

*D R A F T*

Document for a Standard Message-Passing Interface

Message Passing Interface Forum

February 3, 2015

This work was supported in part by NSF and ARPA under NSF contract CDA-9115428 and Esprit under project HPC Standards (21111).

This is the result of a LaTeX run of a draft of a single chapter of the MPIF Final Report document.

# Chapter 4

## Datatypes

Basic datatypes were introduced in Section 3.2.2 and in Section 3.3. In this chapter, this model is extended to describe any data layout. We consider general datatypes that allow one to transfer efficiently heterogeneous and noncontiguous data. We conclude with the description of calls for explicit packing and unpacking of messages.

### 4.1 Derived Datatypes

Up to here, all point to point communications have involved only buffers containing a sequence of identical basic datatypes. This is too constraining on two accounts. One often wants to pass messages that contain values with different datatypes (e.g., an integer count, followed by a sequence of real numbers); and one often wants to send noncontiguous data (e.g., a sub-block of a matrix). One solution is to pack noncontiguous data into a contiguous buffer at the sender site and unpack it at the receiver site. This has the disadvantage of requiring additional memory-to-memory copy operations at both sites, even when the communication subsystem has scatter-gather capabilities. Instead, MPI provides mechanisms to specify more general, mixed, and noncontiguous communication buffers. It is up to the implementation to decide whether data should be first packed in a contiguous buffer before being transmitted, or whether it can be collected directly from where it resides.

The general mechanisms provided here allow one to transfer directly, without copying, objects of various shapes and sizes. It is not assumed that the MPI library is cognizant of the objects declared in the host language. Thus, if one wants to transfer a structure, or an array section, it will be necessary to provide in MPI a definition of a communication buffer that mimics the definition of the structure or array section in question. These facilities can be used by library designers to define communication functions that can transfer objects defined in the host language — by decoding their definitions as available in a symbol table or a dope vector. Such higher-level communication functions are not part of MPI.

More general communication buffers are specified by replacing the basic datatypes that have been used so far with derived datatypes that are constructed from basic datatypes using the constructors described in this section. These methods of constructing derived datatypes can be applied recursively.

A *general datatype* is an opaque object that specifies two things:

- A sequence of basic datatypes
- A sequence of integer (byte) displacements

1 The displacements are not required to be positive, distinct, or in increasing order.  
 2 Therefore, the order of items need not coincide with their order in store, and an item may  
 3 appear more than once. We call such a pair of sequences (or sequence of pairs) a *type*  
 4 *map*. The sequence of basic datatypes (displacements ignored) is the *type signature* of the  
 5 datatype.

6 Let

$$7 \quad \text{Typemap} = \{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

9 be such a type map, where  $type_i$  are basic types, and  $disp_i$  are displacements. Let

$$10 \quad \text{Typesig} = \{type_0, \dots, type_{n-1}\}$$

12 be the associated type signature. This type map, together with a base address `buf`, specifies  
 13 a communication buffer: the communication buffer that consists of  $n$  entries, where the  
 14  $i$ -th entry is at address `buf + dispi` and has type  $type_i$ . A message assembled from such a  
 15 communication buffer will consist of  $n$  values, of the types defined by *Typesig*.

16 Most datatype constructors have replication count or block length arguments. Allowed  
 17 values are non-negative integers. If the value is zero, no elements are generated in the type  
 18 map and there is no effect on datatype bounds or extent.

19 We can use a handle to a general datatype as an argument in a send or receive operation,  
 20 instead of a basic datatype argument. The operation `MPI_SEND(buf, 1, datatype, ...)` will  
 21 use the send buffer defined by the base address `buf` and the general datatype associated  
 22 with `datatype`; it will generate a message with the type signature determined by the `datatype`  
 23 argument. `MPI_RECV(buf, 1, datatype, ...)` will use the receive buffer defined by the base  
 24 address `buf` and the general datatype associated with `datatype`.

25 General datatypes can be used in all send and receive operations. We discuss, in  
 26 Section 4.1.11, the case where the second argument `count` has value  $> 1$ .

27 The basic datatypes presented in Section 3.2.2 are particular cases of a general datatype,  
 28 and are predefined. Thus, `MPI_INT` is a predefined handle to a datatype with type map  
 29  $\{(\text{int}, 0)\}$ , with one entry of type `int` and displacement zero. The other basic datatypes  
 30 are similar.

31 The *extent* of a datatype is defined to be the span from the first byte to the last byte  
 32 occupied by entries in this datatype, rounded up to satisfy alignment requirements. That  
 33 is, if

$$34 \quad \text{Typemap} = \{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

36 then

$$37 \quad \begin{aligned} 38 \quad lb(\text{Typemap}) &= \min_j disp_j, \\ 39 \quad ub(\text{Typemap}) &= \max_j (disp_j + \text{sizeof}(type_j)) + \epsilon, \text{ and} \\ 40 \quad extent(\text{Typemap}) &= ub(\text{Typemap}) - lb(\text{Typemap}). \end{aligned} \quad (4.1)$$

41 If  $type_j$  requires alignment to a byte address that is a multiple of  $k_j$ , then  $\epsilon$  is the least  
 42 non-negative increment needed to round  $extent(\text{Typemap})$  to the next multiple of  $\max_j k_j$ .  
 43 In Fortran, it is implementation dependent whether the MPI implementation computes  
 44 the alignments  $k_j$  according to the alignments used by the compiler in common blocks,  
 45 `SEQUENCE` derived types, `BIND(C)` derived types, or derived types that are neither `SEQUENCE`  
 46 nor `BIND(C)`. The complete definition of *extent* is given by Equation 4.1 Section 4.1.

**Example 4.1** Assume that  $Type = \{(double, 0), (char, 8)\}$  (a `double` at displacement zero, followed by a `char` at displacement eight). Assume, furthermore, that doubles have to be strictly aligned at addresses that are multiples of eight. Then, the extent of this datatype is 16 (9 rounded to the next multiple of 8). A datatype that consists of a character immediately followed by a double will also have an extent of 16.

*Rationale.* The definition of extent is motivated by the assumption that the amount of padding added at the end of each structure in an array of structures is the least needed to fulfill alignment constraints. More explicit control of the extent is provided in Section 4.1.6. Such explicit control is needed in cases where the assumption does not hold, for example, where union types are used. In Fortran, structures can be expressed with several language features, e.g., common blocks, `SEQUENCE` derived types, or `BIND(C)` derived types. The compiler may use different alignments, and therefore, it is recommended to use `MPI_TYPE_CREATE_RESIZED` for arrays of structures if an alignment may cause an alignment-gap at the end of a structure as described in Section 4.1.6 and in Section 17.1.15. (*End of rationale.*)

#### 4.1.1 Type Constructors with Explicit Addresses

In Fortran, the functions `MPI_TYPE_CREATE_HVECTOR`, `MPI_TYPE_CREATE_HINDEXED`, `MPI_TYPE_CREATE_HINDEXED_BLOCK`, `MPI_TYPE_CREATE_STRUCT`, and `MPI_GET_ADDRESS` accept arguments of type `INTEGER(KIND=MPI_ADDRESS_KIND)`, wherever arguments of type `MPI_Aint` are used in C. On Fortran 77 systems that do not support the Fortran 90 `KIND` notation, and where addresses are 64 bits whereas default `INTEGER`s are 32 bits, these arguments will be of type `INTEGER*8`.

#### 4.1.2 Datatype Constructors

**Contiguous** The simplest datatype constructor is `MPI_TYPE_CONTIGUOUS` which allows replication of a datatype into contiguous locations.

`MPI_TYPE_CONTIGUOUS(count, oldtype, newtype)`

IN	count	replication count (non-negative integer)
IN	oldtype	old datatype (handle)
OUT	newtype	new datatype (handle)

```
int MPI_Type_contiguous(int count, MPI_Datatype oldtype,
                       MPI_Datatype *newtype)
```

```
MPI_Type_contiguous(count, oldtype, newtype, ierror)
  INTEGER, INTENT(IN) :: count
  TYPE(MPI_Datatype), INTENT(IN) :: oldtype
  TYPE(MPI_Datatype), INTENT(OUT) :: newtype
  INTEGER, OPTIONAL, INTENT(OUT) :: ierror
```

```
MPI_TYPE_CONTIGUOUS(COUNT, OLDTYPE, NEWTYPE, IERROR)
  INTEGER COUNT, OLDTYPE, NEWTYPE, IERROR
```

1        `newtype` is the datatype obtained by concatenating `count` copies of  
 2        `oldtype`. Concatenation is defined using *extent* as the size of the concatenated copies.

3  
 4        **Example 4.2** Let `oldtype` have type map  $\{(\text{double}, 0), (\text{char}, 8)\}$ , with extent 16, and let  
 5        `count` = 3. The type map of the datatype returned by `newtype` is

6             $\{(\text{double}, 0), (\text{char}, 8), (\text{double}, 16), (\text{char}, 24), (\text{double}, 32), (\text{char}, 40)\}$ ;

7  
 8        i.e., alternating `double` and `char` elements, with displacements 0, 8, 16, 24, 32, 40.

9        In general, assume that the type map of `oldtype` is

10             $\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}$ ,

11        with extent *ex*. Then `newtype` has a type map with `count` · *n* entries defined by:

12             $\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1}), (type_0, disp_0 + ex), \dots, (type_{n-1}, disp_{n-1} + ex),$   
 13             $\dots, (type_0, disp_0 + ex \cdot (\text{count} - 1)), \dots, (type_{n-1}, disp_{n-1} + ex \cdot (\text{count} - 1))\}$ .

14  
 15  
 16  
 17        **Vector** The function `MPI_TYPE_VECTOR` is a more general constructor that allows repli-  
 18        cation of a datatype into locations that consist of equally spaced blocks. Each block is  
 19        obtained by concatenating the same number of copies of the old datatype. The spacing  
 20        between blocks is a multiple of the extent of the old datatype.

21  
 22        `MPI_TYPE_VECTOR(count, blocklength, stride, oldtype, newtype)`

23	IN	count	number of blocks (non-negative integer)
24	IN	blocklength	number of elements in each block (non-negative integer)
25			ger)
26			
27	IN	stride	number of elements between start of each block (integer)
28			ger)
29			
30	IN	oldtype	old datatype (handle)
31	OUT	newtype	new datatype (handle)

32  
 33        `int MPI_Type_vector(int count, int blocklength, int stride,`  
 34            `MPI_Datatype oldtype, MPI_Datatype *newtype)`

35  
 36        `MPI_Type_vector(count, blocklength, stride, oldtype, newtype, ierror)`  
 37            `INTEGER, INTENT(IN) :: count, blocklength, stride`  
 38            `TYPE(MPI_Datatype), INTENT(IN) :: oldtype`  
 39            `TYPE(MPI_Datatype), INTENT(OUT) :: newtype`  
 40            `INTEGER, OPTIONAL, INTENT(OUT) :: ierror`

41        `MPI_TYPE_VECTOR(COUNT, BLOCKLENGTH, STRIDE, OLDTYPE, NEWTYPE, IERROR)`  
 42            `INTEGER COUNT, BLOCKLENGTH, STRIDE, OLDTYPE, NEWTYPE, IERROR`

43  
 44        **Example 4.3** Assume, again, that `oldtype` has type map  $\{(\text{double}, 0), (\text{char}, 8)\}$ , with  
 45        extent 16. A call to `MPI_TYPE_VECTOR(2, 3, 4, oldtype, newtype)` will create the datatype  
 46        with type map,

47             $\{(\text{double}, 0), (\text{char}, 8), (\text{double}, 16), (\text{char}, 24), (\text{double}, 32), (\text{char}, 40),$   
 48             $\dots, (\text{double}, 16 \cdot 2), (\text{char}, 8 \cdot 2), (\text{double}, 16 \cdot 2 + 16), (\text{char}, 24 \cdot 2), (\text{double}, 32 \cdot 2), (\text{char}, 40 \cdot 2)\}$ .

$(\text{double}, 64), (\text{char}, 72), (\text{double}, 80), (\text{char}, 88), (\text{double}, 96), (\text{char}, 104)\}$ .

That is, two blocks with three copies each of the old type, with a stride of 4 elements ( $4 \cdot 16$  bytes) between the the start of each block.

**Example 4.4** A call to `MPI_TYPE_VECTOR(3, 1, -2, oldtype, newtype)` will create the datatype,

$\{(\text{double}, 0), (\text{char}, 8), (\text{double}, -32), (\text{char}, -24), (\text{double}, -64), (\text{char}, -56)\}$ .

In general, assume that `oldtype` has type map,

$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}$ ,

with extent  $ex$ . Let  $bl$  be the blocklength. The newly created datatype has a type map with  $count \cdot bl \cdot n$  entries:

$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1}),$   
 $(type_0, disp_0 + ex), \dots, (type_{n-1}, disp_{n-1} + ex), \dots,$   
 $(type_0, disp_0 + (bl - 1) \cdot ex), \dots, (type_{n-1}, disp_{n-1} + (bl - 1) \cdot ex),$   
 $(type_0, disp_0 + stride \cdot ex), \dots, (type_{n-1}, disp_{n-1} + stride \cdot ex), \dots,$   
 $(type_0, disp_0 + (stride + bl - 1) \cdot ex), \dots, (type_{n-1}, disp_{n-1} + (stride + bl - 1) \cdot ex), \dots,$   
 $(type_0, disp_0 + stride \cdot (count - 1) \cdot ex), \dots,$   
 $(type_{n-1}, disp_{n-1} + stride \cdot (count - 1) \cdot ex), \dots,$   
 $(type_0, disp_0 + (stride \cdot (count - 1) + bl - 1) \cdot ex), \dots,$   
 $(type_{n-1}, disp_{n-1} + (stride \cdot (count - 1) + bl - 1) \cdot ex)\}$ .

A call to `MPI_TYPE_CONTIGUOUS(count, oldtype, newtype)` is equivalent to a call to `MPI_TYPE_VECTOR(count, 1, 1, oldtype, newtype)`, or to a call to `MPI_TYPE_VECTOR(1, count, n, oldtype, newtype)`,  $n$  arbitrary.

**Hvector** The function `MPI_TYPE_CREATE_HVECTOR` is identical to `MPI_TYPE_VECTOR`, except that `stride` is given in bytes, rather than in elements. The use for both types of vector constructors is illustrated in Section 4.1.14. (H stands for “heterogeneous”).

`MPI_TYPE_CREATE_HVECTOR(count, blocklength, stride, oldtype, newtype)`

IN	count	number of blocks (non-negative integer)
IN	blocklength	number of elements in each block (non-negative integer)
IN	stride	number of bytes between start of each block (integer)
IN	oldtype	old datatype (handle)
OUT	newtype	new datatype (handle)

```

1 int MPI_Type_create_hvector(int count, int blocklength, MPI_Aint stride,
2     MPI_Datatype oldtype, MPI_Datatype *newtype)

```

```

3 MPI_Type_create_hvector(count, blocklength, stride, oldtype, newtype,
4     ierror)

```

```

5     INTEGER, INTENT(IN) :: count, blocklength
6     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: stride
7     TYPE(MPI_Datatype), INTENT(IN) :: oldtype
8     TYPE(MPI_Datatype), INTENT(OUT) :: newtype
9     INTEGER, OPTIONAL, INTENT(OUT) :: ierror

```

```

10 MPI_TYPE_CREATE_HVECTOR(COUNT, BLOCKLENGTH, STRIDE, OLDTYPE, NEWTYPE,
11     IERROR)

```

```

12     INTEGER COUNT, BLOCKLENGTH, OLDTYPE, NEWTYPE, IERROR
13     INTEGER(KIND=MPI_ADDRESS_KIND) STRIDE

```

Assume that `oldtype` has type map,

$$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

with extent  $ex$ . Let  $bl$  be the blocklength. The newly created datatype has a type map with  $count \cdot bl \cdot n$  entries:

$$\begin{aligned} & \{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1}), \\ & (type_0, disp_0 + ex), \dots, (type_{n-1}, disp_{n-1} + ex), \dots, \\ & (type_0, disp_0 + (bl - 1) \cdot ex), \dots, (type_{n-1}, disp_{n-1} + (bl - 1) \cdot ex), \\ & (type_0, disp_0 + stride), \dots, (type_{n-1}, disp_{n-1} + stride), \dots, \\ & (type_0, disp_0 + stride + (bl - 1) \cdot ex), \dots, \\ & (type_{n-1}, disp_{n-1} + stride + (bl - 1) \cdot ex), \dots, \\ & (type_0, disp_0 + stride \cdot (count - 1)), \dots, (type_{n-1}, disp_{n-1} + stride \cdot (count - 1)), \dots, \\ & (type_0, disp_0 + stride \cdot (count - 1) + (bl - 1) \cdot ex), \dots, \\ & (type_{n-1}, disp_{n-1} + stride \cdot (count - 1) + (bl - 1) \cdot ex)\}. \end{aligned}$$

**Indexed** The function `MPI_TYPE_INDEXED` allows replication of an old datatype into a sequence of blocks (each block is a concatenation of the old datatype), where each block can contain a different number of copies and have a different displacement. All block displacements are multiples of the old type extent.



```

MPI_TYPE_INDEXED(count, array_of_blocklengths, array_of_displacements, oldtype,
                 newtype)
1
2
IN      count      number of blocks — also number of entries in
3              array_of_displacements and array_of_blocklengths (non-
4              negative integer)
5
IN      array_of_blocklengths  number of elements per block (array of non-negative
6              integers)
7
IN      array_of_displacements  displacement for each block, in multiples of oldtype
8              extent (array of integer)
9
IN      oldtype      old datatype (handle)
10
OUT     newtype      new datatype (handle)
11
12
13
14
int MPI_Type_indexed(int count, const int array_of_blocklengths[], const
                    int array_of_displacements[], MPI_Datatype oldtype,
                    MPI_Datatype *newtype)
15
16
MPI_Type_indexed(count, array_of_blocklengths, array_of_displacements,
                oldtype, newtype, ierror)
17
18
INTEGER, INTENT(IN) :: count, array_of_blocklengths(count),
19
array_of_displacements(count)
20
TYPE(MPI_Datatype), INTENT(IN) :: oldtype
21
TYPE(MPI_Datatype), INTENT(OUT) :: newtype
22
INTEGER, OPTIONAL, INTENT(OUT) :: ierror
23
24
25
MPI_TYPE_INDEXED(COUNT, ARRAY_OF_BLOCKLENGTHS, ARRAY_OF_DISPLACEMENTS,
                OLDTYPE, NEWTYPE, IERROR)
26
27
INTEGER COUNT, ARRAY_OF_BLOCKLENGTHS(*), ARRAY_OF_DISPLACEMENTS(*),
28
OLDTYPE, NEWTYPE, IERROR
29
30

```

**Example 4.5**

Let `oldtype` have type map  $\{(\text{double}, 0), (\text{char}, 8)\}$ , with extent 16. Let  $B = (3, 1)$  and let  $D = (4, 0)$ . A call to `MPI_TYPE_INDEXED(2, B, D, oldtype, newtype)` returns a datatype with type map,

$$\{(\text{double}, 64), (\text{char}, 72), (\text{double}, 80), (\text{char}, 88), (\text{double}, 96), (\text{char}, 104),$$

$$(\text{double}, 0), (\text{char}, 8)\}.$$

That is, three copies of the old type starting at displacement 64, and one copy starting at displacement 0.

In general, assume that `oldtype` has type map,

$$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

with extent  $ex$ . Let  $B$  be the `array_of_blocklengths` argument and  $D$  be the `array_of_displacements` argument. The newly created datatype has  $n \cdot \sum_{i=0}^{\text{count}-1} B[i]$  entries:

$$\{(type_0, disp_0 + D[0] \cdot ex), \dots, (type_{n-1}, disp_{n-1} + D[0] \cdot ex), \dots,$$

```

1      (type0, disp0 + (D[0] + B[0] - 1) · ex), ..., (typen-1, dispn-1 + (D[0] + B[0] - 1) · ex), ...,
2
3      (type0, disp0 + D[count-1] · ex), ..., (typen-1, dispn-1 + D[count-1] · ex), ...,
4
5      (type0, disp0 + (D[count-1] + B[count-1] - 1) · ex), ...,
6
7      (typen-1, dispn-1 + (D[count-1] + B[count-1] - 1) · ex)}.
```

A call to MPI\_TYPE\_VECTOR(count, blocklength, stride, oldtype, newtype) is equivalent to a call to MPI\_TYPE\_INDEXED(count, B, D, oldtype, newtype) where

$$D[j] = j \cdot \text{stride}, \quad j = 0, \dots, \text{count} - 1,$$

and

$$B[j] = \text{blocklength}, \quad j = 0, \dots, \text{count} - 1.$$

**Hindexed** The function MPI\_TYPE\_CREATE\_HINDEXED is identical to MPI\_TYPE\_INDEXED, except that block displacements in array\_of\_displacements are specified in bytes, rather than in multiples of the oldtype extent.

```

21 MPI_TYPE_CREATE_HINDEXED(count, array_of_blocklengths, array_of_displacements,
22                          oldtype, newtype)
```

23	IN	count	number of blocks — also number of entries in array_of_displacements and array_of_blocklengths (non-negative integer)
24			
25			
26			
27	IN	array_of_blocklengths	number of elements in each block (array of non-negative integers)
28			
29	IN	array_of_displacements	byte displacement of each block (array of integer)
30			
31	IN	oldtype	old datatype (handle)
32	OUT	newtype	new datatype (handle)

```

33
34 int MPI_Type_create_hindexed(int count, const int array_of_blocklengths[],
35                             const MPI_Aint array_of_displacements[], MPI_Datatype oldtype,
36                             MPI_Datatype *newtype)
```

```

37 MPI_Type_create_hindexed(count, array_of_blocklengths,
38                          array_of_displacements, oldtype, newtype, ierror)
39 INTEGER, INTENT(IN) :: count, array_of_blocklengths(count)
40 INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) ::
41 array_of_displacements(count)
42 TYPE(MPI_Datatype), INTENT(IN) :: oldtype
43 TYPE(MPI_Datatype), INTENT(OUT) :: newtype
44 INTEGER, OPTIONAL, INTENT(OUT) :: ierror
```

```

45
46 MPI_TYPE_CREATE_HINDEXED(COUNT, ARRAY_OF_BLOCKLENGTHS,
47                          ARRAY_OF_DISPLACEMENTS, OLDTYPE, NEWTYPE, IERROR)
48 INTEGER COUNT, ARRAY_OF_BLOCKLENGTHS(*), OLDTYPE, NEWTYPE, IERROR
```

```
INTEGER(KIND=MPI_ADDRESS_KIND) ARRAY_OF_DISPLACEMENTS(*)
```

Assume that oldtype has type map,

$$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

with extent  $ex$ . Let  $B$  be the `array_of_blocklengths` argument and  $D$  be the `array_of_displacements` argument. The newly created datatype has a type map with  $n \cdot \sum_{i=0}^{count-1} B[i]$  entries:

$$\{(type_0, disp_0 + D[0]), \dots, (type_{n-1}, disp_{n-1} + D[0]), \dots, \\ (type_0, disp_0 + D[0] + (B[0] - 1) \cdot ex), \dots, \\ (type_{n-1}, disp_{n-1} + D[0] + (B[0] - 1) \cdot ex), \dots, \\ (type_0, disp_0 + D[count-1]), \dots, (type_{n-1}, disp_{n-1} + D[count-1]), \dots, \\ (type_0, disp_0 + D[count-1] + (B[count-1] - 1) \cdot ex), \dots, \\ (type_{n-1}, disp_{n-1} + D[count-1] + (B[count-1] - 1) \cdot ex)\}.$$

`Indexed_block` This function is the same as `MPI_TYPE_INDEXED` except that the block-length is the same for all blocks. There are many codes using indirect addressing arising from unstructured grids where the blocksize is always 1 (gather/scatter). The following convenience function allows for constant blocksize and arbitrary displacements.

```
MPI_TYPE_CREATE_INDEXED_BLOCK(count, blocklength, array_of_displacements, oldtype,
                               newtype)
```

IN	count	length of array of displacements (non-negative integer)
IN	blocklength	size of block (non-negative integer)
IN	array_of_displacements	array of displacements (array of integer)
IN	oldtype	old datatype (handle)
OUT	newtype	new datatype (handle)

```
int MPI_Type_create_indexed_block(int count, int blocklength, const
                                  int array_of_displacements[], MPI_Datatype oldtype,
                                  MPI_Datatype *newtype)
```

```
MPI_Type_create_indexed_block(count, blocklength, array_of_displacements,
                              oldtype, newtype, ierror)
```

```
INTEGER, INTENT(IN) :: count, blocklength,
array_of_displacements(count)
TYPE(MPI_Datatype), INTENT(IN) :: oldtype
TYPE(MPI_Datatype), INTENT(OUT) :: newtype
INTEGER, OPTIONAL, INTENT(OUT) :: ierror
```

```
MPI_TYPE_CREATE_INDEXED_BLOCK(COUNT, BLOCKLENGTH, ARRAY_OF_DISPLACEMENTS,
                              OLDTYPE, NEWTYPE, IERROR)
```

```
INTEGER COUNT, BLOCKLENGTH, ARRAY_OF_DISPLACEMENTS(*), OLDTYPE,
NEWTYPE, IERROR
```

1 Hindexed\_block The function `MPI_TYPE_CREATE_HINDEXED_BLOCK` is identical to  
 2 `MPI_TYPE_CREATE_INDEXED_BLOCK`, except that block displacements in  
 3 `array_of_displacements` are specified in bytes, rather than in multiples of the oldtype extent.  
 4

```
5 MPI_TYPE_CREATE_HINDEXED_BLOCK(count, blocklength, array_of_displacements,
6                               oldtype, newtype)
7
8     IN      count                length of array of displacements (non-negative integer)
9     IN      blocklength          size of block (non-negative integer)
10    IN      array_of_displacements byte displacement of each block (array of integer)
11
12    IN      oldtype              old datatype (handle)
13
14    OUT     newtype              new datatype (handle)
```

```
15
16 int MPI_Type_create_hindexed_block(int count, int blocklength, const
17                                 MPI_Aint array_of_displacements[], MPI_Datatype oldtype,
18                                 MPI_Datatype *newtype)
19
20 MPI_Type_create_hindexed_block(count, blocklength, array_of_displacements,
21                               oldtype, newtype, ierror)
22     INTEGER, INTENT(IN) :: count, blocklength
23     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) ::
24     array_of_displacements(count)
25     TYPE(MPI_Datatype), INTENT(IN) :: oldtype
26     TYPE(MPI_Datatype), INTENT(OUT) :: newtype
27     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
28
29 MPI_TYPE_CREATE_HINDEXED_BLOCK(COUNT, BLOCKLENGTH, ARRAY_OF_DISPLACEMENTS,
30                               OLDTYPE, NEWTYPE, IERROR)
31     INTEGER COUNT, BLOCKLENGTH, OLDTYPE, NEWTYPE, IERROR
32     INTEGER(KIND=MPI_ADDRESS_KIND) ARRAY_OF_DISPLACEMENTS(*)
```

33 Struct `MPI_TYPE_CREATE_STRUCT` is the most general type constructor. It further  
 34 generalizes `MPI_TYPE_CREATE_HINDEXED` in that it allows each block to consist of repli-  
 35 cations of different datatypes.  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48

```

MPI_TYPE_CREATE_STRUCT(count, array_of_blocklengths, array_of_displacements,
                        array_of_types, newtype)
1
2
IN      count          number of blocks (non-negative integer) — also num-
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

```

IN count number of blocks (non-negative integer) — also num-  
 ber of entries in arrays `array_of_types`,  
`array_of_displacements` and `array_of_blocklengths`  
 IN `array_of_blocklength` number of elements in each block (array of non-negative  
 integer)  
 IN `array_of_displacements` byte displacement of each block (array of integer)  
 IN `array_of_types` type of elements in each block (array of handles to  
 datatype objects)  
 OUT `newtype` new datatype (handle)

```

int MPI_Type_create_struct(int count, const int array_of_blocklengths[],
                           const MPI_Aint array_of_displacements[], const
                           MPI_Datatype array_of_types[], MPI_Datatype *newtype)
15
16
17
MPI_Type_create_struct(count, array_of_blocklengths,
                       array_of_displacements, array_of_types, newtype, ierror)
18
19
20
21
22
23
24
25
26
MPI_TYPE_CREATE_STRUCT(COUNT, ARRAY_OF_BLOCKLENGTHS,
                       ARRAY_OF_DISPLACEMENTS, ARRAY_OF_TYPES, NEWTYPE, IERROR)
27
28
29
30
31
32

```

**Example 4.6** Let `type1` have type map,

$$\{(\text{double}, 0), (\text{char}, 8)\},$$

with extent 16. Let  $B = (2, 1, 3)$ ,  $D = (0, 16, 26)$ , and  $T = (\text{MPI\_FLOAT}, \text{type1}, \text{MPI\_CHAR})$ . Then a call to `MPI_TYPE_CREATE_STRUCT(3, B, D, T, newtype)` returns a datatype with type map,

$$\{(\text{float}, 0), (\text{float}, 4), (\text{double}, 16), (\text{char}, 24), (\text{char}, 26), (\text{char}, 27), (\text{char}, 28)\}.$$

That is, two copies of `MPI_FLOAT` starting at 0, followed by one copy of `type1` starting at 16, followed by three copies of `MPI_CHAR`, starting at 26. (We assume that a float occupies four bytes.)

In general, let  $T$  be the `array_of_types` argument, where  $T[i]$  is a handle to,

$$\text{typemap}_i = \{(type_0^i, disp_0^i), \dots, (type_{n_i-1}^i, disp_{n_i-1}^i)\},$$

1 with extent  $ex_i$ . Let B be the `array_of_blocklength` argument and D be the  
 2 `array_of_displacements` argument. Let c be the count argument. Then the newly created  
 3 datatype has a type map with  $\sum_{i=0}^{c-1} B[i] \cdot n_i$  entries:

$$\begin{aligned}
 & \{(type_0^0, disp_0^0 + D[0]), \dots, (type_{n_0}^0, disp_{n_0}^0 + D[0]), \dots, \\
 & (type_0^0, disp_0^0 + D[0] + (B[0] - 1) \cdot ex_0), \dots, (type_{n_0}^0, disp_{n_0}^0 + D[0] + (B[0]-1) \cdot ex_0), \dots, \\
 & (type_0^{c-1}, disp_0^{c-1} + D[c-1]), \dots, (type_{n_{c-1}-1}^{c-1}, disp_{n_{c-1}-1}^{c-1} + D[c-1]), \dots, \\
 & (type_0^{c-1}, disp_0^{c-1} + D[c-1] + (B[c-1] - 1) \cdot ex_{c-1}), \dots, \\
 & (type_{n_{c-1}-1}^{c-1}, disp_{n_{c-1}-1}^{c-1} + D[c-1] + (B[c-1]-1) \cdot ex_{c-1})\}.
 \end{aligned}$$

14 A call to `MPI_TYPE_CREATE_HINDEXED(count, B, D, oldtype, newtype)` is equivalent  
 15 to a call to `MPI_TYPE_CREATE_STRUCT(count, B, D, T, newtype)`, where each entry of T  
 16 is equal to oldtype.

### 18 4.1.3 Subarray Datatype Constructor

21 `MPI_TYPE_CREATE_SUBARRAY(ndims, array_of_sizes, array_of_subsizes, array_of_starts,`  
 22 `order, oldtype, newtype)`

24	IN	<code>ndims</code>	number of array dimensions (positive integer)
25	IN	<code>array_of_sizes</code>	number of elements of type <code>oldtype</code> in each dimension of the full array (array of positive integers)
26			
27	IN	<code>array_of_subsizes</code>	number of elements of type <code>oldtype</code> in each dimension of the subarray (array of positive integers)
28			
29	IN	<code>array_of_starts</code>	starting coordinates of the subarray in each dimension (array of non-negative integers)
30			
31	IN	<code>order</code>	array storage order flag (state)
32			
33	IN	<code>oldtype</code>	array element datatype (handle)
34			
35	OUT	<code>newtype</code>	new datatype (handle)

36  
 37 `int MPI_Type_create_subarray(int ndims, const int array_of_sizes[], const`  
 38 `int array_of_subsizes[], const int array_of_starts[], int`  
 39 `order, MPI_Datatype oldtype, MPI_Datatype *newtype)`

40  
 41 `MPI_Type_create_subarray(ndims, array_of_sizes, array_of_subsizes,`  
 42 `array_of_starts, order, oldtype, newtype, ierror)`  
 43 `INTEGER, INTENT(IN) :: ndims, array_of_sizes(ndims),`  
 44 `array_of_subsizes(ndims), array_of_starts(ndims), order`  
 45 `TYPE(MPI_Datatype), INTENT(IN) :: oldtype`  
 46 `TYPE(MPI_Datatype), INTENT(OUT) :: newtype`  
 47 `INTEGER, OPTIONAL, INTENT(OUT) :: ierror`  
 48

```

MPI_TYPE_CREATE_SUBARRAY(NDIMS, ARRAY_OF_SIZES, ARRAY_OF_SUBSIZES,
                          ARRAY_OF_STARTS, ORDER, OLDTYPE, NEWTYPE, IERROR)
INTEGER NDIMS, ARRAY_OF_SIZES(*), ARRAY_OF_SUBSIZES(*),
ARRAY_OF_STARTS(*), ORDER, OLDTYPE, NEWTYPE, IERROR

```

The subarray type constructor creates an MPI datatype describing an  $n$ -dimensional subarray of an  $n$ -dimensional array. The subarray may be situated anywhere within the full array, and may be of any nonzero size up to the size of the larger array as long as it is confined within this array. This type constructor facilitates creating filetypes to access arrays distributed in blocks among processes to a single file that contains the global array, see MPI I/O, especially Section 13.1.1.

This type constructor can handle arrays with an arbitrary number of dimensions and works for both C and Fortran ordered matrices (i.e., row-major or column-major). Note that a C program may use Fortran order and a Fortran program may use C order.

The `ndims` parameter specifies the number of dimensions in the full data array and gives the number of elements in `array_of_sizes`, `array_of_subsizes`, and `array_of_starts`.

The number of elements of type `oldtype` in each dimension of the  $n$ -dimensional array and the requested subarray are specified by `array_of_sizes` and `array_of_subsizes`, respectively. For any dimension  $i$ , it is erroneous to specify `array_of_subsizes[i] < 1` or `array_of_subsizes[i] > array_of_sizes[i]`.

The `array_of_starts` contains the starting coordinates of each dimension of the subarray. Arrays are assumed to be indexed starting from zero. For any dimension  $i$ , it is erroneous to specify `array_of_starts[i] < 0` or `array_of_starts[i] > (array_of_sizes[i] - array_of_subsizes[i])`.

*Advice to users.* In a Fortran program with arrays indexed starting from 1, if the starting coordinate of a particular dimension of the subarray is  $n$ , then the entry in `array_of_starts` for that dimension is  $n-1$ . (*End of advice to users.*)

The `order` argument specifies the storage order for the subarray as well as the full array. It must be set to one of the following:

**MPI\_ORDER\_C** The ordering used by C arrays, (i.e., row-major order)

**MPI\_ORDER\_FORTRAN** The ordering used by Fortran arrays, (i.e., column-major order)

A `ndims`-dimensional subarray (`newtype`) with no extra padding can be defined by the function `Subarray()` as follows:

```

newtype = Subarray(ndims, {size0, size1, ..., sizendims-1},
                  {subsize0, subsize1, ..., subsizendims-1},
                  {start0, start1, ..., startndims-1}, oldtype)

```

Let the typemap of `oldtype` have the form:

$$\{(type_0, disp_0), (type_1, disp_1), \dots, (type_{n-1}, disp_{n-1})\}$$

where  $type_i$  is a predefined MPI datatype, and let  $ex$  be the extent of `oldtype`. Then we define the `Subarray()` function recursively using the following three equations. Equation 4.2 defines the base step. Equation 4.3 defines the recursion step when `order = MPI_ORDER_FORTRAN`, and Equation 4.4 defines the recursion step when `order = MPI_ORDER_C`. These equations use the conceptual datatypes `lb_marker` and `ub_marker`, see Section 4.1.6 for details.

$$\begin{aligned}
& \text{Subarray}(1, \{size_0\}, \{subsize_0\}, \{start_0\}, \\
& \quad \{(type_0, disp_0), (type_1, disp_1), \dots, (type_{n-1}, disp_{n-1})\}) \\
& = \{(lb\_marker, 0), \\
& \quad (type_0, disp_0 + start_0 \times ex), \dots, (type_{n-1}, disp_{n-1} + start_0 \times ex), \\
& \quad (type_0, disp_0 + (start_0 + 1) \times ex), \dots, (type_{n-1}, \\
& \quad \quad disp_{n-1} + (start_0 + 1) \times ex), \dots \\
& \quad (type_0, disp_0 + (start_0 + subsize_0 - 1) \times ex), \dots, \\
& \quad \quad (type_{n-1}, disp_{n-1} + (start_0 + subsize_0 - 1) \times ex), \\
& \quad (ub\_marker, size_0 \times ex)\}
\end{aligned} \tag{4.2}$$

$$\begin{aligned}
& \text{Subarray}(ndims, \{size_0, size_1, \dots, size_{ndims-1}\}, \\
& \quad \{subsize_0, subsize_1, \dots, subsize_{ndims-1}\}, \\
& \quad \{start_0, start_1, \dots, start_{ndims-1}\}, oldtype) \\
& = \text{Subarray}(ndims - 1, \{size_1, size_2, \dots, size_{ndims-1}\}, \\
& \quad \{subsize_1, subsize_2, \dots, subsize_{ndims-1}\}, \\
& \quad \{start_1, start_2, \dots, start_{ndims-1}\}, \\
& \quad \text{Subarray}(1, \{size_0\}, \{subsize_0\}, \{start_0\}, oldtype))
\end{aligned} \tag{4.3}$$

$$\begin{aligned}
& \text{Subarray}(ndims, \{size_0, size_1, \dots, size_{ndims-1}\}, \\
& \quad \{subsize_0, subsize_1, \dots, subsize_{ndims-1}\}, \\
& \quad \{start_0, start_1, \dots, start_{ndims-1}\}, oldtype) \\
& = \text{Subarray}(ndims - 1, \{size_0, size_1, \dots, size_{ndims-2}\}, \\
& \quad \{subsize_0, subsize_1, \dots, subsize_{ndims-2}\}, \\
& \quad \{start_0, start_1, \dots, start_{ndims-2}\}, \\
& \quad \text{Subarray}(1, \{size_{ndims-1}\}, \{subsize_{ndims-1}\}, \{start_{ndims-1}\}, oldtype))
\end{aligned} \tag{4.4}$$

For an example use of `MPI_TYPE_CREATE_SUBARRAY` in the context of I/O see Section [13.11.2](#).

#### 4.1.4 Distributed Array Datatype Constructor

The distributed array type constructor supports HPF-like [1] data distributions. However, unlike in HPF, the storage order may be specified for C arrays as well as for Fortran arrays.

*Advice to users.* One can create an HPF-like file view using this type constructor as follows. Complementary filetypes are created by having every process of a group call this constructor with identical arguments (with the exception of rank which should be set appropriately). These filetypes (along with identical disp and etype) are then used to define the view (via `MPI_FILE_SET_VIEW`), see MPI I/O, especially Section [13.1.1](#) and Section [13.3](#). Using this view, a collective data access operation (with identical offsets) will yield an HPF-like distribution pattern. (*End of advice to users.*)



MPI_TYPE_CREATE_DARRAY(size, rank, ndims, array_of_gsizes, array_of_distrib,			1
array_of_dargs, array_of_psizes, order, oldtype, newtype)			2
IN	size	size of process group (positive integer)	3
IN	rank	rank in process group (non-negative integer)	4
IN	ndims	number of array dimensions as well as process grid dimensions (positive integer)	5
IN	array_of_gsizes	number of elements of type oldtype in each dimension of global array (array of positive integers)	6
IN	array_of_distrib	distribution of array in each dimension (array of state)	7
IN	array_of_dargs	distribution argument in each dimension (array of positive integers)	8
IN	array_of_psizes	size of process grid in each dimension (array of positive integers)	9
IN	order	array storage order flag (state)	10
IN	oldtype	old datatype (handle)	11
OUT	newtype	new datatype (handle)	12
			13
int MPI_Type_create_darray(int size, int rank, int ndims, const			14
int array_of_gsizes[], const int array_of_distrib[], const			15
int array_of_dargs[], const int array_of_psizes[], int order,			16
MPI_Datatype oldtype, MPI_Datatype *newtype)			17
			18
MPI_Type_create_darray(size, rank, ndims, array_of_gsizes,			19
array_of_distrib, array_of_dargs, array_of_psizes, order,			20
oldtype, newtype, ierror)			21
INTEGER, INTENT(IN) :: size, rank, ndims, array_of_gsizes(ndims),			22
array_of_distrib(ndims), array_of_dargs(ndims),			23
array_of_psizes(ndims), order			24
TYPE(MPI_Datatype), INTENT(IN) :: oldtype			25
TYPE(MPI_Datatype), INTENT(OUT) :: newtype			26
INTEGER, OPTIONAL, INTENT(OUT) :: ierror			27
			28
MPI_TYPE_CREATE_DARRAY(SIZE, RANK, NDIMS, ARRAY_OF_GSIZES,			29
ARRAY_OF_DISTRIBS, ARRAY_OF_DARGS, ARRAY_OF_PSIZEs, ORDER,			30
OLDTYPE, NEWTYPE, IERROR)			31
INTEGER SIZE, RANK, NDIMS, ARRAY_OF_GSIZES(*), ARRAY_OF_DISTRIBS(*),			32
ARRAY_OF_DARGS(*), ARRAY_OF_PSIZEs(*), ORDER, OLDTYPE, NEWTYPE, IERROR			33
			34
MPI_TYPE_CREATE_DARRAY can be used to generate the datatypes corresponding			35
to the distribution of an ndims-dimensional array of oldtype elements onto an			36
ndims-dimensional grid of logical processes. Unused dimensions of array_of_psize			37
s should be set to 1. (See Example 4.7.) For a call to MPI_TYPE_CREATE_DARRAY to be correct, the			38
equation $\prod_{i=0}^{ndims-1} array\_of\_psizes[i] = size$ must be satisfied. The ordering of processes			39
in the process grid is assumed to be row-major, as in the case of virtual Cartesian process			40
topologies.			41
			42
			43
			44
			45
			46
			47
			48

1        *Advice to users.* For both Fortran and C arrays, the ordering of processes in the  
 2        process grid is assumed to be row-major. This is consistent with the ordering used in  
 3        virtual Cartesian process topologies in MPI. To create such virtual process topologies,  
 4        or to find the coordinates of a process in the process grid, etc., users may use the  
 5        corresponding process topology functions, see Chapter 7. (*End of advice to users.*)

6        Each dimension of the array can be distributed in one of three ways:

- 8        • MPI\_DISTRIBUTE\_BLOCK - Block distribution
- 10       • MPI\_DISTRIBUTE\_CYCLIC - Cyclic distribution
- 12       • MPI\_DISTRIBUTE\_NONE - Dimension not distributed.

13       The constant MPI\_DISTRIBUTE\_DFLT\_DARG specifies a default distribution argument.  
 14       The distribution argument for a dimension that is not distributed is ignored. For any  
 15       dimension  $i$  in which the distribution is MPI\_DISTRIBUTE\_BLOCK, it is erroneous to specify  
 16        $\text{array\_of\_dargs}[i] * \text{array\_of\_psizes}[i] < \text{array\_of\_gsizes}[i]$ .

17       For example, the HPF layout `ARRAY(CYCLIC(15))` corresponds to  
 18       MPI\_DISTRIBUTE\_CYCLIC with a distribution argument of 15, and the HPF layout `AR-`  
 19       `RAY(BLOCK)` corresponds to MPI\_DISTRIBUTE\_BLOCK with a distribution argument of  
 20       MPI\_DISTRIBUTE\_DFLT\_DARG.

21       The `order` argument is used as in MPI\_TYPE\_CREATE\_SUBARRAY to specify the stor-  
 22       age order. Therefore, arrays described by this type constructor may be stored in Fortran  
 23       (column-major) or C (row-major) order. Valid values for `order` are MPI\_ORDER\_FORTRAN  
 24       and MPI\_ORDER\_C.

25       This routine creates a new MPI datatype with a typemap defined in terms of a function  
 26       called “cyclic()” (see below).

27       Without loss of generality, it suffices to define the typemap for the  
 28       MPI\_DISTRIBUTE\_CYCLIC case where MPI\_DISTRIBUTE\_DFLT\_DARG is not used.

29       MPI\_DISTRIBUTE\_BLOCK and MPI\_DISTRIBUTE\_NONE can be reduced to the  
 30       MPI\_DISTRIBUTE\_CYCLIC case for dimension  $i$  as follows.

31       MPI\_DISTRIBUTE\_BLOCK with  $\text{array\_of\_dargs}[i]$  equal to MPI\_DISTRIBUTE\_DFLT\_DARG  
 32       is equivalent to MPI\_DISTRIBUTE\_CYCLIC with  $\text{array\_of\_dargs}[i]$  set to

$$33 \quad (\text{array\_of\_gsizes}[i] + \text{array\_of\_psizes}[i] - 1) / \text{array\_of\_psizes}[i].$$

34       If  $\text{array\_of\_dargs}[i]$  is not MPI\_DISTRIBUTE\_DFLT\_DARG, then MPI\_DISTRIBUTE\_BLOCK and  
 35       MPI\_DISTRIBUTE\_CYCLIC are equivalent.

36       MPI\_DISTRIBUTE\_NONE is equivalent to MPI\_DISTRIBUTE\_CYCLIC with  
 37        $\text{array\_of\_dargs}[i]$  set to  $\text{array\_of\_gsizes}[i]$ .

38       Finally, MPI\_DISTRIBUTE\_CYCLIC with  $\text{array\_of\_dargs}[i]$  equal to  
 39       MPI\_DISTRIBUTE\_DFLT\_DARG is equivalent to MPI\_DISTRIBUTE\_CYCLIC with  
 40        $\text{array\_of\_dargs}[i]$  set to 1.

41       For MPI\_ORDER\_FORTRAN, an  $\text{ndims}$ -dimensional distributed array (`newtype`) is defined  
 42       by the following code fragment:

```
43
44
45     oldtypes[0] = oldtype;
46     for (i = 0; i < ndims; i++) {
47         oldtypes[i+1] = cyclic(array_of_dargs[i],
48
```

```

array_of_gsizes[i],
r[i],
array_of_psizes[i],
oldtypes[i]);
}
newtype = oldtypes[ndims];

```

For MPI\_ORDER\_C, the code is:

```

oldtypes[0] = oldtype;
for (i = 0; i < ndims; i++) {
    oldtypes[i + 1] = cyclic(array_of_dargs[ndims - i - 1],
array_of_gsizes[ndims - i - 1],
r[ndims - i - 1],
array_of_psizes[ndims - i - 1],
oldtypes[i]);
}
newtype = oldtypes[ndims];

```

where  $r[i]$  is the position of the process (with rank `rank`) in the process grid at dimension  $i$ . The values of  $r[i]$  are given by the following code fragment:

```

t_rank = rank;
t_size = 1;
for (i = 0; i < ndims; i++)
    t_size *= array_of_psizes[i];
for (i = 0; i < ndims; i++) {
    t_size = t_size / array_of_psizes[i];
    r[i] = t_rank / t_size;
    t_rank = t_rank % t_size;
}

```

Let the typemap of `oldtype` have the form:

$$\{(type_0, disp_0), (type_1, disp_1), \dots, (type_{n-1}, disp_{n-1})\}$$

where  $type_i$  is a predefined MPI datatype, and let  $ex$  be the extent of `oldtype`. The following function uses the conceptual datatypes `lb_marker` and `ub_marker`, see Section 4.1.6 for details.

Given the above, the function `cyclic()` is defined as follows:

```

cyclic(darg, gsize, r, psize, oldtype)
= { (lb_marker, 0),
    (type_0, disp_0 + r × darg × ex), ...,
    (type_{n-1}, disp_{n-1} + r × darg × ex),
    (type_0, disp_0 + (r × darg + 1) × ex), ...,
    (type_{n-1}, disp_{n-1} + (r × darg + 1) × ex),

```

```

1      ...
2      (type0, disp0 + ((r + 1) × darg - 1) × ex), ...,
3          (typen-1, dispn-1 + ((r + 1) × darg - 1) × ex),
4
5      (type0, disp0 + r × darg × ex + psize × darg × ex), ...,
6          (typen-1, dispn-1 + r × darg × ex + psize × darg × ex),
7      (type0, disp0 + (r × darg + 1) × ex + psize × darg × ex), ...,
8          (typen-1, dispn-1 + (r × darg + 1) × ex + psize × darg × ex),
9
10     ...
11     (type0, disp0 + ((r + 1) × darg - 1) × ex + psize × darg × ex), ...,
12         (typen-1, dispn-1 + ((r + 1) × darg - 1) × ex + psize × darg × ex),
13
14         ⋮
15     (type0, disp0 + r × darg × ex + psize × darg × ex × (count - 1)), ...,
16         (typen-1, dispn-1 + r × darg × ex + psize × darg × ex × (count - 1)),
17     (type0, disp0 + (r × darg + 1) × ex + psize × darg × ex × (count - 1)), ...,
18         (typen-1, dispn-1 + (r × darg + 1) × ex
19             + psize × darg × ex × (count - 1)),
20
21     ...
22     (type0, disp0 + (r × darg + darglast - 1) × ex
23         + psize × darg × ex × (count - 1)), ...,
24         (typen-1, dispn-1 + (r × darg + darglast - 1) × ex
25             + psize × darg × ex × (count - 1)),
26
27     (ub_marker, gsize * ex)}
28
29

```

where *count* is defined by this code fragment:

```

31     nblocks = (gsize + (darg - 1)) / darg;
32     count = nblocks / psize;
33     left_over = nblocks - count * psize;
34     if (r < left_over)
35         count = count + 1;
36

```

Here, *nblocks* is the number of blocks that must be distributed among the processors. Finally, *darg<sub>last</sub>* is defined by this code fragment:

```

37     if ((num_in_last_cyclic = gsize % (psize * darg)) == 0)
38         darg_last = darg;
39     else {
40         darg_last = num_in_last_cyclic - darg * r;
41         if (darg_last > darg)
42             darg_last = darg;
43         if (darg_last <= 0)
44             darg_last = darg;
45     }
46
47
48

```

**Example 4.7** Consider generating the filetypes corresponding to the HPF distribution:

```

<oldtype> FILEARRAY(100, 200, 300)
!HPF$ PROCESSORS PROCESSES(2, 3)
!HPF$ DISTRIBUTE FILEARRAY(CYCLIC(10), *, BLOCK) ONTO PROCESSES

```

This can be achieved by the following Fortran code, assuming there will be six processes attached to the run:

```

ndims = 3
array_of_gsizes(1) = 100
array_of_distrib(1) = MPI_DISTRIBUTE_CYCLIC
array_of_dargs(1) = 10
array_of_gsizes(2) = 200
array_of_distrib(2) = MPI_DISTRIBUTE_NONE
array_of_dargs(2) = 0
array_of_gsizes(3) = 300
array_of_distrib(3) = MPI_DISTRIBUTE_BLOCK
array_of_dargs(3) = MPI_DISTRIBUTE_DFLT_DARG
array_of_psize(1) = 2
array_of_psize(2) = 1
array_of_psize(3) = 3
call MPI_COMM_SIZE(MPI_COMM_WORLD, size, ierr)
call MPI_COMM_RANK(MPI_COMM_WORLD, rank, ierr)
call MPI_TYPE_CREATE_DARRAY(size, rank, ndims, array_of_gsizes, &
    array_of_distrib, array_of_dargs, array_of_psize, &
    MPI_ORDER_FORTRAN, oldtype, newtype, ierr)

```

#### 4.1.5 Address and Size Functions

The displacements in a general datatype are relative to some initial buffer address. *Absolute addresses* can be substituted for these displacements: we treat them as displacements relative to “address zero,” the start of the address space. This initial address zero is indicated by the constant `MPI_BOTTOM`. Thus, a datatype can specify the absolute address of the entries in the communication buffer, in which case the `buf` argument is passed the value `MPI_BOTTOM`.

The address of a location in memory can be found by invoking the function `MPI_GET_ADDRESS`.

The relative displacement between two absolute addresses can be calculated with the function `MPI_AINT_DIFF`. A new absolute address as sum of an absolute base address and a relative displacement can be calculated with the function `MPI_AINT_ADD`. To ensure portability, arithmetic on absolute addresses should not be performed with the intrinsic operators “-” and “+”. See also Sections 4.1.5 and 4.1.12 on pages 19 and 33.

*Rationale.* Address sized integer values, i.e., `MPI_Aint` or `INTEGER(KIND=MPI_ADDRESS_KIND)` values, are signed integers, while absolute addresses are unsigned quantities. Direct arithmetic on addresses stored in address sized signed variables can cause overflows, resulting in undefined behavior. (*End of rationale.*)

```

1 MPI_GET_ADDRESS(location, address)
2   IN      location          location in caller memory (choice)
3   OUT     address           address of location (integer)
4

```

```

5
6 int MPI_Get_address(const void *location, MPI_Aint *address)
7

```

```

8 MPI_Get_address(location, address, ierror)
9   TYPE(*), DIMENSION(..), ASYNCHRONOUS :: location
10  INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(OUT) :: address
11  INTEGER, OPTIONAL, INTENT(OUT) :: ierror
12

```

```

13 MPI_GET_ADDRESS(LOCATION, ADDRESS, IERROR)
14   <type> LOCATION(*)
15   INTEGER IERROR
16   INTEGER(KIND=MPI_ADDRESS_KIND) ADDRESS
17

```

Returns the (byte) address of location.

*Rationale.* In the `mpi_f08` module, the `location` argument is not defined with `INTENT(IN)` because existing applications may use `MPI_GET_ADDRESS` as a substitute for `MPI_F_SYNC_REG` that was not defined before MPI-3.0. (*End of rationale.*)

**Example 4.8** Using `MPI_GET_ADDRESS` for an array.

```

24
25 REAL A(100,100)
26 INTEGER(KIND=MPI_ADDRESS_KIND) I1, I2, DIFF
27 CALL MPI_GET_ADDRESS(A(1,1), I1, IERROR)
28 CALL MPI_GET_ADDRESS(A(10,10), I2, IERROR)
29 DIFF = MPI_AINT_DIFF(I2, I1)
30 ! The value of DIFF is 909*sizeofreal; the values of I1 and I2 are
31 ! implementation dependent.
32

```

*Advice to users.* C users may be tempted to avoid the usage of `MPI_GET_ADDRESS` and rely on the availability of the address operator `&`. Note, however, that `& cast-expression` is a pointer, not an address. ISO C does not require that the value of a pointer (or the pointer cast to `int`) be the absolute address of the object pointed at — although this is commonly the case. Furthermore, referencing may not have a unique definition on machines with a segmented address space. The use of `MPI_GET_ADDRESS` to “reference” C variables guarantees portability to such machines as well. (*End of advice to users.*)

*Advice to users.* To prevent problems with the argument copying and register optimization done by Fortran compilers, please note the hints in Sections 17.1.10–??. (*End of advice to users.*)

To ensure portability, arithmetic on MPI addresses must be performed using the `MPI_AINT_ADD` and `MPI_AINT_DIFF` functions.

```

MPI_AINT_ADD(base, disp) 1
    IN      base          base address (integer) 2
    IN      disp         displacement (integer) 3
  4
  5

```

```

MPI_Aint MPI_Aint_add(MPI_Aint base, MPI_Aint disp) 6
INTEGER(KIND=MPI_ADDRESS_KIND) MPI_Aint_add(base, disp) 7
    INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: base, disp 8
  9
INTEGER(KIND=MPI_ADDRESS_KIND) MPI_AINT_ADD(BASE, DISP) 10
    INTEGER(KIND=MPI_ADDRESS_KIND) BASE, DISP 11
  12

```

MPI\_AINT\_ADD produces a new MPI\_Aint value that is equivalent to the sum of the base and disp arguments, where base represents a base address returned by a call to MPI\_GET\_ADDRESS and disp represents a signed integer displacement. The resulting address is valid only at the process that generated base, and it must correspond to a location in the same object referenced by base, as described in Section 4.1.12. The addition is performed in a manner that results in the correct MPI\_Aint representation of the output address, as if the process that originally produced base had called:

```

MPI_Get_address((char *) base + disp, &result) 20
  21

```

*Rationale.* MPI\_Aint values are signed integers, while addresses are unsigned quantities. Direct arithmetic on addresses in MPI\_Aint variables can cause overflows, resulting in undefined behavior. (*End of rationale.*)

```

MPI_AINT_DIFF(addr1, addr2) 28
    IN      addr1        minuend address (integer) 29
    IN      addr2        subtrahend address (integer) 30
  31
  32

```

```

MPI_Aint MPI_Aint_diff(MPI_Aint addr1, MPI_Aint addr2) 33
INTEGER(KIND=MPI_ADDRESS_KIND) MPI_Aint_diff(addr1, addr2) 34
    INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: addr1, addr2 35
  36
INTEGER(KIND=MPI_ADDRESS_KIND) MPI_AINT_DIFF(ADDR1, ADDR2) 37
    INTEGER(KIND=MPI_ADDRESS_KIND) ADDR1, ADDR2 38
  39

```

MPI\_AINT\_DIFF produces a new MPI\_Aint value that is equivalent to the difference between addr1 and addr2 arguments, where addr1 and addr2 represent addresses returned by calls to MPI\_GET\_ADDRESS. The resulting address is valid only at the process that generated addr1 and addr2, and addr1 and addr2 must correspond to locations in the same object in the same process, as described in Section 4.1.12. The difference is calculated in a manner that results the signed difference from addr1 to addr2, as if the process that originally produced the addresses had called (char \*) addr1 - (char \*) addr2 on the addresses initially passed to MPI\_GET\_ADDRESS.

The following auxiliary functions provide useful information on derived datatypes.

```

1 MPI_TYPE_SIZE(datatype, size)
2     IN      datatype          datatype (handle)
3     OUT     size              datatype size (integer)
4
5
6 int MPI_Type_size(MPI_Datatype datatype, int *size)
7
8 MPI_Type_size(datatype, size, ierror)
9     TYPE(MPI_Datatype), INTENT(IN) :: datatype
10    INTEGER, INTENT(OUT) :: size
11    INTEGER, OPTIONAL, INTENT(OUT) :: ierror
12
13 MPI_TYPE_SIZE(DATATYPE, SIZE, IERROR)
14     INTEGER DATATYPE, SIZE, IERROR
15
16 MPI_TYPE_SIZE_X(datatype, size)
17     IN      datatype          datatype (handle)
18     OUT     size              datatype size (integer)
19
20
21 int MPI_Type_size_x(MPI_Datatype datatype, MPI_Count *size)
22
23 MPI_Type_size_x(datatype, size, ierror)
24     TYPE(MPI_Datatype), INTENT(IN) :: datatype
25     INTEGER(KIND=MPI_COUNT_KIND), INTENT(OUT) :: size
26     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
27
28 MPI_TYPE_SIZE_X(DATATYPE, SIZE, IERROR)
29     INTEGER DATATYPE, IERROR
30     INTEGER(KIND = MPI_COUNT_KIND) SIZE

```

MPI\_TYPE\_SIZE and MPI\_TYPE\_SIZE\_X set the value of `size` to the total size, in bytes, of the entries in the type signature associated with `datatype`; i.e., the total size of the data in a message that would be created with this `datatype`. Entries that occur multiple times in the `datatype` are counted with their multiplicity. For both functions, if the `OUT` parameter cannot express the value to be returned (e.g., if the parameter is too small to hold the output value), it is set to `MPI_UNDEFINED`.

#### 4.1.6 Lower-Bound and Upper-Bound Markers

It is often convenient to define explicitly the lower bound and upper bound of a type map, and override the definition given on page 23. This allows one to define a `datatype` that has “holes” at its beginning or its end, or a `datatype` with entries that extend above the upper bound or below the lower bound. Examples of such usage are provided in Section 4.1.14. Also, the user may want to override the alignment rules that are used to compute upper bounds and extents. E.g., a C compiler may allow the user to override default alignment rules for some of the structures within a program. The user has to specify explicitly the bounds of the `datatypes` that match these structures.



To achieve this, we add two additional conceptual datatypes, *lb\_marker* and *ub\_marker*, that represent the lower bound and upper bound of a datatype. These conceptual datatypes occupy no space ( $extent(lb\_marker) = extent(ub\_marker) = 0$ ). They do not affect the size or count of a datatype, and do not affect the content of a message created with this datatype. However, they do affect the definition of the extent of a datatype and, therefore, affect the outcome of a replication of this datatype by a datatype constructor.

**Example 4.9** A call to `MPI_TYPE_CREATE_RESIZED(MPI_INT, -3, 9, type1)` creates a new datatype that has an extent of 9 (from -3 to 5, 5 included), and contains an integer at displacement 0. This is the datatype defined by the typemap  $\{(lb\_marker, -3), (int, 0), (ub\_marker, 6)\}$ . If this type is replicated twice by a call to `MPI_TYPE_CONTIGUOUS(2, type1, type2)` then the newly created type can be described by the typemap  $\{(lb\_marker, -3), (int, 0), (int, 9), (ub\_marker, 15)\}$ . (An entry of type *ub\_marker* can be deleted if there is another entry of type *ub\_marker* with a higher displacement; an entry of type *lb\_marker* can be deleted if there is another entry of type *lb\_marker* with a lower displacement.)

In general, if

$$Typemap = \{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\},$$

then the *lower bound* of *Typemap* is defined to be

$$lb(Typemap) = \begin{cases} \min_j disp_j & \text{if no entry has type } lb\_marker \\ \min_j \{disp_j \text{ such that } type_j = lb\_marker\} & \text{otherwise} \end{cases}$$

Similarly, the *upper bound* of *Typemap* is defined to be

$$ub(Typemap) = \begin{cases} \max_j (disp_j + sizeof(type_j)) + \epsilon & \text{if no entry has type } ub\_marker \\ \max_j \{disp_j \text{ such that } type_j = ub\_marker\} & \text{otherwise} \end{cases}$$

Then

$$extent(Typemap) = ub(Typemap) - lb(Typemap)$$

If  $type_i$  requires alignment to a byte address that is a multiple of  $k_i$ , then  $\epsilon$  is the least non-negative increment needed to round  $extent(Typemap)$  to the next multiple of  $\max_i k_i$ . In Fortran, it is implementation dependent whether the MPI implementation computes the alignments  $k_i$  according to the alignments used by the compiler in common blocks, SEQUENCE derived types, BIND(C) derived types, or derived types that are neither SEQUENCE nor BIND(C).

The formal definitions given for the various datatype constructors apply now, with the amended definition of *extent*.

*Rationale.* Before Fortran 2003, `MPI_TYPE_CREATE_STRUCT` could be applied to Fortran common blocks and SEQUENCE derived types. With Fortran 2003, this list was extended by BIND(C) derived types and MPI implementors have implemented the alignments  $k_i$  differently, i.e., some based on the alignments used in SEQUENCE derived types, and others according to BIND(C) derived types. (*End of rationale.*)

*Advice to implementors.* In Fortran, it is generally recommended to use BIND(C) derived types instead of common blocks or SEQUENCE derived types. Therefore it is recommended to calculate the alignments  $k_i$  based on BIND(C) derived types. (*End of advice to implementors.*)

*Advice to users.* Structures combining different basic datatypes should be defined so that there will be no gaps based on alignment rules. If such a datatype is used to create an array of structures, users should also avoid an alignment-gap at the end of the structure. In MPI communication, the content of such gaps would not be communicated into the receiver's buffer. For example, such an alignment-gap may occur between an odd number of `floats` or `REALs` before a `double` or `DOUBLE PRECISION` data. Such gaps may be added explicitly to both the structure and the MPI derived datatype handle because the communication of a contiguous derived datatype may be significantly faster than the communication of one that is non-contiguous because of such alignment-gaps.

Example: Instead of

```

TYPE, :: my_data
  REAL, DIMENSION(3) :: x
  ! there may be a gap of the size of one REAL
  ! if the alignment of a DOUBLE PRECISION is
  ! two times the size of a REAL
  DOUBLE PRECISION :: p
END TYPE

```

one should define

```

TYPE, :: my_data
  REAL, DIMENSION(3) :: x
  REAL :: gap1
  DOUBLE PRECISION :: p
END TYPE

```

and also include `gap1` in the matching MPI derived datatype. It is required that all processes in a communication add the same gaps, i.e., defined with the same basic datatype. Both the original and the modified structures are portable, but may have different performance implications for the communication and memory accesses during computation on systems with different alignment values.

In principle, a compiler may define an additional alignment rule for structures, e.g., to use at least 4 or 8 byte alignment, although the content may have a  $max_i k_i$  alignment less than this structure alignment. To maintain portability, users should always resize structure derived datatype handles if used in an array of structures, see the Example in Section 17.1.15. (*End of advice to users.*)

#### 4.1.7 Extent and Bounds of Datatypes

`MPI_TYPE_GET_EXTENT(datatype, lb, extent)`

IN	<code>datatype</code>	datatype to get information on (handle)
OUT	<code>lb</code>	lower bound of datatype (integer)
OUT	<code>extent</code>	extent of datatype (integer)

```

int MPI_Type_get_extent(MPI_Datatype datatype, MPI_Aint *lb,
                        MPI_Aint *extent)
MPI_Type_get_extent(datatype, lb, extent, ierror)
    TYPE(MPI_Datatype), INTENT(IN) :: datatype
    INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(OUT) :: lb, extent
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror
MPI_TYPE_GET_EXTENT(DATATYPE, LB, EXTENT, IERROR)
    INTEGER DATATYPE, IERROR
    INTEGER(KIND = MPI_ADDRESS_KIND) LB, EXTENT

MPI_TYPE_GET_EXTENT_X(datatype, lb, extent)
    IN      datatype          datatype to get information on (handle)
    OUT     lb                lower bound of datatype (integer)
    OUT     extent            extent of datatype (integer)

int MPI_Type_get_extent_x(MPI_Datatype datatype, MPI_Count *lb,
                          MPI_Count *extent)
MPI_Type_get_extent_x(datatype, lb, extent, ierror)
    TYPE(MPI_Datatype), INTENT(IN) :: datatype
    INTEGER(KIND = MPI_COUNT_KIND), INTENT(OUT) :: lb, extent
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror
MPI_TYPE_GET_EXTENT_X(DATATYPE, LB, EXTENT, IERROR)
    INTEGER DATATYPE, IERROR
    INTEGER(KIND = MPI_COUNT_KIND) LB, EXTENT

Returns the lower bound and the extent of datatype (as defined in Equation 4.1 page
2).
For both functions, if either OUT parameter cannot express the value to be returned
(e.g., if the parameter is too small to hold the output value), it is set to MPI_UNDEFINED.
MPI allows one to change the extent of a datatype, using lower bound and upper bound
markers. This provides control over the stride of successive datatypes that are replicated
by datatype constructors, or are replicated by the count argument in a send or receive call.

MPI_TYPE_CREATE_RESIZED(oldtype, lb, extent, newtype)
    IN      oldtype          input datatype (handle)
    IN      lb                new lower bound of datatype (integer)
    IN      extent            new extent of datatype (integer)
    OUT     newtype           output datatype (handle)

int MPI_Type_create_resized(MPI_Datatype oldtype, MPI_Aint lb, MPI_Aint
                           extent, MPI_Datatype *newtype)

```

```

1 MPI_Type_create_resized(oldtype, lb, extent, newtype, ierror)
2     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: lb, extent
3     TYPE(MPI_Datatype), INTENT(IN) :: oldtype
4     TYPE(MPI_Datatype), INTENT(OUT) :: newtype
5     INTEGER, OPTIONAL, INTENT(OUT) :: ierror

```

```

6 MPI_TYPE_CREATE_RESIZED(OLDTYPE, LB, EXTENT, NEWTYPE, IERROR)
7     INTEGER OLDTYPE, NEWTYPE, IERROR
8     INTEGER(KIND=MPI_ADDRESS_KIND) LB, EXTENT
9

```

10 Returns in `newtype` a handle to a new datatype that is identical to `oldtype`, except that
11 the lower bound of this new datatype is set to be `lb`, and its upper bound is set to be `lb`
12 + `extent`. Any previous `lb` and `ub` markers are erased, and a new pair of lower bound and
13 upper bound markers are put in the positions indicated by the `lb` and `extent` arguments.
14 This affects the behavior of the datatype when used in communication operations, with
15 `count > 1`, and when used in the construction of new derived datatypes.

16

#### 17 4.1.8 True Extent of Datatypes

18

19 Suppose we implement `gather` (see also Section 5.5) as a spanning tree implemented on
20 top of point-to-point routines. Since the receive buffer is only valid on the root pro-
21 cess, one will need to allocate some temporary space for receiving data on intermedi-
22 ate nodes. However, the datatype extent cannot be used as an estimate of the amount
23 of space that needs to be allocated, if the user has modified the extent, for example
24 by using `MPI_TYPE_CREATE_RESIZED`. The functions `MPI_TYPE_GET_TRUE_EXTENT`
25 and `MPI_TYPE_GET_TRUE_EXTENT_X` are provided which return the true extent of the
26 datatype.

27

```

28 MPI_TYPE_GET_TRUE_EXTENT(datatype, true_lb, true_extent)

```

29

30	IN	datatype	datatype to get information on (handle)
31	OUT	true_lb	true lower bound of datatype (integer)
32	OUT	true_extent	true size of datatype (integer)

33

```

34 int MPI_Type_get_true_extent(MPI_Datatype datatype, MPI_Aint *true_lb,
35                             MPI_Aint *true_extent)
36

```

```

37 MPI_Type_get_true_extent(datatype, true_lb, true_extent, ierror)
38     TYPE(MPI_Datatype), INTENT(IN) :: datatype
39     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(OUT) :: true_lb, true_extent
40     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
41

```

```

42 MPI_TYPE_GET_TRUE_EXTENT(DATATYPE, TRUE_LB, TRUE_EXTENT, IERROR)
43     INTEGER DATATYPE, IERROR
44     INTEGER(KIND = MPI_ADDRESS_KIND) TRUE_LB, TRUE_EXTENT
45

```

46

47

48

```

MPI_TYPE_GET_TRUE_EXTENT_X(datatype, true_lb, true_extent) 1
    IN      datatype          datatype to get information on (handle) 2
    OUT     true_lb           true lower bound of datatype (integer) 3
    OUT     true_extent       true size of datatype (integer) 4

```

```

int MPI_Type_get_true_extent_x(MPI_Datatype datatype, MPI_Count *true_lb, 7
                               MPI_Count *true_extent) 8

```

```

MPI_Type_get_true_extent_x(datatype, true_lb, true_extent, ierror) 10
    TYPE(MPI_Datatype), INTENT(IN) :: datatype 11
    INTEGER(KIND = MPI_COUNT_KIND), INTENT(OUT) :: true_lb, true_extent 12
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror 13

```

```

MPI_TYPE_GET_TRUE_EXTENT_X(DATATYPE, TRUE_LB, TRUE_EXTENT, IERROR) 14
    INTEGER DATATYPE, IERROR 15
    INTEGER(KIND = MPI_COUNT_KIND) TRUE_LB, TRUE_EXTENT 16

```

`true_lb` returns the offset of the lowest unit of store which is addressed by the datatype, i.e., the lower bound of the corresponding typemap, ignoring explicit lower bound markers. `true_extent` returns the true size of the datatype, i.e., the extent of the corresponding typemap, ignoring explicit lower bound and upper bound markers, and performing no rounding for alignment. If the typemap associated with `datatype` is

$$Typemap = \{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}$$

Then

$$true\_lb(Typemap) = \min_j \{disp_j : type_j \neq lb\_marker, ub\_marker\},$$

$$true\_ub(Typemap) = \max_j \{disp_j + sizeof(type_j) : type_j \neq lb\_marker, ub\_marker\},$$

and

$$true\_extent(Typemap) = true\_ub(Typemap) - true\_lb(Typemap).$$

(Readers should compare this with the definitions in Section 4.1.6 and Section 4.1.7, which describe the function `MPI_TYPE_GET_EXTENT`.)

The `true_extent` is the minimum number of bytes of memory necessary to hold a datatype, uncompressed.

For both functions, if either OUT parameter cannot express the value to be returned (e.g., if the parameter is too small to hold the output value), it is set to `MPI_UNDEFINED`.

#### 4.1.9 Commit and Free

A datatype object has to be *committed* before it can be used in a communication. As an argument in datatype constructors, uncommitted and also committed datatypes can be used. There is no need to commit basic datatypes. They are “pre-committed.”

```

1 MPI_TYPE_COMMIT(datatype)
2     INOUT    datatype                datatype that is committed (handle)
3
4
5 int MPI_Type_commit(MPI_Datatype *datatype)
6 MPI_Type_commit(datatype, ierror)
7     TYPE(MPI_Datatype), INTENT(INOUT) :: datatype
8     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
9
10 MPI_TYPE_COMMIT(DATATYPE, IERROR)
11     INTEGER DATATYPE, IERROR

```

The commit operation commits the datatype, that is, the formal description of a communication buffer, not the content of that buffer. Thus, after a datatype has been committed, it can be repeatedly reused to communicate the changing content of a buffer or, indeed, the content of different buffers, with different starting addresses.

*Advice to implementors.* The system may “compile” at commit time an internal representation for the datatype that facilitates communication, e.g., change from a compacted representation to a flat representation of the datatype, and select the most convenient transfer mechanism. (*End of advice to implementors.*)

MPI\_TYPE\_COMMIT will accept a committed datatype; in this case, it is equivalent to a no-op.

**Example 4.10** The following code fragment gives examples of using MPI\_TYPE\_COMMIT.

```

26 INTEGER type1, type2
27 CALL MPI_TYPE_CONTIGUOUS(5, MPI_REAL, type1, ierr)
28     ! new type object created
29 CALL MPI_TYPE_COMMIT(type1, ierr)
30     ! now type1 can be used for communication
31 type2 = type1
32     ! type2 can be used for communication
33     ! (it is a handle to same object as type1)
34 CALL MPI_TYPE_VECTOR(3, 5, 4, MPI_REAL, type1, ierr)
35     ! new uncommitted type object created
36 CALL MPI_TYPE_COMMIT(type1, ierr)
37     ! now type1 can be used anew for communication
38
39
40
41 MPI_TYPE_FREE(datatype)
42     INOUT    datatype                datatype that is freed (handle)
43
44
45 int MPI_Type_free(MPI_Datatype *datatype)
46 MPI_Type_free(datatype, ierror)
47     TYPE(MPI_Datatype), INTENT(INOUT) :: datatype
48     INTEGER, OPTIONAL, INTENT(OUT) :: ierror

```

```
MPI_TYPE_FREE(DATATYPE, IERROR)
    INTEGER DATATYPE, IERROR
```

Marks the datatype object associated with `datatype` for deallocation and sets `datatype` to `MPI_DATATYPE_NULL`. Any communication that is currently using this datatype will complete normally. Freeing a datatype does not affect any other datatype that was built from the freed datatype. The system behaves as if input datatype arguments to derived datatype constructors are passed by value.

*Advice to implementors.* The implementation may keep a reference count of active communications that use the datatype, in order to decide when to free it. Also, one may implement constructors of derived datatypes so that they keep pointers to their datatype arguments, rather than copying them. In this case, one needs to keep track of active datatype definition references in order to know when a datatype object can be freed. (*End of advice to implementors.*)

#### 4.1.10 Duplicating a Datatype

```
MPI_TYPE_DUP(oldtype, newtype)
```

IN	oldtype	datatype (handle)
OUT	newtype	copy of oldtype (handle)

```
int MPI_Type_dup(MPI_Datatype oldtype, MPI_Datatype *newtype)
```

```
MPI_Type_dup(oldtype, newtype, ierror)
    TYPE(MPI_Datatype), INTENT(IN) :: oldtype
    TYPE(MPI_Datatype), INTENT(OUT) :: newtype
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror
```

```
MPI_TYPE_DUP(OLDTYPE, NEWTYPE, IERROR)
    INTEGER OLDTYPE, NEWTYPE, IERROR
```

`MPI_TYPE_DUP` is a type constructor which duplicates the existing `oldtype` with associated key values. For each key value, the respective copy callback function determines the attribute value associated with this key in the new communicator; one particular action that a copy callback may take is to delete the attribute from the new datatype. Returns in `newtype` a new datatype with exactly the same properties as `oldtype` and any copied cached information, see Section 6.7.4. The new datatype has identical upper bound and lower bound and yields the same net result when fully decoded with the functions in Section 4.1.13. The `newtype` has the same committed state as the old `oldtype`.

#### 4.1.11 Use of General Datatypes in Communication

Handles to derived datatypes can be passed to a communication call wherever a datatype argument is required. A call of the form `MPI_SEND(buf, count, datatype, ...)`, where `count > 1`, is interpreted as if the call was passed a new datatype which is the concatenation of `count` copies of `datatype`. Thus, `MPI_SEND(buf, count, datatype, dest, tag, comm)` is equivalent to,

```

1 MPI_TYPE_CONTIGUOUS(count, datatype, newtype)
2 MPI_TYPE_COMMIT(newtype)
3 MPI_SEND(buf, 1, newtype, dest, tag, comm)
4 MPI_TYPE_FREE(newtype).

```

5 Similar statements apply to all other communication functions that have a `count` and  
6 `datatype` argument.

7 Suppose that a send operation `MPI_SEND(buf, count, datatype, dest, tag, comm)` is  
8 executed, where `datatype` has type map,

```

9       $\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}$ ,
10

```

11 and extent *extent*. (Explicit lower bound and upper bound markers are not listed in the  
12 type map, but they affect the value of *extent*.) The send operation sends  $n \cdot \text{count}$  entries,  
13 where entry  $i \cdot n + j$  is at location  $addr_{i,j} = \text{buf} + \text{extent} \cdot i + disp_j$  and has type  $type_j$ , for  
14  $i = 0, \dots, \text{count} - 1$  and  $j = 0, \dots, n - 1$ . These entries need not be contiguous, nor distinct;  
15 their order can be arbitrary.

16 The variable stored at address  $addr_{i,j}$  in the calling program should be of a type that  
17 matches  $type_j$ , where type matching is defined as in Section 3.3.1. The message sent contains  
18  $n \cdot \text{count}$  entries, where entry  $i \cdot n + j$  has type  $type_j$ .

19 Similarly, suppose that a receive operation `MPI_RECV(buf, count, datatype, source, tag,`  
20 `comm, status)` is executed, where `datatype` has type map,

```

21       $\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}$ ,
22

```

23 with extent *extent*. (Again, explicit lower bound and upper bound markers are not listed in  
24 the type map, but they affect the value of *extent*.) This receive operation receives  $n \cdot \text{count}$   
25 entries, where entry  $i \cdot n + j$  is at location  $\text{buf} + \text{extent} \cdot i + disp_j$  and has type  $type_j$ . If the  
26 incoming message consists of  $k$  elements, then we must have  $k \leq n \cdot \text{count}$ ; the  $i \cdot n + j$ -th  
27 element of the message should have a type that matches  $type_j$ .

28 Type matching is defined according to the type signature of the corresponding datatypes,  
29 that is, the sequence of basic type components. Type matching does not depend on some  
30 aspects of the datatype definition, such as the displacements (layout in memory) or the  
31 intermediate types used.

32 **Example 4.11** This example shows that type matching is defined in terms of the basic  
33 types that a derived type consists of.

```

34 ...
35 ...
36 CALL MPI_TYPE_CONTIGUOUS(2, MPI_REAL, type2, ...)
37 CALL MPI_TYPE_CONTIGUOUS(4, MPI_REAL, type4, ...)
38 CALL MPI_TYPE_CONTIGUOUS(2, type2, type22, ...)
39 ...
40 CALL MPI_SEND(a, 4, MPI_REAL, ...)
41 CALL MPI_SEND(a, 2, type2, ...)
42 CALL MPI_SEND(a, 1, type22, ...)
43 CALL MPI_SEND(a, 1, type4, ...)
44 ...
45 CALL MPI_RECV(a, 4, MPI_REAL, ...)
46 CALL MPI_RECV(a, 2, type2, ...)
47 CALL MPI_RECV(a, 1, type22, ...)
48 CALL MPI_RECV(a, 1, type4, ...)

```



Each of the sends matches any of the receives. 1

A datatype may specify overlapping entries. The use of such a datatype in a receive operation is erroneous. (This is erroneous even if the actual message received is short enough not to write any entry more than once.) 2 3 4

Suppose that `MPI_RECV(buf, count, datatype, dest, tag, comm, status)` is executed, where `datatype` has type map, 5 6

$$\{(type_0, disp_0), \dots, (type_{n-1}, disp_{n-1})\}. 7 8$$

The received message need not fill all the receive buffer, nor does it need to fill a number of locations which is a multiple of  $n$ . Any number,  $k$ , of basic elements can be received, where  $0 \leq k \leq count \cdot n$ . The number of basic elements received can be retrieved from `status` using the query functions `MPI_GET_ELEMENTS` or `MPI_GET_ELEMENTS_X`. 9 10 11 12 13 14

`MPI_GET_ELEMENTS(status, datatype, count)` 15

IN	status	return status of receive operation (Status)	16
IN	datatype	datatype used by receive operation (handle)	17
OUT	count	number of received basic elements (integer)	18

```
int MPI_Get_elements(const MPI_Status *status, MPI_Datatype datatype,
                    int *count) 19 20 21 22
```

```
MPI_Get_elements(status, datatype, count, ierror) 23 24
    TYPE(MPI_Status), INTENT(IN) :: status 25
    TYPE(MPI_Datatype), INTENT(IN) :: datatype 26
    INTEGER, INTENT(OUT) :: count 27
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror 28
```

```
MPI_GET_ELEMENTS(STATUS, DATATYPE, COUNT, IERROR) 29
    INTEGER STATUS(MPI_STATUS_SIZE), DATATYPE, COUNT, IERROR 30
```

`MPI_GET_ELEMENTS_X(status, datatype, count)` 31 32 33

IN	status	return status of receive operation (Status)	34
IN	datatype	datatype used by receive operation (handle)	35
OUT	count	number of received basic elements (integer)	36

```
int MPI_Get_elements_x(const MPI_Status *status, MPI_Datatype datatype,
                      MPI_Count *count) 37 38 39 40 41
```

```
MPI_Get_elements_x(status, datatype, count, ierror) 42 43
    TYPE(MPI_Status), INTENT(IN) :: status 44
    TYPE(MPI_Datatype), INTENT(IN) :: datatype 45
    INTEGER(KIND = MPI_COUNT_KIND), INTENT(OUT) :: count 46
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror 47
```

```
MPI_GET_ELEMENTS_X(STATUS, DATATYPE, COUNT, IERROR) 48
```

```

1     INTEGER STATUS(MPI_STATUS_SIZE), DATATYPE, IERROR
2     INTEGER(KIND=MPI_COUNT_KIND) COUNT

```

The `datatype` argument should match the argument provided by the receive call that set the `status` variable. For both functions, if the `OUT` parameter cannot express the value to be returned (e.g., if the parameter is too small to hold the output value), it is set to `MPI_UNDEFINED`.

The previously defined function `MPI_GET_COUNT` (Section 3.2.5), has a different behavior. It returns the number of “top-level entries” received, i.e. the number of “copies” of type `datatype`. In the previous example, `MPI_GET_COUNT` may return any integer value  $k$ , where  $0 \leq k \leq \text{count}$ . If `MPI_GET_COUNT` returns  $k$ , then the number of basic elements received (and the value returned by `MPI_GET_ELEMENTS` or `MPI_GET_ELEMENTS_X`) is  $n \cdot k$ . If the number of basic elements received is not a multiple of  $n$ , that is, if the receive operation has not received an integral number of `datatype` “copies,” then `MPI_GET_COUNT` sets the value of `count` to `MPI_UNDEFINED`.

**Example 4.12** Usage of `MPI_GET_COUNT` and `MPI_GET_ELEMENTS`.

```

18     ...
19     CALL MPI_TYPE_CONTIGUOUS(2, MPI_REAL, Type2, ierr)
20     CALL MPI_TYPE_COMMIT(Type2, ierr)
21     ...
22     CALL MPI_COMM_RANK(comm, rank, ierr)
23     IF (rank.EQ.0) THEN
24         CALL MPI_SEND(a, 2, MPI_REAL, 1, 0, comm, ierr)
25         CALL MPI_SEND(a, 3, MPI_REAL, 1, 0, comm, ierr)
26     ELSE IF (rank.EQ.1) THEN
27         CALL MPI_RECV(a, 2, Type2, 0, 0, comm, stat, ierr)
28         CALL MPI_GET_COUNT(stat, Type2, i, ierr)      ! returns i=1
29         CALL MPI_GET_ELEMENTS(stat, Type2, i, ierr)  ! returns i=2
30         CALL MPI_RECV(a, 2, Type2, 0, 0, comm, stat, ierr)
31         CALL MPI_GET_COUNT(stat, Type2, i, ierr)    ! returns i=MPI_UNDEFINED
32         CALL MPI_GET_ELEMENTS(stat, Type2, i, ierr) ! returns i=3
33     END IF

```

The functions `MPI_GET_ELEMENTS` and `MPI_GET_ELEMENTS_X` can also be used after a probe to find the number of elements in the probed message. Note that the `MPI_GET_COUNT`, `MPI_GET_ELEMENTS`, and `MPI_GET_ELEMENTS_X` return the same values when they are used with basic datatypes as long as the limits of their respective count arguments are not exceeded.

*Rationale.* The extension given to the definition of `MPI_GET_COUNT` seems natural: one would expect this function to return the value of the count argument, when the receive buffer is filled. Sometimes `datatype` represents a basic unit of data one wants to transfer, for example, a record in an array of records (structures). One should be able to find out how many components were received without bothering to divide by the number of elements in each component. However, on other occasions, `datatype` is used to define a complex layout of data in the receiver memory, and does not represent a basic unit of data for transfers. In such cases, one needs to use the function `MPI_GET_ELEMENTS` or `MPI_GET_ELEMENTS_X`. (*End of rationale.*)

*Advice to implementors.* The definition implies that a receive cannot change the value of storage outside the entries defined to compose the communication buffer. In particular, the definition implies that padding space in a structure should not be modified when such a structure is copied from one process to another. This would prevent the obvious optimization of copying the structure, together with the padding, as one contiguous block. The implementation is free to do this optimization when it does not impact the outcome of the computation. The user can “force” this optimization by explicitly including padding as part of the message. (*End of advice to implementors.*)

#### 4.1.12 Correct Use of Addresses

Successively declared variables in C or Fortran are not necessarily stored at contiguous locations. Thus, care must be exercised that displacements do not cross from one variable to another. Also, in machines with a segmented address space, addresses are not unique and address arithmetic has some peculiar properties. Thus, the use of *addresses*, that is, displacements relative to the start address `MPI_BOTTOM`, has to be restricted.

Variables belong to the same *sequential storage* if they belong to the same array, to the same `COMMON` block in Fortran, or to the same structure in C. Valid addresses are defined recursively as follows:

1. The function `MPI_GET_ADDRESS` returns a valid address, when passed as argument a variable of the calling program.
2. The `buf` argument of a communication function evaluates to a valid address, when passed as argument a variable of the calling program.
3. If `v` is a valid address, and `i` is an integer, then `v+i` is a valid address, provided `v` and `v+i` are in the same sequential storage.

A correct program uses only valid addresses to identify the locations of entries in communication buffers. Furthermore, if `u` and `v` are two valid addresses, then the (integer) difference `u - v` can be computed only if both `u` and `v` are in the same sequential storage. No other arithmetic operations can be meaningfully executed on addresses.

The rules above impose no constraints on the use of derived datatypes, as long as they are used to define a communication buffer that is wholly contained within the same sequential storage. However, the construction of a communication buffer that contains variables that are not within the same sequential storage must obey certain restrictions. Basically, a communication buffer with variables that are not within the same sequential storage can be used only by specifying in the communication call `buf = MPI_BOTTOM`, `count = 1`, and using a `datatype` argument where all displacements are valid (absolute) addresses.

*Advice to users.* It is not expected that MPI implementations will be able to detect erroneous, “out of bound” displacements — unless those overflow the user address space — since the MPI call may not know the extent of the arrays and records in the host program. (*End of advice to users.*)

*Advice to implementors.* There is no need to distinguish (absolute) addresses and (relative) displacements on a machine with contiguous address space: `MPI_BOTTOM` is zero, and both addresses and displacements are integers. On machines where the

1 distinction is required, addresses are recognized as expressions that involve  
 2 MPI\_BOTTOM. (*End of advice to implementors.*)

### 4.1.13 Decoding a Datatype

5 MPI datatype objects allow users to specify an arbitrary layout of data in memory. There  
 6 are several cases where accessing the layout information in opaque datatype objects would  
 7 be useful. The opaque datatype object has found a number of uses outside MPI. Further-  
 8 more, a number of tools wish to display internal information about a datatype. To achieve  
 9 this, datatype decoding functions are provided. The two functions in this section are used  
 10 together to decode datatypes to recreate the calling sequence used in their initial defini-  
 11 tion. These can be used to allow a user to determine the type map and type signature of a  
 12 datatype.  
 13

15 MPI\_TYPE\_GET\_ENVELOPE(datatype, num\_integers, num\_addresses, num\_datatypes,  
 16 combiner)

17	IN	datatype	datatype to access (handle)
18			
19	OUT	num_integers	number of input integers used in the call constructing
20			combiner (non-negative integer)
21	OUT	num_addresses	number of input addresses used in the call construct-
22			ing combiner (non-negative integer)
23			
24	OUT	num_datatypes	number of input datatypes used in the call construct-
25			ing combiner (non-negative integer)
26	OUT	combiner	combiner (state)
27			

28 int MPI\_Type\_get\_envelope(MPI\_Datatype datatype, int \*num\_integers,  
 29 int \*num\_addresses, int \*num\_datatypes, int \*combiner)

30 MPI\_Type\_get\_envelope(datatype, num\_integers, num\_addresses, num\_datatypes,  
 31 combiner, ierror)

32 TYPE(MPI\_Datatype), INTENT(IN) :: datatype  
 33 INTEGER, INTENT(OUT) :: num\_integers, num\_addresses, num\_datatypes,  
 34 combiner  
 35 INTEGER, OPTIONAL, INTENT(OUT) :: ierror

37 MPI\_TYPE\_GET\_ENVELOPE(DATATYPE, NUM\_INTEGERS, NUM\_ADDRESSES, NUM\_DATATYPES,  
 38 COMBINER, IERROR)

39 INTEGER DATATYPE, NUM\_INTEGERS, NUM\_ADDRESSES, NUM\_DATATYPES, COMBINER,  
 40 IERROR

41  
 42 For the given datatype, MPI\_TYPE\_GET\_ENVELOPE returns information on the num-  
 43 ber and type of input arguments used in the call that created the datatype. The number-of-  
 44 arguments values returned can be used to provide sufficiently large arrays in the decoding  
 45 routine MPI\_TYPE\_GET\_CONTENTS. This call and the meaning of the returned values is  
 46 described below. The combiner reflects the MPI datatype constructor call that was used in  
 47 creating datatype.  
 48

*Rationale.* By requiring that the `combiner` reflect the constructor used in the creation of the `datatype`, the decoded information can be used to effectively recreate the calling sequence used in the original creation. This is the most useful information and was felt to be reasonable even though it constrains implementations to remember the original constructor sequence even if the internal representation is different.

The decoded information keeps track of datatype duplications. This is important as one needs to distinguish between a predefined datatype and a dup of a predefined datatype. The former is a constant object that cannot be freed, while the latter is a derived datatype that can be freed. (*End of rationale.*)

The list in Table 4.1 has the values that can be returned in `combiner` on the left and the call associated with them on the right.

<code>MPI_COMBINER_NAMED</code>	a named predefined datatype
<code>MPI_COMBINER_DUP</code>	<code>MPI_TYPE_DUP</code>
<code>MPI_COMBINER_CONTIGUOUS</code>	<code>MPI_TYPE_CONTIGUOUS</code>
<code>MPI_COMBINER_VECTOR</code>	<code>MPI_TYPE_VECTOR</code>
<code>MPI_COMBINER_HVECTOR</code>	<code>MPI_TYPE_CREATE_HVECTOR</code>
<code>MPI_COMBINER_INDEXED</code>	<code>MPI_TYPE_INDEXED</code>
<code>MPI_COMBINER_HINDEXED</code>	<code>MPI_TYPE_CREATE_HINDEXED</code>
<code>MPI_COMBINER_INDEXED_BLOCK</code>	<code>MPI_TYPE_CREATE_INDEXED_BLOCK</code>
<code>MPI_COMBINER_HINDEXED_BLOCK</code>	<code>MPI_TYPE_CREATE_HINDEXED_BLOCK</code>
<code>MPI_COMBINER_STRUCT</code>	<code>MPI_TYPE_CREATE_STRUCT</code>
<code>MPI_COMBINER_SUBARRAY</code>	<code>MPI_TYPE_CREATE_SUBARRAY</code>
<code>MPI_COMBINER_DARRAY</code>	<code>MPI_TYPE_CREATE_DARRAY</code>
<code>MPI_COMBINER_F90_REAL</code>	<code>MPI_TYPE_CREATE_F90_REAL</code>
<code>MPI_COMBINER_F90_COMPLEX</code>	<code>MPI_TYPE_CREATE_F90_COMPLEX</code>
<code>MPI_COMBINER_F90_INTEGER</code>	<code>MPI_TYPE_CREATE_F90_INTEGER</code>
<code>MPI_COMBINER_RESIZED</code>	<code>MPI_TYPE_CREATE_RESIZED</code>

Table 4.1: `combiner` values returned from `MPI_TYPE_GET_ENVELOPE`

If `combiner` is `MPI_COMBINER_NAMED` then `datatype` is a named predefined datatype.

The actual arguments used in the creation call for a `datatype` can be obtained using `MPI_TYPE_GET_CONTENTS`.

```

1 MPI_TYPE_GET_CONTENTS(datatype, max_integers, max_addresses, max_datatypes,
2     array_of_integers, array_of_addresses, array_of_datatypes)
3     IN      datatype          datatype to access (handle)
4     IN      max_integers      number of elements in array_of_integers (non-negative
5     integer)
6
7     IN      max_addresses     number of elements in array_of_addresses (non-negative
8     integer)
9
10    IN      max_datatypes     number of elements in array_of_datatypes (non-negative
11    integer)
12    OUT     array_of_integers  contains integer arguments used in constructing
13    datatype (array of integers)
14    OUT     array_of_addresses contains address arguments used in constructing
15    datatype (array of integers)
16    OUT     array_of_datatypes contains datatype arguments used in constructing
17    datatype (array of handles)
18

```

```

19 int MPI_Type_get_contents(MPI_Datatype datatype, int max_integers,
20     int max_addresses, int max_datatypes, int array_of_integers[],
21     MPI_Aint array_of_addresses[],
22     MPI_Datatype array_of_datatypes[])
23

```

```

24 MPI_Type_get_contents(datatype, max_integers, max_addresses, max_datatypes,
25     array_of_integers, array_of_addresses, array_of_datatypes,
26     ierror)
27     TYPE(MPI_Datatype), INTENT(IN) :: datatype
28     INTEGER, INTENT(IN) :: max_integers, max_addresses, max_datatypes
29     INTEGER, INTENT(OUT) :: array_of_integers(max_integers)
30     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(OUT) ::
31     array_of_addresses(max_addresses)
32     TYPE(MPI_Datatype), INTENT(OUT) :: array_of_datatypes(max_datatypes)
33     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
34

```

```

35 MPI_TYPE_GET_CONTENTS(DATATYPE, MAX_INTEGERS, MAX_ADDRESSES, MAX_DATATYPES,
36     ARRAY_OF_INTEGERS, ARRAY_OF_ADDRESSES, ARRAY_OF_DATATYPES,
37     IERROR)
38     INTEGER DATATYPE, MAX_INTEGERS, MAX_ADDRESSES, MAX_DATATYPES,
39     ARRAY_OF_INTEGERS(*), ARRAY_OF_DATATYPES(*), IERROR
40     INTEGER(KIND=MPI_ADDRESS_KIND) ARRAY_OF_ADDRESSES(*)
41

```

datatype must be a predefined unnamed or a derived datatype; the call is erroneous if datatype is a predefined named datatype.

The values given for max\_integers, max\_addresses, and max\_datatypes must be at least as large as the value returned in num\_integers, num\_addresses, and num\_datatypes, respectively, in the call MPI\_TYPE\_GET\_ENVELOPE for the same datatype argument.

*Rationale.* The arguments max\_integers, max\_addresses, and max\_datatypes allow for error checking in the call. (*End of rationale.*)

The datatypes returned in `array_of_datatypes` are handles to datatype objects that are equivalent to the datatypes used in the original construction call. If these were derived datatypes, then the returned datatypes are new datatype objects, and the user is responsible for freeing these datatypes with `MPI_TYPE_FREE`. If these were predefined datatypes, then the returned datatype is equal to that (constant) predefined datatype and cannot be freed.

The committed state of returned derived datatypes is undefined, i.e., the datatypes may or may not be committed. Furthermore, the content of attributes of returned datatypes is undefined.

Note that `MPI_TYPE_GET_CONTENTS` can be invoked with a datatype argument that was constructed using `MPI_TYPE_CREATE_F90_REAL`, `MPI_TYPE_CREATE_F90_INTEGER`, or `MPI_TYPE_CREATE_F90_COMPLEX` (an unnamed predefined datatype). In such a case, an empty `array_of_datatypes` is returned.

*Rationale.* The definition of datatype equivalence implies that equivalent predefined datatypes are equal. By requiring the same handle for named predefined datatypes, it is possible to use the `==` or `.EQ.` comparison operator to determine the datatype involved. (*End of rationale.*)

*Advice to implementors.* The datatypes returned in `array_of_datatypes` must appear to the user as if each is an equivalent copy of the datatype used in the type constructor call. Whether this is done by creating a new datatype or via another mechanism such as a reference count mechanism is up to the implementation as long as the semantics are preserved. (*End of advice to implementors.*)

*Rationale.* The committed state and attributes of the returned datatype is deliberately left vague. The datatype used in the original construction may have been modified since its use in the constructor call. Attributes can be added, removed, or modified as well as having the datatype committed. The semantics given allow for a reference count implementation without having to track these changes. (*End of rationale.*)

In the deprecated datatype constructor calls, the address arguments in Fortran are of type `INTEGER`. In the preferred calls, the address arguments are of type `INTEGER(KIND=MPI_ADDRESS_KIND)`. The call `MPI_TYPE_GET_CONTENTS` returns all addresses in an argument of type `INTEGER(KIND=MPI_ADDRESS_KIND)`. This is true even if the deprecated calls were used. Thus, the location of values returned can be thought of as being returned by the C bindings. It can also be determined by examining the preferred calls for datatype constructors for the deprecated calls that involve addresses.

*Rationale.* By having all address arguments returned in the `array_of_addresses` argument, the result from a C and Fortran decoding of a datatype gives the result in the same argument. It is assumed that an integer of type `INTEGER(KIND=MPI_ADDRESS_KIND)` will be at least as large as the `INTEGER` argument used in datatype construction with the old MPI-1 calls so no loss of information will occur. (*End of rationale.*)

The following defines what values are placed in each entry of the returned arrays depending on the datatype constructor used for `datatype`. It also specifies the size of the arrays needed which is the values returned by `MPI_TYPE_GET_ENVELOPE`. In Fortran, the following calls were made:

```

1     PARAMETER (LARGE = 1000)
2     INTEGER TYPE, NI, NA, ND, COMBINER, I(LARGE), D(LARGE), IERROR
3     INTEGER (KIND=MPI_ADDRESS_KIND) A(LARGE)
4     ! CONSTRUCT DATATYPE TYPE (NOT SHOWN)
5     CALL MPI_TYPE_GET_ENVELOPE(TYPE, NI, NA, ND, COMBINER, IERROR)
6     IF ((NI .GT. LARGE) .OR. (NA .GT. LARGE) .OR. (ND .GT. LARGE)) THEN
7         WRITE (*, *) "NI, NA, OR ND = ", NI, NA, ND, &
8             " RETURNED BY MPI_TYPE_GET_ENVELOPE IS LARGER THAN LARGE = ", LARGE
9         CALL MPI_ABORT(MPI_COMM_WORLD, 99, IERROR)
10    ENDIF
11    CALL MPI_TYPE_GET_CONTENTS(TYPE, NI, NA, ND, I, A, D, IERROR)

```

or in C the analogous calls of:

```

14 #define LARGE 1000
15 int ni, na, nd, combiner, i[LARGE];
16 MPI_Aint a[LARGE];
17 MPI_Datatype type, d[LARGE];
18 /* construct datatype type (not shown) */
19 MPI_Type_get_envelope(type, &ni, &na, &nd, &combiner);
20 if ((ni > LARGE) || (na > LARGE) || (nd > LARGE)) {
21     fprintf(stderr, "ni, na, or nd = %d %d %d returned by ", ni, na, nd);
22     fprintf(stderr, "MPI_Type_get_envelope is larger than LARGE = %d\n",
23         LARGE);
24     MPI_Abort(MPI_COMM_WORLD, 99);
25 }
26 MPI_Type_get_contents(type, ni, na, nd, i, a, d);

```

In the descriptions that follow, the lower case name of arguments is used.

If combiner is `MPI_COMBINER_NAMED` then it is erroneous to call `MPI_TYPE_GET_CONTENTS`.

If combiner is `MPI_COMBINER_DUP` then

Constructor argument	C	Fortran location
oldtype	d[0]	D(1)

and ni = 0, na = 0, nd = 1.

If combiner is `MPI_COMBINER_CONTIGUOUS` then

Constructor argument	C	Fortran location
count	i[0]	I(1)
oldtype	d[0]	D(1)

and ni = 1, na = 0, nd = 1.

If combiner is `MPI_COMBINER_VECTOR` then

Constructor argument	C	Fortran location
count	i[0]	I(1)
blocklength	i[1]	I(2)
stride	i[2]	I(3)
oldtype	d[0]	D(1)

and ni = 3, na = 0, nd = 1.

If combiner is `MPI_COMBINER_HVECTOR` then



Constructor argument	C	Fortran location
count	i[0]	I(1)
blocklength	i[1]	I(2)
stride	a[0]	A(1)
oldtype	d[0]	D(1)

and ni = 2, na = 1, nd = 1.

If combiner is MPI\_COMBINER\_INDEXED then

Constructor argument	C	Fortran location
count	i[0]	I(1)
array_of_blocklengths	i[1] to i[i[0]]	I(2) to I(I(1)+1)
array_of_displacements	i[i[0]+1] to i[2*i[0]]	I(I(1)+2) to I(2*I(1)+1)
oldtype	d[0]	D(1)

and ni = 2\*count+1, na = 0, nd = 1.

If combiner is MPI\_COMBINER\_HINDEXED then

Constructor argument	C	Fortran location
count	i[0]	I(1)
array_of_blocklengths	i[1] to i[i[0]]	I(2) to I(I(1)+1)
array_of_displacements	a[0] to a[i[0]-1]	A(1) to A(I(1))
oldtype	d[0]	D(1)

and ni = count+1, na = count, nd = 1.

If combiner is MPI\_COMBINER\_INDEXED\_BLOCK then

Constructor argument	C	Fortran location
count	i[0]	I(1)
blocklength	i[1]	I(2)
array_of_displacements	i[2] to i[i[0]+1]	I(3) to I(I(1)+2)
oldtype	d[0]	D(1)

and ni = count+2, na = 0, nd = 1.

If combiner is MPI\_COMBINER\_HINDEXED\_BLOCK then

Constructor argument	C	Fortran location
count	i[0]	I(1)
blocklength	i[1]	I(2)
array_of_displacements	a[0] to a[i[0]-1]	A(1) to A(I(1))
oldtype	d[0]	D(1)

and ni = 2, na = count, nd = 1.

If combiner is MPI\_COMBINER\_STRUCT then

Constructor argument	C	Fortran location
count	i[0]	I(1)
array_of_blocklengths	i[1] to i[i[0]]	I(2) to I(I(1)+1)
array_of_displacements	a[0] to a[i[0]-1]	A(1) to A(I(1))
array_of_types	d[0] to d[i[0]-1]	D(1) to D(I(1))

and ni = count+1, na = count, nd = count.

If combiner is MPI\_COMBINER\_SUBARRAY then

Constructor argument	C	Fortran location
ndims	i[0]	I(1)
array_of_sizes	i[1] to i[i[0]]	I(2) to I(I(1)+1)
array_of_subsizes	i[i[0]+1] to i[2*i[0]]	I(I(1)+2) to I(2*I(1)+1)
array_of_starts	i[2*i[0]+1] to i[3*i[0]]	I(2*I(1)+2) to I(3*I(1)+1)
order	i[3*i[0]+1]	I(3*I(1)+2)
oldtype	d[0]	D(1)

and ni = 3\*ndims+2, na = 0, nd = 1.

If combiner is MPI\_COMBINER\_DARRAY then

Constructor argument	C	Fortran location
size	i[0]	I(1)
rank	i[1]	I(2)
ndims	i[2]	I(3)
array_of_gsizes	i[3] to i[i[2]+2]	I(4) to I(I(3)+3)
array_of_distribs	i[i[2]+3] to i[2*i[2]+2]	I(I(3)+4) to I(2*I(3)+3)
array_of_dargs	i[2*i[2]+3] to i[3*i[2]+2]	I(2*I(3)+4) to I(3*I(3)+3)
array_of_psizes	i[3*i[2]+3] to i[4*i[2]+2]	I(3*I(3)+4) to I(4*I(3)+3)
order	i[4*i[2]+3]	I(4*I(3)+4)
oldtype	d[0]	D(1)

and ni = 4\*ndims+4, na = 0, nd = 1.

If combiner is MPI\_COMBINER\_F90\_REAL then

Constructor argument	C	Fortran location
p	i[0]	I(1)
r	i[1]	I(2)

and ni = 2, na = 0, nd = 0.

If combiner is MPI\_COMBINER\_F90\_COMPLEX then

Constructor argument	C	Fortran location
p	i[0]	I(1)
r	i[1]	I(2)

and ni = 2, na = 0, nd = 0.

If combiner is MPI\_COMBINER\_F90\_INTEGER then

Constructor argument	C	Fortran location
r	i[0]	I(1)

and ni = 1, na = 0, nd = 0.

If combiner is MPI\_COMBINER\_RESIZED then

Constructor argument	C	Fortran location
lb	a[0]	A(1)
extent	a[1]	A(2)
oldtype	d[0]	D(1)

and ni = 0, na = 2, nd = 1.

#### 4.1.14 Examples

The following examples illustrate the use of derived datatypes.

**Example 4.13** Send and receive a section of a 3D array.

```

REAL a(100,100,100), e(9,9,9)
INTEGER oneslice, twoslice, threeslice, myrank, ierr
INTEGER (KIND=MPI_ADDRESS_KIND) lb, sizeofreal
INTEGER status(MPI_STATUS_SIZE)

C extract the section a(1:17:2, 3:11, 2:10)
C and store it in e(:, :, :).

CALL MPI_COMM_RANK(MPI_COMM_WORLD, myrank, ierr)

CALL MPI_TYPE_GET_EXTENT(MPI_REAL, lb, sizeofreal, ierr)

C create datatype for a 1D section
CALL MPI_TYPE_VECTOR(9, 1, 2, MPI_REAL, oneslice, ierr)

C create datatype for a 2D section
CALL MPI_TYPE_CREATE_HVECTOR(9, 1, 100*sizeofreal, oneslice,
                             twoslice, ierr)

C create datatype for the entire section
CALL MPI_TYPE_CREATE_HVECTOR(9, 1, 100*100*sizeofreal, twoslice,
                             threeslice, ierr)

CALL MPI_TYPE_COMMIT(threeslice, ierr)
CALL MPI_SENDRECV(a(1,3,2), 1, threeslice, myrank, 0, e, 9*9*9,
                  MPI_REAL, myrank, 0, MPI_COMM_WORLD, status, ierr)

```

**Example 4.14** Copy the (strictly) lower triangular part of a matrix.

```

REAL a(100,100), b(100,100)
INTEGER disp(100), blocklen(100), ltype, myrank, ierr
INTEGER status(MPI_STATUS_SIZE)

C copy lower triangular part of array a
C onto lower triangular part of array b

CALL MPI_COMM_RANK(MPI_COMM_WORLD, myrank, ierr)

C compute start and size of each column
DO i=1, 100
  disp(i) = 100*(i-1) + i
  blocklen(i) = 100-i
END DO

C create datatype for lower triangular part
CALL MPI_TYPE_INDEXED(100, blocklen, disp, MPI_REAL, ltype, ierr)

CALL MPI_TYPE_COMMIT(ltype, ierr)

```

```

1      CALL MPI_SENDRECV(a, 1, ltype, myrank, 0, b, 1,
2                          ltype, myrank, 0, MPI_COMM_WORLD, status, ierr)
3

```

4 **Example 4.15** Transpose a matrix.

```

5
6      REAL a(100,100), b(100,100)
7      INTEGER row, xpose, myrank, ierr
8      INTEGER (KIND=MPI_ADDRESS_KIND) lb, sizeofreal
9      INTEGER status(MPI_STATUS_SIZE)
10
11 C    transpose matrix a onto b
12
13      CALL MPI_COMM_RANK(MPI_COMM_WORLD, myrank, ierr)
14
15      CALL MPI_TYPE_GET_EXTENT(MPI_REAL, lb, sizeofreal, ierr)
16
17 C    create datatype for one row
18      CALL MPI_TYPE_VECTOR(100, 1, 100, MPI_REAL, row, ierr)
19
20 C    create datatype for matrix in row-major order
21      CALL MPI_TYPE_CREATE_HVECTOR(100, 1, sizeofreal, row, xpose, ierr)
22
23      CALL MPI_TYPE_COMMIT(xpose, ierr)
24
25 C    send matrix in row-major order and receive in column major order
26      CALL MPI_SENDRECV(a, 1, xpose, myrank, 0, b, 100*100,
27                          MPI_REAL, myrank, 0, MPI_COMM_WORLD, status, ierr)
28
29

```

30 **Example 4.16** Another approach to the transpose problem:

```

31
32      REAL a(100,100), b(100,100)
33      INTEGER row, row1
34      INTEGER (KIND=MPI_ADDRESS_KIND) disp(2), lb, sizeofreal
35      INTEGER myrank, ierr
36      INTEGER status(MPI_STATUS_SIZE)
37
38      CALL MPI_COMM_RANK(MPI_COMM_WORLD, myrank, ierr)
39
40 C    transpose matrix a onto b
41
42      CALL MPI_TYPE_GET_EXTENT(MPI_REAL, lb, sizeofreal, ierr)
43
44 C    create datatype for one row
45      CALL MPI_TYPE_VECTOR(100, 1, 100, MPI_REAL, row, ierr)
46
47 C    create datatype for one row, with the extent of one real number
48      lb = 0

```

```

CALL MPI_TYPE_CREATE_RESIZED(row, lb, sizeofreal, row1, ierr)      1
                                                                    2
CALL MPI_TYPE_COMMIT(row1, ierr)                                  3
                                                                    4
C    send 100 rows and receive in column major order              5
CALL MPI_SENDRECV(a, 100, row1, myrank, 0, b, 100*100,           6
                 MPI_REAL, myrank, 0, MPI_COMM_WORLD, status, ierr) 7
                                                                    8

```

**Example 4.17** We manipulate an array of structures.

```

struct Partstruct                                               10
{                                                                 11
  int    type; /* particle type */                               12
  double d[6]; /* particle coordinates */                       13
  char   b[7]; /* some additional information */                14
};                                                                 15
                                                                    16
struct Partstruct    particle[1000];                             17
                                                                    18
int                  i, dest, tag;                                19
MPI_Comm             comm;                                       20
                                                                    21
                                                                    22
                                                                    23
/* build datatype describing structure */                         24
                                                                    25
MPI_Datatype Particlestruct, Particletype;                       26
MPI_Datatype type[3] = {MPI_INT, MPI_DOUBLE, MPI_CHAR};         27
int          blocklen[3] = {1, 6, 7};                           28
MPI_Aint     disp[3];                                           29
MPI_Aint     base, lb, sizeofentry;                              30
                                                                    31
                                                                    32
/* compute displacements of structure components */             33
                                                                    34
MPI_Get_address(particle, disp);                                  35
MPI_Get_address(particle[0].d, disp+1);                         36
MPI_Get_address(particle[0].b, disp+2);                         37
base = disp[0];                                                 38
for (i=0; i < 3; i++) disp[i] = MPI_Aint_diff(disp[i], base); 39
                                                                    40
MPI_Type_create_struct(3, blocklen, disp, type, &Particlestruct); 41
                                                                    42
/* If compiler does padding in mysterious ways,                43
the following may be safer */                                    44
                                                                    45
/* compute extent of the structure */                            46
MPI_Get_address(particle+1, &sizeofentry);                      47
                                                                    48

```

```

1  sizeofentry = MPI_Aint_diff(sizeofentry, base);
2
3  /* build datatype describing structure */
4
5  MPI_Type_create_resized(Particlestruct, 0, sizeofentry, &Particletype);
6
7
8      /* 4.1:
9      send the entire array */
10
11 MPI_Type_commit(&Particletype);
12 MPI_Send(particle, 1000, Particletype, dest, tag, comm);
13
14
15      /* 4.2:
16      send only the entries of type zero particles,
17      preceded by the number of such entries */
18
19 MPI_Datatype Zparticles; /* datatype describing all particles
20                          with type zero (needs to be recomputed
21                          if types change) */
22 MPI_Datatype Ztype;
23
24 int          zdisp[1000];
25 int          zblock[1000], j, k;
26 int          zzblock[2] = {1,1};
27 MPI_Aint     zzdisp[2];
28 MPI_Datatype zztype[2];
29
30 /* compute displacements of type zero particles */
31 j = 0;
32 for (i=0; i < 1000; i++)
33     if (particle[i].type == 0)
34         {
35             zdisp[j] = i;
36             zblock[j] = 1;
37             j++;
38         }
39
40 /* create datatype for type zero particles */
41 MPI_Type_indexed(j, zblock, zdisp, Particletype, &Zparticles);
42
43 /* prepend particle count */
44 MPI_Get_address(&j, zzdisp);
45 MPI_Get_address(particle, zzdisp+1);
46 zztype[0] = MPI_INT;
47 zztype[1] = Zparticles;
48 MPI_Type_create_struct(2, zzblock, zzdisp, zztype, &Ztype);

```

```

1
MPI_Type_commit(&Ztype);
2
MPI_Send(MPI_BOTTOM, 1, Ztype, dest, tag, comm);
3
4
5
/* A probably more efficient way of defining Zparticles */
6
7
/* consecutive particles with index zero are handled as one block */
8
j=0;
9
for (i=0; i < 1000; i++)
10
    if (particle[i].type == 0)
11
        {
12
            for (k=i+1; (k < 1000)&&(particle[k].type == 0) ; k++);
13
            zdisp[j] = i;
14
            zblock[j] = k-i;
15
            j++;
16
            i = k;
17
        }
18
MPI_Type_indexed(j, zblock, zdisp, Particletype, &Zparticles);
19
20
21
/* 4.3:
22
    send the first two coordinates of all entries */
23
24
MPI_Datatype Allpairs; /* datatype for all pairs of coordinates */
25
26
MPI_Type_get_extent(Particletype, &lb, &sizeofentry);
27
28
/* sizeofentry can also be computed by subtracting the address
29
    of particle[0] from the address of particle[1] */
30
31
MPI_Type_create_hvector(1000, 2, sizeofentry, MPI_DOUBLE, &Allpairs);
32
MPI_Type_commit(&Allpairs);
33
MPI_Send(particle[0].d, 1, Allpairs, dest, tag, comm);
34
35
/* an alternative solution to 4.3 */
36
37
MPI_Datatype Twodouble;
38
39
MPI_Type_contiguous(2, MPI_DOUBLE, &Twodouble);
40
41
MPI_Datatype Onepair; /* datatype for one pair of coordinates, with
42
    the extent of one particle entry */
43
44
MPI_Type_create_resized(Twodouble, 0, sizeofentry, &Onepair );
45
MPI_Type_commit(&Onepair);
46
MPI_Send(particle[0].d, 1000, Onepair, dest, tag, comm);
47
48

```

1 **Example 4.18** The same manipulations as in the previous example, but use absolute  
 2 addresses in datatypes.

```

3
4 struct Partstruct
5 {
6     int    type;
7     double d[6];
8     char   b[7];
9 };
10
11 struct Partstruct particle[1000];
12
13     /* build datatype describing first array entry */
14
15 MPI_Datatype Particletype;
16 MPI_Datatype type[3] = {MPI_INT, MPI_DOUBLE, MPI_CHAR};
17 int          block[3] = {1, 6, 7};
18 MPI_Aint     disp[3];
19
20 MPI_Get_address(particle, disp);
21 MPI_Get_address(particle[0].d, disp+1);
22 MPI_Get_address(particle[0].b, disp+2);
23 MPI_Type_create_struct(3, block, disp, type, &Particletype);
24
25 /* Particletype describes first array entry -- using absolute
26    addresses */
27
28     /* 5.1:
29    send the entire array */
30
31 MPI_Type_commit(&Particletype);
32 MPI_Send(MPI_BOTTOM, 1000, Particletype, dest, tag, comm);
33
34
35     /* 5.2:
36    send the entries of type zero,
37    preceded by the number of such entries */
38
39 MPI_Datatype Zparticles, Ztype;
40
41 int          zdisp[1000];
42 int          zblock[1000], i, j, k;
43 int          zzblock[2] = {1,1};
44 MPI_Datatype zztype[2];
45 MPI_Aint     zzdisp[2];
46
47 j=0;
48 for (i=0; i < 1000; i++)

```



```

    if (particle[i].type == 0)
    {
        for (k=i+1; (k < 1000)&&(particle[k].type == 0) ; k++);
        zdisp[j] = i;
        zblock[j] = k-i;
        j++;
        i = k;
    }
MPI_Type_indexed(j, zblock, zdisp, Particletype, &Zparticles);
/* Zparticles describe particles with type zero, using
   their absolute addresses*/

/* prepend particle count */
MPI_Get_address(&j, zzdisp);
zzdisp[1] = (MPI_Aint)0;
zztype[0] = MPI_INT;
zztype[1] = Zparticles;
MPI_Type_create_struct(2, zblock, zzdisp, zztype, &Ztype);

MPI_Type_commit(&Ztype);
MPI_Send(MPI_BOTTOM, 1, Ztype, dest, tag, comm);

```

**Example 4.19** Handling of unions.

```

union {
    int    ival;
    float  fval;
} u[1000];

int    utype;

/* All entries of u have identical type; variable
   utype keeps track of their current type */

MPI_Datatype  mpi_utype[2];
MPI_Aint      i, extent;

/* compute an MPI datatype for each possible union type;
   assume values are left-aligned in union storage. */

MPI_Get_address(u, &i);
MPI_Get_address(u+1, &extent);
extent = MPI_Aint_diff(extent, i);

MPI_Type_create_resized(MPI_INT, 0, extent, &mpi_utype[0]);

```

```

1 MPI_Type_create_resized(MPI_FLOAT, 0, extent, &mpi_utype[1]);
2
3 for(i=0; i<2; i++) MPI_Type_commit(&mpi_utype[i]);
4
5 /* actual communication */
6
7 MPI_Send(u, 1000, mpi_utype[utype], dest, tag, comm);
8
9
10 Example 4.20 This example shows how a datatype can be decoded. The routine
11 printdatatype prints out the elements of the datatype. Note the use of MPI_Type_free for
12 datatypes that are not predefined.
13
14 /*
15    Example of decoding a datatype.
16
17    Returns 0 if the datatype is predefined, 1 otherwise
18 */
19 #include <stdio.h>
20 #include <stdlib.h>
21 #include "mpi.h"
22 int printdatatype(MPI_Datatype datatype)
23 {
24     int *array_of_ints;
25     MPI_Aint *array_of_adds;
26     MPI_Datatype *array_of_dtypes;
27     int num_ints, num_adds, num_dtypes, combiner;
28     int i;
29
30     MPI_Type_get_envelope(datatype,
31                           &num_ints, &num_adds, &num_dtypes, &combiner);
32     switch (combiner) {
33     case MPI_COMBINER_NAMED:
34         printf("Datatype is named:");
35         /* To print the specific type, we can match against the
36            predefined forms. We can NOT use a switch statement here
37            We could also use MPI_TYPE_GET_NAME if we preferred to use
38            names that the user may have changed.
39         */
40         if (datatype == MPI_INT) printf( "MPI_INT\n" );
41         else if (datatype == MPI_DOUBLE) printf( "MPI_DOUBLE\n" );
42         ... else test for other types ...
43         return 0;
44         break;
45     case MPI_COMBINER_STRUCT:
46     case MPI_COMBINER_STRUCT_INTEGER:
47         printf("Datatype is struct containing");
48         array_of_ints = (int *)malloc(num_ints * sizeof(int));
49         array_of_adds =

```

```

        (MPI_Aint *) malloc(num_adds * sizeof(MPI_Aint));           1
array_of_dtypes = (MPI_Datatype *)                               2
    malloc(num_dtypes * sizeof(MPI_Datatype));                   3
MPI_Type_get_contents(datatype, num_ints, num_adds, num_dtypes,  4
    array_of_ints, array_of_adds, array_of_dtypes);              5
printf(" %d datatypes:\n", array_of_ints[0]);                   6
for (i=0; i<array_of_ints[0]; i++) {                             7
    printf("blocklength %d, displacement %ld, type:\n",          8
        array_of_ints[i+1], (long)array_of_adds[i]);            9
    if (printdatatype(array_of_dtypes[i])) {                     10
        /* Note that we free the type ONLY if it                11
           is not predefined */                                  12
        MPI_Type_free(&array_of_dtypes[i]);                      13
    }                                                             14
}                                                                 15
free(array_of_ints);                                           16
free(array_of_adds);                                           17
free(array_of_dtypes);                                         18
break;                                                         19
... other combiner values ...                                  20
default:                                                       21
    printf("Unrecognized combiner type\n");                      22
}                                                               23
return 1;                                                       24
}                                                               25

```

## 4.2 Pack and Unpack

Some existing communication libraries provide pack/unpack functions for sending noncontiguous data. In these, the user explicitly packs data into a contiguous buffer before sending it, and unpacks it from a contiguous buffer after receiving it. Derived datatypes, which are described in Section 4.1, allow one, in most cases, to avoid explicit packing and unpacking. The user specifies the layout of the data to be sent or received, and the communication library directly accesses a noncontiguous buffer. The pack/unpack routines are provided for compatibility with previous libraries. Also, they provide some functionality that is not otherwise available in MPI. For instance, a message can be received in several parts, where the receive operation done on a later part may depend on the content of a former part. Another use is that outgoing messages may be explicitly buffered in user supplied space, thus overriding the system buffering policy. Finally, the availability of pack and unpack operations facilitates the development of additional communication libraries layered on top of MPI.

```

1 MPI_PACK(inbuf, incount, datatype, outbuf, outsize, position, comm)
2   IN      inbuf          input buffer start (choice)
3
4   IN      incount       number of input data items (non-negative integer)
5
6   IN      datatype     datatype of each input data item (handle)
7
8   OUT     outbuf        output buffer start (choice)
9
10  IN      outsize       output buffer size, in bytes (non-negative integer)
11
12  INOUT   position     current position in buffer, in bytes (integer)
13
14  IN      comm          communicator for packed message (handle)

```

```

15 int MPI_Pack(const void* inbuf, int incount, MPI_Datatype datatype,
16             void *outbuf, int outsize, int *position, MPI_Comm comm)
17
18 MPI_Pack(inbuf, incount, datatype, outbuf, outsize, position, comm, ierror)
19   TYPE(*), DIMENSION(..), INTENT(IN) :: inbuf
20   TYPE(*), DIMENSION(..) :: outbuf
21   INTEGER, INTENT(IN) :: incount, outsize
22   TYPE(MPI_Datatype), INTENT(IN) :: datatype
23   INTEGER, INTENT(INOUT) :: position
24   TYPE(MPI_Comm), INTENT(IN) :: comm
25   INTEGER, OPTIONAL, INTENT(OUT) :: ierror
26
27 MPI_PACK(INBUF, INCOUNT, DATATYPE, OUTBUF, OUTSIZE, POSITION, COMM, IERROR)
28   <type> INBUF(*), OUTBUF(*)
29   INTEGER INCOUNT, DATATYPE, OUTSIZE, POSITION, COMM, IERROR

```

Packs the message in the send buffer specified by `inbuf`, `incount`, `datatype` into the buffer space specified by `outbuf` and `outsize`. The input buffer can be any communication buffer allowed in `MPI_SEND`. The output buffer is a contiguous storage area containing `outsize` bytes, starting at the address `outbuf` (length is counted in *bytes*, not elements, as if it were a communication buffer for a message of type `MPI_PACKED`).

The input value of `position` is the first location in the output buffer to be used for packing. `position` is incremented by the size of the packed message, and the output value of `position` is the first location in the output buffer following the locations occupied by the packed message. The `comm` argument is the communicator that will be subsequently used for sending the packed message.

```

37
38
39
40
41
42
43
44
45
46
47
48

```

```

MPI_UNPACK(inbuf, insize, position, outbuf, outcount, datatype, comm) 1
IN      inbuf      input buffer start (choice) 2
IN      insize     size of input buffer, in bytes (non-negative integer) 3
INOUT   position   current position in bytes (integer) 4
OUT     outbuf     output buffer start (choice) 5
IN      outcount   number of items to be unpacked (integer) 6
IN      datatype   datatype of each output data item (handle) 7
IN      comm       communicator for packed message (handle) 8
                                                9
int MPI_Unpack(const void* inbuf, int insize, int *position, void *outbuf,
               int outcount, MPI_Datatype datatype, MPI_Comm comm) 10
MPI_Unpack(inbuf, insize, position, outbuf, outcount, datatype, comm, 11
           ierror) 12
    TYPE(*), DIMENSION(..), INTENT(IN) :: inbuf 13
    TYPE(*), DIMENSION(..) :: outbuf 14
    INTEGER, INTENT(IN) :: insize, outcount 15
    INTEGER, INTENT(INOUT) :: position 16
    TYPE(MPI_Datatype), INTENT(IN) :: datatype 17
    TYPE(MPI_Comm), INTENT(IN) :: comm 18
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror 19
MPI_UNPACK(INBUF, INSIZE, POSITION, OUTBUF, OUTCOUNT, DATATYPE, COMM, 20
           IERROR) 21
    <type> INBUF(*), OUTBUF(*) 22
    INTEGER INSIZE, POSITION, OUTCOUNT, DATATYPE, COMM, IERROR 23

```

Unpacks a message into the receive buffer specified by `outbuf`, `outcount`, `datatype` from the buffer space specified by `inbuf` and `insize`. The output buffer can be any communication buffer allowed in `MPI_RECV`. The input buffer is a contiguous storage area containing `insize` bytes, starting at address `inbuf`. The input value of `position` is the first location in the input buffer occupied by the packed message. `position` is incremented by the size of the packed message, so that the output value of `position` is the first location in the input buffer after the locations occupied by the message that was unpacked. `comm` is the communicator used to receive the packed message.

*Advice to users.* Note the difference between `MPI_RECV` and `MPI_UNPACK`: in `MPI_RECV`, the count argument specifies the maximum number of items that can be received. The actual number of items received is determined by the length of the incoming message. In `MPI_UNPACK`, the count argument specifies the actual number of items that are unpacked; the “size” of the corresponding message is the increment in `position`. The reason for this change is that the “incoming message size” is not predetermined since the user decides how much to unpack; nor is it easy to determine the “message size” from the number of items to be unpacked. In fact, in a heterogeneous system, this number may not be determined *a priori*. (*End of advice to users.*)

1 To understand the behavior of pack and unpack, it is convenient to think of the data  
2 part of a message as being the sequence obtained by concatenating the successive values sent  
3 in that message. The pack operation stores this sequence in the buffer space, as if sending  
4 the message to that buffer. The unpack operation retrieves this sequence from buffer space,  
5 as if receiving a message from that buffer. (It is helpful to think of internal Fortran files or  
6 `sscanf` in C, for a similar function.)

7 Several messages can be successively packed into one *packing unit*. This is effected  
8 by several successive *related* calls to `MPI_PACK`, where the first call provides `position = 0`,  
9 and each successive call inputs the value of `position` that was output by the previous call,  
10 and the same values for `outbuf`, `outcount` and `comm`. This packing unit now contains the  
11 equivalent information that would have been stored in a message by one send call with a  
12 send buffer that is the “concatenation” of the individual send buffers.

13 A packing unit can be sent using type `MPI_PACKED`. Any point to point or collective  
14 communication function can be used to move the sequence of bytes that forms the packing  
15 unit from one process to another. This packing unit can now be received using any receive  
16 operation, with any datatype: the type matching rules are relaxed for messages sent with  
17 type `MPI_PACKED`.

18 A message sent with any type (including `MPI_PACKED`) can be received using the type  
19 `MPI_PACKED`. Such a message can then be unpacked by calls to `MPI_UNPACK`.

20 A packing unit (or a message created by a regular, “typed” send) can be unpacked into  
21 several successive messages. This is effected by several successive related calls to  
22 `MPI_UNPACK`, where the first call provides `position = 0`, and each successive call inputs the  
23 value of `position` that was output by the previous call, and the same values for `inbuf`, `insize`  
24 and `comm`.

25 The concatenation of two packing units is not necessarily a packing unit; nor is a  
26 substring of a packing unit necessarily a packing unit. Thus, one cannot concatenate two  
27 packing units and then unpack the result as one packing unit; nor can one unpack a substring  
28 of a packing unit as a separate packing unit. Each packing unit, that was created by a related  
29 sequence of pack calls, or by a regular send, must be unpacked as a unit, by a sequence of  
30 related unpack calls.

31  
32 *Rationale.* The restriction on “atomic” packing and unpacking of packing units  
33 allows the implementation to add at the head of packing units additional information,  
34 such as a description of the sender architecture (to be used for type conversion, in a  
35 heterogeneous environment) (*End of rationale.*)

36  
37 The following call allows the user to find out how much space is needed to pack a  
38 message and, thus, manage space allocation for buffers.

MPI_PACK_SIZE(incount, datatype, comm, size)	1	
IN incount	count argument to packing call (non-negative integer)	2
IN datatype	datatype argument to packing call (handle)	3
IN comm	communicator argument to packing call (handle)	4
OUT size	upper bound on size of packed message, in bytes (non-	5
	negative integer)	6
		7
		8

```
int MPI_Pack_size(int incount, MPI_Datatype datatype, MPI_Comm comm,
                 int *size)

```

```
MPI_Pack_size(incount, datatype, comm, size, ierror)
    INTEGER, INTENT(IN) :: incount
    TYPE(MPI_Datatype), INTENT(IN) :: datatype
    TYPE(MPI_Comm), INTENT(IN) :: comm
    INTEGER, INTENT(OUT) :: size
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror

```

```
MPI_PACK_SIZE(INCOUNT, DATATYPE, COMM, SIZE, IERROR)
    INTEGER INCOUNT, DATATYPE, COMM, SIZE, IERROR

```

A call to MPI\_PACK\_SIZE(incount, datatype, comm, size) returns in size an upper bound on the increment in position that is effected by a call to MPI\_PACK(inbuf, incount, datatype, outbuf, outcount, position, comm). If the packed size of the datatype cannot be expressed by the size parameter, then MPI\_PACK\_SIZE sets the value of size to MPI\_UNDEFINED.

*Rationale.* The call returns an upper bound, rather than an exact bound, since the exact amount of space needed to pack the message may depend on the context (e.g., first message packed in a packing unit may take more space). (*End of rationale.*)

**Example 4.21** An example using MPI\_PACK.

```
int position, i, j, a[2];
char buff[1000];

MPI_Comm_rank(MPI_COMM_WORLD, &myrank);
if (myrank == 0)
{
    /* SENDER CODE */

    position = 0;
    MPI_Pack(&i, 1, MPI_INT, buff, 1000, &position, MPI_COMM_WORLD);
    MPI_Pack(&j, 1, MPI_INT, buff, 1000, &position, MPI_COMM_WORLD);
    MPI_Send(buff, position, MPI_PACKED, 1, 0, MPI_COMM_WORLD);
}
else /* RECEIVER CODE */
    MPI_Recv(a, 2, MPI_INT, 0, 0, MPI_COMM_WORLD, MPI_STATUS_IGNORE);

```

**Example 4.22** An elaborate example.

```

1  int   position, i;
2  float a[1000];
3  char  buff[1000];
4
5  MPI_Comm_rank(MPI_COMM_WORLD, &myrank);
6  if (myrank == 0)
7  {
8      /* SENDER CODE */
9
10     int len[2];
11     MPI_Aint disp[2];
12     MPI_Datatype type[2], newtype;
13
14     /* build datatype for i followed by a[0]...a[i-1] */
15
16     len[0] = 1;
17     len[1] = i;
18     MPI_Get_address(&i, disp);
19     MPI_Get_address(a, disp+1);
20     type[0] = MPI_INT;
21     type[1] = MPI_FLOAT;
22     MPI_Type_create_struct(2, len, disp, type, &newtype);
23     MPI_Type_commit(&newtype);
24
25     /* Pack i followed by a[0]...a[i-1]*/
26
27     position = 0;
28     MPI_Pack(MPI_BOTTOM, 1, newtype, buff, 1000, &position, MPI_COMM_WORLD);
29
30     /* Send */
31
32     MPI_Send(buff, position, MPI_PACKED, 1, 0,
33             MPI_COMM_WORLD);
34
35     /* *****
36      One can replace the last three lines with
37      MPI_Send(MPI_BOTTOM, 1, newtype, 1, 0, MPI_COMM_WORLD);
38      ***** */
39 }
40 else if (myrank == 1)
41 {
42     /* RECEIVER CODE */
43
44     MPI_Status status;
45
46     /* Receive */
47
48     MPI_Recv(buff, 1000, MPI_PACKED, 0, 0, MPI_COMM_WORLD, &status);

```



```

1
2  /* Unpack i */
3
4  position = 0;
5  MPI_Unpack(buff, 1000, &position, &i, 1, MPI_INT, MPI_COMM_WORLD);
6
7  /* Unpack a[0]...a[i-1] */
8  MPI_Unpack(buff, 1000, &position, a, i, MPI_FLOAT, MPI_COMM_WORLD);
9
10 }

```

**Example 4.23** Each process sends a count, followed by count characters to the root; the root concatenates all characters into one string.

```

11
12
13
14 int count, gsize, counts[64], totalcount, k1, k2, k,
15     displs[64], position, concat_pos;
16 char chr[100], *lbuf, *rbuf, *cbuf;
17
18 MPI_Comm_size(comm, &gsize);
19 MPI_Comm_rank(comm, &myrank);
20
21     /* allocate local pack buffer */
22 MPI_Pack_size(1, MPI_INT, comm, &k1);
23 MPI_Pack_size(count, MPI_CHAR, comm, &k2);
24 k = k1+k2;
25 lbuf = (char *)malloc(k);
26
27     /* pack count, followed by count characters */
28 position = 0;
29 MPI_Pack(&count, 1, MPI_INT, lbuf, k, &position, comm);
30 MPI_Pack(chr, count, MPI_CHAR, lbuf, k, &position, comm);
31
32 if (myrank != root) {
33     /* gather at root sizes of all packed messages */
34     MPI_Gather(&position, 1, MPI_INT, NULL, 0,
35             MPI_DATATYPE_NULL, root, comm);
36
37     /* gather at root packed messages */
38     MPI_Gatherv(lbuf, position, MPI_PACKED, NULL,
39             NULL, NULL, MPI_DATATYPE_NULL, root, comm);
40
41 } else { /* root code */
42     /* gather sizes of all packed messages */
43     MPI_Gather(&position, 1, MPI_INT, counts, 1,
44             MPI_INT, root, comm);
45
46     /* gather all packed messages */
47     displs[0] = 0;
48     for (i=1; i < gsize; i++)

```

```

1     displs[i] = displs[i-1] + counts[i-1];
2     totalcount = displs[gsize-1] + counts[gsize-1];
3     rbuf = (char *)malloc(totalcount);
4     cbuf = (char *)malloc(totalcount);
5     MPI_Gatherv(lbuf, position, MPI_PACKED, rbuf,
6                 counts, displs, MPI_PACKED, root, comm);
7
8     /* unpack all messages and concatenate strings */
9     concat_pos = 0;
10    for (i=0; i < gsize; i++) {
11        position = 0;
12        MPI_Unpack(rbuf+displs[i], totalcount-displs[i],
13                  &position, &count, 1, MPI_INT, comm);
14        MPI_Unpack(rbuf+displs[i], totalcount-displs[i],
15                  &position, cbuf+concat_pos, count, MPI_CHAR, comm);
16        concat_pos += count;
17    }
18    cbuf[concat_pos] = '\0';
19 }

```

### 4.3 Canonical MPI\_PACK and MPI\_UNPACK

These functions read/write data to/from the buffer in the “external32” data format specified in Section 13.7.2, and calculate the size needed for packing. Their first arguments specify the data format, for future extensibility, but currently the only valid value of the `datarep` argument is “external32.”

*Advice to users.* These functions could be used, for example, to send typed data in a portable format from one MPI implementation to another. (*End of advice to users.*)

The buffer will contain exactly the packed data, without headers. `MPI_BYTE` should be used to send and receive data that is packed using `MPI_PACK_EXTERNAL`.

*Rationale.* `MPI_PACK_EXTERNAL` specifies that there is no header on the message and further specifies the exact format of the data. Since `MPI_PACK` may (and is allowed to) use a header, the datatype `MPI_PACKED` cannot be used for data packed with `MPI_PACK_EXTERNAL`. (*End of rationale.*)

```

MPI_PACK_EXTERNAL(datarep, inbuf, incount, datatype, outbuf, outsize, position) 1
    IN      datarep      data representation (string) 2
    IN      inbuf        input buffer start (choice) 3
    IN      incount      number of input data items (integer) 4
    IN      datatype     datatype of each input data item (handle) 5
    OUT     outbuf       output buffer start (choice) 6
    IN      outsize      output buffer size, in bytes (integer) 7
    INOUT   position    current position in buffer, in bytes (integer) 8
                                                                9
                                                                10
                                                                11
int MPI_Pack_external(const char datarep[], const void *inbuf, int incount, 12
                    MPI_Datatype datatype, void *outbuf, MPI_Aint outsize, 13
                    MPI_Aint *position) 14
                                                                15
MPI_Pack_external(datarep, inbuf, incount, datatype, outbuf, outsize, 16
                position, ierror) 17
    CHARACTER(LEN=*) , INTENT(IN) :: datarep 18
    TYPE(*), DIMENSION(..), INTENT(IN) :: inbuf 19
    TYPE(*), DIMENSION(..) :: outbuf 20
    INTEGER, INTENT(IN) :: incount 21
    TYPE(MPI_Datatype), INTENT(IN) :: datatype 22
    INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: outsize 23
    INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(INOUT) :: position 24
    INTEGER, OPTIONAL, INTENT(OUT) :: ierror 25
                                                                26
MPI_PACK_EXTERNAL(DATAREP, INBUF, INCOUNT, DATATYPE, OUTBUF, OUTSIZE, 27
                POSITION, IERROR) 28
    INTEGER INCOUNT, DATATYPE, IERROR 29
    INTEGER(KIND=MPI_ADDRESS_KIND) OUTSIZE, POSITION 30
    CHARACTER*(*) DATAREP 31
    <type> INBUF(*), OUTBUF(*) 32
                                                                33
                                                                34
MPI_UNPACK_EXTERNAL(datarep, inbuf, insize, position, outbuf, outsize, position) 35
    IN      datarep      data representation (string) 36
    IN      inbuf        input buffer start (choice) 37
    IN      insize       input buffer size, in bytes (integer) 38
    INOUT   position    current position in buffer, in bytes (integer) 39
    OUT     outbuf       output buffer start (choice) 40
    IN      outcount     number of output data items (integer) 41
    IN      datatype     datatype of output data item (handle) 42
                                                                43
                                                                44
                                                                45
int MPI_Unpack_external(const char datarep[], const void *inbuf, 46
                    MPI_Aint insize, MPI_Aint *position, void *outbuf, 47
                    int outcount, MPI_Datatype datatype) 48

```

```

1 MPI_Unpack_external(datarep, inbuf, insize, position, outbuf, outcount,
2     datatype, ierror)
3     CHARACTER(LEN=*), INTENT(IN) :: datarep
4     TYPE(*), DIMENSION(..), INTENT(IN) :: inbuf
5     TYPE(*), DIMENSION(..) :: outbuf
6     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(IN) :: insize
7     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(INOUT) :: position
8     INTEGER, INTENT(IN) :: outcount
9     TYPE(MPI_Datatype), INTENT(IN) :: datatype
10    INTEGER, OPTIONAL, INTENT(OUT) :: ierror
11
12 MPI_UNPACK_EXTERNAL(DATAREP, INBUF, INSIZE, POSITION, OUTBUF, OUTCOUNT,
13     DATATYPE, IERROR)
14    INTEGER OUTCOUNT, DATATYPE, IERROR
15    INTEGER(KIND=MPI_ADDRESS_KIND) INSIZE, POSITION
16    CHARACTER*(*) DATAREP
17    <type> INBUF(*), OUTBUF(*)
18
19
20 MPI_PACK_EXTERNAL_SIZE(datarep, incount, datatype, size)
21     IN        datarep                data representation (string)
22     IN        incount                number of input data items (integer)
23     IN        datatype              datatype of each input data item (handle)
24     OUT       size                  output buffer size, in bytes (integer)
25
26
27 int MPI_Pack_external_size(const char datarep[], int incount,
28     MPI_Datatype datatype, MPI_Aint *size)
29
30 MPI_Pack_external_size(datarep, incount, datatype, size, ierror)
31     TYPE(MPI_Datatype), INTENT(IN) :: datatype
32     INTEGER, INTENT(IN) :: incount
33     CHARACTER(LEN=*), INTENT(IN) :: datarep
34     INTEGER(KIND=MPI_ADDRESS_KIND), INTENT(OUT) :: size
35     INTEGER, OPTIONAL, INTENT(OUT) :: ierror
36
37 MPI_PACK_EXTERNAL_SIZE(DATAREP, INCOUNT, DATATYPE, SIZE, IERROR)
38     INTEGER INCOUNT, DATATYPE, IERROR
39     INTEGER(KIND=MPI_ADDRESS_KIND) SIZE
40     CHARACTER*(*) DATAREP
41
42
43
44
45
46
47
48

```

# Bibliography

- [1] Charles H. Koelbel, David B. Loveman, Robert S. Schreiber, Guy L. Steele Jr., and Mary E. Zosel. *The High Performance Fortran Handbook*. MIT Press, 1993. [4.1.4](#)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

# Index

- Absolute addresses, 19
- addresses, 33
- committed, 27
- CONST:&, 20
- CONST:int, 20
- CONST:MPI\_Aint, 3, 3, 5, 8, 11, 19–21, 24–26, 36, 57, 58
- CONST:MPI\_BOTTOM, 19, 33, 34
- CONST:MPI\_BYTE, 56
- CONST:MPI\_CHAR, 11
- CONST:MPI\_COMBINER\_CONTIGUOUS, 35, 38
- CONST:MPI\_COMBINER\_DARRAY, 35, 40
- CONST:MPI\_COMBINER\_DUP, 35, 38
- CONST:MPI\_COMBINER\_F90\_COMPLEX, 35, 40
- CONST:MPI\_COMBINER\_F90\_INTEGER, 35, 40
- CONST:MPI\_COMBINER\_F90\_REAL, 35, 40
- CONST:MPI\_COMBINER\_HINDEXED, 35, 39
- CONST:MPI\_COMBINER\_HINDEXED\_BLOCK, 35, 39
- CONST:MPI\_COMBINER\_HVECTOR, 35, 38
- CONST:MPI\_COMBINER\_INDEXED, 35, 39
- CONST:MPI\_COMBINER\_INDEXED\_BLOCK, 35, 39
- CONST:MPI\_COMBINER\_NAMED, 35, 38
- CONST:MPI\_COMBINER\_RESIZED, 35, 40
- CONST:MPI\_COMBINER\_STRUCT, 35, 39
- CONST:MPI\_COMBINER\_SUBARRAY, 35, 39
- CONST:MPI\_COMBINER\_VECTOR, 35, 38
- CONST:MPI\_Datatype, 3
- CONST:MPI\_DATATYPE\_NULL, 29
- CONST:MPI\_DISTRIBUTE\_BLOCK, 16
- CONST:MPI\_DISTRIBUTE\_CYCLIC, 16
- CONST:MPI\_DISTRIBUTE\_DFLT\_DARG, 16
- CONST:MPI\_DISTRIBUTE\_NONE, 16
- CONST:MPI\_FLOAT, 11
- CONST:MPI\_INT, 2
- CONST:MPI\_ORDER\_C, 13, 16, 17
- CONST:MPI\_ORDER\_FORTRAN, 13, 16
- CONST:MPI\_PACKED, 50, 52, 56
- CONST:MPI\_Status, 31
- CONST:MPI\_UNDEFINED, 22, 25, 27, 32, 53
- EXAMPLES:Datatype
  - 3D array, 40
  - absolute addresses, 46
  - array of structures, 43
  - elaborate example, 53, 55
  - matching type, 30
  - matrix transpose, 42
  - union, 47
- EXAMPLES:MPI\_Aint, 43
- EXAMPLES:MPI\_Gather, 55
- EXAMPLES:MPI\_Gatherv, 55
- EXAMPLES:MPI\_GET\_ADDRESS, 20
- EXAMPLES:MPI\_Get\_address, 43, 46, 47, 53
- EXAMPLES:MPI\_GET\_COUNT, 32
- EXAMPLES:MPI\_GET\_ELEMENTS, 32
- EXAMPLES:MPI\_Pack, 53, 55
- EXAMPLES:MPI\_Pack\_size, 55
- EXAMPLES:MPI\_RECV, 30
- EXAMPLES:MPI\_SEND, 30
- EXAMPLES:MPI\_Send, 43, 46, 47, 53
- EXAMPLES:MPI\_SENDRECV, 40–42
- EXAMPLES:MPI\_TYPE\_COMMIT, 28, 40–42
- EXAMPLES:MPI\_Type\_commit, 43, 46, 47, 53
- EXAMPLES:MPI\_TYPE\_CONTIGUOUS, 4, 23, 30, 32

- EXAMPLES:MPI\_TYPE\_CREATE\_DARRAY, [56](#) 1
- [18](#) MPI\_PACK\_EXTERNAL(datatype, inbuf, in- 2
- EXAMPLES:MPI\_TYPE\_CREATE\_HVECTOR, count, datatype, outbuf, outsize, po- 3
- [40](#), [42](#) sition), [57](#) 4
- EXAMPLES:MPI\_Type\_create\_hvector, [43](#), MPI\_PACK\_EXTERNAL\_SIZE(datatype, in- 5
- [46](#) count, datatype, size), [58](#) 6
- EXAMPLES:MPI\_TYPE\_CREATE\_STRUCT, MPI\_PACK\_SIZE, [53](#) 7
- [11](#), [23](#), [42](#) MPI\_PACK\_SIZE(incount, datatype, comm, 8
- EXAMPLES:MPI\_Type\_create\_struct, [43](#), [46](#), size), [53](#), [53](#) 9
- [47](#), [53](#) MPI\_RECV, [51](#) 10
- EXAMPLES:MPI\_Type\_get\_contents, [48](#) MPI\_RECV(buf, 1, datatype,...), [2](#) 11
- EXAMPLES:MPI\_Type\_get\_envelope, [48](#) MPI\_RECV(buf, count, datatype, dest, tag, 12
- EXAMPLES:MPI\_TYPE\_GET\_EXTENT, [40](#), comm, status), [31](#) 13
- [42](#) MPI\_RECV(buf, count, datatype, source, tag, 14
- EXAMPLES:MPI\_Type\_get\_extent, [43](#) comm, status), [30](#) 15
- EXAMPLES:MPI\_TYPE\_INDEXED, [7](#), [41](#) MPI\_SEND, [50](#) 16
- EXAMPLES:MPI\_Type\_indexed, [43](#), [46](#) MPI\_SEND(buf, 1, datatype,...), [2](#) 17
- EXAMPLES:MPI\_TYPE\_VECTOR, [4](#), [5](#), [40](#), MPI\_SEND(buf, count, datatype, ...), [29](#) 18
- [42](#) MPI\_SEND(buf, count, datatype, dest, tag, 19
- EXAMPLES:MPI\_Unpack, [53](#), [55](#) comm), [29](#), [30](#) 20
- EXAMPLES:Typemap, [3–5](#), [7](#), [11](#), [18](#) MPI\_TYPE\_COMMIT, [28](#) 21
- extent, [2](#), [23](#) MPI\_TYPE\_COMMIT(datatype), [28](#) 22
- general datatype, [1](#) MPI\_TYPE\_CONTIGUOUS, [3](#), [35](#) 23
- lb, [26](#) MPI\_TYPE\_CONTIGUOUS(2, type1, type2), 24
- lb\_marker, [13](#), [14](#), [17](#), [23](#), [27](#) [23](#) 25
- lower bound, [23](#) MPI\_TYPE\_CONTIGUOUS(count, oldtype, 26
- MPI\_AINT\_ADD, [19–21](#) newtype), [3](#), [5](#) 27
- MPI\_AINT\_ADD(base, disp), [21](#) MPI\_TYPE\_CREATE\_DARRAY, [15](#), [35](#) 28
- MPI\_AINT\_DIFF, [19–21](#) MPI\_TYPE\_CREATE\_DARRAY(size, rank, 29
- MPI\_AINT\_DIFF(addr1, addr2), [21](#) ndims, array\_of\_gsizes, array\_of\_distrib, array\_of\_dar- 30
- MPI\_F\_SYNC\_REG, [20](#) array\_of\_psizes, order, oldtype, new- 31
- MPI\_FILE\_SET\_VIEW, [14](#) type), [15](#) 32
- MPI\_GET\_ADDRESS, [3](#), [19–21](#), [33](#) MPI\_TYPE\_CREATE\_F90\_COMPLEX, [35](#), 33
- MPI\_GET\_ADDRESS(location, address), [20](#) [37](#) 34
- MPI\_GET\_COUNT, [32](#) MPI\_TYPE\_CREATE\_F90\_INTEGER, [35](#), 35
- MPI\_GET\_ELEMENTS, [31](#), [32](#) [37](#) 36
- MPI\_GET\_ELEMENTS(status, datatype, count), [35](#) 37
- [31](#) MPI\_TYPE\_CREATE\_F90\_REAL, [35](#), [37](#) 37
- MPI\_GET\_ELEMENTS\_X, [31](#), [32](#) MPI\_TYPE\_CREATE\_HINDEXED, [3](#), [8](#), [10](#), 38
- MPI\_GET\_ELEMENTS\_X(status, datatype, newtype), [8](#) 39
- count), [31](#) MPI\_TYPE\_CREATE\_HINDEXED(count, ar- 40
- MPI\_PACK, [53](#), [56](#) ray\_of\_blocklengths, array\_of\_displacements, oldtype, 41
- MPI\_PACK(inbuf, incount, datatype, out- MPI\_TYPE\_CREATE\_HINDEXED(count, B, 42
- buf, outcount, position, comm), [53](#) D, oldtype, newtype), [12](#) 43
- MPI\_PACK(inbuf, incount, datatype, out- MPI\_TYPE\_CREATE\_HINDEXED\_BLOCK, 44
- buf, outsize, position, comm), [50](#) [3](#), [10](#), [35](#) 45
- MPI\_TYPE\_CREATE\_HINDEXED\_BLOCK(count, MPI\_TYPE\_CREATE\_HINDEXED\_BLOCK(count, 46
- blocklength, array\_of\_displacements, oldtype, 47

