Driving the New Era of Immersive Experiences
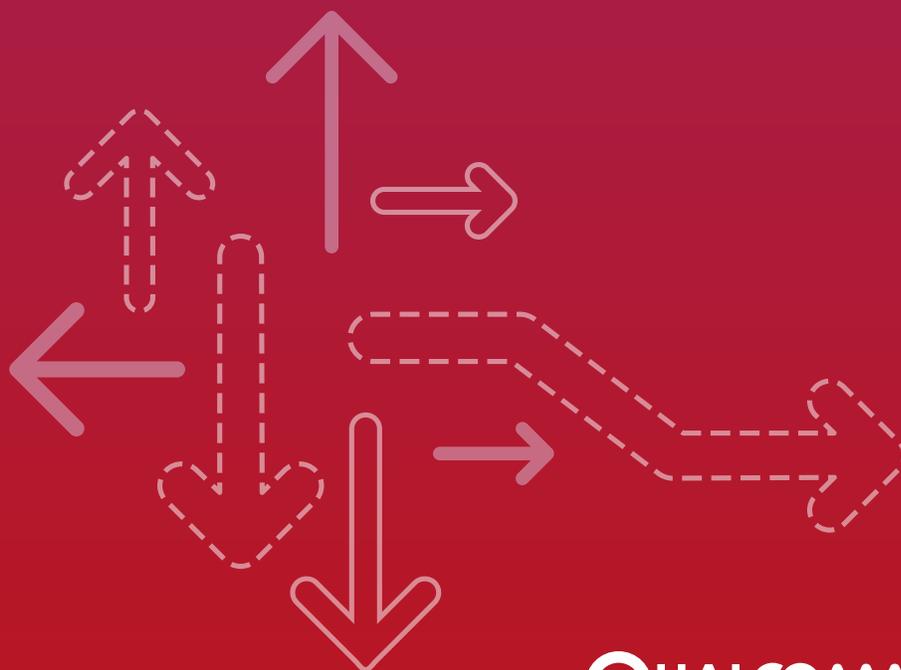
QUALCOMM®

Disclaimer

Qualcomm, Snapdragon, Adreno, Hexagon, and DragonBoard are trademarks of Qualcomm Incorporated, registered in the United States and other countries. Krait, Kryo, Qualcomm Spectra, FastCV, and Zeroth are trademarks of Qualcomm Incorporated. All Qualcomm Incorporated trademarks are used with permission. Other products and brand names may be trademarks or registered trademarks of their respective owners.

Qualcomm Snapdragon, Qualcomm Adreno, Qualcomm Spectra, Qualcomm Kryo, Qualcomm Krait, Qualcomm Hexagon, FastCV, and Qualcomm Zeroth are products of Qualcomm Technologies, Inc. DragonBoard is distributed by Intrinsyc.

Contents

1   Executive summary

Immersion enhances everyday experiences, making them more realistic, engaging, and satisfying. Virtual reality is the ultimate level of immersion, but consumers want immersive experiences on all their devices — whether playing a video game on a smartphone, video conferencing on a tablet, or watching a movie on a TV. The goal is to provide the appropriate level of immersion based on the device form factor, use case, and context. This paper examines to how to make this goal possible.

The three pillars of immersive experiences are visual quality, sound quality, and intuitive interactions. Full immersion can only be achieved by simultaneously focusing on the broader dimensions of each pillar. Too often, the focus has been on specific dimensions, such as pixel quantity, rather than other dimensions, like pixel quality, which may be equally or more important for specific use cases.

The optimal way to enhance these broader dimensions for more immersive experiences requires:

- Taking an end-to-end approach for comprehensive solutions
- Utilizing heterogeneous computing for efficiency
- Applying cognitive technologies for on-device intelligence

Qualcomm Technologies, Inc. (QTI) is uniquely positioned to enhance the broader dimensions of immersive experiences by custom designing specialized engines across the SoC and offering comprehensive ecosystem support.

2   Immersion enhances experiences

The experiences worth having, remembering, and reliving are immersive — experiences like a live sporting event, an exotic vacation, a concert performance, a great movie, or nature. Immersive experiences draw you in, take you to another place, and keep you present in the moment. With more distractions than ever in modern life, an immersive experience makes you focus, filter out noise, and be part of the experience, rather than a passive observer.

Immersive experiences stimulate your senses — your vision, your hearing, your smell, your taste — and your imagination. For example, imagine viewing a sunset at the Grand Canyon. Your eyes might be stimulated by the marvelous colors of the sunset, your ears by the sound of a bird flying by, your nose by the clean smell of fresh air, and your skin by the slight humidity in the air.

Immersion will enhance everyday experiences, making them more realistic, engaging, and satisfying. For example, experiences such as (Figure 1):

- Watching movies, sports, or other types of videos that make you feel like you are actually there, such as at the Super Bowl, the World Cup, or a Rolling Stones concert.
- Playing mobile games with visual user experiences so realistic that they suck you into the action.
- Video conferencing with the family as if you are all in the same room.

- Seamlessly interacting with the user interfaces (UI) of magazines, web pages, and showrooms.
- Augmented reality (AR)[1], where objects, such as a toy or image, are brought to life. AR blurs the lines between physical and digital, allowing users to interact with the physical world around them in new ways.
- Virtual reality (VR)[2] enables users to experience just about anything imaginable, such as explore the Seven Wonders of the World, play video games, and interact with other people in new ways. VR, when done right, is the ultimate level of immersion since it stimulates the human senses with feedback that is so realistic that it convinces the brain that the virtual experience is real.

Figure 1: Immersion enhances everyday experiences across devices

Various levels of immersion are achievable across different device form factors. For example, all these experiences could be completely immersive on a VR headset and highly immersive on a smartphone, while the visual experience on a smartwatch may not be quite as immersive due to the small screen. The goal is to make experiences appropriately immersive by taking into account device constraints. The rest of this paper will explore the key elements of immersive experiences and optimal approaches to making them possible.

3 Focusing on the broader dimensions of immersion

The three pillars of immersive experiences are visual quality, sound quality, and intuitive interactions. While each pillar stands alone at making experiences more immersive — think of the importance of the visual quality when viewing a photo or sound quality when listening to music — they are also complementary and synergistic. Immersion happens when all three pillars are combined together. Within each pillar, there are multiple dimensions that improve the overall quality. Full immersion can only be achieved by simultaneously focusing on the broader dimensions of each pillar.

---

[1]Augmented reality means the use of computer vision to recognize and reconstruct objects and environments in real time, together with the use of the resulting positional data to superimpose content over a real-time image of a real-world scene, such that the content appears to a viewer to be part of the real-world scene.
[2]Virtual reality means using a combination of sensors and/or video and/or 3D graphics and/or audio technology to replicate an environment that simulates physical presence in places in the real world or imagined worlds, and enabling the user interact in that world.
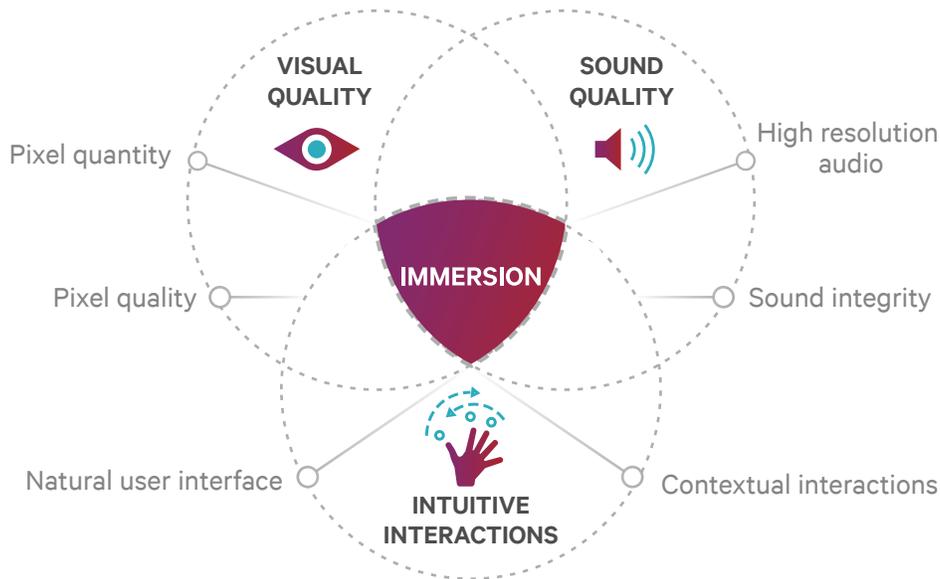
Figure 2: Full immersion by focusing on the broader dimensions

Too often, the focus has been on dimensions that do not necessarily provide the most efficient use of silicon for improving user experience.

3.1   Visual quality — focusing on pixel quality rather than just pixel quantity

The two key dimensions of visual quality are pixel quantity and pixel quality. Although pixel quantity is very important, there are diminishing returns as pixel quantity increases for specific use cases. Pixel quality is equally important for improving visual quality.

Resolution and frame rate are the two key aspects of pixel quantity (Figure 3). Resolution, which is the number of pixels in the horizontal and vertical direction, is the specification most often advertised for cameras and displays. There's been explosive growth in the number of pixels in mobile devices, with tablets and smartphones approaching 4K display resolutions and camera image sensors surpassing 20 megapixels. This improvement in pixels-per-inch has dramatically improved visual quality, since increased resolution results in increased definition and sharpness. Frame rate is a measure of how many frames are processed per second. Displays, for example, are usually refreshed at 60 frames per second. Increased frame rates reduce blurring and ghosting, which is important for fast moving objects, as seen in sports or an action movie.

Pixel quality is a function of color accuracy, contrast, and brightness (Figure 4). Improved color accuracy is achieved through an expanded color gamut, color depth, and color temperature. Color accuracy is very important, whether you are trying to pick the correct matching dress as a bridesmaid or the right paint for the living room. The color gamut is the range of colors that are able to be captured and reproduced, such as the ability to reproduce a specific hue of green. The color gamut of a display and camera is often less than what humans are a capable of seeing in the real world. With color gamuts becoming "wider" — capable of displaying more colors — it is essential to have an increased color depth, which is the number of bits used to actually represent each color. For example, 24-bit RGB color uses 8 bits to represent the red channel. Color temperature is used interchangeably with white balance, which allows a global adjustment of the intensities of the colors to make a displayed image appear to have the same

general appearance as the original scene. Contrast enables more realistic images with a wider dynamic range in light intensity. Brightness provides better viewing in high lighting conditions, like when you are at the beach on a sunny day.

Resolution

Increased definition and sharpness

Frame Rate

Reduced blurring and latency
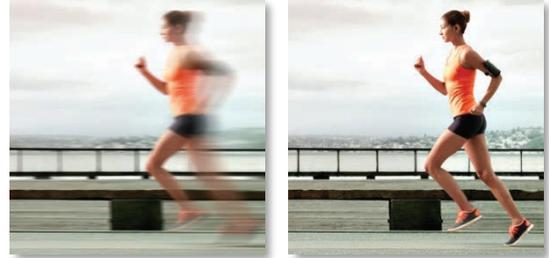
Figure 3: Resolution and frame are the two key aspects of pixel quality

Color Accuracy

More realistic colors through an expanded color gamut, depth, and temperature

Contrast and Brightness

Increased detail through a larger dynamic range and lighting enhancements

Figure 4: Pixel quality is a function of color accuracy, contrast, and brightness

Note that increased resolution and frame rate increase performance, power, and cost requirements of many components of a device, such as the SoC, memory, camera, and display. Focusing on the right dimensions and making the appropriate tradeoffs is essential. For example, focusing more on the quality of each pixel can be a more efficient way to improve visual quality.

3.2 Sound quality — high resolution audio and sound integrity

The two key dimensions of sound quality are high resolution audio and sound integrity. When the sound is realistic and matches the visual, you are truly immersed in the experience. In contrast, when the sound quality is compromised, you immediately notice it — think of a movie where the audio was muffled or mismatched the lip movements. Just as the smell of great food can enhance its perceived taste, so too does clear, realistic, 3D sound make visual user experiences more immersive.

High resolution audio has a sampling rate and precision that is high enough for the full spectrum of human hearing. Increased sampling rates capture both the low frequency sounds, such as water dripping, and high frequency sounds, such as birds chirping, so that the entire audio environment can be reproduced. Increased precision, or bits-per-sample, improve audio fidelity. More bits allow the analog sound signal to be reproduced more precisely when converted to digital.

Sound integrity includes 3D surround sound and clear audio. 3D surround sound provides realistic capture and playback of audio for personalized, immersive experiences. 3D surround sound reaches the left and right ear at the appropriate time and with the appropriate intensity so that the sound conveys the right direction, distance, and volume as it would in the real world. For example, you can hear the direction of a plane flying over your head before you see it on a movie screen or hear where an explosion is coming from in a video game. To capture and play back audio properly in surround sound, the audio industry is moving from channel-based audio to scene-based[3] and object-based audio. Scene-based audio captures the audio scene as a field of sound pressure values at all points in a space over time. Object-based audio individually captures audio from various sound sources in the scene, such as people, animals, vehicles, or weapons. Clear audio allows you to zoom and focus on the sound you want to hear while filtering out the noise by using multiples mics. This is very useful when trying to have a conversation on a phone in a noisy environment or when recording the sound at your child's concert performance.

3.3   Intuitive interactions — natural user interfaces and contextual interactions

Intuitive interactions immerse you in the experience by stimulating your senses with realistic feedback. Humans are very perceptive at noticing things that feel out of place or do not behave in a natural way, which takes you out of the moment. Natural user interfaces and contextual interactions are the two key dimensions for intuitive interactions.

Natural user interfaces, such as gestures and voice, allow you to interact with devices in the most natural way, making the interaction intuitive and efficient. User interfaces have evolved over time from punch cards, to command line interfaces with a keyboard, to graphical user interfaces with a mouse, to touch and other more natural user interfaces. Natural user interfaces should be:

- Seamless and effectively invisible, as if you aren't dealing with an interface. The appropriate user interface needs to be made available based on the user, device form factor, and application.
- Responsive, since any perceived stutter or delay takes you out of immersion and is annoying. For example, consider how a lag in the display updating can ruin an experience when swiping a touch screen, moving a mouse, or moving your head in a virtual reality headset.
- Accurate for the task at hand. For example, very accurate touch is essential for precisely editing a photo on a smartphone with high pixel density.

Contextual interactions allow devices to intelligently interact with users and provide personalized experiences based on context. Devices will be intelligent, notifying you when appropriate and blocking unnecessary interruptions. You can stay immersed and undistracted, knowing that you are not missing something important. For example, your device would block certain notifications, such as an unimportant phone call, especially while watching a movie, playing a game, or driving home. However, there will be times when you want to be interrupted due to safety, convenience, or importance. Imagine being so immersed in an experience, such as

[3]https://www.qualcomm.com/scene-based-audio

virtual reality, that you would walk into a table, miss the baby crying, or miss the doorbell ringing. Based on context, the device will know when to interrupt you, and you'll trust the device to do it appropriately.

Devices will also provide personalized experiences based on useful context to enhance experiences and remove friction. For example, imagine being in a new city, at an amusement park, or at a museum. Your device through augmented reality would suggest activities of interest or provides relevant information, such as a nearby event, restaurant, sale, or friend.

4 The optimal approach to enhance the broader dimensions

Achieving full immersion is very challenging due to the performance, power, thermal, and cost constraints of devices. The optimal way to enhance these broader dimensions of visual quality, sound quality, and intuitive interactions requires taking an end-to-end approach, utilizing heterogeneous computing, and applying cognitive technologies.

4.1 Taking an end-to-end approach for comprehensive solutions

Taking an end-to-end approach means thinking holistically at the system level[4], understanding all the challenges, and working with other companies in the ecosystem to develop comprehensive solutions. For example, maintaining color accuracy, which is a key aspect of visual quality, from camera to display requires an end-to-end approach. Consider how an image is captured and displayed, and the challenges that must be overcome.


Lens and sensor **SoC** **Display**

End-to-end solutions
Component tuning | Image processing | System optimization
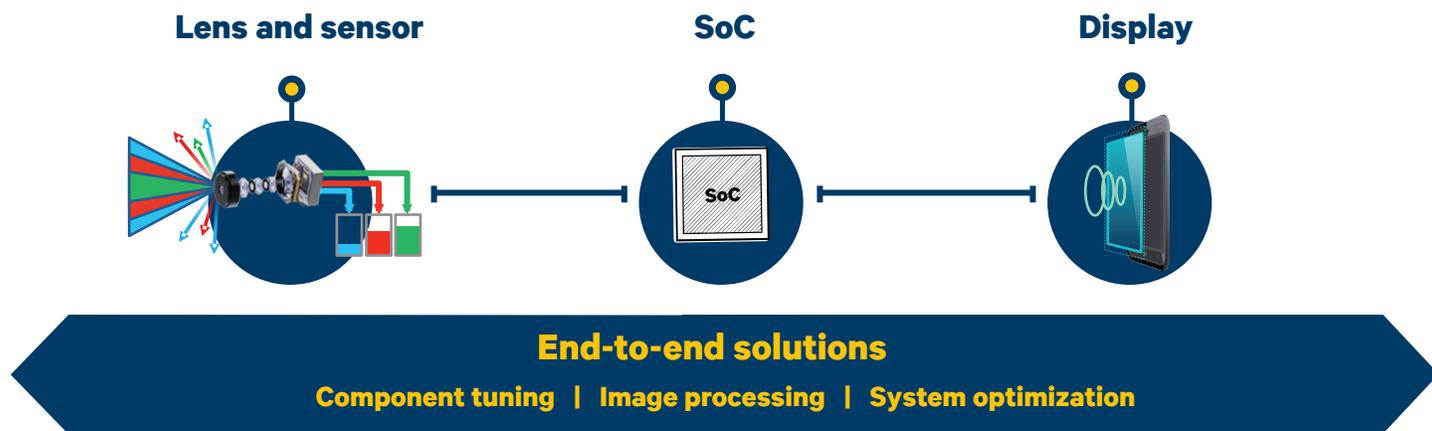
Figure 5: Taking an end-to-end approach to mobile photography

For the camera module, mobile size and cost constraints make it challenging to have large, high-quality glass lenses often found in DSLR cameras. Also, as the resolution of camera sensors has increased, the effective area on the sensor per pixel continues to decrease, resulting in less light gathering per pixel and increased noise. Similarly, there are physical and power limitations of LCD and LED displays. Displays are not yet able to emit the full color gamut that humans can perceive. There are also power constraints

that must be overcome, particularly when trying to improve the ability to view content in bright sunlight. The SoC addresses the camera and display challenges while also enhancing the visual quality through image processing.

End-to-end solutions are required for several system-level visual quality challenges besides color management, such as artifact removal and optimized click-to-shoot time. The holistic way to address these challenges requires:

- Component tuning tools to calibrate the camera and display for key visual quality aspects, such as the proper color gamut.
- Image processing, which involves preserving color accuracy and consistency across the system while efficiently enhancing and processing pixels.
- System optimization and coordination across the entire device, including hardware, software, and components. For example, an enhanced camera experience requires optimizing the latency for auto-focus, click-to-shoot, and shot-to-shot.

Similar to visual quality, sound quality and intuitive interactions also require an end-to-end approach. For sound quality, the journey of audio from the microphone to the speaker is challenging, with many opportunities for optimizations across the system. An end-to-end approach is also required to address many of the intuitive interactions requirements, specifically minimizing latency for natural UIs. For example, consider virtual reality where users interact with their head mounted display in natural ways, such as moving their head. The time from the head movement to the screen being updated, also known as "motion to display", needs to be very fast. If not, the user will see the display stuttering, making the experience less immersive and possibly making the user sick.

(4.2) Utilizing heterogeneous computing for efficiency

To efficiently enable immersive experiences within the challenging constraints of mobile devices, it is essential to utilize heterogeneous computing. Heterogeneous computing[5] runs the appropriate task on specialized engines across the SoC to meet the processing requirements of immersive experiences at low power and thermals.

For example, image processing tasks, such as computational photography[6], are crucial for visual quality and use the majority of the processing engines in the SoC. The ISP, GPU, DSP, CPU, display engine, and memory subsystem are highly utilized and must all work in harmony to efficiently enhance images and provide features, such as high dynamic range (HDR), low-light photography, and focus anywhere capabilities (Figure 6).

Natural user interfaces, such as gesture recognition, use computer vision to interpret hand movements. Computer vision runs on the ISP, CPU, GPU, and DSP to improve efficiency. In addition, voice recognition and processing would primarily run on the DSP and partially use the CPU.


■ High utilization   * Not to scale
■ No utilization
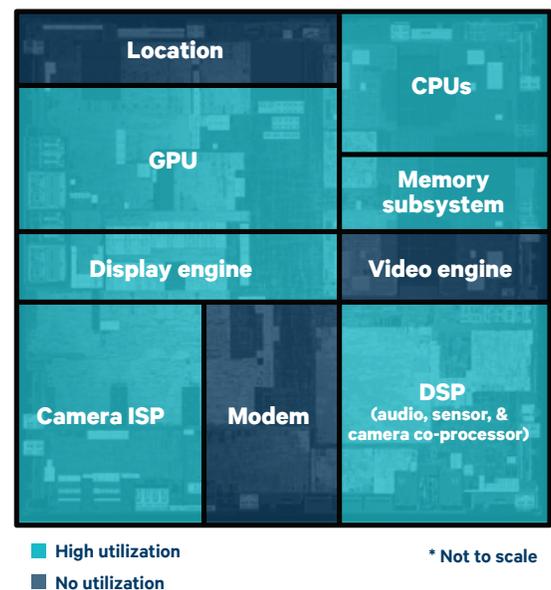
Figure 6: Utilizing heterogeneous computing for computational photography

The key point of heterogeneous computing is to increase efficiency. It should be noted that the CPU, which is often the most talked about processing engine, is only used for some tasks and often is not highly utilized.

Cognitive technologies, like machine learning and computer vision, make experiences more immersive. They enable devices to perceive, reason, and take intuitive actions so that devices can learn your preferences, personalize your experiences, and enable intuitive interactions.

For example, cognitive technologies improve visual and sound quality by automatically capturing better pixels and clearer 3D surround sound. Consider the scenario that you are at your child's play and want to record the special moment (Figure 7). Through machine learning and computer vision, your device would understand the scene — it would know that you are at a play, which child is yours, and that you are most interested in recording your child. The camera would automatically configure its settings, such as the exposure time, white balance, and depth of field, based on the environment at the play. The camera would automatically track, zoom, and focus on areas of importance, which in this case is your child.

Figure 7: Capturing immersive visuals and sound

Similarly, audio sensing allows the device to understand the environment, identify your child's voice, and then automatically adjust its settings. The microphones would track, zoom, and focus on your child's voice, removing audience noise and separating your child's voice from other sounds.

Immersive visual playback allows you to relive the moment in its full glory, personalized to how you like it. You can imagine pixels adjusting to your preferences automatically — such as your preferred color, brightness, or contrast — or based on the sunlight and ambient light to provide the best visual experience. Sound will also be much more personalized. Immersive audio playback provides personalized and realistic 3D surround sound (Figure 8). With just two speakers, it is possible to create 3D surround sound so that you hear sound coming from the correct direction, as it does in real life. Even if the device or your head moves, facial recognition and head tracking can compensate the audio playback to dynamically maintain the 3D surround sound.
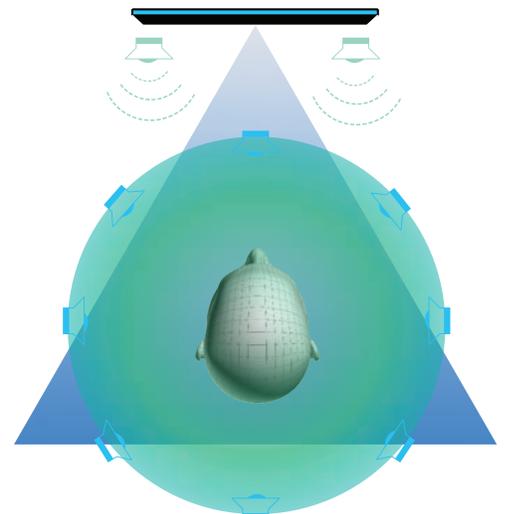

Figure 8: Immersive 3D surround sound even as the device or head moves

Interactions are being made more intuitive through cognitive technologies. Natural user interfaces use cognitive technologies to provide much more intuitive interactions that are adaptive and multimodal. For example, motion and gesture recognition use computer vision and motion sensors to recognize gestures, such as hand or head movements. In addition, the device will adapt the UI to your preferences through machine learning, enriching the experience and making it more immersive. For example, devices will adjust the screen UI to what you want to see, such as your calendar, email, weather, camera, without you even pushing a button.

Contextual awareness will make your device more intelligent and personalized, providing the right level of immersion. Contextual awareness is created through cognitive capabilities, such as scene recognition, sensor fusion, proximal awareness, and learned preferences.

5   QTI is uniquely positioned

Qualcomm Technologies is uniquely positioned to enhance the broader dimensions of immersive experiences. QTI intends to make experiences more immersive by designing efficient solutions that meet the device constraints and help the ecosystem quickly bring products to consumers. We see opportunities to improve immersive experiences by focusing on the three pillars.

- For visual quality, we are focused on consistent accurate color, in-focus images, and low-light video & photography.
- For sound quality, we are focused on realistic 3D surround sound, noise removal, and a dynamic sweet spot.
- For intuitive interactions, we are focused on seamless and responsive user interfaces, while supporting intelligent contextual interactions.

QTI is designing solutions to meet the immersive experiences requirements within the device constraints with regards to performance, power, and thermals. We are positioned to meet the unique challenges of the mobile industry in the areas of:

- Fast development cycles with customers who increasingly require more comprehensive solutions.
- Sleek, passively cooled form factors that become thinner and more challenging to design for each generation.
- Reduced cost of technologies so that our customers can deploy new and more immersive experiences to consumers worldwide.

We enable the industry to commercialize devices and experiences via Snapdragon solutions and ecosystem enablement.

5.1   Qualcomm® Snapdragon™ processors

Snapdragon processors are designed to efficiently support immersive experiences. For Snapdragon solutions, QTI has made the appropriate tradeoffs and focused on the right dimensions to design efficient SoCs. We offer custom-designed processing engines, efficient heterogeneous computing, comprehensive solutions across tiers, and cognitive computing capabilities.

Rather than licensing off-the-shelf processing engines, we have custom designed several processing engines to be optimized for specific tasks, use cases, and efficiency.

For example, the new custom Qualcomm® Adreno™ 530 GPU and Qualcomm Spectra™ camera ISP in Qualcomm Technologies' recently
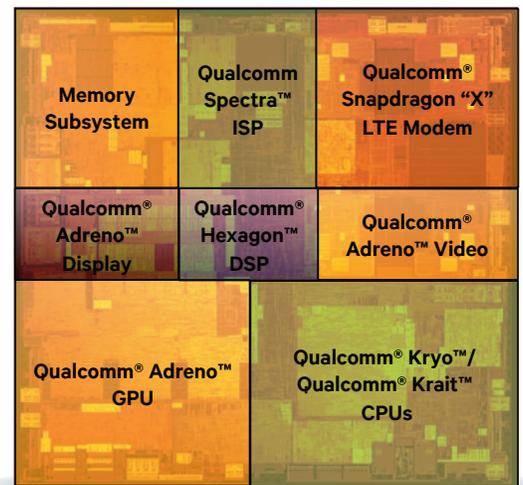
* Not to scale

Figure 9: Snapdragon processors are custom designed with specialized engines[7]

announced Snapdragon 820 processor are engineered to significantly enhance the visual processing capabilities to support next-generation immersive experiences related to computational photography, virtual reality, and photo-realistic graphics.

By custom designing superior processing engines, we also gain tremendous insight that we use to make system-level optimizations. QTI takes a system approach to design an optimal heterogeneous computing solution. We make hardware decisions based on the applications and tasks that need to be executed. Through optimized system-level software, we then run the tasks on the most appropriate engines.

QTI offers comprehensive solutions across tiers. We have four tiers of Snapdragon processors that allow customers to select the product that meets their performance, functionality, and cost needs. In addition, our software often works across tiers, so customers can reduce their development costs and commercialization time.

Qualcomm® Zeroth™ platform is our cognitive computing platform that will ship with Snapdragon processors, starting with the Snapdragon 820 processor. It is a highly-optimized hardware and software platform designed to deliver intuitive experiences and on-device cognitive capabilities by taking full advantage of the:

- Heterogeneous compute capabilities within our highly integrated Snapdragon processors
- Algorithmic innovations in machine learning, computer vision, and low power sensor processing

(5.2) Ecosystem enablement

QTI works closely with Independent Software Vendors (ISVs), Independent Hardware Vendors (IHVs), OEMs, and OS vendors to provide optimized solutions. We enable the ecosystem to quickly commercialize products through comprehensive tools and Snapdragon development platforms.

For app developers, QTI provides comprehensive content creation tools that fit within their software development toolchains. Developers often work through an iterative loop of developing, debugging, and optimizing. We have several tools for each of these areas. For example, we offer software development kits, or SDKs, for content development, such as:

- The Qualcomm® Adreno™ SDK is used for graphics and compute development.
- The FastCV™ software development kit includes a mobile-optimized computer vision library that offers the most frequently used vision processing functions.
- The Qualcomm® Hexagon™ SDK gives access to the DSP, where developers can take advantage of the real-time, efficient signal processing for tasks like audio, computer vision, and sensor fusion.

QTI also offers device optimization tools that are used primarily by OEMs to fully tune devices. These tools are necessary for system level optimization and end-to-end solutions, such as color management as described earlier.
Development devices with real silicon are very important for content creation and device optimization. Developers typically use the Mobile Development Platform (MDP) or a commercially available device powered by a Snapdragon processor to see how applications run on real hardware. IHVs often use the DragonBoard™ development kit for peripheral bring up and optimization.

6   Conclusion

Consumers want their everyday experiences to be more immersive. Full immersion can only be achieved by simultaneously focusing on the broader dimensions of visual quality, sound quality, and intuitive interactions. To meet the performance requirements necessary for next generation immersive experiences while staying within the power and thermal constraints of mobile devices, the right approach is necessary. The optimal way to enhance these broader dimensions requires taking an end-to-end approach, utilizing heterogeneous computing, and apply cognitive technologies.

Qualcomm Technologies is uniquely positioned to enhance the broader dimensions of immersive experiences by taking the optimal approach. We custom design specialized engines across the SoC and offer comprehensive ecosystem support. Enabling the next generation of immersive experiences on mobile devices is yet another example of how Qualcomm Technologies is once again re-inventing the mobile world we live in.

To get most updated information about Immersive Experiences, please visit: **www.qualcomm.com/immersive**