MCS572F96 Supercomputer Performance – Theoretical Homework 1 – Hanson

Homework due Monday 07 October 1996 in class.

1. Find a two-term asymptotic approximation for the *speedup* and *efficiency* of the *problem dependent, nested loop modified Amdahl model*,

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

where $K_m \neq 0$, for loop nests up to depth m with N iterations per loop in the following limits described below. Make graphical sketches of the speedup and efficiency for the following cases:

   (a) *Massively Parallel Processor Limit* as $p \to +\infty$ for fixed N (i.e., show $S_p \sim C_1 \cdot (1 + C_2/p^{\alpha_1})$ giving the constants C_1, C_2 and α_1 explicitly and sketching both speedup and efficiency versus p for fixed N and m).

   (b) *Massive Problem Limit* as $N \to \infty$ for fixed p (i.e., show $S_p \sim C_3 \cdot (1 + C_4 N^{-m})$ giving the constants C_3 and C_4 explicitly and sketching versus N for fixed p and m).

   (c) *Nonuniform Intermediate Limit* along the curve $p = \gamma \cdot N^m$ as $N \to \infty$ with $\gamma > 0$ and constant (i.e., show $S_p \sim C_5 \cdot N^m \cdot (1 + C_6 N^{\alpha_3})$ giving the constants C_5, C_6, and α_3 explicitly and sketching versus N for fixed m).

   Give reasons all steps.

2. Compare *Amdahl's version of parallel speedup* for fixed problem size N

$$S_p(N) = \frac{T_1(N)}{T_p(N)}$$

with *Gustafson's scaled speedup*,

$$SS_p(N) = \frac{p \cdot T_1(N/p)}{T_p(N)},$$

for the parallel overhead model,

$$T_p(N) = \tau \cdot \left[K_1 + K_2 \cdot \frac{N}{p} + K_3 \cdot \frac{\sqrt{p}}{N}\right],$$

where the K_i are positive constants. In your comparison, a sketch of the behavior of both speedups S_p versus SS_p is sufficient. (Note: Amdahl's Speedup is not the same as Amdahl's Performance Law.)

3. Show that when Gustafson's linear scaled speedup

$$SS_p = \sigma' + (1 - \sigma') \cdot p$$

has the same value as Amdahl's speedup

$$S_p = \frac{1}{\sigma + (1 - \sigma)/p},$$

(i.e., $SS_p = \gamma = S_p$, for some $\gamma > 0$), then

$$SS_p = \frac{\sigma'}{\sigma} = S_p$$

and the corresponding efficiencies satisfy

$$SE_p = \frac{\alpha'}{\alpha} = E_p,$$

where σ' and σ are the respective scaled and unscaled scalar fractions, while α' and α are the corresponding parallel fraction complements. Hint: solve for σ and σ', and similarly for α and α', respectively.

4. For the Hockney linear performance model, find the best linear estimate of the asymptotic performance $\widehat{R}_\infty$ and the half-peak problem size $\widehat{N}_{1/2}$ for the pipelined array processor with timing

$$t_{N,k,p} = \tau \left[\sigma + k - 1 + \left\lceil \frac{N}{p} \right\rceil \right],$$

where τ is the constant time scaling, σ is the constant pipeline startup, k is the constant number of pipeline levels, N is the problem size, p is the number of pipelines arrayed in the processor, and $\lceil x \rceil$ denotes the integer roof function of x. 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.

5. For the *general linear idle processor model,*

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

compare the decreasing ($\theta = 0$), uniform ($\theta = 1/2$) and increasing ($\theta = 1$) distributions:

   (a) First find an expression for the normalization constant β in terms of the available number of processors p and linear model parameter θ, in general.

   (b) Find the estimated speedup $\widehat{S}_p = T_1/\overline{t_q}$ for each of the three distributions.

   (c) Compare the estimated efficiency $\widehat{E}_p = \widehat{S}_p/p$ for the three cases and explain why the three cases differ in terms of processor utilization and load balance.

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