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 7 discussion

High Performance Computing

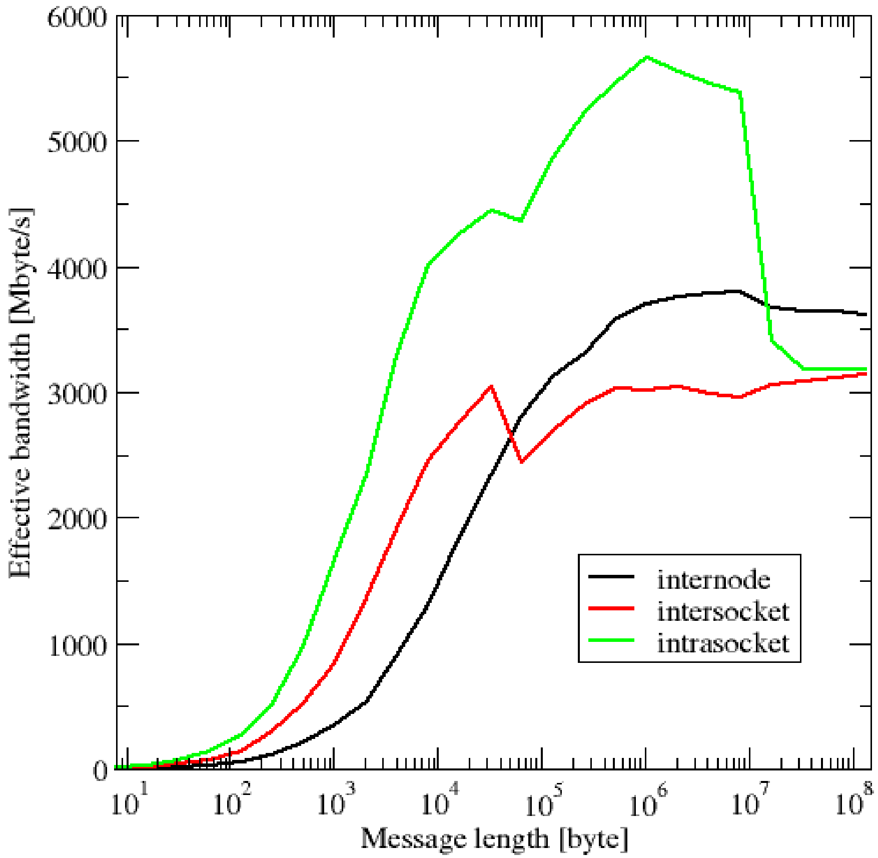
Assignment 7, Task 1 – MPI ping-pong

```
int rank,N;
MPI_Status s;

MPI_Comm_rank(MPI_COMM_WORLD, &rank);

int dest = (rank+1)%2;

if(0==rank) wc = MPI_Wtime();

for(int i=0; i<repeat; ++i) {
  if(0==rank) {
    MPI_Send(buffer, N, MPI_BYTE, dest, 0, MPI_COMM_WORLD);
    MPI_Recv(buffer, N, MPI_BYTE, dest, 0, MPI_COMM_WORLD, &s);
  } else {
    MPI_Recv(...);
    MPI_Send(...);
  }
}

if(0==rank) wc = MPI_Wtime() - wc;
double Beff = 2.0*N/wc; // effective bandwidth
```

MPI ping-pong

Observations:

- Asymptotic BW quite similar across experiments
- Intranode as. BW even lower than internode
- Protocol switch is clearly visible for intranode @ ≈ 30 kB
  → no simple Hockney model

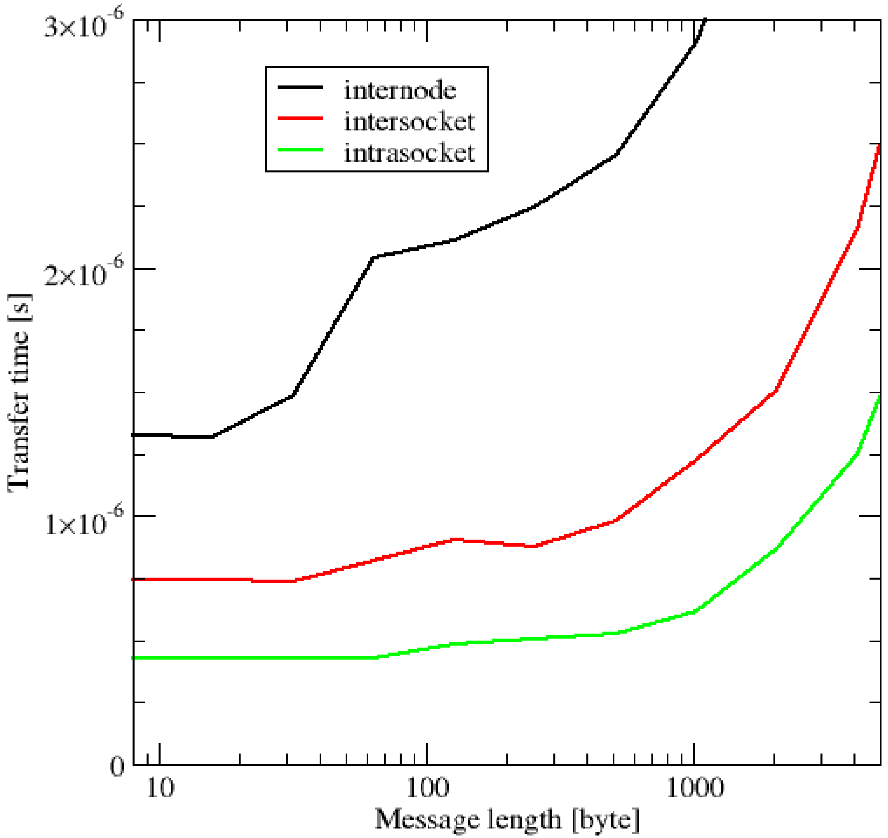


MPI ping-pong

Observations:

- Latency-dominated regime is easily reached @ V=10 byte
- Internode $T_\ell \approx 1.3\ \mu s$
- Intersocket $T_\ell \approx 750\ \mathrm{ns}$
- Intrasocket $T_\ell \approx 450\ \mathrm{ns}$

- Protocol switches visible → no simple Hockney model



Assignment 7 – Task 2: message aggregation

- Parameters:
  asymptotic bandwidth $B = 125\ \text{Mbyte/s}$
  Latency $T_\ell = 30\ \mu s$
  Message length m in bytes
- Time for one 1000-byte message:

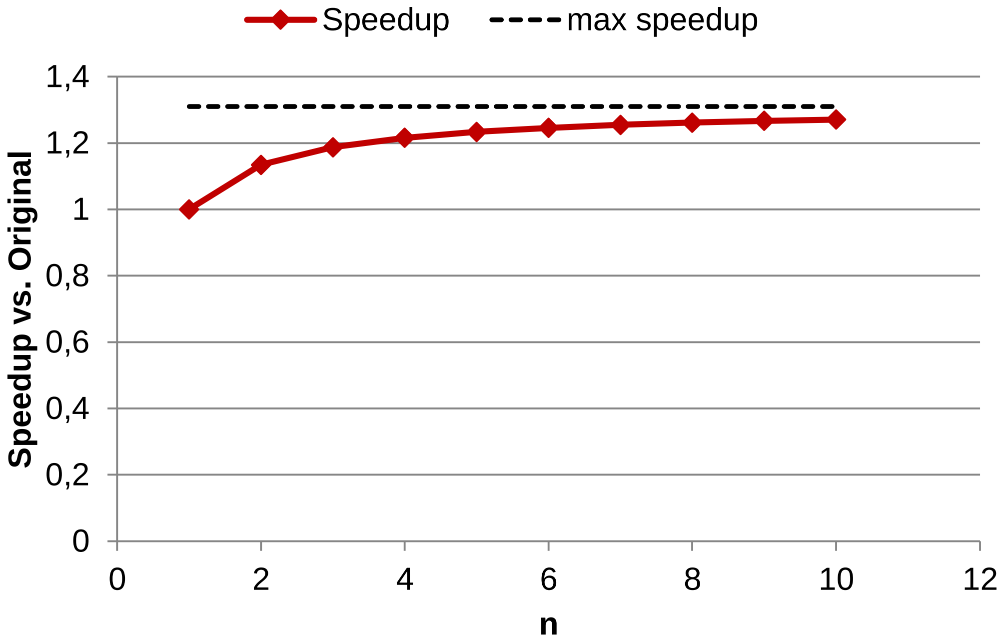
$$T_{msg} = T_\ell + \frac{m}{B} = \left(30 + \frac{1000}{125}\right)\mu s = 38\ \mu s$$

  → dominated by latency!
- Time for n messages
  - without aggregation: nT_{msg}
  - with aggregation: $T_\ell + \frac{nm}{B}$

Message aggregation

Speedup by message aggregation for 30% comm. overhead:

$$S = \frac{T_{orig}}{T_{opt}} = \frac{n \cdot \left(T_l + \frac{m}{B}\right) \cdot \frac{10}{3}}{T_l + n\frac{m}{B} + n \cdot \left(T_l + \frac{m}{B}\right) \cdot \frac{7}{3}} \xrightarrow{n \to \infty} 1.31$$



Assignment 7 – Task 3: π with MPI

```
int i,b,rank,size,count=0,n=200000000,nn;
MPI_Status status;

MPI_Init(&argc, &argv);
MPI_Comm_rank(MPI_COMM_WORLD, &rank);
MPI_Comm_size(MPI_COMM_WORLD, &size);

seed = 2*rank;
double r = 1./RAND_MAX;

// strong scaling
nn = n/size + ((rank >= n%size)? 0:1);

wc = MPI_Wtime();
for(i=0; i<nn; ++i) {
  x = rand_r(&seed)*r;
  y = rand_r(&seed)*r;
  if((x*x+y*y) <1.0) ++count;
}
```

```
if(0==rank) {
    for(i=1; i<size; ++i) {
      MPI_Recv(&b, 1, MPI_INT,
               MPI_ANY_SOURCE,
               MPI_ANY_TAG,
               MPI_COMM_WORLD,
               &status);
      count += b;
    }
  } else {
    MPI_Send(&count, 1, MPI_INT,
             0, 0, MPI_COMM_WORLD);
  }
  wc = MPI_Wtime()-wc;
```

Assignment 7 – Task 3: π with MPI

```
$ for p in `seq 1 80`; do mpirun_rrze -np $p \
   -pin 0_1_2_3_4_5_6_7_8_9_10_11_12_13_14_15_16_17_18_19 \
   ./a.out done
```

4 Emmy nodes, 2. 2 GHz, Intel v17