Erlangen Regional Computing Center

UNIVERSITÄT GREIFSWALD
Wissen lockt. Seit 1456

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Winter term 2020/2021

Parallel Programming with OpenMP and MPI

Dr. Georg Hager
Erlangen Regional Computing Center (RRZE) at Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute of Physics, Universität Greifswald

Assignment 1 discussion

High Performance Computing

Assignment 1, Task 1

- Hockney Model: $T = \lambda + \frac{N}{b}$, where λ is latency, N is message size, and b is bandwidth
- Effective bandwidth: $B_{\text{eff}} = N/T$

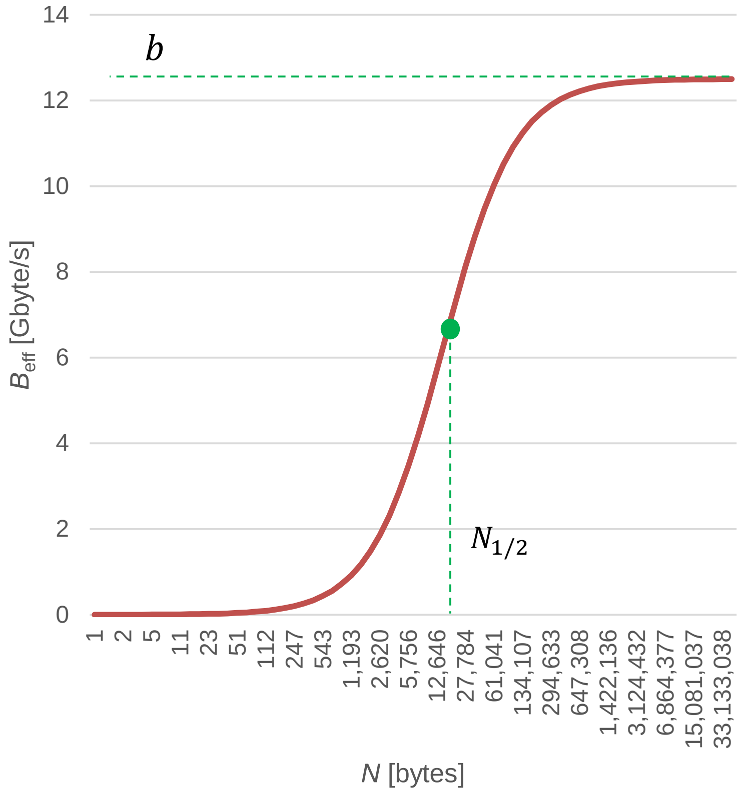
(a) N for half asymptotic bandwidth? → $\frac{N_{1/2}}{\lambda + \frac{N_{1/2}}{b}} = \frac{b}{2}$ → $N_{1/2} = \lambda b$

Lower is better → bandwidth increase will require larger messages to get to half max.!

Assignment 1, Task 1

(b) Effective bandwidth vs. message size

→ $N_{1/2} = 15\ \mathrm{kB}$



Assignment 1, Task 2

- Intel Xeon Phi “Knights Landing” Coprocessor (2016, discontinued)
  - Clock speed: up to **1.4 GHz**
  - SIMD register width: **512 bit**
  - Floating-point superscalarity: **2 FMA** instructions per cycle
  - Cache size up to 36 Mbyte
  - Memory bandwidth (measured): 480 Gbyte/s
  - Number of cores: up to **72**

(a) DP peak: $P_{peak} = 72 \times 8 \times 2 \times 2 \times 1.4 \cdot 10^9 \text{ flop/s} = \mathbf{3.23\ Tflop/s}$

Assignment 1, Task 2

(b) Upper performance bound for

```
for(int i=0; i<10000000; i++)
  a[i] += s*b[i];
```

Minimum execution times:

- $T_{BW} = \frac{24\times10^7 \text{byte}}{480\times10^9 \text{byte/s}} = 500\ \mu s$
- $T_{flops} = \frac{2\times10^7 \text{flop}}{3.23\times10^{12} \text{flop/s}} = 6.19\ \mu s$

Roofline shortcut: per-loop bound for two bottlenecks

$$P_{bound} = \min(P_{peak}, b/B_c),$$

where B_c is the code balance:

$$B_c = \frac{\text{bytes transferred}}{\text{flops executed}}$$

→ Upper performance bound: $P_{bound} = \frac{2\times10^7 \text{flop}}{T_{BW}} = 40\ \text{Gflop/s}$