Main SoC and XBOX One Kinect

John Sell
Patrick O’Connor
Xbox One

Diagram showing the components of the Xbox One, including:
- Game Controller
- Xbox One Kinect
- South Bridge
- Main SoC
- 8 GB DRAM

Key components:
- HDMI In
- Ethernet
- Gigabit Ethernet MAC & PHY
- HDMI SPDIF Out
- USB
- AV Out
- PCIe
- SATA 2
- eMMC4.5
- BluRay
- 500 GB HD
- 8 GB Flash
Main SoC

- 363 mm²
- 28 nm TSMC HP
- 5 billion transistors
- 47 Mbytes of storage on chip
- Power islands and clock gating to 2.5% of full power
SoC Components

15 special purpose processors offload CPU & GPU

- PCIe
- Audio Processors
- Audio DMA
- AV Out Rsz Cmpst
- AV In
- Video Encode
- Video Decode
- Swizzle/LZ Encode
- Swz LZ/MJPG Dcd
- Swizzle Copy
- Swizzle Copy
- Graphics Core

CPU

Any DRAM data can be coherent with CPU caches

CPU-Cache-Coherent Memory Access

30 GB/sec coherent BW

CPU, GPU, special processors, and IO share memory via host-guest MMUs and synchronized page tables

68 GB/sec peak BW
4 x 64 bit

DRAM CNTR

109 GB/sec min
204 GB/sec peak BW
4 x 256 bit read & write

8 MB  8 MB  8 MB  8 MB

2 GB DDR3
2 GB DDR3
2 GB DDR3
2 GB DDR3
CPU

2 MB 16-Way Shared L2 Cache
(8-Way L1 DC, 2-Way L1 IC per Core)

North Bridge
Audio Processors

- Audio codec and signal processing optimized vector and scalar cores
- Two 128-bit SIMD FP vector cores, 15.4 GFLOPs total
- Specialized hardware engines equivalent to 18 G Ops
SoC Summary

- High performance, but power efficient, and very low power modes
- AV in and out media hub
- Specialized audio, graphics, and video processors offload CPU and graphics core
- CPU, GPU, specialized processors, and IO share memory via host-guest MMUs with synchronized page tables
- High bandwidth CPU cache coherency
- 200+ GB/second power efficient memory system balanced to CPU, GPU, specialized processors, and IO requirements
- DX11.1+ graphics core with custom graphics and compute command processors to offload CPU and improve GPGPU
Different Needs – Different Solution

Xbox One Kinect
# A Next Generation Kinect Experience

<table>
<thead>
<tr>
<th>User Experience</th>
<th>Technical Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent performance, kids &amp; adults</td>
<td>Depth resolution of ~1%</td>
</tr>
<tr>
<td>– child’s wrist</td>
<td>Minimum SW-detectable object &lt;2.5cm</td>
</tr>
<tr>
<td>Playspace Flexibility</td>
<td>Full spec operation 0.8m – 4.2m range</td>
</tr>
<tr>
<td>Small or Large rooms, Multiple players</td>
<td>70° Horizontal optical field of view</td>
</tr>
<tr>
<td>Accurate, Responsive, Consistent</td>
<td>&lt;2% accuracy, &lt;20ms latency to SW,</td>
</tr>
<tr>
<td>User Experience</td>
<td>&lt;14ms total exposure time</td>
</tr>
<tr>
<td>Lighting Independent</td>
<td>Depth performance independent of</td>
</tr>
<tr>
<td></td>
<td>room lighting</td>
</tr>
</tbody>
</table>

→ **Microsoft designed a new highly-customized image sensor based on Time of Flight Technology to meet this need**
A Differential Pixel

For an XBOX One Kinect Depth Sensor

Differential Pixel

- \((A+B)\) gives the ambient (room) lighting (‘common mode’) – ‘normal’ grey scale image
- \((A-B)\) gives phase (depth) information after an arctan calculation – depth image
- \(\sqrt{(A-B)^2}\) is the ‘Active’ image – A grey scale image independent of ambient lighting

→ Depth & Image performance is per-pixel, defined by optical & electrical parameters
Sensor System Block Diagram

Time Of Flight Technology can deliver the performance needed
Near & Far, Reflective & Non-Reflective

High dynamic range is required to enable robust multi-player capability

- Performance ~ SNR → Multiple players
  - *Must meet minimum SNR at all points in the room simultaneously*

- Cannot use normal photographic tricks of Aperture / Exposure
- Need a dynamic range of ~2500x
  - *Use 2 or more shutter times, choose which gives the best image*

---

Light Falls Off with $1/R^2$

- Laser Diode & Optics
- Sensor Lens

- **Near:** 0.8m
  - Dark Shirt ~10%
  - Bright Shirt ~95%

- **Far:** 4.2m
Need High Z-resolution over long range

Phase ‘wraps’ at 360° – 3.75m for 80MHz – Must we use a lower frequency?

- Each frequency gives a wrapped estimate of distance
- With the combination, find the unwrapped distance
- Allows high frequency (good for resolution) over long distance
Room Light Conditions Distorts 2D Images

Active IR Provides Consistently Lit Images

- Example: A side-lit face has shadowing that confuses SW
- The Active image is front-lit and insensitive to room lighting

A Regular Color Camera  
Active, same conditions
Real Time Depth Captured @30fps
Thank You!

Microsoft Silicon Development Team