Development of the ARM ecosystem for artificial intelligence, cloud and high performance computing

Valeriy Cherepennikov
Vice-president of
Nizhny Novgorod
Research Center
Contents

1. The Landscape
2. Early signs
3. Ecosystem – long way to go
   ✓ Key resources
   ✓ Russia
   ✓ Industry alliances
4. Huawei products
   ✓ Kunpeng
   ✓ Ascend
The Landscape

- Technology
- Economy
- Politics
- And a bit of Philosophy
Early signs

• **Fujitsu** and RIKEN Take First Place Worldwide in TOP500, HPCG, and HPL-AI with Supercomputer Fugaku

• By the end of this year, **Apple** will have a **MacBook Pro** on sale that is **powered by an ARM-based processor of Apple’s own design**.

• Is NVidia buying ARM?
Supported OS

- AltServer
- SimlyLinux
- Альт Рабочая станция
- Альт Образование
- RedOS (server)
- AstraLinux Special Edition
- CentOS
- Debian
- Ubuntu
- SLES
- openEuler
- NeoKylin
- Deepin
- Linx
- Euler
- Kylin
- BC-Linux

Developer tools

- GCC
- CLANG
- Perf
- GDB
- Git
- Make
- Cmake
- FreePascal
- OpenJDK
- HuaweiJDK

http://ic-openlabs.huawei.com/openlab/#/unioncompaty
Russia Kunpeng Ecosystem in 2020

Kunpeng

Virtualization
- AccentOS (IaaS)
- Horizon-VS

OS
- Ubuntu
- Red OS

Database
- Red DB
- PostgreSQL
- MySQL
- Mango DB

Middleware
- RabbitMQ
- ActiveMQ
- Memcached
- Redis
- Squid
- Varnish
- Lighttpd

Big Data
- Hadoop

Webservice
- Nginx
- JBoss
- Apache
- Tomcat
- HAPROXY
- LVS

Video
- Ceph

SUSE enterprise storage
- SUSE

Applications
- Video Most (Video Analysis)
- Mobility Lab (Enterprise Mobility)
- Posit Tech (Security)
- Infotecs (Security)
- Lampyris (Supply Chain)
- MyOffice
- 1C

SDS
- Aerodisk
- Gluster
- MiniO

AccentOS (VDI, IoT)
- AccentOS
- TSS (Security)
- Bars Group (BI)
- CFT (Bank)
- Diasoft (Security)
- Galaktika (ERP, CRM)
- SKB Kontur
ARM ecosystem needs center of gravity for

- Better scaling
- Resource saving
- Enabling acceleration (5 years instead of 20)
- Efficiency grows nonlinearly(!)
- Influence point
Hi1620 Specifications Overview

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU core</td>
<td>Up to 64 ARMv 8.2 cores, 3.0 GHz, 48-bit physical address</td>
</tr>
<tr>
<td></td>
<td>4 issue OoO superscalar design</td>
</tr>
<tr>
<td></td>
<td>64 KB L1 I Cache and 64 KB L1 D cache</td>
</tr>
<tr>
<td>L2 cache</td>
<td>512 KB private per core, 24 MB total</td>
</tr>
<tr>
<td>L3 cache</td>
<td>48 MB shared for all (1 MB/core), Partitioned</td>
</tr>
<tr>
<td>Memory</td>
<td>8-channel DDR4-2400/2666/2933/3200</td>
</tr>
<tr>
<td></td>
<td>16 ranks/channel, 1DPC and 2DPC configurations</td>
</tr>
<tr>
<td></td>
<td>x4/x8 support</td>
</tr>
<tr>
<td></td>
<td>ECC, SDDC, DDDC</td>
</tr>
<tr>
<td>PCIe</td>
<td>40 lanes of PCIe Gen4.0 16x</td>
</tr>
<tr>
<td>Integrated I/O</td>
<td>8 lanes of ETH, Combo MACs, supporting 2 x 100GE, 2 x 40GE, 8 x 25GE/10GE, 10 x GE, supporting SR-IOV</td>
</tr>
<tr>
<td></td>
<td>RoCEv2/RoCEv1</td>
</tr>
<tr>
<td></td>
<td>x4 USB</td>
</tr>
<tr>
<td></td>
<td>3.0 x8 SAS 3.0 x2 SATA 3.0</td>
</tr>
<tr>
<td>Crypto engine</td>
<td>AES, DES/3DES, MD5, SHA1, SHA2, HMAC, CMAC</td>
</tr>
<tr>
<td></td>
<td>Up to 100 Gbit/s</td>
</tr>
<tr>
<td>Compression</td>
<td>GZIP, LZS, LZ4</td>
</tr>
<tr>
<td></td>
<td>Up to 40 Gbit/s (compress)/100 Gbit/s (decompression)</td>
</tr>
<tr>
<td>RAID</td>
<td>RAID5/6, DIF, XOR, PQ acceleration</td>
</tr>
<tr>
<td>CCIX</td>
<td>Cache coherency interface for accelerator, like Xilinx FPGA</td>
</tr>
<tr>
<td>Scale-up</td>
<td>Coherent SMP interface for 2P/4P</td>
</tr>
<tr>
<td></td>
<td>3*240Gbps bandwidth</td>
</tr>
<tr>
<td>Power</td>
<td>TDP ~150 W (48C 2.6 GHz)</td>
</tr>
</tbody>
</table>
**Application enablement**: full-pipeline services (ModelArts), layered APIs, and pre-integrated solution

**MindSpore**: unified training and inference framework for device-edge-cloud (independent or collaborative)

**CANN**: chip operator library and highly automated operator development tools

**Ascend**: a series of AI IPs and chips with unified and scalable architecture

**Atlas**: AI infrastructure in various form factors, powered by Huawei Ascend AI processors for device-edge-cloud all scenarios
Thank you.

Bring digital to every person, home and organization for a fully connected, intelligent world.

Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.