Software Automatic Tuning

Ken Naono • Keita Teranishi
John Cavazos • Reiji Suda
Editors

Software Automatic Tuning

From Concepts to State-of-the-Art Results

Springer

Editors
Ken Naono
Central Research Laboratory
Hitachi Ltd.
1-280 Higashi-Koigakubo
Kokubunji-shi
Tokyo 185-8601, Japan
ken.naono.aw@hitachi.com

John Cavazos
Department of Computer
& Information Sciences
University of Delaware
101 Smith Hall
Newark, DE 19716, USA
cavazos@cis.udel.edu

Keita Teranishi
Cray Inc.
Suite # 210, 380 Jackson St.
St Paul, MN 55101, USA
keita@cray.com

Reiji Suda
Department of Computer Science
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033, Japan
reiji@is.s.u-tokyo.ac.jp

ISBN 978-1-4419-6934-7 e-ISBN 978-1-4419-6935-4
DOI 10.1007/978-1-4419-6935-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010934406

Preface

Software automatic tuning is a technology paradigm enabling software adaptation to a variety of computational conditions. Originating from the stream of research works on highperformance computing, it is considered to be the most promising approach to the required performance advancements on the next generation supercomputing platforms. Also, as its effectiveness is widely recognized, its scope is expanding from scientific and engineering computations to general purpose computations.

This book is a fruit of international collaboration developed in iWAPT workshop series, where iWAPT stands for International Workshop on Automatic Performance Tuning. The first workshop (iWAPT 2006) has been held in the University of Tokyo on September 12, 2006. It was a 1-day workshop with two invited presentations from USA and four invited presentations from Japan. iWAPT 2007 was a 2-day workshop with three invited presentations, seven refereed oral presentations, and eight poster presentations, held at the University of Tokyo. In 2008, iWAPT was held in conjunction with IEEE Cluster 2008 at Tsukuba, with two invited presentations and seven refereed oral presentations. iWAPT 2009 was a 2-day workshop with two invited presentations seven refereed oral presentations and four poster presentations, held at the University of Tokyo. iWAPT 2010 will be held in conjunction with VECPAR at Berkeley, CA, USA iWAPT is now lead by International Steering Committee, where five members are from Japan, four from USA, and one from Europe (see http://www.iwapt.org).

This book consists of 20 chapters that encompass almost all the areas of automatic tuning research: matrix kernels, FFT, matrix decompositions, iterative solvers, numerical library, scientific computing, GPGPU, parallel processing, autotuning framework, mathematical methods of autotuning, programming languages, and compiler technologies. The first chapter is an introduction to software automatic tuning, written by the editors. Six chapters are invited papers. Two of them are written by invited speakers of iWAPT workshops, and four of them are by members of organizing committee of iWAPT workshops. Thirteen chapters are peerreviewed contributed papers. Six come from iWAPT 2009, two from iWAPT 2007, and the other five papers are newly submitted for this publication. We arrange the chapters in the order of topics, rather than in the order of origins.

The editors appreciate the contributions of the authors of the chapters and the organizers, presenters and participants of the iWAPT workshop series. We are especially grateful to R. Clint Whaley for their invaluable efforts for this publication. We are also thankful to Charles Glaser and Amanda Davis of Springer USA for their help.

We sincerely hope that this book contributes the progress of software automatic tuning technology and world's welfare through information technology.

Tokyo, Japan — Ken Naono
St Paul, MN, USA — Keita Teranishi
Newark, DE, USA — John Cavazos
Tokyo, Japan — Reiji Suda

Contents

Contributors

Charles Audet Department of Mathematics and Industrial Engineering, Ecole Polytechnique, Montréal, QC, Canada
and
GERAD, Montréal, QC, Canada, charles.audet@gerad.ca

Sanjukta Bhowmick Department of Computer Science, University of Nebraska at Omaha, sbhowmick@unomaha.edu

Angela Demke Brown University of Toronto, Canada M5S 2E4, demke@cs.toronto.edu

John Cavazos University of Delaware, Newark, DE, USA, cavazos@cis.udel.edu

Jacqueline Chame Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA, jchame@isi.edu

Chun Chen School of Computing, University of Utah, Salt Lake City, UT 84112, USA, chunchen@cs.utah.edu

Cong-Kien Dang GERAD, Montréal, QC, Canada, kien.cong.dang@gerad.ca

Brian Demsky University of California, Irvine, CA, USA, bdemsky@uci.edu

Masashi Egi Central Research Laboratory, Hitachi Ltd., 1-280, Higashi-koigakubo, Kokubunji, Tokyo, Japan, masashi.egi.zj@hitachi.com

Victor Eijkhout Advanced Computing Center, The University of Texas at Austin, eijkhout@tacc.utexas.edu

Yoav Freund Department of Computer Science and Engineering, University of California, San Diego, yfreund@ucsd.edu

Erika Fuentes Microsoft Inc., efuentes@cs.utk.edu

Takeshi Fukaya Nagoya University, Nagoya, Aichi 464-8603, Japan, t-fukaya@na.cse.nagoya-u.ac.jp

Serban Georgescu Department of Quantum Engineering and Systems Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8654, Japan, serban@nihonbashi.race.u-tokyo.ac.jp

Mary W. Hall School of Computing, University of Utah, Salt Lake City, UT 84112, USA, mhall@cs.utah.edu

Hidehiko Hasegawa University of Tsukuba, Ibaraki 305–8550, Japan, hasegawa@slis.tsukuba.ac.jp

Takashige Hoshiai Graduate School of Global Information and Telecommunication Studies, Waseda University, 1-3-10, Nishiwaseda, Shinjyuku, Tokyo, Japan, hoshiai@pcl.cs.waseda.jp

Paul D. Hovland Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA, hovland@mcs.anl.gov

Yusuke Ishikawa Graduate School of Science and Engineering, Saitama University, Saitama, Japan, s09mm305@mail.saitama-u.ac.jp

Shoji Itoh Information Technology Center, The University of Tokyo, Yayoi 2-11-16, Bunkyo, Tokyo 113-8658, Japan, itosho@cc.u-tokyo.ac.jp

Tamito Kajiyama CITI, DI/FCT, Universidade Nova de Lisboa, Caparica 2829–516, Portugal, t.kajiyama@di.fct.unl.pt

Hidehiro Kanemitsu Graduate School of Global Information and Telecommunication Studies, Waseda University, 1-3-10, Nishiwaseda, Shinjyuku, Tokyo, Japan, kanemih@ruri.waseda.jp

Takahiro Katagiri Information Technology Center, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8658, Japan, katagiri@cc.u-tokyo.ac.jp

David Keyes Department of Applied Physics and Applied Mathematics, Columbia University
and
King Abdullah University of Science and Technology, david.keyes@kaust.edu.sa

Hiroaki Kobayashi Cyberscience Center, Tohoku University, 6-3 Aramaki-aza-aoba, Aoba, Sendai 980-8578, Japan, koba@isc.tohoku.ac.jp

Kazuhiko Komatsu Cyberscience Center, Tohoku University, 6-3 Aramaki-aza-aoba, Aoba, Sendai 980-8578, Japan, komatsu@sc.isc.tohoku.ac.jp

Yutaka Kuwajima Graduate School of Science and Engineering, Saitama University, Saitama, Japan, kuwa@mail.saitama-u.ac.jp

Gilhyon Lee Graduate School of Global Information and Telecommunication Studies, Waseda University, 1-3-10, Nishiwaseda, Shinjyuku, Tokyo, Japan, ghlee@akane.waseda.jp

Hidenori Nakazato Graduate School of Global Information and Telecommunication Studies, Waseda University, 1-3-10, Nishiwaseda, Shinjyuku, Tokyo, Japan, nakazato@waseda.jp

Ken Naono Central Research Laboratory, Hitachi Ltd., 1-280, Higashi-koigakubo, Kokubunji, Tokyo, Japan, ken.naono.aw@hitachi.com

Akira Nishida Kyushu University, Fukuoka 812–8581, Japan,
nishida@cc.kyushu-u.ac.jp

Akira Nukada Tokyo Institute of Technology, Tokyo 152–8552, Japan,
nukada@smg.is.titech.ac.jp

Hiroshi Okuda Research into Artifacts, Center for Engineering (RACE), The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8568, Japan,
okuda@race.u-tokyo.ac.jp

Dominique Orban Department of Mathematics and Industrial Engineering, École Polytechnique, Montréal, QC, Canada
and
GERAD, Montréal, QC, Canada, dominique.orban@gerad.ca

Gennady Pekhimenko Carnegie Mellon University, 5000 Forbes Ave, GHC, Pittsburgh PA 15213, gpekhime@cs.cmu.edu

Takao Sakurai Central Research Laboratory, Hitachi Ltd., 1-280, Higashi-koigakubo, Kokubunji, Tokyo, Japan, takao.sakurai.ju@hitachi.com

Katsuto Sato Graduate School of Information Sciences, Tohoku University, 6-3 Aramaki-aza-aoba, Aoba, Sendai 980-8578, Japan, katuto@sc.isc.tohoku.ac.jp

Yuta Sawa Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan, yuta.sawa.eh@hitachi.com

Lakshminarasimhan Seshagiri Scalable Computing Laboratory, The Ames Laboratory, US DoE, Ames, IA 50011, USA, sln@scl.ameslab.gov

Takaomi Shigehara Graduate School of Science and Engineering, Saitama University, Saitama, Japan, sigehara@nc.ics.saitama-u.ac.jp

Jaewook Shin Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA, jaewook@mcs.anl.gov

Masha Sosonkina Scalable Computing Laboratory, The Ames Laboratory, US DoE, Ames, IA 50011, USA, masha@scl.ameslab.gov

Reiji Suda Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
and
CREST, JST, Tokyo, Japan, reiji@is.s.u-tokyo.ac.jp

Masaaki Sugihara Graduate School of Information Science and Technology, The University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo 113-8656, Japan,
m_sugihara@mist.i.u-tokyo.ac.jp

Daisuke Takahashi Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan,
daisuke@cs.tsukuba.ac.jp

Hiroyuki Takizawa Graduate School of Information Sciences,
Tohoku University, 4F, 6-3 Aramaki-aza-aoba, Aoba-ku, Sendai, 980-8578, Japan,
tacky@isc.tohoku.ac.jp

Junichi Tamura Graduate School of Science and Engineering,
Saitama University, Saitama, Japan, s08mm318@mail.saitama-u.ac.jp

Keita Teranishi Cray Inc., Suite # 210, 380 Jackson st., St Paul, MN 55101, USA,
keita@cray.com

Yoshiyori Urano Graduate School of Global Information and Telecommunication
Studies, Waseda University, 1-3-10, Nishiwaseda, Shinjyuku, Tokyo, Japan,
muranolt@waseda.jp

R. Clint Whaley Department of Computer Science, Univ of TX, San Antonio,
TX 78249, whaley@cs.utsa.edu

Meng-Shiou Wu Scalable Computing Laboratory, The Ames Laboratory,
US DoE, Ames, IA 50011, USA, mswu@scl.ameslab.gov

Yusaku Yamamoto Kobe University, Kobe, Hyogo 657-8501, Japan,
yamamoto@cs.kobe-u.ac.jp

Zhao Zhang Department of Electrical and Computer Engineering, Iowa State
University, Ames, IA 50011, USA, zzhang@iastate.edu

Jin Zhou University of California, Irvine, CA, USA, jzhou1@uci.edu