TEGRA X1: NVIDIA’S NEW MOBILE SUPER-CHIP

Steve Molnar
HPG2015 Hot3D
2014: Tegra K1

NVIDIA’S FIRST SUPERCHIP: Console features and performance on a mobile processor

- Kepler GPU
- 192 CUDA cores
GPU TRAJECTORIES: DISCRETE AND MOBILE

Advances

GPU ARCHITECTURE

MOBILE ARCHITECTURE

Tesla

Fermi

Kepler

Tegra 3

Tegra 4

Tegra K1

Maxwell

Tegra X1
TEGRA X1: NVIDIA’S NEW MOBILE SUPERCHIP

256-core Maxwell GPU | 8-core 64bit ARM CPU | 60 fps 4K Video (H.265/VP9) | 20nm
RAISING THE BAR. AGAIN.

RELATIVE PERFORMANCE

- **GFXBench 3.0 Manhattan 1080p offscreen**
- **GFXBench 2.7 TRex 1080p offscreen**
- **3Dmark 1.3 Icestorm Unlimited**
- **BasemarkX 1.1**
2X ENERGY EFFICIENT

PERFORMANCE

GPU POWER
GFXBench 3.0 Manhattan (1080p, Offscreen)

Tegra X1

Tegra K1
WORLD’S 1ST TERAFLOPS MOBILE PROCESSOR

GRAPHICS HORSEPOWER (GFLOPS)

TIME

0 200 400 600 800 1000 1200

Tegra 2  Tegra 3  Tegra 4  Tegra K1  Tegra X1

ASCI Red
1st Teraflop Supercomputer
THE MOST ADVANCED GPU. AGAIN.

<table>
<thead>
<tr>
<th></th>
<th>TEGRA X1</th>
<th>GEFORCE GTX 980</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenGL ES 3.1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AEP</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OpenGL 4.5</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DX12</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CUDA 6.0</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
1 Processor and memory architecture
2 Maxwell GPU advances
3 Computer vision
4 4K video
5 Shield Android TV / console
6 Demos
TEGRA X1 - OVERVIEW

**GRAPHICS**
Maxwell GeForce - World’s Fastest GPU
2 x SMM units, DX-next, OpenGL 4.4

**CPU**
Octo-Core 64b ARM v8 CPU Complex
4xCA57 Atlas/2MB L2; 4xCA53 Apollo/512KB L2

**MEMORY**
64b / Dual-Quad Channel Memory Interface
LPDDR4-3200, LPDDR3E-1866, DDR3L-1866

**VIDEO**
4K x 2K Encode and Decode
H.264, H.265, VP8, VP9 (dec-only)

**POWER**
Low Power
20nm, HW Offloads, Isolated Pwr Rails, PRISM

**DISPLAY**
4K x 2K 24b @60Hz, 1080p @120Hz
DSI 2x4, eDP, High Speed HDMI 2.0, DP

**IMAGING**
Full Quad Camera imaging, Dual ISP 650Mp/s
Maxwell 16fp imaging GPGPU, HW AO-HDR

**Mobile I/O**
Designed for mobile
e.MMC5.x, USB3.0/2.0/HSIC, SD/SDIO 3.0, CSI-2
TEGRA X1 CPU CONFIGURATION

4 HIGH PERFORMANCE A57 BIG CORES
- 2MB L2 cache
- 48KB L1 instruction cache
- 32KB L1 data cache

4 HIGH EFFICIENCY A53 LITTLE CORES
- 512KB L2 cache
- 32KB L1 instruction cache
- 32KB L1 data cache
2X CPU EFFICIENCY

SPECint score estimates. Both Tegra X1 and Exynos 5433 have ARM A57 cores, and are manufactured using 20nm process. Measured estimates is using SPEC CPU2000. Tegra X1 is measured on NVIDIA reference platform and Exynos 5433 is measured on Samsung Note 4.
LPDDR4

- Memory system capabilities dictated by DRAM technology
- Internal DRAM array structure and speeds change little over generations (~200 MHz internal clock)
- IO speed increasing exponentially over last generations
  - LPDDR 400 Mbit/sec
  - LPDDR2 1066 Mbit/sec
  - LPDDR3 1866 Mbit/sec
  - LPDDR4 3200 Mbit/sec
- LPDDR4 runs at 40% less power than LPDDR3
LPDDR4

- 32-bit channel becomes pair of 16-bit channels with higher burst length (same 32B transfer atom)
- Low Voltage Swing Terminated Logic (LVSTL) I/O interface
- Doubling of the interface speed requires additional training, clocking changes

**Diagram:**

- Tegra K1
  - LPDDR3 32-bit
  - LPDDR3 32-bit
  - 2x 32 bits @ 1866 Mbit/sec

- Tegra X1
  - LPDDR4 32-bit
  - LPDDR4 32-bit
  - 4x 16 bits @ 3200 Mbit/sec

- LPDDR4 32-bit
  - 14.9 GB/sec

- LPDDR4 32-bit
  - 25.6 GB/sec
TEGRA X1 COMPRESSION ADVANCES

NEW END-TO-END COMPRESSION

3RD GEN DELTA COLOR COMPRESSION

System Memory

Compressed Traffic

Maxwell GPU

Compressed Traffic

Display Controller

Display
MEMORY

COMPRESSION

Original Image

Image taken on Half Life 2
MEMORY COMPRESSION

3rd Generation
Delta Color Compression
MEMORY BANDWIDTH REDUCTION

- Enhanced Compression Algorithms
- Enhanced Caching Effectiveness
- New End-to-End Compression
# Tegra X1 DRAM Energy Efficiency

**GFXBench 3.0 Manhattan**

<table>
<thead>
<tr>
<th></th>
<th>Tegra K1</th>
<th>Tegra X1</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bytes per frame efficiency</td>
<td>1x</td>
<td>1.5x</td>
<td>Bandwidth reduction due to Maxwell architecture + end-to-end compression</td>
</tr>
<tr>
<td>Dram pJ/byte efficiency</td>
<td>1x</td>
<td>1.4x</td>
<td>pJ/Byte reduction due to LPDDR4 vs LPDDR3</td>
</tr>
<tr>
<td>DRAM energy efficiency</td>
<td>1x</td>
<td>2.1x</td>
<td></td>
</tr>
</tbody>
</table>
“...faster, less power hungry, and quieter... once again NVIDIA has landed the technical trifecta”

– Anandtech
TEGRA X1 MAXWELL GPU

- 2x performance vs Tegra K1
- 2x perf/watt vs Tegra K1
- 2 SM
- 256 CUDA Cores
- 2 Geometry Units
- 16 Texture Units
- 16 ROP Units
- Maxwell Memory Architecture
- 64-bit LPDDR4
TEGRA X1
MAXWELL SM

- 2x Energy Efficiency vs Tegra K1
- 40% higher perf per core
- Improved scheduler
- New datapath organization
- Enhanced memory hierarchy
- Double speed FP16
NEXT-GENERATION GRAPHICS FEATURES

- Voxel based Global Illumination
- Tiled Resources
- Multi-Projection Acceleration
- Advanced Sampling
- Raster Ordered View
- Conservative Raster
- Tiled Resources
GLOBAL ILLUMINATION
THE GRAND CHALLENGE OF REAL-TIME COMPUTER GRAPHICS
VXGI
VOXEL GLOBAL ILLUMINATION

Dynamic

1-bounce Indirect Diffuse, Specular, Reflections, Area Lights

Available for UE4 and other major engines (Q4)
VR DIRECT
Auto Stereo
Low Latency
Auto Asynchronous Warp
TEGRA X1
DEEP LEARNING AND COMPUTER VISION
BUILT FOR COMPUTER VISION

- High computational horsepower
- Low power consumption
- General purpose programmability
- Multi-camera support
- High performance HD video and image processing
GPUs Revolutionizing Deep Learning Across Industries

“In 2009, Andrew Ng and a team at Stanford realized that GPU chips could run neural networks in parallel. Today neural nets running on GPUs are routinely used by cloud-enabled companies such as Facebook to identify your friends in photos or, in the case of Netflix, to make reliable recommendations for its more than 50 million subscribers.”

*WIRED*
THE PERCEPTRON — THE SIMPLEST MODEL

PERCEPTRON

\[ \sum w_i \cdot x_i \quad f() \]

Output
THE PERCEPTRON — THE SIMPLEST MODEL

PERCEPTRON

\[ \sum w_i x_i \rightarrow f(\cdot) \]

\[ \text{Output} \]

Activation Functions:

- Step Function
- Sign Function
- Sigmoid Function

Learning:

\[ y^{(t)} = f \left( \sum_i w_i^{(t)} x_i^{(t)} \right) \]

\[ \Delta w_i^{(t)} = \varepsilon (d^{(t)} - y^{(t)}) x_i^{(t)} \]

\[ w_i^{(t+1)} = w_i^{(t)} + \Delta w_i^{(t)} \]
HOW A TRAINED MACHINE SEES

TEGRA X1 FOR DEEP LEARNING

OverFeat for Classification

![Graph comparing Tegra K1 and Tegra X1 for OverFeat classification across layers lay0 to lay5.](image-url)
CARS THAT SEE BETTER

NEURAL NET MODEL

“CORNER CASES” TO REFINE MODEL

NVIDIA TESLA SUPERCOMPUTER FOR TRAINING

NVIDIA TEGRA SUPERCHIP FOR CLASSIFICATION
TEGRA X1
4K VIDEO
END-TO-END 4K PIPELINE

WiFi Interface >4K ISP 4K Video Encode 4K Video Decode >4K JPG Encode

Storage Interface 4K Compositor GPU 4K Internal Panel 4K External Display

60 fps 4K STREAMING
## THE FULL 4K EXPERIENCE

<table>
<thead>
<tr>
<th>Feature</th>
<th>Tegra X1</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>60fps 4K VP9 &amp; H.265</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>10-bit 4K H.265</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
## Tegra X1 — Video, Memory, Display Specs

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>Tegra K1</th>
<th>Tegra X1</th>
<th>Tegra X1 vs Tegra K1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Video Decode</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP9</td>
<td></td>
<td>2160p 60</td>
<td>New 2X</td>
</tr>
<tr>
<td>VP8</td>
<td>2160p 30</td>
<td>2160p 60</td>
<td>2X</td>
</tr>
<tr>
<td>H.264</td>
<td>2160p 30</td>
<td>2160p 60</td>
<td>New 5X</td>
</tr>
<tr>
<td>H.265</td>
<td>120MPix/s</td>
<td>2160p 60 (10-bit) 600MPix/s</td>
<td></td>
</tr>
<tr>
<td>JPEG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Video Encode</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP8</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>H.264</td>
<td>✔</td>
<td>✔</td>
<td>New 5X</td>
</tr>
<tr>
<td>H.265</td>
<td></td>
<td>2160p 30 600MPix/s</td>
<td></td>
</tr>
<tr>
<td>JPEG</td>
<td>120MPix/s</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Video Image Compositor</strong></td>
<td>800MPix/s</td>
<td>4000MPix/s</td>
<td>5X</td>
</tr>
<tr>
<td><strong>Internal Display</strong></td>
<td>3200x2000 @60Hz</td>
<td>2160p @60Hz VESA DSC Compression</td>
<td>1.4X New</td>
</tr>
<tr>
<td><strong>External Display</strong></td>
<td>2160p 30 HDMI 1.4b HDCP 1.4</td>
<td>2160p 60 HDMI 2.0 HDCP 2.2</td>
<td>2X New  New</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>LPDDR3 14.9GB/s (LP3)</td>
<td>LPDDR3, LPDDR4 25.6GB/s (LP4)</td>
<td>New 1.7X</td>
</tr>
</tbody>
</table>

Note: New functionalities are marked in green, indicating improvements or new capabilities compared to Tegra K1.
TEGRA X1
SHIELD CONSOLE / ANDROID TV
BATTLE FOR THE LIVING ROOM

HBO NOW STAND-ALONE SERVICE SHAKES UP TV INDUSTRY

Los Angeles Times

Apple is Out to Blow Up the Cable TV Model

Bloomberg

The media industry is racing toward an Internet-TV future at a breathtaking pace

WALL STREET JOURNAL
“We are only at the beginning of this app driven evolution.”

— Chris Louie
VP @ Nielsen, Nov 2014
ANDROID TV MOMENTUM

# of Apps
1 Year Later

Recently Announced

HBO NOW™
HBO GO™
FXNOW
CBS
FOX NOW™
CBS SPORTS
CBS NEWS
Vudu HD Movies
Live Channels
SHIELD

“The flagship of Android TV”

TEGRA X1: MOST ADVANCED MOBILE PROCESSOR

CONNECTED TO A WORLD OF CONTENT IN 4K

AMAZING GAMES
“SHIELD is powered by Tegra X1, meaning it massively overpowers rivals like Fire TV and Apple TV.”

— Tech Radar

![SHIELD's chip](image)

<table>
<thead>
<tr>
<th>RELATIVE PERFORMANCE</th>
<th>Apple TV</th>
<th>Roku 3</th>
<th>Fire TV</th>
<th>SHIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>35</td>
</tr>
</tbody>
</table>

Data: GFXBench T-Rex off-screen
WITH 4K MOVIES AND SHOWS.
WITH Amazing GAMES.
WITH a world of apps.
The FLAGSHIP OF ANDROID TV

- **Best Performance**
  - Tegra X1, 256 CORE NVIDIA GPU, 64-BIT CPU, 3GB

- **Always Up-to-Date**
  - NVIDIA OTAs

- **Voice Search**

- **Private Audio**

- **Most Storage Options**
  - MicroSD slot
  - USB 3.0 x2

- **Fastest Connectivity**
  - Gigabit Ethernet
  - 802.11ac Wi-Fi
1 Processor and memory architecture
2 Maxwell GPU advances
3 Computer vision
4 4K video
5 Shield Android TV / console
6 Demos