Enabling Embedded AI for Healthcare

Shang-Hung Lin
VeriSilicon
September 2020
Presentation Agenda

• Intro - What do we need?
• Device categories for personal healthcare
• What is required to make these devices happen?
• Device categories - examples
• Challenges
• Key elements of an Enabling SoC
• How VeriSilicon can help?
• Conclusion
VeriSilicon – A SiPaaS Company

What we do

• IP-centric
• Platform-based
• End-to-end turnkey service

What we don’t do

• No fab
• No branded product
  • No NRE investment
  • Limited inventory risk

We call it Silicon Platform as a Service, or SiPaaS®
VeriSilicon IP Portfolio – Scalable STAR IPs for Licensing

- NPU
- GPU
- Video
- Compression/Encryption
- Computer Vision Imaging DSP
- Compute
- ISP
- Audio/Voice
- Display
- FLEXA API

Server Class

Automotive

Tablets

Smartphone

Wearables & IoT

© 2020 VeriSilicon
VeriSilicon is The Leader in Embedded NPU IP Since 2016

- Global Leader in Embedded NPU (Neural Processor Unit)
- First To Introduce Programmable, Scalable, High Performance, Low Power NPU
- First NPU To Support Common Neural Network plus OpenCL
- Leading in number of licensed customers (30+)
- Leading in number of SOC designs engaged (50+)
- Leading in embedded products shipped in mass production
- VIP9000 is applied to wide range of applications in AI Vision, AI Voice, and AI Pixels
- VIP9000 expands into all major market segments
Introduction

• Personal healthcare – why?
  • Comfort, privacy
  • Customized service, 24-7 care
  • Great improvement in quality of life

• What we need
  • Advanced AI enabled personal healthcare devices
  • Sensors, monitors, aides
  • Sensing/sensor fusion, Interpret, Alarm / Act
What Device Categories Can Help Us?

• Non-intrusive
  • Wearable, body contact, consumer devices

• Intrusive
  • Capsule camera, implants under the skin or in the body

• Semi-intrusive
  • Earbud, hearing aid
What Is Needed To Make These Devices Happen?

► **Sense**: MEMS sensors

► **Interpret**: AI engines

► **Act**: DSP engines

► **Communicate**: Wireless

► **Design**: low power / always on
Device Examples Today - Non Intrusive

- Fitness Tracker Ring
- Wrist band
- Chest band
Intrusive Devices For Personal Healthcare

• Capsule Camera
  • Battery operated device; low power inference is crucial
  • Location detection
  • Hemorrhage and lesion detection
• Advanced Hearing aid examples:
  • In 1994, Hitachi’s Dr. Makimoto spoke about his vision – a realtime language translator device
  • In 2005, *The Hitchiker’s Guide to the Galaxy* suggested a solution called Babble Fish
  • At CES 2019, Google assistant does real time translation

• Moving forward, we anticipate advanced, AI-powered in-ear devices in market
  • Active noise treatment
  • AI-enabled DSP audio signal enhancement
  • AI-enabled natural user interface
Challenges

• Always On – lowest power
• Small footprint
• Efficient wireless communication
• Wireless charging and energy harvesting
• Safety and reliability to obtain FDA approval
• Security – hacker safe!
Key Elements of An Enabling SoC

• Sensor Fusion

• Wireless communication

• AI engine

• DSP engine

• Optimized memory subsystem for lowest power operation
AI Functions For True Wireless Earbud Systems (TWS)

- Environmental Noise Cancellation (ENC)
  - Suppress background noise
  - Hybrid with ANC
- Wake Word Trigger
  - “OK Google”, “Alexa”...
- Voice Assistant
  - Speaker identification
  - Voice command recognition
  - Speech recognition, machine translation
- Biometric Sensor Fusion
  - Heart rate, temperature, EKG, VO2...
Optimizing System Design – Example: VeriSilicon PicoAI IP

- Tiny AI Solution for Mass Market AIOT devices
  - Voice and Vision Wake up
  - AI Voice, AI Vision Applications
  - 200+ GOPS NPU (VIP9000Pico)
  - 1 mm² in TSMC 12nm
- Extremely low power Always-on Wakeup capability
- Highly integrated, self-contained solution
- All Wakeup operations can be achieved without DDR support
- Extremely low power design and implementation
- Low latency Wakeup
- Light weight software stack - RTOS OS, Bare Metal
System Partitioning Is Crucial – Power Hierarchy And Minimizing Memory Accesses

3-Level Waking Up

- **Sensing**
  - Voice Activity Detection (VAD)
  - ~50 uW
  - 24 hours/day (Always-On)

- **Processing**
  - Key Word Spotting (KWS)
  - ~150 uW
  - ~30 min/day

- **AI inferencing**
  - Voice Command, ASR
  - 1 mW – 10 mW
  - ~1 min/day

- **Key Word Spotting (KWS)**
  - ~150 uW
  - ~30 min/day

- **Voice Command, ASR**
  - 1 mW – 10 mW
  - ~1 min/day
System Software Design Is Crucial – Software System for Deeply Embedded Systems

**Offline Compiling on the Host**

- Tensorflow
- PyTorch
- TF-Lite
- Caffe
- TF Lite Micro
- Darknet

**Run Time On the Target**

- Network Binary
- Graph
- Object Code

**ACUITY SDK**

- Universal Model Converter
- Post-training Quantizer
- NN Compiler

**Network Binary**

**Graph**

**Object Code**

**Network Binary**

**Graph**

**Object Code**

**VIP9000-Pico**

**MCU/DSP**

- VIP-Lite Runtime
- RTOS or Bare Metal

**MEMORY**

© 2020 VeriSilicon
Conclusion

• Challenges are ahead
• Solutions are starting to form
• VeriSilicon is here to help
• Let’s make this Embedded Vision happen together!
Resources

Wearable devices:

• [https://mymotiv.com/the-ring/](https://mymotiv.com/the-ring/)


AI-Powered TWS:


PicoAI, VIP9000, VIP9000Pico:


ACUITY SDK:

• [https://github.com/VeriSilicon/acuity-models](https://github.com/VeriSilicon/acuity-models)

2020 Embedded Vision Summit

• Visit VeriSilicon’s virtual booth to speak with technology experts and watch exciting demos.