Once the system is thermalized, one can start measuring properties of the system. One computes a physical quantity every hundred or every thousand sweeps and takes the average of these measurements. That average is the mean value of the physical quantity at temperature T.

Why does this work? Consider two configurations x and x' which respectively have energies $E = E(x)$ and $E' = E(x')$ and are occupied with probabilities $P_t(x)$ and $P_t(x')$ as the system is thermalizing. If $E' > E$, then the rate $R(x' \to x)$ of going from x' to x is the rate v of choosing to test x when one is at x' times the probability $P_t(x')$ of being at x', that is, $R(x' \to x) = v\, P_t(x')$. The reverse rate is $R(x \to x') = v\, P_t(x)\, e^{-(E'-E)/kT}$ with the same v since the random walk is symmetric. The net rate from $x' \to x$ then is

$$R(x' \to x) - R(x \to x') = v \left(P_t(x') - P_t(x)\, e^{-(E'-E)/kT} \right). \tag{14.8}$$

This net flow of probability from $x' \to x$ is positive if and only if

$$P_t(x')/P_t(x) > e^{-(E'-E)/kT}. \tag{14.9}$$

The probability distribution $P_t(x)$ therefore flows with each sweep toward the Boltzmann distribution $\exp(-E(x)/kT)$. The flow slows and stops when the two rates are equal $R(x' \to x) = R(x \to x')$ a condition called **detailed balance**. At this equilibrium, the distribution $P_t(x)$ satisfies $P_t(x) = P_t(x')\, e^{-(E-E')/kT}$ in which $P_t(x')\, e^{E'/kT}$ is independent of x. So the thermalizing distribution $P_t(x)$ approaches the distribution $P(x) = c\, e^{-E/kT}$ in which c is independent of x. Since the sum of these probabilities must be unity, we have

$$\sum_x P(x) = c \sum_x e^{-E/kT} = 1 \tag{14.10}$$

which means that the constant c is the inverse of the **partition function**

$$Z(T) = \sum_x e^{-E(x)/kT}. \tag{14.11}$$

The thermalizing distribution approaches Boltzmann's distribution (1.345)

$$P_t(x) \to P_B(x) = e^{-E(x)/kT}/Z(T). \tag{14.12}$$

Example 14.2 (Z_2 Lattice Gauge Theory) First, one replaces space-time with a lattice of points in d dimensions. Two nearest neighbor points are separated by the lattice spacing a and joined by a link. Next, one puts an

element U of the gauge group on each link. For the Z_2 gauge group (example 10.4), one assigns an action $S_\square$ to each elementary square or *plaquette* of the lattice with vertices 1, 2, 3, and 4

$$S_\square = 1 - U_{1,2}U_{2,3}U_{3,4}U_{4,1}. \tag{14.13}$$

Then, one replaces $E(x)/kT$ with βS in which the action S is a sum of all the plaquette actions S_p. More details are available at Michael Creutz's website (latticeguy.net/lattice.html). □

Although the generation of configurations distributed according to the Boltzmann probability distribution (1.345) is one of its most useful applications, the Monte Carlo method is much more general. It can generate configurations x distributed according to any probability distribution $P(x)$.

To generate configurations distributed according to $P(x)$, we accept any new configuration x' if $P(x') \geq P(x)$ and also accept x' with probability

$$P(x \to x') = P(x')/P(x) \tag{14.14}$$

if $P(x) > P(x')$.

This works for the same reason that the Boltzmann version works. Consider two configurations x and x'. If the system is thermalized, then the probabilities $P_t(x)$ and $P_t(x')$ have reached equilibrium, and so the rate $R(x \to x')$ from $x \to x'$ must equal that $R(x' \to x)$ from $x' \to x$. If $P(x') < P(x)$, then $R(x' \to x)$ is

$$R(x' \to x) = v\, P_t(x') \tag{14.15}$$

in which v is the rate of choosing $\delta x = x' - x$, while the rate $R(x \to x')$ is

$$R(x \to x') = v\, P_t(x)\, P(x')/P(x) \tag{14.16}$$

with the same v since the random walk is symmetric. Equating the two rates

$$R(x' \to x) = R(x \to x') \tag{14.17}$$

we find that the flow of probability stops when

$$P_t(x) = P(x)\, P_t(x')/P(x') = c\, P(x) \tag{14.18}$$

where c is independent of x'. Thus $P_t(x) \to P(x)$.

So far we have assumed that the rate of choosing $x \to x'$ is the same as the rate of choosing $x' \to x$. In **Smart Monte Carlo** schemes, physicists arrange the rates $v_{x \to x'}$ and $v_{x' \to x}$ so as to steer the flow and speed-up thermalization. To compensate for this asymmetry, they change the second

part of the Metropolis step from $x \to x'$ when $E' = E(x') > E = E(x)$ to accept conditionally with probability

$$P(x \to x') = P(x')\, v_{x' \to x} / \left[P(x)\, v_{x \to x'}\right]. \tag{14.19}$$

Now if $P(x') < P(x)$, then $R(x' \to x)$ is

$$R(x' \to x) = v_{x' \to x}\, P_t(x') \tag{14.20}$$

while the rate $R(x \to x')$ is

$$R(x \to x') = v_{x \to x'}\, P_t(x)\, P(x')\, v_{x' \to x} / \left[P(x)\, v_{x \to x'}\right]. \tag{14.21}$$

Equating the two rates $R(x' \to x) = R(x \to x')$, we find

$$P_t(x') = P_t(x)\, P(x')/P(x). \tag{14.22}$$

That is $P_t(x) = P(x)\, P_t(x')/P(x')$ which gives

$$P_t(x) = N\, P(x) \tag{14.23}$$

where N is a constant of normalization.

14.5 Solving Arbitrary Problems

If you know how to generate a suitably large space of trial solutions to a problem, and you also know how to compare the quality of any two of your solutions, then you can use a Monte Carlo method to solve it. The hard parts of this seemingly magical method are characterizing a big enough space of solutions s and constructing a quality function or functional that assigns a number $Q(s)$ to every solution in such a way that if s is a better solution than s', then

$$Q(s) > Q(s'). \tag{14.24}$$

But once one has characterized the space of possible solutions s and has constructed the quality function $Q(s)$, then one simply generates zillions of random solutions and selects the one that maximizes the function $Q(s)$ over the space of all solutions.

If one can characterize the solutions as vectors of a certain dimension, $s = (x_1, \ldots, x_n)$, then one may use the Monte Carlo method of the previous section (14.4) by replacing $-E(s)$ with $Q(s)$ and kT with a parameter of the same dimension as $Q(s)$, nominally dimensionless.