

# Using Heterogeneous Computing for Mobile and Embedded Vision

Rick Maule, Sr. Director, Product Management

Qualcomm Technologies, Inc.

October 2, 2013

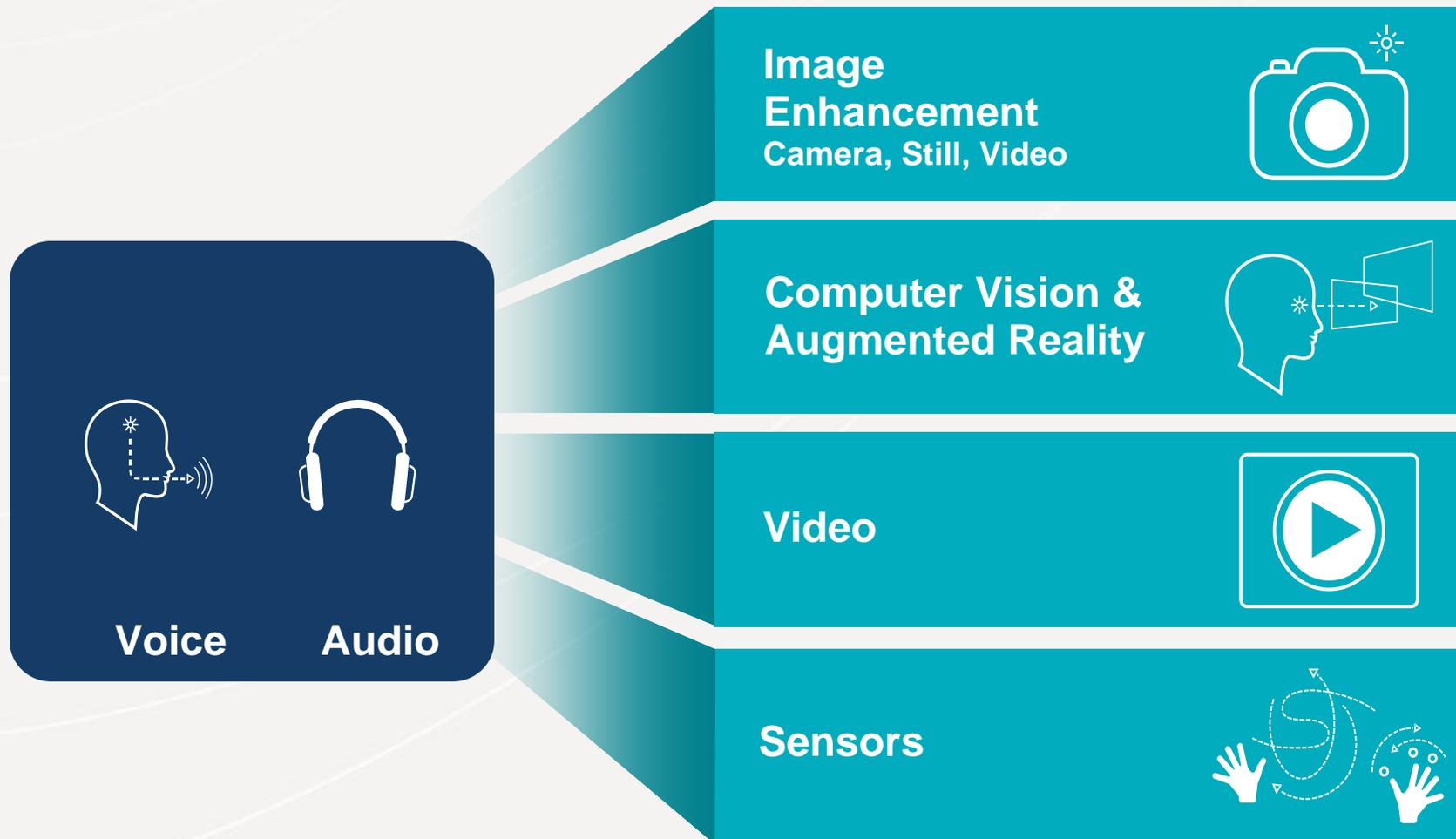


## Evolution in the mobile arena

---

*Revolutionizing mobile visual experiences*





# Enriching the Mobile Experience

## New Apps

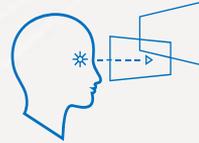
Emerging Workloads

Compute Intensive

Diverse Characteristics



Realistic Physics



Computer Vision



Web Browsing



Computational Photography



Augmented Reality



Natural UI & Gestures



Contextual Awareness



## Mobile Device Constraints

Sleek, Ultra-Light



Long Battery Life

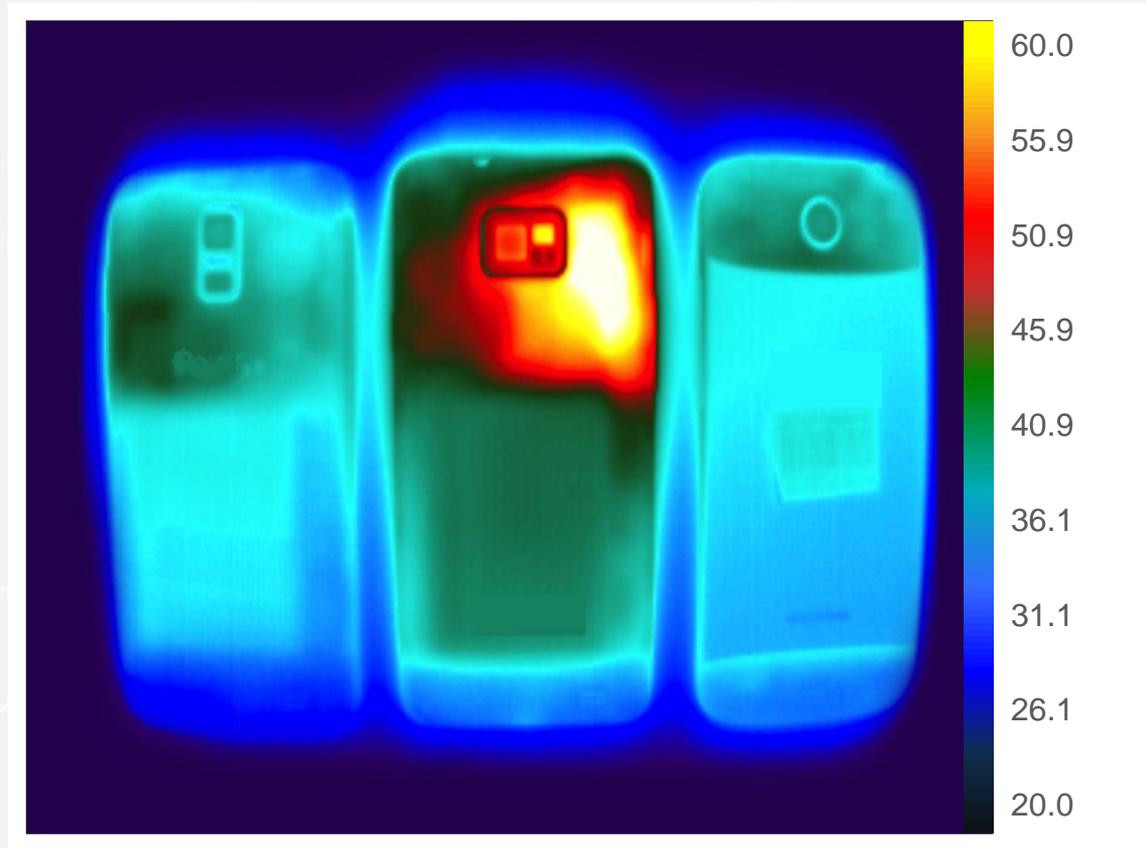


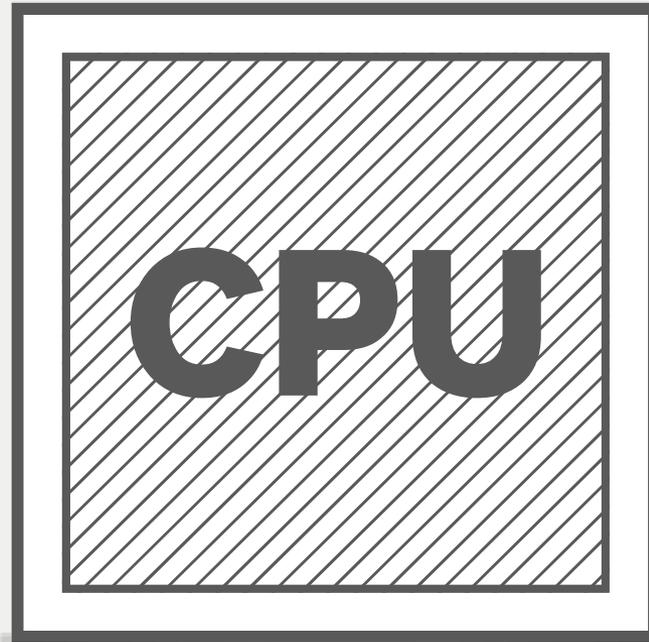
Thermal Efficiency

# Better Thermals = Better User Experience

***Less wasted energy***

***Longer life of device***





## Looking ***beyond the CPU***

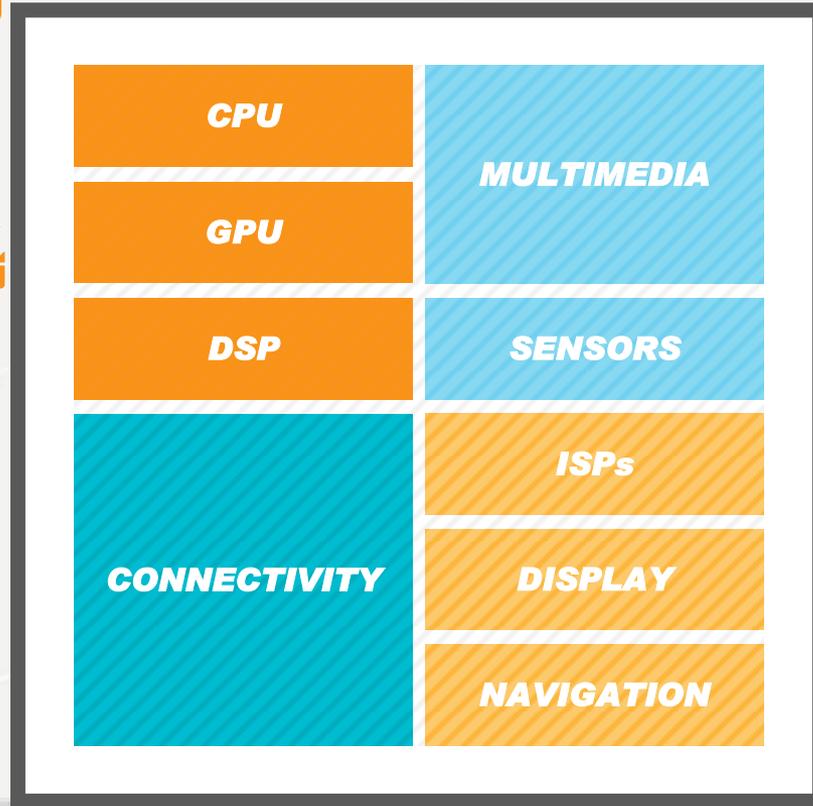
What % of a Mobile SoC is CPU?



15%

Looking ***beyond the CPU***

# Heterogeneous computing



## Right task for the right processing engine

---

*Strength from diversity*



# Diversity exists for a reason

Each programmable engine has its own strengths and weaknesses

## Sequential Control



Browsing



Game AI

## Low-power Real-time



Audio



Object Detection

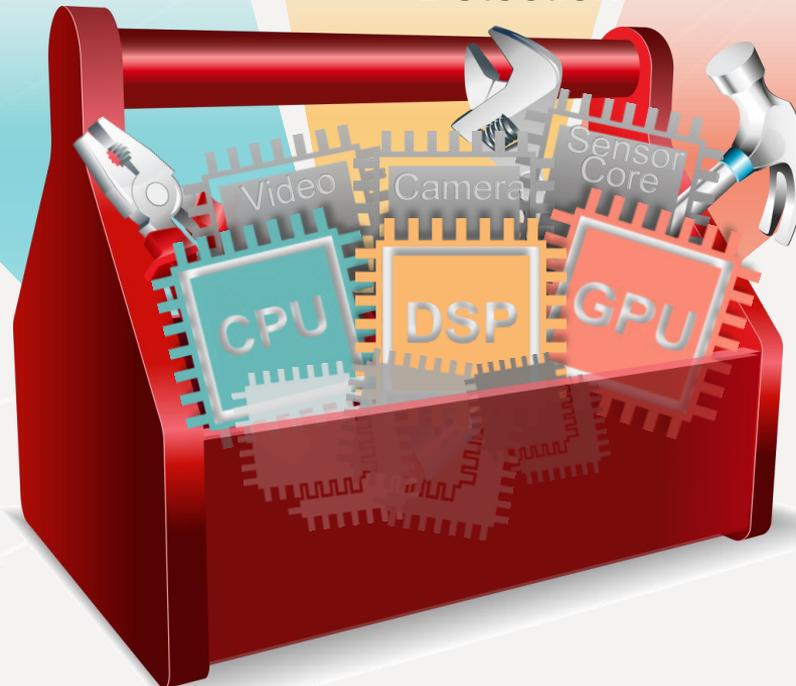
## Streaming Parallel Data



Composition



Image Processing

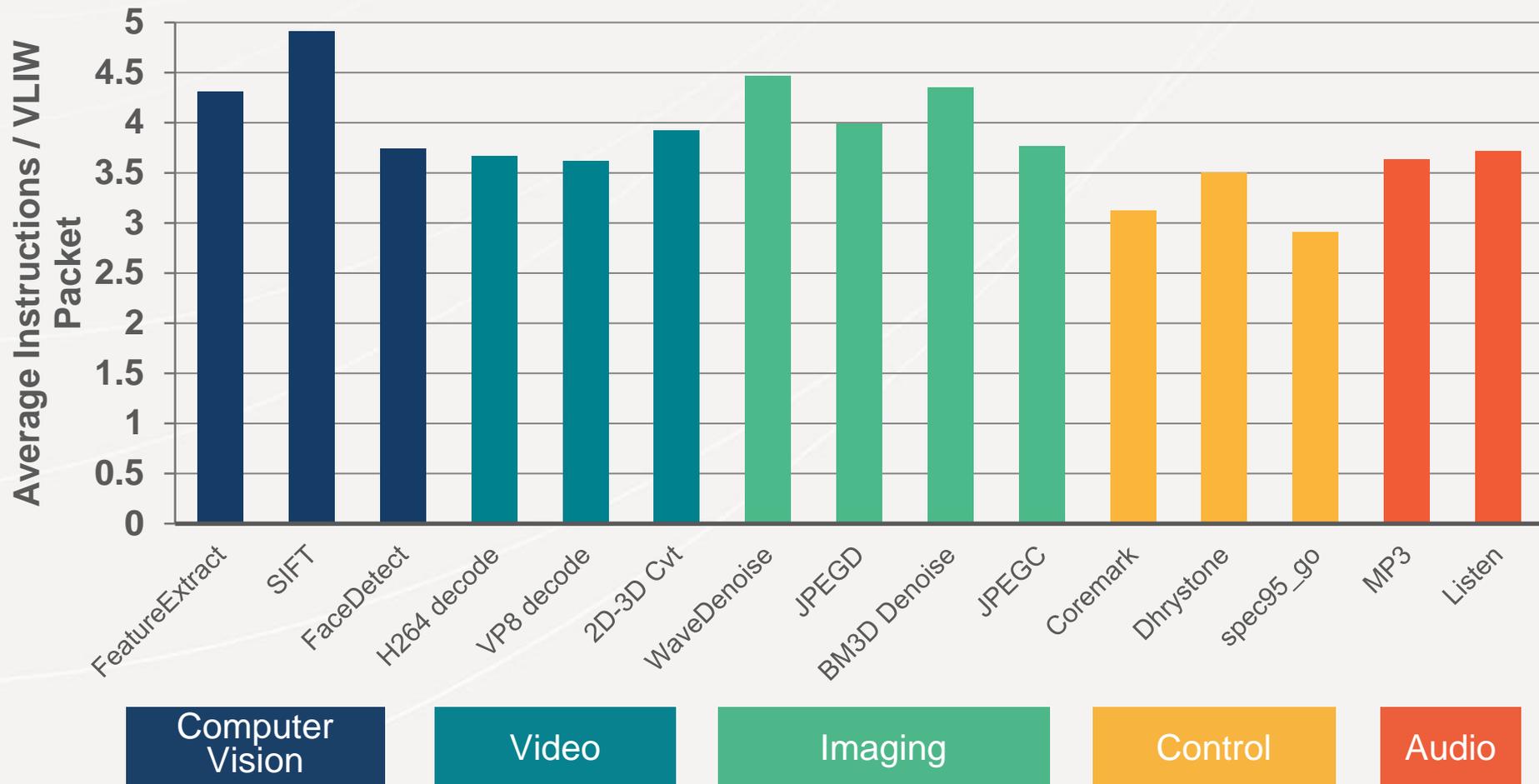


*You don't build a house with just a hammer*

# Each engine has characteristics suited for specific tasks

	CPU Strengths	DSP Strengths	GPU Strengths
<b>Applications</b>	<p><b>Best for:</b></p> <ul style="list-style-type: none"> <li>• Large kernels</li> <li>• Irregular data structures</li> <li>• Complex flow control</li> </ul> <p><b>Examples:</b></p> <ul style="list-style-type: none"> <li>• OS, Web browser</li> <li>• Office productivity apps</li> <li>• Game Engines</li> </ul>	<p><b>Best for:</b></p> <ul style="list-style-type: none"> <li>• Apps with irregular data parallelism (context or data dynamically modifies processing)</li> <li>• Flow control mixed with data processing</li> </ul> <p><b>Examples:</b></p> <ul style="list-style-type: none"> <li>• Object detection &amp; tracking (NCC, OCR)</li> <li>• Video codecs/filters not in dedicated hw</li> <li>• Audio &amp; image signal processing and filters</li> </ul>	<p><b>Best for:</b></p> <ul style="list-style-type: none"> <li>• Arithmetic heavy kernels</li> <li>• Large &amp; parallel data sets</li> <li>• Predication style flow control</li> </ul> <p><b>Examples:</b></p> <ul style="list-style-type: none"> <li>• Games UIs rendering &amp; compositing</li> <li>• Image processing (ALU rich)</li> </ul>
<b>Compute Control</b>	<ul style="list-style-type: none"> <li>• <b>Highest control code performance</b></li> <li>• &gt;10x control capability of GPU</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Moderate amount of control code performance (esp. loop control)</b></li> <li>• Best for throughput of small to medium separate data processing streams</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Much less control performance than CPU</b></li> <li>• Best throughput for large data streams</li> </ul>
<b>Compute Data Parallelism</b>	<p><b>Some</b></p> <ul style="list-style-type: none"> <li>• Multi-core multiprocessing</li> <li>• NEON vector processing</li> </ul>	<p><b>Moderate</b></p> <ul style="list-style-type: none"> <li>• HW for multithreading/multitasking</li> <li>• High SIMD for 8-/16-bit fixed point</li> <li>• Multiple 32-bit fixed/floating point per cycle</li> </ul>	<p><b>Large</b></p> <ul style="list-style-type: none"> <li>• Highest SIMD compute (can be &gt;5x over CPU, depending on chipset and floating point precision)</li> </ul>
<b>Memory Interaction</b>	Many applications tend to improve with low memory latency	Enhances memory latency tolerance with multi-threading and prefetching of data (more tolerant than CPU, less than GPU)	Highly memory latency tolerant, but often requires high bandwidth
<b>Dedicated Functions</b>	<b>None</b>	<b>Yes</b> Special instructions for multimedia acceleration (e.g., CV feature detection)	<b>Yes</b> Dedicated image/texture filtering hardware and Rasterization useable by compute

# DSP Goal: More work at lower frequencies



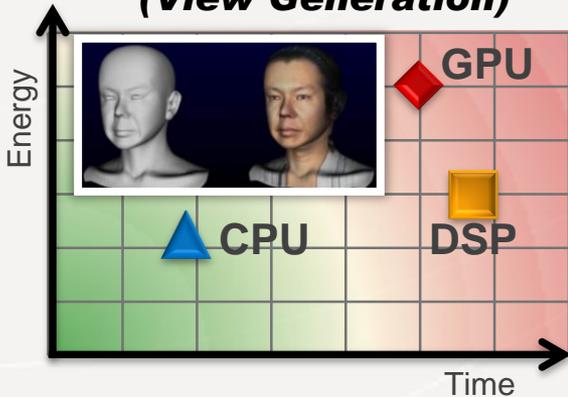
Source: Qualcomm Technologies internal measurements

*Compound instructions count as 2*

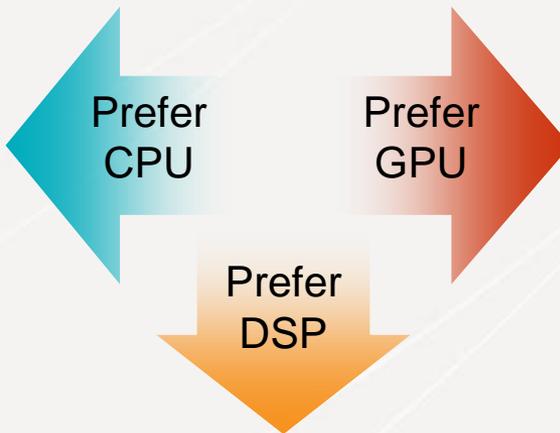
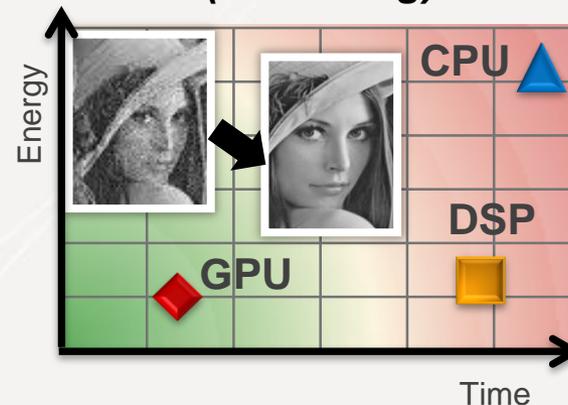
# Right task on the right engine

Different performance & power benefits

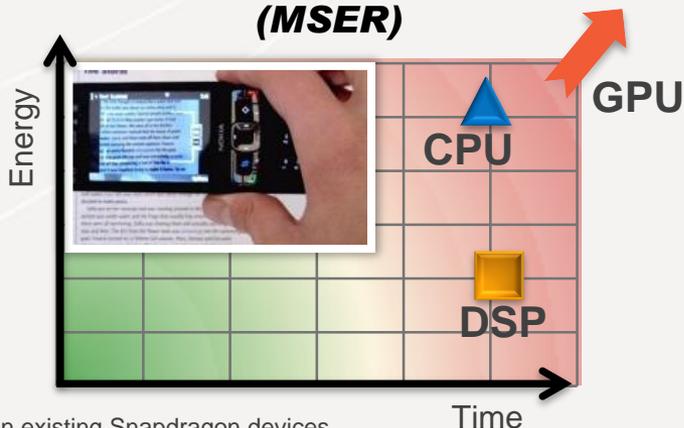
## 2D -> 3D Video Conversion (View Generation)



## Image Processing (Denoising)



## Character Recognition (MSER)



Source: Internal Qualcomm Technologies measurements on existing Snapdragon devices

## Developing “Heterogeneously”

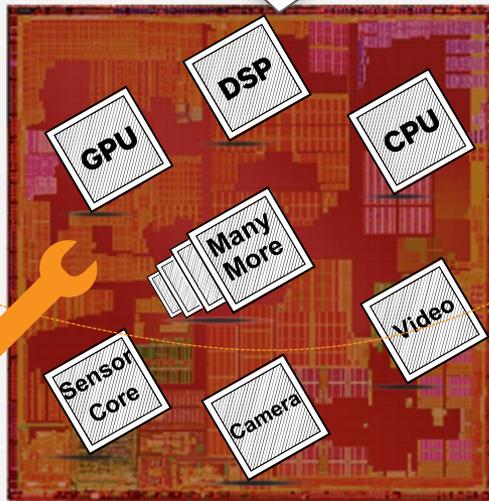
---

*Programming for mobile imaging*

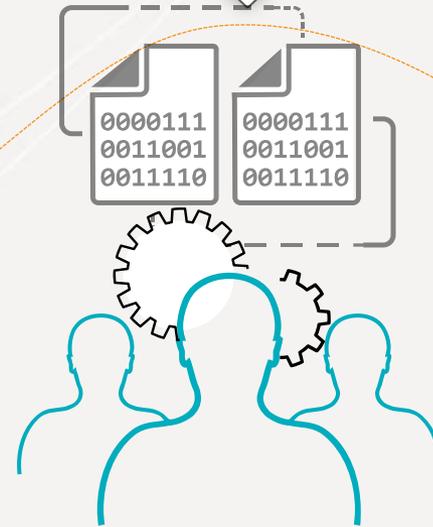


# Mobile heterogeneous computing requires more than engines

***A computing approach that intelligently uses fundamentally different types of processors***



***Assign right task to the right engines***



***Make the engines Accessible with Tools***

# Tools, Languages, APIs that Target Multiple Engines

Languages/APIs

LLVM Compiler Toolchain  
(LLVM = Low Level Virtual Machine)

HW

Developer Tools

OpenCL  
Renderscript  
DSP APIs  
Web Audio  
Fast CV  
OpenCV/VL  
Shader Programs  
(GLSL, HLSL)  
Others...

LLVM  
Front  
Ends

LLVM  
Optimizer

LLVM  
DSP  
Back End

LLVM  
GPU  
Back End

LLVM  
CPU  
Back End

DSP

GPU

CPU

Bitcode

Bitcode

Bitcode

Data

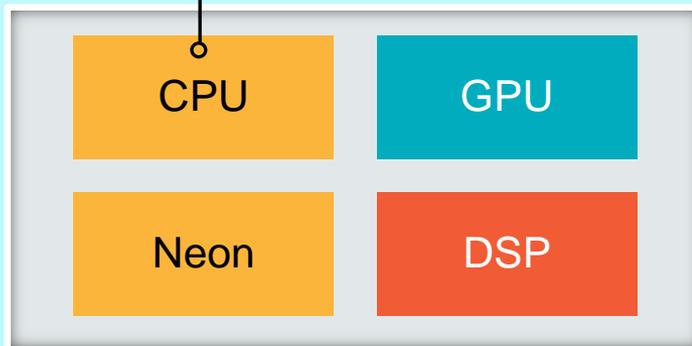
Data

Data

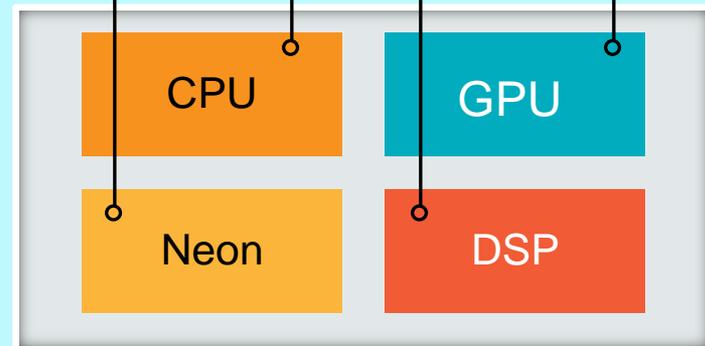
- Debugger
- Emulator
- Profiler
- Parallel Analyzer
- Optimized Libraries
- Code Samples

## Computer vision engine

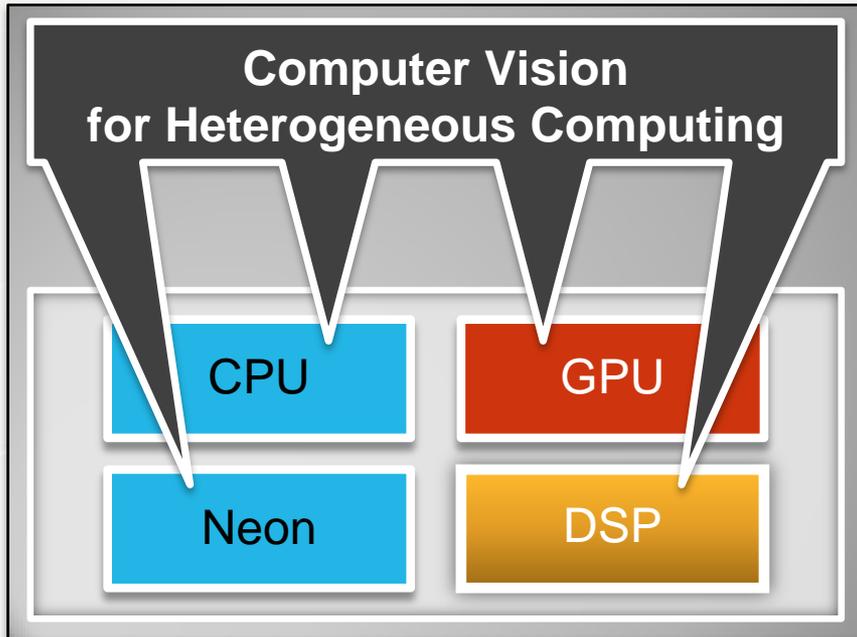
### Computer Vision for CPU



### Computer Vision for Heterogeneous Computing



# Computer Vision API Benefits



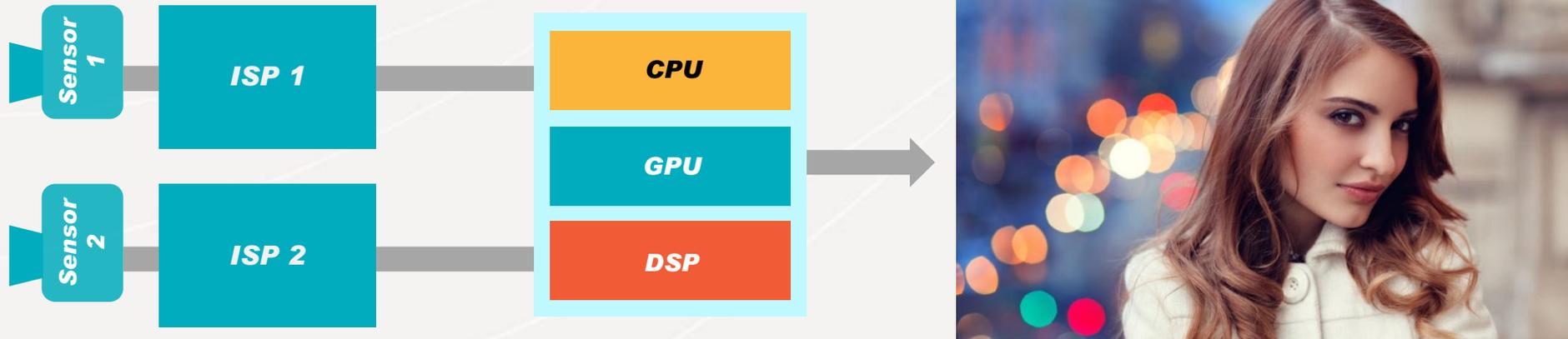
- Designed to provide key functions needed by Real-time Computer Vision (CV) applications
- Functions can be optimized to take advantage of each processing engine as appropriate

Function	Speed up over CPU only
<b>NCC</b>	<b>23.1x</b>
<b>Dot Product 128x4</b>	<b>10.0x</b>
<b>Convert YUV420</b>	<b>1.3x</b>
<b>Sobel</b>	<b>7.8x</b>
<b>Median3x3</b>	<b>51.9x</b>
<b>Threshold</b>	<b>9.7x</b>
....	....

Source: Qualcomm Technologies internal measurements

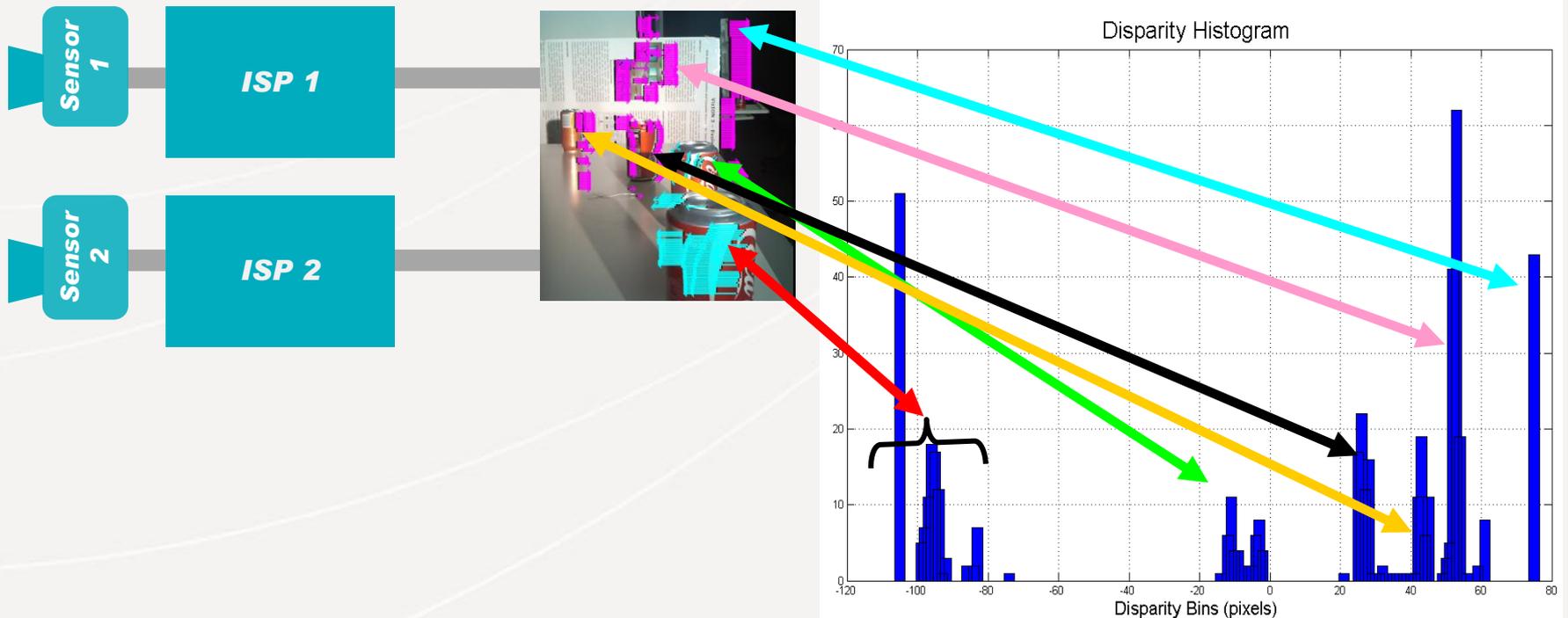
# Computational camera

Depth calculation designed to enable Bokeh effect



# Computational camera

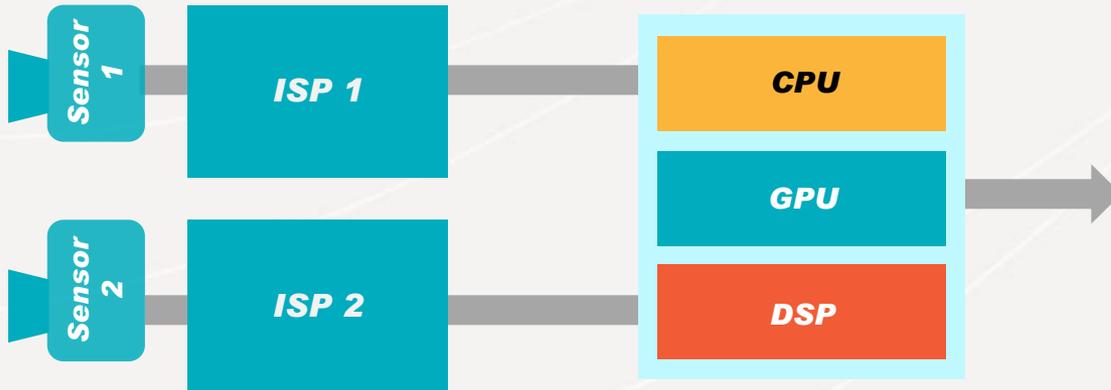
Depth calculation designed to enable range finding for instant auto-focus



Source: Qualcomm Technologies internal data

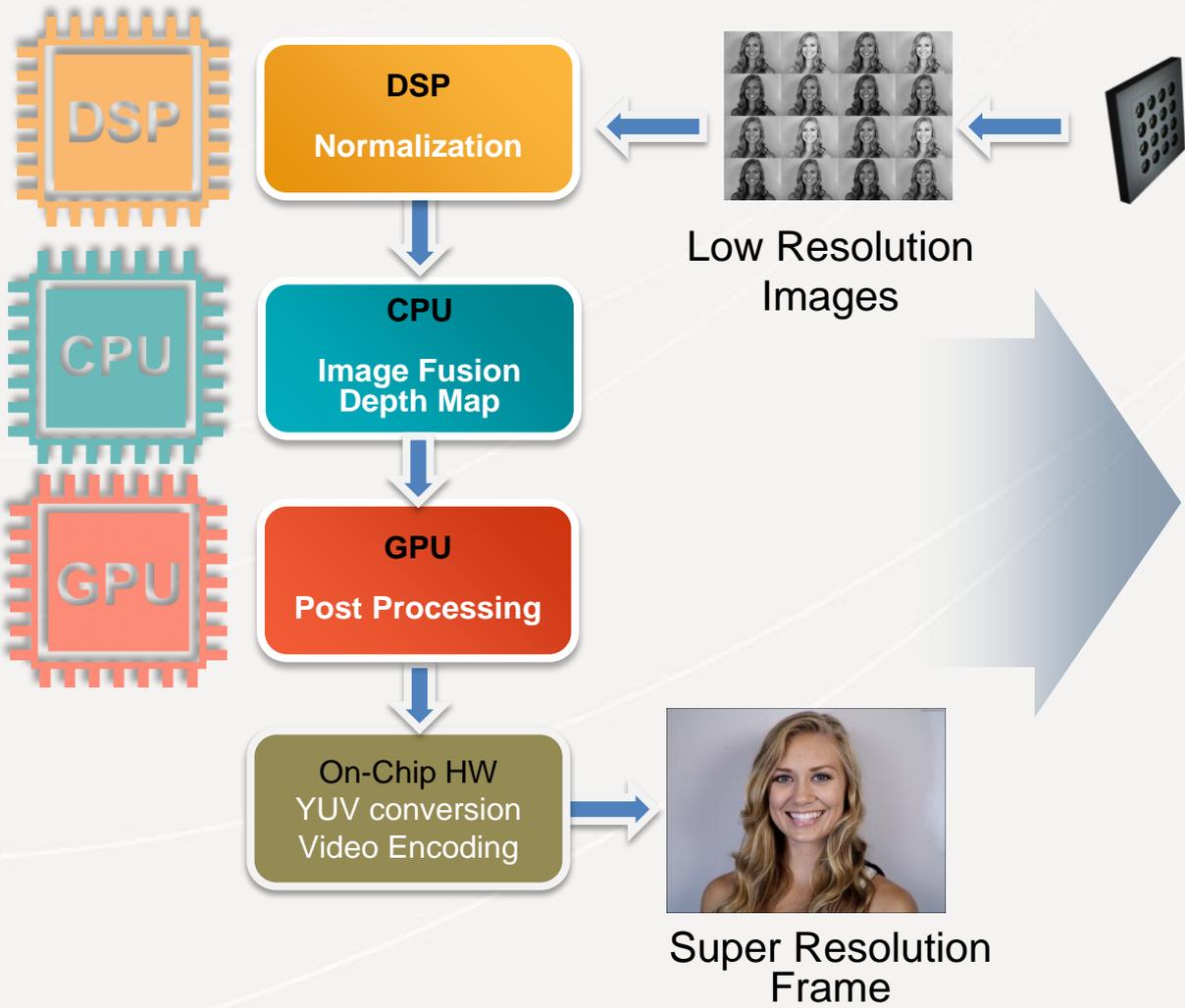
# Computational camera

Depth calculation designed to enable shoot once, focus later



# Heterogeneous collaboration for computational photography

Multiple engine "pipeline"



## Heterogeneity Benefits

### System

> 3x overall improvement  
CPU to distributed

### For some key algorithms

>10x Performance

>3x Energy reduction!



For more details please go to  
[www.pelicanimaging.com](http://www.pelicanimaging.com)

## Why do you care?

---

*It's all about power in mobile*

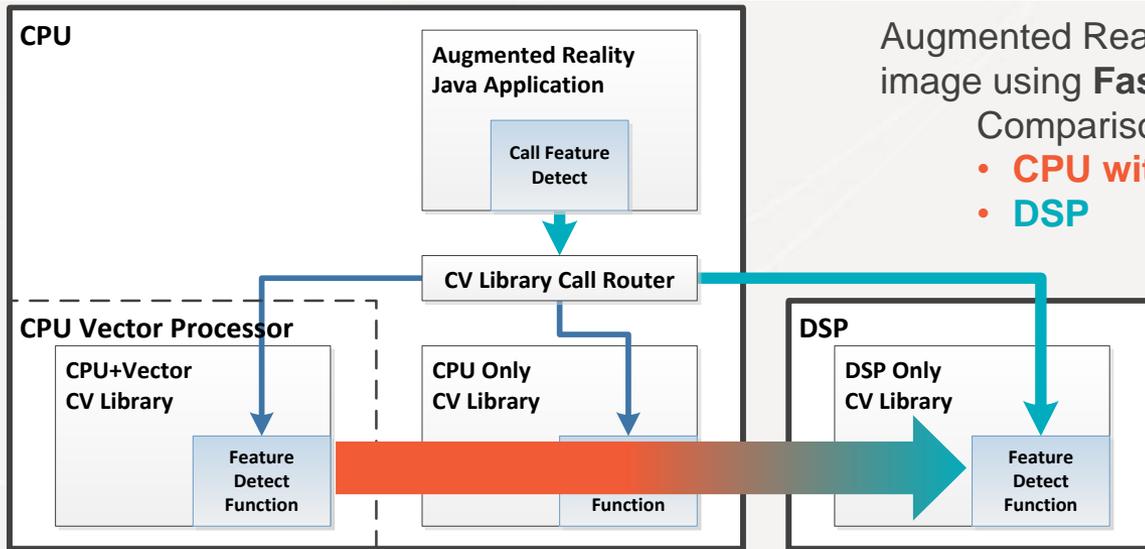


# Computer vision offload - CPU to DSP

Augmented Reality Java App finding objects in image using **FastCV Feature Detect**

Comparison of Feature Detect run on:

- **CPU with Vector Processor**
- **DSP**



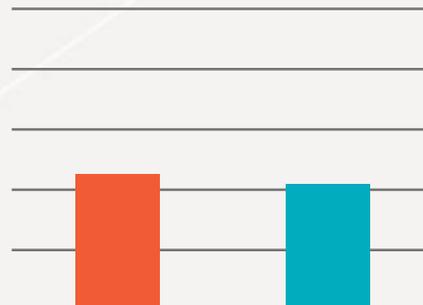
**CPU Utilization (%)**

**Detection Time (%)**

**Total Device Power (%)**



**52% Less CPU**



**7% Less Time**

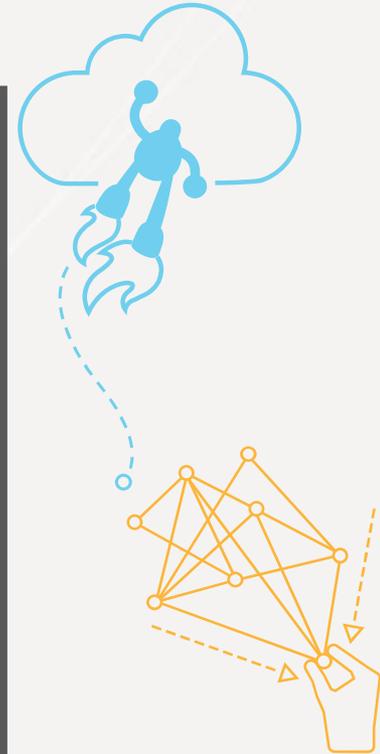
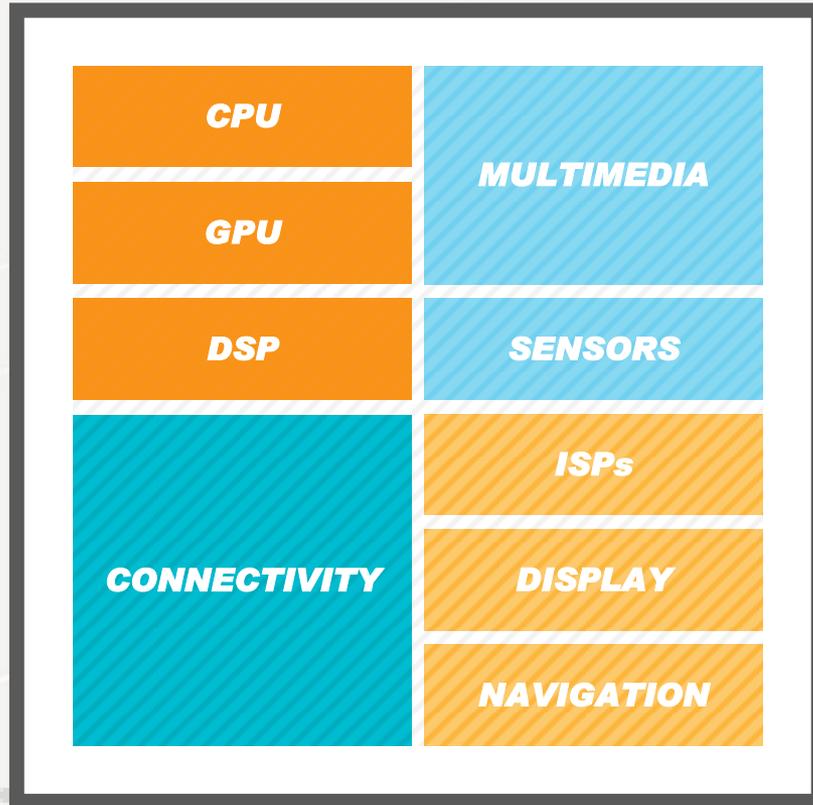


**32% Less Power\***

\* Power measured at the device battery

Source: Qualcomm Technologies internal measurements.

# Heterogeneous computing



## Developer.Qualcomm.com

- Mobile Platform Operating Systems
  - [Android™](#)
  - [Brew®](#)
  - [HTML5 Mobile Web](#)
  - [Windows® RT](#)
  - [Windows Phone](#)
- Mobile Platforms Devices
  - [Snapdragon MDP Smartphone & Tablet](#)
  - [DragonBoard Mobile Development Board](#)
- Performance Tools
  - [Snapdragon LLVM Compiler for Android](#)
  - [Trepp Profiler for Snapdragon Processors](#)
  - [Trepp Plug-in for Eclipse](#)
- Mobile Technologies
  - [3G & 4G Connectivity \(Gobi\)](#)
  - [Augmented Reality \(Vuforia\)](#)
  - [Computer Vision \(FastCV\)](#)
  - [Context Aware \(Gimbal\)](#)
  - [Mobile Gaming & Graphics Optimization \(Adreno\)](#)
  - [Multimedia Optimization \(Hexagon SDK\)](#)
  - [Snapdragon SDK for Android](#)
  - [Peer-to-Peer \(AllJoyn\)](#)
  - [Web Technologies](#)
  - [Wireless Health \(2net\)](#)

# Thank You

For more information on Qualcomm, visit us at:  
[www.qualcomm.com](http://www.qualcomm.com) & [www.qualcomm.com/blog](http://www.qualcomm.com/blog)

© 2013 Qualcomm Technologies, Inc.

Qualcomm, Snapdragon, and Hexagon are trademarks of QUALCOMM Incorporated, registered in the United States and other countries. All QUALCOMM Incorporated trademarks are used with permission. Other product and brand names may be trademarks or registered trademarks of their respective owners. Snapdragon processors and Hexagon are products of Qualcomm Technologies, Inc.

This technical data may be subject to U.S. and international export, re-export, or transfer (“export”) laws. Diversion contrary to U.S. and international law is strictly prohibited.

Qualcomm Incorporated  
5775 Morehouse Drive  
San Diego, CA 92121  
U.S.A.