CS4617 Computer Architecture

Lecture 2a

Dr J Vaughan

September 15, 2014

Die yield

- Fraction of good dies on wafer = die yield
- $Die\ yield =$
  $Wafer\ yield \times 1/(1 + Defects\ per\ unit\ area \times Die\ area)^N$
- This is the Bose-Einstein formula: an empirical model
  - *Wafer yield* accounts for wafers that are completely bad, with no need for testing
  - *Defects per unit area* accounts for random manufacturing defects = 0.016 to 0.057 per cm^2
  - N = *process complexity factor, measures manufacturing difficulty*
    = *11.5 to 15.5 for a 40nm process (in 2010)*

Yield

- Example
  - Find the number of dies per 300mm wafer for a die that is 1.5 cm square.
- Solution

$$
\begin{aligned}
Die\ area &= 1.5 \times 1.5 = 2.25cm^2 \\
Dies\ per\ wafer &= \frac{\pi \times (30/2)^2}{2.25} - \frac{\pi \times 30}{\sqrt{2 \times 2.25}} \\
&= 270
\end{aligned}
$$

Yield

- Example
  - Assuming the wafer yield is 100%, find the die yield for a die that is 1.5 cm square
  - The defect density is 0.031 per cm^2 and the process complexity factor is 13.5
- Solution

$$\begin{aligned} \textit{Die yield} &= \frac{1}{(1 + 0.031 \times 2.25)^{13.5}} \\ &= 0.4 \end{aligned}$$

- Note
  - Defect density per chip $= 0.031 \times 2.25 = .06975 \approx 7\%$
  - Why is yield not 93%?

Other costs

- The computer designer affects die size and cost by including/excluding functions and by the number of I/O pins.
- Mask costs for a high-density fabrication process with 4–6 layers exceed $1M
  - This is a fixed cost that is likely to increase
  - It makes small-volume production very expensive

Problem

- What power savings can be made by cutting processor voltage by 20% and clock frequency by 40%?
- $\frac{Energy_{new}}{Energy_{old}} = \frac{(Voltage \times 0.80)^2}{Voltage^2} = 0.8^2 = 0.64$
- $\frac{Power_{new}}{Power_{old}} = 0.64 \times \frac{Frequency_{switched} \times 0.6}{Frequency_{switched}} = 0.384$

Problem

- An enhancement to a computer accelerates some mode of execution by a factor of 10. The enhanced mode is used 50% of the time, measured as a percentage of the execution time when the enhanced mode is in use. Recall that Amdahl's Law depends on the fraction of the original unenhanced execution time that could make use of enhanced mode. Thus the 50% measurement cannot be used directly to compute speedup with Amdahl's Law.
- What is the speedup obtained from fast mode?
- What percentage of the original execution time has been converted to fast mode?

Approach to solution

- Up to now, we have phrased Amdahl's Law in terms of the fraction of an "old" system that can be *enhanced*. Call this $F_{enhanced}$
- This time, we are given the fraction of execution time that the enhanced component is used *after* the enhancement has taken place. Call this $F_{dehanced}$ to show that we can reverse the process to convert it to the original fraction that can be enhanced
- First, apply Amdahl's Law in the normal "forward" direction to get an expression for the ratio of new to old execution times
- Then, apply Amdahl's Law backwards in time to get an expression for the ratio of old to new execution times

Solution

$$\begin{aligned}
\frac{Exectime_{new}}{Exectime_{old}} &= (1 - F_{enhanced}) + F_{enhanced} \times \frac{1}{speedup_{enhanced}} \\
&= (1 - F_{enhanced}) + 0.1 \times F_{enhanced} \\
&= 1 - 0.9 \times F_{enhanced} \\
\frac{Exectime_{old}}{Exectime_{new}} &= (1 - F_{dehanced}) + F_{dehanced} \times \frac{1}{speedup_{dehanced}} \\
&= 0.5 + \frac{0.5}{0.1} \\
&= 5.5 \\
5.5 &= \frac{1}{1 - 0.9 \times F_{enhanced}} \\
5.5 - 4.95 \times F_{enhanced} &= 1 \\
F_{enhanced} &= \frac{4.5}{4.95} \\
&= 0.91
\end{aligned}$$

Check

$$
\begin{aligned}
Speedup &= \frac{1}{(1-0.91)+0.91\times\frac{1}{10}} \\
&= \frac{1}{0.09+0.091} \\
&= \frac{1}{0.181} \\
&= 5.52
\end{aligned}
$$

Problem

- What is the speedup with N processors if 80% of the application is parallelizable, ignoring the cost of communication?
- $$\frac{Exectime_{old}}{Exectime_{new}} = \frac{1}{(1 - F_{enhanced}) + F_{enhanced} \times \frac{1}{speedup_{enhanced}}}$$
  $$= \frac{1}{0.2 + \frac{0.8}{N}}$$

Problem

- What is the speedup with 8 processors if, for every processor added, the communication overhead is 0.5% of the original execution time?
- N = 8

$$\frac{Exectime_{old}}{Exectime_{new}} = \frac{1}{(1 - F_{enhanced}) + F_{enhanced} \times \frac{1}{speedup_{enhanced}}}$$
$$= \frac{1}{0.2 + \frac{0.8}{8} + 0.005 \times 8}$$
$$= \frac{1}{0.34}$$
$$= 2.94$$

Problem

- What is the speedup with 8 processors if, for every time the number of processors is doubled, the communication overhead is increased by 0.5% of the original execution time?
- N = 8

$$\frac{Exectime_{old}}{Exectime_{new}} = \frac{1}{0.2 + \frac{0.8}{8} + 0.005 \times 3}$$
$$= \frac{1}{0.315}$$
$$= 3.17$$