

2020 IEEE Hot Chips 32 Symposium (HCS 2020)

**Palo Alto, California, USA
16-18 August 2020**



**IEEE Catalog Number: CFP20HCS-POD
ISBN: 978-1-7281-7130-2**

**Copyright © 2020 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:	CFP20HCS-POD
ISBN (Print-On-Demand):	978-1-7281-7130-2
ISBN (Online):	978-1-7281-7129-6
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
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

A TECHNICAL OVERVIEW OF CORTEX-M55 AND ETHOS-U55: ARM'S MOST CAPABLE PROCESSORS FOR ENDPOINT AI	1
<i>Allan Skillman, Tomas Edsö</i>	
A SCALABLE BAYESIAN INFERENCE ACCELERATOR FOR UNSUPERVISED LEARNING.....	11
<i>Glenn Ko, Yuji Chai, Marco Donato, Paul N. Whatmough, Thierry Tambe, Rob A. Rutenbar, Gu-Yeon Wei, David Brooks</i>	
XUANTIE-910: INNOVATING CLOUD AND EDGE COMPUTING BY RISC-V	25
<i>Chen Chen, Xiaoyan Xiang, Chang Liu, Yunhai Shang, Ren Guo, Dongqi Liu, Yimin Lu, Ziyi Hao, Jiahui Luo, Zhijian Chen, Chunqiang Li, Yu Pu, Jianyi Meng, Xiaolang Yan, Yuan Xie, Xiaoning Qi</i>	
AGILEX™ GENERATION OF INTEL® FPGAS.....	35
<i>Ilya K. Ganusov, Mahesh A. Iyer, Ning Cheng, Alon Meisler</i>	
COMPUTE SUBSTRATE FOR SOFTWARE 2.0	48
<i>Ljubisa Bajić, Jasmina Vasiljević</i>	
XILINX VERSAL™ PREMIUM	64
<i>Martin Voogel, Yohan Frans, Matt Ouellette, Jason Coppens, Sagheer Ahmad, Jaideep Dastidar, Ehab Mohsen, Faisal Dada, Mike Thompson, Ralph Wittig, Trevor Bauer, Gaurav Singh</i>	
THE XE GPU ARCHITECTURE.....	87
<i>David Blythe</i>	
XBOX SERIES X SYSTEM ARCHITECTURE.....	101
<i>Jeff Andrews, Mark Grossman</i>	
NVIDIA A100 GPU: PERFORMANCE & INNOVATION FOR GPU COMPUTING	116
<i>Jack Choquette, Wish Gandhi</i>	
AI RESEARCH AT SCALE-OPPORTUNITIES ON THE ROAD AHEAD.....	138
<i>Dan Belov</i>	
VIDEO OF FRANCES ALLEN TRIBUTE PLACEHOLDER	182
<i>N/A</i>	
HANGUANG 800 NPU – THE ULTIMATE AI INFERENCE SOLUTION FOR DATA CENTERS	226
<i>Yang Jiao, Liang Han, Xin Long</i>	
BAIDU KUNLUN AN AI PROCESSOR FOR DIVERSIFIED WORKLOADS	241
<i>Jian Ouyang, Mijung Noh, Yong Wang, Wei Qi, Yin Ma, Canghai Gu, SoonGon Kim, Ki-il Hong, Wang-Keun Bae, Zhibiao Zhao, Jing Wang, Peng Wu, Xiaozhang Gong, Jiaxin Shi, Hefei Zhu, Xueliang Du</i>	
SILICON PHOTONICS FOR ARTIFICIAL INTELLIGENCE ACCELERATION : HOTCHIPS 32.....	250
<i>Carl Ramey</i>	

SOFTWARE CO-DESIGN FOR THE FIRST WAFER-SCALE PROCESSOR (AND BEYOND).....	263
<i>N/A</i>	
GOOGLE'S TRAINING CHIPS REVEALED: TPUV2 AND TPUV3.....	276
<i>Thomas Norrie, Nishant Patil, Doe Hyun Yoon, George Kurian, Sheng Li, James Laudon, Cliff Young, Norman P. Jouppi, David Patterson</i>	
A 4096-CORE RISC-V CHIPLET ARCHITECTURE FOR ULTRA-EFFICIENT FLOATING-POINT COMPUTING.....	311
<i>Florian Zaruba, Fabian Schuiki, Luca Benini</i>	
AMD NEXT GENERATION 7NM RYZEN™ 4000 APU "RENOIR"	323
<i>Sonu Arora, Dan Bouvier, Chris Weaver</i>	
INSIDE TIGER LAKE: INTEL'S NEXT GENERATION MOBILE CLIENT CPU	338
<i>Xavier Vera</i>	
HIGH-DENSITY MULTI-TENANT BARE-METAL CLOUD WITH MEMORY EXPANSION SOC AND POWER MANAGEMENT	351
<i>Xiantao Zxt, Zhengxiao Zx, Justin Song</i>	
THE FUNGIBLE DPU™: A NEW CATEGORY OF MICROPROCESSOR FOR THE DATA-CENTRIC ERA : HOT CHIPS 2020	360
<i>N/A</i>	
DISTRIBUTED SERVICES ARCHITECTURE.....	373
<i>Francis Matus</i>	
INTEL TOFINO2 – A 12.9TBPS P4-PROGRAMMABLE ETHERNET SWITCH.....	382
<i>Anurag Agrawal, Changhoon Kim</i>	
IBM'S POWER10 PROCESSOR	398
<i>William Starke, Brian Thompto</i>	
THE IBM Z15 PROCESSOR CHIP SET	420
<i>Anthony Saporito</i>	
NEW 3RD GEN INTEL® XEON® SCALABLE PROCESSOR (CODENAME: ICE LAKE-SP).....	429
<i>Irma Esmer Papazian</i>	
THUNDERX3 NEXT-GENERATION ARM-BASED SERVER.....	440
<i>Rabin Sugumar</i>	

Author Index