UPPSALA
UNIVERSITET



# Distributed Computing - MPI

Dimitar Lukarski, Jarmo Rantakokko

Division of Scientific Computing
Department of Information Technology
Uppsala Programming for Multicore Architectures Research Center
(UPMARC)
Uppsala University

Programming of Parallel Computers, Jan, 2014



UPPSALA
UNIVERSITET

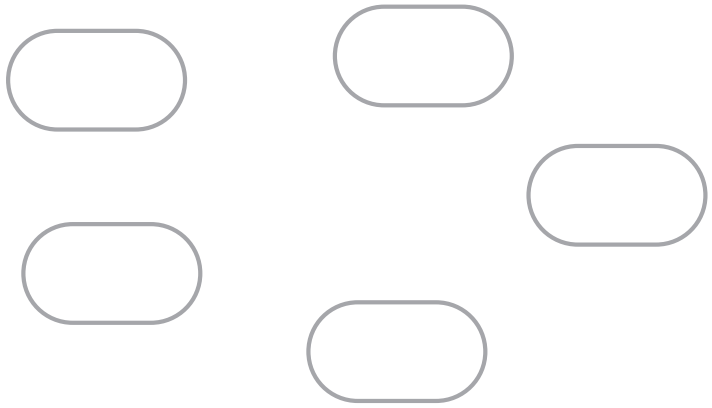
Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

## Outline

UPPSALA
UNIVERSITET



# Distributed computing



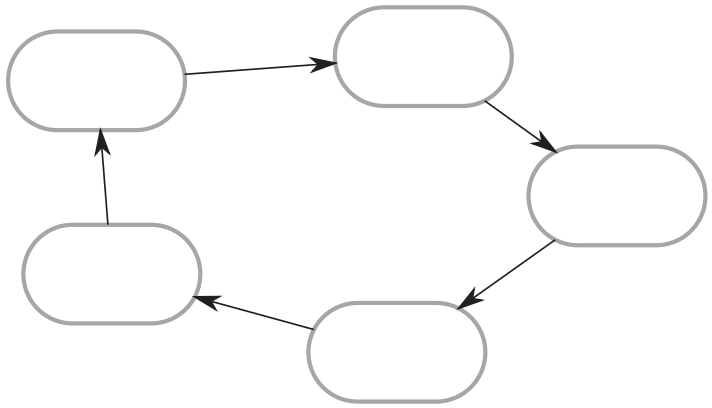




UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

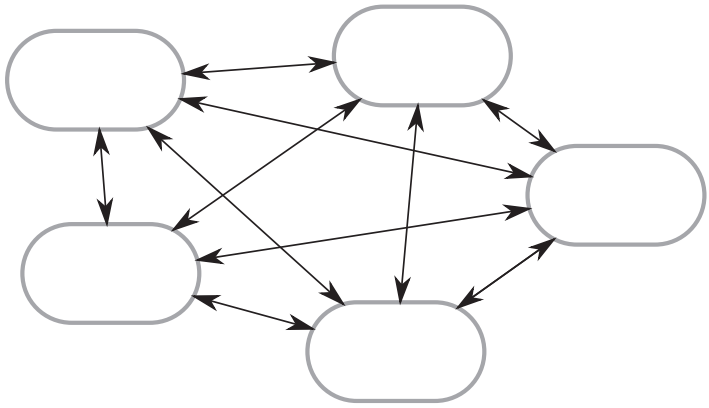
# Distributed computing



UPPSALA
UNIVERSITET



# Distributed computing



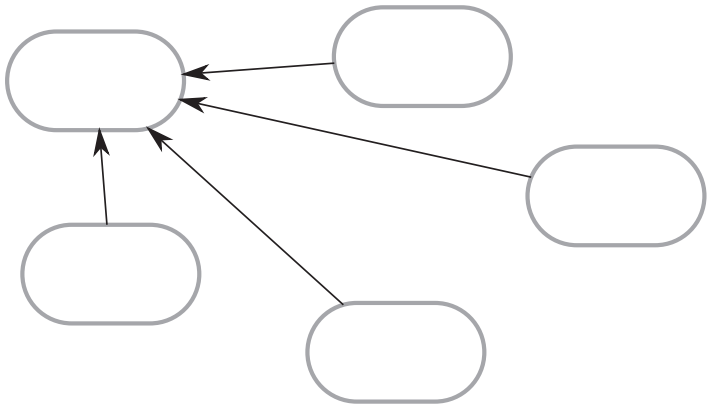




UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# Distributed computing



UPPSALA
UNIVERSITET



## Distributed computing

You know how to communicate with MPI!

- Communicators
- Sending/Receiving data
- Synchronous/Blocking modes
- Other Point-to-Point Functions
- Global for synchronization
- Global for communication



UPPSALA
UNIVERSITET

UPPSALA
UNIVERSITET



# Problems with Send/Recv

What issues are expected when using Send and Recv functions?





UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# Problems with Send/Recv

What issues are expected when using Send and Recv functions?

- Send/Recv match
- Handshaking overhead
- Extra buffers

UPPSALA
UNIVERSITET



# One-sided Communication

- Direct access to the remote memory (RMA)
- No hand-shaking procedure
- Flexible and dynamic data distribution
- Hardware optimization
- No extra buffers





UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

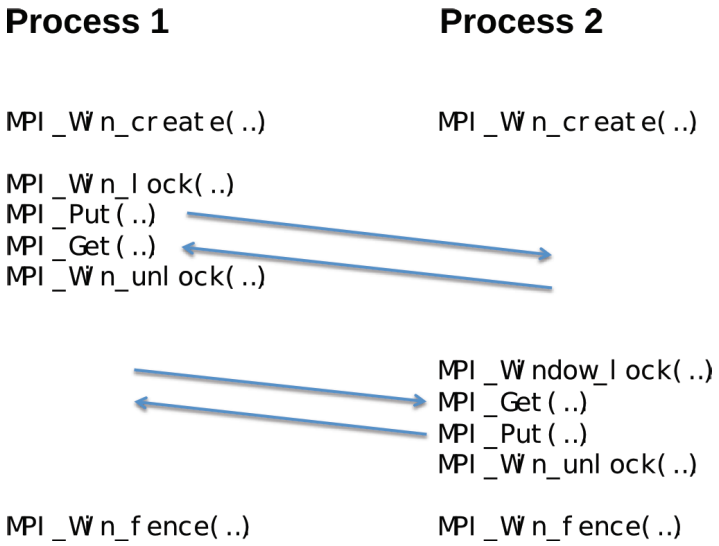
# One-sided Communication

- Create accessing window
- Accessing pattern
- Sync (fence)
- Reduction (MPI_Accumulate())
- Lighter syncs

UPPSALA
UNIVERSITET



## One-sided Communication

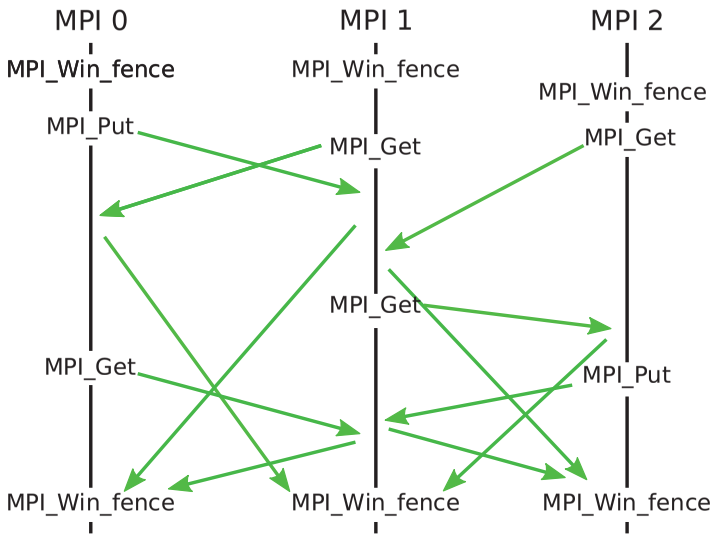





UPPSALA
UNIVERSITET



## One-sided Communication



UPPSALA
UNIVERSITET



## Access Epoch

- Lighter/local than fence
- Epoch: time-frame when a local window may be used for RMA

Origin

- MPI_Win_start
- MPI_Win_complete

Target

- MPI_Win_post
- MPI_Win_wait



UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

UPPSALA
UNIVERSITET



# Dynamic Process Creation

A MPI process can spawn new MPI processes at run time which starts running a new program.

- MPI_Comm_spawn(...);
- MPI_Comm_spawn_multiple(...) (MPMD)



UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# Code Example

```
// Parent program
MPI_Init(...);
MPI_Comm childcomm;
MPI_Comm_spawn("childprog",arg,nproc,...,
               &childcomm);
MPI_Send(data,count,type,0,tag,childcomm);
...
MPI_Recv(data,count,type,0,tag,childcomm,status);
MPI_Comm_free(&childcomm);
MPI_Finalize();
```

UPPSALA
UNIVERSITET



## Code Example

```
// Child program (started by the parent program)
MPI_Init(...);
MPI_Comm parentcomm;
MPI_Get_parent(&parentcomm);
MPI_Get_rank(MPI_COMM_WORLD,&rank);
If (rank==0) MPI_Recv(data,count,type,0,tag,
             parentcomm,status);
...
If (rank==0) MPI_Send(data,count,type,0,tag,
                      parentcomm);
MPI_Finalize();
```

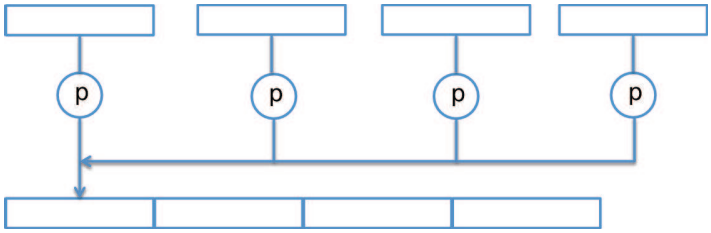


UPPSALA
UNIVERSITET

UPPSALA
UNIVERSITET



# I/O – Approach 1








UPPSALA
UNIVERSITET



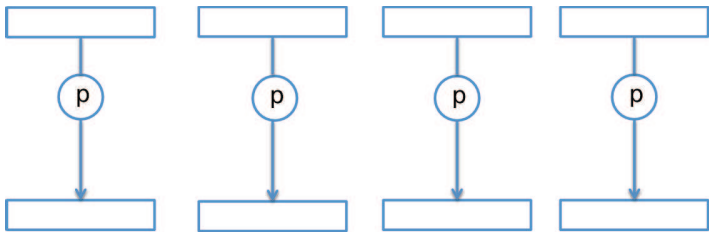
# I/O – Approach 1

Gather all data to process 0 and process 0 writes to file.
Serial operation, lacks scalability and performance.

UPPSALA
UNIVERSITET



# I/O – Approach 2








UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

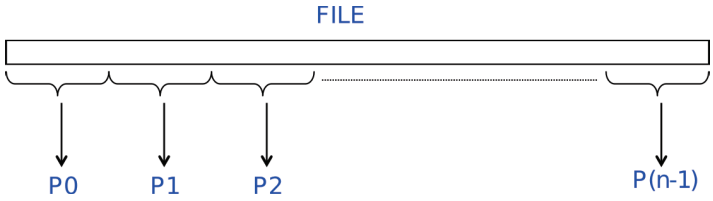
# I/O – Approach 2

Each process writes to a separate file. Good performance but lots of small files to manage. Not a practical solution, what if the number of MPI tasks changes?

UPPSALA
UNIVERSITET



# I/O – Approach 3






UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# I/O – Approach 3

MPI allows for several processes to read and write to the same file in parallel

- MPI_File_open(...)
- MPI_File_read(...)
- MPI_File_write(...)
- etc (more than 60 functions)

Portable, scalable and efficient solution!

UPPSALA
UNIVERSITET





UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# PDE

## 1D Heat problem

$$\begin{aligned} \partial_t u - \nu \partial_x^2 u &= 0 && \text{in } (0,T] \times (0,1) \\ u(0,\cdot) &= u_0(\cdot) && \text{in } (0,1) \quad \text{(Initial condition)} \\ u(\cdot,0) = u(\cdot,1) &= 0 && \text{in } (0,T] \quad \text{(Boundary condition)} \end{aligned}$$

where $\nu > 0$.

Solve it in numerically and in parallel!

UPPSALA
UNIVERSITET



# Head Problem

Math



UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

# Head Problem

Math

- Check compatibility condition

UPPSALA
UNIVERSITET



# Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)



UPPSALA
UNIVERSITET



# Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)

UPPSALA
UNIVERSITET



## Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)
- Discretize the problem (FD)



UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

## Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)
- Discretize the problem (FD)
- Explicit method (no LS)

UPPSALA
UNIVERSITET



## Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)
- Discretize the problem (FD)
- Explicit method (no LS)
- Check Stability condition



UPPSALA
UNIVERSITET

Outline
Recap
One-sided Comm
Dynamic
I/O
PDE

## Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)
- Discretize the problem (FD)
- Explicit method (no LS)
- Check Stability condition $2\nu\tau \leq h^2$

UPPSALA
UNIVERSITET



# Head Problem

Math

- Check compatibility condition
- Create space grid (def space step)
- Create time grid (def time step)
- Discretize the problem (FD)
- Explicit method (no LS)
- Check Stability condition $2\nu\tau \leq h^2$

Black-board writting...