

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Alexey Lastovetsky Tahar Kechadi
Jack Dongarra (Eds.)

Recent Advances in Parallel Virtual Machine and Message Passing Interface

15th European PVM/MPI Users' Group Meeting
Dublin, Ireland, September 7-10, 2008
Proceedings

Volume Editors

Alexey Lastovetsky
School of Computer Science and Informatics
University College Dublin
Belfield, Dublin 4, Ireland
E-mail: alexey.lastovetsky@ucd.ie

Tahar Kechadi
School of Computer Science and Informatics
University College Dublin
Belfield, Dublin 4, Ireland
E-mail: tahar.kechadi@ucd.ie

Jack Dongarra
Computer Science Department
University of Tennessee
Knoxville, TN, USA
E-mail: dongarra@cs.utk.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): D.1.3, D.3.2, F.1.2, G.1.0, B.2.1, C.1.2

LNCS Sublibrary: SL 2 – Programming and Software Engineering

ISSN	0302-9743
ISBN-10	3-540-87474-7 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-87474-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12511464 06/3180 5 4 3 2 1 0

Preface

Current thinking about state-of-the-art infrastructure for computational science is dominated by two concepts: computing clusters and computational grids. Cluster architectures consistently hold the majority of slots on the list of Top 500 supercomputer sites, and computational Grids, in both experimental and production deployments, have become common in academic, government and industrial research communities around the world. The message passing is the dominant programming paradigm for high-performance scientific computing on these architectures. MPI and PVM have emerged as standard programming environments in the message-passing paradigm. The EuroPVM/MPI conference series is the premier research event for high-performance parallel programming in the message-passing paradigm. Applications using parallel message-passing programming, pioneered in this research community, are having significant impact in the areas of computational science, such as bioinformatics, atmospheric sciences, chemistry, physics, astronomy, medicine, banking and finance, energy, etc.

EuroPVM/MPI is a flagship conference for this community, established as the premier international forum for researchers, users and vendors to present their latest advances in MPI and PVM. EuroPVM/MPI is the forum where fundamental aspects of message passing, implementations, standards, benchmarking, performance and new techniques are presented and discussed by researchers, developers and users from academia and industry.

EuroPVM/MPI 2008 was organized by the UCD School of Computer Science and Informatics in Dublin, September 7–10, 2008. This was the 15th issue of the conference, which takes place each year at a different European location. Previous meetings were held in Paris (2007), Bonn (2006), Sorrento (2005), Budapest (2004), Venice (2003), Linz (2002), Santorini (2001), Balatonfured (2000), Barcelona (1999), Liverpool (1998), Krakow (1997), Munich (1996), Lyon (1995), and Rome (1994).

The main topics of the meeting were formal verification of message passing programs, collective operations, parallel applications using the message passing paradigm, one-sided and point-to-point communication, MPI standard extensions or evolution, tools for performance evaluation and optimization, MPI-I/O, multi-core and multithreaded architectures, and heterogeneous platforms.

For this year's conference, the Program Committee Co-chairs invited seven outstanding researchers to present lectures on different aspects of the message-passing and multithreaded paradigms: George Bosilca, one of the leading members of OpenMPI, presented "The Next Frontier," Franck Cappello, one of the leading experts in the fault-tolerant message passing, presented "Fault Tolerance for PetaScale Systems: Current Knowledge, Challenges and Opportunities," Barbara Chapman, the leader of the OpenMP community, presented "Managing Multi-core with OpenMP," Al Geist, one of the authors of PVM, presented

“MPI Must Evolve or Die,” William Gropp, one of the leaders of MPICH, presented “MPI and Hybrid Programming Models for Petascale Computing,” Rolf Rabenseifner, one of the leading experts in optimization of collective operations, presented “Some Aspects of Message Passing on Future Hybrid Systems,” and Vaidy Sunderam, one of the authors of PVM, presented “From Parallel Virtual Machine to Virtual Parallel Machine: The Unibus System.”

In addition to the conference main track, the meeting featured the seventh edition of the special session “ParSim 2008 - Current Trends in Numerical Simulation for Parallel Engineering Environments.” The conference also included a full-day tutorial on “Using MPI-2: A Problem-Based Approach” by Ewing Rusty Lusk and William Gropp.

The response to the call for papers was very good: we received 55 full papers submitted to EuroPVM/MPI from 22 countries including Italy, Thailand, China, Germany, India, Greece, Spain, Japan, Switzerland, Ireland, Canada, Poland, Russia, Brazil, Denmark, Belgium, Mexico, Austria, Israel, Iran, France, and USA. Out of the 55 papers, 29 were selected for presentation at the conference. Each submitted paper was assigned to four members of the Program Committee (PC) for evaluation. The PC members either reviewed the papers themselves, or solicited external reviewers. The reviewing process went quite smoothly, and almost all reviews were returned, providing a solid basis for the Program Chairs to make the final selection for the conference program. The result was a well-balanced, focused and high-quality program. Out of the accepted 29 papers, four were selected as outstanding contributions to EuroPVM/MPI 2008, and were presented at special, plenary sessions:

- “Non-Data-Communication Overheads in MPI: Analysis on Blue Gene/P” by Pavan Balaji, Anthony Chan, William Gropp, Rajeev Thakur and Ewing Lusk
- “Architecture of the Component Collective Messaging Interface” by Sameer Kumar, Gabor Dozsa, Jeremy Berg, Bob Cernohous, Douglas Miller, Joseph Ratterman, Brian Smith and Philip Heidelberger
- “X-SRQ - Improving Scalability and Performance of Multi-Core InfiniBand Clusters” by Galen Shipman, Stephen Poole, Pavel Shamis and Ishai Rabinovitz
- “A Software Tool for Accurate Estimation of Parameters of Heterogeneous Communication Models” by Alexey Lastovetsky, Maureen O’Flynn and Vladimir Rychkov

Information about the conference can be found at the conference website: <http://pvmmpi08.ucd.ie>, which will be kept available.

The EuroPVM/MPI 2008 logo was designed by Alexander Kourinniiy.

The Program and General Chairs would like to thank all who contributed to making EuroPVM/MPI 2008 a fruitful and stimulating meeting, be they technical paper or poster authors, PC members, external referees, participants or sponsors. We would like to express our gratitude to all the members of the PC and the additional reviewers, who ensured the high quality of Euro PVM/MPI 2008 with their careful work.

Finally, we would like to thank the University College Dublin and the UCD School of Computer Science and Informatics for their support and efforts in organizing this event. In particular, we would like to thank Vladimir Rychkov (UCD), Alexander Ufimtsev (UCD), Angela Logue (UCD), An Nhien LeKhac (UCD) and Clare Comerford (UCD). Special thanks go to all the School of Computer Science and Informatics PhD students who helped in the logistics of the conference.

September 2008



Alexey Lastovetsky
Tahar Kechadi
Jack Dongarra

Organization

EuroPVM/MPI 2008 was organized by the School of Computer Science and Informatics, University College Dublin.

General Chair

Jack J. Dongarra

University of Tennessee, Knoxville, USA

Program Chairs

Alexey Lastovetsky

University College Dublin, Ireland

Tahar Kechadi

University College Dublin, Ireland

Program Committee

George Almasi

IBM, USA

Lamine Aouad

University College Dublin, Ireland

Ranieri Baraglia

ISTI-CNR, Italy

Richard Barrett

ORNL, USA

Gil Bloch

Mellanox, USA

George Bosilca

University of Tennessee, USA

Franck Cappello

INRIA, France

Brian Coghlan

Trinity College Dublin, Ireland

Yiannis Cotronis

University of Athens, Greece

Jean-Christophe Desplat

ICHEC, Ireland

Frederic Desprez

INRIA, France

Erik D'Hollander

University of Ghent, Belgium

Beniamino Di Martino

Second University of Naples, Italy

Jack Dongarra

University of Tennessee, USA

Edgar Gabriel

University of Houston, USA

Al Geist

Oak Ridge National Laboratory, USA

Patrick Geoffray

Myricom, USA

Michael Gerndt

Technische Universität München, Germany

Sergei Gorlatch

Universität Münster, Germany

Andrzej Goscinski

Deakin University, Australia

Richard L. Graham

ORNL, USA

William Gropp

Argonne National Laboratory, USA

Rolf Hempel

German Aerospace Center DLR, Germany

Thomas Herault

INRIA/LRI, France

Yutaka Ishikawa

University of Tokyo, Japan

Alexey Kalinov

Cadence, Russia

Tahar Kechadi	University College Dublin, Ireland
Rainer Keller	HLRS, Germany
Stefan Lankes	RWTH Aachen, Germany
Alexey Lastovetsky	University College Dublin, Ireland
Laurent Lefevre	INRIA, Universite de Lyon, France
Greg Lindahl	Blekkio, Inc., USA
Thomas Ludwig	University of Heidelberg, Germany
Ewing Rusty Lusk	Argonne National Laboratory, USA
Tomas Margalef	Universitat Autònoma de Barcelona, Spain
Jean-François Méhaut	IMAG, France
Bernd Mohr	Forschungszentrum Jülich, Germany
John P. Morrison	University College Cork, Ireland
Matthias Müller	Dresden University of Technology, Germany
Raymond Namyst	University of Bordeaux, France
Salvatore Orlando	University of Venice, Italy
Christian Perez	IRISA, France
Neil Pundit	Sandia National Laboratories, USA
Rolf Rabenseifner	HLRS, Germany
Thomas Rauber	Universität Bayreuth, Germany
Ravi Reddy	University College Dublin, Ireland
Casiano Rodriguez-Leon	University of La Laguna, Spain
Martin Schulz	Lawrence Livermore National Laboratory, USA
Andy Shearer	NUI Galway, Ireland
Jeffrey Squyres	Cisco, Inc., USA
Jesper Larsson Träff	C&C Research Labs, NEC Europe, Germany
Carsten Trinitis	Technische Universität München, Germany
Roland Wismueller	Universität Siegen, Germany
Felix Wolf	Forschungszentrum Jülich, Germany
Joachim Worringer	Dolphin Interconnect Solutions, Germany

Conference Organization

Alexey Lastovetsky	University College Dublin, Ireland
Tahar Kechadi	University College Dublin, Ireland
Vladimir Rychkov	University College Dublin, Ireland

External Referees

Erika Abraham	Stephen Childs
Olivier Aumage	Carsten Clauss
Boris Bierbaum	Camille Coti
Aurelien Bouteiller	Maurizio D'Arienzo
Ron Brightwell	Jan Duennweber
Michael Browne	Hubert Eichner

Markus Geimer
Ludovic Hablot
Mauro Iacono
Yvon Jégou
Julian Kunkel
Diego Latella
Pierre Lemarinier
Andreas Liehr
Stefano Marrone
Alberto F. Martín Huertas
Alastair McKinstry
Guillaume Mercier
Ruben Niederhagen
Ron Oldfield

Alexander Ploss
Marcela Printista
German Rodriguez-Herrera
John Ryan
Maraike Schellmann
Carlos Segura
Andy Shearer
Daniel Stodden
Honore Tapamo
Georg Wassen
Zhaofang Wen
Niall Wilson
Brian Wylie
Mathijs den Burger

Sponsors

The conference would have been significantly more expensive and much less pleasant to organize without the generous support of our industrial sponsors. EuroPVM/MPI 2008 gratefully acknowledges the contributions of the sponsors to a successful conference.

Platinum Level Sponsors



Gold Level Sponsors



Table of Contents

Invited Talks

The Next Frontier	1
<i>George Bosilca</i>	
Fault Tolerance for PetaScale Systems: Current Knowledge, Challenges and Opportunities	2
<i>Franck Cappello</i>	
Managing Multicore with OpenMP	3
<i>Barbara Chapman</i>	
MPI Must Evolve or Die	5
<i>Al Geist</i>	
MPI and Hybrid Programming Models for Petascale Computing	6
<i>William D. Gropp</i>	
Some Aspects of Message-Passing on Future Hybrid Systems	8
<i>Rolf Rabenseifner</i>	
From Parallel Virtual Machine to Virtual Parallel Machine: The Unibus System	11
<i>Vaidy Sunderam</i>	

Tutorial

EuroPVM/MPI Full-Day Tutorial. Using MPI-2: A Problem-Based Approach	12
<i>William Gropp and Ewing Lusk</i>	

Outstanding Papers

Non-data-communication Overheads in MPI: Analysis on Blue Gene/P	13
<i>Pavan Balaji, Anthony Chan, William Gropp, Rajeev Thakur, and Ewing Lusk</i>	
Architecture of the Component Collective Messaging Interface	23
<i>Sameer Kumar, Gabor Dozsa, Jeremy Berg, Bob Cernohous, Douglas Miller, Joseph Ratterman, Brian Smith, and Philip Heidelberg</i>	

X-SRQ – Improving Scalability and Performance of Multi-core InfiniBand Clusters	33
<i>Galen M. Shipman, Stephen Poole, Pavel Shamis, and Ishai Rabinovitz</i>	
A Software Tool for Accurate Estimation of Parameters of Heterogeneous Communication Models	43
<i>Alexey Lastovetsky, Vladimir Rychkov, and Maureen O’Flynn</i>	
Applications	
Sparse Non-blocking Collectives in Quantum Mechanical Calculations	55
<i>Torsten Hoeftler, Florian Lorenzen, and Andrew Lumsdaine</i>	
Dynamic Load Balancing on Dedicated Heterogeneous Systems	64
<i>Ismael Galindo, Francisco Almeida, and José Manuel Badía-Contelles</i>	
Communication Optimization for Medical Image Reconstruction Algorithms	75
<i>Torsten Hoeftler, Maraike Schellmann, Sergei Gorlatch, and Andrew Lumsdaine</i>	
Collective Operations	
A Simple, Pipelined Algorithm for Large, Irregular All-gather Problems	84
<i>Jesper Larsson Träff, Andreas Ripke, Christian Siebert, Pavan Balaji, Rajeev Thakur, and William Gropp</i>	
MPI Reduction Operations for Sparse Floating-point Data	94
<i>Michael Hofmann and Gudula Rünger</i>	
Library Internals	
A Prototype Implementation of MPI for SMARTMAP	102
<i>Ron Brightwell</i>	
Gravel: A Communication Library to Fast Path MPI	111
<i>Anthony Danalis, Aaron Brown, Lori Pollock, Martin Swamy, and John Cavazos</i>	
Message Passing for Multi-core and Multithreaded Architectures	
Toward Efficient Support for Multithreaded MPI Communication	120
<i>Pavan Balaji, Darius Buntinas, David Goodell, William Gropp, and Rajeev Thakur</i>	

MPI Support for Multi-core Architectures: Optimized Shared Memory Collectives	130
<i>Richard L. Graham and Galen Shipman</i>	

MPI Datatypes

Constructing MPI Input-output Datatypes for Efficient Transpacking	141
<i>Faisal Ghias Mir and Jesper Larsson Träff</i>	
Object-Oriented Message-Passing in Heterogeneous Environments	151
<i>Patrick Heckeler, Marcus Ritt, Jörg Behrend, and Wolfgang Rosenstiel</i>	

MPI I/O

Implementation and Evaluation of an MPI-IO Interface for GPFS in ROMIO	159
<i>Francisco Javier García Blas, Florin Isailă, Jesús Carretero, and Thomas Großmann</i>	
Self-consistent MPI-IO Performance Requirements and Expectations	167
<i>William D. Gropp, Dries Kimpe, Robert Ross, Rajeev Thakur, and Jesper Larsson Träff</i>	

Synchronisation Issues in Point-to-Point and One-Sided Communications

Performance Issues of Synchronisation in the MPI-2 One-Sided Communication API	177
<i>Lars Schneidenbach, David Böhme, and Bettina Schnor</i>	
Lock-Free Asynchronous Rendezvous Design for MPI Point-to-Point Communication	185
<i>Rahul Kumar, Amith R. Mamidala, Matthew J. Koop, Gopal Santhanaraman, and Dhabaleswar K. Panda</i>	

Tools

On the Performance of Transparent MPI Piggyback Messages	194
<i>Martin Schulz, Greg Bronevetsky, and Bronis R. de Supinski</i>	
Internal Timer Synchronization for Parallel Event Tracing	202
<i>Jens Doleschal, Andreas Knüpfer, Matthias S. Müller, and Wolfgang E. Nagel</i>	

A Tool for Optimizing Runtime Parameters of Open MPI	210
<i>Mohamad Chaarawi, Jeffrey M. Squyres, Edgar Gabriel, and Saber Feki</i>	
MADRE: The Memory-Aware Data Redistribution Engine	218
<i>Stephen F. Siegel and Andrew R. Siegel</i>	
MPIBlib: Benchmarking MPI Communications for Parallel Computing on Homogeneous and Heterogeneous Clusters	227
<i>Alexey Lastovetsky, Vladimir Rychkov, and Maureen O'Flynn</i>	

Verification of Message Passing Programs

Visual Debugging of MPI Applications	239
<i>Basile Schaeli, Ali Al-Shabibi, and Roger D. Hersch</i>	
Implementing Efficient Dynamic Formal Verification Methods for MPI Programs	248
<i>Sarvani Vakkalanka, Michael DeLisi, Ganesh Gopalakrishnan, Robert M. Kirby, Rajeev Thakur, and William Gropp</i>	
ValiPVM – A Graphical Tool for Structural Testing of PVM Programs	257
<i>Paulo Lopes de Souza, Eduardo T. Sawabe, Adenilso da Silva Simão, Silvia R. Vergilio, and Simone do Rocio Senger de Souza</i>	
A Formal Approach to Detect Functionally Irrelevant Barriers in MPI Programs	265
<i>Subodh Sharma, Sarvani Vakkalanka, Ganesh Gopalakrishnan, Robert M. Kirby, Rajeev Thakur, and William Gropp</i>	
Analyzing BLODFLOW: A Case Study Using Model Checking to Verify Parallel Scientific Software	274
<i>Stephen F. Siegel and Louis F. Rossi</i>	

ParSim

7th International Special Session on Current Trends in Numerical Simulation for Parallel Engineering Environments: New Directions and Work-in-Progress (ParSim 2008)	283
<i>Carsten Trinitis and Martin Schulz</i>	
LibGeoDecomp: A Grid-Enabled Library for Geometric Decomposition Codes	285
<i>Andreas Schäfer and Dietmar Fey</i>	
Using Arithmetic Coding for Reduction of Resulting Simulation Data Size on Massively Parallel GPGPUs	295
<i>Ana Balevic, Lars Rockstroh, Marek Wroblewski, and Sven Simon</i>	

Benchmark Study of a 3d Parallel Code for the Propagation of Large Subduction Earthquakes	303
<i>Mario Chavez, Eduardo Cabrera, Raúl Madariaga, Narciso Perea, Charles Moulinec, David Emerson, Mike Ashworth, and Alejandro Salazar</i>	

Posters Abstracts

Vis-OOMPI: Visual Tool for Automatic Code Generation Based on C++/OOMPI	311
<i>Chantana Phongpensri (Chantrapornchai) and Thanarat Rungthong</i>	
A Framework for Deploying Self-predefined MPI Communicators and Attributes	313
<i>Carsten Clauss, Boris Bierbaum, Stefan Lankes, and Thomas Bemmerl</i>	
A Framework for Proving Correctness of Adjoint Message-Passing Programs	316
<i>Uwe Naumann, Laurent Hascoët, Chris Hill, Paul Hovland, Jan Riehme, and Jean Utke</i>	
A Compact Computing Environment for a Windows Cluster: Giving Hints and Assisting Job Execution	322
<i>Yuichi Tsujita, Takuya Maruyama, and Yuhei Onishi</i>	
Introduction to Acceleration for MPI Derived Datatypes Using an Enhancer of Memory and Network	324
<i>Noboru Tanabe and Hironori Nakajo</i>	
Efficient Collective Communication Paradigms for Hyperspectral Imaging Algorithms Using HeteroMPI	326
<i>David Valencia, Antonio Plaza, Vladimir Rychkov, and Alexey Lastovetsky</i>	
An MPI-Based System for Testing Multiprocessor and Cluster Communications	332
<i>Alexey N. Salnikov and Dmitry Y. Andreev</i>	
MPI in Wireless Sensor Networks	334
<i>Yannis Mazzer and Bernard Tourancheau</i>	

Erratum

Dynamic Load Balancing on Dedicated Heterogeneous Systems	E1
<i>Ismael Galindo, Francisco Almeida, Vicente Blanco, and José Manuel Badía-Contelles</i>	

Author Index	341
--------------------	-----