Accelerating Image Processing for Object Storage

James Zhang, Product Manager, CTAccel
Sean Gardner, Sr. Product Manager, Xilinx
We Are Xilinx
Founded in 1984

$2.54B FY18 Revenue
56% Market Share
4,000+ Employees*
20,000+ Customers

*~3,000 Engineers

Headquarters
Sales & Support
Research & Development
Manufacturing

>> 2
Xilinx Innovation & Industry Firsts

- World’s First Fabless Semiconductor Company
- World’s First FPGA
- First integrated processor in an FPGA
- First HW/SW Programmable SoC
- World’s First 2.5D IC FPGA
- First ASIC-Strength Design Suite
- First Multi-Processing SoC (MPSoC)
- SDx Development Environments
- First RFSoC
- Acceleration Stacks & Frameworks
<table>
<thead>
<tr>
<th>Performance vs. Server CPUs</th>
<th>Real FaaS Use Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Machine Learning Inference</strong></td>
<td></td>
</tr>
<tr>
<td>Speech Recognition</td>
<td></td>
</tr>
<tr>
<td><strong>Genomics</strong></td>
<td></td>
</tr>
<tr>
<td>20 min vs. 33 hours for whole genome analysis</td>
<td></td>
</tr>
<tr>
<td><strong>Video Streaming</strong></td>
<td></td>
</tr>
<tr>
<td>Frame rate for HEVC encoding</td>
<td></td>
</tr>
<tr>
<td><strong>Big Data Analytics</strong></td>
<td></td>
</tr>
<tr>
<td>40 min vs. 60 hours for logfile query</td>
<td></td>
</tr>
</tbody>
</table>

Our Mission: Building the Adaptable Intelligent World
**Storage Acceleration supercharges compute**

<table>
<thead>
<tr>
<th>Applications</th>
<th>Genomics</th>
<th>Security</th>
<th>Machine Learning</th>
<th>Video Transcoding</th>
<th>Big Data Analytics</th>
<th>Financial Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics</td>
<td>![DNA]</td>
<td></td>
<td><img src="brain.png" alt="Brain" /></td>
<td><img src="video.png" alt="Video" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>![Shield]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Transcoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Data Analytics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial Analytics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Compute Acceleration**

- Machine Learning
- Video & Image Processing
- Data Analytics
- Genomics

**Inline Storage Acceleration**

---

**SDx™**
Software-Defined Development Environments

**Solution Development**

**Cloud, Enterprise to Edge-Computing**
CTAccel

- Founded in 2016, in HK and Shenzhen
- Main staffs from CUHK, HKUST, FDU, CAS, etc.
- Focusing on FPGA data center acceleration computing technology
- Core technology has obtained U.S. patents.
Market Demand and Challenges

More Data
- Modern electronics has higher resolution in capture and display
- Users generate more image and video everyday

Better Quality
- Users crave better viewing experience

Faster Access
- Customers demand instant access to the resource

Challenges:
- Huge consumption of computational and storage resource
- Server and storage performance IS NOT KEEPING PACE

Data storage supply and demand worldwide, from 2009 to 2020 (in exabytes)*
CTAccel solutions to accelerating Image Processing for Object Storage

CTAccel Image Processing (CIP) effectively accelerates:
- Thumbnail Generation/Transcoding
- Image processing (sharpen/color filter)

CIP includes the following FPGA-based accelerated functions:
- Decoder: JPEG, Lepton, HEIC
- Pixel processing: Resizing/Crop
- Encoder: JPEG, WebP, Lepton

Software compatibility:
- OpenCV
- ImageMagicks

End-user
- Mobile phone/PAD
- Camera
- ... PC

CIP
- JPEG 2 Webp
- JPEG 2 Lepton
- ... Resize

Applications
- JPEG decode
- JPEG thumbnail
- Sharpen
- Main color
- Watermark
- Brightness-Contrast
- HEIC decode
- Lepton Encode
- WebP Encode
**Customer**
Famous O2O service provider in China

**Scenario**
Transcoding from JPEG to WebP

**Advantages**

**High performance**
- 1 server with one CIP accelerator has the equivalent computing capability with 3 servers without CIP

**TCO reduction**
- Reduce TCO by 50% at least

**Improve customer experience**
- Reduce latency by 64%
- Deterministic timing: the same low latency at full-load and no-load

---

<table>
<thead>
<tr>
<th>Throughput (MB/s)</th>
<th>Latency (ms)</th>
<th>Power Consumption (Watt)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FPGA</strong></td>
<td><strong>CPU</strong></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>27.5</td>
<td>20</td>
</tr>
<tr>
<td>441</td>
<td>1596</td>
<td>80</td>
</tr>
</tbody>
</table>

5x **Increase concurrency**
4x **Reduce latency**
4x **Save energy**
Take-Aways

- **High Throughput**: Enables better cost efficiency in Machines
- **Low Latency**: Provides better user experience
- **Energy Saving**: Makes the Internet Data Centre greener
- **Simple Maintenance**: Remote upgrade on module with worry-free maintenance
Adaptable.
Intelligent.