

2021 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC 2021)

**St. Louis, Missouri, USA
15 November 2021**



**IEEE Catalog Number: CFP21J40-POD
ISBN: 978-1-6654-1125-7**

**Copyright © 2021 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP21J40-POD
ISBN (Print-On-Demand):	978-1-6654-1125-7
ISBN (Online):	978-1-6654-1124-0
ISSN:	2768-4237

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2021 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC) **MLHPC 2021**

Table of Contents

Message from the Workshop Chairs	vi
Workshop Organization	vii

MLHPC21 Technical Papers

Semantic-Aware Lossless Data Compression for Deep Learning Recommendation Model (DLRM)	1
<i>Sarunya Pumma (Advanced Micro Devices, Inc. (AMD)) and Abhinav Vishnu (Advanced Micro Devices, Inc. (AMD))</i>	
Colmena: Scalable Machine-Learning-Based Steering of Ensemble Simulations for High Performance Computing	9
<i>Logan Ward (Argonne National Laboratory, USA; University of Chicago, USA), Ganesh Sivaraman (Argonne National Laboratory, USA), J. Gregory Pauloski (University of Chicago, USA), Yadu Babuji (University of Chicago, USA), Ryan Chard (Argonne National Laboratory, USA), Naveen Dandu (University of Chicago, USA), Paul C. Redfern (University of Chicago, USA), Rajeev S. Assary (University of Chicago, USA), Kyle Chard (University of Chicago, USA), Larry A. Curtiss (University of Chicago, USA), Rajeev Thakur (Argonne National Laboratory, USA), and Ian Foster (Argonne National Laboratory, USA; University of Chicago, USA)</i>	
Production Deployment of Machine-Learned Rotorcraft Surrogate Models on HPC	21
<i>Wesley Brewer (DoD HPCMP PET/GDIT, USA), Daniel Martinez (Science and Technology Corporation, USA), Mathew Boyer (DoD HPCMP PET/GDIT, USA), Dylan Jude (US Army DEVCOM AvMC DSE, USA), Andy Wissink (US Army DEVCOM AvMC DSE, USA), Ben Parsons (DoD HPCMP, USA), Junqi Yin (Oak Ridge Leadership Computing Facility, USA), and Valentine Anantharaj (Oak Ridge Leadership Computing Facility, USA)</i>	

MLPerfTM HPC: A Holistic Benchmark Suite for Scientific Machine Learning on HPC Systems 33

Steven Farrell (Lawrence Berkeley National Laboratory, USA), Murali Emani (Argonne National Laboratory), Jacob Balma (Maxwell Labs), Lukas Drescher (Swiss National Supercomputing Center, Switzerland), Aleksandr Drozd (RIKEN Center for Computational Science (R-CCS), Japan), Andreas Fink (Swiss National Supercomputing Centre (CSCS)), Geoffrey Fox (University of Virginia), David Kanter (MLCommons), Thorsten Kurth (NVIDIA, Switzerland), Peter Mattson (Google LLC, USA), Dawei Mu (National Center for Supercomputing Applications (NCSA), USA), Amit Ruhela (Texas Advanced Computing Center (TACC)), Kento Sato (RIKEN Center for Computational Science (R-CCS), Japan), Koichi Shirahata (Fujitsu Ltd, Japan), Tsuguchika Tabaru (Fujitsu Ltd, Japan), Aristeidis Tsaris (Oak Ridge National Laboratory (ORNL), USA), Jan Balewski (Lawrence Berkeley National Laboratory (LBNL), USA), Ben Cumming (Swiss National Supercomputing Centre (CSCS), Switzerland), Takumi Danjo (Fujitsu Ltd, Japan), Jens Domke (RIKEN Center for Computational Science (R-CCS), Japan), Takaaki Fukai (RIKEN Center for Computational Science (R-CCS), Japan), Naoto Fukumoto (Fujitsu Ltd, Japan), Tatsuya Fukushima (Fujitsu Ltd, Japan), Balazs Gerofi (RIKEN Center for Computational Science (R-CCS), Japan), Takumi Honda (Fujitsu Ltd, Japan), Toshiyuki Imamura (RIKEN Center for Computational Science (R-CCS), Japan), Akihiko Kasagi (Fujitsu Ltd, Japan), Kentaro Kawakami (Fujitsu Ltd, Japan), Shuhei Kudo (RIKEN Center for Computational Science (R-CCS), Japan), Akiyoshi Kuroda (RIKEN Center for Computational Science (R-CCS), Japan), Maxime Martinasso (Swiss National Supercomputing Centre (CSCS), Switzerland), Satoshi Matsuoka (RIKEN Center for Computational Science (R-CCS), Japan), Henrique Mendonça (Swiss National Supercomputing Centre (CSCS), Switzerland), Kazuki Minami (RIKEN Center for Computational Science (R-CCS), Japan), Prabhat Ram (Microsoft, USA), Takashi Sawada (Fujitsu Ltd, Japan), Mallikarjun Shankar (Oak Ridge National Laboratory (ORNL), USA), Tom St. John (Cruise, USA), Akihiro Tabuchi (Fujitsu Ltd, Japan), Venkatram Vishwanath (Argonne National Laboratory), Mohamed Wahib (National Institute of Advanced Industrial Science and Technology (AIST), Japan), Masafumi Yamazaki (Fujitsu Ltd, Japan), and Junqi Yin (Oak Ridge National Laboratory (ORNL))

High-Performance Deep Learning Toolbox for Genome-Scale Prediction of Protein Structure and Function 46

Mu Gao (Georgia Institute of Technology), Peik Lund-Andersen (University of Idaho), Alex Morehead (University of Missouri), Sajid Mahmud (University of Missouri), Chen Chen (University of Missouri), Xiao Chen (University of Missouri), Nabin Giri (University of Missouri), Raj S. Roy (University of Missouri), Farhan Qadir (University of Missouri), T. Chad Effler (Oak Ridge National Laboratory), Ryan Prout (Oak Ridge National Laboratory), Subil Abraham (Oak Ridge National Laboratory), Wael Elwasif (Oak Ridge National Laboratory), N. Quentin Haas (Oak Ridge National Laboratory), Jeffrey Skolnick (Georgia Institute of Technology), Jianlin Cheng (Georgia Institute of Technology; University of Missouri), and Ada Sedova (Georgia Institute of Technology; Oak Ridge National Laboratory)

HPCFAIR: Enabling FAIR AI for HPC Applications	58
<i>Gaurav Verma (Stony Brook University, USA), Murali Emani (Argonne National Laboratory, USA), Chunhua Liao (Lawrence Livermore National Laboratory, USA), Pei-Hung Lin (Lawrence Livermore National Laboratory, USA), Tristan Lucas Vanderbruggen (Lawrence Livermore National Laboratory, USA), Xipeng Shen (North Carolina State University, USA), and Barbara Chapman (Stony Brook University, USA)</i>	
HPC Ontology: Towards a Unified Ontology for Managing Training Datasets and AI Models for High-Performance Computing	69
<i>Chunhua Liao (Lawrence Livermore National Laboratory, USA), Pei-Hung Lin (Lawrence Livermore National Laboratory, USA), Gaurav Verma (Stony Brook University, USA), Tristan Vanderbruggen (Lawrence Livermore National Laboratory, USA), Murali Emani (Argonne National Laboratory, USA), Zifan Nan (North Carolina State University, USA), and Xipeng Shen (North Carolina State University, USA)</i>	
HYPPO: A Surrogate-Based Multi-Level Parallelism Tool for Hyperparameter Optimization	81
<i>Vincent Dumont (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA), Casey Garner (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA; Department of Mathematics, University of Minnesota Twin Cities, USA), Anuradha Trivedi (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA; Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, USA), Chelsea Jones (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA; Department of Statistical Sciences & Operations Research, Virginia Commonwealth University, USA), Vidya Ganapati (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA; Department of Engineering, Swarthmore College, USA), Julianne Mueller (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA), Talita Perciano (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA), Mariam Kiran (Computing Sciences Area, Lawrence Berkeley National Laboratory, USA), and Marc Day (Computational Science Center, National Renewable Energy Laboratory, USA)</i>	
Is Disaggregation Possible for HPC Cognitive Simulation?	94
<i>Michael Wyatt (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory), Valen Yamamoto (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory), Zoe Tosi (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory), Ian Karlin (Intel Corporation), and Brian Van Essen (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory)</i>	
Author Index	107