magnets.

Homework, due 10/02. *Taking this interpretation of the probability compute the distribution of energy in the system $P(E)$. What is the name of this distribution function. What are the fluctuation of the energy and magnetization? Discuss what happens as N increases.*

One can invert Eq. (55) without assumption about small temperature. Then

$$\delta U \equiv N\gamma H + U = 2N\gamma H \frac{1}{\exp\left[\frac{2\gamma H}{T}\right] + 1}. \tag{68}$$

As the energy further increases both entropy and temperature increase until we reach maximally probable unconstrained configuration where $S = N \ln 2$ and $T = \infty$. After that the entropy starts to decrease and the temperature becomes negative. This is actually the artifact of our model with bounded spectrum. Usually the energy spectrum is unbounded from above and thus infinite temperature state corresponds to the infinite energy, so it can not be reached. However, this example shows an important distinction between temperature showing how entropy changes with energy and the energy itself.

Homework, due 10/02. *Show that one can formally derive Eq. (68) by assuming that the probabilities of spin pointing down and up satisfy Boltzmann's distribution: $p_\downarrow/p_\uparrow = \exp[-2\gamma H/T]$ and of course $p_\downarrow + p_\uparrow = 1$. Repeat calculations for the previous problem with this distribution. Does it become narrower or wider if at given N we increase temperature? Explain.*

C. First law of thermodynamics.

Let us assume that the state of the system depends on some external parameter x, which can be volume, magnetic field etc. If we do reversible (adiabatic) transformation on the system the energy change is characterized by the work $\delta W = J\delta x$. For the reversible process the entropy can not change. One has to be a little careful that after such transformation the microcanonical ensemble might not be any longer microcanonical, i.e. the idea that all states at a fixed energy are equally probable does not quite work. However, for large systems the fluctuations of all quantities including energy should remain small and thus we should deal with the narrow energy distribution even after the transformation. Then we can write

$$\Delta S = 0 = S(E + J\delta x, x + \delta x) \approx \left.\frac{\partial S}{\partial E}\right|_x J\delta x + \left.\frac{\partial S}{\partial x}\right|_E \delta x. \tag{69}$$

Note that it is crucial that we have reversible process because otherwise we will heat the system and the energy change will depend on the details of the process. Thus

$$\left.\frac{\partial S}{\partial x}\right|_E = -\frac{J}{T}. \tag{70}$$

Now for arbitrary process, where entropy can increase we have

$$dS = \left.\frac{\partial S}{\partial E}\right|_x dE + \left.\frac{\partial S}{\partial x}\right|_E dx = \frac{dE}{T} - \frac{Jdx}{T}. \tag{71}$$

Therefore

$$dE = TdS + Jdx. \tag{72}$$

We thus can identify TdS with the heat. Note that $-J$ is the generalized force.

Some justification of $\Delta S = 0$ for a reversible process: In quantum mechanics if we change external adiabatically (slowly) we do not cause transitions between energy levels. The energy levels also generally stay adjacent to each other. So if we occupy them maximally within some energy window then we will continue to do so thus entropy does not change. In classical mechanics there are no transitions between orbits so the phase space also remains contingent and the entropy does not change. Other justification is that if we have a macroscopic degree of freedom and can change coupling between this degree of freedom and the system then we should have some surface of constant entropy characterized by this coupling. The entropy can not spontaneously change otherwise reversible motion is not possible. Note that large degree of freedom bears very little entropy.

D. Second law of thermodynamics

Now assume that we start from two systems. Each one is in equilibrium, but they are not necessarily in equilibrium with each other. Then after any process the entropy can either increase or stay the same.

The proof of this statement follows from the fact that at equilibrium (but the postulate of the statistical physics) the system occupies the largest phase space volume, i.e.

$$\Omega_1(E_1^\star)\Omega_2(E_2^\star) \geq \Omega_1(E_1)\Omega_2(E_2), \tag{73}$$

where E_1 and E_2 are the energies of the systems 1 and 2 before they are brought to the contact, while $E_1^\star$ and $E_2^\star$ are the equilibrium values of energy. Note that if no external work is done then $E_1^\star + E_2^\star = E_1 + E_2$.

Consequence: Second law in the Clausius form. Assume that we put the systems in contact for a short time so that the entropy change is small then

$$\delta S = \left(\frac{\partial S_1}{\partial E_1}\bigg|_{x_1}\right)\delta E_1 - \frac{\partial S_2}{\partial E_2}\bigg|_{x_2}\delta E_1 = \delta E_1\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \geq 0. \tag{74}$$

So we see that the heat flows from hotter to colder body. Similarly the second law in the Kelvin's form immediately follows from the fact that the entropy does not change in the adiabatic (reversible) cyclic process thus it can only increase if the process is irreversible.

Stability condition:

$$\frac{\partial^2 S_1}{\partial E_1^2} + \frac{\partial^2 S_2}{\partial E_2^2} \leq 0 \tag{75}$$

which implies that $\partial T/\partial E_x \geq 0$. This means that the heat capacity must be positive.

E. Third law of thermodynamics

Entropy is zero at zero temperature. We saw how the third law works at a particular example. The third law of thermodynamics is purely quantum mechanical. It is related to the fact that the lowest energy state is always non-degenerate.

V. PARTITION FUNCTION, CANONICAL ENSEMBLE, VON-NEUMANN ENTROPY

In the previous lecture we derived the properties of a spin system assuming it is **isolated**, and therefore **conserved energy**. This means that we worked in the **micro-canonical ensemble**.

Our approach was to start from the energy and derive entropy and temperature as its derivatives. However, this approach is not convenient. The total energy is something which is usually hard to measure because we are usually dealing with open systems with unclear boundaries. On the other hand temperature is something easy to measure. We know that if we put a spin system (thermometer) to a contact with the reservoir then the temperatures of both thermometer and the system will be the same. If the reservoir is big then presumably the state of the reservoir will not be affected by the thermometer and the temperature will be that of the reservoir. Once we start from temperature we can compute the magnetization and compare it to what we measure. This way is actually also much easier for doing calculations.

Like dealing with energy usually assumes isolated system (where energy is conserved) and thus with canonical ensembles, dealing with temperature assumes that there is an energy exchange with environment. I.e. our (sub)system is a part of a bigger system. Then the energy of the subsystem is not conserved and can fluctuate. Thus we have to deal with a different ensemble, which is called **canonical**.

Return to our spin system. Consider a subsystem of the system with N_1 spins in it (then $N - N_2$ systems) constitute environment. Now we assume that $N_1 \ll N$. We start from the extreme situation where $N_1 = 1$. Let us compute the phase space volumes corresponding to each of these situations:

$$\Omega_\downarrow = \Omega_1(E_1 = \gamma H)\Omega_2(E - E_1) \approx \exp\left[S_2 + \frac{\partial S_2}{\partial E_2}(-\gamma H)\right] = C\exp[-\gamma H/T] \qquad (76)$$

similarly

$$\Omega_\uparrow = \Omega_1(E_1 = -\gamma H) = C\exp[\gamma H/T] \qquad (77)$$

If we are interested only in the state of the particular spin we worry only about this numbers, which give the probabilities to occupy "down" or "up" states. So we can write

$$p_\uparrow = \frac{1}{Z}\exp[\gamma H/T],\ p_\downarrow = \frac{1}{Z}\exp[-\gamma H/T], \qquad (78)$$

where

$$Z = \exp[\gamma H/T] + \exp[-\gamma H/T] = \sum_n \exp[-\mathcal{E}_n/T] \qquad (79)$$

is called the partition function. It is customary to introduce the notation $\beta = 1/T$.

With this partition function we can calculate various quantities. For example average energy will be

$$\langle\mathcal{E}\rangle = \sum_n p_n\mathcal{E}_n = \frac{1}{Z}\sum_n \mathcal{E}_n\exp[-\beta\mathcal{E}_n] = -\frac{\partial \ln Z}{\partial \beta}. \qquad (80)$$

Let us play with our specific example a bit more

$$Z = 2\cosh[\beta\gamma H]. \tag{81}$$

Then the mean energy of a single spin is

$$\langle \mathcal{E} \rangle = -\gamma H \frac{\sinh(\beta\gamma H)}{\cosh(\beta\gamma H)} = -\gamma H + \frac{2\gamma H}{\exp[2\beta\gamma H] + 1}. \tag{82}$$

Exactly what we had before. At small magnetic fields (high temperatures) we get $\langle \mathcal{E} \rangle \approx -\frac{\gamma^2 H^2}{T}$. The average magnetization is

$$\langle S_z \rangle = -\frac{\langle \mathcal{E} \rangle}{\gamma H} = \tanh \beta\gamma H \tag{83}$$

The other important quantity is the magnetic susceptibility

$$\chi = \frac{\partial M}{\partial H} = \beta\gamma^2 \frac{1}{\cosh^2(\beta\gamma\hat{\mathcal{H}})}. \tag{84}$$

At large temperatures (small magnetic fields) we have $\chi \approx \gamma^2/T$ famous Curie law.

Specific heat:

$$C = \frac{\partial E}{\partial T} = \frac{(\beta\gamma H)^2}{\cosh^2(\beta\gamma H)} \tag{85}$$

Note that at small temperatures $\beta \gg 1$ the heat capacity is exponentially small: this is a general property of gapped systems. In the opposite limit $\beta \ll 1$ ($T \gg 1$) the heat capacity is small again. This is related to the maximum in the density of states: we can not add more energy to the system by increasing the temperature.

Homework, due 10/02. *Consider a system of noninteracting spins with arbitrary S:* $S_z = S, S-1, ..., -S$ *in a magnetic field with the Hamiltonian* $\mathcal{H} = -hS_z$. *Calculate the partition function for this system at finite temperature. Calculate average energy, magnetization, susceptibility and heat capacity. Discuss its behavior at large S. Now consider a classical spin in the magnetic field with the energy* $\mathcal{E} = -hS_z = -hS\cos\theta$. *Define the classical partition function* $Z_{cl} = \int \sin\theta d\theta d\phi \exp[-\beta\mathcal{E}(\theta,\phi)]$. *Calculate the same quantities as for the quantum problem. Compare your results with* $S = 1/2$ *quantum case and large S quantum case. Discuss which of the computed quantities are similar for* $S = 1/2$ *and the classical case and which quantities are qualitatively different in these two situations.*

This single spin partition function is good for describing properties of a single spin. But what if we are interested in two-spin observables or even observables which include many spins. Then we have to consider larger subsystems. But the reasoning stays the same. The probability of each microscopic configuration is given by the Boltzmann factor $C\exp[-\beta\mathcal{E}]$. Indeed the corresponding contribution to the partition function is $\Omega_1(\mathcal{E})\exp[-\beta\mathcal{E}]$. But $\Omega_1(\mathcal{E})$ is precisely the number of microscopic configurations of the subsystem 1 having the energy $\mathcal{E}$. So quite generally we can write

$$p(\mathcal{E}) = \frac{1}{Z}\exp[-\beta\mathcal{E}] \tag{86}$$

Note that there is nothing special about energy except this is a conserved quantity (for the total system). If there are other additive conserved quantities we should have similar "Boltzmann's" weights to them (everything follows from the expansion of entropy around the maximum). For example if the total number of particles is allowed to fluctuate between the subsystem and environment then

$$p(\mathcal{E},\mathcal{N}) = \frac{1}{Z}\exp[-\beta(\mathcal{E}-\mu\mathcal{N})], \tag{87}$$

where

$$Z = \sum_{\mathcal{N}}\sum_{\mathcal{E}(\mathcal{N})}\exp[-\beta(\mathcal{E}-\mu\mathcal{N})], \tag{88}$$

$\mu\beta = \partial S/\partial\mathcal{N}|_{\mathcal{E}}$ or $\mu = T\partial S/\partial\mathcal{N}|_{\mathcal{E}}$. So one can write

$$d\mathcal{E} = TdS - \mu d\mathcal{N}. \tag{89}$$

One can write similar expressions if there are other additive conserved quantities (integrals of motion). The corresponding distributions are caller generalized Gibbs ensembles. The ensemble where only energy and number of particles are allowed to fluctuate (Eq. (87)) is called grand canonical.

A. Entropy

Starting from temperature we managed to derive energy and reproduce our earlier results. Next we need to find entropy. First let us make a formal derivation. We know that

$$TdS = dE \;\Leftrightarrow\; dS = \beta dE = d(\beta E) - Ed\beta. \tag{90}$$

Using that $E = -d\ln Z/d\beta$ we find that $S = \beta E + \ln Z$. The arbitrary integration constant is fixed by the third law of thermodynamics $S = 0$ at $T = 0$. Let us introduce the notation $F = -T\ln Z$, free energy. Then we find that

$$S = \frac{E - F}{T} \;\Leftrightarrow\; F = E - TS. \tag{91}$$

From thermodynamics recall that F is defined as the free energy. Now we have its expression through the partition function (note that both F and E are expressed through Z, which in turn is the function of temperature. If we change temperature (keeping other external parameters fixed) then

$$dF = dE - TdS - SdT = -SdT \tag{92}$$

so we also find that

$$S = -\frac{\partial F}{\partial T} = \frac{\partial (T\ln Z)}{\partial T}, \tag{93}$$

which of course agrees with our previous formula.

How about the probabilistic interpretation of the entropy? Let us again return to our example and then generalize. Recall that for noninteracting spins

$$Z = \prod_j \sum_{n_j} \exp[-\beta E_j^{(n)}] = \left(\sum_n \exp[-\beta \mathcal{E}_n]\right)^N \tag{94}$$

So the free energy is

$$F = -TN\ln\left[\sum_n \exp[-\beta\mathcal{E}_n]\right] \tag{95}$$

and thus the entropy

$$S = -dF/dT = N\ln(z) + \frac{N}{zT}\sum_n \mathcal{E}_n \exp[-\beta\mathcal{E}_n] = -\frac{N}{z}\sum_n \ln\left[\frac{\exp[-\beta\mathcal{E}_n]}{z}\right]\exp[-\beta\mathcal{E}_n] = -N\sum_n p_n \ln p_n, \tag{96}$$

where z is the partition function corresponding to the single spin and $= p_n = \exp[-\beta\mathcal{E}_n]/z$ is the probability to occupy this state. Note that we actually did not have to assume that our subsystem contains a single spin. Generally we would have

$$S = -\sum_n p_n \ln p_n, \tag{97}$$

where $p_n = \exp[-\beta\mathcal{E}_n]/Z$. This is the von Neumann entropy for the system in thermal equilibrium.

Let us try to get this answer purely probabilistically. Suppose we split the system into many subsystems and will start filling it with particles. The probability of the particle to go to the subsystem 1 is p_1 and so on. Of course the most probable configuration is that we have Np_1 particles in the first subsystem, Np_2 particles in the second subsystem etc. Let us find the number of ways Ω we can achieve the most probable configuration:

$$\Omega = \frac{N!}{(Np_1)!(Np_2)!\dots(Np_m)!} \tag{98}$$

This is pure combinatorics: first we choose the first particle for the first subsystem: N-ways then the second particle for the first subsystem: $N-1$ ways and so on. Then we will need to divide by number of equivalent permutations within each subsystem because this permutations lead to equivalent configurations. Now let us use Stirling's formula and take the logarithm of the expression above:

$$S = \ln\Omega \approx N\ln N - N - \sum_n (p_n N \ln(p_n N) - p_n N) = N(\ln N - \sum_n p_n \ln(Np_n)) = -\sum_n p_n \ln p_n. \tag{99}$$

So indeed the von Neumann entropy has a purely probabilistic nature.

Let us recall the whole story again. We started from the microcanonical ensemble and the basic assumption that all states with the same energy are equally probable. We termed entropy the logarithm of number of available microstates. Then by splitting the system into a subsystem and environment we showed that microcanonical distribution naturally leads to the canonical distribution for the subsystem where the probability to occupy a certain microstate with energy $\mathcal{E}$ is proportional to $\exp[-\beta\mathcal{E}]$. Then we computed the number of ways one can distribute particles over these microstates took the logarithm and derived von Neumann entropy.

Homework. Due 10/06. *Derive the expression for the von Neumann entropy for the spin one half system in the magnetic field and compare it with Eq. (55. Note that you will have to express either von Neumann entropy through energy or the entropy given by Eq. (55) through temperature.*

B. Information theory and entropy.

Information theory was invented by Claude Shannon at AT&T Bell Labs beginning in the 1940s. Information theory attempts to quantify the information content of messages and studies the reliability of transmission of that information over noisy channels of finite frequency bandwidth.

What is information? Suppose you have a deck of cards. I randomly take on card the question is how much information I can transmit by revealing what the card is. Think the deck of cards being the next hot thing in Wall street. For how much you can sell this information? If there is only one card in the deck then there is no information content - by saying you took ace you do not say much if there is only one ace in the deck (by saying it is oil you do not say much if you trade only with oil). Clearly the more cards in the deck (the more trading areas you see) the more information you reveal by opening your card (the better price you can get for revealing this information). So information should be monotonic function of N - the number of cards in a deck.

$$I(N+1) \geq I(N) \tag{100}$$

If you have another deck of M cards then if you reveal two cards one from first deck and one from the second deck your information is additive. Your earning will add up. So we should have

$$I(NM) = I(N) + I(M) \tag{101}$$

The sensitive definition of I satisfying this is

$$I(N) = \log_2(N) \tag{102}$$

The base 2 is just the tradition. So far we assumed that all outcomes are equally probable. But suppose that one outcome has very large probability. Then by picking this outcome you do not reveal much information. This outcome is almost expected. If the outcome is not very probable you are revealing much more information but the probability of this outcome is small. What is the sensible definition of the information in this case?

Suppose now that we are using a transmission line sending a signal encoded in zeros and ones. Each zero occurs with the probability p and one with probability q. What is the amount of information we transmit in N bits. Assume N is large. Obviously if $p = 1$ or $p = 0$ our signal would consist of only zeros and one and we transmit no information. In general we will have total number of zeros pN and total number of ones qN. The total number of way we can arrange for having this many zeros and this many ones is

$$\Gamma = \frac{N!}{(Np)!(Nq)!} \tag{103}$$

The information encoded in a particular sequence is defined as $I = \log_2 \Gamma$ (note that Γ plays the role of number of cards in the deck) so

$$I = -N(p \log_2 p + q \log_2 q) \tag{104}$$

So the information content per bit is $I_1 = -p \log_2 p - q \log_2 q$. This obviously generalizes to more outcomes:

$$I_1 = -\sum_i p_i \log_2 p_i \tag{105}$$

So the Shannon information is directly related to the von Neumann entropy.

C. Fluctuations of energy

Let us return to Boltzmann distribution and compute the energy fluctuations.

$$\delta E^2 = \langle E^2 \rangle - \langle E \rangle^2 = \frac{1}{Z} \sum_n E_n^2 \exp[-\beta E_n] - \left(\frac{\partial \ln Z}{\partial \beta} \right)^2 = \frac{\partial^2 \ln Z}{\partial^2 \beta} = -\frac{\partial E}{\partial \beta} = T^2 \frac{\partial E}{\partial T} = T^2 C_V. \tag{106}$$

Note that C_V is the heat capacity. We put index V here to emphasize that this quantity is computed at fixed external parameters like volume. We already saw that C_V must be positive in order to have stable equilibrium: maximum entropy versus minimum or saddle point. Here we see that the same quantity C_V determines the energy fluctuations in the system. This is one of the examples of the fluctuation-dissipation theorem: fluctuations of some quantity is related to a certain susceptibility of this quantity. We will return to this issue in detail later.

Similarly in the grand canonical ensemble one can study fluctuations of the number of particles. We need to change β to $-\beta\mu$ and repeat the analysis:

$$\langle N \rangle = T \frac{\partial \ln Z}{\partial \mu} \tag{107}$$

and

$$\delta N^2 = T^2 \frac{\partial^2 \ln Z}{\partial \mu^2} = T \frac{\partial N}{\partial \mu} = TV \frac{\partial n}{\partial \mu} \tag{108}$$

The latter quantity is called compressibility and we see that it must be positive in order to ensure the equilibrium in the system. Recall the chemical potential is a measure of how entropy of the system changes with the number of particles. We will see later that in ideal gases chemical potential is related to the pressure. High pressure means a high desire of the system to expand, i.e. the system can gain a lot of entropy expanding. So positive compressibility basically tells us that as we decrease volume of the system (increase n) we necessarily increase pressure. We will derive this explicitly when we talk about ideal gases.

VI. THE IDEAL GAS. THERMODYNAMIC POTENTIALS.

So far we dealt with the situations where the Hilbert space was finite and the partition function

$$Z=\sum_n \exp[-\beta\mathcal{E}_n] \tag{109}$$

was well defined. What shall we do if the spectrum is continuous. Let us consider the ideal one dimensional gas in the box of size L. Then from quantum mechanics we know that

$$\mathcal{E}_n=\frac{\hbar^2 k_n^2}{2m}, \tag{110}$$

where $k_n=2\pi n/L$. If L is large then for all practical purposes we can change summation over n to integration over k. Note that

$$1=\delta n=\frac{L}{2\pi}\delta k=\frac{L}{2\pi\hbar}\delta p \tag{111}$$

Therefore in this situations

$$Z=\sum_n \exp[-\beta\mathcal{E}_n]=\int\frac{dxdp}{2\pi\hbar}\exp[-\beta p^2/2m]. \tag{112}$$

We could have guessed this result from the beginning except for the factor $2\pi\hbar$. In classical physics it is arbitrary. But it is very convenient because it makes the partition function dimensionless and connects quantum and classical partition functions. This expression obviously generalizes to N-particles in d-dimensions

$$Z=\prod_i\int\frac{d\mathbf{x}_i d\mathbf{p}_i}{(2\pi\hbar)^d}\exp[-\beta\mathcal{E}(\mathbf{x}_i,\mathbf{p}_i)]. \tag{113}$$

Let us now compute the partition function for the ideal gas

$$Z=\left(\int\frac{d\mathbf{x}d\mathbf{p}}{(2\pi\hbar)^d}\exp[-\beta\mathbf{p}^2/2m]\right)^N=\left(\frac{V}{(2\pi\hbar)^d}\left[\int_{-\infty}^{\infty}\exp[-\beta p^2/2m]dp\right]^d\right)^N=\left(V\left[\frac{mT}{2\pi\hbar}\right]^{d/2}\right)^N. \tag{114}$$

From this we can define free energy

$$F=-T\ln Z=-TN\ln\left(V\left[\frac{mT}{2\pi\hbar}\right]^{d/2}\right), \tag{115}$$

similarly

$$E = -\frac{\partial \ln Z}{\partial \beta} = NT\frac{d}{2}. \tag{116}$$

So each particle has an average energy $Td/2$ ($3/2T$ in three dimensions). This result we could have obtained just using single particle Boltzmann's probability distribution.

From usual thermodynamics we should remember that the pressure is defined as

$$P = -\frac{\partial E}{\partial V}|_S = -\frac{\partial F}{\partial V}|_T. \tag{117}$$

So for the ideal gas we find

$$P = NT/V. \tag{118}$$

Should sound familiar. Interestingly this is a very robust law: it does not depend on the dimensionality, mass of the particles their structure etc. It relies only on the fact that the particles are not interacting and the partition function is proportional to V^N.

We can actually also define the pressure in a usual way from the principle of maximum of the entropy. In the same way we defined temperature and chemical potential. Imagine that we have two gases separated by a membrane which can move. We will allow for the energy exchange to fix the temperature but will not allow for particle exchange. Then the total $V = V_1 + V_2$ is fixed so we need to maximize the entropy with respect to V_1. Thus

$$\frac{\partial S_1}{\partial V_1}|_{N_1,T} = \frac{\partial S_2}{\partial V_2}|_{N_2,T} \tag{119}$$

Let us make the formal definition

$$P = T\frac{\partial S}{\partial V}|_{T,N} \tag{120}$$

Now we can generally write (for the subsystem 1) that

$$dS_1 = \frac{\partial S_1}{\partial E_1}|_{V_1} dE_1 + \frac{\partial S_1}{\partial V_1}|_{E_1} dV_1 = \frac{1}{T_1} dE_1 + \frac{P_1}{T_1} dV_1 \tag{121}$$

From this we indeed recover that

$$dE = TdS - PdV, \quad dF = -SdT - PdV \tag{122}$$

So the two definitions of the pressure coincide.

Homework, due 10/06. *Consider a system of spins in a magnetic field. Define the magnetization as $m = -1/V(\partial F/\partial H)$. Illustrate explicitly that this magnetization is indeed proportional to average S_z. Now define the magnetization in a different way using entropy.*

Assume that the magnet is split into two subsystems and the total magnetic flux coming through this subsystems is fixed, i.e. $H_1A_1 + H_2A_2 = \text{const}$. *Here* A_1 *and* A_2 *are the surface areas of the subsystems. Now demand that the entropy is maximum of possible partitionings of the flux between the subsystem. Show this should leave to a quantity which is identical between the two subsystems. Show that this quantity (up to a prefactor) is the magnetization defined in the beginning of the problem.*

Homework from Kardar, due 10/06. *4.1, 4.2, 4.4,4.7, 4.11*

A. Gibbs paradox.

The expression for the free energy we obtained has a problem: it is not extensive. Suppose we increase the volume and the number of particles by a factor of two. Then $N \to 2N$ and $V \to 2V$ so that $F \to 2F - TN\ln 2$. Note that the energy itself is extensive: $E \to 2E$. So does the heat capacity. Pressure which is $\partial F/\partial V$ is also not affected by this extra term. Yet it is annoying since we rather deal with extensive quantities. Note that

$$F = E - TS \tag{123}$$

since E is not affected by this extra term then the fact that F is not extensive means that entropy is not extensive. This is quite contrary to our previous discussion and the spin chain example.

To make the point even more transparent consider two subsystems containing identical ideal gases at the same temperature and pressure. Suppose one gas contains N_1 particles and the other N_2 particles. Then the initial entropy (while the gases are separated) is

$$S_1 + S_2 = -\frac{\partial F_1}{\partial T} - \frac{\partial F_2}{\partial T} = N_1 \ln\left(V_1 \left[\frac{mT}{2\pi\hbar}\right]^{d/2}\right) + N_2 \ln\left(V_2 \left[\frac{mT}{2\pi\hbar}\right]^{d/2}\right) + \frac{d}{2}(N_1 + N_2) \tag{124}$$

Suppose that $N_1 = N_2 = N/2$ and $V_1 = V_2 = V/2$. Then

$$S_1 + S_2 = N \ln\left(\frac{V}{2}\sigma\right) + \frac{Nd}{2}, \tag{125}$$

$\sigma = \sqrt{(mT/2\pi\hbar)^d}$. Let us pretend for a second and assume that gases in two parts are different. Then when we mix them (remove the barrier) the entropy will definitely increase. Clearly the final entropy is

$$S = N \ln(V\sigma) + \frac{Nd}{2} \tag{126}$$

so the difference between the two is

$$\Delta S_{mix} = S - S_1 - S_2 = N \ln 2 \tag{127}$$

This difference is called the mixing entropy. The fact that it is positive illustrates a well known point that once we remove the barrier the particles will mix with each other and never return to the original part of the container.

But what if the particles are identical? By removing the barrier we really did not do anything so the entropy increase is not real. The problem with our definition of Z is that we are over-counting states. We count the state where the particle 1 was in the phase space point $\mathbf{x}_1, \mathbf{p}_1$ and the particle 2 was in the phase space point $\mathbf{x}_2, \mathbf{p}_2$ and vice verso as two different states. But if the particles are indistinguishable this is one state. In quantum mechanics this issue is naturally resolved because the wave function of identical particles is always the superposition of say $|1\langle|2\rangle$ and $|2\rangle|1\rangle$. So this state in quantum mechanics would be counted once. So better definition of Z for identical particles is

$$Z = \frac{1}{N!} \prod_i \int \frac{d\mathbf{x}_i d\mathbf{p}_i}{(2\pi\hbar)^d} \exp[-\beta \mathcal{E}(\mathbf{x}_i, \mathbf{p}_i)]. \tag{128}$$

This gives additional contribution to the free energy which makes it extensive (without affecting energy, heat capacity, pressure, etc.) After all this prefactor is just a constant. So we get

$$F = -T \ln Z = -TN \ln \left(\frac{V}{N} \left[\frac{mT}{2\pi\hbar} \right]^{d/2} \right) - TN, \tag{129}$$

Now the free energy is explicitly extensive. So is the entropy

$$S = N \ln \left(\frac{V}{N} \sigma \right) + \frac{N(d+2)}{2}. \tag{130}$$

Note that whether we treat particles as identical or distinguishable is a matter of convenience as long as our measurements (or interactions) are insensitive to differences between them. Taking different isotopes should not result in any physical difference as long as there is nothing which tells us that the isotopes are different. And indeed we saw that measurable quantities like energy, heat capacity, pressure do not care about this extra term coming from $N!$. We will return to this issue again when we will study quantum gases.

Let us compute heat capacity. From $C_V = dE/dT|_V$ we find that

$$C_V = N \frac{d}{2}. \tag{131}$$

Sometimes it can be of interest to know heat capacity when we fix pressure rather then volume fixed. Note that the heat is defined as TdS (it is equal to dE if we fix the volume) then

$$C_P = T\frac{dS}{dT}|_P = \left(\frac{\partial E}{\partial T}|_P + P\frac{\partial V}{\partial T}|_P\right) = \frac{\partial(E+PV)}{\partial T}|_P = N\frac{d+2}{2}. \tag{132}$$

Note the expression $E+PV$, which is called enthalpy is convenient thermodynamic potential if we work at constant pressure.

B. Thermodynamic potentials. Relations between derivatives.

We already encountered energy and free energy $F = E - TS$. We know that they satisfy

$$dE = TdS - PdV, \quad dF = -SdT - PdV \tag{133}$$

From these relations we see that

$$P = -\frac{\partial E}{\partial V}\Big|_S = - - \frac{\partial F}{\partial V}\Big|_T \tag{134}$$

Usually the relations (133) are interpreted as that energy is the function of S and V, while free energy is the function of T and V. Of course this is not literary true since S, T and V are not independent variables but rather they are connected by the equation of state (see Eq. (130)). It is just the matter if convenience: if temperature is fixed (and volume changes) i.e. if we perform an isothermal process, then the free energy change is equal to $-PdV$. On the other hand if the entropy is fixed (adiabatic process) than the change of energy is equal to $-PdV$. On the other hand change of the energy in the first case and change of the free energy in the second are more complicated. For example

$$E = F + TS = F - T\frac{\partial F}{\partial T} \tag{135}$$

is explicitly defined as a function of temperature and volume.

We can assume that there is some other external parameter λ. Then obviously if we change it infinitesimally we have

$$dE = TdS - PdV + \Lambda d\lambda, \tag{136}$$

where $\Lambda = (\partial E/\partial\lambda)|_{S,V}$. Similarly

$$dF = -SdT - PdV + \Lambda d\lambda. \tag{137}$$

Thus we see that $\Lambda = (\partial F/\partial\lambda)_{T,V}$. We can interpret these equalities also in the following way

$$\delta E|_{S,V} = \delta F|_{T,V} \tag{138}$$

I.e. if we infinitesimally change some external parameter in the adiabatic process then the change of the energy would be identical to the change of the free energy in the corresponding isothermal process.

One can also introduce other potentials like enthalpy $H = E + PV$ and Gibbs free energy $G = H - TS = E - TS + PV$. We will obviously get $dH = TdS + VdP$ and $dG = -SdT + VdP$. These two potentials are convenient if we use S, P or T, P as independent variables respectively.

Using general thermodynamic relations we can obtain identities between various quantities. For example

$$\left.\frac{\partial C_V}{\partial V}\right|_T = T\frac{\partial^2 S}{\partial V\partial T} = -T\frac{\partial^3 F}{\partial T^2\partial V} = T\left.\frac{\partial^2 P}{\partial T^2}\right|_V \tag{139}$$

Similarly we can find

$$\left.\frac{\partial C_P}{\partial P}\right|_T = -T\left.\frac{\partial^2 V}{\partial T^2}\right|_P. \tag{140}$$

One can derive many similar relations like this.

Very often if we analyze various derivatives it is convenient to go from one set of variables to another e.g. from S and V to T and P. This can be done using Jacobians:

$$\frac{\partial(x,y)}{\partial(u,v)} \equiv \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v. \end{vmatrix} \tag{141}$$

Jacobians have a number of properties:

$$\frac{\partial(x,y)}{\partial(x,v)} = \left.\frac{\partial y}{\partial v}\right|_x \tag{142}$$

also

$$\frac{\partial(x,y)}{\partial(f,g)} = \frac{\partial(x,y)}{\partial(u,v)}\frac{\partial(u,v)}{\partial(f,g)}. \tag{143}$$

In particular

$$1 = \frac{\partial(x,y)}{\partial(x,y)} = \frac{\partial(x,y)}{\partial(u,v)}\frac{\partial(u,v)}{\partial(x,y)}. \tag{144}$$

Using these properties we can easily derive various thermodynamic relations. E.g.

$$C_V = T\frac{\partial S}{\partial T}\Big| V = T\frac{\partial(S,V)}{\partial(T,V)} = T\frac{\partial(S,V)}{\partial(T,P)}/\partial(T,V)/\partial(T,P) = T\left(\frac{\partial S}{\partial T}|_P\frac{\partial V}{\partial P}|_T - \frac{\partial S}{\partial P}|_T\frac{\partial V}{\partial T}|_P\right)/(\partial V/\partial P)|_T \tag{145}$$

Next note that

$$\left.\frac{\partial S}{\partial P}\right|_T = \frac{\partial^2 G}{\partial T \partial P} = -\left.\frac{\partial V}{\partial T}\right|_T. \tag{146}$$

Thus we find

$$C_P - C_V = -T\left(\left.\frac{\partial V}{\partial T}\right|_T\right)^2 /(\partial V/\partial P)|_T \tag{147}$$

Because as we found out earlier $\partial P/\partial V|_T \leq 0$ we find that $C_P \geq C_V$ in accord to Eqs. (131) and (132) for ideal gases. But now we see that this inequality is a general result of thermodynamics. It is valid for quantum or classical systems, interacting or not.

Homework, due 10/13. *Prove that*

$$\left.\frac{\partial V}{\partial P}\right|_S = \frac{C_V}{C_P}\left.\frac{\partial V}{\partial P}\right|_T. \tag{148}$$

Give a physical interpretation of this result. What does it mean? Check this relation explicitly for an ideal gas.

C. Le Chatelet principle.

Le Chatelet principle: any external action stimulates process which reduces the action. This principle is so general that it often goes beyond the realm of equilibrium statistical mechanics. The idea is very simple if no opposition to a certain external action is expected then the corresponding process would likely happen by itself, without any need of the external action. Look into society, biology. In some sense a third Newton law is in accord with this principle. Let us see how this principle agrees with the main postulate of thermodynamics.

Consider a system consisting of two parts with e.g. volumes V_1 and V_2 such that $V_1 + V_2 = V$. There is no particle exchange but there is a membrane so that V_1 and V_2 can change keeping the total volume fixed. Instead of volume we can use any other parameter: temperature, magnetization, etc. Now suppose we rapidly change V_1: $V_1 \to V_1 + \delta V$. Then we let the system relax back to thermal equilibrium. Le Chatelet principle tells us that after the system equilibrates $\delta V_1(\infty) < \delta V$. Let us prove this is the case. Then after the initial change we have V_2, P_2 and T_2 are unchanged while V_1, P_1, and T_1 do change. After the system relaxes to the equilibrium the pressures in both sides of the container will equilibrate. During this process the total entropy of the system must increase. The volume V_1 will then have an additional change $\delta V'$.

Let us compute changes in the entropy of subsystems 1 and 2 after the system re-equilibrates

$$\delta S_1 = \left.\frac{\partial S_1}{\partial E_1}\right|_{V_1} \delta E_1 + \left.\frac{\partial S_1}{\partial V_1}\right|_{E_1} \delta V'$$
$$\delta S_2 = -\left.\frac{\partial S_2}{\partial E_2}\right|_{V_2} \delta E_1 - \left.\frac{\partial S_2}{\partial V_2}\right|_{E_2} \delta V' \tag{149}$$

So we get

$$\delta S = \delta E_1 \left(\frac{1}{T_1} - \frac{1}{T_2}\right) + \delta V' \left(\frac{P_1}{T_1} - \frac{P_2}{T_2}\right). \tag{150}$$

Let us assume for simplicity that $T_1 = T_2 = T$, i.e. we initially expanded the volume isothermally. Then $\delta S = (P_1 - P_2)/T\delta V' \geq 0$. Let us note that

$$P_1 = P_2 + \left.\frac{\partial P_1}{\partial V_1}\right|_{T_1} \delta V_1 \tag{151}$$

The derivative $\partial P_1/\partial V_1|_T$ is always negative so we find that $\delta V \delta V' < 0$ i.e. the system counter reacts on changing the volume by reducing this change.

Homework. Due 10/16. Leave extra time for this homework!!!.

(i) Repeat the previous discussion for the situation where we initially change the volume in the first subsystem adiabatically, i.e. conserving S_1.

(ii) Consider a different situation. Assume that we heat the subsystem one at constant volume increasing its temperature by δT. *Prove that after re-equilibration the temperature in the subsystem one will reduce.*

(iii) Consider our Ising spin system in equilibrium at some fixed magnetization M. *You may assume that* $M \ll N$, *where* N *is the total number of spins. Now imagine that in the subsystem one consisting of half of the spins we increased the average magnetization by the amount* ΔM, *which is not necessarily much smaller than* M, *such that the subsystem one is in local thermal equilibrium (but not in equilibrium with the subsystem two). Calculate the new magnetization and temperature in the system after the two subsystems re-equilibrate. Do your results agree with the Le-Chatelet principle?*

Let us discuss how $C_P > C_V$ follows from the Le-Chatelet principle. Assume that we have a gas at a fixed volume V and we heat it by temperature ΔT. Then we increase its energy by the amount $\Delta Q = C_V \Delta T$. By doing this we also increase its pressure. Now assume the system is in the contact with the environment (no energy exchange). Then it will expand

and cool until the pressures become identical. So in order to heat at constant pressure by the same amount δT we need to do additional heating. Thus $C_P > C_V$.

Example of solution of the problem similar to homework. Let us prove that under conditions (i) of the homework above that after adiabatic volume change in part 1 the relaxation will cause pressure to change according to Le-Chatelet principle. Let us rewrite Eq. (150) in the following way

$$\delta S \approx -\delta E_1 \frac{\Delta T}{T^2} - \frac{\Delta T}{T^2} P \delta V_1 + \frac{\Delta P}{T} \delta V_1 = -\frac{\Delta T}{T} \delta S_1 + \frac{\Delta P}{T} \delta V_1. \tag{152}$$

Here we used the notation $\Delta T = T_1 - T_2$ is the temperature change due to initial change of volume. Let us now note that

$$\Delta T = \left(\frac{\partial T}{\partial V}\right)_S \Delta V,\ \Delta P = \left(\frac{\partial P}{\partial V}\right)_S \Delta V. \tag{153}$$

From the Maxwell's relation (obvious from $dE = TdS - PdV$) we have

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V \tag{154}$$

Thus we obtain

$$\delta S \approx \frac{\Delta V}{t}\left[\left(\frac{\partial P}{\partial S}\right)_V \delta S_1 + \left(\frac{\partial P}{\partial V}\right)_S \delta V_1\right] = \frac{\Delta V \delta P}{T} \geq 0 \tag{155}$$

Now we see that because ΔP and ΔV have opposite signs (it follows from the Eq. (148) and the fact that compressibility is positive as well as C_V and C_P are also positive) δP and ΔP must have opposite signs.

VII. INTERACTING SYSTEMS. HIGH TEMPERATURE EXPANSION

So far we were dealing with, in some sense, trivial situations of non-interacting systems. In reality we always have to deal with systems with interactions. Then the partition function can not be written as a product of partition functions corresponding to different degrees of freedom. What shall we do in this situations? Of course one can try to do numerical simulations. However, this is not always possible because we can simulate only a limited number of degrees of freedom and in quantum systems we have to sample over the exponentially large Hilbert space. It is often efficient to use a Monte-Carlo algorithm. There are however subtleties in using Monte-Carlo. E.g. in quantum systems sometimes Monte-Carlo simulations are not efficient because of the sign problem (one needs to evaluate average over many terms with opposite signs). Also we generally understand better qualitative properties if we have analytic results.

A. Ising model.

We will try to use a different root of using perturbation theory. For now e stick purely to classical systems. Suppose we have an ensemble of interacting spins with the Hamiltonian (energy):

$$\mathcal{H} = -h\sum_j \sigma_j - \frac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j = \mathcal{H}_0 + \mathcal{H}_{int}, \tag{156}$$

where $\sigma_j = \pm 1$ is the z-component of the spin. For practical purposes we will mostly deal with the nearest neighbor interaction where $J_{ij} = J$ if i and j form pairs of nearest neighbors in a d-dimensional lattice and zero otherwise. The factor of $1/2$ is to avoid double counting of all bonds. Note that for $J = 0$ our models reduces to the one we studied earlier in Sec. IV.A (we now use slightly different notations).

We need to evaluate the partition function

$$Z = \sum_{\{\sigma_j\}} \exp[-\beta\mathcal{H}], \tag{157}$$

which as we know contains all information about equilibrium properties of our system. Let us note that we expect qualitatively different behavior of the system depending on the value of the product βJ (for simplicity we assume for now that $h = 0$). At high temperatures when $\beta J \ll 1$ we expect that the interaction will be a small perturbation over the equilibrium noninteracting state where all spins have random orientation independent of each other. In the opposite limit we expect that all spins will align to minimize the energy and the state will be ordered. So we can expect that as one changes βJ there can be a qualitative change in behavior characterized by a spontaneous formation of magnetization $\langle\sigma\rangle$ will spontaneously acquire some expectation value at small T. This qualitative change in behavior is called a phase transition and we will come back to this important example later on.

For now we will be interested in a simpler question of understanding hat happens when $\beta J \ll 1$. Let us rewrite the partition function in the following way

$$Z = \sum_{\{\sigma_j\}} \exp\left[\frac{\beta}{2}\sum_{\langle ij\rangle} J_{ij}\sigma_i\sigma_j\right] \exp[-\beta\mathcal{H}_0(\{\sigma_j\})] = Z_0 \left\langle \exp\left[\frac{\beta}{2}\sum_{\langle ij\rangle} J_{ij}\sigma_i\sigma_j\right]\right\rangle_0 \equiv Z_0\langle \exp[-\beta\mathcal{H}_{int}]\rangle_0, \tag{158}$$

where Z_0 is the noninteracting partition function, and $\langle\ldots\rangle_0$ is the expectation value weighted with the noninteracting equilibrium probability distribution. We basically found an exact identity that the interacting partition function Z is the product of the interacting partition function Z_0 and the average of the exponent of $-\beta$ times interaction energy weighted with non-interacting probability distribution. There is a similar expression in quantum mechanics, but it is more subtle because $\mathcal{H}_0$ and $\mathcal{H}$ in general do not commute.

Since we assumed that βJ is small we can expand the exponent in Eq. (158) into the Taylor series and find

$$Z \approx Z_0\left(1+\frac{\beta}{2}\sum_{\langle ij\rangle} J_{ij}\langle\sigma_i\sigma_j\rangle_0+\frac{1}{2!}\frac{\beta^2}{4}\sum_{\langle ij\rangle_0}\sum_{\langle kl\rangle} J_{ij}J_{kl}\langle\sigma_i\sigma_j\sigma_k\sigma_l\rangle_0+\ldots\right) \tag{159}$$

So the partition function can be written in terms of series of moments of $\mathcal{H}_{int}$. Likewise we can write the expansion for the free energy

$$F=-T\ln Z=F_0-T\ln\langle\exp[-\beta\mathcal{H}_{int}]\rangle_0=F_0-T\sum_n\frac{(-\beta)^n}{n!}\langle\mathcal{H}_{int}^n\rangle_0^c. \tag{160}$$

So expansion of the free energy is the cumulant expansion.

First let us evaluate Z and F to the leading order in J. Note that for $i\neq j$ we have $\langle\sigma_i\sigma_j\rangle_0=\langle\sigma_i\rangle_0\langle\sigma_j\rangle_0$, for $i=j$ on the other hand $\langle\sigma_i^2\rangle_0=1$. Therefore

$$Z\approx Z_0\left(1+\frac{\beta}{2}\left[\langle\sigma\rangle_0\right]^2\sum_{i\neq j}J_{ij}+\frac{\beta}{2}\sum_j J_{jj}\right), \tag{161}$$

where N is the total number of spins. For the nearest neighbor model on the cubic lattice (with $2d$ nearest neighbors) we can simplify the expression above further

$$Z\approx Z_0\left(1+\beta NJd\left[\langle\sigma\rangle_0\right]^2\right). \tag{162}$$

Note that in the absence of external magnetic field $\langle\sigma\rangle_0=0$ and there is no correction to the partition function to the linear order in J. Likewise for the free energy we have

$$F\approx F_0-JdN\left[\langle\sigma\rangle_0\right]^2. \tag{163}$$

Using results from Sec. IV.A we find

$$F\approx -TN\ln 2\cosh\beta h-JdN\tanh^2\beta h. \tag{164}$$

So for example the magnetization (recall last lecture: either differentiate the energy with respect to h at a constant entropy or the free energy with respect to h at a constant temperature)

$$M = -\frac{1}{V}\frac{\partial F}{\partial h}\bigg|_T = n\tanh\beta h + 2Jdn\beta\frac{\sinh\beta h}{\cosh^3\beta h}. \tag{165}$$

At high temperatures this yields

$$M \approx n\beta h(1 + 2Jd/T) \tag{166}$$

at low temperatures the correction to the magnetization due to J is exponentially small. Can you explain these results? What is the effective magnetic field each spins sees due to interactions with other spins? Try to argue that at small βh the effective field is $h_{eff} = h + 2Jd\beta h$. This should explain this result for the magnetization.

Let us move to the next order in J. We need to evaluate

$$\langle\sum_{ij}\sum_{kl} J_{ij}J_{kl}\sigma_i\sigma_j\sigma_k\sigma_l\rangle_0 \tag{167}$$

To simplify our life we consider $h = 0$. Then note that the expectation value is non zero only if $i = k$, $j = l$ or $i = l$ and $j = k$. This term gives a trivial contribution since $\langle\sigma_i^2\rangle_0 = 1$. So to the second order in βJ and at $h = 0$ for the nearest neighbor case we find

$$Z \approx Z_0\left(1 + \frac{\beta^2}{8}N4dJ^2\right) = 2^N\left(1 + \frac{\beta^2 J^2 Nd}{2}\right) \tag{168}$$

Here we used the fact that there are $2d$ neighbors for each site and we have two possible allowed permutations with $i = k$ and $i = l$. For the free energy the expansion is similar because at $h = 0$ the second cumulant of $\mathcal{H}_{int}$ is the same as the second moment.

$$\mathcal{F} \approx -TN\ln(2) - N\frac{J^2 d}{2T}. \tag{169}$$

From free energy we can compute quantities like entropy, energy, and specific heat.

$$S = -\frac{\partial F}{\partial T} \approx N\ln(2) - N\frac{dJ^2}{2T^2}. \tag{170}$$

The entropy decreases with J which indicates that the system is more ordered. Note that this is true for both signs of J.

$$E = F + TS = \approx -TN\ln(2) - N\frac{J^2 d}{2T} + TN\ln(2) - N\frac{J^2 d}{2T} = -N\frac{J^2 d}{T}. \tag{171}$$

Finally for the specific heat we have

$$C = \frac{1}{V}\frac{\partial E}{\partial T} = n\frac{J^2 d}{T^2} \tag{172}$$

Homework, due 10/23. Please allow for extra time to do this homework! *Find expansion of the partition function and the free energy for the Ising model up to the order J^2 in non-zero magnetic field in one and two dimensions. Pay special attention to how various terms scale with N. You should find some non-extensive contributions, proportional to N^2 in the partition function, but if you do algebra correctly then such contributions should cancel out in the cummulant expansion for the free energy. So that the free energy remains extensive. You can assume $N \gg 1$. From your expansion find to the same order magnetization and magnetic susceptibility. Interpret your results.*

As an exercise let us compute the next correction. It is easy to convince oneself that there is no contribution to Z and F in the order of J^3. In the order of J^4 we need to compute

$$\frac{1}{4!2^4}(J\beta)^4 \sum_{ij}\sum_{kl}\sum_{mn}\sum_{pq}\langle \sigma_i\sigma_j\sigma_k\sigma_l\sigma_m\sigma_n\sigma_p\sigma_q\rangle_0. \tag{173}$$

This some is non-zero if and only if indices coincide pairwise. It is convenient to compute all various contractions graphically. Each bond is represented by the line, horizontal or vertical. One can show that only loops matter everything else combines into $2^N(\cosh\beta J)^{Nd}$ and the expansion parameter is not βJ but $\tanh\beta J$.

Note that the only closed loops that matter are squares. There are 8 ways to arrange for a square and there are $6 * 2^3 = 48$ ways to make various contractions $j = k$, $l = m$, $n = p$, $i = q$ (first choose the bond then each bond except the first one can be oriented in an arbitrary way) like. So in total squares give the contribution $8 * 48 = 384$. And up to the fourth order we get

$$Z \approx \cosh^{Nd}(\beta J) 2^N (1 + N\tanh^4\beta J) \tag{174}$$

To see that only closed loops matter let us observe that

$$\exp[\beta J\sigma_i\sigma_j] = \cosh(\beta J) + \sinh(\beta J)\sigma_i\sigma_j = \cosh\beta J(1 + \sigma_i\sigma_j\tanh\beta J). \tag{175}$$

So the partition function is indeed written as

$$Z = \cosh^{Nd}(\beta J)\sum_{\{\sigma_j\}}(1 + \tanh(\beta J)\sigma_i\sigma_j). \tag{176}$$

Note that each link in the product appears once and thus repeated loops are excluded. You can use this trick in your homework!

Computing the partition function in a non-zero magnetic field is not a very pleasant exercise (see homework). This is so because once we introduce nonzero h the spin inversion symmetry is gone and many terms, which are zero at $h = 0$ become nonzero. However, very often we are not interested in full statistical information for arbitrary h. For example we can be interested the response of the system to a weak magnetic field, which is characterized by the magnetic susceptibility (per particle):

$$\chi = \frac{1}{N}\frac{\partial M}{\partial h}|_{h=0} = -\frac{1}{N}\frac{\partial F^2}{\partial h^2}\Big|_{h=0} = -\frac{T}{N}\frac{1}{Z^2}\left(\frac{\partial Z}{\partial h}\Big|_{h=0}\right)^2 + \frac{T}{N}\frac{1}{Z}\frac{\partial^2 Z}{\partial h^2}\Big|_{h=0}. \tag{177}$$

Note that the first term in the sum is zero, because it is proportional to the square of magnetization at $h = 0$, which is zero by symmetry. The second term gives

$$\chi = \frac{1}{NT}\frac{1}{Z}\sum_{\{\sigma_{ij}\}}\sum_{ij}\sigma_i\sigma_j\exp[-\beta H] = \frac{1}{T}\sum_j\langle\sigma_j\sigma_0\rangle = \frac{1}{T} + \frac{1}{T}\sum_{j\neq 0}\langle\sigma_j\sigma_0\rangle, \tag{178}$$

where we used the fact the the system is translationally invariant. Note that the expectation value is taken with respect to the full partition function. We know how to expand the denominator in terms of J. We also need to expand the nominator in a similar fashion and read the result.

In the zeros order in J we recover usual Curie susceptibility:

$$\chi = \frac{1}{T} \tag{179}$$

In the first order in J there is no correction to Z so we do not worry about the denominator. In nominator only j which is the nearest neighbor contributes:

$$\chi \approx \frac{1}{T} + \frac{J}{2T^2}\sum_j\sum_{k,l}\langle\sigma_j\sigma_0\sigma_k\sigma_l\rangle_0 = \frac{1}{T} + \frac{2Jd}{T^2}. \tag{180}$$

This is of course what we had before. Let us go to the second order. In denominator we will have to evaluate sums with six spins:

$$\frac{J^2}{8T^3}\sum_j\sum_{kl}\sum_{mn}\langle\sigma_j\sigma_0\sigma_k\sigma_l\sigma_m\sigma_n\rangle_0. \tag{181}$$

Note that in this sum each site should appear twice, otherwise the average is zero. There are two types of terms. The first type is $j = 0$ and ($k = m$, $l = n$ or $k = n$, $l = m$) (this terms is automatically excluded if we restrict summation to $j \neq 0$). This would correspond to a disconnected graph where contractions in the sum coming from the partition functions are independent of the external spins in the sum. Such terms are dangerous because they are not-extensive (they contain extra factor of N). However such terms are precisely canceled by similar terms in denominator Z. So we should not worry about such terms. This is a general statement. The other terms are such where all indices are linked together It is easy to check that in 1D there are only two possibilities to connect next nearest neighbors: $j = 2$, $k = 0$, $m = 1$, $l = 1$, $n = 2$ (plus seven equivalent permutations) and similarly with $j = -2$. In two dimensions there are 12 different permutations, in three dimensions there are $6 + 24 = 30$ permutations etc. So to the next order in J we find (to be specific let us stick to two dimensions)

$$\chi \approx \frac{1}{T}\left(1 + \frac{4J}{T} + \frac{12J^2}{T^2} + \dots\right) \tag{182}$$

B. High temperature expansion for interacting gas.

Let us assume that we have a gas of weakly interacting particles described by the Hamiltonian:

$$\mathcal{H} = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \sum_{i<j} U(q_i - q_j) \tag{183}$$

For simplicity we assume that we include only two-body interactions. In dilute limit particles are unlikely to meet at one point. Long range coulomb interactions can also be written in this two-body form. Tho shorten the notations we will denote the last sum (interaction energy) by $U(\mathbf{q})$, where $\mathbf{q}$ spans coordinates of all particles.

We are going to use the same procedure as for the Ising model:

$$Z(T,V,N) = Z_0(T,V,N)\langle \exp[-\beta U(\mathbf{q})]\rangle_0 = Z_0(T,V,N)\sum_n \frac{(-\beta)^n}{n!}\langle U(\mathbf{q})^n\rangle_0 \tag{184}$$

and similarly

$$F(T,V,N) = F_0(T,V,N) + T\sum_n \frac{(-\beta)^n}{n!}\langle U(\mathbf{q})^n\rangle_0^c. \tag{185}$$

Let us now compute the first and the second order corrections to the free energy. In the first order in U the cumulant is equal to the expectation value:

$$\langle U\rangle_0^c = \sum_{i<j}\int \frac{d^3q_i}{V}\frac{d^3q_j}{V} U(q_i - q_j) = \frac{N(N-1)}{2V}\int d^dq\, U(q). \tag{186}$$

Note that the factors of $1/V$ come from Z_0 and since there is no coordinate dependence of the non-interacting weight all coordinates are equally probable. So in the leading order in interaction we are getting only the uniform shift.

The second order correction is written as the sum

$$\langle U^2\rangle_0^c = \sum_{i<j,\, k<l} \left[\langle U(q_i - q_j)U(q_k - q_l)\rangle_0 - \langle U(q_i - q_j)\rangle_0 \langle U(q_k - q_l)\rangle_0\right]. \tag{187}$$

Note that there is no contribution to this sum if all indices i, j, k, l are different because all coordinates are equally probable. If there is a one common index, e.g. $i = k$ then the same situation is again true. This is because we can choose this coordinate to be equal to zero (by changing variables) and the two remaining coordinates j and l are again uniformly distributed. So the average of the product again factorizes into the product of averages (cf. with Ising model). The only nontrivial contribution comes from the situations where $i = k$ and $j = l$ (note that because $i < j$ and $k < l$ there are no additional permutations (alternatively we can include them adding extra factors of 1/2). So we find

$$\langle U^2\rangle_c^0 = \frac{N(N-1)}{2}\left[\int \frac{d^dq}{V} U^2(q) - \left(\int \frac{d^dq}{V} U(q)\right)^2\right]. \tag{188}$$

Note that the second term is smaller than the first one by a factor of $1/V$ and can be ignored in thermodynamic limit.

So up to the second order in U we have

$$F(T,V,N) \approx F_0(T,V,N) + \frac{N^2}{2V}\left[\int d^dq U(q) - \frac{\beta}{2}\int d^dq U(q)^2\right]. \tag{189}$$

This expression already allows us to calculate pressure, entropy, etc. Unfortunately it is not very useful because interatomic potentials are typically very strong when particles are close to each other. So the integrals diverge at short distances. In these situations instead of expansion in U it is much better to expand in terms of density (small density corresponds to the dilute noninteracting limit). The problem can be addressed by partially resumming the series or by equivalently using the trick simiar to the one we used for the Ising model:

$$\exp[-\sum_{i<j} U_{ij}] = \prod_{i<j}(1 - 1 + \exp[-U_{ij}]) \equiv \prod_{i<j}[1 + f_{ij}], \tag{190}$$

where $f_{ij} = \exp[-U_{ij}] - 1$. Note that f is a much better expansion parameter than U because when the interaction is weak then f essentially coincides with U while if interaction is strong

f approaches -1 instead of ∞. Clearly each order in the expansion in f_{ij} brings us the extra factor of $1/V$ - we need to integrate over more points, and extra factor of N because there are more permutations, so we have the expansion in density as wanted. Physically this cluster expansion means that we solve two particle problem exactly and expand only in the number of events where more and more particles meat together within the interaction range. For example in the leading order on density one sums only over the clusters consisting of two different points i and j:

$$\ln Z = \ln Z_0 + \sum_{n\geq 1} \frac{(-\beta)^n}{n!}\frac{N(N-1)}{2}\int \frac{d^dq}{V}V(q)^p + O\left(\frac{N^3\beta^2U^3}{V^2}\right) = \ln Z_0 + \frac{N(N-1)}{2V}\int d^dqf(q) + O\Bigg(\frac{N^3\beta^2U^3}{V^2} \tag{191}$$

So we see that indeed the first term in the density expansion from re-summation of the series coincides with the first term in the cluster expansion.

In higher orders in density we will find terms of the order of Nn^2 etc. See Kardar for formal rules and diagrams. We will stop our expansion at this order and find corrections to the equation of state. We have

$$F(T,V,N) \approx F_0(T,V,N) - \frac{TN^2}{2V}\int d^dqf(q) \tag{192}$$

Note that if the potential is weak and we expand $f(q)$ to the second order in V we will reproduce Eq. (189). Let us use the following model (van der Walls) potential:

$$U(r) = \begin{cases} \infty & r < r_0 \\ -u_0(r_0/r)^6 & r > r_0 \end{cases} \tag{193}$$

Then we find that

$$-\int d^3qf(q) \approx \frac{4\pi r_0^3}{3}\left(1-\beta u_0\right), \tag{194}$$

where we used high temperature expansion $\beta u_0 \ll 1$ to evaluate the integral. Therefore

$$F \approx F_0 + \frac{N^2}{V}\frac{2\pi r_0^3}{3}(T-u_0) = F_0 + N\frac{\Omega N}{2V}(T-u_0), \tag{195}$$

where $\Omega = \frac{4\pi r_0^3}{3}$ is the excluded volume. Note that our expansion relies on the fact that integral over the potential converges at long distances. So potential should decrease faster than $1/r^3$. Otherwise the virial expansion does not work. We see that for short ranged potential the small parameter is the ratio of the excluded volume and the volume.

From the free energy we can get pressure, entropy etc. In particular

$$P = -\frac{\partial F}{\partial V}_T = \frac{NT}{V} + \frac{N^2\Omega}{2V^2}(T-u_0) + \ldots = nT + \frac{\Omega}{2}n^2(T-u_0) + \ldots \tag{196}$$

This we can rearrange as

$$P + \frac{u_0\Omega}{2}n^2 \approx \frac{nT}{1-\Omega n/2} \tag{197}$$

The correction in the denominator in the second term can be interpreted as the correction to the volume. The factor of one half implies that this is the total excluded volume for all particles: for the first nothing is excluded, for the last $N\Omega$ is excluded so we have $\Omega/2$ on average per particle (remember the homework.) The correction to the pressure comes from the interactions: the particles attract to each other so there is a lower pressure on the boundaries. The equation above is the van der Waals equation with parameters $a = u_0\Omega/2$ and $b = \Omega/2$. (See hand-waving derivation of the van der Waals equation in Kardar).

Van der Waals isotherms have a fundamental flaw that they have a region of negative compressibility: the isotherm $P(V)$ can be non-monotonic. Thus we can find that

$$\frac{1}{\kappa} = -\frac{1}{V}\left(\frac{\partial P}{\partial V}\right)_T = \frac{nT}{(1-nb)^2} - 2an^2 \tag{198}$$

If b is negligible we see that at $T \approx 2an$ we get unstable regions.

The correct phase diagram can be obtained from these isotherms using Maxwell's construction. Note that

$$\Phi = E - TS + PV \tag{199}$$

is the function of T, P, and N (in the sense as we discussed earlier). This implies that

$$\mu = \left(\frac{\partial \Phi(P,T,N)}{\partial N}\right)_{T,P} \tag{200}$$

Because T and P are not extensive we can integrate this equality and we see that $\mu N = E - TS + PV$. Therefore $\Omega = E - TS - \mu N = -PV$. Note that $d\Omega = -SdT - PdV - Nd\mu = -PdV - VdP$ therefore

$$SdT + Nd\mu = VdP \tag{201}$$

If we are on isotherm we can obtain the chemical potential and thus Φ by integrating

$$d\mu = \frac{V}{N}dP \tag{202}$$

along the isotherm. Because the curve is anisotropic we will get three possible values for the chemical potential at the same value of T and P. So the chemical potential is not a unique function of T and P. This indicates there is a phase coexistence (another possibility of course is that van der Walls equations are not very accurate). Which of the phases is

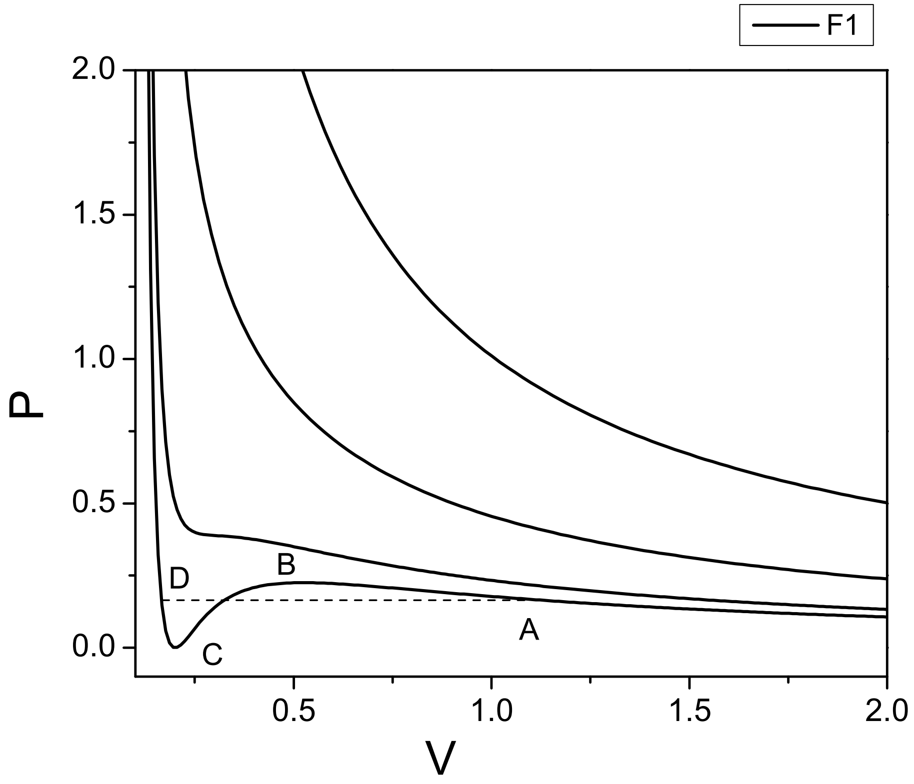


FIG. 3 Van der Waals isotherm for a particular set of parameters $a = b = 0.1$, $N = 1$. The dashed line illustrates Maxwell construction.

correct? In equilibrium if we keep T and P fixed then Φ should be minimal (we will prove this statement later). So we want to choose the minimal μ. The phase transition occurs when to values of the chemical potential intersect:

$$\Delta\mu = \int \frac{V}{N} dP = 0, \tag{203}$$

where the integral is taken along the path $A \to B \to C \to D$, which is nothing but the difference of the areas above and below straight line connecting points A and D. This is the value of the pressure where the phase transition between liquid and gas occurs. In between we have a coexistence region.

Homework, due 11/06. *Problems from Kardar, chapter 5: 5.1, 5.2, 5.4, 5.5, 5.10. You will need to look through Sec. 4.8 to solve the problem 5.5.*

VIII. NONINTERACTING QUANTUM SYSTEMS

Let us first compute the grand canonical partition function for a single degree of freedom:

$$\mathcal{Z}_{\vec{k}} = \sum_{N_{\vec{k}}=0}^{N_{max}} e^{-(\epsilon_{\vec{k}}-\mu)\beta} \tag{204}$$

the upper limit of the sum really depends on whether our particles are fermions or bosons. For bosons the sum stretches to infinity, and we get:

$$\mathcal{Z}^B_{\vec{k}} = \sum_{N_{\vec{k}}=0}^{\infty} e^{-(\epsilon_{\vec{k}}-\mu)\beta} = \frac{1}{1-e^{-(\epsilon_{\vec{k}}-\mu)\beta}} \tag{205}$$

For fermions only up to 1, since there could be at most one particle in the same quantum state:

$$\mathcal{Z}^F_{\vec{k}} = \sum_{N_{\vec{k}}=0}^{1} e^{-(\epsilon_{\vec{k}}-\mu)\beta} = 1+e^{-(\epsilon_{\vec{k}}-\mu)\beta}. \tag{206}$$

For many particles we have to take the double sum over all states and all occupations of the states (note that we are not summing over different permutations each quantum state is counted once)

$$\Omega = -T\ln Z = \sum_k \Omega_k, \tag{207}$$

where $\Omega_k = -T\ln(1+\exp[-\beta\epsilon_k])$ for fermions and $\Omega_k = T\ln(1-\exp[-\beta\epsilon_k])$ for bosons.

From the thermodynamic potential Ω we can find energy:

$$\mathcal{E} = -\frac{\partial lnZ}{\partial\beta} = \sum_k \epsilon_k n_k, \tag{208}$$

where

$$n_k = \frac{1}{\exp[\beta(\epsilon_k-\mu)]\pm 1} \tag{209}$$

is the distribution function of fermions "+" sign or bosons "−" sign.

The two distributions are called Fermi-Dirac and Bose-Einstein respectively. For a fixed number of particles we must have

$$N = \sum_k n_k \tag{210}$$

This equality can be also derived directly by differentiating Ω with respect to μ. Clearly for bosons we must have $\mu \le 0$, where 0 is the minimal possible energy. If the bosons are interacting this is no longer true. The classical Boltzmann's statistic is obtained when

$n_k \ll 1$ which means that the argument of the exponent is large and positive (large negative chemical potential). Then both for fermions and bosons we have

$$n_k \approx \exp[-\beta(\epsilon_k - \mu)]. \tag{211}$$

For bosons there is also another classical limit corresponding to classical waves. TO be specific consider a single oscillator with the Hamiltonian

$$\mathcal{H} = \frac{p^2}{2m} + \frac{m\omega^2}{2} q^2 \tag{212}$$

Then the classical partition function

$$Z_{vib} = \int \frac{dpdq}{2\pi\hbar} \exp[-\beta\mathcal{H}] = \frac{T}{\hbar\omega}. \tag{213}$$

This partition function gives e.g. equipartitioning of different energies. The quantum partition function is

$$Z_{vib}^q = \frac{\exp[-\beta\hbar\omega/2]}{1 - \exp[-\beta\hbar\omega]} \tag{214}$$

In the limit $\beta\hbar\omega \ll 1$ this function becomes identical to the classical one. So now the classical wave (vibrational) limit corresponds to large mode occupancy. The same is true if we have many modes. If the occupancy per mode is large we can treat these modes as classical waves. If the occupancy per mode is small we treat them as particles. We will be more specific later when we discuss superfluidity.

A. Noninteracting Fermions

Let us now focus on noninteracting fermions. It turns out that despite Coulomb interactions the free fermion model is very adequate for describing metals.First let us analyze the zero temperature limit What does this function look like at zero temperature? Assuming that our energy spectrum stretches from zero up, (which we can always do by setting the lowest energy eigenvalue in the problem to zero), if $\mu < 0$ then:

$$f(\epsilon_{\vec{k}}) = \frac{1}{e^{(\epsilon_{\vec{k}} - \mu)/T} + 1} \to 0 \tag{215}$$

since the exponent is large. For positive μ it becomes:

$$f(\epsilon_{\vec{k}}) = \frac{1}{e^{(\epsilon_{\vec{k}} - \mu)/T} + 1} = \begin{cases} 1 & \epsilon_{\vec{k}} < \mu \\ 0 & \epsilon_{\vec{k}} > \mu \end{cases} \tag{216}$$

It is a step like function. Only states below the chemical potential are occupied, we call the chemical potential in this case - the Fermi energy. Let's calculate it. We equate the number of states up to the fermi energy. In the case of a spherically-symmetric dispersion, we can write the k-integral very simply:

$$n = (2s+1)\int\limits_0^{k_F} \frac{d^dk}{(2\pi)^d} = (2s+1)\int_0^{E_F} \rho(\epsilon)d\epsilon = (2s+1)\Omega_d \frac{1}{(2\pi)^d}\frac{k_F^d}{d}, \tag{217}$$

where we included the factor of $2s+1$ to take into account spin degeneracy, $\rho(\epsilon)$ is the single particle density of states:

$$\rho(\epsilon) = \int \frac{d^dk}{(2\pi)^d}\delta(\epsilon - \epsilon(\mathbf{k})) = \Omega_d \left(\frac{\sqrt{2m}}{2\pi\hbar}\right)^d \epsilon^{d/2-1} \tag{218}$$

Sometimes the spin factor $2s+1$ is included in ρ. In three dimensions we have

$$n = \frac{(2s+1)}{4\pi}k_F^2 \tag{219}$$

We can of-course write this in terms of energy:

$$n = \frac{(2s+1)}{4\pi}\frac{2mE_F}{\hbar^2} \tag{220}$$

This defines the fermi energy, but the cool thing lies in the k-space determination: upto a factor of π and the spin-related degeneracy, the fermi wave number, k_F, is just the inverse inter-particle distance (by Pauli principle). In a metal, usually there is just one particle per atom. The distance between atoms is of the order of 5Å. What is the Fermi energy?

$$n = \frac{1}{(2\pi)^3}\frac{4\pi}{3}k_F^3 \cdot 2 \approx \frac{1}{\pi^3}k_F^3 \tag{221}$$

and using the famous near-law of the state of Indiana: $\pi = 3$. Actually, it was $\pi = 3.2$, and the bill almost got through in the year of 1897. It passed a first reading, but not a second one. The final result is then:

$$k_F = \pi\frac{1}{a} = \frac{\pi}{5 \cdot 10^{-10}m} \tag{222}$$

and energy:

$$E_F = \frac{\hbar^2\pi^2}{2 \cdot 10^{-30}kg(5 \cdot 10^{-10}m)^2} \approx 2 \cdot 10^{-19}J = 1.25eV = 1.5 \cdot 10^4 K. \tag{223}$$

Indeed the fermi energy of most metals is very high, and is of the order of between 10^4 and 10^5 K. A more amazing number is the fermi velocity:

$$v_F = \frac{\hbar k_F}{m} = \frac{\pi\hbar}{ma} \approx 10^6 m/s \tag{224}$$

Quite fast for something that is actually at zero temperature!

Let us next compute pressure:

$$-PV = \Omega(T,\mu,V) = -T(2s+1)\sum_{\vec{k}} \ln\left(\mathcal{Z}_k\right) = -V(2s+1)\int \frac{d^dk}{(2\pi)^d} \ln\left(1+e^{-(\epsilon_{\vec{k}}-\mu)\beta}\right) \quad (225)$$

At very small temperatures the integrand in the expression above is 0 for $\epsilon_{\vec{k}} > \mu = E_F$ and is dominated by the exponent in the opposite case so we find that in three dimensions

$$P \approx (2s+1)\int_{\epsilon_{\vec{k}}<E_f} \frac{d^3k}{(2\pi)^3}(E_f - \epsilon_{\vec{k}}) \approx \frac{2}{5}nE_f. \quad (226)$$

To find this result we used that

$$n = (2s+1)\int\int_{\epsilon_{\vec{k}}<E_f} \frac{d^3k}{(2\pi)^3} \quad (227)$$

Another useful property of the ideal Fermi gas is the isothermal compressibility:

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_{T,N} = \frac{1}{n}\left(\frac{\partial n}{\partial P}\right)_{T,N} = \frac{1}{n}\left(\frac{\partial n}{\partial \mu}\right)_T \left(\frac{\partial \mu}{\partial P}\right)_T. \quad (228)$$

Recall that $SdT + Nd\mu = VdP$ therefore

$$\kappa = \frac{1}{n^2}\left(\frac{\partial n}{\partial \mu}\right)_T. \quad (229)$$

At zero temperature the compressibility is equal to

$$\kappa(T=0) = \frac{1}{n^2}\rho_0(E_f) = \frac{3}{2nE_f} = \frac{3}{5}\frac{1}{P}. \quad (230)$$

It is very hard to compress fermions at high densities.

We can relate compressibility also to number fluctuations. Recall that (see Eq. (108))

$$\left.\frac{\partial n}{\partial \mu}\right|_T = \frac{1}{TV}\left(\langle N\rangle^2 - \langle N\rangle^2\right) \quad \Rightarrow \quad \kappa = \frac{n}{T}\frac{\delta N^2}{N} \quad (231)$$

For the ideal Fermi gas we have

$$\delta N^2 = \sum_k f(\epsilon_{\vec{k}}) - f^2(\epsilon_{\vec{k}}) \quad (232)$$

Clearly the fluctuations come only from energies close to E_f because if $f(\epsilon_{\vec{k}}) = 0$ and $f(\epsilon_{\vec{k}}) = 1$ we have $f - f^2 \approx 0$.

Homework, due to 11/13. *Show that Eqs. (231) and (232) reproduce the result (230).*

Another important quantity is the entropy:

$$S = -\frac{\partial \Omega}{\partial T}\bigg|_{\mu,V} \tag{233}$$

Using the explicit form of the thermodynamic potential we can find

$$S = V\int d\epsilon \rho_0(\epsilon)\left[\ln\left(1+\exp[-\beta(\epsilon-\mu)]\right)+\beta(\epsilon-\mu)f(\epsilon-\mu)\right] = -V\int d\epsilon\rho_0(\epsilon)\left[f(\epsilon)\ln(f(\epsilon))+(1-f(\epsilon))\ln(\tag{234}$$

One can recognize that this is indeed the usual von-Neumann entropy where $f(\epsilon)$ serves as the probability that the state is occupied and $1-f(\epsilon)$ corresponds the probability of having the empty state.

Specific heat:

$$c_v = \frac{T}{V}\frac{\partial S}{\partial T}\bigg|_V = \frac{\partial E}{\partial T}\bigg|_V = -\int d\epsilon\rho_0(\epsilon)(\epsilon-\mu)^2\frac{\partial f(\epsilon)}{\partial \epsilon}. \tag{235}$$

The specific heat is also determined by energies close to μ. It vanishes in the zero-temperature limit.

Sommerfeld expansion. Suppose now we want to analyze our thermodynamic functions at small but finite temperature. We need to be able to evaluate integrals like

$$I = -\int_0^\infty d\epsilon G(\epsilon)\frac{\partial f(\epsilon)}{\epsilon}. \tag{236}$$

There is a general way to do this using Sommerfeld expansion. Note that if we have a low temperature $T \ll E_F$ then the Fermi-Dirac distribution is 1 or 0 outside a region $|E-\mu| \lesssim 2T$ so that the derivative of the Fermi fnction is essentially zero. If the temperature is small compared to E_f then this region is very narrow. Note that in metals where E_f 10^4K room temperatures satisfy this criterion $T \ll E_f$. So it makes sense to expand various smooth quantities into Taylor series around $E=\mu$:

$$G(\epsilon) \approx G(\mu) + G'(\mu)(\epsilon-\mu) + \dots \tag{237}$$

With the exponential accuracy in E_F/T the integral over ϵ can be extended to $-\infty$. Then it is easy to see that

$$I(T) \approx \sum_{n\geq 0} b_n T^n \frac{\partial G^n}{\partial \epsilon^n}\bigg|_{\epsilon=\mu}, \tag{238}$$

where

$$b_n = -\frac{1}{n!}\int_{-\infty}^{\infty} dx x^n \frac{\partial f(x)}{\partial x}. \tag{239}$$

The integrals above are tabulated:

$$b_0 = 1, \quad b_2 = \frac{\pi^2}{6}, \quad b_4 = \frac{7\pi^4}{360}. \tag{240}$$

As an example of using the Sommerfeld expansion let us compute energy (per unit volume) as a function of temperature and from this heat capacity and entropy:

$$E = \int d\epsilon \rho(\epsilon)\epsilon f(\epsilon) \tag{241}$$

Let us integrate by parts. Introduce the notation

$$G(\epsilon) = \int_0^{\epsilon} \epsilon' \rho(\epsilon') d\epsilon' \tag{242}$$

Then

$$E = G(\epsilon) f(\epsilon) \Big|_0^{\infty} - \int_0^{\infty} d\epsilon G(\epsilon) \frac{\partial f}{\partial \epsilon} \tag{243}$$

The first terms is clearly zero and the second is of precisely the form (236). So we find

$$E \approx G(\mu) + \frac{\pi^2}{6} T^2 \frac{d^2 G}{d\epsilon^2}\Big|_{\mu} = \int_0^{\mu} \epsilon \rho(\epsilon) + \frac{\pi^2}{6} T^2 (\rho(\mu) + \mu \rho'(\mu)) \tag{244}$$

Note that the second term is of the order T^2 so there we can safely substitute μ by E_F. In the first term we can not do it because μ itself also acquires corrections of the order of T^2. Indeed

$$n = \int d\epsilon \rho(\epsilon) f(\epsilon) \approx \int_0^{\mu} \rho(\epsilon) d\epsilon + \frac{\pi^2}{6} T^2 \rho'(\mu) \tag{245}$$

But n is independent of temperature (we want to find energy vs. temperature at fixed N and V). Note that $\int_0^{E_F} \rho(\epsilon) d\epsilon = n$ and writing $\mu = E_f + \delta\mu$ we find

$$n \approx n + \rho(E_F)\delta\mu + \frac{\pi^2}{6} T^2 \rho'(E_F). \tag{246}$$

Therefore up to T^2 we have

$$\delta\mu \approx -\frac{\pi^2}{6} T^2 \frac{\rho'(E_F)}{\rho(E_F)} \tag{247}$$

Note that if the density of states increases with E then the correction to the chemical potential is negative. This is so because if we keep chemical potential constant, by broadening Fermi function, we would populate more higher energy states than deplete lower energy states. Now we can find energy (per unit volume)

$$E \approx E_F + \delta\mu E_F \rho(E_F) + \frac{\pi^2}{6} T^2 (\rho(E_F) + E_F \rho'(E_F))) \approx E_F + \frac{\pi^2}{6} T^2 \rho(E_F) \tag{248}$$

Thus the heat capacity is

$$c_v = \frac{\partial E}{\partial T}\Big|_{V,N} \approx \frac{\pi^2}{3} T\rho(E_F) = \frac{\pi^2}{2} n \frac{T}{E_F}, \tag{249}$$

where the last equality is true for the free fermions in three dimensions. This is in accord with our previous estimates. Note that

$$c_v = T\frac{\partial S}{\partial T}. \tag{250}$$

Because c_v is linear in T we find that at low temperatures $S \approx c_v$.

Homework, due to 11/13. *Problems from Chapter 7 of Kardar, 7.6, 7.8, 7.17*

B. Noninteracting electrons in magnetic field.

1. Pauli Susceptibility of an electron gas

There are two effects of the magnetic field on electrons: orbital - vector potential couples to the momentum via the term $(\mathbf{p} - e/c\mathbf{A})^2/2m$ and to the spin via the term

$$\hat{\mathcal{H}}_{spin} = -g\mu_B \vec{h}\cdot\hat{\vec{\sigma}} = -g\mu_B h\sigma_z, \tag{251}$$

where we chose the spin-quantization access along the magnetic field. Now, spin up has less energy then spin down. The chemical potential of the electron gas is still the same for the two spins, but this implies that the fermi energy - which here we distinguish from the chemical potential by saying that it is the *kinetic* energy up to which electron states are filled - is not the same for the two electron flavors.

For down electrons, the total energy is:

$$E_\downarrow = \epsilon_{\vec{k}} + hg\mu_B \tag{252}$$

while

$$E_\uparrow = \epsilon_{\vec{k}} - hg\mu_B \tag{253}$$

Now we need to find the chemical potential at which the up and down systems are at equilibrium. Formally we can do this by requiring that

$$n = \int_0^\infty \rho(\epsilon)\frac{1}{\exp[\beta(\epsilon + g\mu_B h - \mu)]} + \int_0^\infty \rho(\epsilon)\frac{1}{\exp[\beta(\epsilon - g\mu_B h - \mu)]} \tag{254}$$

Here $\rho(\epsilon)$ is the density of states for spinless particles (i.e. not multiplied by $2s+1=2$). Note that at zero temperature and small magnetic fields the density of states can be treated as a constant so

$$n \approx \rho(E_F)(\mu - g\mu_B h + \mu + g\mu_B h) = 2\mu\rho(E_F). \tag{255}$$

Therefore there is no correction to the chemical potential in the linear order in h therefore $\mu \approx E_F$. This means $E_{F\uparrow} \approx E_F - g\mu_B h$ and $E_{F\downarrow} \approx E_F + g\mu_B h$. This relations can be understood also in the following way. Since the energy shift for up is minus that of down, some electrons from the high-energy down spins, would flip to point in the direction of the magnetic field. This will happen until we have

$$E_{F\uparrow} - E_{F\downarrow} = 2hg\mu_B \tag{256}$$

After that it is already energetically unfavorable to flip flop electrons.

So we have an excess of upspins, and we can easily calculate how many:

$$n_\uparrow - n\downarrow = \int\limits_{\mu - hg\mu_B}^{\mu + hg\mu_B} \rho(\epsilon) d\epsilon = 2\rho(E_F)hg\mu_B \tag{257}$$

and the excess spin magnetization is:

$$g\mu_B(n_\uparrow - n\downarrow) = 2\rho(E_F)h(g\mu_B)^2 \tag{258}$$

So the susceptibility is:

$$\chi_{Pauli} = 2\rho(E_F)(g\mu_B)^2 = 3n\frac{(g\mu_B)^2}{E_F}. \tag{259}$$

It is a direct measure of the density of states at the Fermi surface. Free electron spins tend to point in the direction of the field, but don't actually have magnetization in zero field. Hence this is called **paramagnetism**.

Homework, due 11/13. *Assume that temperature is small compared to E_F but finite. Find the temperature correction to the magnetization (still to the linear order in h. Argue that this result can be also understood using picture of two fermi gases (up and down) where we start from the unpolarized state and flip electrons until the free energy reaches the minimum.*

Now, in addition to this effect there is also going to be an orbital effect on the general magnetization, and the question rises - how do we measure the Pauli susceptibility alone? This is done using the Knight shift - The energy level of nuclear spins are coupled to the electronic spins through the contact term of the hyper-fine hamiltonian of an atom. This coupling is much stronger than the coupling of nuclei to a raw magnetic field (since their Bohr magneton is three orders of magnitude smaller than that of the electron's). Using NMR one can get very accurate readings of nuclear energy levels, and there one see's precisely this quantity.

2. Orbital effects of a magnetic field - Shubnikov-De-Haas-van Alpen oscillations and Landau diamagnetism

Note that classically there is no diamagnetism. The partition function of free classical electron gas is

$$Z = \frac{1}{N!}\left[\int \frac{d\mathbf{p}d\mathbf{q}}{(2\pi\hbar)^3}\exp[-\beta(\mathbf{p}-e/c\mathbf{A})^2/2m]\right]. \tag{260}$$

The vector potential or the magnetic field can be removed by a simple change of variables $\mathbf{p} \to \mathbf{p} + e/c\mathbf{A}$. However, quantum mechanically the situation is quite different. Quantum mechanically we can introduce a new energy scale $\hbar\omega_c$, where

$$\omega_c = \frac{eB}{mc} \tag{261}$$

is the cyclotron frequency, which is the frequency of rotation of a classical particle in a magnetic field (recall $m\omega^2 R = e/c\omega RB$ so that $\omega = eB/(mc)$). In SI units there is no c in denominator. Since the energies must be quantized into multiples of that frequency we expect that at least at small temperatures $T \ll \hbar\omega_c$ there will be a strong difference between quantum and classical pictures. The discrete energy levels of electrons in magnetic field are called Landau-levels. Like in an Harmonic oscillator, the allowed energy eigenvalues of a particle in a magnetic field are:

$$E_n = \left(\frac{1}{2}+n\right)\hbar\omega_c \tag{262}$$

In two dimensions, instead of having a flat density of states, we suddenly have a comb like shape with spikes at these energies.

We can estimate the degeneracy of each level for each spin N_L from a simple argument. (Let us do it in 2D for simplicity). The total number of states with and without magnetic field should be the same therefore

$$N_L = A\int_{0\leq\epsilon_{\vec{k}}\leq\hbar\omega_c}\frac{d^2k}{(2\pi)^2} = \frac{A}{2\pi\hbar^2}m\hbar\omega_c = \frac{A}{2\pi\hbar^2}m\hbar\frac{eB}{mc} = \frac{\Phi}{\Phi_0}, \tag{263}$$

where $\Phi = Ah$ is the total magnetic flux and

$$\Phi_0 = \frac{2\pi\hbar c}{e} \tag{264}$$

is the flux quantum. There is an alternative semiclassical explanation to why the Landau level degeneracy is such. When you let an electron go in a closed orbit, the phase of the wave function (by gauge invariance) acquires the value:

$$\varphi \sim \frac{e}{\hbar}\int d\vec{l}\cdot\vec{A} = 2\pi\frac{\Phi}{h/e} \tag{265}$$

Whenever φ is a multiple of 2π, the electron can be at rest. So each electron orbit should encircle at least one flux quantum. The area of each orbit is therefore Φ_0/B. So the total number of electrons we can fit to area A is thus $N_L = A/(\Phi_0/B) = \Phi/\Phi_0$.

A magnetic field is like a flux density. The most useful way is to think about it as the field for a given 2d fluxon density:

$$B = \frac{\Phi_0}{b} \tag{266}$$

as it turns out, for each landau level, there is a density of states which is exactly b - just as though there is one electronic state for each fluxon. This means that the density of states function for an electron gas in 2d is (ignoring spin) :

$$\rho(\epsilon) = \sum_{n=0}^{\infty}\delta(\epsilon - (\frac{1}{2}+n)\hbar\omega_c)\frac{B}{\Phi_0} \tag{267}$$

The integer quantum hall effect occurs when the number of electrons fits exactly an integer number of landau levels. We define the filling factor:

$$\nu = \frac{n}{B/\Phi_0} \tag{268}$$

ν tells us how many landau levels are full. The Qauntum-Hall effects occur when ν hits the magic numbers $1, 2, 3\ldots$ for integer, and a variety of rational fractions - most notably $1/3$, for which Laughlin Stoermer and Tsui won their nobel prize for.

But those regime of small ν, ie, of high field, they are complex, and have a lot of structure, which , unfortunately, lies beyond the scope of this class. We will concentrate on the relatively simple, but nevertheless fascinating phenomena at low fields.

Recall that in the case of the Pauli susceptibility, everything happens near the edge of the fermi surface. Looking at the 2d electron gas, when we turn on a magnetic field, the electronic states just bunch into delta-functions. All those states that are filled - instead of being a continuum, they are just bunched together. Even if we turn on a spin-field, as in the Pauli susceptibility, the depolarization occurs. It is clear that near the Fermi surface - the thing that really matters once a field is on, is how filled the last landau level is. This is:

$$mod(\nu) \tag{269}$$

Every quantity of the system will depend strongly on $mod(\nu)$, but only weakly on what ν is, or on the magnetic field. Thus, every quantity will undergo an oscillation, whenever ν goes through 1. But This means:

$$\nu = \int_{\epsilon_{\vec{k}}<E_F} \frac{d^2k}{(2\pi^2} \frac{\Phi_0}{B} = \frac{1}{2\pi^2} \frac{V_K \Phi_0}{B}, \tag{270}$$

where V_K is the area of the Fermi surface in the momentum space. Note that in 3D the momentum orthogonal to magnetic field is conserved and thus we this area law for each z-component of the wave vector. The largest effect to the oscillations will be whenever the density of states is largest, i.e. when the area satisfying ν =integer reaches extremum (maximum or minimum on the Fermi surface). So as a function of $1/B$ we'll get oscillations in any quantity that are with 'wave-length':

$$\Delta(1/B) = \frac{(2\pi)^2}{V_K \Phi_0} \tag{271}$$

which corresponds to $1/B$ changes that move ν to $\nu - 1$. These oscillations are called Shubnikov-De-Haas-Van-Alpen oscillations. This result is remarkable - if you measure the resistance of a metal in a varying field, plot the result with respect to $1/B$, we'll see oscillations, which frequency, corresponds to the area enclosed in the fermi surface (note the shape of the Fermi surface may not be necessarily a sphere in real crystals).

There's another thing. Because the dispersion of the electrons changes, so does their energy. In fact it rises:

$$E = E_0 + \frac{1}{2} \chi_L B^2 \tag{272}$$

The extra kinetic energy due to the magnetic field means also that the electrons, in their orbits, produce a magnetization contrary to the external field and try to reduce it. This contribution to the magnetization is called Landau-diamagnetism. In the problem set you'll find:

$$\chi_L = -\frac{1}{3}\chi_{Pauli} \tag{273}$$

in a free electron gas.

A bit of terminology - free spins tend to point at the direction of the external field, but not to reorder ferromagnetically on their own - this is paramagnetism, and it implies a negative magnetic energy. Diamagnetism is the opposite - the magnetic moment tries to reduce the external field - which is diamagnetism. The difference between diamagnetism and AFM is that without a field there isn't any spontaneous canted internal field.

Homework due 11/20. *Landau diamagnetism*

Begin Homework

Landau diamagnetism. The electrons are constrained to move in a plane, the electronic states become Landau levels. Using that their energies are $\epsilon_n = \hbar\omega_c(n + 1/2)$ and the degeneracy of each level is AB/Φ_0 find the energy increase per area due to magnetic field:

$$\Delta E \approx -\frac{1}{2}\chi_l h^2 \tag{274}$$

(where χ_L) is the Landau diamagnetic susceptibility. In this problem we will calculate it. (a) What is the energy of the electron gas in a magnetic field B in terms of its temperature and chemical potential? No need to evaluate the infinite sum yet. Ignore the Zeeman spin-splitting due to the magnetic field. (b) To evaluate the total energy, it is necessary to convert the sum to an integral. Prove the following approximation:

$$\eta\sum_{n=0}^{\infty} f(\eta(n+1/2)) \approx \int_0^{\infty} f(x)dx - \frac{\eta^2}{24}\int_0^{\infty} f''(x)dx = \int_0^{\infty} f(x)dx + \frac{\eta^2}{24}f'(0) \tag{275}$$

to second order in h. (c) Using the approximation of 1b, find χ_L for a noninteracting electron gas. Assume T close to zero, but take the limit $T \to 0$ only once you obtain an expression for χ_L, i.e. consider the order of limits where the magnetic field vanishes first (at finite T) and then temperature is set to zero.

End Homework

C. Bose-Einstein statistics

The distribution for bosons is quite different from that of fermions, although the numerical expression is quite similar:

$$n(\epsilon) = \frac{1}{e^{(\epsilon-\mu)\beta} - 1} \tag{276}$$

The big difference comes from the minus sign. It is possible to have a divergence of the distribution. The properties of bosonic gasses are mostly determined by what happens near this divergence.

First consider two archetypical bosonic systems. Photons and phonons. and we will derive the black body radiation formulas, and also the Debye theory for heat capacity. Both of these are important examples for the success of quantum mechanics in explaining deviations of experimental measurements from classical results.

1. Black body radiation

The phenomena of black-body radiation requires little introduction. By the end of the nineteenth century, from observing hot metals, as well as stars, it was clear that:

- Hot bodies emit electromagnetic radiation that is unrelated to their spectrum.
- the maximum wavelength of the radiation is inversely proportional to their temperature:

  $$\lambda_m T = constant \tag{277}$$

  which was known as Wien's law.
- The total radiation emitted per unit area of a hot body (body hotter than its environment) is given by:

  $$I/A = \sigma T^4 \tag{278}$$

  where $\sigma = 5.67 \cdot 10^{-8} \frac{W}{m^2 K^4}$, which is known as Stefan's law.

Let us try to understand this result classically. Like the wave function of a particle in a box, the electromagnetic field has solutions of the maxwell equations, or the E-M wave equation:

$$\frac{\partial^2 \vec{A}}{\partial t^2} = \nabla^2 \vec{A} \tag{279}$$

which are $e^{i\vec{k}\cdot\vec{x}}$. There are two solutions for each k value - two polarizations $\vec{n}\cdot\vec{k}=0$.

Each value of $\vec{k}$ was considered to be a harmonic mode, and therefore, just like a spring, each mode, by equipartition theorem gets:

$$\langle\epsilon_k\rangle = T, \tag{280}$$

where ϵ_k is the energy associated with each mode (It can be obtained from writing classical energy density $(\mathbf{E}^2+\mathbf{B}^2)/8\pi$ in terms of modes in the box. Note these are purely classical considerations.

The dispersion of light as we know:

$$\omega = c|\vec{k}| \tag{281}$$

and therefore the energy in the range ω, $\omega+d\omega$ is the number of modes in the spherical shell $k=\omega/c$ to $k+dk=(\omega+d\omega)/c$, times T:

$$f(T,\omega)d\omega = 2\cdot\frac{4\pi k^2}{(2\pi)^3}dkT = \frac{1}{\pi^2}T\frac{\omega^2}{c^3}d\omega \tag{282}$$

Now we need to integrate this expression over all modes and get

$$f(T) = \int_0^\infty d\omega\frac{1}{\pi^2}T\frac{\omega^2}{c^3} \to \infty \tag{283}$$

So there is something wrong in this purely classical formulation. We need to use correct Bose-Einstein distribution which does not give equipartition.

Instead of the equipartition theorem we need to put our better guess. Let's calculate quickly the partition function for a photonic state. A state can have one photon, two photons, etc. Each photon in state k has energy:

$$\epsilon_{\vec{k}} = \hbar\omega_k = \hbar ck. \tag{284}$$

The partition function is:

$$Z = 1 + e^{-\epsilon_{\vec{k}}\beta} + e^{-2\epsilon_{\vec{k}}\beta}\ldots = \frac{1}{1-e^{-\epsilon_{\vec{k}}\beta}}. \tag{285}$$

Note that there is no chemical potential. The latter appeared because we needed to work with fixed (or slightly fluctuating number of particles). Alternatively we split the system into two subsystems and introduced chemical potential because we needed to maximize entropy with respect to exchanging particles between the two subsystems. That was the reason that $\partial S/\partial N_1$ had to be maximized and that derivative we called chemical potential. Here number of photons is not conserved (like number of quanta in an oscillator) so we automatically some over all occupancies of all modes without any constraints.

The number of photons in the state, is therefore like the BE distribution, only with $\mu = 0$. We differentiate the log with respect to β and obtain:

$$n_k = -\frac{\partial \ln Z}{\partial \beta} = \frac{1}{e^{\epsilon_{\vec{k}}\beta} - 1} \tag{286}$$

At this point let us look at this expression. If $\epsilon_{\vec{k}} = \hbar\omega_k$ is much smaller than T,

$$\epsilon_{\vec{k}} = \hbar\omega_k \ll T, \tag{287}$$

we can expand the exponential in a power series, and get:

$$n_k \approx \frac{T}{\epsilon_{\vec{k}}} \tag{288}$$

and therefore an energy:

$$\epsilon_{\vec{k}} n_k \approx T, \tag{289}$$

which is indeed equipartition. But if the energy is much larger than T:

$$\epsilon_{\vec{k}} = \hbar\omega_k \gg T, \tag{290}$$

the exponent in the denominator of the BE distribution will dominant over the 1, and we would have:

$$\epsilon_{\vec{k}} n_k \approx \epsilon_{\vec{k}} e^{-\epsilon_{\vec{k}}\beta} \tag{291}$$

Note that a very similar argument works in other situations. For instance, a diatomic molecule has a rotational degree of freedom and also a vibrational degree of freedom. The first excited state of the rotational degree of fredom is:

$$\frac{L^2}{2I} \sim \frac{\hbar^2}{2md^2} \sim \frac{10^{-68}}{10^{-26}(10^{-10})^2} = 10^{-22}J = 10K \tag{292}$$

So in room temperature we can forget about quantization of angular momentum and treat them classically (equipartition works). The vibrational degrees of freedom, that have a spectrum very much like that of photons in a k state, will only get excited therefore when T is of order of their energy. The frequency of chemical bonds is such, that they start playing a role at temperatures of the order of thousands of kelvins. So at room temperatures vibrational degrees of freedom are completely frozen and do not contribute to e.g. specific heat.

Using the correct Bose-Einstein distribution we find

$$f(T,\omega)d\omega = 2\cdot\frac{4\pi k^2}{(2\pi)^3}dk\frac{1}{e^{\epsilon_{\vec{k}}\beta}-1} = \frac{1}{\pi^2}\frac{\omega^2}{c^3}d\omega\frac{\omega\hbar}{e^{\omega\hbar\beta}-1} \tag{293}$$

This is the Planck formulae for the power spectrum of the black body. To get the total energy density we need to integrate over all frequencies:

$$f(T) = \frac{1}{\pi^2}\int\limits_0^\infty\frac{\omega^2}{c^3}d\omega\frac{1}{e^{\omega\hbar\beta}-1} \tag{294}$$

by changing variables we can get all the dependencies outside the integral. The change of variables is dictated by the exponent, and we define

$$x = \omega\beta\hbar \tag{295}$$

we then obtain

$$f(T) = \frac{1}{\pi^2}\int\limits_0^\infty\frac{\omega^2}{(\hbar\beta c)^3}d\omega\frac{1}{e^x-1}\frac{1}{\beta} = \frac{T^4}{\pi^2\hbar^3c^3}\int\limits_0^\infty dx x^3\frac{1}{e^x-1} \tag{296}$$

The integral above can be evaluated exactly:

$$\int\limits_0^\infty dx x^3\frac{1}{e^x-1} = \frac{\pi^4}{15} \tag{297}$$

so that

$$f(T) = \frac{\pi^2}{15}\frac{T^4}{c^3\hbar^3}. \tag{298}$$

From this density we still need to get the radiance per unit area. Assume that we have a small hole in a black body. Then the average velocity of outgoing photons is

$$\langle c_\perp\rangle = \frac{c}{4\pi}\int_0^{\pi/2}d\theta\int_0^{2\pi}d\phi\sin\theta\,\cos\theta = \frac{c}{4}. \tag{299}$$

So the average flux of energy per unit area is

$$I = \frac{c}{4}f(T) = \sigma T^4, \tag{300}$$

where

$$\sigma = \frac{\pi^2}{60}\frac{k_B^4}{\hbar^3c^2} = 5.67\,10^{-8}\frac{W}{m^2K^4} \tag{301}$$

We can also compute pressure by black body radiation:

$$P = \frac{1}{3}\frac{E}{V} = \frac{1}{3}f \approx 2.5\,10^{-18}\,T^4\,\mathrm{Pa}, \tag{302}$$

where 1/3 comes from the relativistic dispersion (there is also a Casimir pressure due to zero point energy, which is beyond the scope of this course). Note that normal atmospheric pressure is about 10^5 Pa.

2. Debye theory of heat-capacity of solids

Bose-Einstein statistics is closely related to the statistics of waves. Of course electromagnetic waves are one of the primary examples. Another important example are the phonons or the sound waves in solids and liquids. We will concentrate on solid where phonons are just lattice vibrations or displacement waves. Let's consider a simple example of a chain of masses and springs in 1d. Each mass has the equation of motion:

$$m\frac{\partial^2 x_n}{\partial t^2} = -\kappa(x_{n+1} - x_n) + \kappa(x_n - x_{n-1}) \tag{303}$$

And let us assume periodic boundary conditions - like the particle in the box, this is solved by:

$$x_n = e^{iqna - i\omega t} \tag{304}$$

and the dispersion relation becomes:

$$\omega = \sqrt{\frac{\kappa}{m}(1 - \cos qa)} \tag{305}$$

where a is the lattice constant, and we use q as the wave vector.

Now, what is the range of q? Note that if we add $2\pi/a$ to q then the solution (304) will not change so we can choose the range

$$-\pi < qa < \pi \tag{306}$$

There is another reason of why we the domain of q is bounded. Let us compute the total number of allowed q. The latter is quantized in units

$$dq = \frac{2\pi}{L} \tag{307}$$

If the modes are restricted to be between $\pm\pi/a$, then we have:

$$g = \frac{2\pi}{a} / \frac{2\pi}{L} = \frac{L}{a} = N \tag{308}$$

with N being the number of particles. That's right! can't have more modes than particles!

Another feature of the dispersion is the low q, or low a, behavior. When qa is small, we can expand the cosine and obtain:

$$\omega = \sqrt{\frac{k}{m}} qa \tag{309}$$

This implies that the force equation can also be approximated by:

$$m\frac{\partial^2 x}{\partial t^2} = a^2 k \frac{\partial^2 x}{\partial r^2} \tag{310}$$

Note that here r is the physical location of the mass (instead of x_n we use $x(r)$ in the continuum language) and x is its displacement from the equilibrium position. So e.g. x is always small while r can be arbitrary number between 0 and L. We see that at low frequencies, we can approximate lattice vibrations like a sound wave.

This equation is easily generalizable to higher dimensional solids:

$$\frac{\partial^2 x}{\partial t^2} = c^2 \frac{\partial^2 x}{\partial r^2} \tag{311}$$

with c the speed of sound. In a solid though, there would be three sound modes for every k - two transverse and one longitudinal. Except for that, sound is just like light!

Another comment before starting to work is that - every mode has its own frequency. therefore each mode is like a quantum harmonic oscillator, with energies:

$$E_n = (n + 1/2)\hbar\omega \tag{312}$$

This can be also understood if we work with momentum and coordinate operators and go to normal modes using the quantum Hamiltonian. The nice thing about hamronic systems is that quantum and classical equations of motion coincide. Now, let's discuss the energy stored in these vibrational modes in a 3d solid. We have the following info to go by:

- Generalizing from the 1-d example: we have a total number of $3N$ vibrational modes - since we have $3N$ degrees of freedom.

- At low wave numbers or frequency, the dispersion is linear, and we can approximate the vibrations as sound waves.

- each mode acts like a quantum harmonic oscillator:

$$E_n = (n + 1/2)\hbar\omega \tag{313}$$

n is the number of *phonons* in the state.

Let's start with high temperatures. By high we mean temperature higher than the largest energy of the phonon (corresponding to $q = \pi/a$ in our case). Let us call this energy scale as Debye temperature:

$$T_D = \alpha\hbar\frac{c}{a} \tag{314}$$

This is an approximation for the highest frequency that a vibrational mode can hold, α is a numerical coefficient of order 1.

If $T \gg T_D$, we can forget about quantum mechanics, and we have equipartition, T energy in each mode.

$$U = 3NT \tag{315}$$

this is called the *Dulong and Petit law.* This what we would get for light if the spectrum of photons was bounded. The heat capacity at high-T is thus:

$$C_V = 3N \tag{316}$$

At very low temperatures we can use the sound-wave approximation because only small frequency modes corresponding to small q are populated. Let us assume, for simplicity, that the speed of sound for all three sound modes is the same, c, although they are unlikely to be so. If we have three sound modes than, at low T, the energy density for phonons is just like it is for photons:

$$U/V = 3\int_0^\infty \frac{4\pi\omega^2 d\omega}{(2\pi c)^3}\frac{\hbar\omega}{e^{\hbar\omega\beta} - 1} = \frac{\pi^4}{15}\frac{T^4}{\pi^2 c^3\hbar^3}, \tag{317}$$

where the prefactor 3 corresponds to three different modes. This expression is very similar to what we got for photons. The heat capacity now is the famous Debye law:

$$C_V = \frac{\pi^4}{5}\frac{T^3}{\pi^2 c^3\hbar^3} \tag{318}$$

So the heat capacity is constant and high temperatures and rapidly decreases to zero as T^3 at low temperatures. In between we should get some kind of interpolation which is sensitive to the details of the dispersion relations.

3. Bose statistics in systems with conserved number of particles. Bose-Einstein condensation.

Recall that the occupation number per mode, according to Bose-Einestein statistics is

$$f_b(\epsilon) = \frac{1}{\exp[\beta(\epsilon - \mu)] - 1} \tag{319}$$

We need to find μ by fixing the total density of particles n:

$$n = \int \frac{d^d q}{(2\pi)^d} \frac{1}{\exp[\beta(\epsilon_q - \mu)] - 1}. \tag{320}$$

At high temperatures we expect μ to be large and negative. Indeed if this is so we have

$$n \approx \exp[\beta\mu] \int \frac{d^d q}{(2\pi)^d} \exp[-\beta \frac{\hbar^2 q^2}{2m}] \approx \exp[\beta\mu] \left(\frac{m}{2\pi^2\hbar^2}\right)^{d/2} T^{d/2} \int d^d\xi \exp[-\xi^2] \tag{321}$$

So we see that at large temperatures $T \gg T_D \approx \hbar^2 n^{2/d}/(2m)$ indeed $\exp[\beta\mu] \ll 1$. What happens as T approaches the degeneracy temperature? Note that T_D is the same object (up to various factors of π) as the Fermi energy. This is the scale where de Broglie wavelength associated with the average energy of the particles T is comparable to inter-particle spacing. This is the regime when particle start to acquire wave-like phenomena, they can not be treated as point objects any more. This is also the regime where statistics bosonic vs. fermionic becomes important.

Obviously as T decreases chemical potential increases approaching zero. Eventually as $T \to 0$ it should reach the asymptotic value, which for bosons must be zero (since contrary to fermions it can not cross it). So we must have

$$n \to \int \frac{d^d q}{(2\pi)^d} \frac{1}{\exp[\beta\hbar^2 q^2/(2m)] - 1} \tag{322}$$

Note that in $1D$ and $2D$ this integral diverges at small q meaning that we should introduce small negative chemical potential to make it convergent. As temperature goes to zero so does chemical potential as T in $1D$ and as $\exp[-1/T]$ in $2D$. However in $3D$ this integral converges. So the chemical potential vanishes at finite temperature such that

$$n \approx \zeta_{3/2} \left(\frac{mT}{(2\pi\hbar)}\right)^{3/2}, \tag{323}$$

which defines the temperature of Bose-Einstein condensation:

$$T_c = \frac{(2\pi\hbar)^2}{m} \left(\frac{n}{\zeta_{3/2}}\right)^{2/3}, \tag{324}$$

here $\zeta_{3/2} \approx 2.612$ is the zeta function:

$$\zeta_m = \frac{1}{\Gamma(m)} \int_0^\infty \frac{dx x^{m-1}}{\exp[x] - 1}. \tag{325}$$

As T becomes smaller than T_c we must have $\mu = 0$ and one state (the ground state) gets macroscopic occupation. Recall that we have to sum over all states (different k) rather then integrate over them. The summation is equivalent to integration only if the function $f_b \epsilon_k$ is smooth. If one state is macroscopically occupied this is no longer true.

Below T_c the Bose gas has unusual properties. The population of excited modes does not depend on the number of particles. So we are essentially dealing with an ensemble where N is not conserved like phonons or photons. In other words the condensate is like a reservoir of particles out of which we can take particles for free: it does not cost any entropy or energy. The dispersion of this particles is quadratic (not linear) but it is an artifact of non-interacting model. With interactions the dispersion becomes linear. The low energy modes are macroscopically occupied so we can treat them as waves (in quantum mechanics such modes are described by the coherent states where both number and phase can be well defined). The gas pressure for $T < T_c$ is

$$P = \frac{2}{3} \int \frac{d^3 q}{(2\pi)^3} \frac{\epsilon_q}{\exp[\beta \epsilon_q] - 1} = T \zeta_{5/2} \frac{1}{\lambda_D^3} \propto T^{5/2} \tag{326}$$

is independent on the number of particles. Here $\lambda_D = \sqrt{2\pi\hbar/mT}$ is the de Broglie wavelength. This means for example that below T_c pressure is independent of volume (again like in the case of photons).

Homework. Due 12/02. *Read Chapters 7.1, 7.2 from Kardar. Re-derive Eqs. (7.18)-(7.21) for two dimensions. Problems from Kardar: 7.13 (see also solution to problem 1.5 for Clausius-Clapeyron law), 7.15, 7.16*

We will discuss relation of Bose-Einstein condensation and superfluidity in the next chapter.

IX. BROKEN SYMMETRY: MEAN-FIELD AND VARIATIONAL APPROACH.

This section closely follows notes by S. Girvin (Yale).

A. Ising model

Let us return back to the classical Ising model:

$$\mathcal{H} = -J\sum_{\langle ij\rangle}\sigma_i\sigma_j - h\sum_i \sigma_i. \tag{327}$$

Recall that the magnetization

$$m = -\frac{1}{V}\frac{\partial F}{\partial h} = \frac{T}{V}\frac{\partial \ln Z}{\partial h} = \frac{1}{V}\sum_j \sigma_j \tag{328}$$

In the absence of magnetic field the model has the spin-inversion symmetry $\sigma_j \to -\sigma_j$: two directions of spins are equivalent. However in the presence of magnetic field this symmetry is broken. So h is the symmetry-breaking field.

Now let us introduce a very important concept of spontaneous symmetry breaking. When $h = 0$ we must have $m = 0$. But what happens if $h \to 0$ and $V \to \infty$. We can have to possible limits

$$\lim_{h\to 0}\lim_{V\to\infty} m \tag{329}$$

and

$$\lim_{V\to\infty}\lim_{h\to 0} m \tag{330}$$

The latter limit is always zero. But this is not true about the first one. If the first limit is not-zero the system is said to be in spontaneously broken symmetry state. Clearly the order of limits can be different if the free energy is a non-analytic function of h in the limit $V \to \infty$.

To illustrate the idea let us consider spin system at very low temperature. First start from two spins. Assume that J is very large both compared to h and to T. The two favorable spin configurations are $|\uparrow,\uparrow\rangle$ and $|\downarrow,\downarrow\rangle$. The energy difference between them is $\Delta\mathcal{E} = 4h$. Once this energy difference becomes much smaller than temperature (at least according to naive criterion) we obviously have $m \to 0$. This happens at $h \sim T/4$. Now consider instead 4 spins. By the same argument the energy difference between two favorite configurations $|\uparrow,\uparrow,\uparrow,\uparrow\rangle$ and $|\downarrow,\downarrow,\downarrow,\downarrow\rangle$ is $\Delta\mathcal{E} = 8h$ so the spins become disordered at $h \approx T/8$. We see that as the number of spins increases the "naive" argument predicts that spins remain ordered up to smaller and smaller fields, which vanish as the number of spins become macroscopic. Similarly one can imagine that the times required to go from one macroscopic configuration to another become exponentially long.

We will see later that the naive argument is qualitatively correct in dimensions higher than one, where at small temperatures entropy is not important and we can focus only on the two optimal-energy configurations. The ordering is clearly a *collective* effect. Noninteracting spins will never order! At high temperatures $T \gg J$ on the other hand spins are almost independent from each other and weak magnetic field can not have a pronounced effect. So we expect that there is a transition between ordered regime (where symmetry is spontaneously broken) and disordered regime. How do we find this transition? There are several available tools: meanfield approach, variational approach, numerical simulations, analysis of asymptotics of high-temperature expansion, renormalization group, exact solution for solvable models, ... In this course we will focus on the first two methods.

1. Self-consistent meanfield approach.

When we did high temperature expansion in the Ising model we argued that each spin feels an effective magnetic field $h_{eff} = h(1 + 2Jd/T)$ (see Eq. (166)). Note that this result can be understood from self-consistent arguments: without interactions the average magnetization per spin is h. Therefore to the lowest order the effect of interactions on a given spin with say $j = 0$ is:

$$-J \sum_{\langle j0 \rangle} \sigma_j \sigma_0 \approx -2Jdh\sigma_0. \tag{331}$$

Let us extend this argument below perturbation theory in J/T and assume that we can substitute the true interaction term by the mean-field one:

$$-J \sum_{\langle j0 \rangle} \sigma_j \sigma_0 \to -h_{eff}\sigma_0 \tag{332}$$

so the only result of all other spins on spin 0 is that the latter feels some effective average field from the other neighboring spins. Clearly we are making an approximation because the problem is correlated: what nearest neighbors of site 0 do depends on spin σ_0 itself in a complicated way. However, it is intuitively clear that in high-dimensions (large coordination numbers) this effect should be small. Other spins have too many nearest neighbors to care a lot about what a single spin is doing. However, at low dimensions, especially in 1D this approximation is very crude, and in fact incorrect.

Clearly $h_{eff} = 2Jd\langle s \rangle$ (assuming there is no external field) and

$$\langle \sigma_0 \rangle = \tanh(\beta h_{eff}) \tag{333}$$

To be consistent we must have

$$h_{eff} = 2Jd\tanh(\beta h_{eff}). \tag{334}$$

because all spins should have the same average magnetization. Denoting βh_{eff} as x we see that this equation becomes:

$$x = 2Jd\beta\tanh x \tag{335}$$

It can be solved graphically. Clearly for $2Jd\beta < 1$ (high temperatures) we find that the only solution is $x = 0$. While for $2Jd\beta > 1$ there are two additional solutions corresponding to non-zero magnetization. In the limit $\beta \to 0$ we obviously have the nonzero solutions at $x \approx \pm 2Jd\beta$ corresponding to $\langle\sigma\rangle \approx \pm 1$. It is intuitively clear that non-zero solutions are the correct ones by continuity argument. In the next section we will give a more rigorous justification to this statement.

B. Variational approach

From quantum mechanics we know that the energy of the ground state is larger than the expectation value of the Hamiltonian in any state:

$$E_{gs} \leq \langle\psi|\mathcal{H}|\psi\rangle \tag{336}$$

for any ψ. This statement serves as the basis for the variational principle, where we choose any trial state with some unfixed parameters and minimize the expectation value of the Hamiltonian with respect to this state. Can we generalize this principle to finite temperatures? Obviously we are no longer minimizing energy because the entropy should play the role.

It turns out that the generalization does exist and it can be formulated as follows. Let us choose some arbitrary trial Hamiltonian $\mathcal{H}_0$. Then the corresponding grand potential (free energy in the canonical case) satisfies the following inequality:

$$\Omega \equiv -T\ln Z \leq \Omega_{\rm tr} = \Omega_0 + \langle\mathcal{H} - \mathcal{H}_0\rangle_0, \tag{337}$$

where the expectation value is taken with respect to the equilibrium partition function corresponding to $\mathcal{H}_0$. Let us check that this variational principle reduces to (336) in the zero temperature limit. In this case $\Omega = E_{gs}$, $\Omega_0 = E^0_{gs}$ so we have $E_{gs} \leq \langle\mathcal{H}\rangle_0$, which indeed follows from Eq. (336). The only difference is that the trial state is no assumed to be the ground state of some particular trial Hamiltonian. This Hamiltonian can be always found for any $|\psi\rangle$ by an appropriate choice of potential.

Let us prove this statement, first in the classical case.

$$Z = \int dxdp \exp[-\beta(\mathcal{H}_0(x,p) - V(x,p))] = \frac{1}{Z_0}\langle \exp[-\beta V(x,p)]\rangle_0, \tag{338}$$

where by x, p we schematically denoted all multi-dimensional phase space and $V = \mathcal{H}(x,p) - \mathcal{H}_0(x,p)$. The function $\exp[-\beta v]$ is convex (it has a positive second derivative). For any convex function we have $f(\overline{v}) \leq \overline{f(v)}$ (this is called Jensen's inequality) in particular $f((v_1 + v_2)/2) \leq (f(v_1) + f(v_2))/2$. This most easily can be seen from the picture. The general prove is beyond the purpose of this course, but for a particular exponential function we are interested in the proof is very simple. Let us prove that

$$g(\beta) = \left\langle \exp[-\beta(V - \langle V\rangle_0)]\right\rangle_0 \geq 1 \tag{339}$$

To do this we note that

$$g(\beta) = g(0) + \int\limits_0^\beta \frac{dg}{d\beta'} d\beta' = g(0) - \int\limits_0^\beta \left\langle (V - \langle V\rangle_0)\, \mathrm{e}^{-\beta'(V - \langle V\rangle_0)}\right\rangle_0 d\beta' = g(0) - \int\limits_0^\beta \Big\langle (V - \langle V\rangle_0) \left(\mathrm{e}^{-\beta'(V - \langle V\rangle_0)} - 1\right) \tag{340}$$

The integrand in the last expression is never positive so we see that $g(\beta) \geq g(0) = 1$. The proof in the quantum case is similar if V commutes with $\mathcal{H}_0$. Otherwise it is more complicated but still correct (to prove the statement in general quantum case one can invoke the matrix generalization of Jensen's inequality for convex function: $\mathrm{Tr}(f(A) - f(B)) \leq \mathrm{Tr}\left[(A - B)f'(B)\right]$).

We can use Eq. (337) to build a variational approach at finite temperature: choose a simple trial Hamiltonian with free parameters, from this find a trial thermodynamic potential Ω_{tr}. Minimize this partition function. The state where it is equal to minimal is closest to the true equilibrium state of the system.

Let us apply this method to the Ising model

$$\mathcal{H} = -J \sum_{\langle ij\rangle} \sigma_i \sigma_j. \tag{341}$$

Let us choose a trial Hamiltonian to be noninteracting:

$$\mathcal{H}_0 = -\lambda \sum_j \sigma_j, \tag{342}$$

where λ is the variational parameter. Then

$$\Omega_0 = -TN\ln[2\cosh\beta\lambda], \quad \langle\mathcal{H}_0\rangle_0 = -\lambda N\tanh(\beta\lambda), \quad \langle\mathcal{H}\rangle = -\frac{NJ2d}{2}\langle\sigma\rangle_0^2 = -NJd\tanh^2(\beta\lambda). \tag{343}$$

Therefore

$$\Omega_{\rm tr} = -TN\ln[2\cosh(\beta\lambda)] + \lambda N\tanh(\beta\lambda) - NJd\tanh^2(\beta\lambda) \tag{344}$$

Minimize with respect to λ:

$$-N\tanh(\beta\lambda)+N\tanh(\beta\lambda)+\frac{N\lambda\beta}{\cosh^2(\beta\lambda)}-2NJd\beta\frac{\tanh(\beta\lambda)}{\cosh^2(\beta\lambda)} = 0, \quad \Leftrightarrow \quad \beta\lambda = 2Jd\beta\tanh(\beta\lambda). \tag{345}$$

The last equality exactly reproduces the mean field equation (335) with λ playing the role of $h_{\rm eff}$. But now we can immediately see that for $2Jd\beta > 1$ the broken symmetry solution corresponds to the minimum of free energy. Indeed if we expand $\Omega_{\rm tr}(\lambda)$ around $\lambda = 0$ we will find

$$\Omega_{\rm tr} \approx {\rm const} + N\lambda^2\left[-\frac{\beta}{2} + \beta - \frac{J2d\beta^2}{2}\right] = {\rm const} + \frac{N\lambda^2\beta}{2}[1 - 2Jd\beta] \tag{346}$$

So we find that indeed for $2Jd\beta > 1$ the symmetry broken state has lower (trial) free energy and thus this would be closer to the true equilibrium state). We can make the expansion to the fourth order in λ and get that

$$\frac{\Omega_{\rm tr}\beta}{N} \approx {\rm const} + \frac{\alpha}{2}\lambda^2 + \frac{\gamma}{4}\lambda^4, \tag{347}$$

where $\alpha = 1 - 2\beta Jd$ and $\gamma = 8\beta Jd/3 - 1$. Note that near the transition $2\beta Jd = 1$ the coefficient γ is positive $\gamma \approx (1/3)$ so the effective free energy has a Mexican-hat type form. Because physically λ plays the role of magnetization we can view $\Omega_{\rm tr}(\lambda)$ as the expansion of free energy as a function of magnetization, which is a particular example of phenomenological Landau theory of phase transitions.

Homework, due 12/4/08 *Compare the result for magnetic susceptibility obtained from high temperature expansion and from meanfield approach up to the second order in J/T. For the meanfield approach you need to assume that there is a small magnetic field h and find the effective magnetic field (obtained self-consistently) to the linear order in h but including corrections in J/T.*

Use variational approach to find $\Omega_{\rm tr}$ *at fixed magnetization. Essentially you need to repeat the steps but minimize* $\Omega_{\rm tr}$ *using magnetization* $m = 1/N\sum_{\langle}\sigma_j\rangle$ *as a Lagrange multiplier. Derive Taylor expansion of* $\Omega_{\rm tr}(m)$ *to the second order in* m*, discuss your result.*

C. Interacting many-particle systems.

1. Quick introduction to second quantization

When we deal with many-particle systems it is usually not very convenient to deal with a many-body phase space. There is a much more convenient way to relabel many-particle basis using (complete) single-particle basis and distribution of different particles in this basis. So the eigenstates will be labeled by strings like this

$$|2,0,1,3,0,0,0,1,\dots\rangle \tag{348}$$

for bosons and by strings like

$$|1,0,1,1,0,0,0,1,\dots\rangle \tag{349}$$

for fermions. The numbers here indicate that e.g. the first state contains two particles in the single-particle ground state, 0 particles in the first excited state, etc. If there are no interactions, this basis labels eigenstates of the many-particle system. We actually used this basis when we considered ideal quantum gases. However, if the systems are interacting this basis does not describe energy eigenstates. Any Hamiltonian in this basis can be described by matrix elements between different states. It turns out that the second quantized language allows us to write these matrix elements in a very compact form. Let us introduce the creation and annihilation operators denoted as $a_j^\dagger$ and a_j, which are responsible for increasing or decreasing number of particles in a given single-particle state. In principle the choice of the matrix elements of this operators with other states is arbitrary. However, it it convenient and conventional to stick to the following rules:

- For bosons we require that creation and annihilation operators corresponding to different single particle states commute with each other. This automatically gives us symmetric wave-functions. E.g.

$$\psi(1,3) = |1,0,1,0,\dots\rangle = a_1^\dagger a_3^\dagger|0,0,0,0,\dots\rangle = a_3^\dagger a_1^\dagger|0,0,0,0,\dots\rangle = \psi(3,1). \tag{350}$$

Equivalently we have $a_1^\dagger a_3^\dagger = a_3^\dagger a_1^\dagger$. Similarly for fermions it is convenient to choose creation and annihilation operators corresponding to different states as anti commuting

$$\psi(1,3) = |1,0,1,0,\dots\rangle = c_1^\dagger c_3^\dagger |0,0,0,0,\dots\rangle = -c_3^\dagger c_1^\dagger |0,0,0,0,\dots\rangle = -\psi(3,1), \quad (351)$$

or $c_1^\dagger c_3^\dagger = -c_3^\dagger c_1^\dagger$. It is easy to check that similar (anti)commutation relations between creation and annihilation operators, as well as annihilation and creation operators ensure that all basis functions are always properly symmetrized.

- Now we need to decide what to do with commutation relations for a_j and $a_j^\dagger$ corresponding to the same single particle state. For bosons it is convenient to use the same convention as for harmonic oscillators, i.e.

$$[a_j, a_j^\dagger] = a_j a_j^\dagger - a_j^\dagger a_j = 1. \quad (352)$$

This requirement immediately yields familiar matrix elements for the creation and annihilation operators in the harmonic oscillator:

$$a_j^\dagger |n_j\rangle = \sqrt{n_j + 1}|n_j + 1\rangle. \quad (353)$$

This can be seen e.g. from normalization conditions:

$$\langle 1|1\rangle = \langle 0|a_j a_j^\dagger|0\rangle = \langle 0|1 - a_j^\dagger a_j|0\rangle = 1 \quad (354)$$

and

$$\langle 1|a_j a_j^\dagger|1\rangle = \langle 0|a_j a_j a_j^\dagger a_j^\dagger|0\rangle = \langle 0|a_j(1 + a_j^\dagger a_j)a_j^\dagger|0\rangle = 1 + \langle 0|(1 + a_j^\dagger a_j)(1 + a_j^\dagger a_j)|0\rangle = 2 \quad (355)$$

From this and normalization $\langle 2|2\rangle = 1$ we see that $a_j^\dagger|1\rangle = \sqrt{2}|2\rangle$. And so on. Similarly we can find that $a_j|n_j\rangle = \sqrt{n_j}|n_j - 1\rangle$. This commutation relations immediately yield that $n_j = a_j^\dagger a_j$ is the number operator: $a_j^\dagger a_j|n_j\rangle = a_j^\dagger \sqrt{n_j}|n_j - 1\rangle = n_j|n_j\rangle$. So the Hamiltonian of any noninteracting bosonic system can be written as

$$\mathcal{H} = \sum_j \epsilon_j n_j = \sum_j \epsilon_j a_j^\dagger a_j \quad (356)$$

Note that we never used any details of the Hamiltonian to write down this expression. Harmonic oscillator shows us an example how these operators can be explicitly constructed from the coordinate and momentum operators.

So bosonic operators satisfy

$$[a_i, a_j^\dagger] = \delta_{i,j} \tag{357}$$

- For fermions it is convenient to choose creation and annihilation operators corresponding to the same state satisfying $\{c_j, c_j^\dagger\} \equiv c_j c_j^\dagger + c_j^\dagger c_j = 1$. We must also have $c_j c_j = c_j^\dagger c_j^\dagger = 0$ because by Pauli exclusion principle it is impossible to have more than one particle in the same state. This choice automatically yields $c_j^\dagger|0\rangle = |1\rangle$ and $c_j|1\rangle = |0\rangle$. Indeed then $\langle 1|1\rangle = \langle 0|c_j c_j^\dagger|0\rangle = \langle 0|1 - c_j^\dagger c_j|0\rangle = 1$. The quantity $c_j^\dagger c_j$ also serves as the number operator so for the noninteracting particles the Hamiltonian still takes the form (356) with $a_j \leftrightarrow c_j$. Obtaining explicit form of the fermionic operators is much less trivial than the bosonic. However, there is such a representation of fermionic operators through products of 2×2 matrices. This representation is non-local; it was derived by Jordan and Wigner (you can search for Jordan-Wigner transformation or look in e.g. Subir Sachdev, *Quantum Phase Transitions*).

- The form of the Hamiltonian (356) is consistent with our earlier expressions for the average energy:

$$E = \langle \mathcal{H} \rangle = \sum_j \epsilon_j f(\epsilon_j), \tag{358}$$

  where $f(\epsilon_j)$ is the bosonic or fermionic distribution function. Similarly one can check that we can derive correct expressions for fluctuations of energy, etc.

- It is important to be able to change the basis and know how creation and annihilation operators transform. Let us consider the following single-particle transformation:

$$|j\rangle = \sum_k U_{jk}|k\rangle, \tag{359}$$

  where U is some unitary matrix. On the other hand this is equivalent to

$$a_j^\dagger|0\rangle = \sum_k U_{jk} b_k^\dagger|0\rangle \tag{360}$$

  We use notations b_k to denote annihilation operators in the different basis. So

$$a_j^\dagger = U_{jk} b_k^\dagger, \quad \Leftrightarrow \quad a_j = U_{kj}^\dagger b_k \tag{361}$$

  or in the compact matrix notation $a = bU^\dagger$. Note that this unitary transformation does not change commutation relations:

$$[a_j^\dagger, a_i] = U_{jk} U_{qi}^\dagger [b_k^\dagger, b_q] = U_{jk} U_{k,i}^\dagger = \delta_{ji}. \tag{362}$$

For example if j stands for the (discrete) position and k stands for the momentum we have

$$a_j = \frac{1}{\sqrt{L}} \sum_k a_k \exp[ikj] \tag{363}$$

- If we now write the single-particle Hamiltonian (or any other single particle operator) in arbitrary basis than in the second quantized form it becomes:

  $$\mathcal{H} = \sum h_{ij} a_i^\dagger a_j, \tag{364}$$

  where h_{ij} are the matrix elements of the single particle Hamiltonian between the states i and j. This can be obtained e.g. from the Eq. (356) and the unitary transformation (361). This expression has a very clear interpretation: single particle operators can annihilate a particle at one state and create at some other state, with which the matrix element of the corresponding operator is nonzero. For example if the Hamiltonian has a kinetic energy and some external potential then the off-diagonal part of H corresponds to scattering between different momentum states.

- Our final ingredient is to add interactions to this picture. Typically we deal with two body interactions corresponding to two-particle collisions. Then the corresponding contribution to the Hamiltonian reads:

  $$\frac{1}{2} \sum_{ijkl} c_i^\dagger c_j^\dagger V_{ij,kl} c_k c_l, \tag{365}$$

  where $V_{ij,kl}$ is the matrix element of the interaction between the properly symmetrized states where the two particles are in states $|i\rangle$, $|j\rangle$ and $|k\rangle$, $|l\rangle$ respectively. The factor of 1/2 is usually written for convenience. Let us prove that this is indeed the case for a stationary density density interaction:

  $$V(x_1, x_2, \dots x_N) = \sum_{i<j} v(x_j - x_j) = \frac{1}{2} \sum_{i,j} v(x_j - x_j) = V = \frac{1}{2} \int dx_1 dx_2 v(x_1 - x_2) \rho(x_1) \rho(x_2), \tag{366}$$

  where the sum is taken over all particle pairs and $\rho(x) = \sum_i \delta(x - x_i)$. In the second quantized language the density operator becomes simply $a^\dagger(x)a(x)$. This is obvious if we treat x as the discrete coordinate, then density is just the number operator (up to volume factors) corresponding to the given position. So we have

  $$V = \frac{1}{2} \int \int dx_1 dx_2 v(x_1 - x_2) a^\dagger(x_1) a(x_1) a^\dagger(x_2) a(x_2) = \frac{1}{2} v(0) N + \frac{1}{2} \int \int dx_1 dx_2 a^\dagger(x_1) a^\dagger(x_2) v(x_1 - x_2) a(\tag{367}$$

Note that in this form the expression works both for bosons and for fermions. The first term is trivial shift of energy and we can always eliminate it reabsorbing either into the chemical potential (in grand canonical systems) or to the redefinition of the ground state energy (in canonical systems). We will thus not worry about this term. In the second term in this often convenient to go to the momentum basis. Then

$$V = \frac{1}{2L^2} \sum_{q_1,q_2,q_3,q_4} a^\dagger_{q_1} a^\dagger_{q_2} a(q_3) a(q_4) \int dx_1 dx_2 v(x_1 - x_2) \exp[iq_1 x_1 + iq_2 x_2 - iq_3 x_2 - iq_4 x_1]. \tag{368}$$

By changing variables to $x = x_1 - x_2$ and $X = (x_1 + x_2)/2$ we find that the integral over X gives $q_1 + q_2 = q_3 + q_4$ - total momentum is conserved and the integral over x gives the Fourier transform of $v(x)$ so

$$V = \frac{1}{2} \sum_{k,p,q} a^\dagger_{k+q} a^\dagger_{p-q} v(q) a_p a_k \tag{369}$$

The physical interpretation of this term is very simple. Interaction takes two particles with momenta k and p and scatters them to the two-particles into two other states conserving the total momentum (which is the consequence of the translational invariance of the interaction. Note again that Eq. (369) is valid again both for bosons and for fermions.

This is the end of our short introduction to the second quantization. We never used any details from the Shrödinger first-quantized picture. The second quantized approach is very powerful for the many-particle systems. It also gives very convenient way for seeing particle-wave duality. As we will see later for classical waves (phonons, photons, condensates,...) it is a_j which becomes a classical field: $a_j \sim \sqrt{\rho_j} \exp[i\phi_j]$, where ρ_j is the density and ϕ_j is the phase. One should not be deceived, by the simplicity of the interaction term (369). It is a compact notation of the interaction operator in the exponentially large Hilbert space. The problem of solving the Schrödinger equation is not becoming easier if we just change notations. However, manipulations and building approximate schemes in the second quantized form becomes much more transparent.

2. Interacting Fermi systems.

Both mean-field principle and the variational approach are straightforwardly generalized to interacting quantum systems (in fact quantum or classical does not make a big difference, but quantum is more general). In general we are dealing not with ideal quantum or classical gases but rather with systems with interactions. In the second quantized language we are dealing with the Hamiltonians:

$$\mathcal{H} = K + U, \tag{370}$$

where

$$\mathcal{H} = \sum_k (\varepsilon_k - \mu) c_k^\dagger c_k \tag{371}$$

and

$$U = \frac{1}{2}\int dr dr' \rho(r) u(r - r')\rho(r') = \frac{1}{2}\sum_{k,p,q} c_{k+q}^\dagger c_{p-q}^\dagger u(q) c_p c_k. \tag{372}$$

Understanding the properties of the system with the general hamiltonian $\mathbf{H}$ is the realm of the condensed matter physics. Here I will flash several examples, showing how from variational approach and mean-field theory we can understand many different phases and phase transitions, from theory of Fermi liquids to such phenomena as crystallization and superfluidity (superconductivity). The trick is to determine the possible broken phase and write a simple trial Hamiltonian (or simple mean-field). For us simple means non-interacting. The most general noninteracting Hamiltonian we can write is

$$\mathcal{H}_0 = \sum_{k,k'} \Lambda_{k,k'} c_k^\dagger c_{k'} + \Gamma_{k,k'} c_k^\dagger c_{k'}^\dagger + \Gamma_{k,k'}^\star c_k c_{k'}, \tag{373}$$

where we treat $\Lambda_{k,k'}$, $\Gamma_{k,k'}$, and $\Gamma_{k,k'}^\star$ as variational parameters. This is substantially more complicated than the simple Ising model because instead of one variational parameter we have many. Yet the variational problem is much simpler than the exact one because instead of exponentially large number of coefficients of the density matrix we have only a power law number of terms.

Let us see what are the symmetries which can be broken within this trial Hamiltonian. Let us write operators for the momentum and number:

$$P = \sum_k \hbar k c_k^\dagger c_k, \quad N = \sum_k c_k^\dagger c_k. \tag{374}$$

Both operators commute with the full Hamiltonian. But note that

$$[P, \mathcal{H}_0] = \sum_{k,k'} \hbar(k - k')\Lambda_{k,k'} c_k^\dagger c_{k'} + \hbar(k + k')\Gamma_{k,k'} c_k^\dagger c_{k'}^\dagger + c.c. \tag{375}$$

So we see that the translational symmetry is not broken if we insist that $\Lambda_{k,k'} = \Lambda_k \delta_{k,k'}$, $\Gamma_{k,k'} = \Gamma_k \delta_{k,-k'}$. This is more or less intuitively clear we can either kill and create a particle with the same momentum or create or kill the pair of particles with opposite momenta so the total momentum is conserved. Similarly

$$[N,\mathcal{H}_0]=2\sum_{k,k'}\Gamma_{k,k'}c_k^\dagger c_{k'}^\dagger+c.c. \tag{376}$$

So the total number of particles is conserved (or the corresponding symmetry is non-broken) is equivalent to having $\Gamma\equiv 0$.

3. Fermi liquid theory

Let us assume that there are no broken symmetries. Then the trial Hamiltonian is

$$\mathcal{H}_0=\sum_k(\epsilon_k+\Sigma_k-\mu)c_k^\dagger c_k, \tag{377}$$

where Σ_k is the self energy. For us it is just the variational parameter. Next we need to find a trial free energy. We have several ingredients:

$$\Omega_0=-T\sum_k\ln[1+\exp[-\beta(\epsilon_k+\Sigma_k-\mu)]], \tag{378}$$

$$\langle\mathcal{H}_0\rangle_0=\sum_k(\epsilon_k+\Sigma_k-\mu)f(\epsilon_k+\Sigma_k-\mu), \tag{379}$$

$$\langle K\rangle_0=\sum_k(\epsilon_k-\mu)f(\epsilon_k+\Sigma_k-\mu) \tag{380}$$

The last ingredient is

$$\langle U\rangle_0=\frac{1}{2}\sum_{k,p,q}v_q\langle c_{k+q}^\dagger c_{p-q}^\dagger c_p c_k\rangle_0. \tag{381}$$

The only two possibilities that the expectation value above is nonzero are i) $q=0$ and ii) $q=p-k$ so

$$\langle U\rangle_0=\frac{1}{2}v(0)\sum_{k,p}f(\epsilon_k+\Sigma_k-\mu)f(\epsilon_p+\Sigma_p-\mu)-\frac{1}{2}\sum_{k,p}v(k-p)f(\epsilon_k+\Sigma_k-\mu)f(\epsilon_p+\Sigma_p-\mu). \tag{382}$$

For bosons we will get the same result but with a "+" sign. Next we need to minimize the trial free energy with respect to Σ_k. It is easy to verify that this minimization yields

$$\Sigma_k=v(0)\sum_p f(\epsilon_p+\Sigma_p-\mu)-\sum_p v(p-k)f(\epsilon_p+\Sigma_p-\mu). \tag{383}$$

For bosons there will be gain a "+" sign. This is a system of self-consistent equations which needs to be solved. Still hard but much easier than the original problem. Combining all ingredients we find that

$$\Omega_{\rm tr}=-T\sum_k\ln[1+\exp[-\beta(\epsilon_k+\Sigma_k-\mu)]]+\frac{1}{2}\sum_k\Sigma_k f(\epsilon_k+\Sigma_k-\mu). \tag{384}$$

The trial free energy (thermodynamic potential) depends only on the combination $\epsilon_k + \Sigma_k$. Finding Σ_k requires solving complicated coupled nonlinear equations. Yet it is much easier than solving the full problem. One can anticipate that at small k we have $\Sigma_k = \text{const} + \alpha k^2$: the linear term in k must vanish because of inversion symmetry. Thus at small k we can write that

$$\epsilon_k + \Sigma_k \approx \frac{\hbar^2 k^2}{2m^\star}, \tag{385}$$

where

$$\frac{m}{m^\star} = 1 + \left.\frac{\partial \Sigma_k}{\partial \epsilon_k}\right|_{\epsilon_k = 0}. \tag{386}$$

The quantity $m^\star$ is called the effective mass. For liquid ^4He it is equal to $1.58m$. Similar expansion can be made for interacting nearly degenerate fermions:

$$\epsilon_k + \Sigma_k \approx E_f + \frac{\hbar^2 k_f k}{m^\star}, \tag{387}$$

where

$$\frac{m}{m^\star} = 1 + \left.\frac{\partial \Sigma_k}{\partial \epsilon_k}\right|_{\epsilon_k = E_f}. \tag{388}$$

Note that the same equation (383) can be obtained using meanfield approach.

$$U = \frac{1}{2}\sum_{k,p,q} c^\dagger_{k+q} c^\dagger_{p-q} u(q) c_p c_k \to U_{mf} = \frac{1}{2}\sum_{k,p,q} u(q)[c^\dagger_{k+q} \langle c^\dagger_{p-q} c_p\rangle\, c_k + \langle c^\dagger_{k+q} c_k\rangle\, c^\dagger_{p-q} c_p - c^\dagger_{k+q} c_p \langle c^\dagger_{p-q} c_k\rangle - \langle c^\dagger_{k+q} c_p\rangle\, c \tag{389}$$

Note that if the translational symmetry is not broken the averages are non-zero only if $q = 0$ in the first two terms and $p = k + q$ in the last two. So we have

$$U_{mf} = u(0)\sum_k [c^\dagger_k c_k \sum_p f(p) - c^\dagger_k c_k \sum_p u(k-p) f(p)]. \tag{390}$$

From this we easily recover the equation (383) for the self energy. Note that the first term in Eq. (390) is the classical mean-field potential: electrons feel effective potential from the average density of other electrons surrounding it. This term called after Hartry. The second (named after Fock) is the exchange contribution (negative for electrons and positive for bosons). Note that for short range interactions $v(q) = \text{const}$. The Hartry and Fock terms exactly cancel each other for fermions (electrons do not see each other due to Pauli principle) and these two terms are equal to each other for bosons. So interactions for bosons are enhanced compared to classical particles: Bose bunching, they like to be with each other.

4. Examples of various broken symmetry states

We can now understand how to look into various symmetry broken states in interacting electron (or boson) systems.

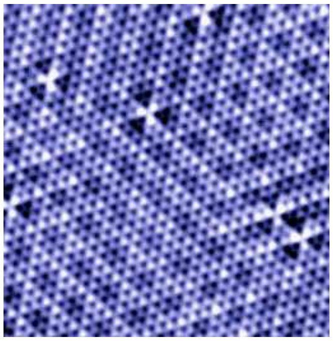
Broken translational symmetry. For example, if we use the trial Hamiltonian of the form

$$\mathcal{H}_0 = \sum_k (\epsilon_k - \mu) c_k^\dagger c_k + \Sigma_{k,k+Q} c_k^\dagger c_{k+Q} + c.c. \tag{391}$$

If the variational state with non-zero Σ is better then with zero Σ then we have a transition to the charge density wave state. Within the mean-field approach this corresponds to nonzero expectation value of

$$c_k^\dagger c_{k+Q} \tag{392}$$

In Fig. 4 we show the example of CDW in NaSe$_2$. In classical gases crystals provide us with the example of systems with spontaneously broken translational symmetry. This phase is characterized by divergent response to the perturbation



CDW in NaSe$_2$,
K. Mc Elroy, U. of Colorado, Boulder

FIG. 4 Example of charge density wave transition NaSe$_2$.

$$\delta H = \lambda \int dx \rho(x) \cos(Qx + \theta) \tag{393}$$

Broken time reversal symmetry: ferromagnetism We can break time reversal symmetry resulting in ferromagnetism. The trial Hamiltonian would be

$$\mathcal{H}_0 = \sum_k (\epsilon_k - \mu + \Sigma_k^+)(c_{k\uparrow}^\dagger c_{k\uparrow} + c_{k\downarrow}^\dagger c_{k\downarrow}) + \Sigma_k^- (c_{k\uparrow}^\dagger c_{k\uparrow} - c_{k\downarrow}^\dagger c_{k\downarrow}) \tag{394}$$

When Σ_k^- becomes nonzero (or equivalently $\langle c_{k\uparrow}^\dagger c_{k\uparrow} - c_{k\downarrow}^\dagger c_{k\downarrow}\rangle$ becomes nonzero in the meanfield language) then we have spontaneous magnetization. The ferromagnetism (as in the case of the Ising model) can be understood as the divergent response to the magnetic field:

$$\delta H = h\sum_k c_{k\uparrow}^\dagger c_{k\uparrow} - c_{k\downarrow}^\dagger c_{k\downarrow}. \tag{395}$$

Broken time reversal + translational symmetries: antiferromagnetism (or more generally spin density waves). If we have a lattice then we can break both symmetries using the trial Hamiltonian

$$\mathcal{H}_0 = \sum_k (\epsilon_k - \mu + \Sigma_k^+)(c_{k\uparrow}^\dagger c_{k\uparrow} + c_{k\downarrow}^\dagger c_{k\downarrow}) \sum_j (-1)^j \Sigma_j (c_{j\uparrow}^\dagger c_{j\uparrow} - c_{j\downarrow}^\dagger c_{j\downarrow}), \tag{396}$$

where j enumerates different positions in the lattice. When Σ_j is non-zero we have divergent response to a staggered magnetic field (which alternates signs between neighboring lattice sites).

D. Broken number symmetry. Superfluids and superocnductors

Before we were dealing with the situations where N was always conserved so that in all thermodynamic ensembles number was a diagonal operator. Also in all previous variational examples the trial Hamiltonian was commuting with N and in the meanfield Hamiltonians only number conserving quantities acquired expectation values.

1. Weakly interacting Bose gases.

But are there situations where the number symmetry is broken? We had a hint from studying Bose-Einstein condensation in ideal Bose gases. Recall that active part of the particles, which participated in specific heat, entropy, pressure etc. was insensitive to the total number of particles which went to the condensate part. So for the active part of the Bose gas below the condensation temperature the number symmetry is effectively broken. Particles can appear from and disappear to the condensate.

Note that if the number of particles in the condensate is large then the corresponding creation and annihilation operators can be treated as classical (complex) numbers. For example

$$a|N\rangle \approx \sqrt{N}|N-1\rangle, \quad a^2|N\rangle \approx \sqrt{N(N-1)}|N-2\rangle \approx \sqrt{N^2}|N-2\rangle \sim N^2|N-2\rangle \tag{397}$$

The last equality implies that the state with $N-2$ particles is very similar to the original state of N particles. One can also introduce phase, which is conjugate to number

$$|\phi\rangle = \sum_{N\geq 0} \exp[i\phi N]|N\rangle \tag{398}$$

Clearly phase and number are like coordinate and momentum: when one is defined the other is not. There is one subtlety that $N \geq 0$ so the sum above is not the Fourier transform, but this subtlety is not important when average N is large. When the modes are highly occupied the best states to work with are the coherent states, where both number and phase are defined. These states are analogues of the minimum uncertainty (gaussian) states for coordinate-momentum. Formally these states are defined as eigenstates of annihilation operators:

$$a|\lambda\rangle_c = \lambda|\lambda\rangle_c \tag{399}$$

One can check by explicit calculation that the state $|\lambda\rangle_c$ can be written as the series:

$$|\lambda\rangle_c = |0\rangle + \frac{\lambda}{1}a^\dagger|0\rangle + \frac{\lambda^2}{2!}(a^\dagger)^2|0\rangle + \ldots = \exp[\lambda a^\dagger]|0\rangle. \tag{400}$$

Properly normalized coherent states are $|\lambda\rangle_c = \exp[-|\lambda|^2/2]\exp[\lambda a^\dagger]|0\rangle$.

For the ideal Bose below T_c we concluded that all condensed particles should go to the zero momentum mode. But because there is so huge degeneracy between states with different N we can equally say that condensed particles are in the coherent state: $|\sqrt{N_0}\exp[i\theta_0]\rangle_c$, where θ is some arbitrary phase. In this state a_0 can be treated as a classical variable:

$$a_0|\sqrt{N_0}\exp[i\theta_0]\rangle_c = \sqrt{N_0}\exp[i\theta_0]|\sqrt{N_0}\exp[i\theta_0]\rangle_c. \tag{401}$$

In the meanfield language we can say that a_0 spontaneously acquired expectation value. This corresponds to the broken global gauge symmetry. If we define the operator

$$S = \exp\left[i\chi a_0^\dagger a_0\right] = \exp\left[i\chi N\right], \tag{402}$$

where N is the total number operator, then we have

$$S a_0 S^\dagger = \exp[-i\chi]a_0. \tag{403}$$

Normally the equilibrium state should be invariant under the global gauge transformations since S commutes with the Hamiltonian. However, if the symmetry is broken then this is not longer the case and thus the equilibrium state can loose this invariance. This what happens if the system picks the coherent state with a fixed phase. We can still define a gauge-symmetry breaking perturbation

$$\delta\mathcal{H} = -\lambda \int dx (\mathrm{e}^{i\theta} a^\dagger(\mathbf{r}) + \mathrm{e}^{-i\theta} a(\mathbf{r})) \tag{404}$$

and see when the response diverges, i.e. when this symmetry is spontaneously broken. Clearly for the ideal Bose gas below T_c these gauge symmetry is broken. For example at zero temperature the expectation value of the perturbation δH in the coherent state with phase θ_0 scales as $-2\lambda\sqrt{N}\cos(\theta - \theta_0)$, while the same expectation value of this perturbation in the number state is simply zero. In the limit $N \to \infty$ the coherent (broken symmetry) state with $\theta = \theta_0$ obviously wins at arbitrarily small λ.

Let us now consider an interacting system and proceed with the variational calculation. We assume that the zero momentum mode is in the coherent state $|\phi\rangle_c$ and the rest are described by the Hamiltonian

$$\mathcal{H}_0 = \sum_{k\neq 0} \xi_k a_k^\dagger a_k + \lambda_k a_k^\dagger a_{-k}^\dagger + \lambda_k^\star a_k a_{-k} \tag{405}$$

The motivation for this (Bogoliubov) Hamiltonian is the following: (i) it is quadratic, (ii) it preserves translational invariance but it breaks global gauge symmetry, (iii) it incorporates our prior result that in the non-interacting limit bosons can condense into the lowest energy mode, which is macroscopically occupied. All parameters here are variational. Physically one can interpret the $a_k^\dagger a_{-k}^\dagger$ terms as creating two particles from the condensate. They should be thus multiplied by a_0^2, but since the condensate is the coherent state these are just multiplication by ϕ^2.

To simplify derivation even more we can first keep only the $k = 0$ term and assume that all the modes with $k \neq 0$ are empty. Since there is no entropy associated with one mode we have $\Omega_0 = \langle H_0\rangle_0$ and so

$$\Omega_{\mathrm{tr}} = \langle H\rangle_0 = (\epsilon_0 - \mu)|\phi|^2 + \frac{u(0)}{2}|\phi|^4 \tag{406}$$

By minimizing with respect to the variational parameter ϕ we find

$$\mu = \epsilon_0 + u(0)|\phi|^2. \tag{407}$$

Recall that $|\phi|^2$ is actually the number of particles in the condensate.

If we include quadratic terms the we have to deal with the Bogoliubov Hamiltonian. It can be diagonalized by the Bogoliubov transformation

$$a_k^\dagger = \cosh\theta_k \alpha_k^\dagger - \sinh\theta_k \alpha_{-k}, a_k = \cosh\theta_k \alpha_k - \sinh\theta_k \alpha_{-k}^\dagger \tag{408}$$

and the inverse transformation

$$\alpha_k^\dagger = \cosh\theta_k a_k^\dagger + \sinh\theta_k a_{-k}, \alpha_k = \cosh\theta_k a_k + \sinh\theta_k a_{-k}^\dagger \tag{409}$$

Note that both $\alpha_k^\dagger$ and $a_k^\dagger$ increase momentum by k. One can check that α_k and $\alpha_k^\dagger$ still satisfy bosonic commutation relations.

Homework,optional. *Prove that* $[\alpha_k^\dagger, \alpha_q^\dagger] = 0$ *and* $[\alpha_k, \alpha_q^\dagger] = \delta_{k,q}$.

Now we will try to choose θ_k so that anomalous terms cancel so that

$$\mathcal{H}_0 = \sum_k E_k \alpha_k^\dagger \alpha_k. \tag{410}$$

This is indeed possible if we choose

$$\tanh(2\theta_k) = -2\frac{\lambda_k}{\xi_k} \tag{411}$$

If we now express $\Omega_{\rm tr}$ through θ_k and E_k (instead of λ_k and ξ_k and minimize it with respect to these parameters we find

$$\tanh 2\theta_k = \frac{-u|\phi|^2}{\epsilon_k - \mu + 2u|\phi|^2} = \frac{-u|\phi|^2}{\epsilon_k + u|\phi|^2} \tag{412}$$

and

$$E_k = \sqrt{\epsilon_k^2 + 2u|\phi|^2\epsilon_k} \tag{413}$$

At small k we get linear dispersion

$$E_k \approx \sqrt{\frac{\hbar^2 u_0 \rho_0}{m}}k, \tag{414}$$

where $u_0 = u(k=0)L^3$ and $\rho_0 = |\phi|^2/L^3$. The latter is the condensate density.

This mode is in fact collective Goldstone mode corresponding to the slow variation of phase. If we use $a(r) \approx \sqrt{\rho(r)\exp[i\theta(r)]}$ and substitute this to the Hamiltonian then we will find

$$\mathcal{H} \approx \int d\mathbf{r}\, \frac{\hbar^2}{8m\rho_0}|\nabla\rho|^2 + \frac{\hbar^2}{2m}|\rho_0||\nabla\theta|^2 + \frac{U}{2}\rho^2. \tag{415}$$

One can check that $a^\dagger(r)$ and $a(r)$ being bosonic fields implies that $\rho(r)$ and $\theta(r)$ are conjugate variables like coordinate and momentum. There are some subtleties involved, which are not important for our purposes. Then the Hamiltonian above is equivalent to that of a coupled Harmonic oscillators with the dispersion (413).

Superfluids have many fascinating properties. They all basically originate from (i) non-dissipativity of the flow and (ii) flux (macroscopic phase) quantization: if we make a circle then the phase θ should change by a multiple integer of 2π. Let us give a simple Landau argument why there is no dissipation in superfluids. Assume that the fluid is flowing around an obstacle, or conversely an obstacle is moving in a still superfluid. In order to create excitation we need to satisfy energy and momentum conservation

$$\frac{Mv^2}{2} = \frac{Mv'^2}{2} + E_{\rm ex}(k) M\mathbf{v} = M\mathbf{v}' + \hbar\mathbf{k} \tag{416}$$

Solving these two and assuming that M is large we find

$$\mathbf{v}\hbar\mathbf{k} = E_{\rm ex} \tag{417}$$

so that

$$v \geq \frac{E_{\rm ex}(k)}{\hbar k} \tag{418}$$

If the ratio above has a minimum then the flow with velocities smaller than this minimum can not lead to dissipation. In conventional liquids this minimum is always zero because there are essentially zero energy transverse modes at arbitrary high momentum. In superfluids the minimum is nonzero, it is called the Landau critical velocity (above this velocity superfluidity becomes unstable). In the example we considered of a weakly interacting Bose gas the Landau critical velocity is the same as the sound velocity. In general this is not true.

2. Weakly attractive Fermi gas. BCS theory of superconductivity.

From what we considered above it may seem that the superfluidity is a feature of Bosonic particles. Indeed at zero temperature non-interacting Bose particles condense while the non-interacting Fermions form highly energetic Fermi see. However, it turns out that at low temperatures many metals become superconducting (superfluid) and the superfluidity of fermions is very closely connected to superfluidity of bosons. It took many years before the mechanism of superconductivity was understood by Bardeen, Cooper, and Schrieffer.

Basically the mechanism behind superconductivity is still bosonic: fermions join into Cooper pairs, i.e. weakly bound states, which in turn can superconduct. This is of course a very loose explanation since the size of Cooper pairs is typically much bigger then interparticle distance. So they can not be literally treated as bosonic particles. However, we won't care. We will just apply our standard mean field - variational argument and see what happens. Also we ignore that Fermions are charged. For electromagnetic response (Meisnet effect) this fact is crucial. However for establishing the mechanism of the frmionic superfluidity and for the BCS theory the fact the fermions are charged is not important.

For simplicity we will assume that the interaction between fermions is local and the Hamiltonian can be written as

$$\mathcal{H}=\sum_{k,\sigma}(\epsilon_k-\mu)c^\dagger_{k,\sigma}c_{k,\sigma}-\frac{U}{2}\sum_r c^\dagger_\uparrow(r)c^\dagger_\downarrow(r)c_\downarrow(r)c_\uparrow(r), \tag{419}$$

where $\sigma=\uparrow,\downarrow$. Note that for the local interaction potential only electrons with opposite spins can interact because of Pauli principle. The "$-$" sign in the Hamiltonian above explicitly tells that the interactions are attractive (usually such U originates from electron-phonon interactions). Similarly to the bosonic case we will consider the following trial Hamiltonian

$$\mathcal{H}_0=\sum_k \xi_k c^\dagger_{k,\sigma}c_{k,\sigma}+\sum_k(\lambda_k c^\dagger_{k,\uparrow}c^\dagger_{-k,\downarrow}+h.c.) \tag{420}$$

This Hamiltonian preserves translational invariance and violates global gauge symmetry (or global number conservation). It can be shown that in the meanfield language this variational Hamiltonian is equivalent to

$$\mathcal{H}_{mf}=\sum_k(\epsilon_k-\mu)c^\dagger_{k,\sigma}c_{k,\sigma}+\Delta c^\dagger_{k,\uparrow}c^\dagger_{-k,\downarrow}+h.c., \tag{421}$$

where

$$\Delta=U/2\,\langle c_\uparrow(r)c_\downarrow(r)\rangle. \tag{422}$$

One can find the spectrum of this Hamiltonian using the Bogoliubov transformation similar to the bosonic case introducing new fermionic operators:

$$\gamma^\dagger_{k,\uparrow}=\cos\theta_k c^\dagger_{k,\uparrow}+\sin\theta_k c_{-k,\downarrow},\gamma_{k,\downarrow}=\cos\theta_k c_{k,\downarrow}+\sin\theta_k c^\dagger_{-k,\uparrow} \tag{423}$$

and requiring that the Hamiltonian does not contain anomalous $\gamma^\dagger_{k,\uparrow}\gamma^\dagger_{-k,\downarrow}$ terms. There is however a more elegant (but equivalent) way to find the spectrum using the Nambu notations:

$$\psi_{k,\uparrow}=c_{k,\uparrow},\quad \psi_{k,\downarrow}=c^\dagger_{-k,\downarrow}. \tag{424}$$

One can check that ψ is a proper fermionic operator satisfying correct commutation relations. In terms of ψ the mean field Hamiltonian is written as

$$\mathcal{H}_{mf} = \sum_k \mathcal{H}_k, \tag{425}$$

where

$$\mathcal{H}_k = \psi^\dagger_{k,\alpha} \left[(\epsilon_k - \mu)\sigma^z_{\alpha,\alpha'} + \Delta \sigma^x_{\alpha,\alpha'} \right] \psi_{k,\alpha'}. \tag{426}$$

The eigen energies are clearly the eigenvalues of the 2 × 2 matrix:

$$E_k = \sqrt{(\epsilon_k - \mu)^2 + \Delta^2}. \tag{427}$$

And finally the self-consistency condition (422) gives

$$\Delta = \frac{U}{V} \sum_k \langle \psi^\dagger_{k,\uparrow} \psi_{k,\downarrow} \rangle = \frac{U}{V} \sum_k \frac{\Delta \tanh(\beta(\epsilon_k - \mu)/2}{(\epsilon_k - \mu)}. \tag{428}$$

At zero temperature this reduces to

$$1 = U\rho(E_F) \int_0^{\hbar\omega_D} \frac{d\epsilon}{\sqrt{\epsilon^2 + \Delta^2}}, \tag{429}$$

where $\hbar\omega_D$ is some high energy cutoff, which is usually the highest energy of the phonons participating in the superconductivity. This gives

$$\Delta \approx 2\hbar\omega_D \exp\left[-\frac{1}{U\rho(E_F)} \right] \ll \hbar\omega_D. \tag{430}$$

Equation (428) can also be used to find the highest temperature above which there is no solution with $\Delta \neq 0$. Thus temperature is given by

$$T_c \approx 1.14 \hbar\omega_D \exp\left[-\frac{1}{U\rho(E_F)} \right]. \tag{431}$$

Let us make a couple of quick comments. By dimensional analysis (also analyzing the behavior of the correlation function $c^\dagger_\uparrow(r) c^\dagger_\downarrow(r'))$ one can estimate the size of the Cooper pair as

$$\xi \sim \frac{\hbar v_F}{\Delta} \gg n^{1/3} \tag{432}$$

The parameter, which justifies the mean field approximation, is the number of particles within the coherent volume: $n\xi^3$ in typical metal superconductors this is a huge number so the meanfield works very well.