

# **2019 IEEE Hot Chips 31 Symposium (HCS 2019)**

**Cupertino, California, USA  
18 – 20 August 2019**



**IEEE Catalog Number: CFP19HCS-POD  
ISBN: 978-1-7281-2090-4**

**Copyright © 2019 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:	CFP19HCS-POD
ISBN (Print-On-Demand):	978-1-7281-2090-4
ISBN (Online):	978-1-7281-2089-8
ISSN:	2573-203X

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# TABLE OF CONTENTS

<b>ZION: FACEBOOK NEXT-GENERATION LARGE MEMORY TRAINING PLATFORM</b> .....	1
<i>Misha Smelyanskiy</i>	
<b>DAVINCI: A SCALABLE ARCHITECTURE FOR NEURAL NETWORK COMPUTING</b> .....	12
<i>Heng Liao ; Jiajin Tu ; Jing Xia ; Xiping Zhou</i>	
<b>DEEP LEARNING TRAINING AT SCALE SPRING CREST DEEP LEARNING ACCELERATOR (INTEL® NERVANA™ NNP-T)</b> .....	34
<i>Andrew Yang</i>	
<b>WAFER-SCALE DEEP LEARNING</b> .....	44
<i>Sean Lie</i>	
<b>HABANA LABS</b> .....	60
<i>Eitan Medina</i>	
<b>AMD “ZEN 2”</b> .....	75
<i>David Suggs ; Dan Bouvier ; Michael Clark ; Kevin Lepak ; Mahesh Subramony</i>	
<b>ARM NEOVERSE N1 CLOUD-TO-EDGE INFRASTRUCTURE SOCS</b> .....	87
<i>Andrea Pellegrini ; Chris Abernathy</i>	
<b>IBM’S NEXT GENERATION POWER PROCESSOR</b> .....	98
<i>Jeff Stuecheli ; Scott Willenborg ; William Starke</i>	
<b>THE TRUE PROCESSING IN MEMORY ACCELERATOR</b> .....	108
<i>Fabrice Devaux</i>	
<b>A PROGRAMMABLE EMBEDDED MICROPROCESSOR FOR BIT-SCALABLE IN-MEMORY COMPUTING</b> .....	120
<i>Hongyang Jia ; H. Valavi ; Y. Tang ; J. Zhang ; N. Verma</i>	
<b>INTEL OPTANE DATA CENTER PERSISTENT MEMORY</b> .....	135
<i>Lily Looi ; Jianping Jane Xu</i>	
<b>DELIVERING THE FUTURE OF HIGH-PERFORMANCE COMPUTING</b> .....	148
<i>Lisa Su</i>	
<b>CREATING AN AGILE HARDWARE FLOW</b> .....	170
<i>Keyi Zhang</i>	
<b>ML BENCHMARK DESIGN CHALLENGES</b> .....	186
<i>Peter Mattson</i>	
<b>LAKEFIELD: HYBRID CORES IN 3D PACKAGE</b> .....	204
<i>Sanjeev Khushu ; Wilfred Gomes</i>	
<b>JINTIDE®: A HARDWARE SECURITY ENHANCED SERVER CPU WITH XEON® CORES UNDER RUNTIME SURVEILLANCE BY AN IN-PACKAGE DYNAMICALLY RECONFIGURABLE PROCESSOR</b> .....	214
<i>Leibo Liu ; Ao Luo ; Guanhua Li ; Jianfeng Zhu ; Yong Wang ; Gang Shan ; Jianfeng Pan ; Shouyi Yin ; Shaojun Wei</i>	
<b>RTX ON – THE NVIDIA TURING GPU</b> .....	227
<i>John Burgess</i>	
<b>7NM “NAVY” GPU - A GPU BUILT FOR PERFORMANCE AND EFFICIENCY</b> .....	241
<i>Mike Mantor</i>	
<b>SILICON AT THE HEART OF HOLOLENS 2</b> .....	255
<i>Elene Terry</i>	
<b>CYW89459: HIGH PERFORMANCE AND LOW POWER WI-FI AND BLUETOOTH 5.1 COMBO CHIP FOR IOT AND AUTOMOTIVE</b> .....	268
<i>Kamesh Medepalli</i>	
<b>OUROBOROS: AN INFERENCE ENGINE FOR DEEP LEARNING BASED TTS ON EMBEDDED DEVICES</b> .....	277
<i>Jiansong Zhang ; Lixue Xia ; Zhao Jiang ; Hao Liang ; Jiaoyan Chen ; Shouda Liu ; Wei Lin ; Yuan Xie</i>	
<b>COMPUTE AND REDUNDANCY SOLUTION FOR THE FULL SELF-DRIVING COMPUTER</b> .....	291
<i>Pete Bannon ; Ganesh Venkataramanan ; Debjit Das Sarma ; Emil Talpes</i>	
<b>A 0.11 PJ/OP, 0.32-128 TOPS, SCALABLE MULTI-CHIP-MODULE-BASED DEEP NEURAL NETWORK ACCELERATOR DESIGNED WITH A HIGH-PRODUCTIVITY VLSI METHODOLOGY</b> .....	302
<i>Rangharajan Venkatesan ; Yakun Sophia Shao ; Brian Zimmer ; Jason Clemons ; Matthew Fojtik ; Nan Jiang ; Ben Keller ; Alicia Klinefelter ; Nathaniel Pinckney ; Priyanka Raina ; Stephen G. Tell ; Yanqing Zhang ; William J. Dally ; Joel S. Emer ; C. Thomas Gray ; Stephen W. Keckler ; Brucek Khailany</i>	

<b>XILINX FIRST 7NM DEVICE: VERSAL AI CORE (VC1902)</b> .....	314
<i>Sagheer Ahmad ; Sridhar Subramanian ; Vamsi Boppana ; Shankar Lakka ; Fu-Hing Ho ; Tomai Knopp ; Juanjo Noguera ; Gaurav Singh ; Ralph Wittig</i>	
<b>SPRING HILL (NNP-I 1000) INTEL'S DATA CENTER INFERENCE CHIP</b> .....	328
<i>Ofri Wechsler ; Michael Behar ; Bharat Daga</i>	
<b>IC TECHNOLOGY – WHAT WILL THE NEXT NODE OFFER US?</b> .....	334
<i>H.-S. Philip Wong ; R. Willard ; Inez Kerr Bell</i>	
<b>GEN-Z CHIPSET FOR EXASCALE FABRICS</b> .....	360
<i>Patrick Knebel ; Dan Berkram ; Al Davis ; Darel Emmot ; Paolo Faraboschi ; Gary Gostin</i>	
<b>TERAPHY: A CHIPLLET TECHNOLOGY FOR LOW-POWER, HIGH-BANDWIDTH IN-PACKAGE OPTICAL I/O</b> .....	371
<i>Mark Wade</i>	
<b>Author Index</b>	