

2021 IEEE/ACM Workshop on Memory Centric High Performance Computing (MCHPC 2021)

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



**IEEE Catalog Number: CFP21W51-POD
ISBN: 978-1-6654-2251-2**

**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:	CFP21W51-POD
ISBN (Print-On-Demand):	978-1-6654-2251-2
ISBN (Online):	978-1-6654-1417-3

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 Memory Centric High Performance Computing (MCHPC) **MCHPC 2021**

Table of Contents

Message from the Workshop Chairs	iv
MCHPC'21 Invited Talks	v
Workshop Organization	vii

Advanced Techniques for Using Heterogeneous Memory

FreeLunch: Compression-based GPU Memory Management for Convolutional Neural Networks	1
<i>Shaurya Patel (University of Massachusetts, Amherst; University of British Columbia), Tongping Liu (University of Massachusetts, Amherst), and Hui Guan (University of Massachusetts, Amherst)</i>	
Using Bandwidth Throttling to Quantify Application Sensitivity to Heterogeneous Memory	9
<i>Clément Foyer (Univ. Bordeaux, France) and Brice Goglin (Univ. Bordeaux, France)</i>	

Application and Memory Optimization Techniques

Performance and Energy Improvement of ECP Proxy App SW4lite under Various Workloads	17
<i>Xingfu Wu (The University of Chicago, USA), Valerie Taylor (The University of Chicago, USA), and Zhiling Lan (Illinois Institute of Technology, USA)</i>	
Memory Optimizations for Sparse Linear Algebra on GPU Hardware	25
<i>Aaron Walden (NASA Langley Research Center, USA), Mohammad Zubair (Old Dominion University, USA), Christopher P. Stone (National Institute of Aerospace, USA), and Eric J. Nielsen (NASA Langley Research Center, USA)</i>	
Author Index	33