MPI 2.1
at
MPI Forum
Chicago, March 10-12, 2008
Version 1.3 – Combined Document

Rolf Rabenseifner
rabenseifner@hlrs.de
(Chairman of MPI 2.1 Task)

University of Stuttgart
High-Performance Computing-Center Stuttgart (HLRS)
www.hlrs.de

MPI-1.3 – Working Plan

- MPI-1.3 draft Mar 04, 2008 is available
- Reviews are done / will be done by the MPI 1.3 reviewing group:

1. **Bill Gropp**	(@meeting Jan-2008)
2. **Rolf Rabenseifner**	(@meeting Jan-2008)
3. **Adam Moody**	(@meeting Jan-2008)
4. **Puri Bangalore**	(@meeting Jan-2008)
5. **Terry Dontje**	(@meeting Mar-2008)

- In the final version of MPI-1.3, also the MPI-2.1 Ballot 4 items 5, 10.e, 14, and 15 will be included (if voted positive, March 11)
- Based on current available reviews, final version will be done until Mar 16, 2008
- Discussion only if differences between views of reviewers & editor
- Final review should be done until Mar 23, 2008. Okay **?**
- If there are still some open issues → reiteration
- Final version → Official reading at April, 1st vote June, 2nd vote Sep.

Annex: Slides from Jan. 2008 meeting

Schedule based on official rules

Rules and Procedures

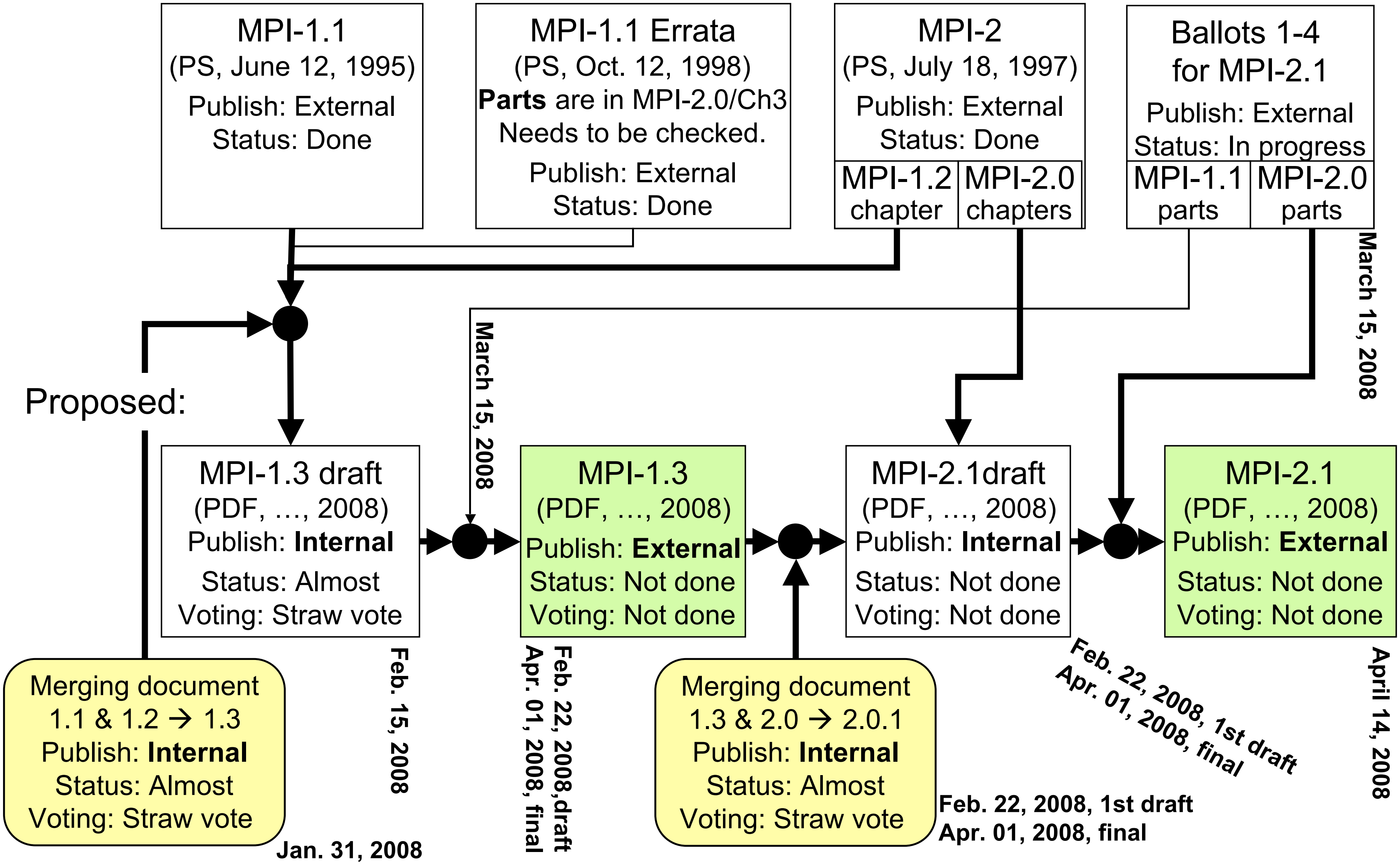
1. Here is a reminder of the traditional MPI voting rules, which have served us well. These rules have been extended to the email discussion of MPI erratas and have been applied to the errata ballots. We expect to adapt these rules, preserving their spirit, as we go forward.
2. One vote per organization
3. To vote, an organization must have been present at the last two MPI Forum meetings.
4. **Votes are taken twice, at separate meetings. Votes are preceded by a reading at an earlier meeting, to familiarize everyone with the issues.**
5. Measures pass on a simple majority.
6. Only items consistent with the charter can be considered.

From **http://www.mpi-forum.org/mpi2_1/index.htm**

For **MPI x.x combined documents**:
This reading at the MPI Forum meetings will be substituted by a review report through a review group. Each Forum member can be part of this group.
With the 1st official vote on a combined document (at next meeting), this modification of the voting rules is accept for that document.

MPI Standards document plan

Existing and in progress:



MPI-1.1
(PS, June 12, 1995)
Publish: External
Status: Done

MPI-1.1 Errata
(PS, Oct. 12, 1998)
Parts are in MPI-2.0/Ch3
Needs to be checked.
Publish: External
Status: Done

MPI-2
(PS, July 18, 1997)
Publish: External
Status: Done
MPI-1.2 chapter | MPI-2.0 chapters

Ballots 1-4
for MPI-2.1
Publish: External
Status: In progress
MPI-1.1 parts | MPI-2.0 parts

March 15, 2008

Proposed:

MPI-1.3 draft
(PDF, …, 2008)
Publish: **Internal**
Status: Almost
Voting: Straw vote

Feb. 15, 2008

March 15, 2008

MPI-1.3
(PDF, …, 2008)
Publish: **External**
Status: Not done
Voting: Not done

Feb. 22, 2008,draft
Apr. 01, 2008, final

MPI-2.1draft
(PDF, …, 2008)
Publish: **Internal**
Status: Not done
Voting: Not done

Feb. 22, 2008, 1st draft
Apr. 01, 2008, final

MPI-2.1
(PDF, …, 2008)
Publish: **External**
Status: Not done
Voting: Not done

April 14, 2008

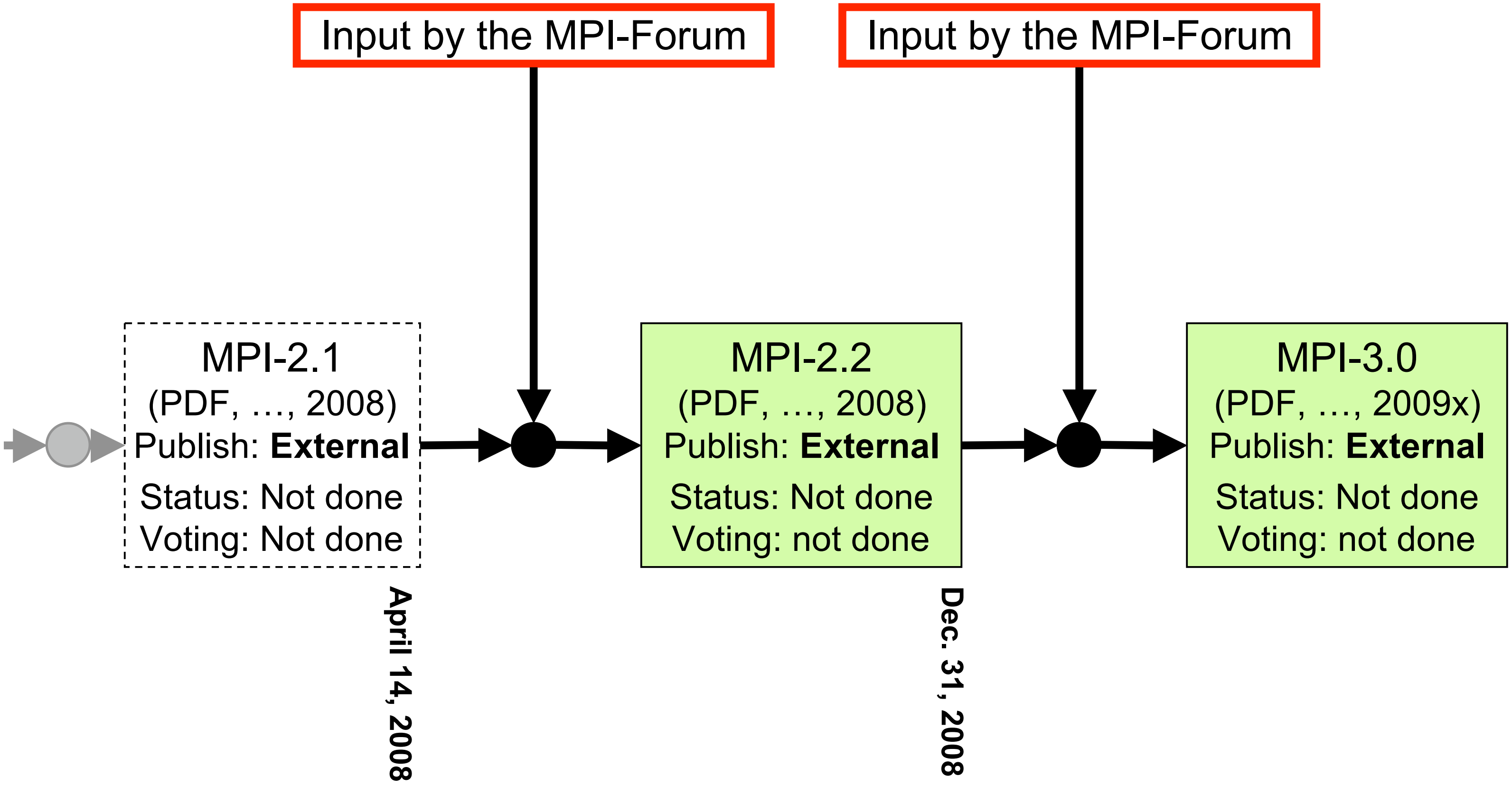
Merging document
1.1 & 1.2 → 1.3
Publish: **Internal**
Status: Almost
Voting: Straw vote

Jan. 31, 2008

Merging document
1.3 & 2.0 → 2.0.1
Publish: **Internal**
Status: Almost
Voting: Straw vote

Feb. 22, 2008, 1st draft
Apr. 01, 2008, final

MPI Standards document plan


Input by the MPI-Forum

Input by the MPI-Forum

MPI-2.1
(PDF, …, 2008)
Publish: **External**
Status: Not done
Voting: Not done

MPI-2.2
(PDF, …, 2008)
Publish: **External**
Status: Not done
Voting: not done

MPI-3.0
(PDF, …, 2009x)
Publish: **External**
Status: Not done
Voting: not done

April 14, 2008

Dec. 31, 2008

MPI 1.2.1 or MPI 1.3

- Should we name it MPI 1.3 instead of 1.2.1,
  including the change in MPI_GET_VERSION to MPI 1.3?
  - Yes: all-11
  - No: 2
  - Abstain: 9

- The rest of the document plan is okay?
  - Yes: all
  - No: 0
  - Abstain: 0

MPI 1.3 combined document

action point

- MPI 1.1 **+** Chap. 3 of MPI-2 (Version 1.2 of MPI) + some errata will be combined to → **MPI 1.3 combined document**
  - Jan.08 meeting:
    Short discussion and **defining a review group** who is reviewing the **MPI 1.3 merging plan** (printed copies available) and the **MPI 1.3 combined document**
  - See e-mail: From: Rainer Keller, Subject: Re: [mpi-21] Documents
    Date: Mon, 7 Jan 2008 12:13:14 +0100
  - Reporting by e-mail on mpi-21 reflector
  - Corrections if necessary
  - Final report of the reviewers at March 2008 meeting
  - **1st vote** by the MPI Forum at April 2008 meeting
  - **2nd (final) vote** by the MPI Forum at June 2008 meeting

MPI 1.3 combined document

Question:

Should we include

MPI-2-chapter3 +all new MPI-1 errata into the combined MPI 1.3?

Yes:

All-1

No:

0

Abstain:

1

- Do we want to include the MPI 1.1 errata already into this MPI 1.3 document?
- Pro:
  - This document is a "final" document telling the MPI-1 standard.
- Con:
  - Formally, it is not the right place. New stuff must be in MPI 2.1.
- My recommendation:
  - The "pro" outweighs the "con".

MPI 1.3 combined document – the “merging document”

Merge of MPI-1.1 (June 1995) and MPI-1.2 (July 1997) plus new Errata (MPI 1.2.1, 2008)

Versions-History page:

Version 1.3: ?????, 2008. This document combines the previous documents MPI 1.1 (June 12, 1995) and the MPI 1.2 Chapter in MPI-2 (July 18, 1997). Additional errata collected by the MPI Forum referring to MPI 1.1 and MPI 1.2 are also included in this document.

New text

Version 1.2: July, 18 1997. The MPI-2 Forum introduced MPI 1.2 as Chap.3 in the standard "MPI-2: Extensions to the Message-Passing Interface", July 18, 1997.“ This section contains clarifications and minor corrections to Version 1.1 of the MPI Standard. The only new function in MPI-1.2 is one for identifying to which version of the MPI Standard the implementation conforms. There are small differences between MPI-1 and MPI-1.1. There are very few differences ~~(only those discussed in this chapter)~~ between MPI-1.1 and MPI-1.2, but large differences ~~(the rest of this document)~~ between MPI-1.2 and MPI-2.

This text is from MPI 2.0, page 21, lines 14-19, but parentheses removed

Version 1.1: June, 1995. Beginning in March, 1995, the Mes…

Version 1.0: June, 1994. The Message Passing Interface Forum (MPIF), with participation from over 40 organizations, …

Existing MPI 1.1 text

Question:

Is the new history text okay?

Yes:

all

No:

0

Abstain:

0

MPI 1.2 combined document – the “merging document”

- 3.1: Integrated MPI_Get_version into Environmental Section, Inquiries -- from MPI-2, p. 21
(changes to appLang.tex and inquiry.tex)
  * the section title in MPI-2 is "Version Number", should not be changed?
  * MPI-2.0 Sect. 3.1 page 21 line 21 - page 22 line 2
added as new Sect. 7.1.1 in MPI-1.1
before current MPI-1.1 Sect 7.1.1 on page 190 line 21
remove last sentence on MPI-2.0 page 22 line 2:
"Its C++ binding can be found in the Annex, Section B.11."

- 3.2: MPI-1.0 and MPI-1.1 Clarifications
  * MPI-2.0 page 22 lines 4-10 not used (removed)

- 3.2.1: MPI_INITIALIZED: -- from MPI-2, p. 21 lines 14-15
  * added in MPI-1.1 page 200 line 11.
  * MPI-1.1 page 200 lines 10-11 must be modified because MPI_GET_VERSION
  * maybe also called before MPI_Init (And MPI_FINALIZED in MPI-2.0):
Changed: "is the only function that may be called before" to
"It is one of the few routines that “

- 3.2.2: Include clarification of MPI_FINALIZE -- from MPI-2, p. 22 line 18 - p. 24 line 48:
Replaces MPI-1.1 paragraph page 199 lines 46-48

MPI 1.2 combined document – the “merging document”

- 3.2.3 Clarification of status after MPI_WAIT and MPI_TEST -- from MPI-2, p. 25 lines 2-12
  Position in standard not completely obvious.
  Fits best after the definition of empty statuses in MPI-1, 3.7.3
  * i.e., after MPI-1.1 page 41 line 20

- 3.2.4 Clarification of MPI_INTERCOMM_CREATE -- from MPI-2, p. 25.
  Added to the section on Inter-Communication
  * Delete the text in parenthesis on MPI-1.1 page 158 line 31.
  * Substitute the sentence MPI-1.1 page 155 lines 36-37
    by MPI-2.0 page 25 lines 37-47

- 3.2.5 Clarification of MPI_INTERCOMM_MERGE -- from MPI-2, p. 26 lines 2-4
  Added paragraph on errorhandlers to MPI_INTERCOMM_MERGE
  * after MPI-1.1 page 160 line 13

- 3.2.6 Clarification of MPI_TYPE_SIZE -- from MPI-2, p. 26 lines 11-13
  Added advice to users
  * after MPI-1.1 page 70 line 43

MPI 1.2 combined document – the “merging document”

Question:

Is the merging decision for MPI-2 Sect.3.2.7 okay?

See next slide!

- 3.2.7 Clarification of MPI_REDUCE -- from MPI-2, p. 26
  Required extensive modification:

* MPI-2.0 page 26 lines 22-28 is substituting the text on MPI-1.1 page 114 lines 25-26.

* The sentence MPI 1.1, page 114, lines 26-27 “User-defined operators may operate on general, derived datatypes.” is not removed.

* MPI-2.0 page 26 lines 29-35 must be added after MPI-1.1 page 114 line 30.

* No need for additional new text "This is further explained in Section 4.9.4"

Review of this proposal on the next slide.

Question:

Is the merging decision for MPI-2 Sect.3.2.7 okay?

Yes:

all

No:

0

Abstain:

0

MPI 1.2 combined document – the “merging document” New proposal on Jan 2008 meeting

- 3.2.7 Clarification of MPI_REDUCE -- from MPI-2, p. 26

Blue: MPI 1.1, page 114, lines 25-30 Purple: MPI-2.0, page 26, lines 22-34

~~The datatype argument of MPI_REDUCE must be compatible with op. Predefined operators work only with the \MPI/ types listed in Sec. \ref{coll-predefined-op} and Sec. \ref{coll-minloc-maxloc}.~~ The datatype argument of MPI_REDUCE must be compatible with op. Predefined operators work only with the MPI types listed in Section ~~4.9.2~~ \ref{coll-predefined-op} and Section ~~4.9.3~~ \ref{coll-minloc-maxloc}. Furthermore, the datatype and op given for predefined operators must be the same on all processes.

Note that it is possible for users to supply different user-defined operations to MPI_REDUCE in each process. MPI does not define which operations are used on which operands in this case. **User-defined operators may operate on general, derived datatypes**. In this case, each argument that the reduce operation is applied to is one element described by such a datatype, which may contain several basic values. This is further explained in Section~\ref{subsec:coll-user-ops}.

> *Advice to users.* Users should make no assumptions about how MPI_REDUCE is implemented. Safest is to ensure that the same function is passed to MPI_REDUCE by each process. *(Advice to users.)*

Overlapping datatypes are permitted in ``send'' buffers. Overlapping datatypes in ``receive'' buffers are erroneous and may give unpredictable results.

This sentence is kept although MPI-2 requires deleting. (The content is correct)

MPI 1.2 combined document – the “merging document”

- 3.2.8 Clarification of Error Behaviour of Attribute Callback Function -- from MPI-2, p. 26 lines 38-39

Added to section 5.7.1, right after definition of delete_fn

* i.e., after MPI-1.1 page 170 line 7

MPI 1.2 combined document – the “merging document”

Question:

Should we keep the rationale MPI-2 Sect.3.2.9 page 27 line 1-32?

See next slides.

- 3.2.9 Clarification of MPI_PROBE and MPI_IPROBE -- from MPI-2, p. 27
  Replaced text, left out rationale...

  * The rationale may be kept, but all references should be
  * referencing the MPI 1.1 document (and not the new combined document)

  * TODO: Decision on Rationale must be done by MPI-2.1 Forum.

    The location for the rationale would be directly after the paragraph with the substituted text, i.e., after MPI 1.1, page 52, line 4.

Decision by the MPI-Forum (on next slides): The rationale is removed.

MPI 1.2 combined document – the “merging document”

- 3.2.9 Clarification of MPI_PROBE and MPI_IPROBE -- from MPI-2, p. 27

Page 52, lines 1 thru 3 (of MPI 1.1, the June 12, 1995 version without changebars)

> **A subsequent receive executed with the same context, and the source and tag returned in status by MPI_IPROBE will receive the message that was matched by the probe, if no other intervening receive occurs after the probe. If the receiving process is multi-threaded, it is the user's responsibility to ensure that the last condition holds.**

become:

> **A subsequent receive executed with the same communicator, and the source and tag returned in status by MPI_IPROBE will receive the message that was matched by the probe, if no other intervening receive occurs after the probe, and the send is not successfully cancelled before the receive. If the receiving process is multi-threaded, it is the user's responsibility to ensure that the last condition holds.**

Rationale.

The following program shows that the original MPI-1.1 definitions of cancel and probe are in conflict:

MPI 1.2 combined document – the “merging document”

- 3.2.9 Clarification of MPI_PROBE and MPI_IPROBE -- from MPI-2, p. 27

Rationale.

The following program shows that the original MPI-1.1 definitions of cancel and probe are in conflict:

```
Process 0                         Process 1
----------                        ----------
MPI_Init();                       MPI_Init();
MPI_Isend(dest=1);

                                  MPI_Probe();
MPI_Barrier();                    MPI_Barrier();
MPI_Cancel();
MPI_Wait();
MPI_Test_cancelled();
MPI_Barrier();                    MPI_Barrier();
                                  MPI_Recv();
```

MPI 1.2 combined document – the “merging document”

- 3.2.9 Clarification of MPI_PROBE and MPI_IPROBE -- from MPI-2, p. 27

Question:

Should we keep the rationale MPI-2 Sect.3.2.9 page 27 line 1-32?

Yes:

0

No:

All-7

Abstain:

7

Since the send has been cancelled by process 0, the wait must be local (MPI 1.1, page 54, line 13) and must return before the matching receive. For the wait to be local, the send must be successfully cancelled, and therefore must not match the receive in process 1 (MPI 1.1, page 54 line 29).

However, it is clear that the probe on process 1 must eventually detect an incoming message. MPI 1.1, p~~P~~age 52 line 1 makes it clear that the subsequent receive by process 1 must return the probed message.

The above are clearly contradictory, and therefore the text “…and the send is not successfully cancelled before the receive” must be added to MPI 1.1, line 3 of page 54.

An alternative solution (rejected) would be to change the semantics of cancel so that the call is not local if the message has been probed. This adds complexity to implementations, and adds a new concept of “state” to a message (probed or not). It would, however, preserve the feature that a blocking receive after a probe is local.

(End of rationale.)

MPI 1.2.1 combined document – Review Group

- The review group has to check the merging locations shown in the "merging document" from Rainer Keller
- They have to check the final "combined document", whether it implements the decisions in the "merging document"
- Proposal:
  - At least 4 persons to check the "merging document" and the final combined document based on the decisions in the merging document
  - MPI 1.3 reviewing group:

action point

1.	**Bill Gropp**	(@meeting Jan-2008)
2.	**Rolf Rabenseifner**	(@meeting Jan-2008)
3.	**Adam Moody**	
4.	**Puri Bangalore**	
5.	**Terry Dontje**	(not @meeting Jan-2008)
6.	**William Yu**	(not @meeting Jan-2008)

action point