Video Codecs In An AI World

Dr Doug Ridge
Amphion Semiconductor
The Proliferance of Video in Networks

- Video produces huge volumes of data
- According to Cisco “By 2021 video will make up 82% of network traffic”
  - Equals 3.3 zetabytes of data annually
  - $3.3 \times 10^{21}$ bytes
  - 3.3 billion terabytes
AI Engines Overview

• Example AI network types include Artificial Neural Networks, Spiking Neural Networks and Self-Organizing Feature Maps

• Learning and processing are automated
  • Processing
    • AI engines designed for processing huge amounts of data quickly
    • High degree of parallelism
    • Much greater performance and significantly lower power than CPU/GPU solutions
  • Learning and Inference
    • AI ‘learns’ from masses of data presented
    • Data presented as Input-Desired Output or as unmarked input for self-organization

• AI network can start processing once initial training takes place
Typical Applications of AI

• Reduce data to be sorted manually
  • Example application in analysis of mammograms
  • 99% reduction in images send for analysis by specialist
  • Reduction in workload resulted in huge reduction in wrong diagnoses

• Aid in decision making
  • Example application in traffic monitoring
  • Identify areas of interest in imagery to focus attention
  • No definitive decision made by AI engine

• Perform decision making independently
  • Example application in security video surveillance
  • Alerts and alarms triggered by AI analysis of behaviours in imagery
  • Reduction in false alarms and more attention paid to alerts by security staff
Typical Video Surveillance System
Video Camera Chip Considerations

- Texas Instruments TMS320DM369
  - 1xHDp30 (AVC only)
- HiSilicon Hi3519 V101
  - 4xHDp30
- Ambarella CV2S
  - 8xHDp30
  - Need to decode streams from supported camera chips
    - Multi-format decoder necessary
    - Support camera resolution and frame rate
    - Support multiple camera streams
    - AV1, VP9 support required in future

Diagram:

- Blue circle: Texas Instruments TMS320DM369
- Green triangle: HiSilicon Hi3519 V101
- Red diamond: Ambarella CV2S

Axes:
- X-axis: AVC/H.264 to HEVC/H.265
- Y-axis: 4Kp60, 4Kp30, HDp30
System Implementation

- CPU
- GPU
- FPGA
- ASIC

Flexibility ➔ Efficiency
Multi-Stream Video Decoding

• Multi-stream Operation
  • Time sliced between streams
  • Context switch at frame boundary
    • Negligible switch time
    • Firmware saves & restores hardware internal context
  • Single datapath decoder processes 8xHDp30 video channels in 28nm technology

• Architecture needs to cater for different stream structures
  • Single feed consisting of multiple streams
  • Multiple streams from multiple sources to be processed by single decoder
  • Stream buffering and management necessary prior to the decoder
Video Decoder Considerations

• Handling large number of concurrent streams
  • Typical video surveillance streams are 1080p30
  • Single SoC expected to handle up to 32 streams
  • Single decoder instance within the SoC must decode multiple concurrent streams (typically up to 8)
    • Minimize system cost, number of instantiations and system complexity

• Memory bandwidth challenges
  • Many variables impact this but could be up to ~16GBps for decode in a 32 HDp30 stream system
  • Additional memory for ISP/AI processing and display
  • Frame Buffer Compression (FBC) a possible option to reduce memory bandwidth

• Real value in these SoCs is in the AI engine and associated software
  • Video codec IP handles standard streams so no added value by developing internally
  • Focus engineering effort on differentiation with AI block
  • Video codec IP maturity is important in reducing development risk

• Low latency required where the SoC is in a control loop (e.g. ADAS)
Multi-Format Video Decoder

- CS8142 ‘Malone’ Video Decoder Core
  - Supported formats
    - *AV1 Main Profile @L5.1
    - VP9 Profile 0, 2 @L5.1
    - H.265 HEVC MP@L5.1
    - H.264 AVC BP/MP/HP @L4.2
    - VC-1 SP/MP/AP
    - MPEG-2 MP/HL
    - MPEG-4.2 SP/ASP
    - H.263 / Sorenson Spark
    - DivX 3.11 + GMC
    - China AVS-1 up to L6.1, AVS+
    - Real Media RV8/RV9/RV10
    - ON2 / Google VP6 / VP8
    - BL JPEG / MJPEG
  - Multi stream
    - Up to 8 streams of HDp30 HEVC video at 28nm

*AV1 Under Development (unlikely to be used in camera chips for a few years due to lack of realtime AV1 encoders)
Silicon Area and Power Consumption

• ‘Brains’ and value-add of the chip is the AI engine and associated software
  • Deliver differentiation

• Need the video encoder and decoder to be minimal size and minimal power
  • Allow more resources to be dedicated to the AI engine
  • Achieved by efficient design with an experienced team
  • Minimize the video codec impact on unit cost and power consumption

• Processor subsystem to increase flexibility of the solution
  • Firmware control of top level functions
  • Custom functionality added through firmware
  • Single processor can control multiple decoders
System Level Challenges

• Memory system
  • What happens to the data once it has been processed by the AI engine?
  • How to process multiple streams from multiple sources
  • Sharing memory to reduce system costs
  • Collaboration between the video codec IP vendor and the SoC designer is key

• Decide on what camera chips to support before deciding on video decode engine
  • Support for existing chips and known planned devices
  • Future-proofing the design by including support for new and emerging formats
Summary

• Multi-format decoder essential
  • Support for wide range of camera chipsets
  • Future-proof design by including latest and emerging formats such as VP9

• Multi-stream
  • Decoder needs to meet performance required for multiple streams
  • Memory and core architecture important in order to handle multiple streams

• Efficient design
  • Small silicon footprint to minimize per unit cost
  • Low power consumption

• IP maturity essential to de-risk projects