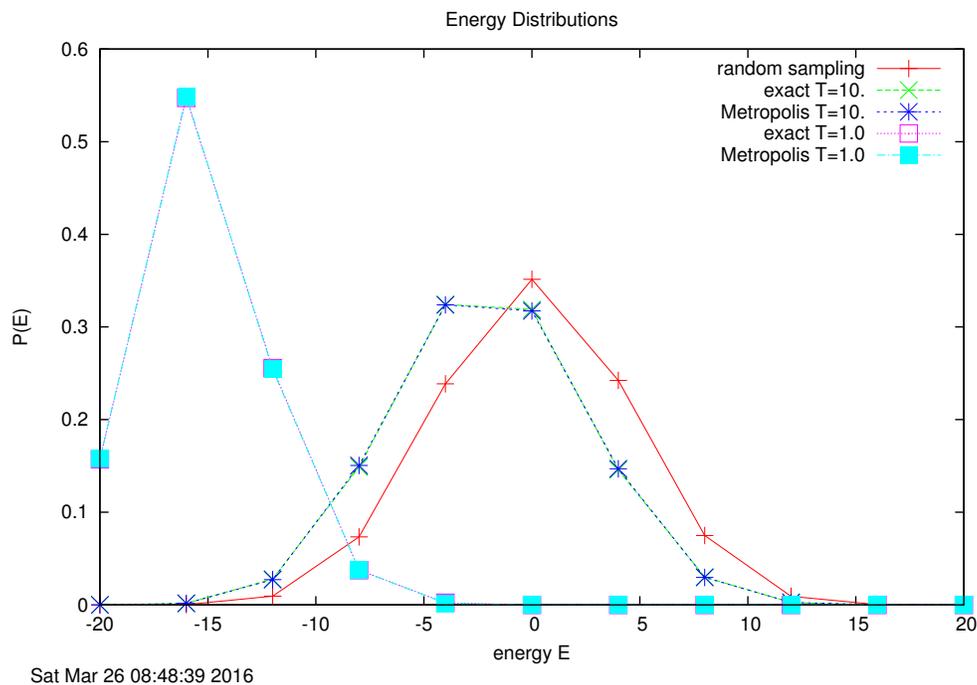


## 14. 6810 Session 14

### a. Follow-ups to Session 13

- **Comment on the codes from Session 13.** The codes used in Session 13 use various “tricks” or shortcuts to make them more compact. In general, I recommend *not* coding this way (so these codes should be rewritten!). For example, the choice of energy units is exploited so that energies are integers and therefore can be used as array indices. But if we added an external field, this would no longer work, so a different approach should be implemented. Another example is the `next_configuration` function, which uses recursion (i.e., the function calls itself) to count upward in binary. This is clever, but really hard to decipher!
- **Recap of energy distributions figure.**

Here is a typical figure you might have obtained from the `sampling_test` program:



Some observations:

- The “random sampling” selects configurations (recall that a configuration is one particular specification of spins in the chain) at random, so each one is equally likely to be chosen. Then the energy is calculated and histogrammed. Since there are only two out of about one million configurations that have the lowest energy of  $-20$ , the curve will be almost zero there. The peak in the middle means that there are more ways to have zero energy than any other energy. The curve is symmetric because there are equal numbers of configurations with a given positive energy as with the corresponding negative energy.

- The exact  $P(E)$  at temperature  $T$  combines two factors: the Boltzmann factor  $e^{-E/kT}/Z$  and the number of states with energy  $E$ . For  $T = 1$  (we use units where  $k = 1$ ), the first factor is largest for  $E = -20$  but the second is smallest there. The product is a moderate probability. For  $E = -16$ , the first factor is smaller but the second is sufficiently larger that the net result is the maximum probability. Even though the second factor is largest at  $E = 0$ , the Boltzmann factor is very small (remember that  $Z$  is large), so there is only a very small probability  $P(0)$ .
  - For larger temperatures, the Boltzmann factor is more uniform, and so the factor of the number of states with energy  $E$  dominates. Thus, the random sampling distribution corresponds to very high temperature.
  - The Metropolis distribution uses only a representative subset of the full set of configurations to estimate  $P(E)$ . But it does a very good job of reproducing the exact distribution at a given temperature.
- **Thermal averages.** Recall the basic idea for evaluating thermal averages. We generate a set of “configurations,” which we’ll label  $\mathbf{x}_0, \mathbf{x}_1, \dots$  or collectively as  $\{\mathbf{x}_i\}$ , that are distributed according to an appropriate probability distribution function (PDF). One of the most physical PDF’s is the Boltzmann distribution. In this case, the thermal average of  $A(\mathbf{x})$  is approximated by

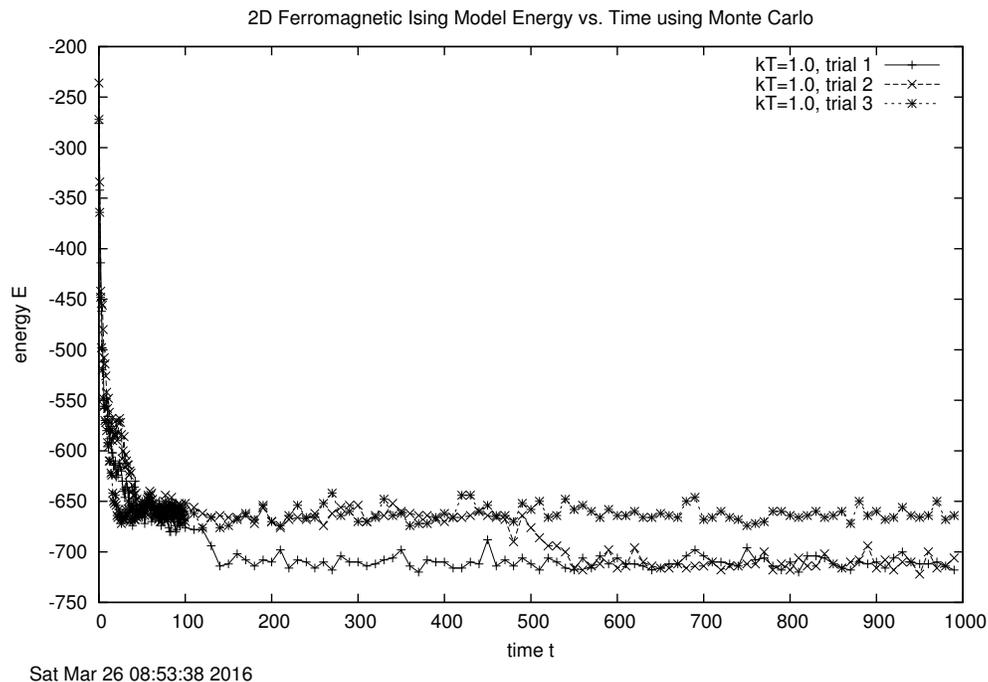
$$\langle A \rangle_T \equiv \frac{\int d\mathbf{x} A(\mathbf{x}) e^{-\mathcal{H}(\mathbf{x})/kT}}{\int d\mathbf{x} e^{-\mathcal{H}(\mathbf{x})/kT}} \approx \sum_i A(\mathbf{x}_i). \quad (14.1)$$

Thus, we find the average simply by summing the quantity  $A$  (e.g., the magnetization) over our representative set of  $\mathbf{x}_i$ ’s (e.g., spin configurations). At high temperatures, there will be a large spread of configurations contributing while at low temperature, states close to the ground state will increasingly dominate (i.e., appear more times in the set of  $\mathbf{x}_i$ ’s).

- **Ensuring representative configurations.** When carrying out the sum over configurations, we want to use ones that are actually representative. Three things we need to do:
  1. Throw out configurations from the beginning until the system has “equilibrated.” How might we tell when this has happened?
  2. Don’t use each configuration resulting from a MCS (“Monte Carlo Step”) but skip every  $n$  configurations to avoid correlations. Basically,  $n$  should be enough MCS’s so that the next configuration used has lost its memory of the previous one used. You can experimentally determine an appropriate “ $n$ ” by looking at an autocorrelation function. We’ll see how this works in Session 14.
  3. Adjust the “step size” between successive  $\mathbf{x}$ ’s so that the Metropolis success rate is reasonable. A rate around 50% is the canonical trade-off between computational efficiency and exploring enough of the configuration space.
- **Efficiencies.** When looking at `ising_model.cpp` compared to `ising_opt.cpp`, please note what has been done to speed up the code. All of the upgrades described in the Session 13 guide are helpful. Why does having a look-up table speed things up? Answer: if there are only a fixed number of energies, we don’t need to calculate the Boltzmann ratio  $e^{-\Delta E/kT}$  over

and over again. Exponentials are expensive (i.e., they take a relatively long time to calculate) while looking up values in an array is cheap. The general idea of a look-up table is one that is frequently useful in speeding up codes.

- **Phase transition.** The optimized Ising model code with a ferromagnetic interaction has a phase transition in two dimensions. There is a “Curie temperature” below which the magnetization is nonzero. Can you see evidence of this phase transition from running `ising_opt`? In the “Phase Transition” section of Session 13, you are asked to calculate the absolute value of the magnetization, which is most usefully defined for this problem as the average spin *per site* (i.e.,  $M = \frac{1}{N} \sum_{i=0}^{N-1} S_i$  with  $S_i = \pm 1$  and  $N$  is the number of lattice sites). To study the phase transition, you should look at the time *average* of this quantity, which means you average the result over all the Monte Carlo steps (after equilibration). If you are well below the phase transition, the spins will be either all up or all down, so you should find  $M = \pm 1$ . The reason to take the absolute value is that, for small systems (like the ones with  $L \leq 20$ ), in the course of a trial you could go from  $M = +1$  to  $M = -1$ . In this case, the average would be less than 1, which does not correctly reflect the physics. This problem is fixed by taking the absolute value. As a function of temperature, you should see the transition from disorder (small average magnetization) to order ( $\langle |M| \rangle = 1$ ) as the temperature is lowered.
- **More Explanation of Metropolis, etc.** The excerpt from the Binder/Heermann book (and the rest of the book) has good suggestions for the practical implementation of Monte Carlo approaches. You are strongly encouraged to look at the notes by Morten Hjorth-Jensen [2], which are linked on the 6810 webpage under “Supplementary Readings.”
  - Chapter 9 is “Random walks and the Metropolis algorithm.” Hjorth-Jensen explains how a Markov process (which is the basis of the Metropolis algorithm) is actually a discretized version of the diffusion equation. There is a discussion of entropy and equilibrium and ergodicity in Sect. 9.4 and a clear and concise discussion of the Metropolis algorithm in Sect. 9.5.
  - Chapter 10 is “Monte Carlo methods in statistical physics,” which has a review of the important thermodynamics as well as an explicit and detailed discussion of the Ising model. (See Fig. 10.2 for a graph of the absolute value of the average magnetization per spin vs. temperature for different size lattices.)
  - Chapter 11 is “Quantum Monte Carlo methods,” which includes a thorough discussion of variational Monte Carlo.
- **Equilibration and Cooling.** Consider the 2D Ising model with ferromagnetic  $J$ . On the next page is an example of what you might find from running the `ising_model.cpp` code with  $kT = 1$  three times. We can see a rapid decrease in energy at small times  $t$ , which reaches a plateau by  $t = 50$  about which successive configurations fluctuate. This is the equilibration or thermalization time (so these are the configurations we should throw out). But we see later that in one trial the energy drops to a lower plateau starting around  $t = 100$  and another drops to the same lower plateau starting around  $t = 500$ . We might imagine the third trial will eventually drop down to the lower energy, but we can’t predict when. This problem of equilibrating to a higher energy (and there may be many other possibilities



if we run for a long time) is *not* fixed by skipping a larger number of configurations at the beginning. (What would you pick? Up to  $t = 100, 500, 2000$ ?) Rather, we give the system a chance to settle into the lowest energy by *cooling*. That is, we start the simulation at a higher temperature with a random configuration and equilibrate at that temperature. Then we use the final configuration as the starting configuration (rather than starting again with a random configuration) for a somewhat lower temperature until it equilibrates. And so on, until we get to the desired temperature. How much to lower the temperature at each stage depends on the problem. (We'll experiment with this rate in a slightly different context in Session 15 with simulated annealing.)

## b. The Autocorrelation Function (based on Pang, *Computational Physics*)

When we apply the Metropolis algorithm to generate a set of configurations that we wish to use to evaluate a physical quantity  $A$  (e.g., the magnetization of a spin system), we do not use all of them. We skip some at the beginning to eliminate the effects of the initial configuration (“thermalization” or “equilibration”) but we also don’t want to use consecutive configurations after that because they are likely to be highly correlated. So we skip some configurations before keeping another one. But how many should we skip? To find out, we analyze the *autocorrelation function* of  $A$ .

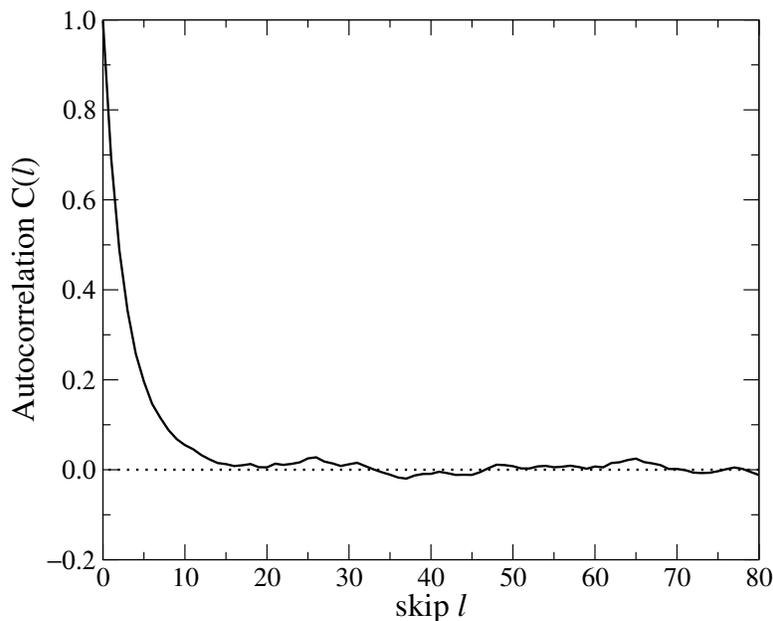
In Session 14 you will calculate the autocorrelation function  $C(l)$ , where

$$C(l) = \frac{\langle A_{n+l}A_n \rangle - \langle A_n \rangle^2}{\langle A_n^2 \rangle - \langle A_n \rangle^2} \quad (14.2)$$

with

$$\langle A_n \rangle = \frac{1}{N_{\text{spins}}} \sum_{n=1}^{N_{\text{spins}}} A_n \quad \text{and} \quad \langle A_{n+l} A_n \rangle = \frac{1}{N_{\text{spins}}} \sum_{n=1}^{N_{\text{spins}}} A_{n+l} A_n, \quad (14.3)$$

where  $A_n$  is the value of  $A$  in the  $n^{\text{th}}$  configuration. (Note that  $\langle A_n^2 \rangle$  is just  $\langle A_{n+l} A_n \rangle$  with  $l = 0$ .) Because  $n$  labels the Monte Carlo step, this correlation is effectively in time (rather than space, for example). Thus, we can use  $C(l)$  to see whether the value at  $n + l$  is related to the value at  $n$ , averaged over  $n = 1$  to  $N_{\text{spins}}$  (strictly speaking, to  $N_{\text{spins}} - l$ ). It is clear that  $C(0) = 1$ , but what do we expect as  $n$  increases? If quantities  $A$  and  $B$  are *uncorrelated*, then the average of their product is the product of their averages:  $\langle AB \rangle = \langle A \rangle \langle B \rangle$ . So if the configurations become uncorrelated, then  $\langle A_{n+l} A_n \rangle \approx \langle A_{n+l} \rangle \langle A_n \rangle \approx \langle A_n \rangle^2$  (since it shouldn't matter whether we average starting at 0 or  $l$ ) and  $C(l) \rightarrow 0$ . Note that it will not go exactly to zero, but fluctuate with each Monte Carlo step about zero (see the figure below). The correlation time (how many configurations to skip when averaging) is roughly how long it takes to go from 1 to the fluctuation region.



In the figure, the correlation time is about 15.

### c. Exchange Integral $J$

[This summary is based on the discussion in Shang-Keng Ma's book on *Statistical Mechanics*.]

Let's look at the physical origin of the exchange integral  $J > 0$  for a ferromagnet. The short-distance exchange interaction is the strongest among electron spins, and comes from the interplay of electrostatic repulsion and the Pauli exclusion principle. Consider the interaction between electrons at two sites. If it were just a magnetic dipole interaction, then  $\uparrow\uparrow$  would not be energetically favored and we would not have ferromagnets! To find the true origin of the interaction, we must turn to

quantum mechanics.

Consider the wave function of electrons 1 and 2 as a product of a spatial part  $\varphi$  and a spin part  $\chi$  (this factorization is possible if there is negligible spin-orbit coupling, which we assume):

$$\psi(1, 2) = \varphi(\mathbf{r}_1, \mathbf{r}_2) \chi(s_1, s_2) , \quad (14.4)$$

where the  $\mathbf{r}_i$ 's are the positions of the electrons and the  $s_i$ 's are the spins. Since electrons are identical *fermions*, the overall wavefunction must be *antisymmetric* under particle exchange (i.e.,  $\psi(2, 1) = -\psi(1, 2)$ ). That leaves two possibilities:  $\varphi$  is symmetric while  $\chi$  is antisymmetric, or *vice versa*. For example,

$\chi$	implication for $\varphi$
$\uparrow\uparrow$	$\varphi(\mathbf{r}_1, \mathbf{r}_2) = -\varphi(\mathbf{r}_2, \mathbf{r}_1)$
$\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow)$	$\varphi(\mathbf{r}_1, \mathbf{r}_2) = +\varphi(\mathbf{r}_2, \mathbf{r}_1)$

Thus the symmetry of the spins dictates the symmetry of the spatial part; this is the key!

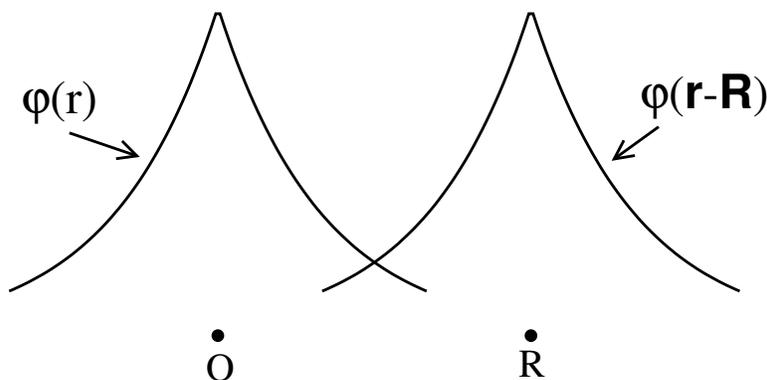
Consider first the non-interacting case and imagine the spatial wavefunctions are like hydrogen ground state wavefunctions centered at the origin  $O$  and point  $R$ . So we make the identification

$$\varphi_1(\mathbf{r}) = \varphi(r) , \quad \varphi_2(\mathbf{r}) = \varphi(\mathbf{r} - \mathbf{R}) , \quad (14.5)$$

as pictured in the figure below. Then the noninteracting wave function is either the symmetric ( $\varphi_+$ ) or antisymmetric ( $\varphi_-$ ) combination:

$$\varphi_{\pm}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) \pm \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)] \quad (14.6)$$

Notice that  $\varphi_-$  vanishes when  $\mathbf{r}_1 = \mathbf{r}_2$ .



Now consider the Coulomb repulsion between the electrons:

$$U(\mathbf{r}_1 - \mathbf{r}_2) = \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} . \quad (14.7)$$

If we calculate the change in energy  $\delta E$  in first-order perturbation theory, then

$$\delta E = \langle \varphi(1, 2) | U | \varphi(1, 2) \rangle . \quad (14.8)$$

The spin part is diagonal and thus contributes only in the sense of telling us whether  $\varphi_+$  or  $\varphi_-$  is the spatial wave function. We can label the two possibilities  $\delta E_{\pm}$ , with

$$\delta E_{\pm} \equiv \langle \varphi_{\pm} | U | \varphi_{\pm} \rangle = \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 |\varphi_1(\mathbf{r}_1)|^2 |\varphi_2(\mathbf{r}_2)|^2 \frac{2e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \pm \frac{J}{2}, \quad (14.9)$$

which defines the *exchange integral*  $J$  as

$$J \equiv \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) \frac{2e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} > 0. \quad (14.10)$$

So what do we learn?

- $J$  is positive definite (for our example), so  $J > 0$  means *lower* energy for  $E_-$ .
- But  $E_-$  implies  $\varphi_-$ , which means an anti-symmetric spatial wave function and therefore symmetric (e.g., parallel) spins.
- Thus the parallel alignment is preferred to lower the energy! In short, if the spins are parallel, the electrons are less likely to be near each other, and therefore will have a lower average potential energy from electrostatic repulsion.
- We can evaluate Eq. (14.10) using hydrogenic wave functions to get a quantitative estimate for  $J$ .

#### d. Variational Monte Carlo

[Note: Please read Chapter 11 of Ref. [2], which is linked on the Computational Physics webpage. It has many details on variational Monte Carlo.]

Here's the basic idea of the variational method for estimating quantum mechanics energies. If we have a guess  $\psi_{\alpha}(\mathbf{R})$  for a ground-state wavefunction that depends on a set of parameters  $\alpha = (\alpha_1, \dots, \alpha_s)$  and the positions of  $N$  particles  $\mathbf{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ , then

$$\langle E_{\alpha} \rangle = \frac{\langle \psi_{\alpha} | H | \psi_{\alpha} \rangle}{\langle \psi_{\alpha} | \psi_{\alpha} \rangle} = \frac{\int d\mathbf{R} \psi_{\alpha}^*(\mathbf{R}) H(\mathbf{R}) \psi_{\alpha}(\mathbf{R})}{\int d\mathbf{R} \psi_{\alpha}^*(\mathbf{R}) \psi_{\alpha}(\mathbf{R})} \quad (14.11)$$

is an estimate of the energy, which is always an *upper bound* to the exact ground-state energy because it is variational. (Note that for generality we don't assume that  $\psi_{\alpha}(\mathbf{R})$  is normalized, which is why in (14.11) the denominator is included.) Thus our best estimate is obtained by *minimizing*  $\langle E_{\alpha} \rangle$  with respect to the parameters  $\alpha$ .

Start with the simplest case, that of a single particle (i.e.,  $N = 1$ ). We might take as our trial wavefunction

$$\psi_{\alpha}(\mathbf{r}) = C(1 + br^2)e^{-ar}, \quad (14.12)$$

in which case  $\alpha = \{a, b\}$ . Note that the normalization constant  $C$  is irrelevant (and is *not* an element of  $\alpha$ ), since it cancels between numerator and denominator in evaluating  $\langle E_{\alpha} \rangle$ . We calculate

the three-dimensional integrals in Eq. (14.11) and minimize the result with respect to the two parameters  $a$  and  $b$ .

The expression for  $\langle E_\alpha \rangle$  is just a multi-dimensional integral, with  $3N$  integrations in general. In realistic systems, the wave function  $\langle \mathbf{R} | \psi_\alpha \rangle$  for the many-body system is very small for most configurations  $\mathbf{R}$  (in this context, a “configuration” just means a specification of the positions of all the particles). Solution? Use importance sampling with the Metropolis algorithm!

The strategy is to use a collection of “random walkers” to build a set of  $\mathbf{R}$ ’s. Boltzmann weighting is not relevant in this problem; what do we use instead? We need a probability distribution, which naturally suggests using the wave function squared. Thus we’ll distribute points according to

$$\rho(\mathbf{R}) \equiv \frac{\psi_\alpha^2(\mathbf{R})}{\int d\mathbf{R} \psi_\alpha^2(\mathbf{R})} \quad (14.13)$$

(we assume  $\psi_\alpha$  is real). Note that

$$\int \rho(\mathbf{R}) d\mathbf{R} = 1 \quad (14.14)$$

by construction. If we define the *local energy*

$$E_L(\mathbf{R}) = \frac{H\psi_\alpha(\mathbf{R})}{\psi_\alpha(\mathbf{R})}, \quad (14.15)$$

then

$$\langle E \rangle = \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R}) \quad (14.16)$$

is the integral we treat with Metropolis. This is carried out for a practice case in Session 14. (Note: We might suspect there will be a problem with Eq. (14.15) if the wave function has nodes. When will this be an issue?)

The basic procedure is to compare probabilities when deciding whether to move from  $\mathbf{R}$  to  $\mathbf{R}'$ . That is, we compare  $\rho(\mathbf{R}')$  to  $\rho(\mathbf{R})$  and use the Metropolis decision algorithm:

1. If  $\rho(\mathbf{R}')/\rho(\mathbf{R}) > 1$ , then we accept the move;
2. if  $\rho(\mathbf{R}')/\rho(\mathbf{R}) < 1$ , then we generate a uniform random number  $s \in [0, 1]$  and accept the step if

$$\frac{\rho(\mathbf{R}')}{\rho(\mathbf{R})} \geq s. \quad (14.17)$$

See the `variational_SH0` code for a sample implementation.

## e. References

- [1] R.H. Landau and M.J. Paez, *Computational Physics: Problem Solving with Computers* (Wiley-Interscience, 1997). [See the 6810 info webpage for details on a new version.]
- [2] M. Hjorth-Jensen, *Computational Physics* (2013). These are notes from a course offered at the University of Oslo. See the 6810 webpage for links to excerpts.

- [3] W. Press *et al.*, *Numerical Recipes in C++*, 3rd ed. (Cambridge, 2007). Chapters from the 2nd edition are available online from <http://www.nrbook.com/a/>. There are also Fortran versions.