

Algorithm Engineering

Jens K. Mueller

`jkm@informatik.uni-jena.de`

Department of Mathematics and Computer Science
Friedrich Schiller University Jena

Monday 2nd February, 2015

Theoretical Limits

Theoretical and Effective Performance

- ▶ Theoretical Performance
 - ▶ Number of nodes: 1
 - ▶ Number of CPUs: 1
 - ▶ Number of Cores: 1
 - ▶ CPU frequency: 2.66 GHz
 - ▶ Number of operations per cycle: 4 FLOP/cycle (single precision)

Theoretical (peak) performance

$$2.66 \text{ GHz} \cdot 4 \text{ FLOP/cycle} \approx 10.6 \text{ GFLOP/s}$$

- ▶ Effective performance determines the number of operations per time

Theoretical and Effective Bandwidth

- ▶ Theoretical bandwidth
 - ▶ Memory clock: 1066 MHz
 - ▶ Type of memory: DDR2 (double data rate)
 - ▶ Number of channels: 2
 - ▶ Memory bus: 64 bit

$$1066 \text{ MHz} \cdot 2 \cdot 2 \cdot 64 \text{ bit} / 8 = 34\,112 \text{ MB/s} \approx 34.1 \text{ GB/s}$$

- ▶ Effective bandwidth

$$(B_r + B_w) / t,$$

where B_r is the number of bytes read, B_w number of bytes written and t the spent time.

System Information

- ▶ CPU

```
$ lshw -C cpu
```

- ▶ Memory

```
$ lshw -short -C memory
```

IO bound vs Compute bound

- ▶ Average #operations per in-/output

$$\frac{\text{number of operations}}{\text{size of in- and output}}$$

- ▶ Number of operations needed to avoid CPU stalls (due to memory accesses)

Amdahl's Law

Fraction x of program optimized by a factor s , then

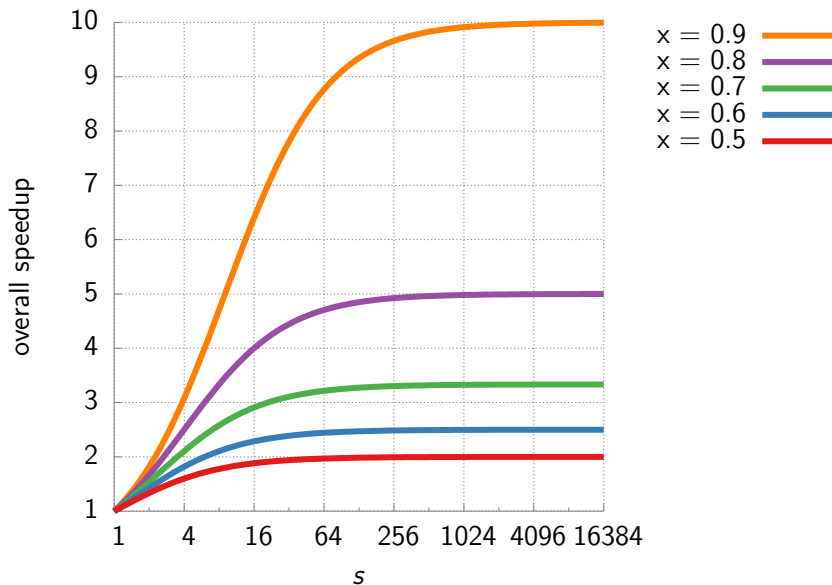
$$T_s = \underbrace{(1 - x)}_{\text{unoptimized}} + \underbrace{\frac{x}{s}}_{\text{optimized}}$$

Assume $T = 1$ (unoptimized program's running time), then overall speedup is

$$\frac{T}{T_s} = \frac{1}{(1 - x) + \frac{x}{s}}$$

- ▶ $x = 1$ yields overall speed up of s
- ▶ Strong scaling
Overall speed up for fixed input size

Amdahl's Law (cont.)



Gustafson's Law

Embarrassingly parallel fraction x of program, then

$$T = (1 - x) + p \cdot x$$

Parallel running time is

$$T_p = (1 - x) + x = 1$$

Hence, speed up is

$$\frac{T}{T_p} = (1 - x) + p \cdot x = 1 + (p - 1) \cdot x$$

- ▶ Parallel part grows linear with the number of processors
- ▶ Input sizes increase
- ▶ Weak scaling

Overall speed up for increasing input size

Gustafson's Law (cont.)

