mpiJava 1.2: API Specification

Bryan Carpenter, Geoffrey Fox,
Sung-Hoon Ko, Sang Lim

Northeast Parallel Architectures Centre,
Syracuse University,
111 College Place,
Syracuse, New York 13244-410
{dbc,gef,shko,slim}@pac.syr.edu
This document defines the API of mpiJava, a Java language binding for MPI 1.1. The document is not a standalone specification of the behaviour of MPI—it is meant to be read in conjunction with the MPI standard document [2]. Subsections are laid out in the same way as in the standard document, to allow cross-referencing. Where the mpiJava binding makes no significant change to a particular section of the standard document, we will just note here that there are no special issues for the Java binding. This does not mean that the corresponding section of the standard is irrelevant to the Java binding—it may mean it is 100% relevant! Where practical the API is modelled on the MPI C++ interface defined in the MPI standard version 2.0 [3].

Changes to the mpiJava 1.1 interface:

- The **MPI.OBJECT** basic type has been added.
- The interface for **MPI.Buffer_detach** has been corrected.
- The API of **UserFunction** has been changed.
- Attributes cached in communicators are now assumed to have integer values. **Attr_put** and **Attr_delete** have been removed.
- The interface to **CartcommdimsCreate** has been corrected.
- **ErrorHandler_set, ErrorHandler_get** have changed from static members of **MPI** to instance methods on **Comm**.
- The method **Is_null** has been added to class **Comm**.
- The initialization method **MPI.Init** now returns the command line arguments set by **MPI**.
- MPI exception classes are now specified to be subclasses of **MPIException** rather than **IOException**. Methods are now declared to throw **MPIException** (see section 8).

The current API is viewed as an interim measure. Further significant changes are likely to result from the efforts of the Message-passing working group of the Java Grande Forum.
Contents

1 Introduction to MPI ........................................... 5

2 MPI Terms and Conventions .................................. 6
   2.1 Document Notation ........................................ 6
   2.2 Procedure Specification .................................. 6
   2.3 Semantic terms ........................................... 6
   2.4 Data types ............................................... 6
   2.5 Language Binding ......................................... 6
   2.6 Processes ................................................. 8
   2.7 Error Handling ............................................ 9

3 Point-to-Point Communication ............................... 10
   3.1 Introduction .............................................. 10
   3.2 Blocking Send and Receive operations ..................... 10
   3.3 Data type matching and data conversion .................... 12
   3.4 Communication Modes ..................................... 12
   3.5 Semantics of point-to-point communication ............... 13
   3.6 Buffer allocation and usage ................................ 13
   3.7 Nonblocking communication ............................... 14
   3.8 Probe and Cancel ......................................... 19
   3.9 Persistent communication requests ....................... 20
   3.10 Send-receive ............................................ 22
   3.11 Null processes ........................................... 23
   3.12 Derived datatypes ....................................... 23
   3.13 Pack and unpack ......................................... 32

4 Collective Communication .................................... 34
   4.1 Introduction and Overview ................................ 34
   4.2 Communicator argument .................................... 34
   4.3 Barrier synchronization ................................... 34
   4.4 Broadcast ................................................ 34
   4.5 Gather .................................................... 34
   4.6 Scatter ................................................... 35
   4.7 Gather-to-all ............................................. 36
   4.8 All-to-All Scatter/Gather .................................. 37
   4.9 Global Reduction Operations ............................. 38
   4.10 Reduce-Scatter ............................................ 40
   4.11 Scan ..................................................... 40
   4.12 Correctness .............................................. 41
1 Introduction to MPI

Evidently, this document adds Java to the C and Fortran bindings defined in the MPI standard. Otherwise no special issues for the Java binding.
2 MPI Terms and Conventions

2.1 Document Notation

No special issues for Java binding.

2.2 Procedure Specification

In general we use *italicized* names to refer to entities in the MPI language independent procedure definitions, and *typewriter* font for concrete Java entities.

As a rule Java argument names are the same as the corresponding language independent names. In instance methods of Comm, Status, Request, Datatype, Op or Group (and subclasses), the class instance generally stands for the argument called *comm, status, request, datatype, op* or *group*, respectively in the language independent procedure definition.

2.3 Semantic terms

No special issues for Java binding.

2.4 Data types

Opaque objects are presented as Java objects. This introduces the option of simplifying the user’s task in managing these objects. MPI destructors can be absorbed into Java object destructors, which are called automatically by the Java garbage collector. We adopt this strategy as the general rule. Explicit calls to MPI destructor functions are typically omitted from the Java user interface (they are absorbed into finalize methods). Exceptions are made for the Comm and Request classes. *MPI_Comm_FREE* is a collective operation, so the user must ensure that calls are made at consistent times by all processors involved—the call can’t be left to the vagaries of the garbage collector. A similar case can be made for *MPI_Request_FREE*.

2.5 Language Binding

**Naming Conventions** All MPI classes belong to the package mpi. Conventions for capitalization, etc, in class and member names generally follow the C++ MPI bindings.

**Restrictions on struct derived type.** Some options allowed for *derived data types* in the C and Fortran binding are deleted in mpiJava. The Java VM does not incorporate a concept of a global linear address space. Passing physical addresses to data type definitions is not allowed. The use of the *MPI_TYPE_STRUCT* datatype constructor is also restricted in a way that
makes it impossible to send mixed basic datatypes in a single message. Since, however, the set of basic datatypes recognised by MPI is extended to include serializable Java objects, this should not be a serious restriction in practice.

**Multi-dimensional arrays and offsets.** The C and Fortran languages define a straightforward mapping (or “sequence association”) between their multi-dimensional arrays and equivalent one-dimensional arrays. So in C or Fortran a multidimensional array passed as a message buffer argument is first interpreted as a one-dimensional array with the same element type as the original multidimensional array. Offsets in the buffer (such as offsets occurring in derived data types) are then interpreted in terms of the effective one-dimensional array (or—equivalent up to a constant factor—in terms of physical memory). In Java the relationship between multidimensional arrays and one dimensional arrays is different. An \( n \)-dimensional array is equivalent to a one-dimensional array of \( (n-1) \)-dimensional arrays. In mpiJava, message buffers are always one-dimensional arrays. The element type may be an object, which may have array type. Hence multidimensional arrays can appear as message buffers, but the interpretation is subtly different. In distinction to the C or Fortran case offsets in multidimensional message buffers are always interpreted as offsets in the outermost one-dimensional array.

**Start of message buffer.** C and Fortran both have devices for treating a section of an array, offset from the beginning of the array, as if it was an array in its own right. Java doesn’t have any such mechanism. To provide the same flexibility, an offset parameter is associated with any buffer argument. This defines the position of the first actual buffer element in the Java array.

**Error codes.** Unlike the standard C and Fortran interfaces, the mpiJava interfaces to MPI calls do not return explicit error codes. The Java exception mechanism will be used to report errors.

*Rationale.* The exception mechanism is very widely used by Java libraries. It is inconvenient to use the single return value of a Java function with an error code. (Java doesn’t allow function arguments to be passed by reference, so returning multiple values tends to be more clumsy than in other languages.) *(End of rationale.)*

**Multiple return values.** A few functions in the MPI interface return multiple values, even after the error code is eliminated. This is dealt with in mpiJava in various ways. Sometimes an MPI function initializes some elements in an array and also returns a count of the number of elements modified. In Java we typically return an array result, omitting the count. The count can be obtained subsequently from the length member of the array. Sometimes an MPI
function initializes an object conditionally and returns a separate flag to say if the operation succeeded. In Java we typically return an object reference which is `null` if the operation fails. Occasionally extra internal state is added to an existing MPI class to hold extra results—for example the `Status` class has extra state initialized by functions like `Waitany` to hold the `index` value. Rarely none of these methods work and we resort to defining auxiliary classes to hold multiple results from a particular function.

**Array count arguments.** The mpiJava binding often omits array size arguments, because they can be picked up within the function by reading the `length` member of the array argument. A major exception is for message buffers, where an explicit count is always given.

Rationale. In the mpiJava, message buffers have explicit `offset` and `count` arguments whereas other kinds of array argument typically do not. Message buffers aside, typical array arguments to MPI functions (eg, vectors of request structures) are small arrays. If subsections of these must be passed to an MPI function, the sections can be copied to smaller arrays at little cost. In contrast message buffers are typically large and copying them is expensive, so it is worthwhile to pass the extra arguments. Also, if derived data types are being used, the required value of the `count` argument is always different to the buffer length. *(End of rationale.)*

**Concurrent access to arrays.** In JNI-based wrapper implementations it may be necessary to impose some non-interference rules for concurrent read and write operations on arrays. When an array is passed to an MPI method such as a send or receive operation, the wrapper code will probably extract a pointer to the contents of the array using a JNI `Get...ArrayElements` routine. If the garbage collector does not support “pinning” (temporarily disabling run-time relocation of data for specific arrays—see [1] for more discussion), the pointer returned by this `Get` function may be to a temporary copy of the elements. The copy will be written back to the true Java array when a subsequent call to `Release...ArrayElements` is made. If two operations involving the same array are active concurrently, this copy-back may result in failure to save modifications made by one or more of the concurrent calls.

Such an implementation may have to enforce a safety rule such as: *when several MPI send or receive (etc) operations are active concurrently, if any one of those operations writes to a particular array, none of the other operations must read or write any portion of that array.*

If the garbage collector supports pinning, this problem does not arise.

2.6 Processes

No special issues for Java binding.
2.7 Error Handling

As explained in section 2.5, the Java methods do not return error codes. The Java exceptions thrown instead are defined in section 7.3.
### 3 Point-to-Point Communication

#### 3.1 Introduction

In general the mpiJava binding of point-to-point communication operations realizes the MPI functions as methods of the Comm class. The basic point-to-point communication operations are send and receive. Their use is illustrated in the example below.

```java
import mpi.*;

class Hello {
    static public void main(String[] args) throws MPIException {
        MPI.Init(args);

        int myrank = MPI.COMM_WORLD.Rank();
        if (myrank == 0) {
            char[] message = "Hello, there".toCharArray();
            MPI.COMM_WORLD.Send(message, 0, message.length, MPI.CHAR, 1, 99);
        } else {
            char[] message = new char[20];
            MPI.COMM_WORLDRecv(message, 0, 20, MPI.CHAR, 0, 99);
            System.out.println("received:" + new String(message) + ":");
        }

        MPI.Finalize();
    }
}
```

#### 3.2 Blocking Send and Receive operations

```java
void Comm.Send(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
```

- **buf**: send buffer array
- **offset**: initial offset in send buffer
- **count**: number of items to send
- **datatype**: datatype of each item in send buffer
- **dest**: rank of destination
- **tag**: message tag

Blocking send operation. Java binding of the MPI operation MPLSEND. The data part of the message consists of a sequence of count values, each of the type indicated by datatype. The actual argument associated with buf must be an
array. The value offset is a subscript in this array, defining the position of the first item of the message.

The elements of buf may have primitive type or class type. If the elements are objects, they must be serializable objects. If the datatype argument represents an MPI basic type, its value must agree with the element type of buf: the basic MPI datatypes supported, and their correspondence to Java types, are as follows

<table>
<thead>
<tr>
<th>MPI datatype</th>
<th>Java datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI.BYTE</td>
<td>byte</td>
</tr>
<tr>
<td>MPI.CHAR</td>
<td>char</td>
</tr>
<tr>
<td>MPI.SHORT</td>
<td>short</td>
</tr>
<tr>
<td>MPI.BOOLEAN</td>
<td>boolean</td>
</tr>
<tr>
<td>MPI.INT</td>
<td>int</td>
</tr>
<tr>
<td>MPI.LONG</td>
<td>long</td>
</tr>
<tr>
<td>MPI.FLOAT</td>
<td>float</td>
</tr>
<tr>
<td>MPI.DOUBLE</td>
<td>double</td>
</tr>
<tr>
<td>MPI.OBJECT</td>
<td>Object</td>
</tr>
</tbody>
</table>

If the datatype argument represents an MPI derived type, its base type must agree with the element type of buf (see section 3.12).

Rationale. The datatype argument is not redundant in mpiJava, because we include support for MPI derived types. If it was decided to remove derived types from the API, datatype arguments could be removed from various functions, and Java runtime inquiries could be used internally to extract the element type of the buffer, or methods like Send could be overloaded to accept buffers with elements of the 9 basic types.

(End of rationale.)

If a data type has MPI.OBJECT as its base type, the objects in the buffer will be transparently serialized and unserialized inside the communication operations.

Status Comm.Recv(Object buf, int offset, int count, Datatype datatype, int source, int tag)

buf      receive buffer array
offset   initial offset in receive buffer
count    number of items in receive buffer
datatype datatype of each item in receive buffer
source   rank of source
tag       message tag

returns: status object

11
Blocking receive operation. Java binding of the MPI operation `MPI_RECV`. The actual argument associated with `buf` must be an array. The value `offset` is a subscript in this array, defining the position into which the first item of the incoming message will be copied.

The elements of `buf` may have primitive type or class type. If the `datatype` argument represents an MPI basic type, its value must agree with the element type of `buf`; if `datatype` represents an MPI derived type, its `base type` must agree with the element type of `buf` (see section 3.12).

The MPI constants `MPI_ANY_SOURCE` and `MPI_ANY_TAG` are available as `MPI_ANY_SOURCE` and `MPI_ANY_TAG`.

The source and tag of the received message are available in the publicly accessible `source` and `tag` fields of the returned object. The following method can be used to further interrogate the return status of a receive operation.

```java
int Status.Get_count(Datatype datatype)
```

`datatype` datatype of each item in receive buffer

returns: number of received entries

Java binding of the MPI operation `MPI_GET_COUNT`.

### 3.3 Data type matching and data conversion

The Java language definition places quite detailed constraints on the representation of its primitive types—for example it requires conformance with IEEE 754 for `float` and `double`. There may still be a requirement for representation conversion in heterogenous systems. For example, source and destination computers (or virtual machines) may have different endianness.

### 3.4 Communication Modes

```java
void Comm.Bsend(Object buf, int offset, int count,
    Datatype datatype, int dest, int tag)
```

`buf` send buffer array
`offset` initial offset in send buffer
`count` number of items to send
`datatype` datatype of each item in send buffer
`dest` rank of destination
`tag` message tag

Send in buffered mode. Java binding of the MPI operation `MPI_BSEND`. Further comments as for `send`.

12
void Comm.Ssend(Object buf, int offset, int count,
    Datatype datatype, int dest, int tag)

buf        send buffer array
offset     initial offset in send buffer
count      number of items to send
datatype   datatype of each item in send buffer
dest       rank of destination
tag        message tag

Send in synchronous mode. Java binding of the MPI operation \texttt{MPI\_SSEND}. Further comments as for \texttt{send}.

void Comm.Rsend(Object buf, int offset, int count,
    Datatype datatype, int dest, int tag)

buf        send buffer array
offset     initial offset in send buffer
count      number of items to send
datatype   datatype of each item in send buffer
dest       rank of destination
tag        message tag

Send in ready mode. Java binding of the MPI operation \texttt{MPI\_RSEND}. Further comments as for \texttt{send}.

3.5 Semantics of point-to-point communication

No special issues for Java binding.

3.6 Buffer allocation and usage

void MPI.Buffer_attach(byte [] buffer)

buffer     buffer array

Provides to MPI a buffer in user’s memory to be used for buffering outgoing messages. Java binding of the MPI operation \texttt{MPI\_BUFFER\_ATTACH}.

byte [] MPI.Buffer_detach()
returns: buffer array

Detach the buffer currently associated with MPI and return it. Java binding of the MPI operation \texttt{MPI\_BUFFER\_DETACH}. If the currently associated buffer is system-defined, returns \texttt{null}.

The MPI constant \texttt{MPI\_BSEND\_OVERHEAD} is available as \texttt{MPI\_BSEND\_OVERHEAD}.

### 3.7 Nonblocking communication

Nonblocking communications use methods of the \texttt{Request} class to identify communication operations and match the operation that initiates the communication with the operation that terminates it.

\begin{verbatim}
Request Comm.Isend(Object buf, int offset, int count, 
    Datatype datatype, int dest, int tag)

buf       send buffer array
offset    initial offset in send buffer
count     number of items to send
datatype  datatype of each item in send buffer
dest      rank of destination
tag       message tag

returns: communication request
\end{verbatim}

Start a standard mode, nonblocking send. Java binding of the MPI operation \texttt{MPI\_ISEND}. Further comments as for \texttt{send}.

\begin{verbatim}
Request Comm.Ibsend(Object buf, int offset, int count, 
    Datatype datatype, int dest, int tag)

buf       send buffer array
offset    initial offset in send buffer
count     number of items to send
datatype  datatype of each item in send buffer
dest      rank of destination
tag       message tag

returns: communication request
\end{verbatim}

Start a buffered mode, nonblocking send. Java binding of the MPI operation \texttt{MPI\_IBSEND}. Further comments as for \texttt{send}. 

14
Request Comm.Issend(Object buf, int offset, int count, 
Datatype datatype, int dest, int tag)

buf send buffer array
offset initial offset in send buffer
count number of items to send
datatype datatyp e of each item in send buffer
dest rank of destination
tag message tag

returns: communication request

Start a synchronous mode, nonblocking send. Java binding of the MPI operation 
MPI_IsSEND. Further comments as for send.

Request Comm.Irsend(Object buf, int offset, int count, 
Datatype datatype, int dest, int tag)

buf send buffer array
offset initial offset in send buffer
count number of items to send
datatype datatype of each item in send buffer
dest rank of destination
tag message tag

returns: communication request

Start a ready mode, nonblocking send. Java binding of the MPI operation 
MPI_IRSEND. Further comments as for send.

Request Comm.Irecv(Object buf, int offset, int count, 
Datatype datatype, int source, int tag)

buf receive buffer array
offset initial offset in receive buffer
count number of items in receive buffer
datatype datatype of each item in receive buffer
source rank of source
tag message tag

returns: communication request
Start a nonblocking receive. Java binding of the MPI operation \texttt{MPI\_IRECV}.

Further comments as for recv.

The following functions are used to complete nonblocking communication operations (and also communications started using the persistent communication requests—subclass \texttt{Request}—introduced later). We use the following terminology. A request is “active” if it is associated with an ongoing communication. Otherwise it is inactive. An inactive instance of the base class \texttt{Request} is called a “void request”. (Note, however, that an inactive instance of the \texttt{Prequest} subclass is not said to be “void”, because it retains detailed information about a communication pattern even when no corresponding communication is ongoing.)

\textit{Rationale.} A “void request” corresponds to what is called a “null handle” in the C and Fortran MPI bindings. It seems impractical to have completion operations like \texttt{wait} set request object references to null references in the Java sense (because Java methods cannot directly modify references passed to them as arguments). To avoid a confusing semantic distinction between null MPI handles and null Java references we introduce the terminology of a “void request object”. If an explicit reference to a void request is needed, one is available as \texttt{MPI\_REQUEST\_NULL}. The inquiry \texttt{Request\_Is\_null} can be used to determine if a particular request is void. (End of rationale.)

\texttt{Status Request\_Wait()}

\textit{returns:} status object

Blocks until the operation identified by the request is complete. Java binding of the MPI operation \texttt{MPI\_WAIT}. After the call returns, the request object becomes inactive.

\texttt{Status Request\_Test()}

\textit{returns:} status object or null reference

Returns a status object if the operation identified by the request is complete, or a null reference otherwise. Java binding of the MPI operation \texttt{MPI\_TEST}. After the call, if the operation is complete (i.e., if the return value of \texttt{test} is non-null), the request object becomes an inactive request.

\texttt{boolean Request\_Is\_null()}

16
returns: true if the request object is void, false otherwise

Note that Is.null is always false on instances of the subclass Request.

void Request.Free()

Set the request object to be void. Java binding of the MPI operation MPI_REQUEST_FREE.

static Status Request.Waitany(Request [] array_of_requests)
array_of_requests array of requests
returns: status object

Blocks until one of the operations associated with the active requests in the array has completed. Java binding of the MPI operation MPI_WAITANY. The index in array_of_requests for the request that completed can be obtained from the status object from the publically accessible Status.index field. The corresponding element of array_of_requests becomes inactive.

The array_of_requests may contain inactive requests. If the list contains no active requests, the method immediately returns a status in which the index field is MPI.UNDEFINED.

static Status Request.Testany(Request [] array_of_requests)
array_of_requests array of requests
returns: status object or null reference

Tests for completion of either one or none of the operations associated with active requests. Java binding of the MPI operation MPI_TESTANY. If some request completed, the index in array_of_requests of that request can be obtained from the status object through the Status.index field. The corresponding element of array_of_requests becomes inactive. If no request completed, testAny returns a null reference.

The array_of_requests may contain inactive requests. If the list contains no active requests, the method immediately returns a status in which the index field is MPI.UNDEFINED.

static Status [] Request.Waitall(Request [] array_of_requests)
**array of requests**  array of requests

**returns:**  array of status objects

Blocks until all of the operations associated with the active requests in the array have completed. Java binding of the MPI operation `MPI_WAITALL`. The result array will be the same size as `array of requests`. On exit, requests become inactive. If the input value of `array of requests` contains any inactive requests, corresponding elements of the result array will contain null status references.

```java
static Status Request.Testall(Request [] array_of_requests)
```

**array of requests**  array of requests

**returns:**  array of status objects, or a null reference

Tests for completion of all of the operations associated with active requests. Java binding of the MPI operation `MPI_TESTALL`. If all operations have completed, the exit values of the argument array and the result array are as for `Waitall`. If any operation has not completed, the result value is null and no element of the argument array is modified.

```java
static Status Request.Waitsome(Request [] array_of_requests)
```

**array of requests**  array of requests

**returns:**  array of status objects

Blocks until at least one of the operations associated with the active requests in the array has completed. Java binding of the MPI operation `MPI_WAITSOME`. The size of the result array will be the number of operations that completed. The index in `array of requests` for each request that completed can be obtained from the `index` field of the returned status objects. The corresponding elements in `array of requests` become inactive.

If `array of requests` list contains no active requests, `testAll` immediately returns a null reference.

```java
static Status Request_testsome(Request [] array_of_requests)
```

**array of requests**  array of requests

**returns:**  array of status objects
Behaves like `waitSome`, except that it returns immediately. Java binding of the MPI operation `MPI_TESTSOME`. If no operation has completed, `Testsome` returns an array of length zero and elements of `arrayOfrequests` are unchanged. Otherwise, arguments and return value are as for `Waitsome`.

### 3.8 Probe and Cancel

**Status Comm.Iprobe(int source, int tag)**

- **source**  
  source rank
- **tag**  
  tag value

**returns:** status object or null reference

Check if there is an incoming message matching the pattern specified. Java binding of the MPI operation `MPI_IProbe`. If such a message is currently available, a status object similar to the return value of a matching `Recv` operation is returned. Otherwise a null reference is returned.

**Status Comm.Probe(int source, int tag)**

- **source**  
  source rank
- **tag**  
  tag value

**returns:** status object or null reference

Wait until there is an incoming message matching the pattern specified. Java binding of the MPI operation `MPI_PROBE`. Returns a status object similar to the return value of a matching `Recv` operation.

**void Request_Cancel()**

Mark a pending nonblocking communication for cancellation. Java binding of the MPI operation `MPI_CANCEL`.

**boolean Status_Test_cancelled()**

**returns:** true if the operation was succesfully cancelled, false otherwise

Test if communication was cancelled. Java binding of the MPI operation `MPI TEST_CANCELLED`.

19
3.9 Persistent communication requests

Prequest Comm.Send_init(Object buf, int offset, int count, 
Datatype datatype, int dest, int tag)

buf   send buffer array
offset    initial offset in send buffer
count    number of items to send
datatype    datatype of each item in send buffer
dest    rank of destination
tag    message tag

returns: persistent communication request

Creates a persistent communication request for a standard mode send. Java
binding of the MPI operation MPI_SEND_INIT. Further comments as for Send.

Prequest Comm.Bsend_init(Object buf, int offset, int count, 
Datatype datatype, int dest, int tag)

buf   send buffer array
offset    initial offset in send buffer
count    number of items to send
datatype    datatype of each item in send buffer
dest    rank of destination
tag    message tag

returns: persistent communication request

Creates a persistent communication request for a buffered mode send. Java bind-
ing of the MPI operation MPI_BSEND_INIT. Further comments as for Send.

Prequest Comm.Ssend_init(Object buf, int offset, int count, 
Datatype datatype, int dest, int tag)

buf   send buffer array
offset    initial offset in send buffer
count    number of items to send
datatype    datatype of each item in send buffer
dest    rank of destination
tag    message tag

returns: persistent communication request
Creates a persistent communication request for a synchronous mode send. Java binding of the MPI operation \textit{MPI\_SSEND\_INIT}. Further comments as for \textit{Send}.

\begin{verbatim}
Prequest Comm.Rsend_init(Object buf, int offset, int count,
                         Datatype datatype, int dest, int tag)
\end{verbatim}

- \texttt{buf} - send buffer array
- \texttt{offset} - initial offset in send buffer
- \texttt{count} - number of items to send
- \texttt{datatype} - datatype of each item in send buffer
- \texttt{dest} - rank of destination
- \texttt{tag} - message tag

\textit{return}s: persistent communication request

Creates a persistent communication request for a ready mode send. Java binding of the MPI operation \textit{MPI\_RSEND\_INIT}. Further comments as for \textit{Send}.

\begin{verbatim}
Prequest Comm.Recv_init(Object buf, int offset, int count,
                         Datatype datatype, int source, int tag)
\end{verbatim}

- \texttt{buf} - receive buffer array
- \texttt{offset} - initial offset in receive buffer
- \texttt{count} - number of items in receive buffer
- \texttt{datatype} - datatype of each item in receive buffer
- \texttt{source} - rank of source
- \texttt{tag} - message tag

\textit{return}s: persistent communication request

Creates a persistent communication request for a receive operation. Java binding of the MPI operation \textit{MPI\_RECV\_INIT}. Further comments as for \textit{Recv}.

\begin{verbatim}
void Prequest.Start()
\end{verbatim}

Activate a persistent communication request. Java binding of the MPI operation \textit{MPI\_START}. The communication is completed by using the request in one of the operations \textit{Request.Wait}, \textit{Request.Test}, \textit{Request.Waitany}, \textit{Request.Testany}, \textit{Request.Waitall}, \textit{Request.Testall}, \textit{Request.Waitsome}, or \textit{Request.Testsome}. On successful completion the request becomes inactive again. It can be reactivated by a further call to \texttt{Start}.

\begin{verbatim}
static void Prequest.Startall(Prequest [] array_of_requests)
\end{verbatim}
array_of_requests  array of persistent communication requests

Activate a list of communication requests. Java binding of the MPI operation
MPI_{STARTALL}.

3.10 Send-receive

Status Comm.Sendrecv(Object sendbuf, int sendoffset,
    int sendcount, Datatype sendtype,
    int dest, int sendtag,
    Object recvbuf, int recvoffset,
    int recvcount, Datatype recvtype,
    int source, int recvtag)

sendbuf    send buffer array
sendoffset initial offset in send buffer
sendcount  number of items to send
sendtype   datatype of each item in send buffer
dest       rank of destination
sendtag    send tag
recvbuf    receive buffer array
recvoffset initial offset in receive buffer
recvcount  number of items in receive buffer
recvtype   datatype of each item in receive buffer
source     rank of source
recvtag    receive tag

returns:   status object

Execute a blocking send and receive operation. Java binding of the MPI operation
MPI_{SENDRECV}. Further comments as for Send andRecv.

Status Comm.Sendrecv_replace(Object buf, int offset,
    int count, Datatype datatype,
    int dest, int sendtag,
    int source, int recvtag)
buf       buffer array
offset    initial offset in buffer
count     number of items to send
datatype  datatype of each item in buffer
dest      rank of destination
sendtag   send tag
source    rank of source
recvtag   receive tag

returns:  status object

Execute a blocking send and receive operation, receiving message into send
buffer. Java binding of the MPI operation MPI_SENDRECV_REPLACE. Further comments as for send and recv.

3.11  Null processes

The constant MPI_PROC_NULL is available as MPI_PROC_NULL.

3.12  Derived datatypes

In C or Fortran bindings of MPI, derived datatypes have two roles. One is to
allow messages to contain mixed types (for example they allow an integer count
followed by a sequence of real numbers to be passed in a single message). The
other is to allow noncontiguous data to be transmitted. In mpiJava the first
role is abandoned. Any derived type can only include elements of a single basic
type.

Rationale.  In the C binding of MPI, for example, the MPI_TYPE_-
STRUCT constructor for derived types might be used to describe the
physical layout of a struct containing mixed types. This will not work
in Java, because Java does not expose the low-level layout of its objects.
In C and Fortran another use of MPI_TYPE_STRUCT involves incorpo-
rating offsets computed as differences between absolute addresses, so that
parts of a message can come from separately declared entities. It might be
possible to contrive something analogous in a Java binding, somehow en-
coding object references instead of physical addresses. Such a contrivance
is unlikely to be very natural—even in C and Fortran the mechanism is not
particularly elegant. Meanwhile, the effect of either of these applications of
MPI_TYPE_STRUCT can be achieved by using MPI_OBJECT as the buffer
type, and relying on Java object serialization. (End of rationale.)

This leaves description of noncontiguous buffers as the essential role for derived
data types in mpiJava.
Every derived data type constructable in mpiJava has a uniquely defined base type. This is one of the 9 basic types enumerated in section 3.2. Derived types inherit their base types from their precursors in a straightforward way.

In mpiJava a **general datatype** is an object that specifies two things:

- A base type
- A sequence of integer displacements

In contrast to the C and Fortran bindings the displacements are in terms of subscripts in the buffer array argument, *not* byte displacements.

The base types for the predefined MPI datatypes are

<table>
<thead>
<tr>
<th>MPI datatype</th>
<th>base type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI.BYTE</td>
<td>byte</td>
</tr>
<tr>
<td>MPI.CHAR</td>
<td>char</td>
</tr>
<tr>
<td>MPI.SHORT</td>
<td>short</td>
</tr>
<tr>
<td>MPI.BOOLEAN</td>
<td>boolean</td>
</tr>
<tr>
<td>MPI.INT</td>
<td>int</td>
</tr>
<tr>
<td>MPI.LONG</td>
<td>long</td>
</tr>
<tr>
<td>MPI.FLOAT</td>
<td>float</td>
</tr>
<tr>
<td>MPI.DOUBLE</td>
<td>double</td>
</tr>
<tr>
<td>MPI.OBJECT</td>
<td>Object</td>
</tr>
<tr>
<td>MPI.LB</td>
<td>⊥</td>
</tr>
<tr>
<td>MPI.UB</td>
<td>⊥</td>
</tr>
<tr>
<td>MPI.PACKED</td>
<td>byte</td>
</tr>
</tbody>
</table>

The symbol ⊥ is a special undefined value. The displacement sequences for the predefined types (other than MPI.LB, MPI.UB) consist of a single zero.

If the displacement sequence of a datatype is

$$\text{DispSeq} = \{\text{disp}_0, \ldots, \text{disp}_{n-1}\}$$

we define

$$\text{lb}(\text{DispSeq}) = \min_{j} \text{disp}_j,$$

$$\text{ub}(\text{DispSeq}) = \max_{j} (\text{disp}_j + 1), \text{ and}$$

$$\text{extent}(\text{DispSeq}) = \text{ub}(\text{DispSeq}) - \text{lb}(\text{DispSeq})$$

*Rationale.* This definition of the extent differs from the definition in the C or Fortran. It is in units of the buffer array index, *not* in units of bytes.

(End of rationale.)

As discussed at the end of this section, these definitions have to be modified if the type construction involves MPI.LB, MPI.UB.
static Datatype Contiguous(int count, Datatype oldtype)

    count    replication count
oldtype   old datatype

    returns:  new datatype

Construct new datatype representing replication of the old datatype into contiguous locations. Java binding of the MPI operation MPI_TYPE_CONTIGUOUS. The base type of the new datatype is the same as the base type of the old type. Assume the displacement sequence of the old type is

\{disp_0, \ldots, disp_{n-1}\}

with extent \(ex\). Then the new datatype has a displacement sequence with \(count \cdot n\) entries defined by:

\[
\begin{align*}
&\{ disp_0, \ldots, disp_{n-1}, \\
&disp_0 + ex, \ldots, disp_{n-1} + ex, \\
&\ldots, \\
&disp_0 + ex \cdot (count - 1), \ldots, disp_{n-1} + ex \cdot (count - 1) \}
\end{align*}
\]

static Datatype Vector(int count, int blocklength, int stride, Datatype oldtype)

    count    number of blocks
blocklength number of elements in each block
stride      number of elements between start of each block
oldtype    old datatype

    returns:  new datatype

Construct new datatype representing replication of the old datatype into locations that consist of equally spaced blocks. Java binding of the MPI operation MPI_TYPEVECTOR. The base type of the new datatype is the same as the base type of the old type. Assume the displacement sequence of the old type is

\{disp_0, \ldots, disp_{n-1}\}

with extent \(ex\). Let \(bl\) be \(blocklength\). Then the new datatype has a displacement sequence with \(count \cdot bl \cdot n\) entries defined by:

\[
\{ disp_0, \ldots, disp_{n-1} \}.
\]
\begin{verbatim}
disp_0 + ex, \ldots, disp_{n-1} + ex,
\ldots,
disp_0 + ex \cdot (bl - 1), \ldots, disp_{n-1} + ex \cdot (bl - 1),

disp_0 + ex \cdot \text{stride}, \ldots, disp_{n-1} + ex \cdot \text{stride},
disp_0 + ex \cdot (\text{stride} + 1), \ldots, disp_{n-1} + ex \cdot (\text{stride} + 1),
\ldots,
disp_0 + ex \cdot (\text{stride} + bl - 1), \ldots, disp_{n-1} + ex \cdot (\text{stride} + bl - 1),
\ldots,

\text{disp}_0 + ex \cdot \text{stride} \cdot (\text{count} - 1), \ldots, disp_{n-1} + ex \cdot \text{stride} \cdot (\text{count} - 1),
\text{disp}_0 + ex \cdot \text{stride} \cdot (\text{count} - 1) + 1, \ldots,
disp_{n-1} + ex \cdot \text{stride} \cdot (\text{count} - 1) + 1,
\ldots,
\text{disp}_0 + ex \cdot \text{stride} \cdot (\text{count} - 1) + bl - 1, \ldots,
disp_{n-1} + ex \cdot \text{stride} \cdot (\text{count} - 1) + bl - 1)
\end{verbatim}

static Datatype Datatype.Hvector(int count,
    int blocklength, int stride,
    Datatype oldtype)

\begin{description}
\item[count] number of blocks
\item[blocklength] number of elements in each block
\item[stride] number of elements between start of each block
\item[oldtype] old datatype
\end{description}

returns: new datatype

Identical to Vector except that the stride is expressed directly in terms of the
buffer index, rather than the units of the old type. Java binding of the MPI
operation MPI\_TYPE\_HVECTOR. Unlike other language bindings, the value of
\text{stride} is \textit{not} measured in bytes. The displacement sequence of the new type
is:

\begin{verbatim}
\{ \text{disp}_0, \ldots, \text{disp}_{n-1},
\text{disp}_0 + ex, \ldots, \text{disp}_{n-1} + ex,
\ldots,
\text{disp}_0 + ex \cdot (bl - 1), \ldots, \text{disp}_{n-1} + ex \cdot (bl - 1),
\end{verbatim}
static Datatype Datatype.Indexed(int [] array_of_blocklengths, 
    int [] array_of_displacements, 
    Datatype oldtype)

array_of_blocklengths number of elements per block
array_of_displacements displacement of each block in units of old type
oldtype old datatype

returns: new datatype

Construct new datatype representing replication of the old type into a sequence of blocks where each block can contain a different number of copies and have a different displacement. Java binding of the MPI operation MPI_TYPE_INDEXED. The number of blocks is taken to be size of the array_of_blocklengths argument. The second argument, array_of_displacements, should be the same size. The base type of the new datatype is the same as the base type of the old type. Assume the displacement sequence of the old type is

{disp₀, ..., dispₙ₋₁}

with extent ex. Let B be the array_of_blocklengths argument and D be the array_of_displacements argument. Then the new datatype has a displacement sequence with \( n \cdot \sum_{i=0}^{count-1} B[i] \) entries:

\[
\{ disp₀ + D[0] \cdot ex, ..., dispₙ₋₁ + D[0] \cdot ex, \\
      disp₀ + (D[0] + 1) \cdot ex, ..., dispₙ₋₁ + (D[0] + 1) \cdot ex, \\
\}
\]
Here, \( \text{count} \) is the number of blocks.

\[
\begin{align*}
\text{disp}_0 + \text{D}[0] + \text{B}[0] - 1 \cdot \text{ex}, & \ldots, \\
\text{disp}_{n-1} + \text{D}[0] + \text{B}[0] - 1 \cdot \text{ex}, \\
\text{disp}_0 + \text{D}[\text{count} - 1] \cdot \text{ex}, & \ldots, \\
\text{disp}_{n-1} + \text{D}[\text{count} - 1] \cdot \text{ex}, \\
\text{disp}_0 + \text{D}[\text{count} - 1] + \text{B}[\text{count} - 1] - 1 \cdot \text{ex}, & \ldots, \\
\text{disp}_{n-1} + \text{D}[\text{count} - 1] + \text{B}[\text{count} - 1] - 1 \cdot \text{ex}
\end{align*}
\]

Identical to indexed except that the displacements are expressed directly in terms of the buffer index, rather than the units of the old type. Java binding of the MPI operation MPI\_TYPE\_HINDEXED. Unlike other language bindings, the values in \text{array\_of\_displacements} are not measured in bytes. The displacement sequence of the new type is:

\[
\begin{align*}
\{ & \text{disp}_0 + \text{D}[0], \ldots, \\
& \text{disp}_{n-1} + \text{D}[0], \\
& \text{disp}_0 + \text{D}[0] + \text{ex}, \ldots, \\
& \text{disp}_{n-1} + \text{D}[0] + \text{ex}, \\
& \ldots, \\
& \text{disp}_0 + \text{D}[\text{count} - 1], \ldots, \\
& \text{disp}_{n-1} + \text{D}[\text{count} - 1], \\
& \text{disp}_0 + \text{D}[\text{count} - 1] + \text{ex}, \ldots, \\
& \text{disp}_{n-1} + \text{D}[\text{count} - 1] + \text{ex}, \\
& \ldots
\end{align*}
\]
\[ \text{disp}_0 + D[\text{count} - 1] + (B[\text{count} - 1] - 1) \cdot ex, \ldots, \]
\[ \text{disp}_{n-1} + D[\text{count} - 1] + (B[\text{count} - 1] - 1) \cdot ex \]

static Datatype Datatype::Struct(int [] array_of_blocklengths,
                                 int [] array_of_displacements,
                                 Datatype [] array_of_types)

array_of_blocklengths number of elements per block
array_of_displacements displacement in buffer for each block
array_of_types type of elements in each block
returns: new datatype

The most general type constructor. Java binding of the MPI operation MPI_TYPE_STRUCT. The number of blocks is taken to be size of the array_of_blocklengths argument. The second and third arguments, array_of_displacements and array_of_types, should be the same size. Unlike other language bindings, the values in array_of_displacements are not measured in bytes. All elements of array_of_types with definite base types must have the same base type; this will be the base type of new datatype. Let \( T \) be the array_of_types argument. Assume the displacement sequence of the old type \( T[i] \) is

\[ \{ \text{disp}_0^i, \ldots, \text{disp}_{n-1}^i \} \]

with extent \( ex \). Let \( B \) be the array_of_blocklengths argument and \( D \) be the array_of_displacements argument. Then the new datatype has a displacement sequence with \( \sum_{i=0}^{c-1} B[i] \cdot n_i \) entries:

\[ \{ \text{disp}_0^0 + D[0], \ldots, \text{disp}_{n-1}^0 + D[0], \]
\[ \text{disp}_0^0 + D[0] + ex_0, \ldots, \text{disp}_{n-1}^0 + D[0] + ex_0, \]
\[ \ldots, \]
\[ \text{disp}_0^c - 1 + D[c - 1], \ldots, \text{disp}_{n-1}^c - 1 + D[c - 1], \]
\[ \text{disp}_0^c - 1 + D[c - 1] + ex_{c-1}, \ldots, \text{disp}_{n-1}^c - 1 + D[c - 1] + ex_{c-1}, \]
\[ \ldots, \]
\[ \text{disp}_0^{c-1} + D[c - 1] + (B[c - 1] - 1) \cdot ex_{c-1}, \ldots, \]
\[ \text{disp}_{n-1}^{c-1} + D[c - 1] + (B[c - 1] - 1) \cdot ex_{c-1} \} \]

Here, \( c \) is the number of blocks.
If any elements of array_of_types are MPI.LB or MPI.UB, the corresponding displacements are omitted in the displacement sequence. These displacements only affect the computation of Datatype.Lb, Datatype.Ub and DatatypeExtent, as explained below.

**Revised definition of general datatype.** In the presence of MPI.LB, MPI.UB component types, an mpiJava general datatype can be represented by four things:

- A base type
- A sequence, DispSeq, of proper displacements.
- A set, LBDisps, of pseudo-displacements for MPI.LB markers.
- A set, UBDisps, of pseudo-displacements for MPI.UB markers.

For basic datatypes other than MPI.LB, MPI.UB the displacements take the form

\[
\begin{align*}
\text{DispSeq} &= \{0\} \\
\text{LBDisps} &= \emptyset \\
\text{UBDisps} &= \emptyset
\end{align*}
\]

For MPI.LB they are

\[
\begin{align*}
\text{DispSeq} &= \emptyset \\
\text{LBDisps} &= \{0\} \\
\text{UBDisps} &= \emptyset
\end{align*}
\]

For MPI.UB they are

\[
\begin{align*}
\text{DispSeq} &= \emptyset \\
\text{LBDisps} &= \emptyset \\
\text{UBDisps} &= \{0\}
\end{align*}
\]

The two sets of pseudo-displacements are propagated to derived types by formulae identical to the ones given above for proper displacements. Below we will use the definition

\[
\text{AllDisps} = \text{DispSeq} \cup \text{LBDisps} \cup \text{UBDisps}
\]

\[
\text{intDatatypeExtent()}
\]
returns: datatype extent

Returns the extent of a datatype. Java binding of the MPI operation `MPI_TYPE_EXTENT`. Return value is equal to

\[ \text{Ub}() - \text{Lb}() \]

\begin{itemize}
  \item \textbf{int Datatype.Lb()}
  \end{itemize}

\textit{returns:} displacement of lower bound from origin

Find the lower bound of a datatype. Java binding of the MPI operation `MPI_TYPE_LB`. If \textit{LBDisps} is non-empty the return value of \textit{Lb} is the least element of that set. Otherwise it is the least element of \textit{AllDisps}.

\begin{itemize}
  \item \textbf{int Datatype.Ub()}
  \end{itemize}

\textit{returns:} displacement of upper bound from origin

Find the upper bound of a datatype. Java binding of the MPI operation `MPI_TYPE_UB`. If \textit{UBDisps} is non-empty the return value of \textit{Ub} is the greatest element of that set. Otherwise it is

\[ \max_{\text{disp} \in \text{AllDisps}} \text{disp} + 1 \]

\begin{itemize}
  \item \textbf{int Datatype.Size()}
  \end{itemize}

\textit{returns:} datatype size

Returns the total size of the type. Java binding of the MPI operation `MPI_TYPE_SIZE`. Size is defined as the total number of buffer elements incorporated by the data type, or equivalently as the length of the displacement sequence. \textit{Unlike other language bindings, the size is not measured in bytes.}

\begin{itemize}
  \item \textbf{void Datatype.Commit()}
  \end{itemize}

Commit a derived datatype. Java binding of the MPI operation `MPI_TYPE_COMMIT`.

\begin{itemize}
  \item \textbf{void Datatype.finalize()}
  \end{itemize}

\footnote{If \textit{AllDisps} is empty (which could happen for a derived datatype created with replication count of zero, for instance) the results of \textit{Lb}, \textit{Ub} and thus \textit{Extent} are undefined.}
Destructor. Java binding of the MPI operation `MPI_TYPE_FREE`.

```java
int Status.Get_elements(Datatype datatype)
```

**datatype** datatype used by receive operation

**returns:** number of received basic elements

Retrieve number of basic elements from status. Java binding of the MPI operation `MPI_GET_ELEMENTS`.

### 3.13 Pack and unpack

```java
int Comm.Pack(Object inbuf, int offset, int incount,
               Datatype datatype,
               byte [] outbuf, int position)
```

**inbuf** input buffer array

**offset** initial offset in input buffer

**incount** number of items in input buffer

**datatype** datatype of each item in input buffer

**outbuf** output buffer

**position** initial position in output buffer

**returns:** final position in output buffer

Packs message in send buffer `inbuf` into space specified in `outbuf`. Java binding of the MPI operation `MPI_PACK`. The return value is the output value of `position`—the initial value incremented by the number of bytes written.

```java
int Comm.Unpack(byte [] inbuf, int position,
                Object outbuf, int offset, int outcount,
                Datatype datatype)
```

**inbuf** input buffer

**position** initial position in input buffer

**outbuf** output buffer array

**offset** initial offset in output buffer

**outcount** number of items in output buffer

**datatype** datatype of each item in output buffer

**returns:** final position in input buffer
Unpacks message in receive buffer `outbuf` into space specified in `inbuf`. Java binding of the MPI operation `MPI.UNPACK`. The return value is the output value of `position`—the initial value incremented by the number of bytes read.

```java
int Comm.Pack_size(int incount, Datatype datatype)
```

- `incount`: number of items in input buffer
- `datatype`: datatype of each item in input buffer

*returns:* upper bound on size of packed message

Returns an upper bound on the increment of `position` effected by `pack`. Java binding of the MPI operation `MPI.PACK_SIZE`. *It is an error to call this function if the base type of `datatype` is `MPI.OBJECT`.*
4 Collective Communication

4.1 Introduction and Overview

In general the mpiJava bindings of collective communication operations realize
the MPI functions as members of the IntraComm class.

4.2 Communicator argument

No special issues for Java binding.

4.3 Barrier synchronization

void Intracomm.Barrier()

A call to Barrier blocks the caller until all processes in the group have called
it. Java binding of the MPI operation MPI_BARRIER.

4.4 Broadcast

void Intracomm.Bcast(Object buffer, int offset, int count,
Datatype datatype, int root)

buf buffer array
offset initial offset in buffer
count number of items in buffer
datatype datatype of each item in buffer
dest rank of broadcast root

Broadcast a message from the process with rank root to all processes of the
group. Java binding of the MPI operation MPI_BCAST.

4.5 Gather

void Intracomm.Gather(Object sendbuf, int sendoffset,
int sendcount, Datatype sendtype,
Object recvbuf, int recvoffset,
int recvcount, Datatype recvtype, int root)
Each process sends the contents of its send buffer to the root process. Java binding of the MPI operation `MPI_GATHER`.

```java
void Intracomm.Gatherv(Object sendbuf, int sendoffset,
                        int sendcount, Datatype sendtype,
                        Object recvbuf, int recvoffset,
                        int [] recvcounts, int [] displs,
                        Datatype recvtype, int root)
```

Extends functionality of `Gather` by allowing varying counts of data from each process. Java binding of the MPI operation `MPI_GATHERV`. The sizes of arrays `recvcounts` and `displs` should be the size of the group. Entry $i$ of `displs` specifies the displacement relative to element `recvoffset` of `recvbuf` at which to place incoming data. Note that if `recvtype` is a derived data type, elements of `displs` are in units of the derived type extent, (unlike `recvoffset`, which is a direct index into the buffer array).

### 4.6 Scatter

```java
void Intracomm.Scatter(Object sendbuf, int sendoffset,
                        int sendcount, Datatype sendtype,
                        Object recvbuf, int recvoffset,
```


int recvcount, Datatype recvtype, int root)

void Intracomm.Scatterv(Object sendbuf, int sendoffset,
                        int [] sendcounts, int [] displs,
                        Datatype sendtype,
                        Object recvbuf, int recvoffset,
                        int recvcount, Datatype recvtype, int root)

void Intracomm.Allgather(Object sendbuf, int sendoffset,
                          int sendcount, Datatype sendtype,
                          Object recvbuf, int recvoffset,
                          int recvcount, Datatype recvtype)

Inverse of the operation Gather. Java binding of the MPI operation MPI_SCATTER.

4.7 Gather-to-all
sendbuf  send buffer array
sendoffset initial offset in send buffer
sendcount number of items sent to each process
sendtype datatype of send buffer items
recvbuf  receive buffer array
recvoffset initial offset in receive buffer
recvcount number of items in receive buffer
recvtype datatype of receive buffer items

Similar to Gather, but all processes receive the result. Java binding of the MPI operation MPI\_ALLGATHER.

```java
void Intracomm.Allgatherv(Object sendbuf, int sendoffset,  
    int sendcount, Datatype sendtype,  
    Object recvbuf, int recvoffset,  
    int [] recvcounts, int [] displs,  
    Datatype recvtype)
```

sendbuf  send buffer array
sendoffset initial offset in send buffer
sendcount number of items to send
sendtype  datatype of each item in send buffer
recvbuf receive buffer array
recvoffset initial offset in receive buffer
recvcounts number of elements received from each process
displs displacements at which to place incoming data
recvtype  datatype of each item in receive buffer

Similar to Gather, but all processes receive the result. Java binding of the MPI operation MPI\_GATHERV.

4.8 All-to-All Scatter/Gather

```java
void Intracomm.ALLTOALL(Object sendbuf, int sendoffset,  
    int sendcount, Datatype sendtype,  
    Object recvbuf, int recvoffset,  
    recvcount, Datatype recvtype)
```
void Intracomm.Alltoallv(Object sendbuf, int sendoffset,  
    int[] sendcounts, int[] sdispls,  
    Datatype sendtype,  
    Object recvbuf, int recvoffset,  
    int[] recvcounts, int[] rdispls,  
    Datatype recvtype)  

void Intracomm.Reduce(Object sendbuf, int sendoffset,  
                       Object recvbuf, int recvoffset,  
                       int count, Datatype datatype,  
                       Op op, int root)
sendbuf send buffer array
sendoffset initial offset in send buffer
recvbuf receive buffer array
recvoffset initial offset in receive buffer
count number of items in send buffer
datatype data type of each item in send buffer
op reduce operation
dest rank of root process

Combine elements in input buffer of each process using the reduce operation, and return the combined value in the output buffer of the root process. Java binding of the MPI operation MPI_REDUCE.

The predefined operations are available in Java as MPI_MAX, MPI_MIN, MPI_SUM, MPI_PROD, MPI_BAND, MPI.BOR, MPI.LXOR, MPI.BXOR, MPI.MINLOC and MPI.MAXLOC.

The handling of MINLOC and MAXLOC is modelled on the Fortran binding. The extra predefined types MPI.SHORT2, MPI.INT2, MPI.LONG2, MPI.FLOAT2, MPI.DOUBLE2 describe pairs of Java numeric primitive types.

void Op.initialize()
Destructor. Java binding of the MPI operation \texttt{MPI\_OP\_FREE}.

\begin{verbatim}
void Intracomm.\texttt{Allreduce}(Object sendbuf, int sendoffset,
                          Object recvbuf, int recvoffset,
                          int count, Datatype datatype,
                          Op op)
sendbuf   send buffer array
sendoffset initial offset in send buffer
recvbuf   receive buffer array
recvoffset initial offset in receive buffer
count     number of items in send buffer
datatype  data type of each item in send buffer
op        reduce operation
\end{verbatim}

Same as \texttt{Reduce} except that the result appears in receive buffer of all processes in the group. Java binding of the MPI operation \texttt{MPI\_ALLREDUCE}.

\section*{4.10 \texttt{Reduce-Scatter}}

\begin{verbatim}
void Intracomm.\texttt{Reduce\_scatter}(Object sendbuf, int sendoffset,
                                        Object recvbuf, int recvoffset,
                                        int [] recvcounts,
                                        Datatype datatype, Op op)
sendbuf   send buffer array
sendoffset initial offset in send buffer
recvbuf   receive buffer array
recvoffset initial offset in receive buffer
recvcounts numbers of result elements distributed to each process
datatype  data type of each item in send buffer
op        reduce operation
\end{verbatim}

Combine elements in input buffer of each process using the reduce operation, and scatter the combined values over the output buffers of the processes. Java binding of the MPI operation \texttt{MPI\_REDUCE\_SCATTER}.

\section*{4.11 \texttt{Scan}}

\begin{verbatim}
void Intracomm.\texttt{Scan}(Object sendbuf, int sendoffset,
                          Object recvbuf, int recvoffset,
                          int count, Datatype datatype,
                          Op op)
\end{verbatim}
sendbuf    send buffer array
sendoffset initial offset in send buffer
recvbuf    receive buffer array
recvoffset initial offset in receive buffer
count      number of items in input buffer
datatype   data type of each item in input buffer
op         reduce operation

Perform a prefix reduction on data distributed across the group. Java binding of the MPI operation `MPI_SCAN`.

4.12 Correctness

No special issues for Java binding.
5 Groups, Contexts and Communicators

5.1 Introduction
No special issues for Java binding.

5.2 Basic Concepts
The constant `MPI_GROUP_EMPTY` is available as `MPI_GROUP_EMPTY`. The constants `MPI_COMM_WORLD`, `MPI_COMM_SELF` are available as `MPI_COMM_WORLD`, `MPI_COMM_SELF`.

5.3 Group Management

```java
int Group.Size()
returns:  number of processors in the group
Size of group. Java binding of the MPI operation `MPI_GROUP_SIZE`.

int Group.Rank()
returns:  rank of the calling process in the group
Rank of this process in group. Java binding of the MPI operation `MPI_GROUP_RANK`. Result value is `MPI.UNDEFINED` if this process is not a member of the group.

static int [] Group.Translate_ranks(Group group1, int [] ranks1, 
  Group group2)
  group1  first group
  ranks1  array of valid ranks in group1
  group2  second group
returns:  array of corresponding ranks in group2
Translate ranks within first group to ranks within second group. Java binding of the MPI operation `MPI_GROUP_TRANSLATE_RANKS`. Result elements are `MPI.UNDEFINED` where no correspondence exists.

static int Group.Compare(Group group1, Group group2)
```
group1  first group
group2  second group

returns:  result

Compare two groups. Java binding of the MPI operation MPI\_GROUP\_COMPARE. MPI\_IDENT results if the group members and group order are exactly the same in both groups. MPI\_SIMILAR results if the group members are the same but the order is different. MPI\_UNEQUAL results otherwise.

Group Comm.Group()

returns:  group corresponding to this communicator

Return group associated with a communicator. Java binding of the MPI operation MPI\_COMM\_GROUP.

static Group Group\_Union(Group group1, Group group2)

group1  first group
group2  second group

returns:  union group

Set union of two groups. Java binding of the MPI operation MPI\_GROUP\_UNION.

static Group Group\_Intersection(Group group1, Group group2)

group1  first group
group2  second group

returns:  intersection group

Set intersection of two groups. Java binding of the MPI operation MPI\_GROUP\_INTERSECTION.

static Group Group\_Difference(Group group1, Group group2)

group1  first group
group2  second group

returns:  difference group
Result contains all elements of the first group that are not in the second group. Java binding of the MPI operation \texttt{MPI\_GROUP\_DIFFERENCE}.

\begin{verbatim}
Group Group.Incl(int [] ranks)
  ranks  ranks from this group to appear in new group
  returns: new group

Create a subset group including specified processes. Java binding of the MPI operation \texttt{MPI\_GROUP\_INCL}.

Group Group.Excl(int [] ranks)
  ranks  ranks from this group \textit{not} to appear in new group
  returns: new group

Create a subset group excluding specified processes. Java binding of the MPI operation \texttt{MPI\_GROUP\_EXCL}.

Group Group.Range_incl(int [][] ranges)
  ranges  array of integer triplets
  returns: new group

Create a subset group including processes specified by strided intervals of ranks. Java binding of the MPI operation \texttt{MPI\_GROUP\_RANGE\_INCL}. The triplets are of the form (first rank, last rank, stride) indicating ranks in this group to be included in the new group. The size of the first dimension of \texttt{ranges} is the number of triplets. The size of the second dimension is 3.

Group Group.Range_excl(int [][] ranges)
  ranges  array of integer triplets
  returns: new group

Create a subset group excluding processes specified by strided intervals of ranks. Java binding of the MPI operation \texttt{MPI\_GROUP\_RANGE\_EXCL}. Triplet array is defined as for \texttt{Range\_incl}, the ranges indicating ranks in this group to be excluded from the new group.
\end{verbatim}
void Group.finalize()

Destructor. Java binding of the MPI operation `MPI_GROUP_FREE`.

### 5.4 Communicator Management

**int Comm.Size()**

*returns:* number of processors in the group of this communicator

Size of group of this communicator. Java binding of the MPI operation `MPI_COMM_SIZE`.

**int Comm.Rank()**

*returns:* rank of the calling process in the group of this communicator

Rank of this process in group of this communicator. Java binding of the MPI operation `MPI_COMM_RANK`.

**static int Comm.Compare(Comm comm1, Comm comm2)**

*comm1* first communicator

*comm2* second communicator

*returns:* result

Compare two communicators. Java binding of the MPI operation `MPI_COMM_COMPARE`. `MPI_IDENT` results if the `comm1` and `comm2` are references to the same object (i.e., if `comm1 == comm2`). `MPI_CONGRUENT` results if the underlying groups are identical but the communicators differ by context. `MPI_SIMILAR` results if the underlying groups are similar but the communicators differ by context. `MPI_UNEQUAL` results otherwise.

**Object Comm.clone()**

*returns:* copy of this communicator

Duplicate this communicator. Java binding of the MPI operation `MPI_COMM_DUP`. The new communicator is "congruent" to the old one, but has a different context.
**Rationale.** The decision to use the standard Java clone method means the static result type must be `Object`. The dynamic type will be that of the `Comm` subclass of the parent. MPI-defined and user-defined subclasses of `Comm` will generally override `clone` to ensure all relevant attributes are copied. *(End of rationale.)*

**Intracomm Intracomm.Create(Group group)**

- ***group***: group which is a subset of the group of this communicator

- **returns**: new communicator

Create a new communicator. Java binding of the MPI operation `MPI_COMM_CREATE`.

**Intracomm Intracomm.Split(int color, int key)**

- ***color***: control of subset assignment
- ***key***: control of rank assignment

- **returns**: new communicator

Partition the group associated with this communicator and create a new communicator within each subgroup. Java binding of the MPI operation `MPI_COMM_SPLIT`.

**void Comm.Free()**

Destroy this communicator. Java binding of the MPI operation `MPI_COMM_FREE`.

**Rationale.** An explicitly called `Free` method is required rather than an implicitly called `finalize` method, because `MPI_COMM_FREE` is a collective operation. We cannot assume that the Java garbage collector will call a `finalize` method synchronously on all processors. *(End of rationale.)*

**boolean Comm.Is_null()**

- **returns**: true if the communicator object has been freed, false otherwise

Replaces comparison with `MPI_COMM_NULL`. 
5.5 Motivating Examples

No special issues for Java binding.

5.6 Inter-Communication

boolean Comm.Test_inter()

returns: true if this is an inter-communicator, false otherwise

Test if this communicator is an inter-communicator. Java binding of the MPI operation MPI_COMM_TEST_INTER.

int Intercomm.Remote_size()

returns: number of process in remote group of this communicator

Size of remote group. Java binding of the MPI operation MPI_COMM_REMOTE_SIZE.

Group Intercomm.Remote_Group()

returns: remote group of this communicator

Return the remote group. Java binding of the MPI operation MPI_COMM_REMOTE_GROUP.

Intercomm Comm.Create_intercomm(Comm local_comm, int local_leader, int remote_leader, int tag)

local_comm  local intra-communicator
local_leader rank of local group leader in localComm
remote_leader rank of remote group leader in this communicator
tag  “safe” tag

returns: new inter-communicator

Create an inter-communicator. Java binding of the MPI operation MPI_INTER_COMM_CREATE.

Rationale. This operation is defined as a method on the “peer communicator”, making it analogous to a Send orRecv communication with the remote group leader. (End of rationale.)

47
**Intracomm Intercomm.Merge(boolean high)**

**high** true if the local group has higher ranks in combined group

**returns:** new intra-communicator

Create an intra-communicator from the union of the two groups in the inter-communicator. Java binding of the MPI operation `MPI_INTERCOMM_MERGE`.

### 5.7 Caching

It is assumed that to achieve the effect of caching attributes in user-customized communicators programmers will create subclasses of the library-defined communicator classes with suitable additional fields. These fields may be copied or deleted by suitably overridden `clone` and `finalize` methods.

Hence the only “caching” operation surviving here is the binding of `MPI_ATTR_GET`, which is needed to access values of attributes predefined by the implementation. According the standard, the key values for such attributes include `MPI.TAG`, `MPI.HOST`, `MPI.IO` and `MPI.WTIME_IS_GLOBAL`.

**int Comm.Attr_get(int keyval)**

**keyval** one of the key values predefined by the implementation

**returns:** attribute value

Retrieves attribute value by key. Java binding of the MPI operation `MPI_ATTR_GET`.
6 Process Topologies

6.1 Introduction

Communicators with Cartesian or graph topologies will be realized as instances of the subclasses Cartcomm or Graphcomm, respectively of Intracomm.

6.2 Virtual Topologies

No special issues for Java binding.

6.3 Embedding in MPI

No special issues for Java binding.

6.4 Overview of the Functions

No special issues for Java binding.

6.5 Topology Constructors

Cartcomm Intracomm.Create_cart(int [] dims, boolean [] periods, boolean reorder)

- dims the number of processes in each dimension
- periods true if grid is periodic, false if not, in each dimension
- reorder true if ranking may be reordered, false if not

returns: new Cartesian topology communicator

Create a Cartesian topology communicator whose group is a subset of the group of this communicator. Java binding of the MPI operation MPI_CART_CREATE. The number of dimensions of the Cartesian grid is taken to be the size of the dims argument. The array periods must be the same size.

static Cartcomm.Dims_create(int nnodes, int [] dims)

- nnodes number of nodes in a grid
- dims array specifying the number of nodes in each dimension

Select a balanced distribution of processes per coordinate direction. Java binding of the MPI operation MPI_DIMS_CREATE. Number of dimensions is the size of is dims. Note that dims is an inout parameter.
Graphcomm Intracomm.Create_graph(int [] index, int [] edges,
   boolean reorder)

  index   node degrees
  edges   graph edges
  reorder true if ranking may be reordered, false if not

  returns: new graph topology communicator

Create a graph topology communicator whose group is a subset of the group of
this communicator. Java binding of the MPI operation MPI_GRAPH_CREATE.
The number of nodes in the graph, nnodes, is taken to be size of the index
argument. The size of array edges must be index[nnodes - 1].

int Comm.Topo_test()

  returns: topology type of communicator

Returns the type of topology associated with the communicator. Java bind-
ing of the MPI operation MPI_TOPO_TEST. The return value will be one of
MPI_GRAPH, MPI_CART or MPI.UNDEFINED.

GraphParms Graphcomm.Get()

  returns: object defining node degrees and edges of graph

Returns graph topology information. Java binding of the MPI operations MPI-
GRAPHDIMS_GET and MPI_GRAPH_GET. The class of the returned object is

   public class GraphParms {
     public int [] index ;
     public int [] edges ;
   }

The number of nodes and number of edges can be extracted from the sizes of
the index and edges arrays.

CartParms Cartcomm.Get()

  returns: object containing dimensions, periods and local coordinates

Returns Cartesian topology information. Java binding of the MPI operations
MPI_CARTDIM_GET and MPI_CART_GET. The class of the returned object is
public class CartParms {
    public int [] dims ;
    public boolean [] periods ;
    public int [] coords ;
}

The number of dimensions can be obtained from the size of (eg) the dims array.

Rationale. The inquiries MPI_GRAPHDIMS_GET, MPI_GRAPH_GET, MPI_CART_DIMS_GET, and MPI_CART_GET are unusual in returning multiple independent values from single calls. This is a problem in Java. The Java binding could split these inquiries into several independent ones, but that would complicate JNI-based wrapper implementations. Hence we introduced the auxiliary classes GraphParms and CartParms to hold multiple results. (End of rationale.)

int Cartcomm.Rank(int [] coords)
    coords Cartesian coordinates of a process
    returns: rank of the specified process

Translate logical process coordinates to process rank. Java binding of the MPI operation MPI_CART_RANK.

int [] Cartcomm.Cords(int rank)
    coords rank of a process
    returns: Cartesian coordinates of the specified process

Translate process rank to logical process coordinates. Java binding of the MPI operation MPI_CART_COORDS.

int [] Graphcomm.Neighbours(int rank)
    coords rank of a process in the group of this communicator
    returns: array of ranks of neighbouring processes to one specified

Provides adjacency information for general graph topology. Java binding of the MPI operations MPI_GRAPH_NEIGHBOURS_COUNT and MPI_GRAPH_NEIGHBOURS. The number of neighbours can be extracted from the size of the result.
ShiftParms Cartcomm.Shift(int direction, int disp)

direction  coordinate dimension of shift
disp      displacement

returns:  object containing ranks of source and destination processes

Compute source and destination ranks for "shift" communication. Java binding of the MPI operation MPI_CARTSHIFT. The class of the returned object is

```java
public class ShiftParms {
    public int rankSource ;
    public int rankDest ;
}
```

Cartcomm Cartcomm.Sub(boolean [] remainDims)

remainDims  by dimension, true if dimension is to be kept, false otherwise

returns:  communicator containing subgrid including this process

Partition Cartesian communicator into subgroups of lower dimension. Java binding of the MPI operation MPI_CARTSUB.

int Cartcomm.Map(int [] dims, boolean [] periods)

dims     the number of processes in each dimension
periods  true if grid is periodic, false if not, in each dimension

returns:  reordered rank of calling process

Compute an optimal placement. Java binding of the MPI operation MPI_CART_MAP. The number of dimensions is taken to be size of the dims argument.

int Graphcomm.Map(int [] index, int [] edges)

index    node degrees
edges    graph edges

returns:  reordered rank of calling process

Compute an optimal placement. Java binding of the MPI operation MPI_GRAPH_MAP. The number of nodes is taken to be size of the index argument.
7 MPI Environmental Management

7.1 Implementation information

The constants $\text{MPI\_TAG\_UB}$, $\text{MPI\_HOST}$ and $\text{MPI\_IO}$ are available as $\text{MPI\_\text{-}\_\text{-}\_\text{-}}\text{TAG\_UB, MPI\_HOST, MPI\_IO}$.

```java
static String MPI\_Get\_processor\_name()

    returns: A unique specifier for the actual node.

Returns the name of the processor on which it is called. Java binding of the MPI operation $\text{MPI\_GET\_PROCESSOR\_NAME}$. 

7.2 Error handling

The constants $\text{MPI\_ERRORS\_ARE\_FATAL}$, $\text{MPI\_ERRORS\_RETURN}$ are available as $\text{MPI\_ERRORS\_ARE\_FATAL, MPI\_ERRORS\_RETURN}$.

If the effective error handler is $\text{MPI\_ERRORS\_RETURN}$, the wrapper codes will throw appropriate Java exceptions (see section 7.3).

Currently mpiJava omits an interface for creating new MPI error handlers (the detailed interface of the handler function depends on unstandardized features of the MPI implementation).

```java
static void Comm\_Errorhandler\_set(Errhandler errhandler)

    errhandler new MPI error handler for communicator

Associates a new error handler with communicator at the calling process. Java binding of the MPI operation $\text{MPI\_ERRORHANDLER\_SET}$. 

```java
static Errhandler Comm\_Errorhandler\_get()

    returns: MPI error handler currently associated with communicator

Returns the error handler currently associated with the communicator. Java binding of the MPI operation $\text{MPI\_ERRORHANDLER\_GET}$. 

7.3 Error codes and classes

The $\text{MPIException}$ subclasses

```java
```
MPIErrBuffer
MPIErrCount
MPIErrType
MPIErrTag
MPIErrComm
MPIErrRank
MPIErrRequest
MPIErrRoot
MPIErrGroup
MPIErrOp
MPIErrTopology
MPIErrDims
MPIErrArg
MPIErrUnknown
MPIErrTruncate
MPIErrOther
MPIErrIntern

correspond to the standard MPI error classes. [Not implemented in the current release.]

7.4 Timers

static double MPI.Wtime()

    returns: elapsed wallclock time in seconds since some time in the past

Returns wallclock time. Java binding of the MPI operation MPI_WTIME.

static double MPI.Wtick()

    returns: resolution of wtime in seconds.

Returns resolution of timer. Java binding of the MPI operation MPI_WTICK.

7.5 Startup

static String [] MPI.Init(String[] argv)

    argv arguments to main method.

    returns: command line arguments returned by MPI.

Initialize MPI. Java binding of the MPI operation MPI_INIT.
static void MPI.Finalize()

Finalize MPI. Java binding of the MPI operation \textit{MPI\_FINALIZE}.

static boolean MPI.Initialized()

\textit{return}: \texttt{true} if \texttt{init} has been called, \texttt{false} otherwise.

Test if MPI has been initialized. Java binding of the MPI operation \textit{MPI\_INITIALIZED}.

void Comm.Abort(int errorcode)

\texttt{errorcode} error code for Unix or POSIX environments

Abort MPI. Java binding of the MPI operation \textit{MPI\_ABORT}. 

55
8 Full public interface of classes

Section names appearing in comments refer to the preceding appendix. Specification of the methods immediately following those comments should be found in the referenced section.
8.1 MPI

public class MPI {
    public static Intracomm COMM_WORLD;

    public static Datatype BYTE, CHAR, SHORT, BOOLEAN, INT, LONG,
    FLOAT, DOUBLE, OBJECT, PACKED, LB, UB ;

    public static int ANY_SOURCE, ANY_TAG ;
    public static int PROC_NULL ;
    public static int BSEND_OVERHEAD ;
    public static int UNDEFINED ;
    public static Op MAX, MIN, SUM, PROD, LAND, BAND,
    LOR, BOR, LXOR, BXOR, MINLOC, MAXLOC ;
    public static Datatype SHORT2, INT2, LONG2, FLOAT2, DOUBLE2 ;
    public static Group GROUP_EMPTY ;
    public static Comm COMM_SELF ;
    public static int IDENT, CONGRUENT, SIMILAR, UNEQUAL ;
    public static int GRAPH, CART ;
    public static ErrHandler ERRORS_ARE_FATAL, ERRORS_RETURN ;
    public static int TAG_UB, HOST, IO ;

    // Buffer allocation and usage
    public static void Buffer_attach(byte [] buffer)
        throws MPIException {...}

    public static byte [] Buffer_detach() throws MPIException {...}

    // Environmental Management
    public static String [] Init(String [] argv) throws MPIException {...}

    public static void Finalize() throws MPIException {...}

    57
public static String Get_processor_name() throws MPIException {...}

public static double Wtime() {...}

public static double Wtick() {...}

public static boolean Initialized() throws MPIException {...}

...

}

public class Errhandler {

...

}
8.2 Comm

public class Comm {

    // Communicator Management
    public int Size() throws MPIException {...}
    public int Rank() throws MPIException {...}
    public Group Group() throws MPIException {...}
        // (see "Group management")
    public static int Compare(Comm comm1, Comm comm2)
        throws MPIException {...}
    public Object clone() {...}
    public void Free() throws MPIException {...}
    public boolean Is_null() {...}

    // Inter-communication
    public boolean Test_inter() throws MPIException {...}
    public Intercomm Create_intercomm(Comm local_comm, int local_leader,
            int remote_leader, int tag)
        throws MPIException {...}

    // Caching
    public Object Attr_get(int keyval) throws MPIException {...}

    // Blocking Send and Receive operations
    public void Send(Object buf, int offset, int count,
            Datatype datatype, int dest, int tag)
        throws MPIException {...}
    public Status Recv(Object buf, int offset, int count,
            Datatype datatype, int source, int tag)
        throws MPIException {...}

    // Communication Modes
public void Bsend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public void Ssend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public void Rsend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

// Nonblocking communication
public Request Isend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public Request Ibsend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public Request Issend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public Request Irsend(Object buf, int offset, int count,
        Datatype datatype, int dest, int tag)
        throws MPIException {...}

public Request Irecv(Object buf, int offset, int count,
        Datatype datatype, int source, int tag)
        throws MPIException {...}

// Probe and cancel
public Status Iprobe(int source, int tag) throws MPIException {...}
public Status Probe(int source, int tag) throws MPIException {...}

// Persistent communication requests
public Request Send_init(Object buf, int offset, int count,
Datatype datatype, int dest, int tag
throws MPIException {...}

public Prequest Bsend_init(Object buf, int offset, int count,
Datatype datatype, int dest, int tag)
throws MPIException {...}

public Prequest Ssend_init(Object buf, int offset, int count,
Datatype datatype, int dest, int tag)
throws MPIException {...}

public Prequest Rsend_init(Object buf, int offset, int count,
Datatype datatype, int dest, int tag)
throws MPIException {...}

public Prequest Recv_init(Object buf, int offset, int count,
Datatype datatype, int source, int tag)
throws MPIException {...}

// Send-receive

public Status Sendrecv(Object sendbuf, int sendoffset, int sendcount,
Datatype sendtype, int dest, int sendtag,
Object recvbuf, int recvoffset, int recvcount,
Datatype recvtype, int source, int recvtag)
throws MPIException {...}

public Status Sendrecv_replace(Object buf, int offset, int count,
Datatype datatype, int dest, int sendtag,
int source, int recvtag)
throws MPIException {...}

// Pack and unpack

public int Pack(Object inbuf, int offset, int incount,
Datatype datatype,
byte [] outbuf, int position)
throws MPIException {...}

public int Unpack(byte [] inbuf, int position,
Object outbuf, int offset, int outcount,
Datatype datatype) throws MPIException {...}
public int Pack_size(int incount, Datatype datatype)
    throws MPIException {...}

// Process Topologies
int Topo_test() throws MPIException {...}

// Environmental Management
public void Errorhandler_set(Errhandler errhandler)
    throws MPIException {...}

public Errhandler Errorhandler_get() throws MPIException {...}

void Abort(int errorcode) throws MPIException {...}

...
8.3 Intracomm and Intercomm

public class Intracomm extends Comm {

    public Object clone() { ... }

    public Intracomm Create(Group group) throws MPIException {...}

    public Intracomm Split(int colour, int key) throws MPIException {...}

    // Collective communication

    public void Barrier() throws MPIException {...}

    public void Bcast(Object buffer, int offset, int count,
                       Datatype datatype, int root)
                        throws MPIException {...}

    public void Gather(Object sendbuf, int sendoffset,
                        int sendcount, Datatype sendtype,
                        Object recvbuf, int recvoffset,
                        int recvcount, Datatype recvtype, int root)
                        throws MPIException {...}

    public void Gatherv(Object sendbuf, int sendoffset,
                         int sendcount, Datatype sendtype,
                         Object recvbuf, int recvoffset,
                         int recvcount, Datatype recvtype, int root)
                         throws MPIException {...}

    public void Scatter(Object sendbuf, int sendoffset,
                         int sendcount, Datatype sendtype,
                         Object recvbuf, int recvoffset,
                         int recvcount, Datatype recvtype, int root)
                         throws MPIException {...}

    public void Scatterv(Object sendbuf, int sendoffset,
                          int[] sendcount, int[] displs,
                          Datatype sendtype,
                          Object recvbuf, int recvoffset,
                          int recvcount, Datatype recvtype, int root)
                          throws MPIException {...}

    public void Allgather(Object sendbuf, int sendoffset,
                           int sendcount, Datatype sendtype,
                           Object recvbuf, int recvoffset,
                           int recvcount, Datatype recvtype)

public void Allgatherv(Object sendbuf, int sendoffset,
int sendcount, Datatype sendtype,
Object recvbuf, int recvoffset,
int[] recvcounts, int[] displs,
Datatype recvtype) throws MPIException {...}

public void Alltoall(Object sendbuf, int sendoffset,
int sendcount, Datatype sendtype,
Object recvbuf, int recvoffset,
int[] recvcounts, Datatype recvtype)
throws MPIException {...}

public void Alltoallv(Object sendbuf, int sendoffset,
int[] sendcount, int[] sdispls,
Datatype sendtype,
Object recvbuf, int recvoffset,
int[] recvcount, int[] rdispls,
Datatype recvtype) throws MPIException {...}

public void Reduce(Object sendbuf, int sendoffset,
Object recvbuf, int recvoffset,
int count, Datatype datatype,
Op op, int root) throws MPIException {...}

public void Allreduce(Object sendbuf, int sendoffset,
Object recvbuf, int recvoffset,
int count, Datatype datatype,
Op op) throws MPIException {...}

public void Reduce_scatter(Object sendbuf, int sendoffset,
Object recvbuf, int recvoffset,
int[] recvcounts, Datatype datatype,
Op op) throws MPIException {...}

public void Scan(Object sendbuf, int sendoffset,
Object recvbuf, int recvoffset,
int count, Datatype datatype,
Op op) throws MPIException {...}

// Topology Constructors

public Graphcomm Create_graph(int[] index, int[] edges,
boolean reorder)
throws MPIException {...}
public Cartcomm Create_cart(int dims, boolean periods, 
    boolean reorder) 
    throws MPIException {...}

... 
}

public class Intercomm extends Comm {
    
    public Object clone() { ... }
    
    // Inter-communication
    
    public int Remote_size() throws MPIException {...}
    
    public Group Remote_group() throws MPIException {...}
    
    public Intracomm Merge(boolean high) throws MPIException {...}
    
    ...
}

65
8.4 Op

public class Op {
    Op(User_function function, boolean commute) throws MPIException {
        ...
    }
    void finalize() throws MPIException {
        ...
    }
    ...
}

8.5 Group

public class Group {

    // Group Management

    public int Size() throws MPIException {...}
    public int Rank() throws MPIException {...}
    public static int Translate_ranks(Group group1, int[] ranks1, Group group2)
        throws MPIException {...}
    public static int Compare(Group group1, Group group2)
        throws MPIException {...}
    public static Group Union(Group group1, Group group2)
        throws MPIException {...}
    public static Group Intersection(Group group1, Group group2)
        throws MPIException {...}
    public static Group Difference(Group group1, Group group2)
        throws MPIException {...}
    public Group Incl(int[] ranks) throws MPIException {...}
    public Group Excl(int[] ranks) throws MPIException {...}
    public Group Range_incl(int[][] ranges) throws MPIException {...}
    public Group Range_excl(int[][] ranges) throws MPIException {...}
    public void finalize() throws MPIException {...}

...}

67
8.6 Status

public class Status {
    public int source;
    public int tag;
    public int index;
    // Blocking Send and Receive operations
    public int Get_count(Datatype datatype) throws MPIException {...}
    // Probe and Cancel
    public boolean Test_cancelled() throws MPIException {...}
    // Derived datatypes
    public int Get_elements(Datatype datatype) throws MPIException {...}
    ...
}

8.7 Request and Prequest

public class Request {

   // Nonblocking communication

   public Status Wait() throws MPIException {...}
   public Status Test() throws MPIException {...}
   public void Free() throws MPIException {...}
   public boolean Is_null() {...}

   public static Status Waitany(Request[] array_of_requests)
       throws MPIException {...}

   public static Status Testany(Request[] array_of_requests)
       throws MPIException {...}

   public static Status[] Waitall(Request[] array_of_requests)
       throws MPIException {...}

   public static Status[] Testall(Request[] array_of_requests)
       throws MPIException {...}

   public static Status[] Waitsome(Request[] array_of_requests)
       throws MPIException {...}

   public static Status[] Testsome(Request[] array_of_requests)
       throws MPIException {...}

   // Probe and cancel

   public void Cancel() throws MPIException {...}

   ...
}

public class Prequest extends Request {

   // Persistent communication requests

   public void Start() throws MPIException {...}

   public static void Startall(Prequest[] array_of_requests)
       throws MPIException {...}

69
...
public class Datatype {

    // Derived datatypes

    public static Datatype Contiguous(int count, Datatype oldtype)
        throws MPIException {...}

    public static Datatype Vector(int count, int blocklength, int stride, Datatype oldtype)
        throws MPIException {...}

    public static Datatype Hvector(int count, int blocklength, int stride, Datatype oldtype)
        throws MPIException {...}

    public static Datatype Indexed(int [] array_of_blocklengths,
        int [] array_of_displacements,
        Datatype oldtype)
        throws MPIException {...}

    public static Datatype Hindexed(int [] array_of_blocklengths,
        int [] array_of_displacements,
        Datatype oldtype)
        throws MPIException {...}

    public static Datatype Struct(int [] array_of_blocklengths,
        int [] array_of_displacements,
        Datatype [] array_of_types)
        throws MPIException {...}

    public int Extent() throws MPIException {...}

    public int Lb() throws MPIException {...}

    public int Ub() throws MPIException {...}

    public int Size() throws MPIException {...}

    public void Commit() throws MPIException {...}

    public void finalize() throws MPIException {...}

    ...
}
8.9 Classes for virtual topologies

public class Cartcomm extends Intracomm {
    public Object clone() { ... }

    // Topology Constructors

    static public Dims create(int nnodes, int [] dims)
                           throws MPIException { ... }*

    public CartParms Get() throws MPIException { ... }

    public int Rank(int [] coords) throws MPIException { ... }

    public int [] Coords(int rank) throws MPIException { ... }

    public ShiftParms Shift(int direction, int disp) throws MPIException { ... }

    public Cartcomm Sub(boolean [] remainDims) throws MPIException { ... }

    public int Map(int [] dims, boolean [] periods) throws MPIException { ... }
}

public class CartParms {
    // Return type for Cartcomm.get()

    public int [] dims;
    public boolean [] periods;
    public int [] coords;
}

public class ShiftParms {
    // Return type for Cartcomm.shift()

    public int rankSource;
    public int rankDest;
}

public class Graphcomm extends Intracomm {
    public Object clone() { ... }

    // Topology Constructors

    public GraphParms Get() throws MPIException { ... }
public int Neighbours(int rank) throws MPIException {...}

public int Map(int[] index, int[] edges) throws MPIException {...}
}

public class GraphParms {

    // Return type for Graphcomm.get()

    public int[] index;
    public int[] edges;
}


References

