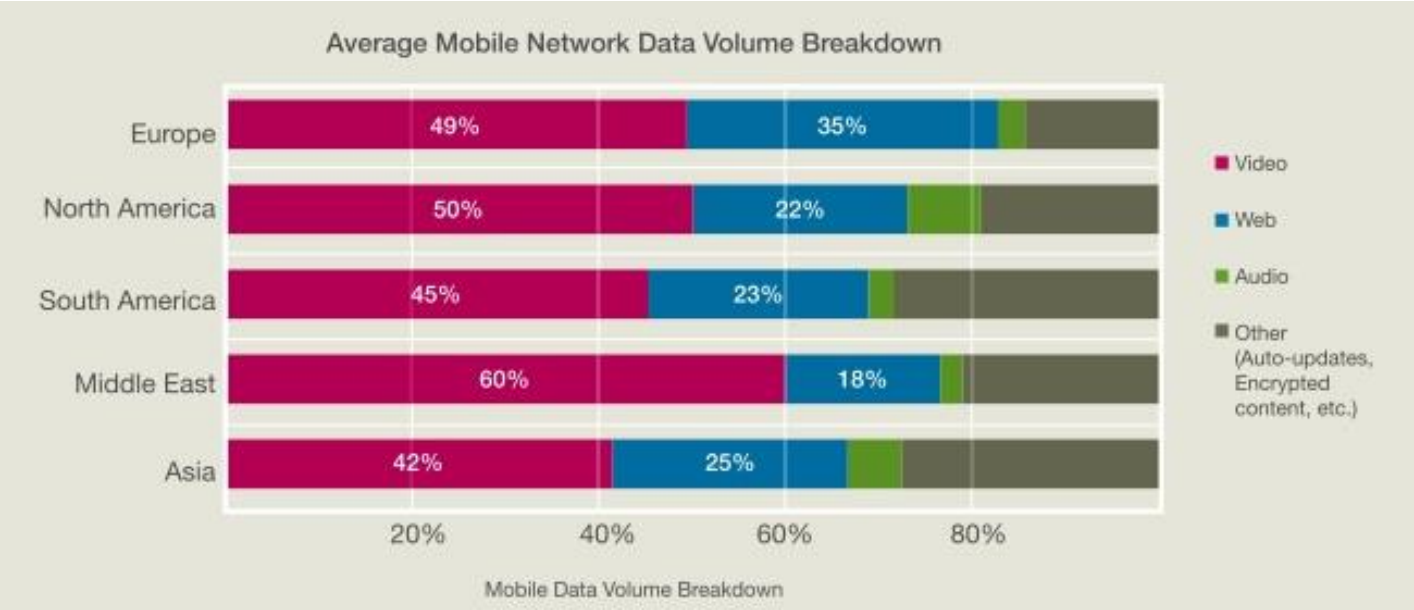
Mali™ GPU acceleration of HEVC and VP9 Decoder

Ittiam

Web Video continues to grow !!!

- Video accounted for 50% of the mobile traffic in 2012
  - Citrix ByteMobile's 4Q 2012 Analytics Report.



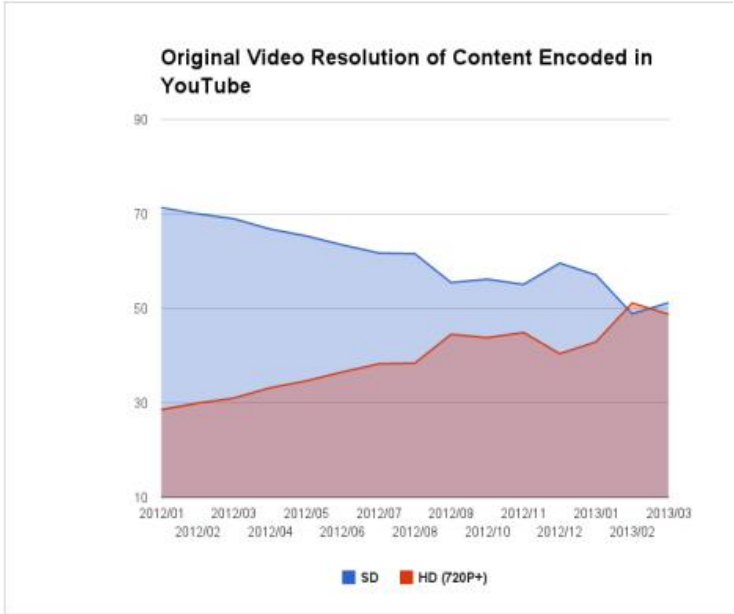
- Globally, IP video traffic will be 79 percent of all consumer Internet traffic in 2018, up from 66 percent in 2013
  - Cisco Visual Networking Index: Forecast and Methodology, 2013-2018, Cisco Systems, Inc

Web Video continues to grow !!!

YouTube Video Growth

A global-scale video platform:

- 1B+ monthly users
- 4B+ video views per day
- 6B+ hours watched per month
- 72+ hours of video uploaded each minute
  - HD becoming dominant
- 25% mobile consumption



Slide Source : Google I/O 2013

Video compression is the solution !!!

- The existing videos have to be compressed more..

Offers upto 50% more compression

- HEVC and VP9 are two new video coding standards

HEVC standard

- **HEVC aka H.265** is a video compression standard, jointly developed by **ISO/IEC MPEG and ITU-T VCEG**

- Licensing cost to patent holders – ~ $0.20 per unit as Royalty

Courtesy : Qualcomm Inc

VP9 standard

- **VP9** is a video compression standard, developed by **Google**

- **No patent licensing cost – Royalty FREE**

Courtesy : Google I/O 2013

HEVC and VP9 computational complexity

- Higher compression ratio comes with a higher computational complexity

- Multiple factors makes HEVC and VP9 more complex
  - Recursive Coding block structure
  - Non-square inter prediction block sizes
  - Higher size of Transforms
  - Higher number of intra prediction angles
  - Higher taps in filters
  - Additional tools such as SAO etc.,

- Implementing HEVC and VP9 decoders on mobile offers multiple choices and hence multiple challenges

HEVC and VP9 on Mobile processors

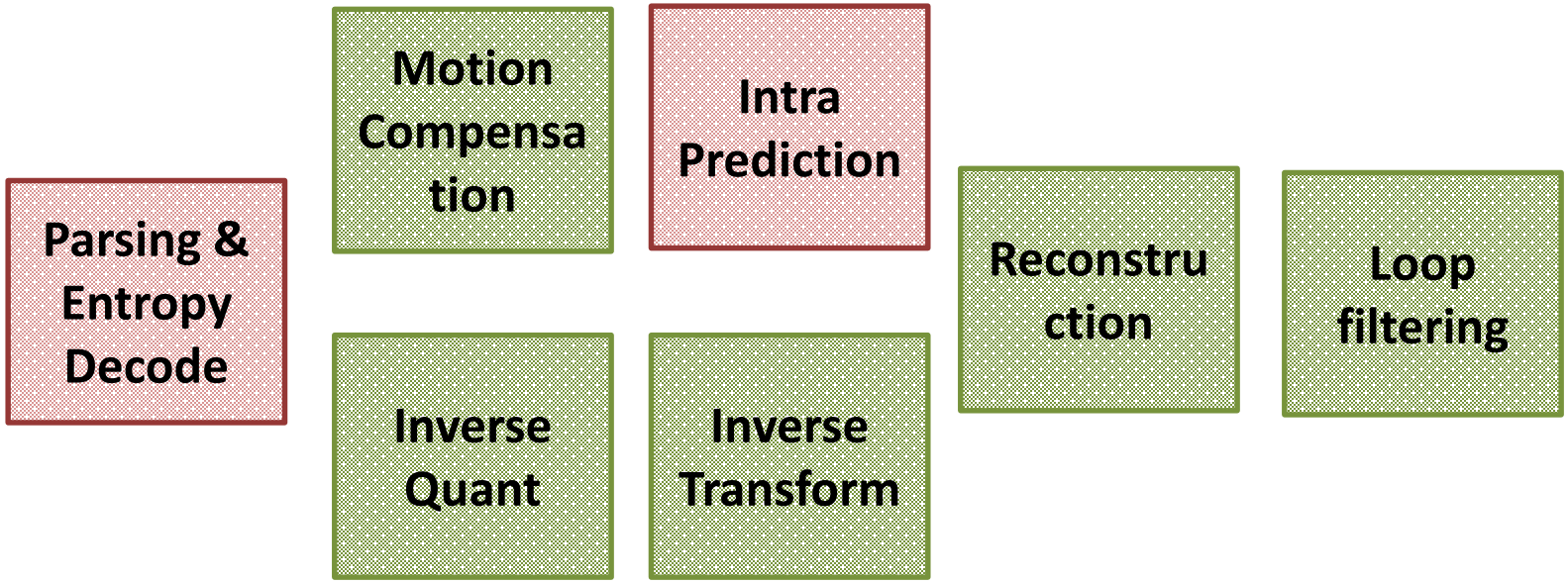
Key Vectors	CPU Only	DSP Only	H/W based	CPU + GPU / Open CL
Power Consumption	High	Moderate	Low	Moderate
Silicon Cost	Low	High	High	Low
Development Lead Time	Low	High	High	Low
Upgradability	High	High	Low	High
Portability	High	Low	None	High

CPU + GPU offers a good trade-off!!!

HEVC/VP9 decoder – GPU acceleration

- GPUs are generally idle during video playout
- GPUs are massively multithreaded devices. i.e., GPUs will require hundreds or thousands of threads to be executed in parallel at any given time
- So only highly data parallel algorithms inside a video codec can be efficiently offloaded to the GPU
- Offloading the video decode task to GPU, would enable the CPU to perform other tasks

HEVC/VP9 Decoder – with GPU Acceleration

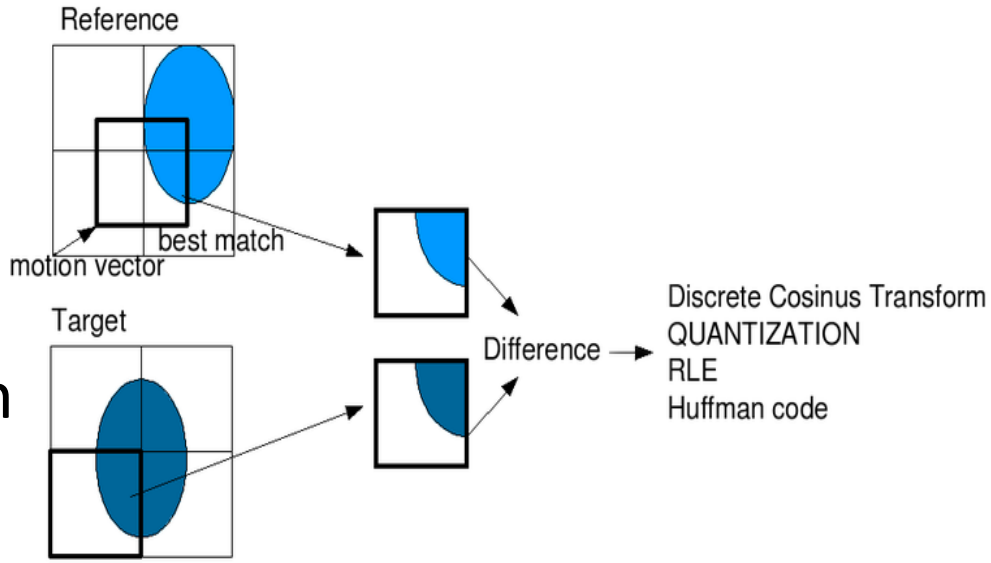


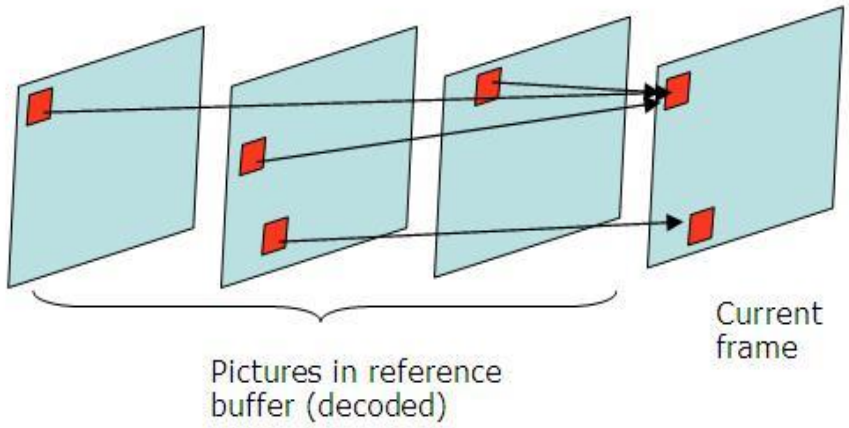
Not suitable for GPU execution

Data parallel execution possible, suitable for GPU execution

Motion compensation

- The current picture/frame pixels is predicted from the reference frame's pixels
- The reference picture can be from past or future
- The prediction happens on a block-by-block basis
- And there can be multiple reference frames for each block

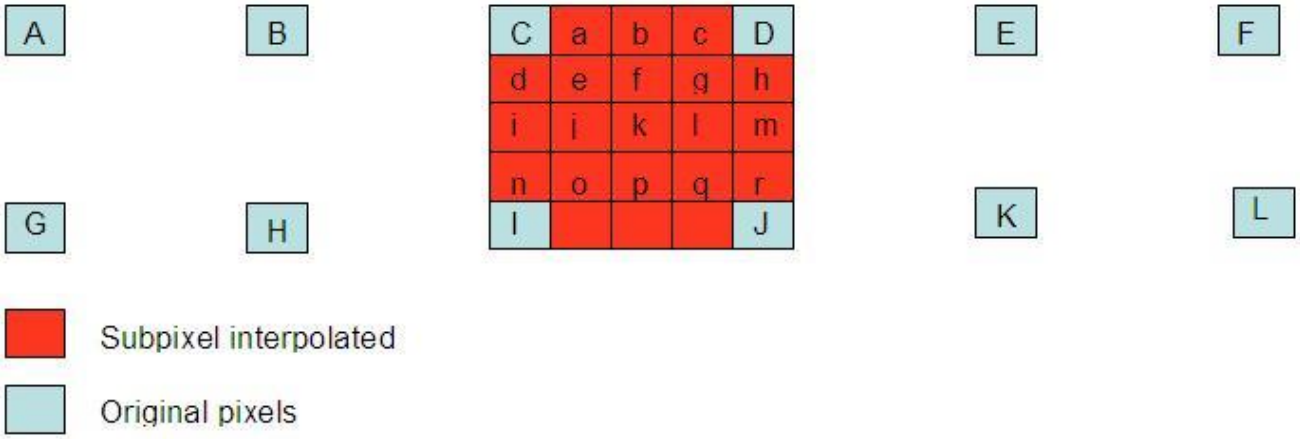




Motion compensation

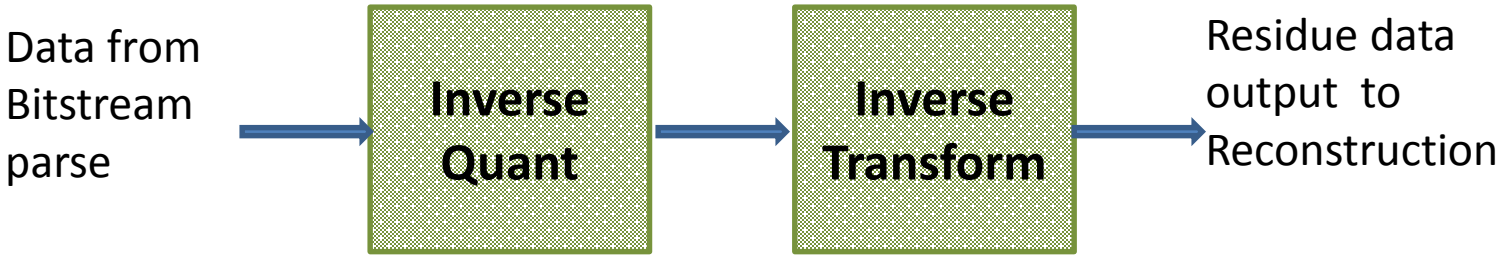
- The most compute intensive part of Motion compensation is sub-pixel interpolation
  - Luma – 8 filter
  - Chroma – 4 tap filter in HEVC and 8 tap filter in VP9

- Sub pixel interpolation is data parallel, i.e., interpolation of each block within a frame can happen in parallel and hence suited for GPU computing



Inverse quantization and transforms

- The residue value need to be Inverse quantized
- 2-D Inverse DCT transformations should be performed over the inverse quantized data
- Both these process are data parallel and 2-D Inverse DCT in particular is highly compute intensive and hence suitable for GPU computing

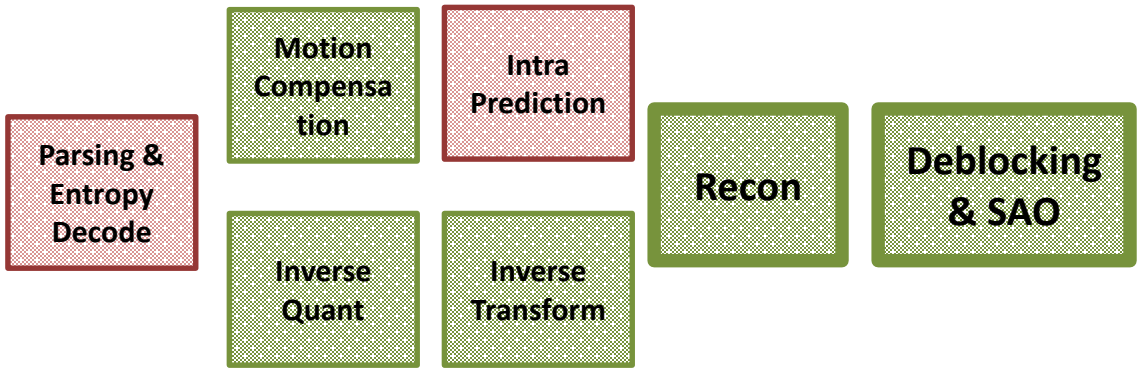


Reconstruction and In-loop filters

- Reconstruction : The output from the Motion compensation and intra prediction should be added with the output from Inverse transform

- In loop filtering such as Deblocking and SAO(only HEVC) filters are applied over reconstructed samples

- Again these are data parallel and compute intensive jobs that can be accelerated with GPU



Benefits of Mali™ GPUs for Video

- Mali GPUs are highly suited for Video acceleration, with a easy-to-optimize architecture
- The 128-bit vector processing
  - Suits DSP algorithms like Video processing
- Presence of GPU cache instead of Local memory
  - No requirement for data transfers from/to global memory. Can be understood just like a CPU
- Flexible OpenCL workgroup size
  - Works optimizally for a large range of OpenCL workgroup sizes. Multiple block sizes in a Video frame can be handled efficiently
- No divergent threads
  - Similar to CPU code, conditional code can be used in OpenCL kernels as well. Different kinds of filter types, filter lengths etc., in video decode can be handled efficiently
- CCI/CCN support
  - Mali GPUs supports ARM® CCI/CCN technologies for cache coherent access of data between CPU and GPU. This will reduce the DDR memory stalls and cache synchronization overheads significantly

Challenges

- Efficient Partitioning of work between CPU and GPU
  - The effective FPS of decoder will be the minimum of the FPS achieved by the CPU and GPU for their respective work
  - So the partitioning needs to be efficient so that both of them perform their respective work at almost the same speed(FPS)
- Efficient pipelining data between CPU and GPU
  - The algorithms running on CPU will depend on the output of algorithms from GPU and/or vice versa
  - A good design should make sure neither the CPU nor the GPU spend any time waiting for the output of the other
- Cache coherency
  - Cache coherency between CPU and GPU data need to ensured.

Summary

- Online video traffic increases rapidly
- HEVC and VP9 answers the video traffic problem, by offering upto 50% more compression on video
- HEVC and VP9 decoding on mobile is challenging
  - CPU + GPU based solutions offers a good trade off
- GPU accelerated video decoding is feasible, but with certain challenges
- The power and performance results of GPU accelerated video decoder is very impressive

THANK YOU