SAGE

Streaming Programming Systems for Exascale

Stefano Markidis, Ivy Bo Peng, Erwin Laure

KTH, Stockholm, Sweden

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671500

Exascale computing systems:

- **distributed memory nodes** with a large number of (possibly heterogeneous) compute cores.
- **parallel threads of execution** split between multiple layers of hardware parallelism.
- **memory architectures** highly fragmented.

Exascale programming system is an API that can address all these parallel threads across all these hardware layers and memory spaces with **maximum performance** + **highly productive**

With the first Exascale hardware in 5-10 years, Exascale applications will make use of combinations of existing programming APIs:

- specific to one or two layers of hardware parallelism.
- well standardized to ensure portability and sustainability

Realistically, MPI can be good candidate for exascale programming systems:

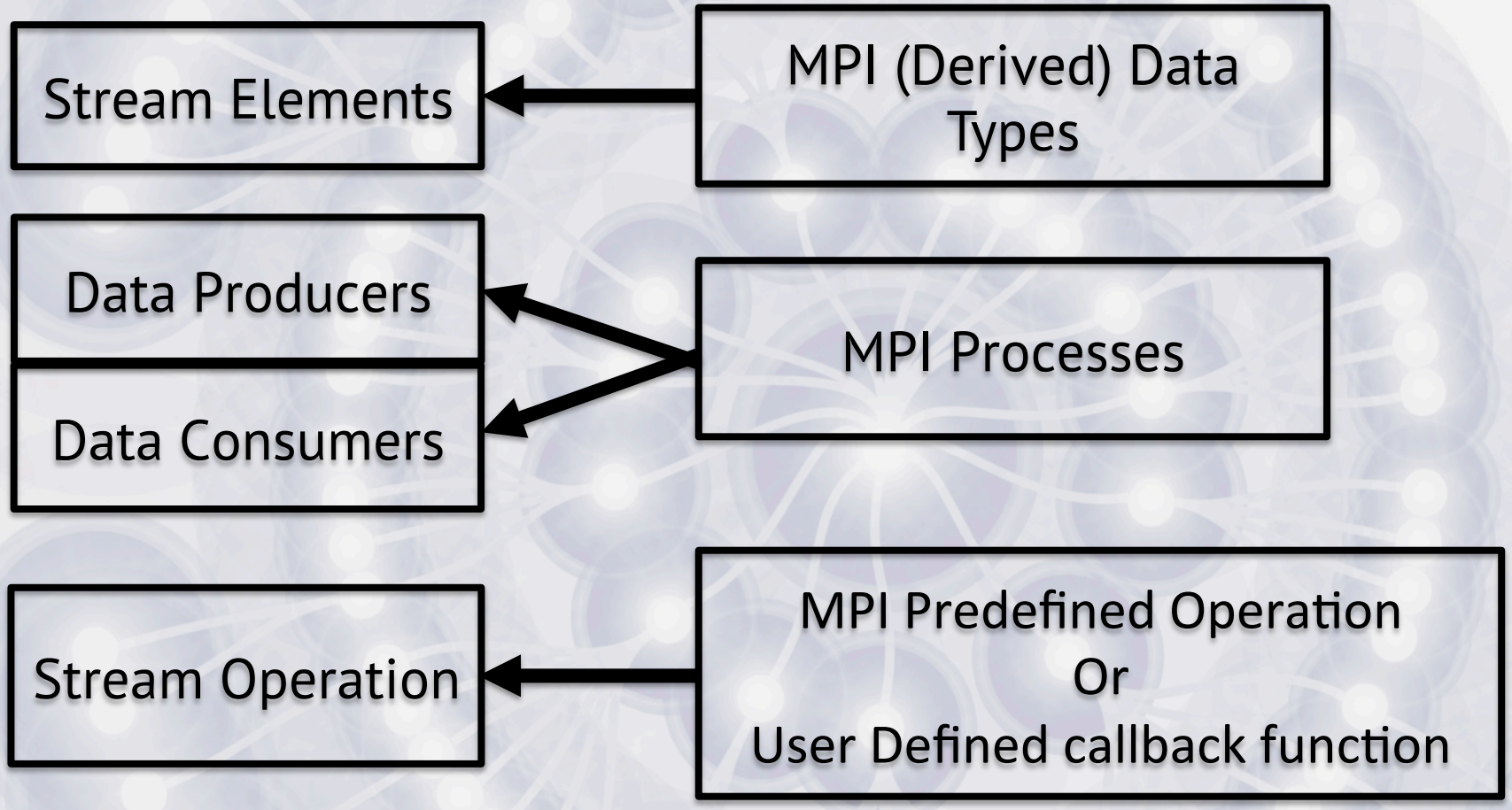
- Largely adopted by HPC community and it allows for incremental porting of applications to new computing models
- Standardized

However, MPI was not designed for emerging computing paradigms, such as streaming model.

The goal of this work is to investigate the possibility of enhancing MPI with streaming capabilities.

Motivations:

- Data-centric problems (that map conveniently to streaming models) are becoming more and more common in HPC.
- Streaming model in MPI can boost the adoption of streaming paradigms in HPC community.
- Streaming frameworks using MPI underneath can benefit from streaming model support in MPI.

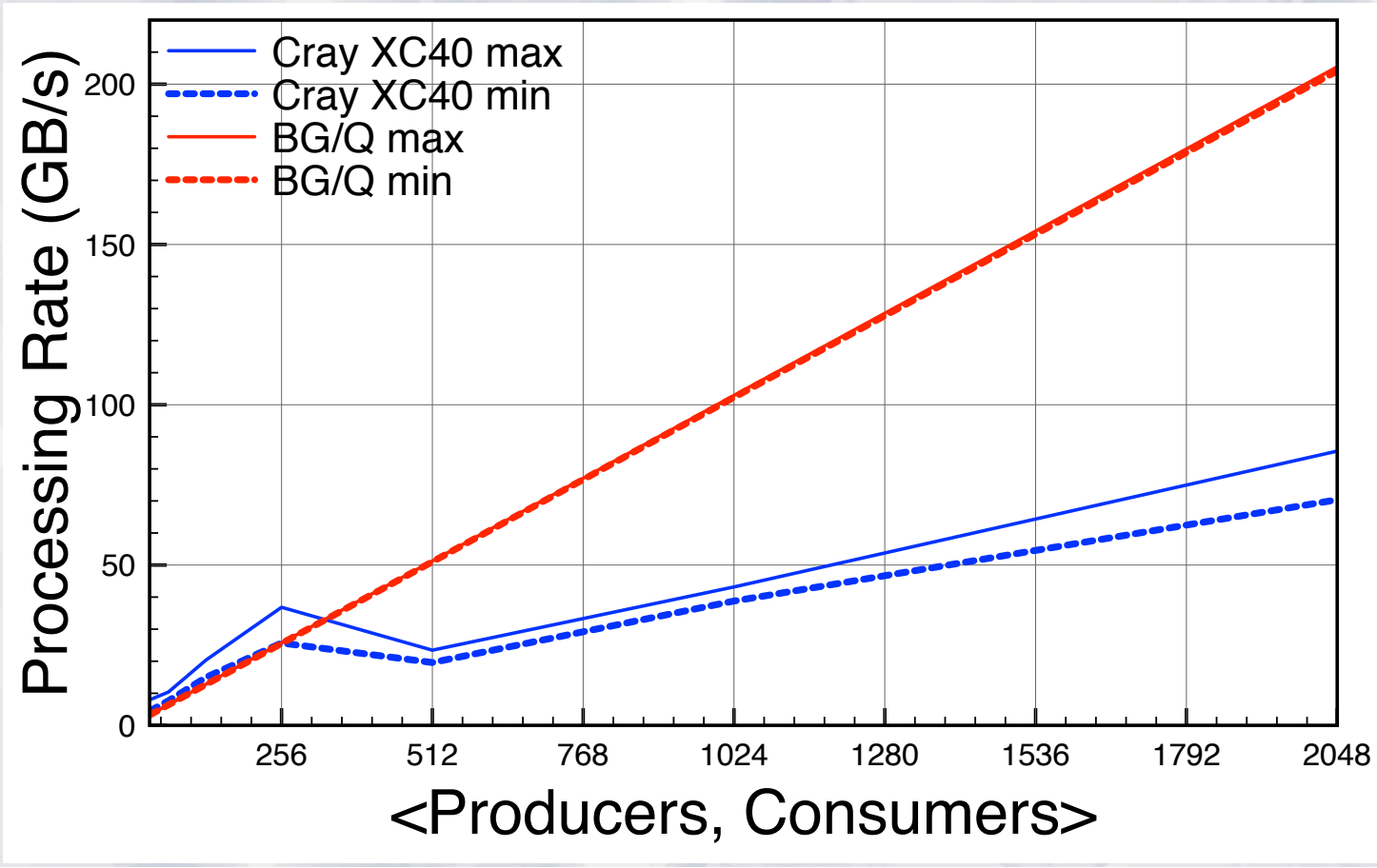


SAGE

Mapping MPI to Stream Concepts

- Written in C on the top of MPI
- Set-up function: `MPIStream_CreateChannel(),MPIStream_Attach()`
- Data Producers main functions:
  - `MPIStream_Send/MPIStream_Isend()`
  - `MPIStream_Terminate()`
  - `MPIStream_Query()`
- Data consumers main functions:
  - `MPIStream_Operate/MPIStream_Ioperate()`

Ivy Bo Peng, S. Markidis, E. Laure, D. Holmes, and M. Bull, *A Data Streaming Model in MPI*, ExaMPI 2015 Workshop in SC'15

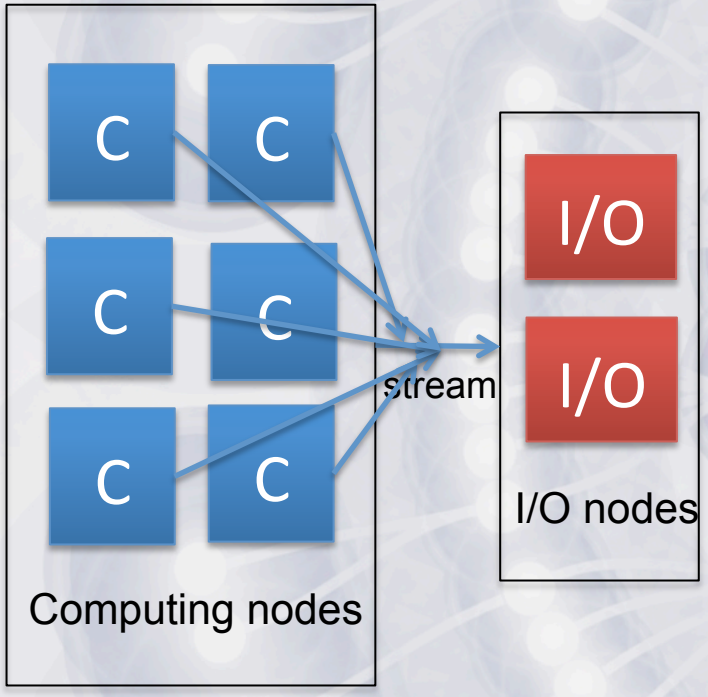


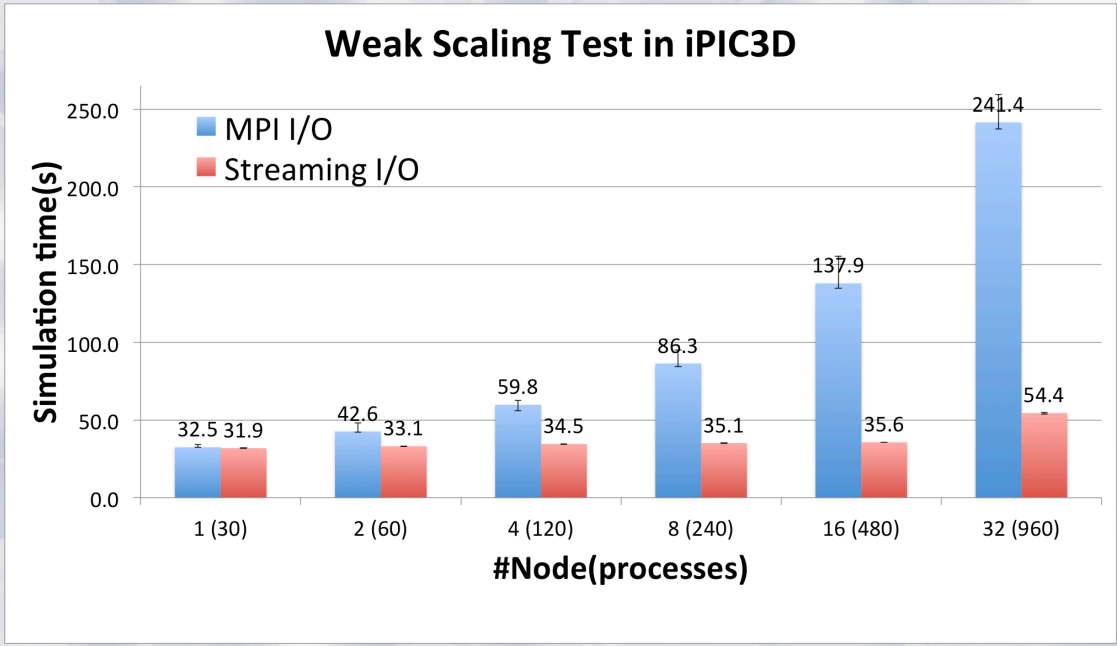
Processing rate in GB/s varying the number of data producers and consumers using scale benchmark on the Cray XC40 and Blue Gene/Q supercomputers

SAGE

STREAM Benchmark Results

Stream out irregular and fine-grained data for parallel I/O





SAGE

HPC Use Case: Parallel I/O

- MPI likely to be an exascale programming system
- We investigated the possibility of enhancing MPI with streaming computing capabilities
- Traditional HPC MPI-based applications can benefit from streaming computing as a mean to off-load I/O, data analytics to dedicated data consumer processes.
- Streaming model in MPI can boost the adoption of streaming paradigms in HPC community.
- Streaming framework using MPI underneath can benefit from streaming model in MPI.

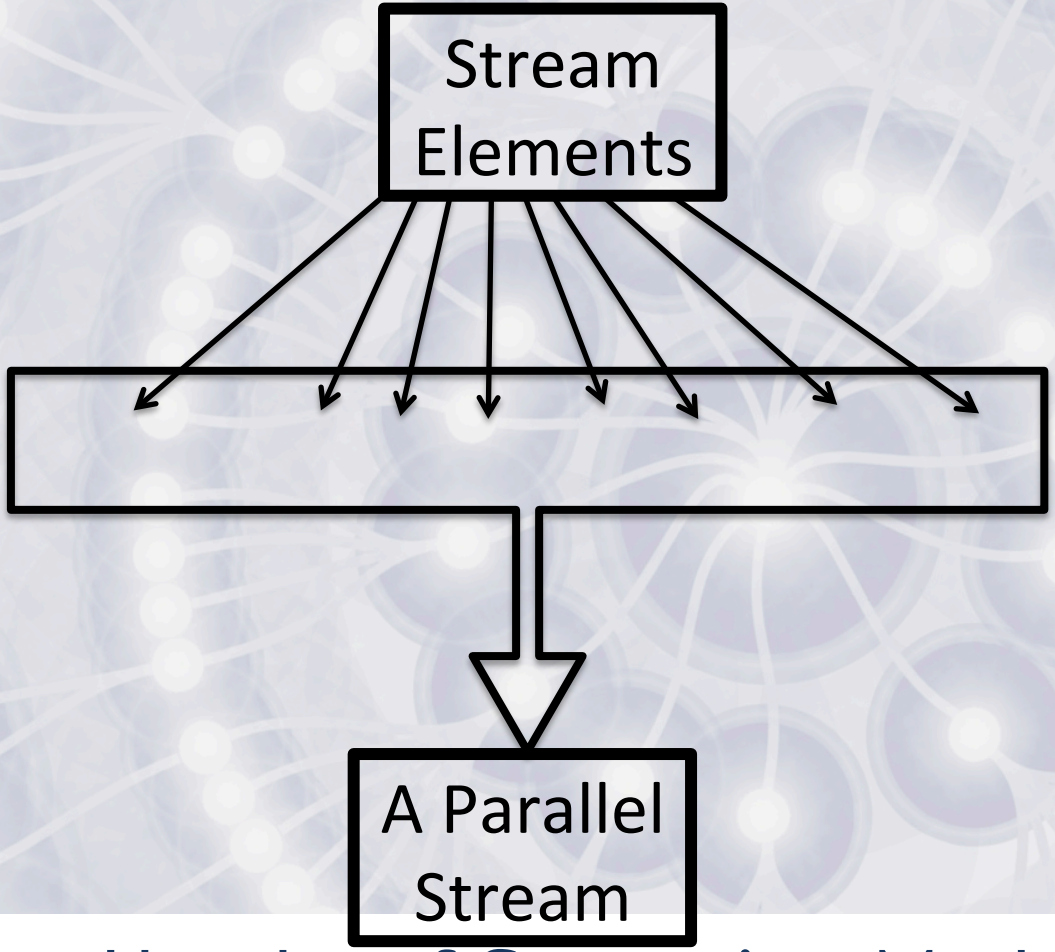
SAGE

Conclusions

Additional Slides

- A Parallel Stream is a a sequence of Stream Elements:
  - A parallel Stream is divided over a group of producers
  - Data producers perform HPC applications and stream out elements occasionally
  - Stream Element is the unit of transmission
- One **Operation** is attached to one Parallel Stream:
  - Performed by Data Consumers on first-come-first-serve basis
  - Consists of one **Intermediate** Operation (per stream element) and one **Terminal** Operation (all stream elements)

A Streaming Model in MPI



SAGE

Uptake of Streaming Models in HPC