Overview of AUTOPACK 1.4

Ray Loy

Approaches to Packing

- Manually by user
- MPI_Pack/MPI_Unpack
  - Unnecessary memory-memory copying
- MPI Derived datatypes
  - Relies on scatter/gather capability
- a library...

What is Autopack?

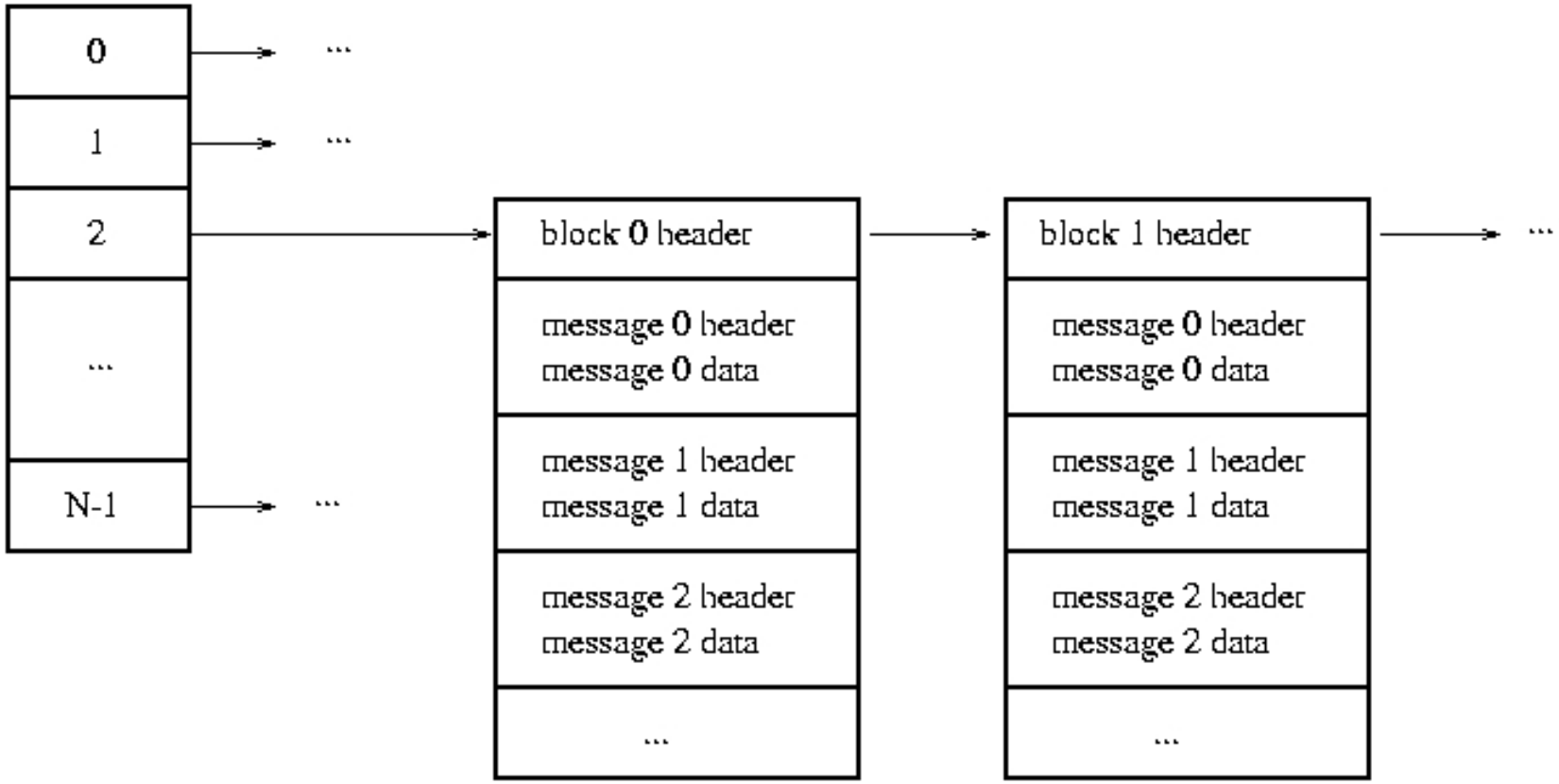
- A message passing library that transparently packs messages into fewer, larger ones for more efficient transport by MPI
  - Makes message amalgamation more convenient
  - Simple adjustment of underlying message size for best efficiency
  - Ideal for adaptive mesh calculations

It does the work, so you don't have to

Native-mode Autopack

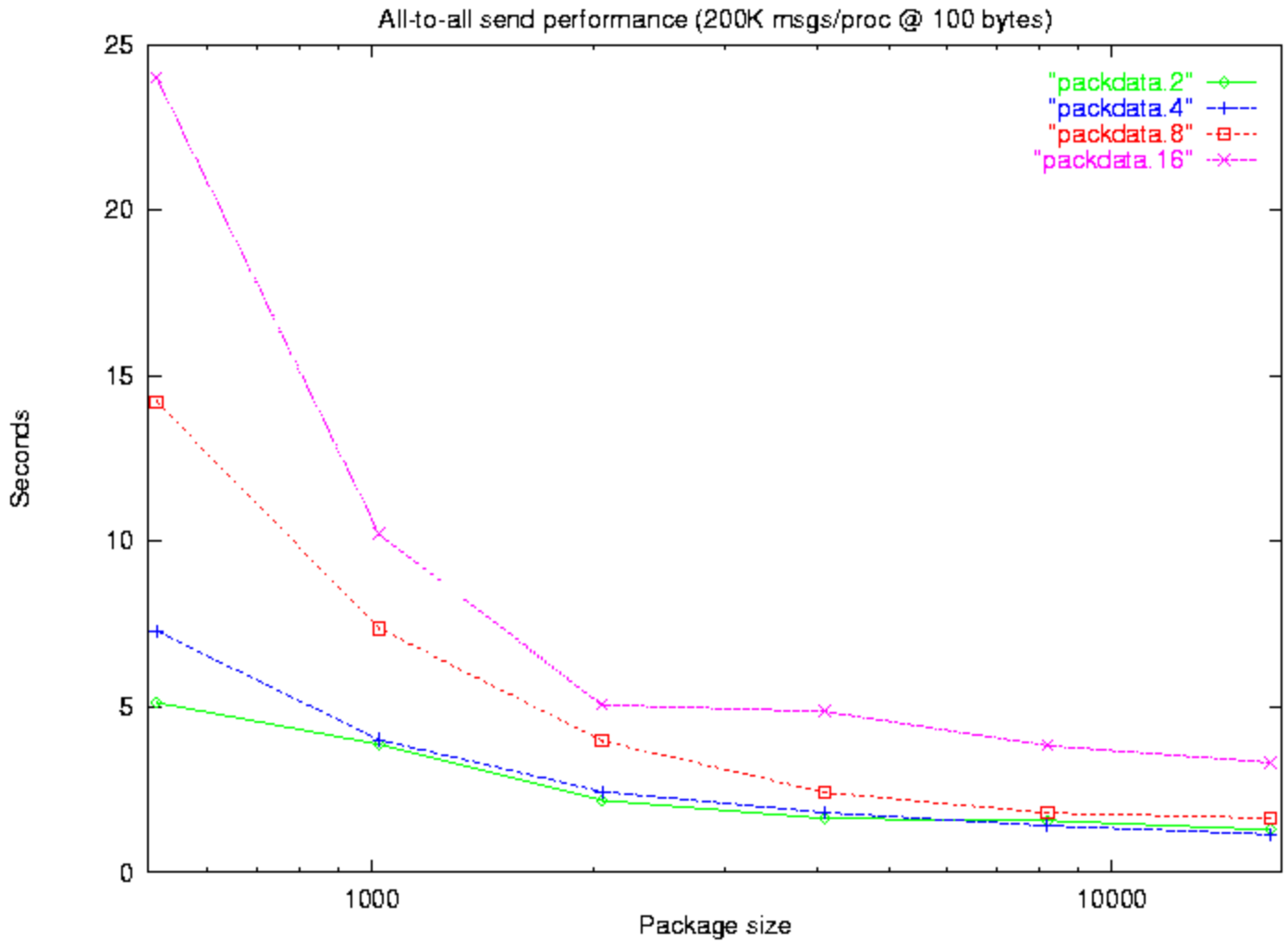
```
for (...) {
    msg= (usertype *) AP_alloc(destination,size,tag);
    msg->field1=...;
    AP_send(msg);
}
AP_flush();

while (AP_recv(..., &msg, ...) {
    result=msg->field1;
    AP_free(msg);
}
```

Internal Data Structure



Additional Functions

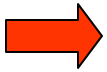
- Flush unsent messages
- Poll/block on message *send* completion
- transparent memory buffer management
- Asynchronous reduction operation
  - Determine expected number of receives


All-to-all send performance (200K msgs/proc @ 100 bytes)
Seconds
Package size
0
5
10
15
20
25
1000
10000
"packdata.2"
"packdata.4"
"packdata.8"
"packdata.16"

Extension: MPI-like Interface

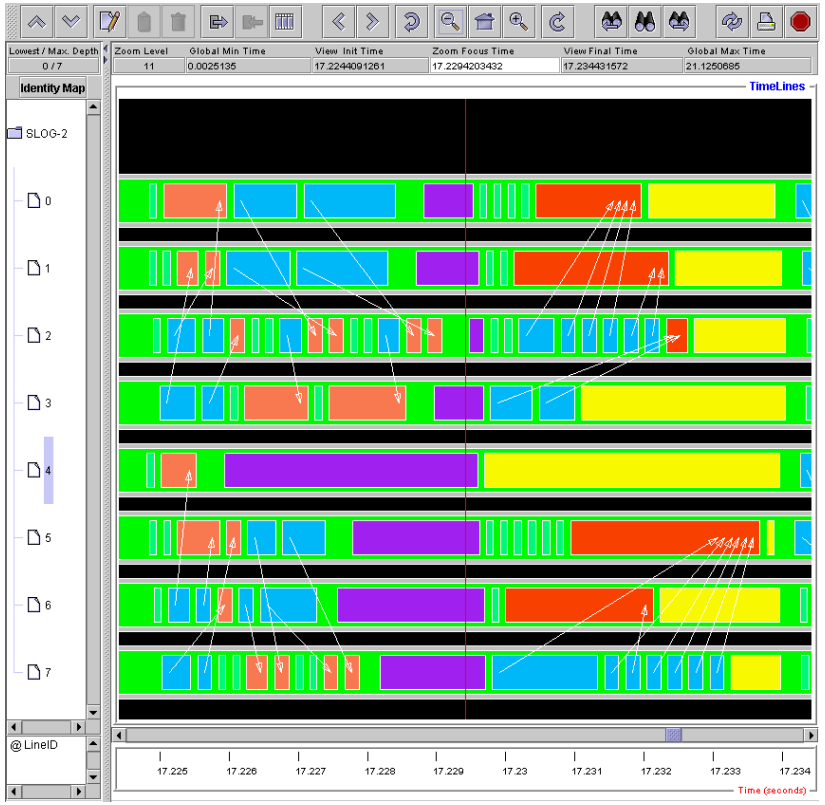
Drop-in Replacement

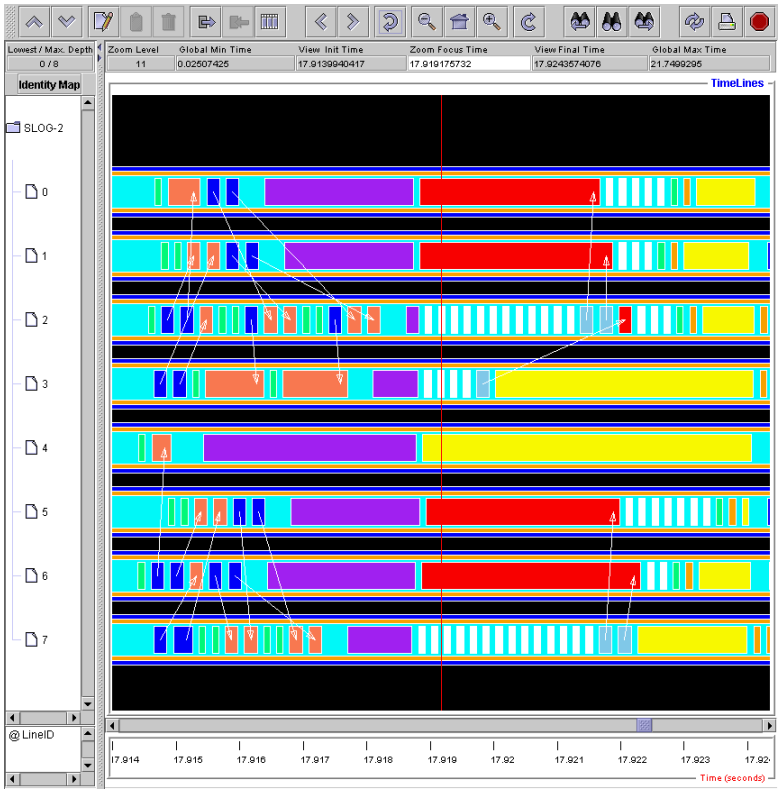
```
do ...
  call MPI_IRECV(...)
end do
do ...
  call MPI_SEND(...)
end do
call MPI_WAITALL(...)
```



```
do ...
  call AP_MPI_IRECV(...)
end do
do ...
  call AP_MPI_SEND(...)
end do
call AP_FLUSH
call AP_MPI_WAITALL(...)
call AP_CHECK_SENDS(...)
```

Using Autopack to Combine Messages

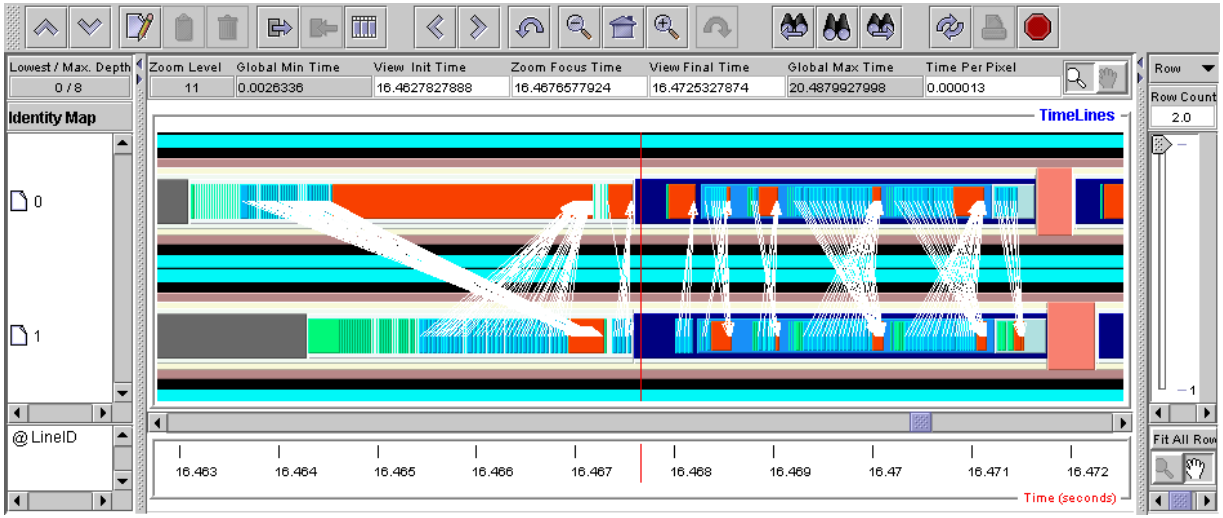




- Before
- After

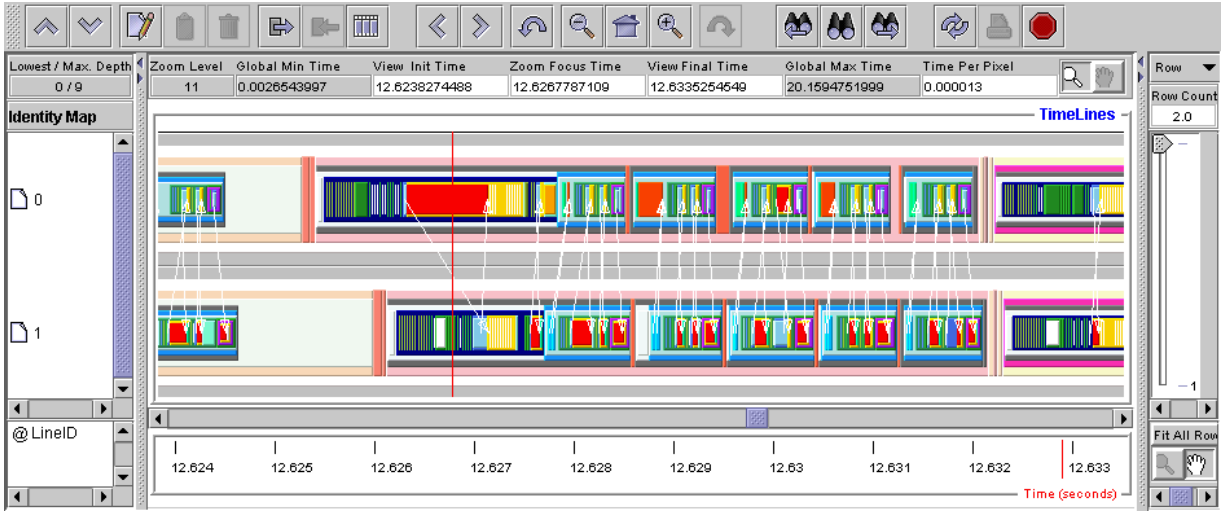
Autopack in FLASH

Without Autopack:
Solve time: 15 sec
Msg time: 9 sec



Communication during FLASH implicit solve

With Autopack:
Solve time: 8 sec
Msg time: 2 sec



(64-process run on Chiba)

New Functions in v1.4

- AP_MPI_Wait() / AP_MPI_Test()
- AP_MPI_Waitany() / AP_MPI_Testany()
- AP_MPI_Waitall () / AP_MPI_Testall()
- AP_MPI_Waitsome() / AP_MPI_Testsome()

Autopack distribution

- Currently used in
  - Flash 2.2/Paramesh 2.0
  - RPI codes AOMD and Trellis
  - MCS DataVise Tool
- Available now on request
  - *web page under construction*

Current/Future Work

- Parameter settings
  - e.g. package size, number of isends
- Fortran Interface
  - Handle/status issues; need *full* MPI 1.2
  - *should be done in full F90 style*
- Can MPI derived datatypes be leveraged?
- POP performance
  - on Cray X1, improved using alternate MPI code, also Co-Array Fortran
  - *Autopack more general purpose*