

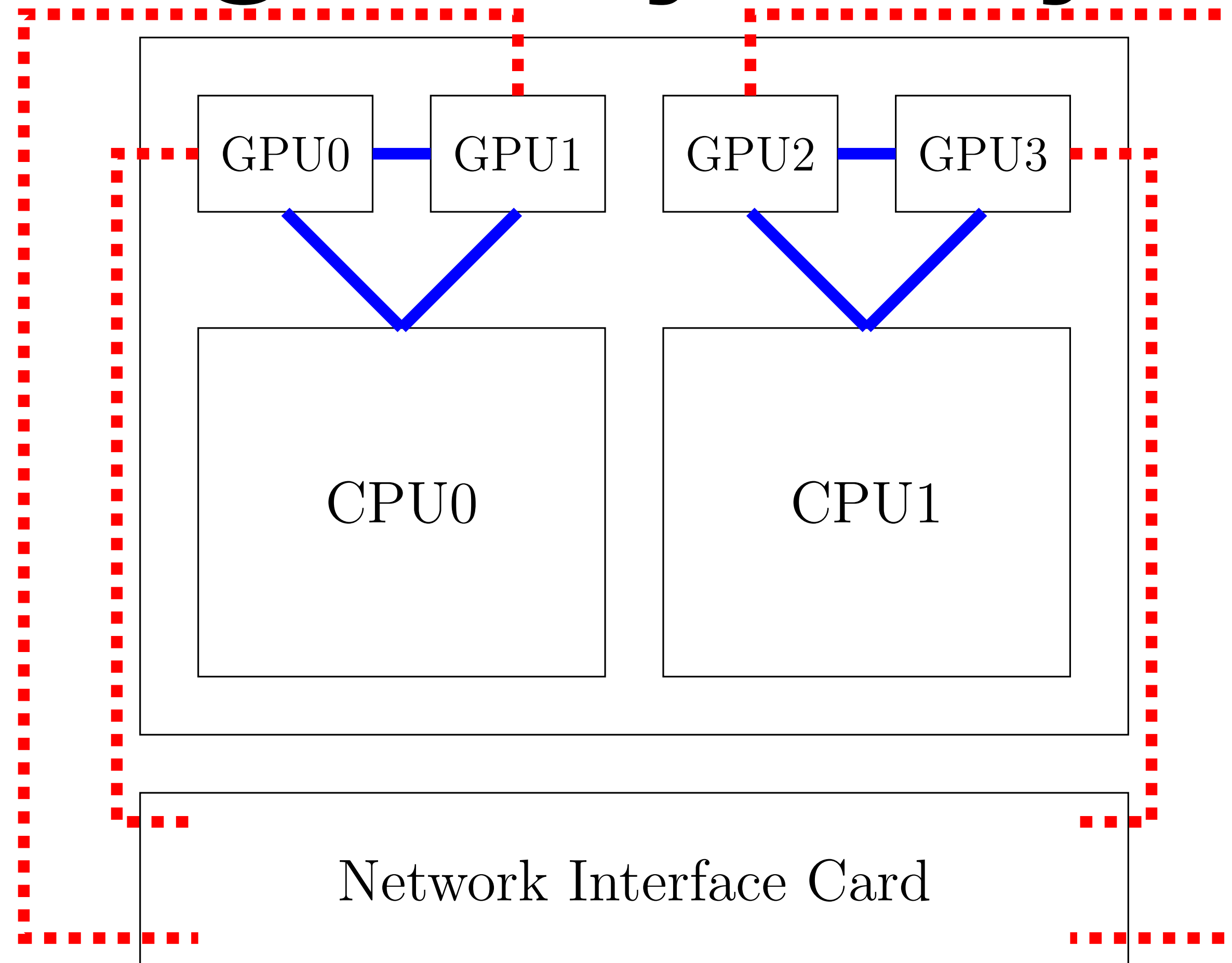
# Introduction to Parallel Processing

Lecture 24 : Performance of Heterogeneous  
Systems

11/28/2022

Professor Amanda Bienz

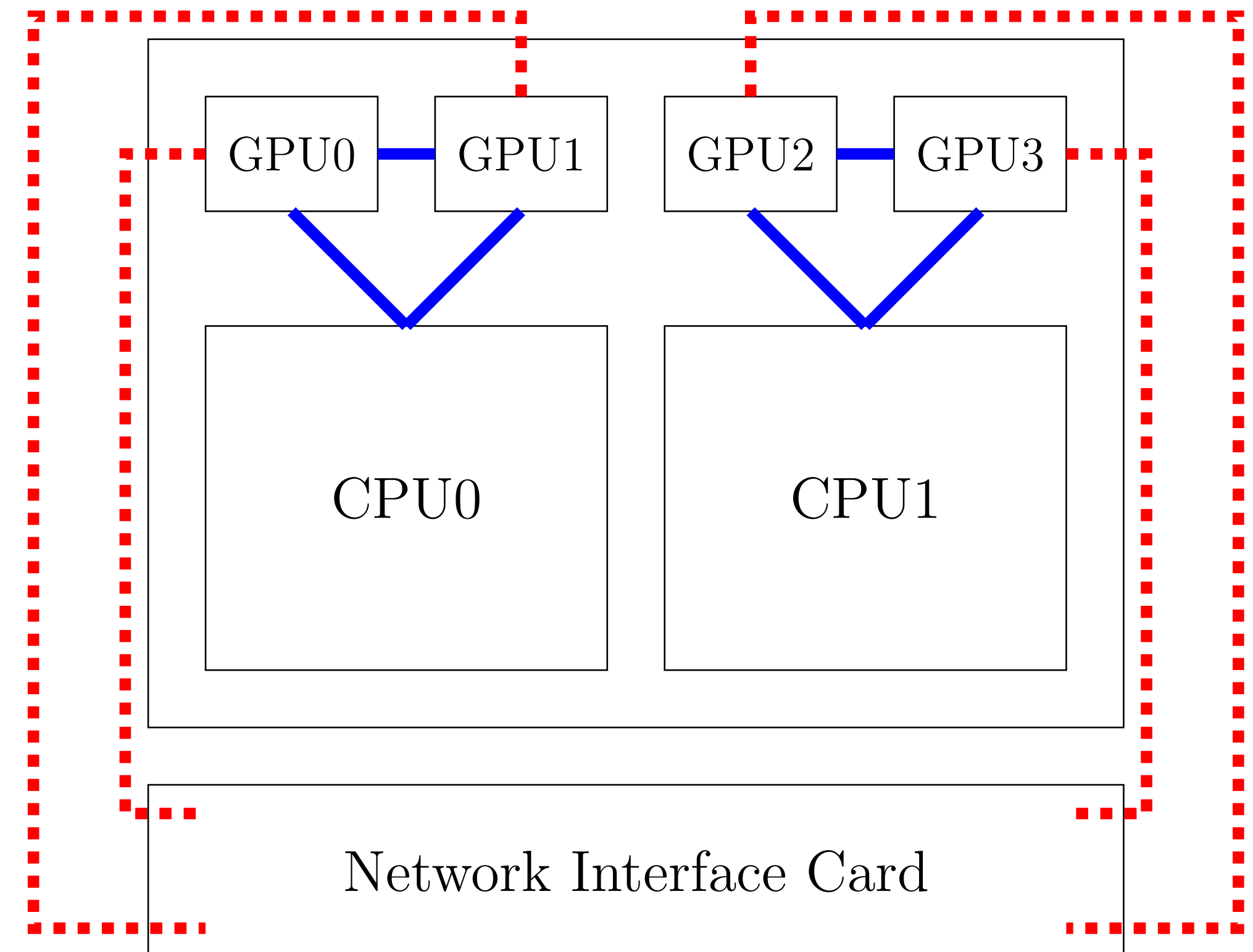
# Heterogeneity in Systems



Assuming we need to communicate between two GPUs, there are many different ways to move data

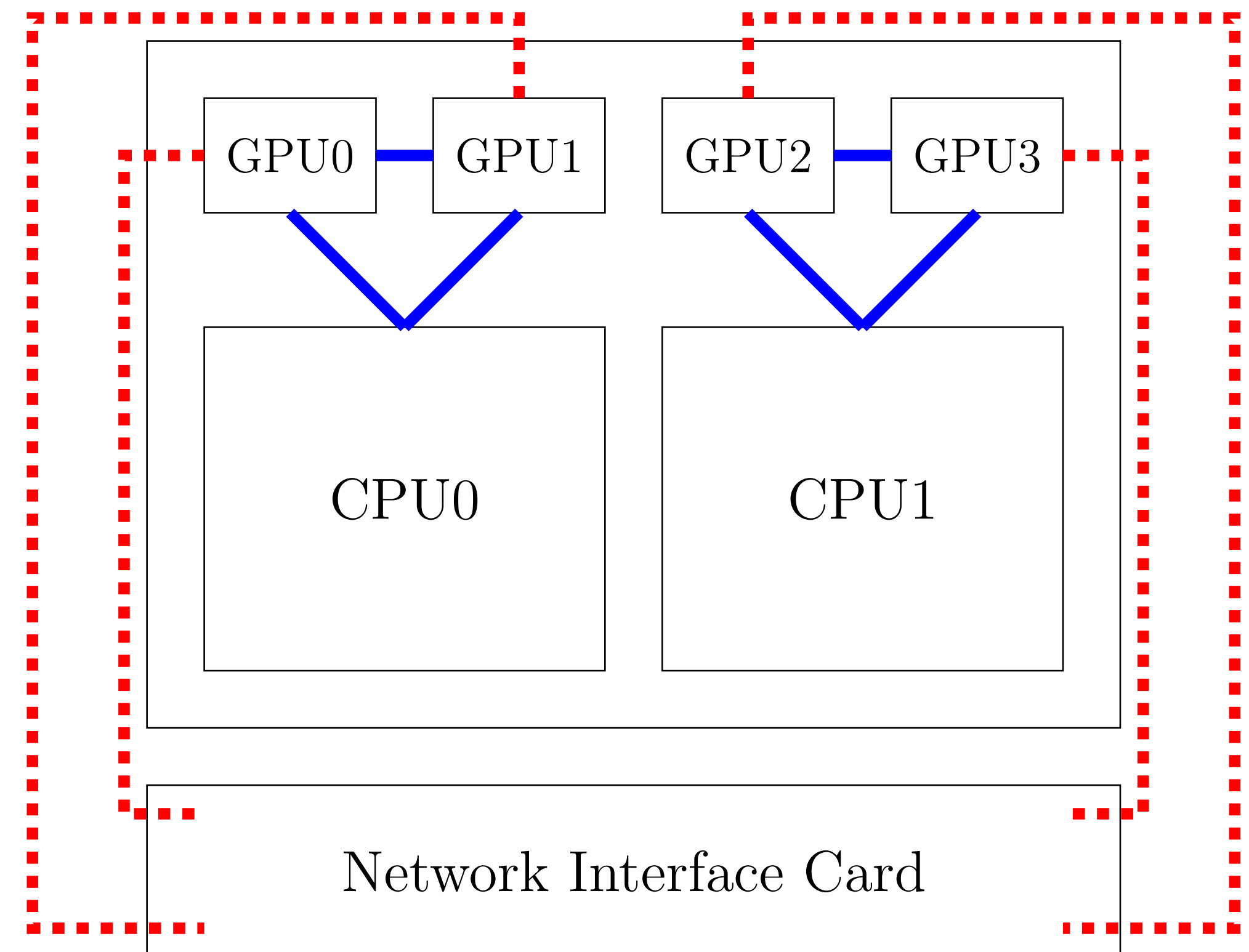
# Supercomputer Architecture

- Summit and Lassen (Power9) :
  - IBM + NVIDIA (SpectrumMPI)
  - 4-6 NVIDIA Tesla V100 GPUs per node
  - 2 22-Core Power9 CPUs per node

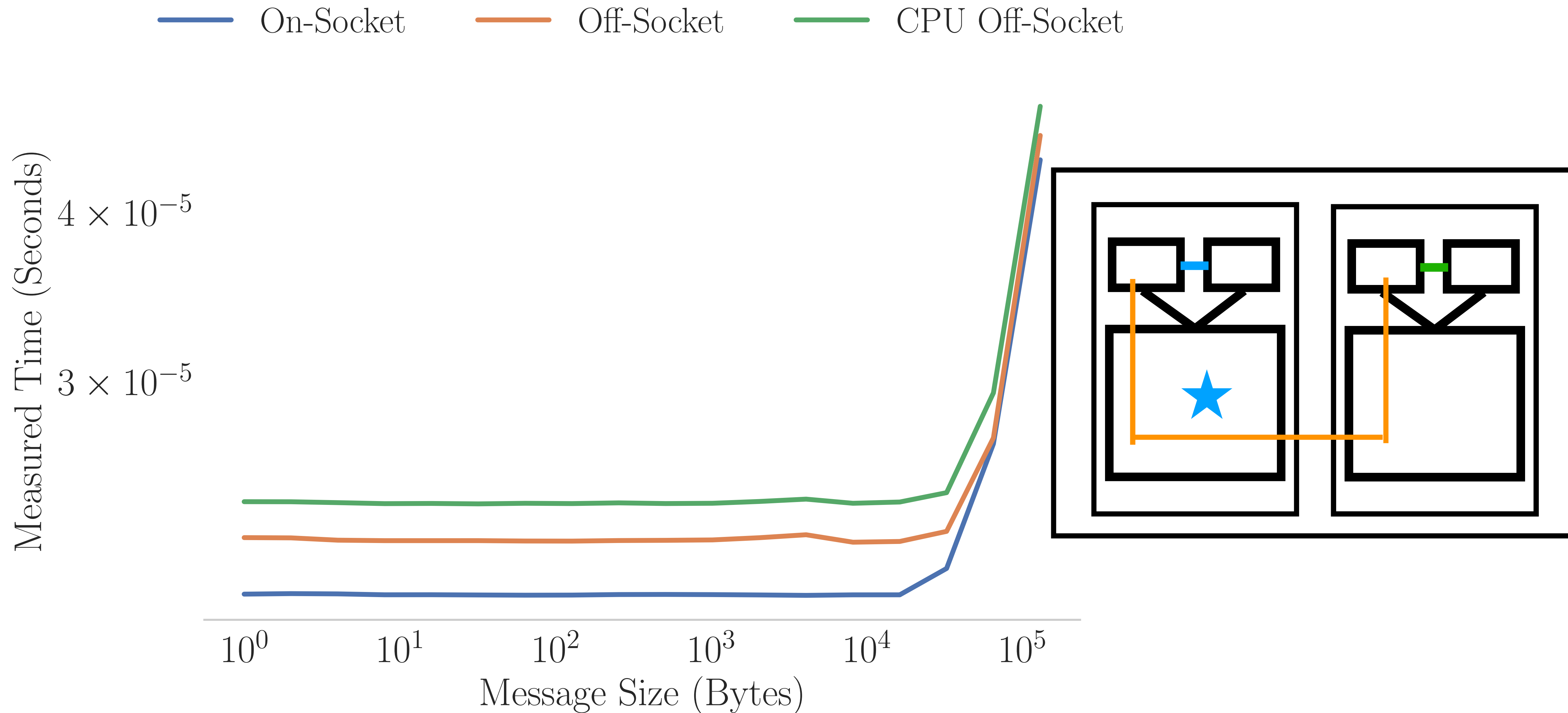


# Supercomputer Architecture

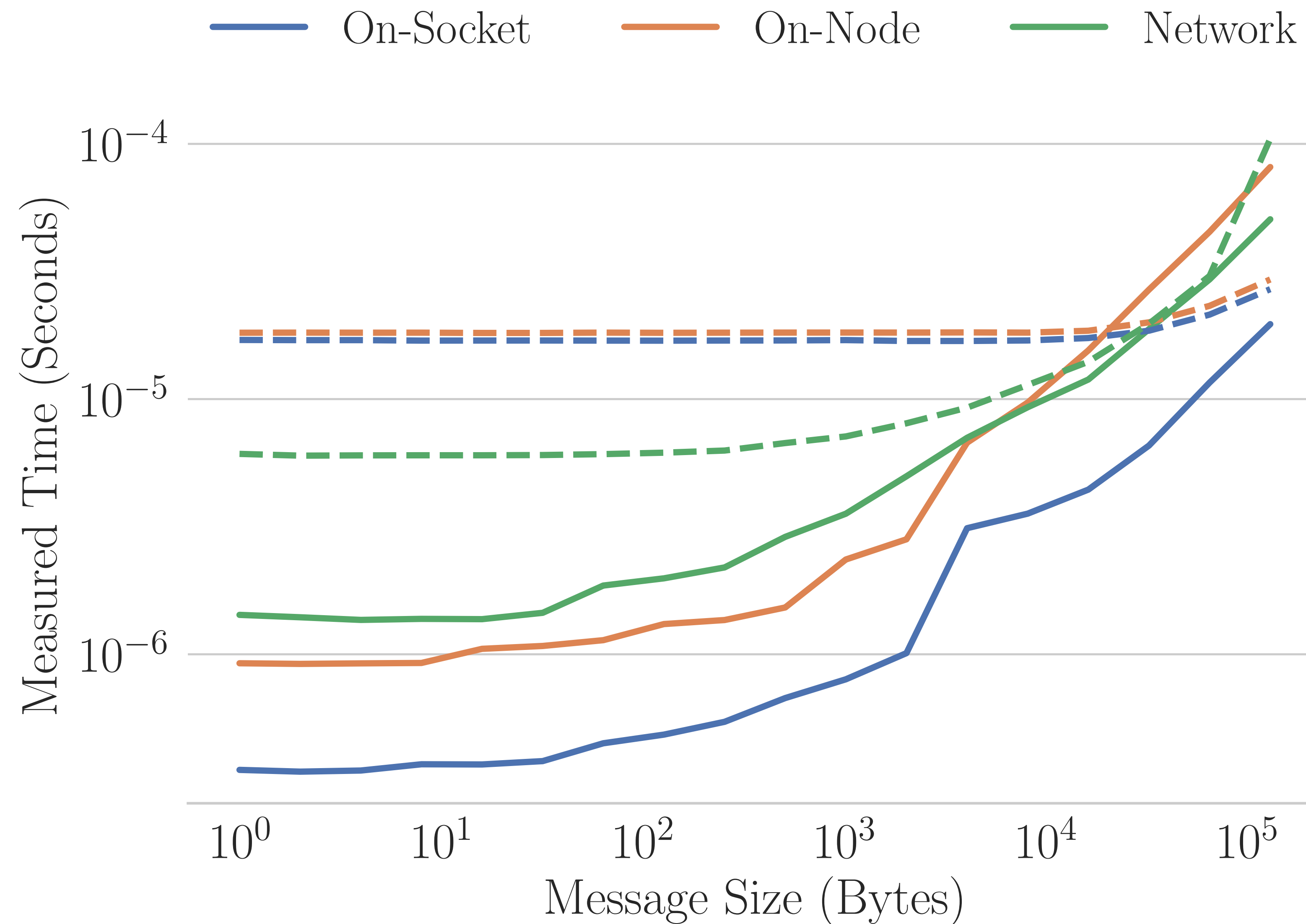
- All GPUs on a socket are connected by physical links (can communicate between them without copying data to the CPU)
- All GPUs are connected directly to the NIC, supporting GPUDirect communication



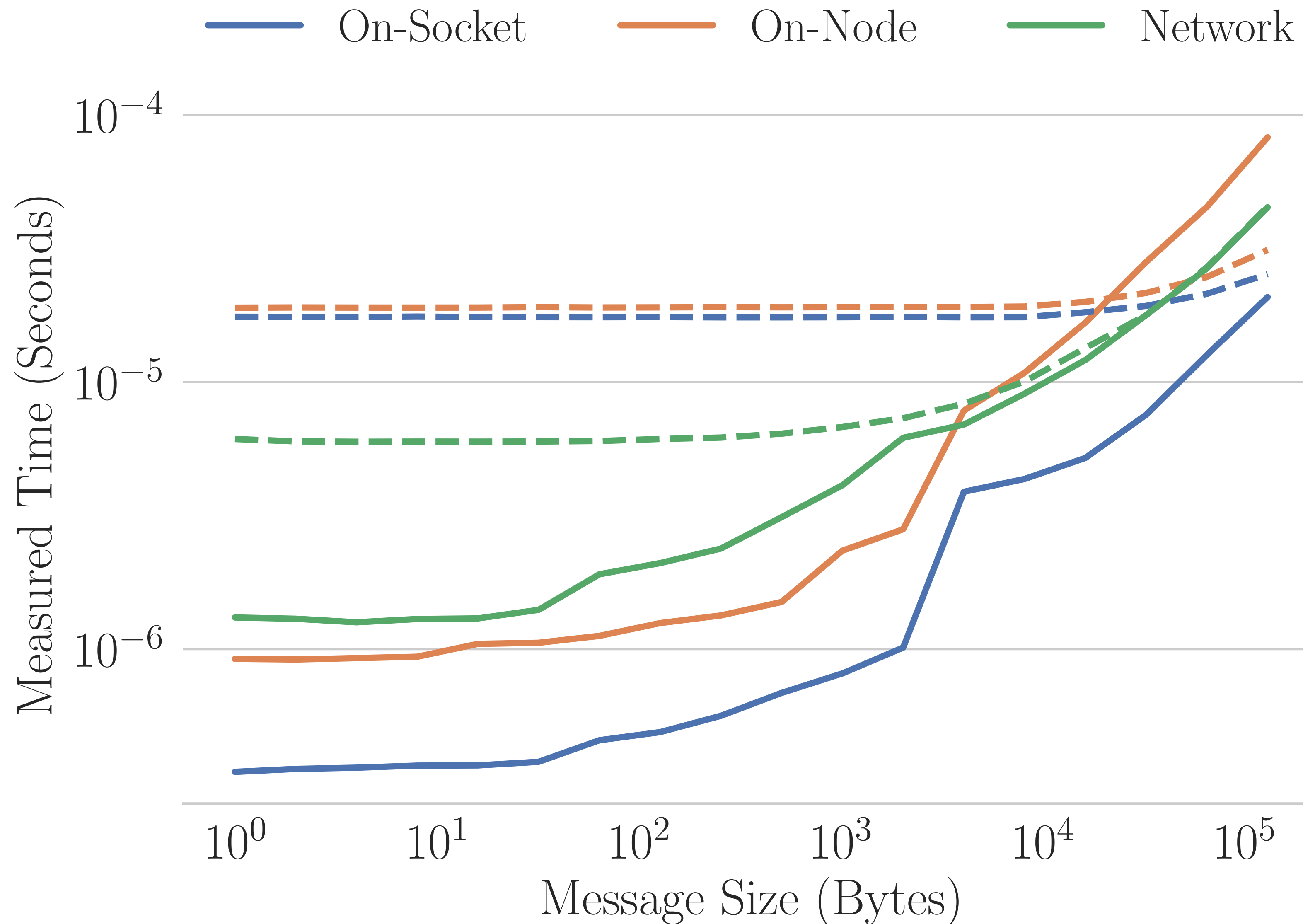
# Summit : on-node inter-GPU memcpy



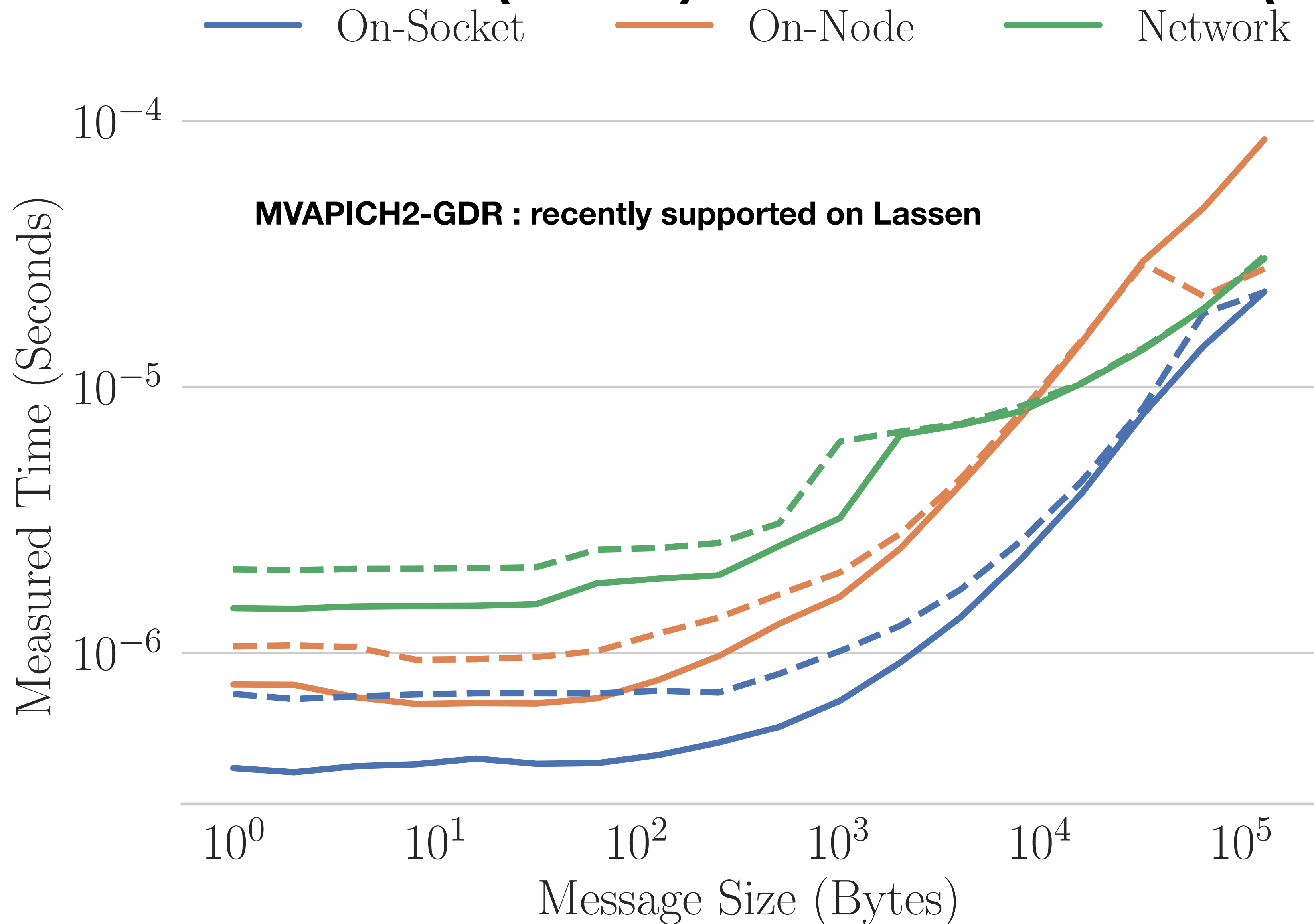
# Summit : inter-CPU (solid) and inter-GPU (dotted)



# Lassen : inter-CPU (solid) and inter-GPU (dotted)

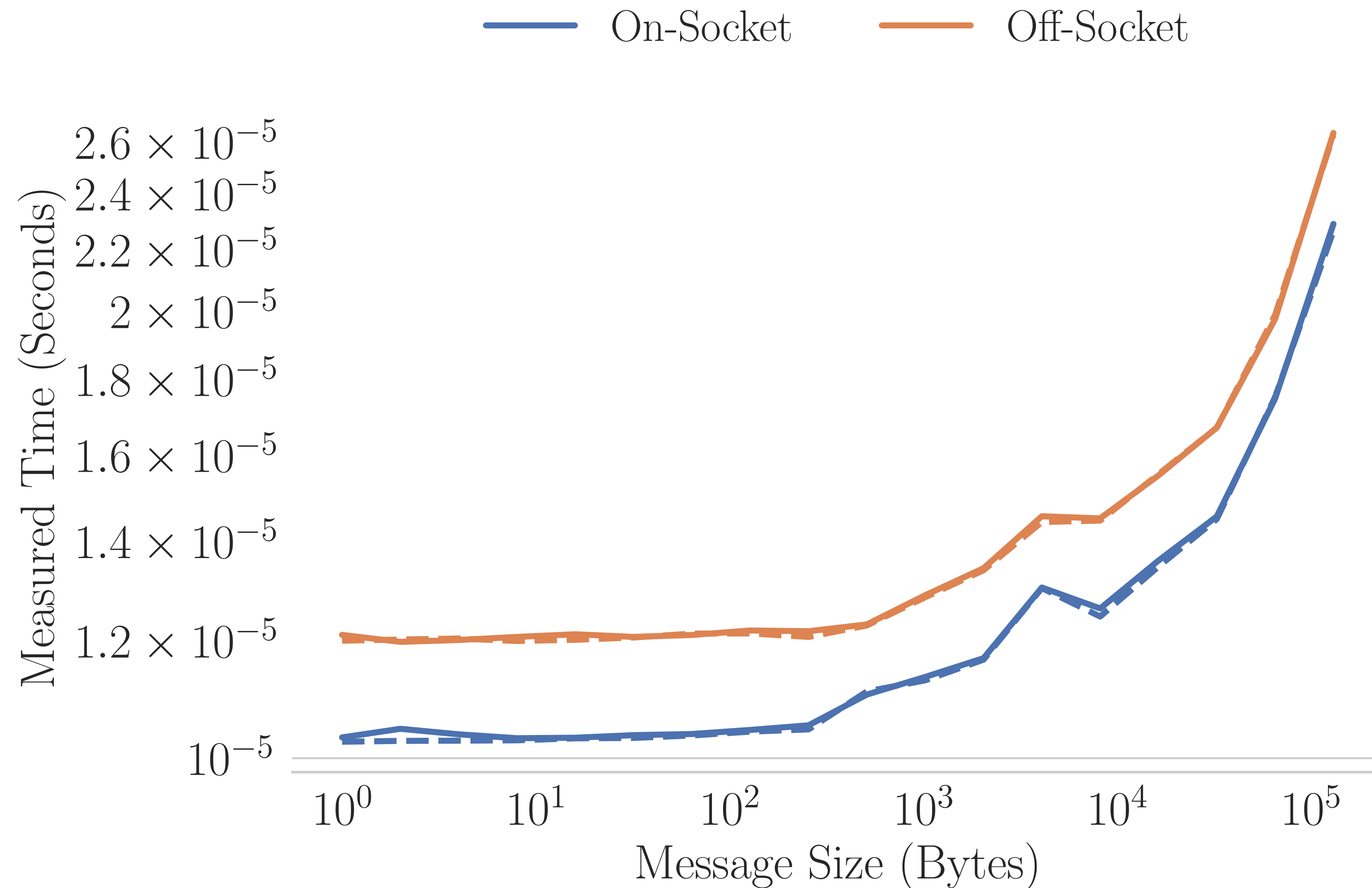


# Lassen : inter-CPU (solid) and inter-GPU (dotted)

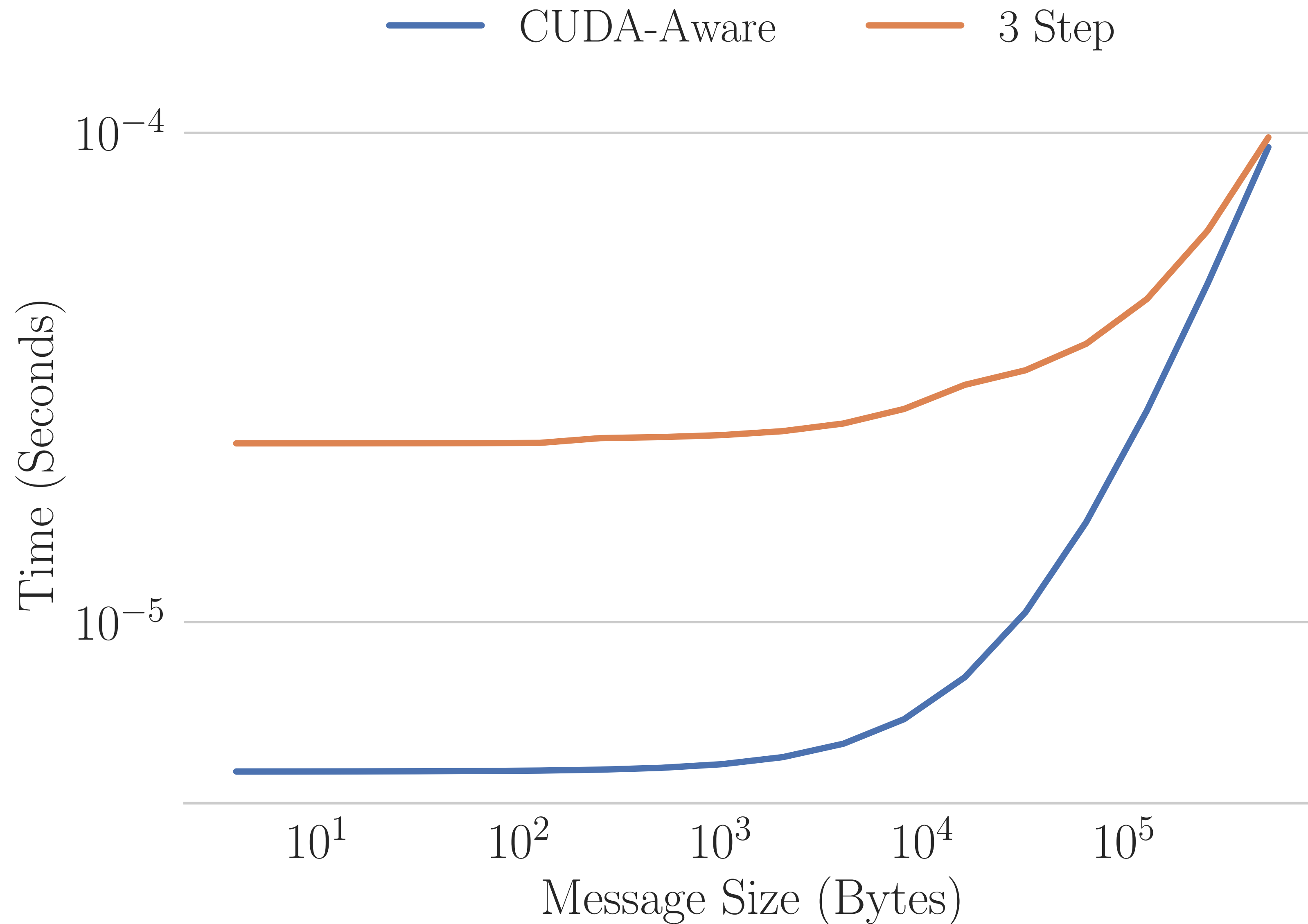




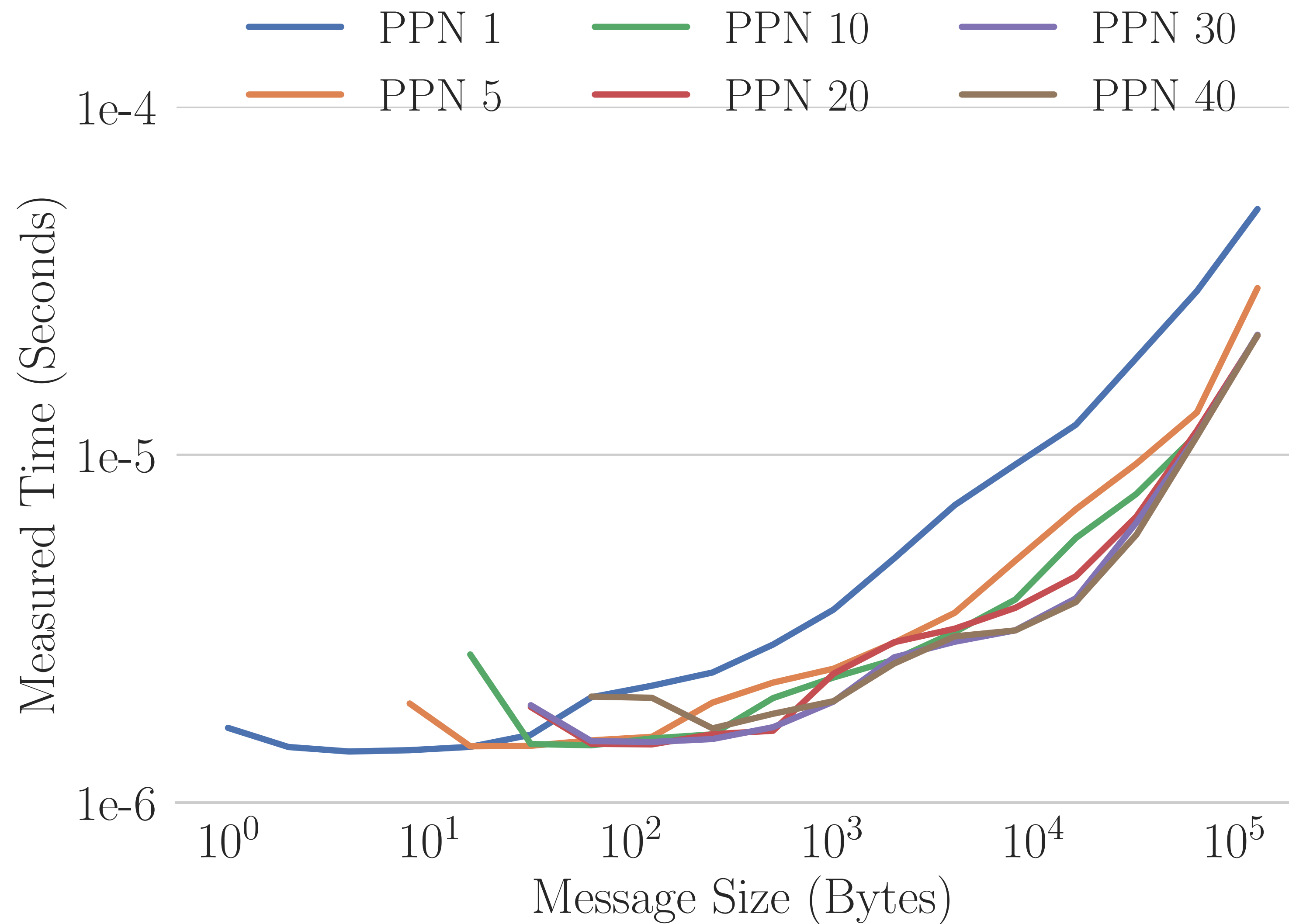
# Summit : Memcpy



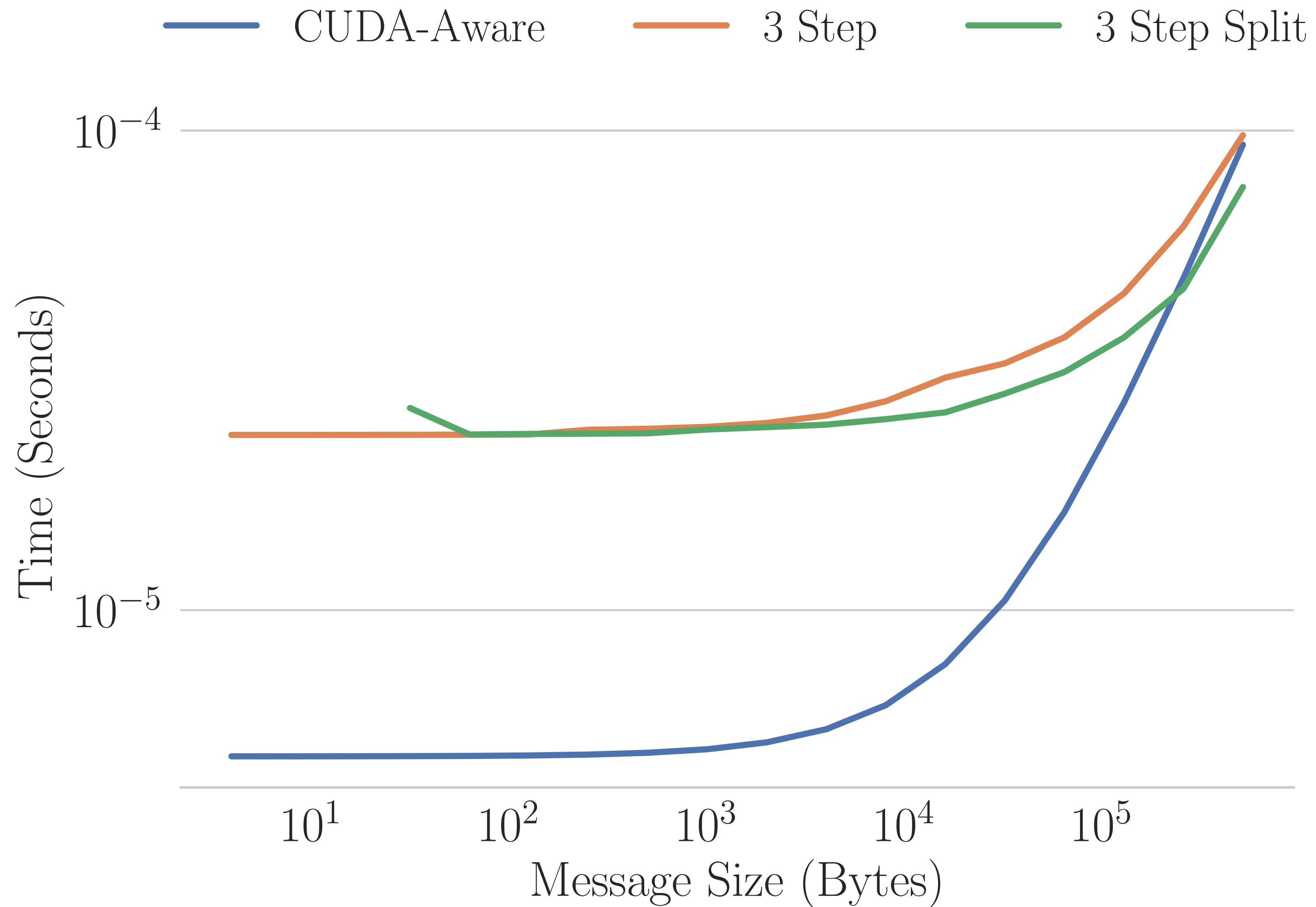
# Summit : Copy to CPU vs GPUDirect



# Summit : Inter-CPU PPN



# Summit : Copy to CPU vs GPUDirect



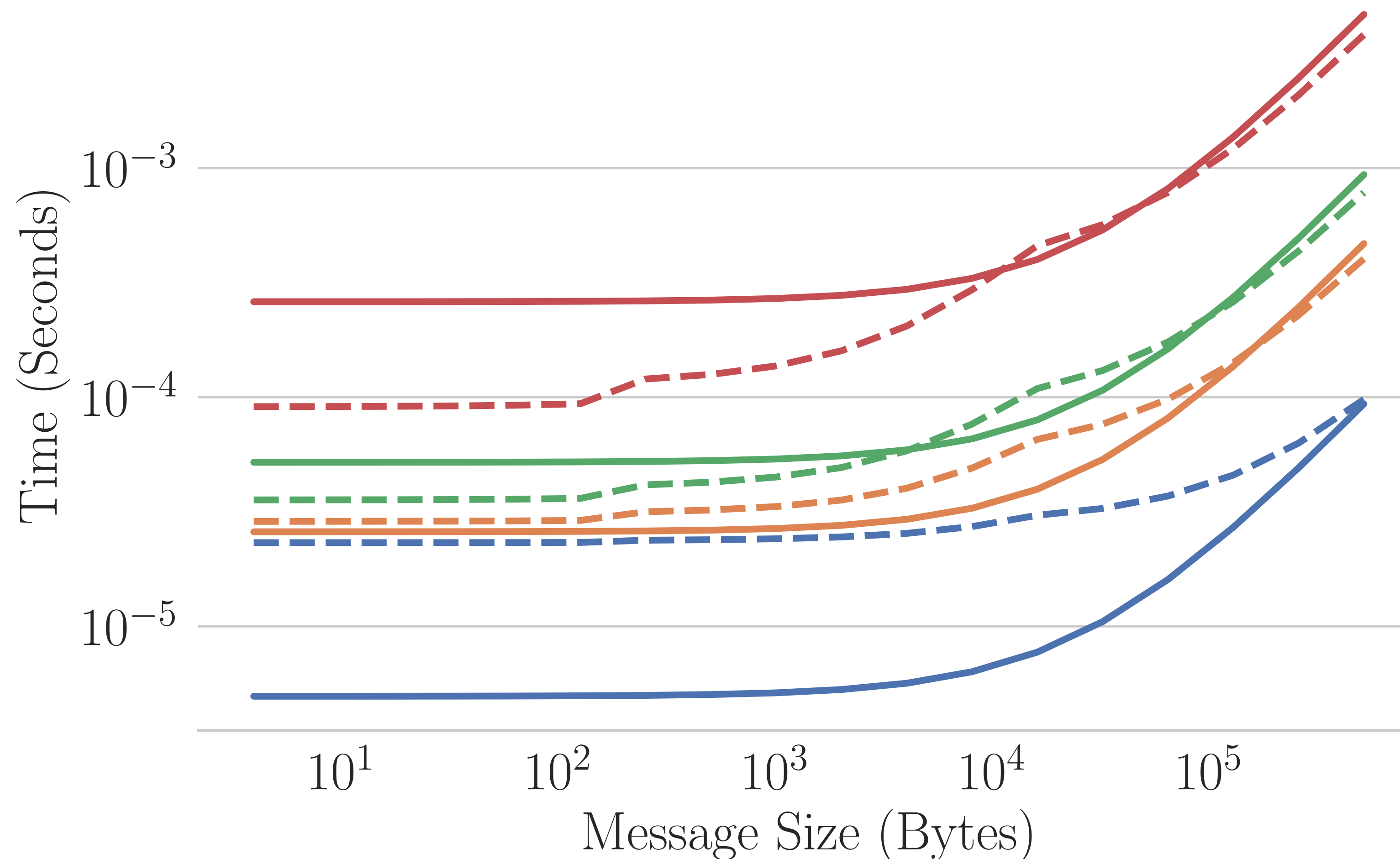
# GPUDirect vs Copy To CPU

- When sending multiple messages :
  - GPU Direct will send one message at a time between GPUs
  - Copy to CPU : only need to copy all data to CPU once, then send all inter-CPU messages, and finally perform a single copy back to GPU
- Inter-CPU messages are much cheaper than inter-GPU
- At some point, it will be cheaper to copy to the CPU, rather than use GPUDirect

# Summit : Multiple Messages

**GPUDirect : solid lines**  
**Copy to CPU : dotted lines**

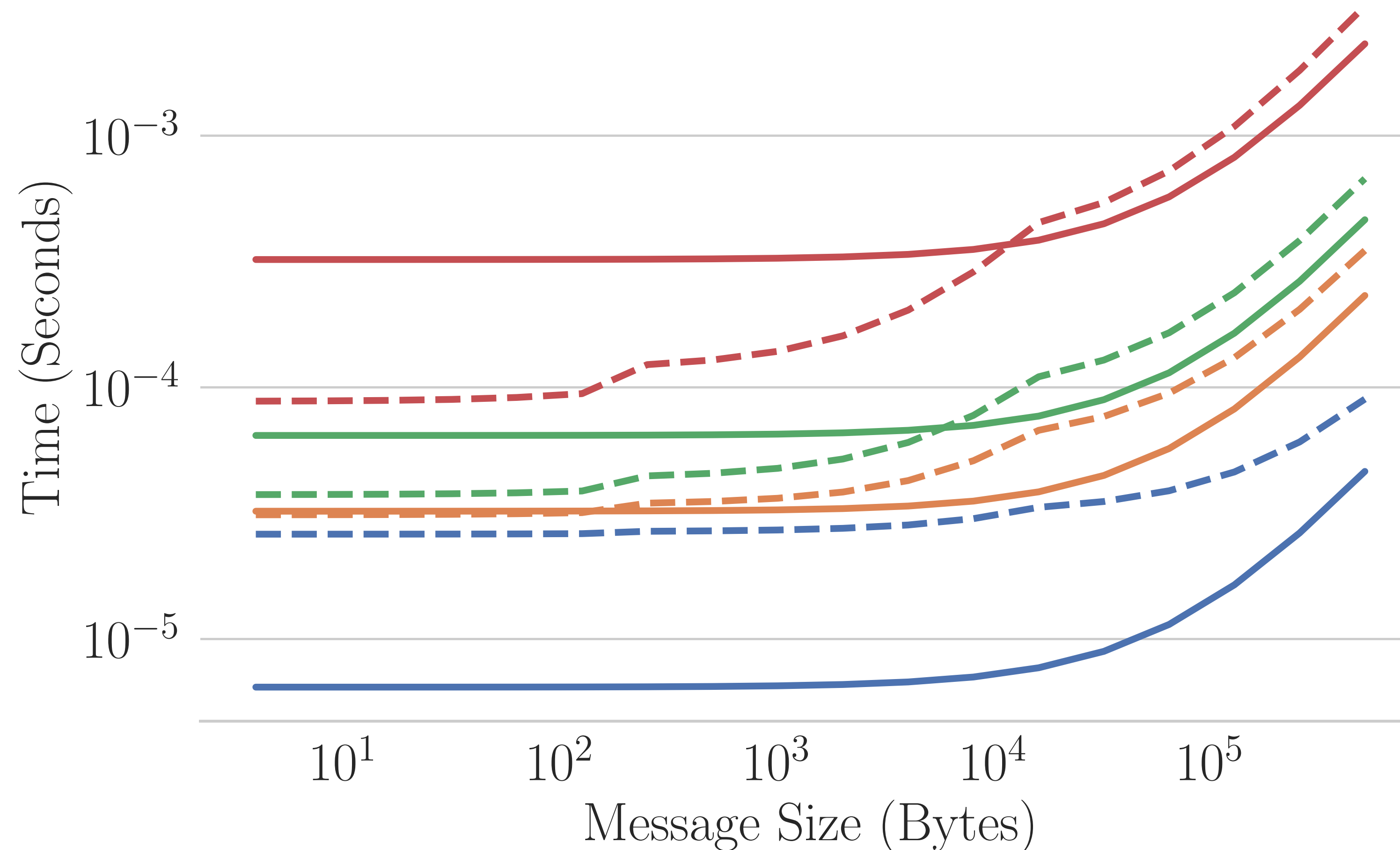
1 Msgs      10 Msgs  
5 Msgs      50 Msgs



# Lassen : Multiple Messages

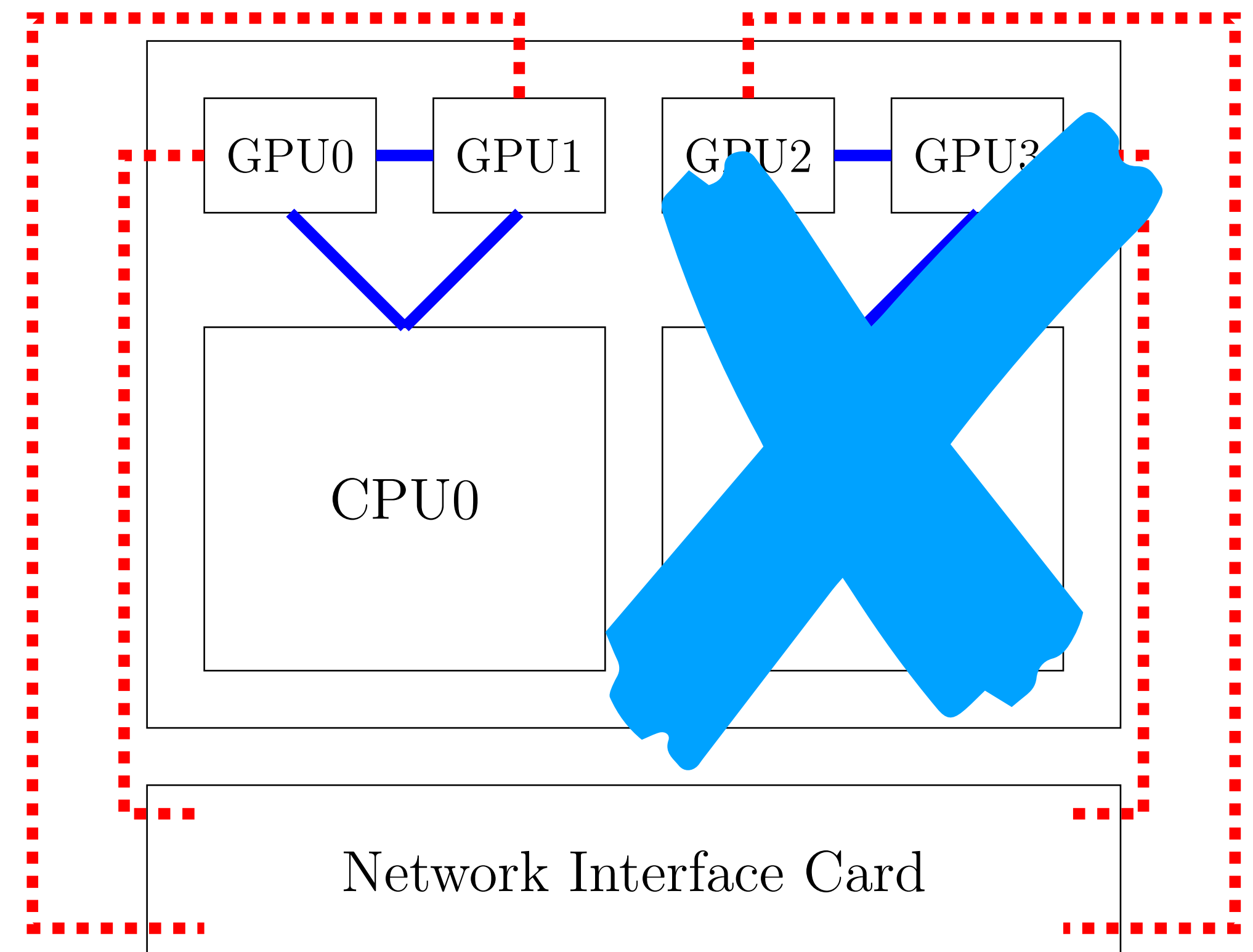
**GPUDirect : solid lines**  
**Copy to CPU : dotted lines**

1 Msgs      10 Msgs  
5 Msgs      50 Msgs



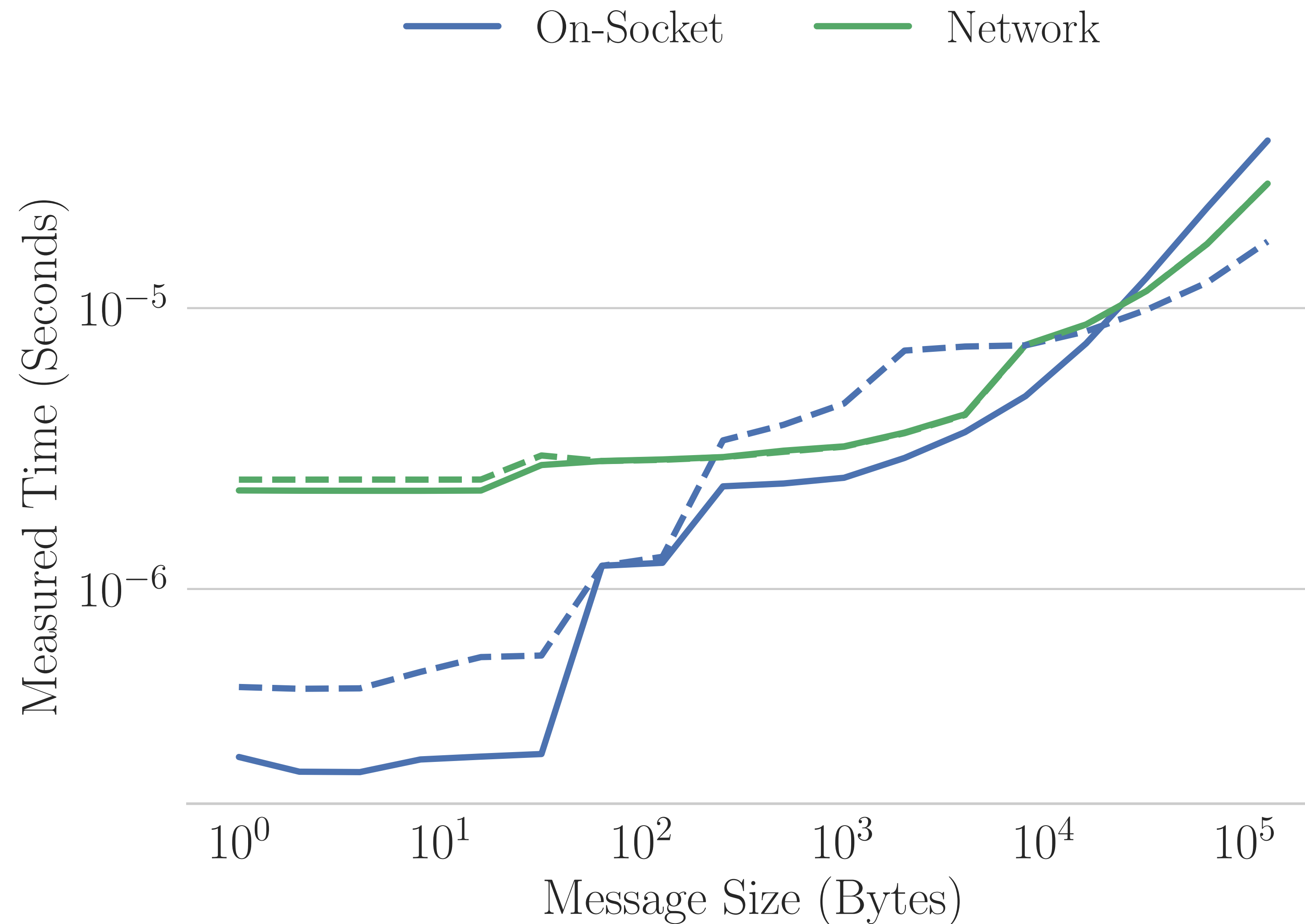
# Supercomputer Architecture

- pre-Frontier System:
  - Cray + AMD (CrayMPI)
  - AMD MI-250X GPUs
  - 1 64-core AMD Trento CPU per node

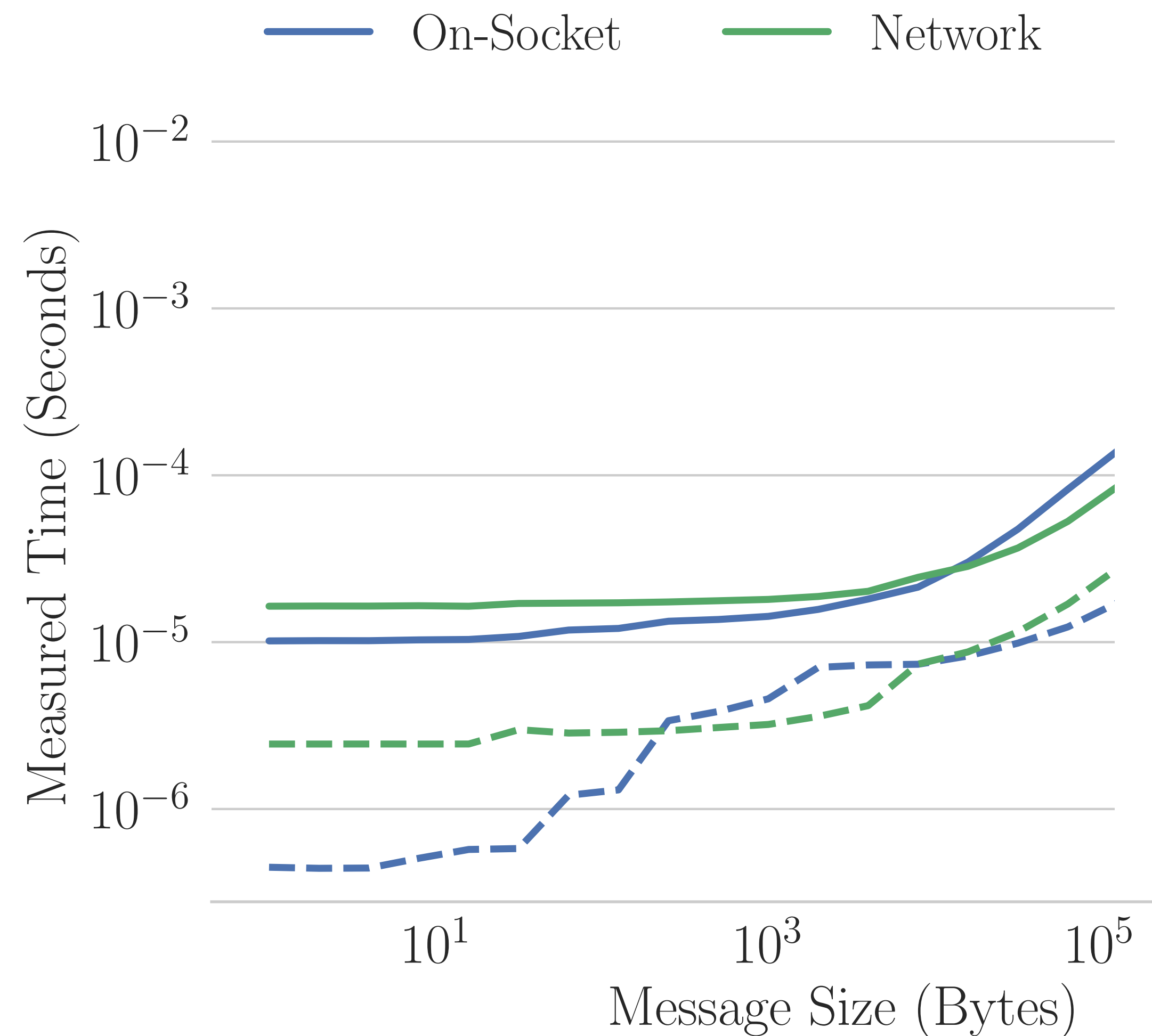




# pre-Frontier : inter-CPU (solid) and inter-GPU (dotted)



# pre-Frontier : copy to CPU (solid) and GPU-Aware (dotted)



# Next Class

- We will discuss algorithms and collectives on current and future heterogeneous systems
  - How are they currently implemented
  - How can they be improved
  - Current research directions