

# MPI: Message Passing Interface

Arquitectura de Computadores II  
Ingeniería en Informática (4º Curso)  
Universidad Carlos III de Madrid

# Definición

- MPI es una interfaz de paso de mensaje que representa un esfuerzo prometededor de mejorar la disponibilidad de un software altamente eficiente y portable para satisfacer las necesidades actuales en la computación de alto rendimiento a través de la definición de un estándar de paso de mensajes universal.

**William D. Gropp et al.**

# ¿Que es MPI?

- Es un estándar de ley.
- Estándar para una biblioteca de paso de mensajes.
- Implementaciones de MPI:
  - [www-unix.mcs.anl.gov/mpi/mpich/](http://www-unix.mcs.anl.gov/mpi/mpich/)
  - [www.lam-mpi.org/](http://www.lam-mpi.org/)
- Otras direcciones:
  - [www-unix.mcs.anl.gov/mpi/](http://www-unix.mcs.anl.gov/mpi/)
  - [www.mpi-forum.org/](http://www.mpi-forum.org/)

# Introducción: Enfoque práctico

- Imaginemos que queremos ejecutar el siguiente programa en 4 máquinas distintas de forma paralela:

```
int main(){  
    printf("Hola mundo\n");  
}
```

```
gcc -Wall hola.c -o hola  
./hola
```

- Una opción, es ir a las 4 máquinas, y ejecutar el programa en ellas (**POCO EFICIENTE**).

# Introducción: Enfoque práctico

- Otra opción es utilizar MPI, y desde una maquina, ejecutar el siguiente código (**MÁS EFICIENTE**)

```
Int main() {  
    MPI_Init(&argc,&argv);  
    printf("Hola mundo\n");  
    MPI_Finalize();  
}
```

```
mpicc -o hola hola.c  
mpirun -n 4 hola
```

# Generalidades

- MPI 1.2 tiene 129 funciones
- Las funciones principales de MPI son:
  - MPI\_Init
  - MPI\_Finalize
  - MPI\_Comm\_size
  - MPI\_Comm\_rank
  - MPI\_Send
  - MPI\_Recv

# Proceso

- La *unidad básica* en MPI son los **procesos**.
- Tienen espacios de memoria independientes.
- Intercambio de información por paso de mensajes.
- Introduce en concepto de comunicadores (grupo de procesos más contexto).
- Cada proceso se le asigna un identificador interno propio de MPI (rank).

# Mi primer programa con MPI

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[]) {
    int myrank, size;
    MPI_Init(&argc, &argv);
    MPI_Comm_rank(MPI_COMM_WORLD,
&myrank);
    MPI_Comm_size(MPI_COMM_WORLD, &size);
    printf("Hola soy el proceso %d de %d\n",
myrank, size);
    MPI_Finalize();
    exit(0);
}
```



# Comunicadores

- MPI agrupa los procesos implicados en una ejecución paralela en comunicadores.
- Un comunicador agrupa a procesos que pueden intercambiarse mensajes.
- El comunicador `MPI_COMM_WORLD` está creado por defecto y engloba a todos los procesos.

# MPI\_Init, MPI\_Finalize

## **int MPI\_Init(int \*argc, char \*\*argv)**

- Primera llamada de cada uno de los procesos MPI
- Establece entorno
- Un proceso solo puede hacer una llamada MPI\_INIT

## **int MPI\_Finalize(void)**

- Termina la ejecución en MPI
- Libera los recursos utilizados por MPI

# Identificación de procesos

- **MPI\_Comm\_rank (comm, &pid);**
  - Devuelve en *pid* el identificador del proceso dentro del comunicador comm especificado
  - (ej.: MPI\_COMM\_WORLD).
- **MPI\_Comm\_size (comm, &npr);**
  - Devuelve en *npr* el número de procesos del comunicador especificado.

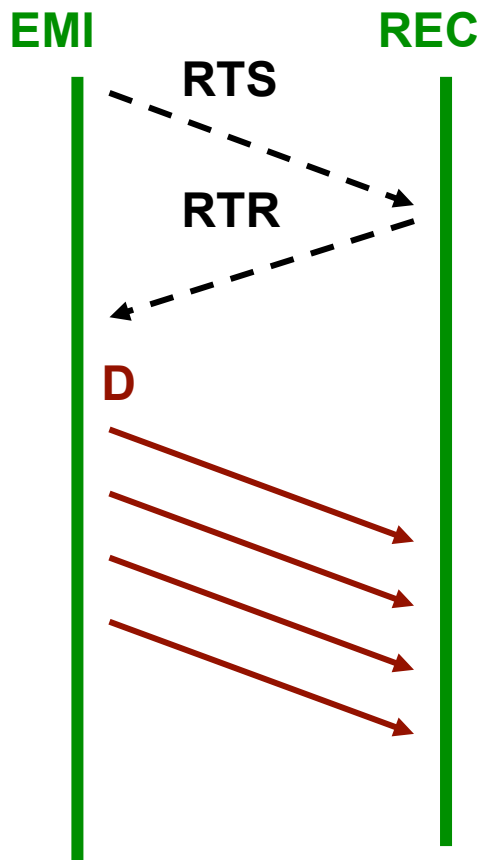
# Tipos de mensajes (I)

- Punto a punto: **1 emisor -1 receptor**
  - Síncronos: El emisor se espera a que la comunicación se produzca (o que el buffer del receptor este disponible). **BLOQUEANTE**
    - Send, Recv, SendRecv, Bsend, Ssend, ...
  - Asíncronos: El emisor NO se espera a que la comunicación se produzca y se comprueba más tarde si se ha producido. **NO BLOQUEANTES**
    - Isend, Irecv, ...

# Tipo de mensajes (II)

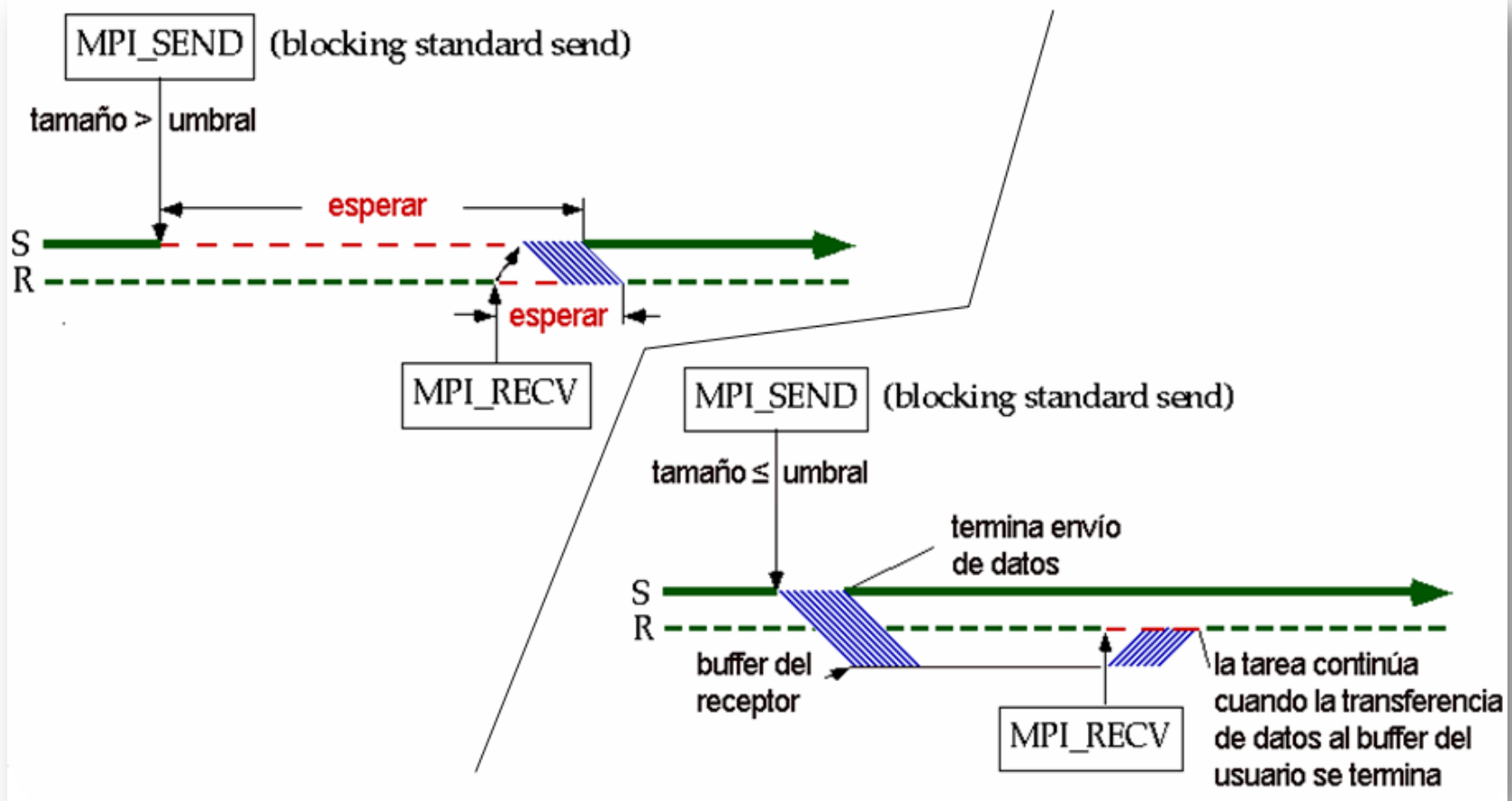
- Multipunto: **1 o N emisores- 1 o N receptores**
  - Bcast, Reduce, Barrier, Gather, Scatter, AllReduce, AllAllgather, ...

# Ejemplo de comunicación



- Petición de transmisión
- Aceptación de transmisión
- Envío de datos (de buffer de usuario a buffer de usuario)

# MPI\_Send, MPI\_Recv (I)



# MPI\_Send, MPI\_Recv (II)

**MPI\_Send(buf, count, datatype, dest, tag, comm)**

**MPI\_Recv(buf, count, datatype, source, tag, comm)**

- **buf:** Dirección donde comienza el mensaje.
- **count:** número de elementos del mensaje.
- **datatype:** tipo del mensaje.
- **dest/source:** posición relativa del proceso fuente/destino dentro del comunicador.  
MPI\_ANY\_SOURCE: permite recibir mensaje de cualquier fuente



# Ejemplo de Send y Recv (I)

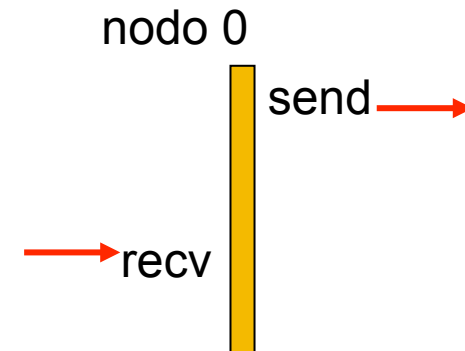
```
#include "mpi.h"
#include <stdio.h>
int main(int argc, char *argv[]) {
    int numtasks, rank, dest, source, rc, tag=1;
    char inmsg, outmsg='x'; MPI_Status Stat;
    MPI_Init(&argc,&argv);
    MPI_Comm_size(MPI_COMM_WORLD, &numtasks);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);

    if (rank == 0) {
        outmsg="r";
        dest = 1;
        printf("Soy %d envio a %d\n",rank,dest);
        rc = MPI_Send(&outmsg, 1, MPI_CHAR, dest, tag, MPI_COMM_WORLD);
    }
    else
    {
        source=0;
        rc = MPI_Recv(&inmsg, 1, MPI_CHAR, source, tag, MPI_COMM_WORLD,&Stat);
        printf("Soy %d recibo de %d la siguiente letra %c\n",rank,source,inmsg);
    }
    MPI_Finalize();
}
```

# Ejemplo de Send y Recv (II)

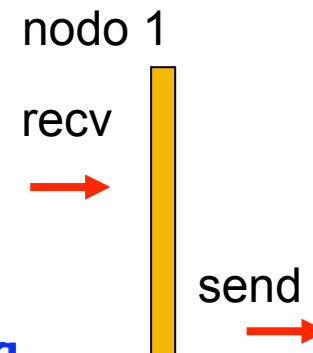
```
#include "mpi.h"
#include <stdio.h>
int main(int argc, char *argv[]) {
    int numtasks, rank, dest, source, rc, tag=1;
    char inmsg, outmsg='x'; MPI_Status Stat;
    MPI_Init(&argc,&argv);
    MPI_Comm_size(MPI_COMM_WORLD, &numtasks);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);

    if (rank == 0) {
        dest = 1;
        source = numtasks-1;
        printf("Soy %d envio a %d\n",rank,dest);
        rc = MPI_Send(&outmsg, 1, MPI_CHAR, dest, tag,
MPI_COMM_WORLD);
        printf("Soy %d espero a %d\n",rank,source);
        rc = MPI_Recv(&inmsg, 1, MPI_CHAR, source, tag,
MPI_COMM_WORLD,&Stat);
        printf("Soy %d recibo de %d\n",rank,source);
    }
}
```

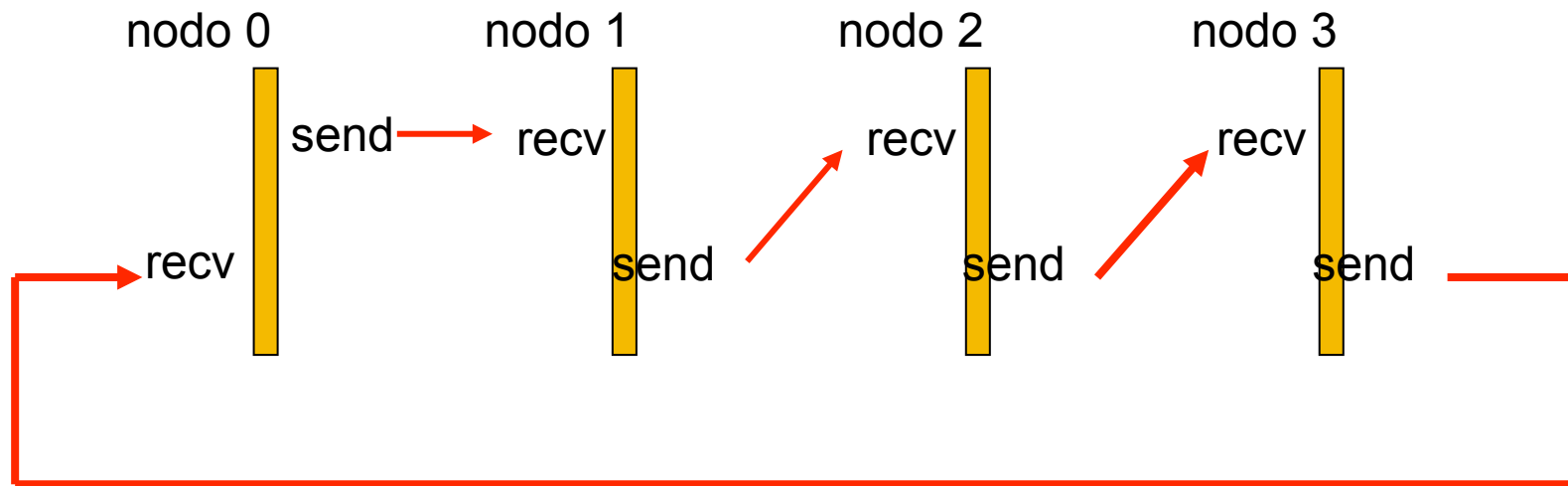


# Ejemplo de Send y Recv (III)

```
else {  
    dest = rank+1;  
    if(dest == numtasks){  
        dest = 0;  
    }  
    source = rank - 1;  
    printf("Soy %d espero a %d\n",rank,source);  
    rc = MPI_Recv(&inmsg, 1, MPI_CHAR, source, tag,  
MPI_COMM_WORLD, &Stat);  
    printf("Soy %d recibo de %d\n",rank,source);  
    printf("Soy %d envio a %d\n",rank,dest);  
    rc = MPI_Send(&outmsg, 1, MPI_CHAR, dest, tag,  
MPI_COMM_WORLD);  
}  
  
MPI_Finalize();  
}
```



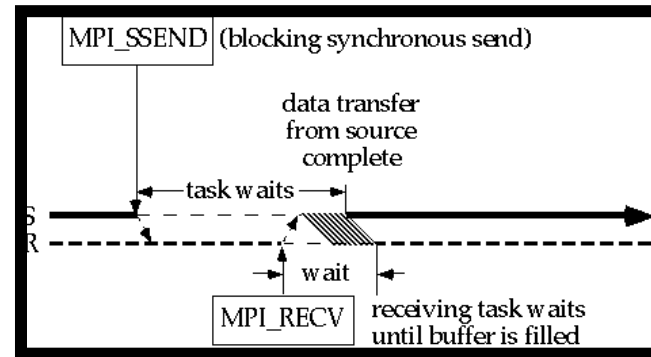
# Ejemplo de Send y Recv (IIII)



# MPI\_Send

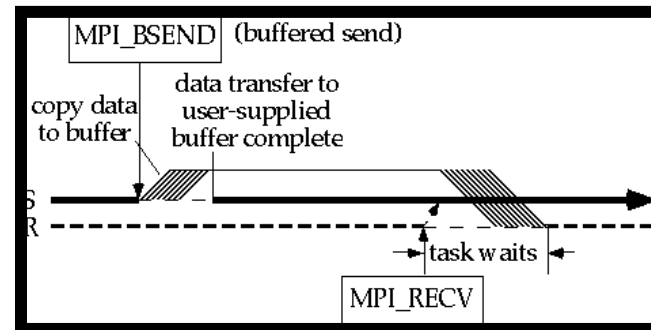
## Síncrono: **MPI\_Ssend(.....)**

La operación se da por terminada sólo cuando el mensaje ha sido recibido en destino



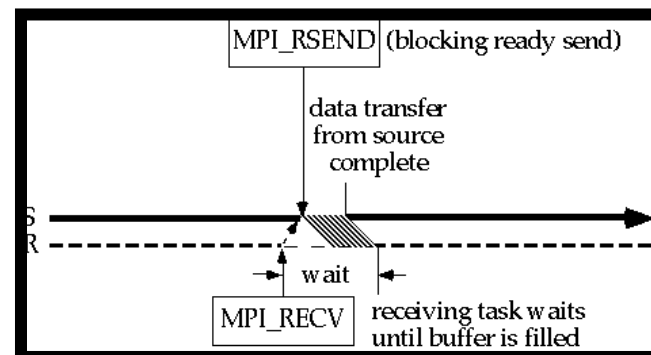
## Buffered: **MPI\_Bsend(.....)**

Cuando se hace un envío con buffer se guarda inmediatamente, en un buffer al efecto en el emisor, una copia del mensaje. La operación se da por completa en cuanto se ha efectuado esta copia. Si no hay espacio en el buffer, el envío fracasa.

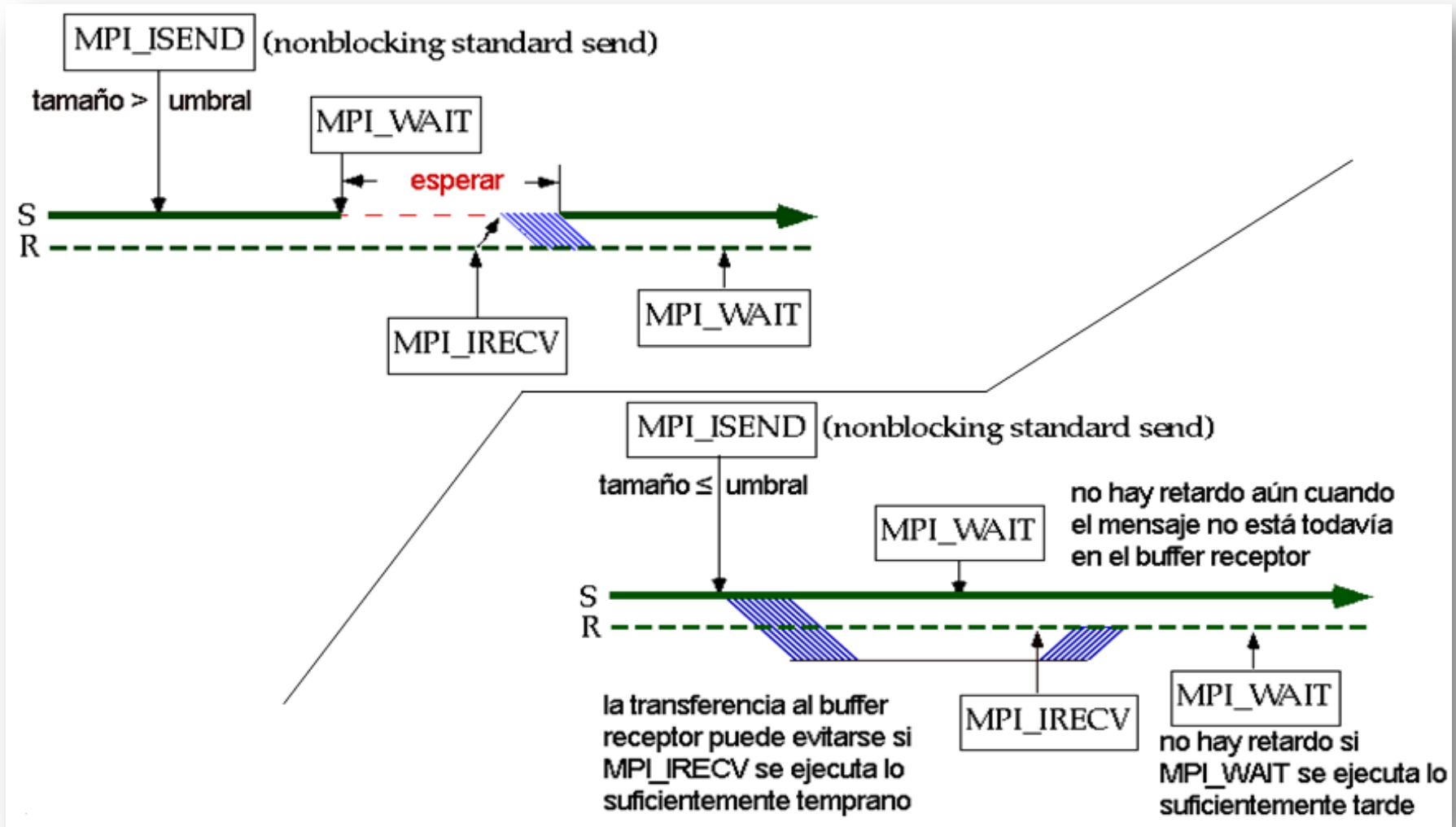


## Ready: **MPI\_Rsend(.....)**

Sólo se puede hacer si antes el otro extremo está preparado para una recepción inmediata. No hay copias adicionales del mensaje (como en el caso del modo con buffer), y tampoco podemos confiar en bloquearnos hasta que el receptor esté preparado.



# MPI\_ISEND, MPI\_Irecv (I)



# MPI\_Isend, MPI\_Irecv (II)

**MPI\_Isend(buf, count, datatype, dest, tag, comm, request)**

**MPI\_Irecv(buf, count, datatype, source, tag, comm, request)**

- **request:** Se usa para saber si la operación ha acabado.
- **MPI\_Wait():** vuelve si la operación se ha completado, espera hasta que se complete.
- **MPI\_Test():** devuelve una bandera que indica si la operación se ha completado.

# Llamadas colectivas (I)

**MPI\_Barrier( )**

**MPI\_Bcast( )**

**MPI\_Gather( )**

**MPI\_Scatter( )**

**MPI\_Alltoall( )**

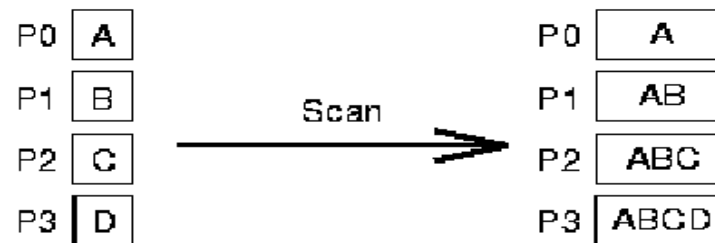
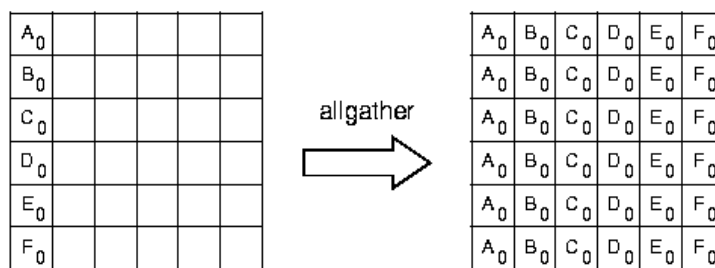
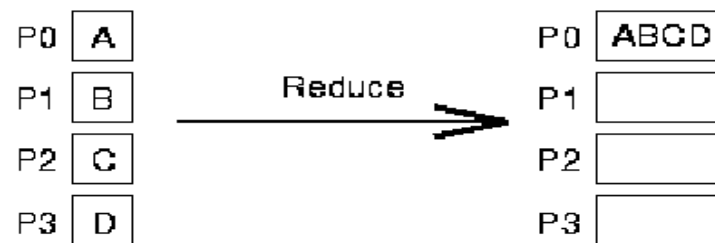
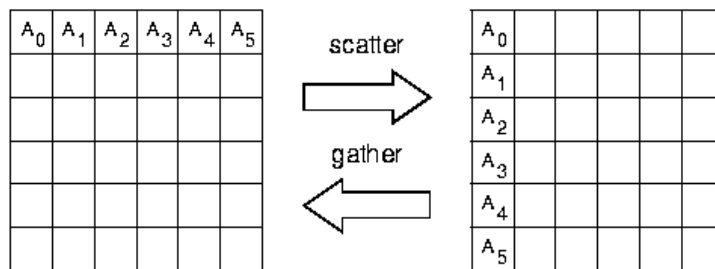
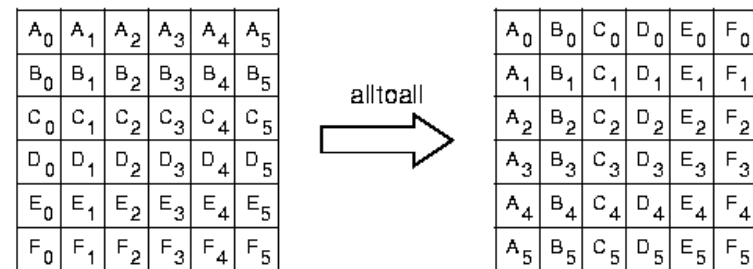
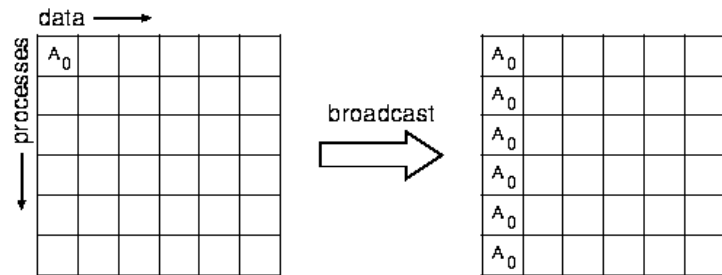
**MPI\_Reduce( )**

**MPI\_Reduce\_scatter( )**

**MPI\_Scan( )**

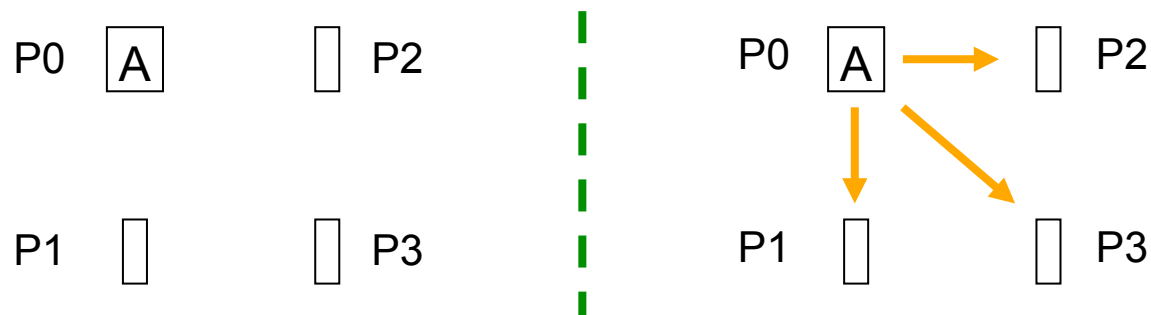


# Llamadas colectivas (II)



# Broadcast

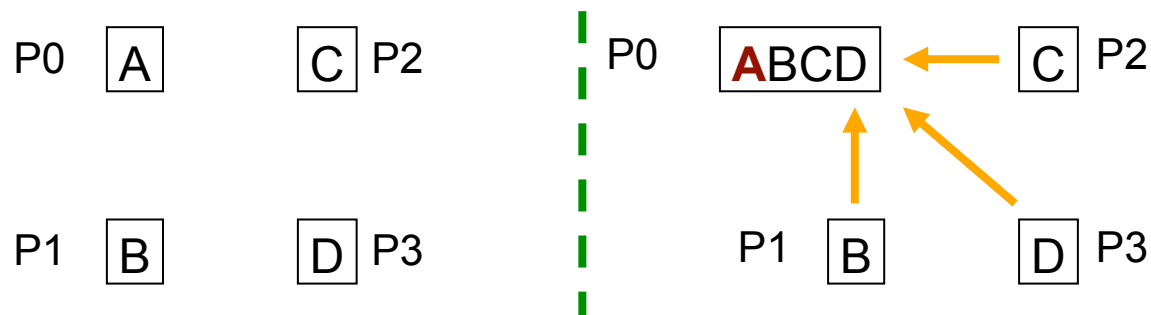
- BROADCAST: envío de datos desde un proceso (root) a todos los demás.



- `MPI_Bcast (&message, count, datatype, root, comm);`

# Gather

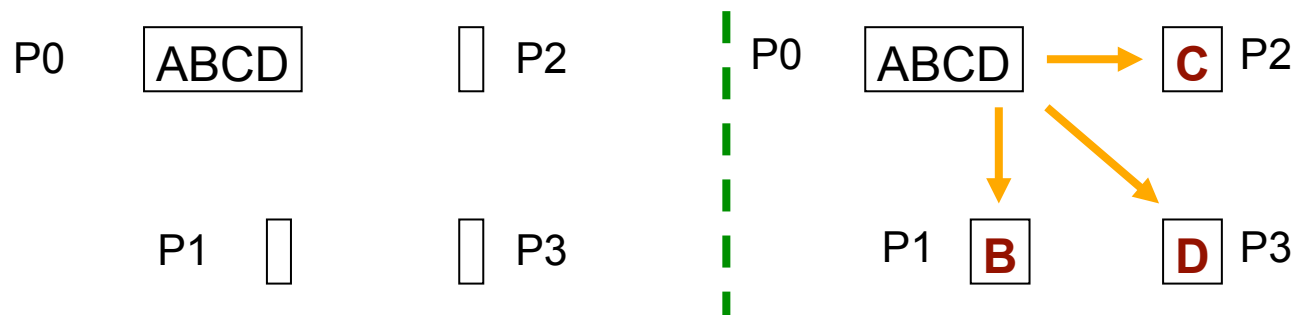
- GATHER: recolección de datos de todos los procesos en uno de ellos (orden estricto pid).



```
MPI_Gather (&send_data, send_count,  
send_type,&recv_data, recv_count,  
recv_type,root, comm);
```

# Scatter

- SCATTER: distribución de datos de un proceso entre todos los demás.



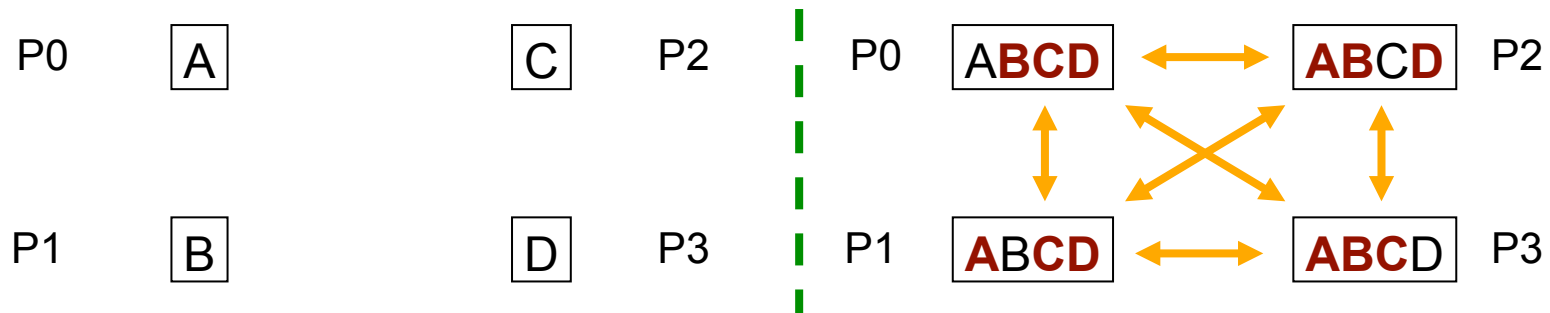
- `MPI_Scatter (&send_data, end_count, send_type, &recv_data, recv_count, recv_type, root, comm);`

# Scatterv

- MPI\_Scatterv() se usa si los datos a enviar:
  - no están almacenados de forma consecutiva en memoria, o
  - los bloques a enviar a cada proceso no son todos del mismo tamaño
- `int MPI_Scatterv(void* sendbuf, int *sendcounts, int *displs, MPI_Datatype sendtype, void* recvbuf, int recvcount, MPI_Datatype recvtype, int root, MPI_Comm comm);`

# Allgather

- ALLGATHER: gather de todos a todos



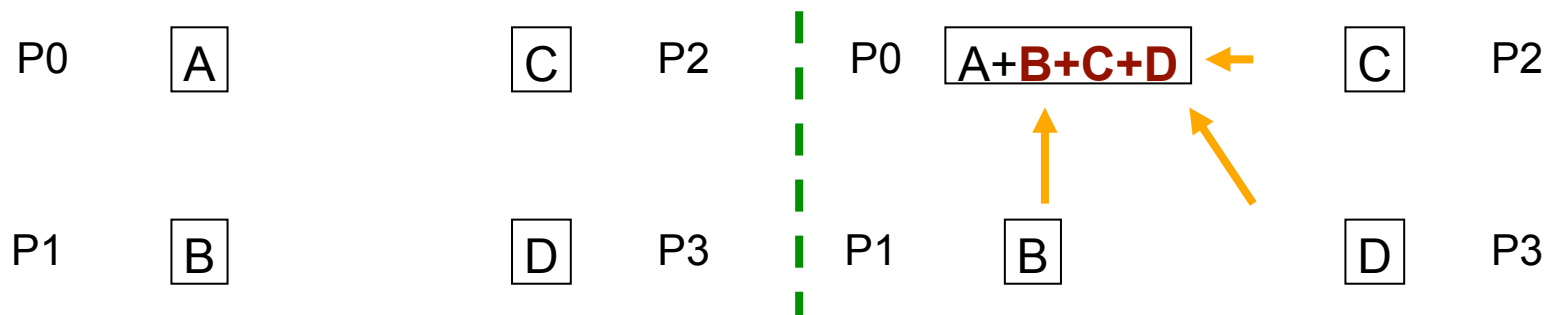
- `MPI_Allgather(&send_data, send_count, send_type, &recv_buf, recv_count, recv_type, comm);`

# Barrier

- BARRIER: sincronización global entre los procesadores del comunicador.
- MPI\_Barrier (comm);
  - La función se bloquea hasta que todos los procesos del comunicador la ejecutan.

# Reduce

- REDUCE: una operación de reducción con los datos de cada procesador, dejando el resultado en uno de ellos (root)



```
MPI_Reduce (&operand, &result, count,  
datatype, operator, root, comm);
```



# Operadores

MPI Operator

Operation

---

MPI_MAX	maximum
MPI_MIN	minimum
MPI_SUM	sum
MPI_PROD	product
MPI_LAND	logical and
MPI_BAND	bitwise and
MPI_LOR	logical or
MPI_BOR	bitwise or
MPI_LXOR	logical exclusive or
MPI_BXOR	bitwise exclusive or
MPI_MAXLOC	max value and location
MPI_MINLOC	min value and location

# Ejemplo MPI\_Reduce

```
#include <stdio.h>
#include <string.h>
#include "mpi.h"
main(int argc, char* argv[])
{
    int my_rank;
    int num_procs;
    int num, prod;
    MPI_Init(&argc, &argv);
    MPI_Comm_rank(MPI_COMM_WORLD, &my_rank);
    MPI_Comm_size(MPI_COMM_WORLD, &num_procs);

    num = my_rank+1;
    MPI_Reduce(&num,&prod,1,MPI_INT,MPI_PROD,
0,MPI_COMM_WORLD);
    if (my_rank==0)
        printf("Hola, soy el proceso %d y %d nodos calculamos el producto:
        %d\n", my_rank ,num_procs, prod);
    MPI_Finalize();
}
```

```
int MPI_Reduce (
    void *sendbuf,
    void *recvbuf,
    int count,
    MPI_Datatype datatype,
    MPI_Op op,
    int root,
    MPI_Comm comm )
```

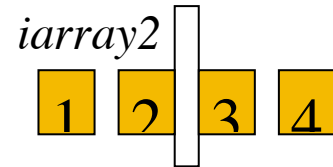
# Ejemplo Bcast

```
#include <stdio.h>
#include <string.h>
#include "mpi.h"
main(int argc, char* argv[])
{
    int my_rank;
    int n;
    MPI_Init(&argc, &argv);
    MPI_Comm_rank(MPI_COMM_WORLD, &my_rank);
    if (my_rank == 0) {
        printf("Hola, ingresa un número: ");
        scanf("%d",&n);
    }
    MPI_Bcast(&n,1,MPI_INT,0,MPI_COMM_WORLD);
    printf("Hola, soy el proceso   %d y el número es el %d
        \n",my_rank,n);
    MPI_Finalize();
}
```

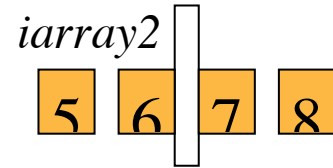
```
int MPI_Bcast (
    void *buffer,
    int count,
    MPI_Datatype datatype,
    int root,
    MPI_Comm comm )
```

# Ejemplo uso de MPI\_Gather

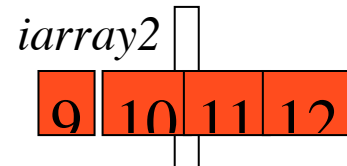
Proceso 0



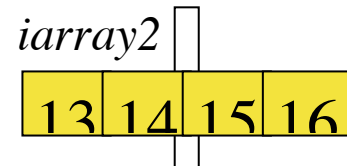
Proceso 1



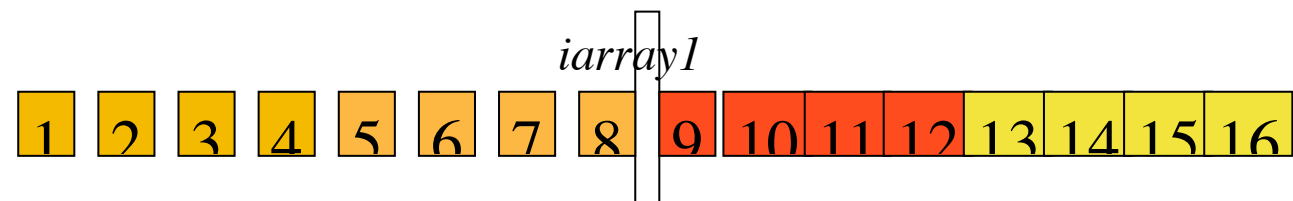
Proceso 2



Proceso 3



Proceso 0



```
call MPI_GATHER( iarray2, 4, MPI_INTEGER,  
                iarray1, 4, MPI_INTEGER,  
                0, MPI_COMM_WORLD)
```

# Ejemplo uso de MPI\_GATHER

```
#include <stdio.h>
#include <string.h>
#include "mpi.h"

main(int argc, char* argv[])
{
    int i,j,k,l;
    int iarray1[16];
    int iarray2[4];
    int myid, numprocs, ierr;
    int status(MPI_STATUS_SIZE);

    MPI_INIT( &argc,&argv );
    MPI_COMM_RANK( MPI_COMM_WORLD, myid);
    MPI_COMM_SIZE( MPI_COMM_WORLD, numprocs);

    printf("Soy el nodo %d de %d procesos\n", myid,numproc);
    for (j=0;j<16;i++)
        iarray1[j]=0;
        for(j=0;j<4;j++)
            iarray2[j]=(myid*4)+j;

    MPI_GATHER(iarray2,4,MPI_INTEGER, iarray1,4,MPI_INTEGER,0,MPI_COMM_WORLD);
    if (myid==0)
        for(i=0;i<16;i++)
            printf("%d\n",iarray1[i]);
    MPI_FINALIZE();
}
```

```
int MPI_Gather (
    void *sendbuf,
    int sendcnt,
    MPI_Datatype sendtype,
    void *recvbuf,
    int recvcount,
    MPI_Datatype recvttype,
    int root,
    MPI_Comm comm )
```

# Tipos de Datos en MPI

<b>MPI_CHAR</b>	<b>signed char</b>
<b>MPI_SHORT</b>	<b>signed short int</b>
<b>MPI_INT</b>	<b>signed int</b>
<b>MPI_LONG</b>	<b>signed long int</b>
<b>MPI_UNSIGNED_CHAR</b>	<b>unsigned char</b>
<b>MPI_UNSIGNED_SHORT</b>	<b>unsigned short int</b>
<b>MPI_UNSIGNED</b>	<b>unsigned int</b>
<b>MPI_UNSIGNED_LONG</b>	<b>unsigned long int</b>
<b>MPI_FLOAT</b>	<b>float</b>
<b>MPI_DOUBLE</b>	<b>double</b>
<b>MPI_LONG_DOUBLE</b>	<b>long double</b>
<b>MPI_BYTE</b>	
<b>MPI_PACKED</b>	

# Crear nuevos tipos de datos

- MPI, nos permite crear nuevos tipos de datos (DataType):
  - n elementos de un vector con stride s
    - MPI\_Type\_vector → Stride s fijo
    - MPI\_Type\_contiguous → Strides s==1
    - MPI\_Type\_indexed → Stride s variable
  - generales (struct)
  - Empaquetamiento
- MPI\_Type\_commit

Hay que llamar a esta rutina antes de poder usar el nuevo tipo de datos que hemos creado.

# Tipos de datos derivados

- **MPI\_Type\_vector**

Elementos espaciados a intervalos regulares.

## MPI\_Type\_vector

```
count = 4; blocklength = 1; stride = 4;  
MPI_Type_vector(count, blocklength, stride, MPI_FLOAT,  
&columnstype);
```

1.0	2.0	3.0	4.0
5.0	6.0	7.0	8.0
9.0	10.0	11.0	12.0
13.0	14.0	15.0	16.0

a[4][4]

**Ejemplo: enviar la columna 2 de la matriz A, de P0 a P1:**

```
float a[4][4];  
MPI_Datatype columstype;  
MPI_Type_vector(4, 1, 4, MPI_FLOAT, columstype);  
MPI_Type_commit(&columstype);  
if (pid==0)  
MPI_Send(&a[0][1], 1, columstype, 1, 0, MPI_COMM_WORLD);  
else if (pid==1)  
MPI_Recv(&a[0][1], 1, columstype, 0, 0, MPI_COMM_WORLD, &info);
```

```
MPI_Send(&a[0][1], 1, columstype, dest, tag, comm);
```

2.0	6.0	10.0	14.0
-----	-----	------	------

1 element of  
columnstype



# Tipos de datos derivados

- **MPI\_Type\_contiguous**

Colección de elementos de un tipo básico, todos del mismo tamaño y almacenados consecutivamente en memoria.

## MPI\_Type\_contiguous

```
count = 4;  
MPI_Type_contiguous(count, MPI_FLOAT, &rowtype);
```

1.0	2.0	3.0	4.0
5.0	6.0	7.0	8.0
9.0	10.0	11.0	12.0
13.0	14.0	15.0	16.0

a[4][4]

```
MPI_Send(&a[2][0], 1, rowtype, dest, tag, comm);
```

9.0	10.0	11.0	12.0
-----	------	------	------

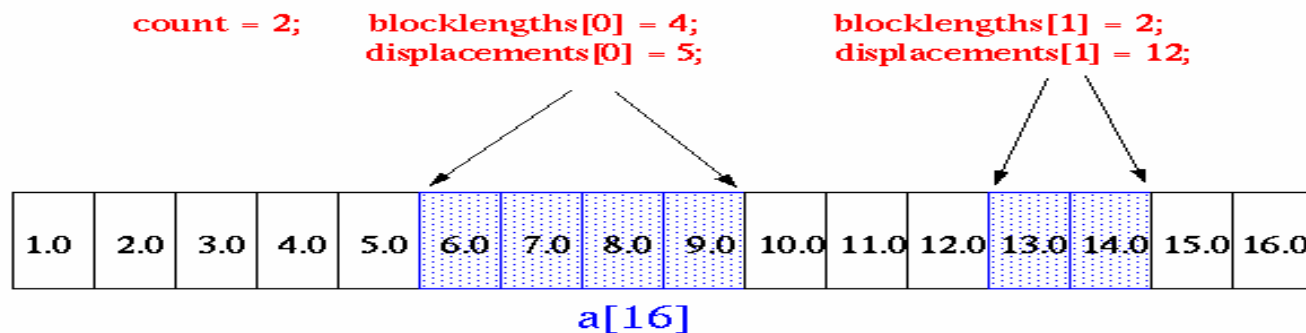
1 element of  
rowtype

# Tipos de datos derivados

- **MPI\_Type\_indexed**

Elementos son entradas arbitrarias de un vector.

## MPI\_Type\_indexed



```
MPI_Type_indexed(count, blocklengths, displacements, MPI_FLOAT, &indextype);
```

```
MPI_Send(&a, 1, indextype, dest, tag, comm);
```



1 element of  
indextype

# Tipos de datos derivados heterogéneos

El nuevo tipo de dato se forma de acuerdo a una agrupación de tipos de datos

**MPI\_TYPE\_STRUCT(  
count,  
blockcounts(),  
offsets(),  
oldtypes,  
newtype,  
ierr)**

## MPI\_Type\_struct

```
typedef struct { float x,y,z,velocity; int n,type; } Particle;  
Particle particles[NELEM];
```

```
MPI_Type_extent(MPI_FLOAT,&extent);
```

```
count = 2; oldtypes[0] = MPI_FLOAT; oldtypes[1] = MPI_INT  
offsets[0] = 0; offsets[1] = 4 * extent;  
blockcounts[0] = 4; blockcounts[1] = 2;
```



```
MPI_Type_struct(count, blockcounts, offsets, oldtypes, &particletype);
```

```
MPI_Send(particles, NELEM, particletype, dest, tag, comm);
```

Sends entire (NELEM) array of particles, each particle being comprised four floats and two integers.

# Tipos de datos derivados

- **MPI\_Type\_Struct** : Nos permite crear un Datatype de con bloques de distinto tipos de datos. Podemos generar un tipo que agrupe datos de tipo float, char y enteros en un mensaje y efectuar un único envío.

```
int MPI_Type_struct(  
    int count,  
    int blocklens[],  
    MPI_Aint indices[],  
    MPI_Datatype old_types[],  
    MPI_Datatype *newtype )  
  
    blen[0] = 1; indices[0] = 0;           oldtypes[0] = MPI_INT;  
    blen[1] = 1; indices[1] = &foo.b - &foo; oldtypes[1] = MPI_CHAR;  
    blen[2] = 1; indices[2] = sizeof(foo); oldtypes[2] = MPI_UB;  
    MPI_Type_struct( 3, blen, indices, oldtypes, &newtype );
```

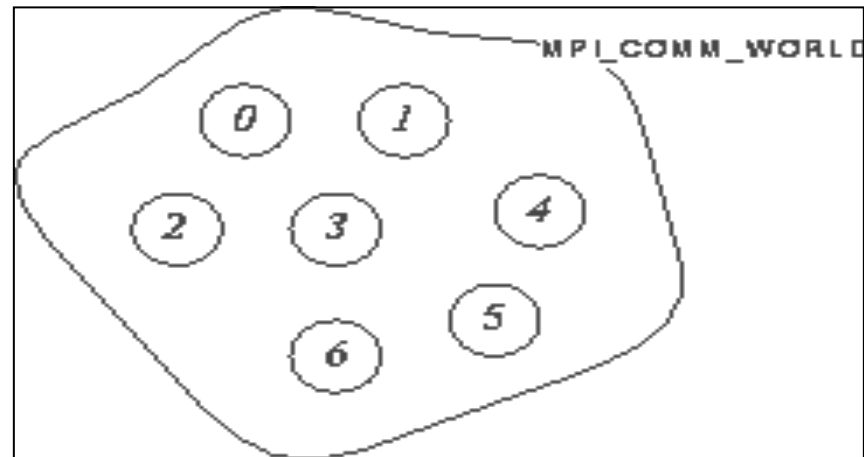
# Empaquetar y desempaquetar

- **MPI\_Pack**  
Empaqueta datos en un *buffer* continuo en memoria para ser enviados como un mensaje único.
- **MPI\_Unpack**  
Desempaqueta los datos recibidos en un mensaje.

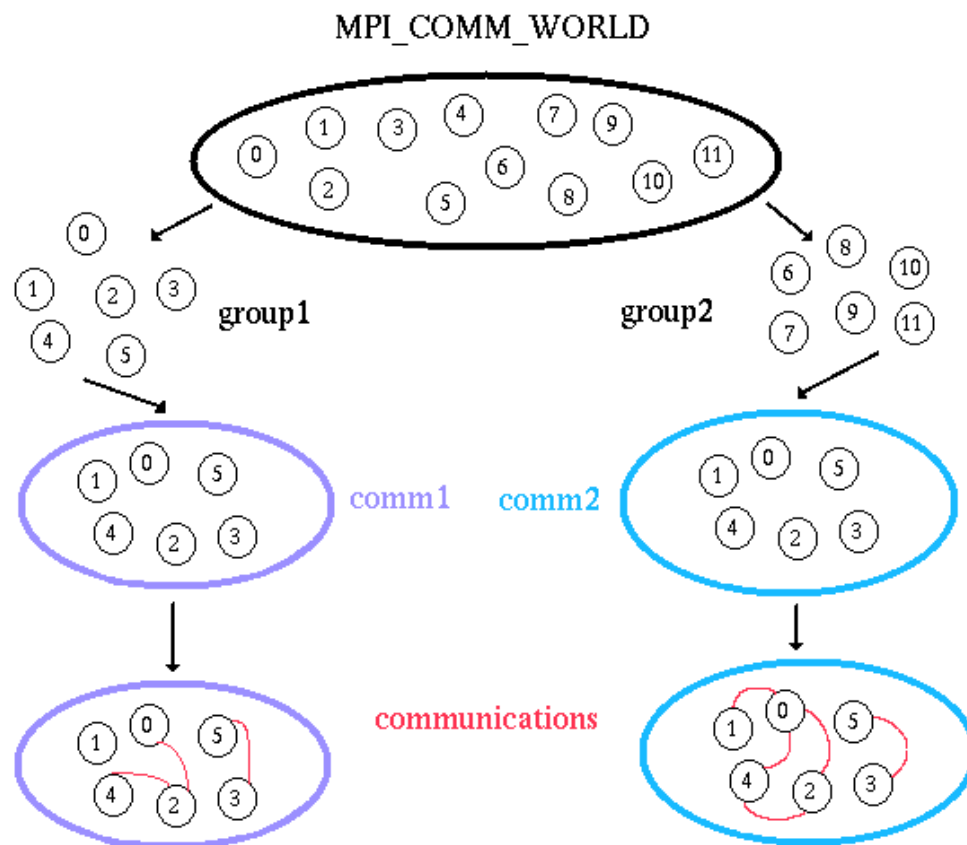
# Cuando utilizar cada tipo

- Empaquetar y desempaquetar se suele utilizar cuando se quieren mandar pocos datos, del mismo tipo y de una sola vez.
- Si los datos a enviar son muchos y no consecutivos, y se envían repetidamente, lo mejor sería definir un tipo específico:
  - datos homogéneos: contiguous, vector, indexed
  - datos heterogéneos: struct

# Comunicadores (I)



# Comunicadores(I)



- Extraer un *handle* de `MPI_COMM_WORLD` usando `MPI_Comm_group`
- Formar nuevo grupo usando `MPI_Group_{include, exclude,...}`
- Crear un nuevo comunicador para el nuevo grupo con `MPI_Comm_create`
- Realizar las comunicaciones con las rutinas de paso de mensajes
- Al terminar libere el comunicador con `MPI_Comm_free` y `MPI_Group_free`



# Comunicadores (II)

- Creación

**MPI\_Comm\_create( ↓Old\_Comm, ↓Group\_Proc,  
↑New\_Comm)**

**MPI\_Comm\_dup( ↓Comm, ↑New\_Comm)**

**MPI\_Comm\_split( ↓Old\_Comm, ↓split\_key, ↑New\_Comm)**

- Eliminación

- **MPI\_Comm\_free( ↓↑Comm)**

- Información

- **MPI\_Comm\_rank( ↓Comm, ↑Rank)**

- **MPI\_Comm\_compare( ↓Comm1, ↓Comm2, ↑Result)**

# Ejemplo usando comunicadores

```
#include "mpi.h";
MPI_Comm comm_world, comm_worker;
MPI_Group group_world, group_worker;
int ierr;
comm_world = MPI_COMM_WORLD;
MPI_Comm_group(comm_world, &group_world);
MPI_Group_excl(group_world, 1, 0, &group_worker);
MPI_Comm_create(comm_world, group_worker, &comm_worker);
/* ACCIONES */
MPI_Comm_free(&comm_worker);
```

# Grupos (I)

- Creación

**MPI\_Comm\_incl( ↓Group, ↓NP, ↓Who[], ↑New\_Group)**

**MPI\_Comm\_excl( ↓Group, ↓NP, ↓Who[], ↑New\_Group)**

**MPI\_Group\_range\_incl( ↓Group, ↓NP, ↓Rango[][3],  
↑New\_Group)**

**MPI\_Group\_range\_excl( ↓Group, ↓NP, ↓Rango[][3],  
↑New\_Group)**

**MPI\_Group\_difference( ↓Group1, ↓Group2, ↑New\_Group)**

**MPI\_Group\_intersection( ↓Group1, ↓Group2, ↑New\_Group)**

**MPI\_Group\_union( ↓Group1, ↓Group2, ↑New\_Group)**

# Grupos (II)

- Eliminación

**MPI\_Comm\_free( ↓ Group)**

- Información

**MPI\_Comm\_group( ↓ Comm, ↑ Group)**

**MPI\_Group\_rank( ↓ Group, ↑ Rank)**

**MPI\_Group\_compare( ↓ Group1, ↓ Group2, ↑ Result)**

**MPI\_Group\_translate\_ranks(**

**↓ Group1, ↓ NP, ↓ Ranks1[], ↓ Group2, ↑ Ranks2)**

# Tiempo

- MPI Wtime()  
double start, finish, time;  
MPI\_Barrier(MPI\_COMM\_WORLD);  
start = **MPI\_Wtime();**  
...  
...  
MPI\_Barrier(MPI\_COMM\_WORLD);  
finish = **MPI\_Wtime();**  
time = finish - start;

# Referencia de MPICH2

- La herramienta **gcc** invoca al compilador de C nativo (por ejemplo cc) y provee el path de los ficheros de cabecera, librerías y enlaces implícitos de MPICH2. La librería MPI se enlaza explícitamente:
  - **% gcc -o programa programa.c -Impich**
- Iniciamos una aplicación MPI invocando al comando **mpiexec**.
  - Este comando provee la información sobre los nodos e identificadores de proceso necesarios para que los procesos se localicen entre si.

# Referencia de MPICH2

- La sintaxis para el comando ***mpiexec*** es la siguiente

**% mpiexec.mpich2 -n<#> programa**

- ***-n <#>*** indica el número de procesos que se crearán.
- ***Programa*** es el fichero ejecutable (incluido el path para llegar hasta el)

# Referencia de MPICH2

- La herramienta ***mpdboot*** inicializa MPICH2 en el cluster especificado.
  - **% mpdboot -n <num máquinas> -f <boot schema>**
  - La opción **-v** permite imprimir mensajes antes de cada paso dado.
- Para comprobar que MPICH2 está corriendo normalmente en el cluster:
  - **% mpdtrace**



# Referencia de MPICH2

- Podemos monitorizar cierta información de los procesos MPI, en tiempo de ejecución, mediante los comandos:
  - **mpdlistjobs** visualiza información sobre los procesos MPI, tal como los ranks (identificadores) de la fuente y el destino, el comunicador, y la función que se está ejecutando

# Referencia de MPICH2

- Mediante el comando ***mpdkilljob***, podemos eliminar todos los procesos y mensajes sin necesidad de abortar MPI.
  - **% mpdkilljob <job id>**
- Por último, la orden ***mpdallexit*** elimina todos los procesos y demonios de MPICH2 en las máquinas que estamos utilizando.
  - **% mpdallexit**

# Seguridad

- Para ejecutar las prácticas hacer antes:
  - `cd ~/.ssh`
  - `ssh-keygen -t dsa`
  - `ssh-keygen -t rsa`
  - `cat *.pub > authorized_keys`

# Compilando y ejecutando

## Compilando

- *mpicc -o hola hola.c*

## Ejecutando

- Fichero con máquinas: *machines*
- *mpdboot -n 2 -f machines*
- *mpiexec -n 2 hola*

# Enlaces

- [http://mpi.deino.net/mpi\\_functions/](http://mpi.deino.net/mpi_functions/)
- <http://www.mpi-forum.org>
- <http://www.mcs.anl.gov/mpi/mpich/>
- <http://www.lam-mpi.org/>
- [http://www.arcos.inf.uc3m.es/~ii\\_ac2/03-04/ejemplos\\_mpi.zip](http://www.arcos.inf.uc3m.es/~ii_ac2/03-04/ejemplos_mpi.zip)
- <http://www.mhpcc.edu/training/workshop/mpi/MAIN.html>
- [http://www.urjc.es/cat/hidra/manuales/MPI/tutorial\\_MPI.pdf](http://www.urjc.es/cat/hidra/manuales/MPI/tutorial_MPI.pdf)
- [http://www.cnb.uam.es/~carazo/practica\\_mpi.html](http://www.cnb.uam.es/~carazo/practica_mpi.html)
- [http://numerix.us.es/pers/denk/\\_parallel/](http://numerix.us.es/pers/denk/_parallel/)
- <http://www ldc.usb.ve/~ibanez/docencia/MPI/>
- <http://www.tc.cornell.edu/services/edu/topics/parallelvw.asp>
- <http://web.tiscali.it/Moncada/documenti/mpi.ppt>
- <http://www-copa.dsic.upv.es/docencia/iblanque/lcp/>
- <http://www.arcos.inf.uc3m.es/~mimpi/>
- <http://www.df.uba.ar/users/marcelo/>
- <http://www.dirinfo.unsl.edu.ar/~prgparal/Teorias/>
- <http://www.cecalc.ula.ve/documentacion/tutoriales/mpi/mpi.ps>