1.3 [20/20] <1.6> Your colleague at AMD suggests that, since the yield is so poor, you might make chips more cheaply if you released multiple versions of the same chip, just with different numbers of cores. For example, you could sell Phoenix^8, Phoenix^4, Phoenix^2, and Phoenix^1, which contain 8, 4, 2, and 1 cores on each chip, respectively. If all eight cores are defect-free, then it is sold as Phoenix^8. Chips with four to seven defect-free cores are sold as Phoenix^4, and those with two or three defect-free cores are sold as Phoenix^2. For simplification, calculate the yield for a single core as the yield for a chip that is 1/8 the area of the original Phoenix chip. Then view that yield as an independent probability of a single core being defect free. Calculate the yield for each configuration as the probability of at the corresponding number of cores being defect free.

   a. [20] <1.6> What is the yield for a single core being defect free as well as the yield for Phoenix^8, Phoenix^4 and Phoenix^1?

   b. [5] <1.6> Using your results from part a, determine which chips you think it would be worthwhile to package and sell, and why.

   c. [10] <1.6> If it previously cost $20 dollars per chip to produce Phoenix^8, what will be the cost of the new Phoenix chips, assuming that there are no additional costs associated with rescuing them from the trash?

   d. [20] <1.6> You currently make a profit of $30 for each defect-free Phoenix^8, and you will sell each Phoenix^4 chip for $25. How much is your profit per Phoenix^8 chip if you consider (i) the purchase price of Phoenix^4 chips to be entirely profit and (ii) apply the profit of Phoenix^4 chips to each Phoenix^8 chip in proportion to how many are produced? Use the yields calculated from part Problem 1.3a, not from problem 1.1a.

1.5 [10/10/10] <1.5> As mentioned in Exercise 1.4, cell phones run a wide variety of applications. We'll make the same assumptions for this exercise as the previous one, that it is 0.5 W per core and that a quad core runs email 3× as fast.

   a. [10] <1.5> Imagine that 80% of the code is parallelizable. By how much would the frequency and voltage on a single core need to be increased in order to execute at the same speed as the four-way parallelized code?

   b. [10] <1.5> What is the reduction in dynamic energy from using frequency and voltage scaling in part a?

   c. [10] <1.5> How much energy is used with a dark silicon approach? In this approach, all hardware units are power gated, allowing them to turn off entirely (causing no leakage). Specialized ASICs are provided that perform the same computation for 20% of the power as the general-purpose processor. Imagine that each core is power gated. The video game requires two ASICs and two cores. How much dynamic energy does it require compared to the baseline of parallelized on four cores?
1.9  [10/10/20/20] <1.5> Server farms such as Google and Yahoo! provide enough compute capacity for the highest request rate of the day. Imagine that most of the time these servers operate at only 60% capacity. Assume further that the power does not scale linearly with the load; that is, when the servers are operating at 60% capacity, they consume 90% of maximum power. The servers could be turned off, but they would take too long to restart in response to more load. A new system has been proposed that allows for a quick restart but requires 20% of the maximum power while in this “barely alive” state.

a. [10] <1.5> How much power savings would be achieved by turning off 60% of the servers?

b. [10] <1.5> How much power savings would be achieved by placing 60% of the servers in the “barely alive” state?

c. [20] <1.5> How much power savings would be achieved by reducing the voltage by 20% and frequency by 40%?

d. [20] <1.5> How much power savings would be achieved by placing 30% of the servers in the “barely alive” state and 30% off?

1.12  [20/10/10/15] <1.9> In this exercise, assume that we are considering enhancing a quad-core machine by adding encryption hardware to it. When computing encryption operations, it is 20 times faster than the normal mode of execution. We will define percentage of encryption as the percentage of time in the original execution that is spent performing encryption operations. The specialized hardware increases power consumption by 2%.

a. [20] <1.9> Draw a graph that plots the speedup as a percentage of the computation spent performing encryption. Label the y-axis “Net speedup” and label the x-axis “Percent encryption.”

b. [10] <1.9> With what percentage of encryption will adding encryption hardware result in a speedup of 2?

c. [10] <1.9> What percentage of time in the new execution will be spent on encryption operations if a speedup of 2 is achieved?

d. [15] <1.9> Suppose you have measured the percentage of encryption to be 50%. The hardware design group estimates it can speed up the encryption hardware even more with significant additional investment. You wonder whether adding a second unit in order to support parallel encryption operations would be more useful. Imagine that in the original program, 90% of the encryption operations could be performed in parallel. What is the speedup of providing two or four encryption units, assuming that the parallelization allowed is limited to the number of encryption units?