MCS572 Introduction to Supercomputing Spring 2003 – Hanson

Theoretical Individual Homework: Supercomputer Performance

Homework due Monday 24 February 2003 in class.

1. For the *Problem Dependent, Nested Loop Model of Hanson (modified Amdahl model)*, find a two-term (two non-zero terms) asymptotic approximation for the *speedup* and *efficiency*, when the time using p processors is

$$T_p(m,N) = \tau \cdot \left[K_0 + \frac{1}{p} \sum_{i=1}^{m} K_i N^i \right],$$

with fixed time scale per flop τ, constant flop counts K_0 for serial terms, K_i at nest depth i, for $i = 1 : m$ with $K_m > 0$, for loop nests up to fixed depth m with a common N iterations in each loop for the following limits described below. Make graphical sketches of the speedup and efficiency for the following cases to demonstrate the the appropriate limits, either massively parallel as $p \to \infty$ or massive workload as $N \to \infty$:

   (a) *Massively Parallel Limit (MPL)* as $p \to +\infty$ for fixed problem size N, i.e., show that

   $$S_p(m,N) \sim c_1 \cdot (1 - c_2/p^{\alpha_1})$$

   by explicitly finding the constants c_1, c_2 and α_1 and by sketching both speedup and efficiency versus p for fixed N and m based upon the asymptotic approximation.

   (b) *Massive Work Limit (MWL)* as $N \to \infty$ for fixed p and m, i.e., show that

   $$S_p(m,N) \sim c_3 \cdot (1 - c_4/N^{\alpha_2})$$

   by explicitly finding the constants c_3, c_4 α_2 and sketching both speedup and efficiency versus N for fixed p and m.

   (c) *Intermediate, Nonuniform Limit* along the curve $p = \gamma \cdot N^m$ as $N \to \infty$ for fixed $\gamma > 0$, i.e., show

   $$S_p(m,N) \sim c_5 \cdot N^m \cdot (1 + c_6/N^{\alpha_3})$$

   explicitly finding the constants c_5, c_6, and α_3 and sketching both speedup and efficiency versus N for fixed m. (Hint: α_3 *is not* m *and beware of massive cancellation since* c_6 *depends on* K_{m-1} *too.*)

   Justify your steps by giving reasons for all parts. Also, note the dominant term in a polynomial as its argument becomes large is the largest power (degree) term.

2. For the *General Linear Idle Processor Model*, the frequency or probability that there are q busy processors is

$$f_q(\theta) = \beta \cdot [(1-\theta) \cdot (p+1) + (2\theta - 1) \cdot q],$$

where $1 \le q \le p$, compare the decreasing ($\theta = 0$), uniform ($\theta = 0.5$) and increasing ($\theta = 1$) distributions:

(a) Find an expression for the normalization constant β to conserve total probability in terms of the available number of processors p and linear model parameter θ, in general.

(b) Find the estimated speedup $\widehat{S}_p(\theta) = T_1/\overline{t_q}$ for each of the three distributions for sufficiently large p, when the time on q ideally parallelized processors is $t_q = T_1/q$. Rank the three cases in estimated speedup performance.

(c) Compare the estimated efficiency $\widehat{E}_p(\theta) = \widehat{S}_p(\theta)/p$ for the three cases. Explain in words why the three cases ($\theta = 0, 0.5, 1$) differ in terms of processor utilization and work load balance for sufficiently large p. Rank the three cases in efficiency.

(*If necessary, it may be assumed that* $\sum_{q=1}^{p}(1/q) \sim \ln(p) + 1$ *for sufficiently large* p.)

3. For the *Hockney Linear Performance Model*, find the best linear (averaged) estimate of the performance $\widehat{R}_n$ and its asymptotic form $\widehat{R}_\infty$ as $n \to \infty$ and the half-peak problem size $\widehat{n}_{\frac{1}{2}}$ for the parallel processor with multiple pipelines per processor, having timing

$$T_{n,k,m,p} = \tau \left[\sigma + k - 1 + \left\lceil \frac{n}{m \cdot p} \right\rceil \right],$$

where τ is the constant time scaling, σ is the constant pipeline startup, k is the constant number of pipeline stages, n is the problem size here in elements, p is the number of processors, m is the number of pipelines per processor, and $\lceil x \rceil$ denotes the integer ceiling function of x. Assume there are K floating point operations per problem element. You must show details of your calculation and explicitly relate the Hockney model parameters to the original timing model parameters. The units of $T_{n,k,p}$ and τ are nanoseconds, while $\widehat{R}_\infty$ is in GigaFlops. Derive the various asymptotic behaviors of the $\widehat{R}_{n,k,m,p}$ as n, p and m approach infinity, respectively.

4. Using the simple or *Cray Weak Data Dependency Test*, demonstrate by suitable loop examples that the two loop cases PLD (Previous, Lesser, Decreasing) and PGI (Previous, Greater, Increasing) for unit stride (± 1) generate no data dependent conflict in parallel synchronization compared to serial execution standard of accuracy, modulo floating point precision accuracy. Discuss the logical equivalence of PGI and PLD. Confirm the conclusions from your loop examples by applying the *Cray Strong Dependency Test*. Hint: *Here "Previous" refers to the loop area preceding the definition of the element* i *of the array* a*, "Greater/Lesser" refers to the subscript of the argument reference to* a *in that area, and "Decreasing/Increasing" refers to the stride direction of the loop index iteration.*

Web Source: http://www.math.uic.edu/∼hanson/mcs572/hw1mcs572s03.pdf
Email Comments or Corrections or Questions to hanson@uic.edu