Design of A Low Power SoC Testchip for Wearables and IoTs

May Wu, Ravi Iyer, Yatin Hoskote, Steven Zhang, Julio Zamora, German Fabila, Ilya Klotchkov, Mukesh Bhartiya
August, 2015
Legal Notices and Disclaimers

Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer. Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice. Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

Copyright © 2015 Intel Corporation. All rights reserved. Intel, the Intel logo, Core, and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.
Wearables and IoT: Redefining computing

- >50 Billion
- 50 Billion
- 200 Billion
- 75 Billion

Morgan Stanley
SIGHT and SOUND: The New Frontiers
INTEGRATED INTELLIGENCE:
Always Listening. Always Watching.
We are limited by high power consumption

*Power is the greatest barrier for intelligent “always on” devices*
Power consumption is nowhere near where we need to be.

<table>
<thead>
<tr>
<th>Component</th>
<th>Power Consumption (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>~100mW</td>
</tr>
<tr>
<td>Processing</td>
<td>200~400mW</td>
</tr>
<tr>
<td>Transmission</td>
<td>200~250mW</td>
</tr>
<tr>
<td>Storage</td>
<td>Tens of mW</td>
</tr>
<tr>
<td>Display</td>
<td>Tens~Hundreds of mW</td>
</tr>
</tbody>
</table>

Wearable Device with A/V Capabilities Today

1~2 Hours Battery Life

- Lower MPixels Type
- Looxcie measure Estimate
- Best WiFi on publication
Rethinking enabling technologies for *always on* IoT and Wearable devices.
Better Ways for Data Capturing and Transmission

- Pushing intelligence close to sensing side
- Threshold / Buffer
- Activity triggers
- Context-Ware Encoding
- Adaptive and cooperative communication
Intelligent IoT and Wearables Demand a Better SoC
Intel Labs: Testchip for Always-On Devices

- Introduced in Q1 of 2015, undergoing further development
- 14nm Intel Process
- From ~2mW keyword recognition to few 10s mW A/V processing
Key Features:
Testchip for Always-On Devices

- Always-Watching: with a Vision Processing Engine (e.g. Gesture, Scene Detect)
- Always-Listening: Voice Activity Detect; Short phrase(s) Recognition; Speaker ID
- Low Power Embedded Communication
- Light-Weight Security Framework and Processing
High Level Architecture Overview of the Testchip

- **Embedded Comm**
  - Antenna
  - PDM Mics

- **Always-Listening**
  - PDM Mic IF
  - Dynamic Noise Acoustic FE
  - Voice Activity Detect
  - Data Packer
  - Short Phrase Recognition
  - Speaker ID

- **Always-Watching**
  - Intra-Frame Encoding
  - Image Sensor IF Low Pwr ISP
  - Vision Processing Engine

- **System Fabric**
  - A/V Fabric
  - Small Core with Light Signal Processing

- **Shared Memory**
  - Shared Mem Management
  - Security Access Ctrl

- **Small Host Processor**

- **Microcontroller Unit (MCU)**

- **Clock Unit**

- **Power Management Unit**

- **GPIO, SPI (Sensors)**
  - Display, Speaker, Sensors; UARTs; SDIO

- **Crystals or XOs**

- **Optional External Flash or Pseudo SRAM for test**

- **Today’s Intro Focus**
  - Always - Listening
  - Always - Watching
Our Design Strategy for “Always-Watching” Devices

Two Key Advances

1. Vision-Driven LP Imaging
   - Very aggressive image sensor power gating
   - Race to halt
   - Light-detect assisted auto-exposure processing
   - Intra-frame and data analysis driven encoding

2. Optimized Common Neural Network Processing for Multiple Applications
   - Shifted Neural Network for the classification
     - Shift operations, fixed point, approx sigmoid/hyperbolic tangent functions, etc
   - Memory optimized convolutional layers for vision recognition feature extraction
Always-Watching – Multiple Applications Common Solutions

**Application**

- Tailored Preprocessing
  - Sobel + Convolutional layers
  - Haar or convolutional layers
  - Otsu + Convolutional layers
- HOG

**Neural Network**

- One NN engine shared
- Changing weights and topology

**Output**

- Animal crossing
- Face at (50,30)
- Text: “7” (Input)
- Action: “Close Application” (Command)
Always-Watching - Vision Processing

Frame Analysis
What's interesting?

Quick Evaluation
Should it be evaluated?

Feature Extraction
- Recognize digits
- Detect faces
- Hand Gesture

Identify what is interesting to reduce NN evaluations
<table>
<thead>
<tr>
<th>Processing Phase</th>
<th>Processing Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Analysis and Segmentation</td>
<td>Small Core with Lite Signal Processing Acceleration</td>
</tr>
<tr>
<td>Quick Evaluation</td>
<td>Small Core with Lite Signal Processing Acceleration</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>• Small core with Lite Signal Processing Acceleration</td>
</tr>
<tr>
<td></td>
<td>• CNN Acceleration IPs for feature extraction</td>
</tr>
<tr>
<td>Classification</td>
<td>Highly optimized Shifted NN</td>
</tr>
</tbody>
</table>
Always-Watching Vision Processing: Hand Gesture Experiment

- Histogram and K-means calculation
- Quick Analysis
- Subsampling
- CNN Evaluation

2 fps or lower; QVGA or lower; YUV; Distance 20cm~1m; Response time 200ms or lower; Recognition processing power <1mW to several mW
Always Listening Speech Processing Pipeline: Design for Power Reduction

End-to-End H/SW Partitioning with Low power Always-Listening on Device

Voice Activity Detector, Enhancement
Acoustic Front-end, Keyword recognition
Command & Control Recognition
Large Vocabulary Continuous Speech Recognition
Natural Language Processing
Text-to-Speech

On the device
In the cloud
Always Listening Block Overview

Key Advances

- Noise reduction tailored for speech recognition
- 0.1~0.2mW DNR+AFE+VAD
- Single digit mW for tens of command and control recognition, with accurate VAD support

- **2 audio channels** can be on/off independently
- **Audio sampled** as 16-bit@16kHz/8KHz for voice activity detect and short phrase recognition
- **Acoustic Front End and Voice Activity Detect process** 1 audio channel in 160-sample frames, 100 frames/s, and produce frames of 12 features + voice&quiet flags
The Testchip
On Intel 14nm

4mmX8mm
Shared Die
(low utilization)

Embedded Comm
(Including RF and
Other Analog
Circuit)

Host Core
SubSystem

Audio/Speech

MCU & macros

Clock/
Ring Oscillator

Other testchip
Projects shared
the same
Die
Area

Shared MEM

Vision&
Small core with
signal processing

Io family
## Test 1: Always-Listening VAD and Keyword Recognition

<table>
<thead>
<tr>
<th></th>
<th>Voice Activity Detect Stage</th>
<th>Keyword Recognition Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>With 1 digital Mic at low performance mode (customized Mic)</td>
<td>~1mW (including Mic)</td>
<td>~1.9mW (including Mic)</td>
</tr>
<tr>
<td>With 1 digital Mic at standard performance mode (regular Mic)</td>
<td>~1.5mW (including Mic)</td>
<td>~2.5mW (including Mic)</td>
</tr>
</tbody>
</table>
Test 2: On Chip \(~22\text{mW}\) A/V capturing, Hand posture and Always-listening

<table>
<thead>
<tr>
<th>Functions</th>
<th>Pwr (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio/Speech</td>
<td>0.9\text{~}0.946</td>
</tr>
<tr>
<td>Imaging</td>
<td>1.65\text{~}3.322</td>
</tr>
<tr>
<td>Vision Recognition</td>
<td>5.478\text{~}5.566</td>
</tr>
<tr>
<td>Host Core with Memory</td>
<td>5.082\text{~}7.04</td>
</tr>
<tr>
<td>Shared Memory</td>
<td>3.586\text{~}3.674</td>
</tr>
<tr>
<td>Fabric&amp;Peri</td>
<td>2.002</td>
</tr>
</tbody>
</table>
Experiment Platform and Usage Examples

Gesture-Based Control

Digit Recognition Example

Speech /Vision
Integrated Pipeline

Gesture Recognition Flow

The Testchip Form Factor Test Board
Immediate Next steps & Longer Term Directions

- Usages trend to require systems to make “human-like” decisions (bots, drones, kids play, …)
- Adaptive vision+speech+sensor capabilities for ULP recognition & understanding (VU/SU)
- Autonomous radio technologies (ULP wideband radio for sensing, wake-up radios, etc)
Drive the Always-On Revolution

IoT and Wearable Usage Tailored Power Efficiency

Reducing data transmitted

New SoC to open unprecedented drops in power consumption