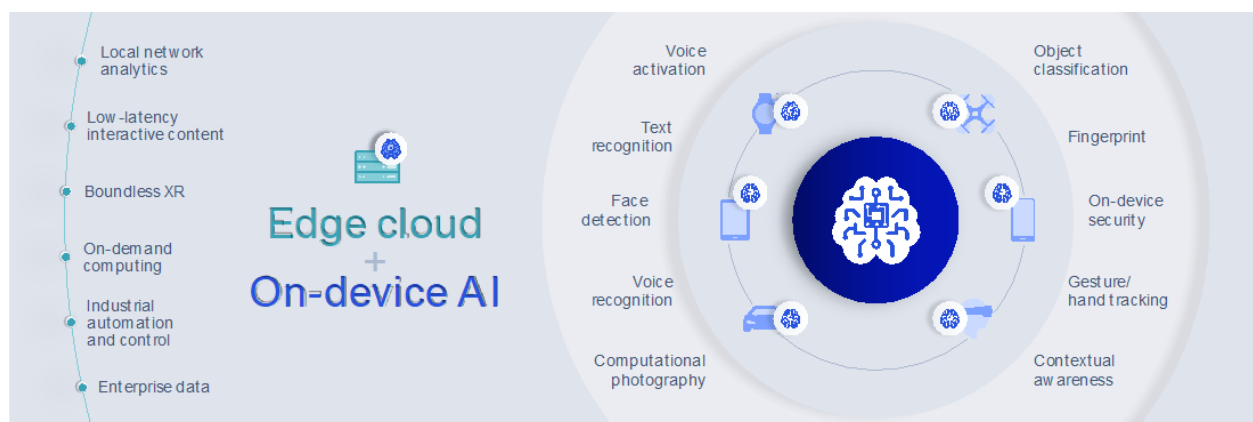


# QUALCOMM'S NEW DISTRIBUTED INTELLIGENCE PLATFORM

MOBILE CHIP LEADER ENABLES MORE ON-DEVICE AI AND PAVES THE WAY FOR DISTRIBUTED INTELLIGENCE OVER 5G

Qualcomm has been innovating artificial intelligence (AI) features for years in their Snapdragon products, and the next generation Snapdragon 865 & 765 5G Mobile Platform, due in 2020, are designed to take AI performance to the next level and enable distributed intelligence over 5G wireless. AI is becoming ubiquitous on smartphones, and although many consumers may be unaware their phones are running AI, it is becoming indispensable in primary applications such as photography, voice recognition, extended reality, gaming and even real-time spoken language translation. As a result, the battle for smartphone market share is shifting from displays to AI, and Qualcomm intends to extend its lead with these new fast chips and software.

FIGURE 1: QUALCOMM'S VISION FOR DISTRIBUTED INTELLIGENCE



Qualcomm's vision for distributed intelligence depends on fast smartphones, with performance on the edge rising to the level needed for tasks such as on-device voice recognition, hand-tracking, and contextual awareness while connecting to the cloud over 5G wireless networking.

Source: Qualcomm

## DISTRIBUTED INTELLIGENCE BEGINS WITH A GREAT SMARTPHONE

In our [previous research note](#) exploring Qualcomm AI and 5G, Moor Insights & Strategy described the company's vision for distributed intelligence. In this vision, AI is executed close to or at the edge to maximize interactive value and minimize bothersome latency to the cloud, while seamlessly harnessing the cloud to access more computational

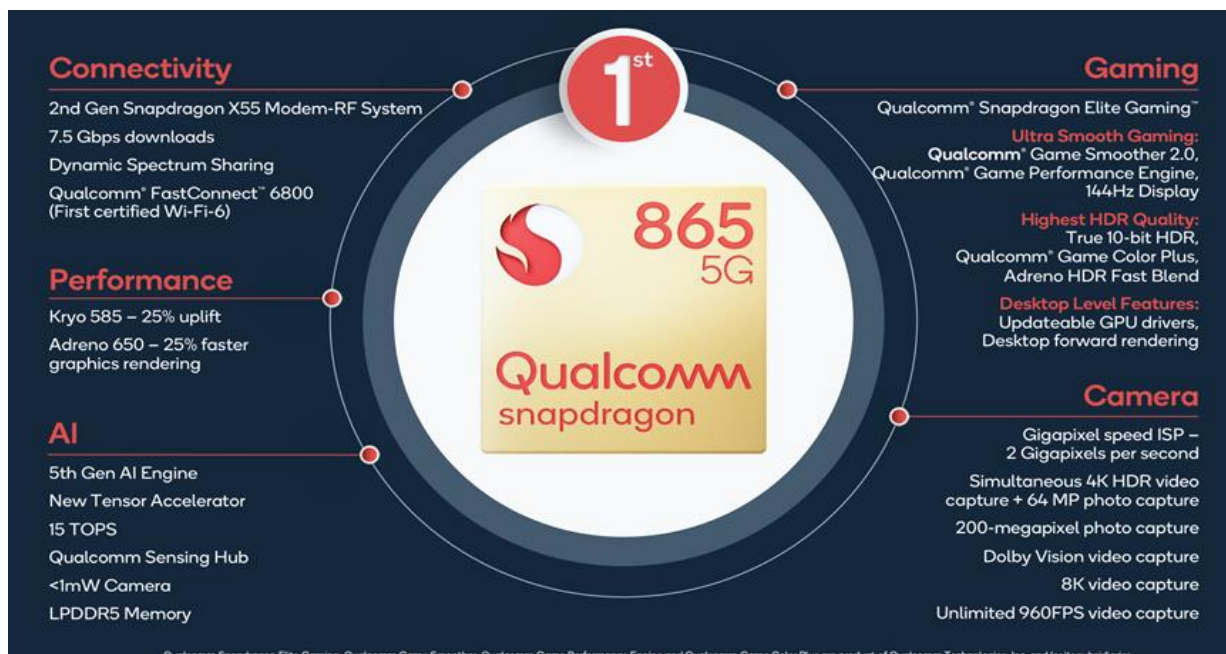
power and robust data sets. This paper analyzes how the newly announced Snapdragon 765 and 865 should help deliver on that vision.

Given Qualcomm’s portfolio in both mobile AI technology and 5G networks, Moor Insights & Strategy believes Qualcomm enjoys a significant lead on its closest challengers. It is the only company to have announced its intent to build distributed intelligence, extending its mobile technology into the data center for AI accelerators.

## INTRODUCING THE NEW SNAPDRAGON FAMILY

Typically, semiconductor vendors introduce a new architecture with one product at the high end, then cascade down to lower price points over time. However, to accelerate the adoption ramp of 5G, Qualcomm launched both the premium Snapdragon 865 and midrange 765 simultaneously, with handsets expected to be available in the first half of 2020. This decision builds on Qualcomm’s momentum in 5G, where it already has over 40 carriers and 230 5G design wins, and has deployed in more than 20 countries.

FIGURE 2: THE SNAPDRAGON 865 FEATURES



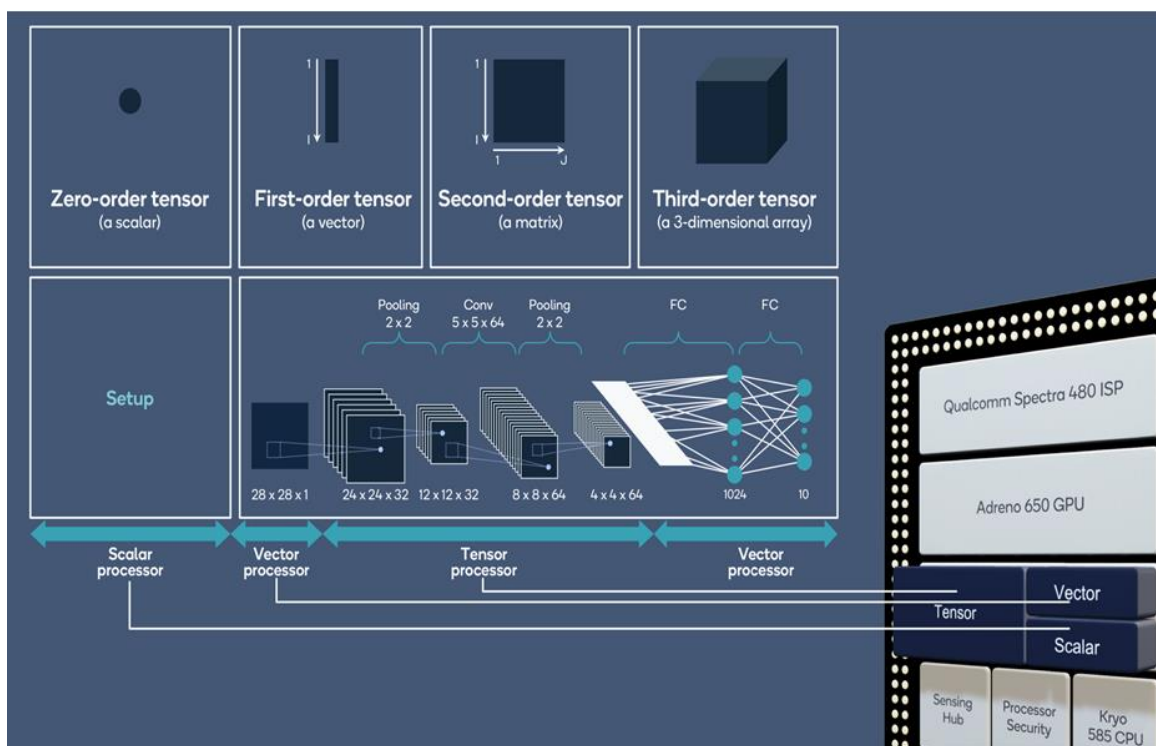
The flagship Snapdragon 865 features technology for the industry’s most advanced smartphones, as well as embedded and datacenter intelligence.

Source: Qualcomm

The flagship Snapdragon 865 model improves features and performance over the already impressive Snapdragon 855, which powers smartphones such as the Google Pixel 4, and doubles the performance of the embedded AI engine. The Snapdragon 865 is manufactured by TSMC, while the Snapdragon 765 is built on Samsung’s 7nm fab line. The Snapdragon 865 uses an off-die, second-generation X55 5G modem, and the mid-range Snapdragon 765 includes the X52 on the Snapdragon die.

By utilizing an off-die modem in the high-end variant, all the die area on the 865 is dedicated to maximizing performance at low power and includes the company’s 5<sup>th</sup> generation AI engine. It’s important to note that the AI engine is comprised of the 8-core Kryo ARM CPU, the Adreno GPU, the Hexagon Processor and the newly supported LP-DDR5 memory. It sits alongside the Spectra image signal processor for up to 7 (yes seven!) cameras and a security module.

**FIGURE 3: ACCELERATORS TO OPTIMIZE PERFORMANCE**



By offering a wide variety of efficient accelerators, Qualcomm is helping the software and DNN model developer optimize performance while minimizing power consumption.

Source: Qualcomm

The Snapdragon 865 and 765 also feature a new Sensing Hub for low-power capabilities. Beyond a typical sensor hub, which collects information from the different sensors in the smartphone, Qualcomm added capabilities to allow the Sensing Hub to collect data for audio, voice, Wi-Fi, location, and even camera capabilities at less than 1mW – virtually enabling always-on operation. The DSP within the Sensing hub can also support AI algorithms at low power for various use cases. Always-on will facilitate new ease-of-use enhancements, while the integrated sensing hub will simplify software development. The optional low-power camera will give a smartphone always-on sight without compromising privacy. One of the many things possible with a low-power camera includes recognizing QR codes in standby mode or face authentication without first waking the phone.

The AI engine performance in the new generation increases from 7 to 15 Trillion Operations Per Second (TOPS), and has an enhanced tensor accelerator joining the scalar and vector extensions to process deep neural networks. Both the GPU and Hexagon DSP now support mixed precision to deliver more efficient AI processing (16- and 32-bit float, and 8- and 16-bit integer).

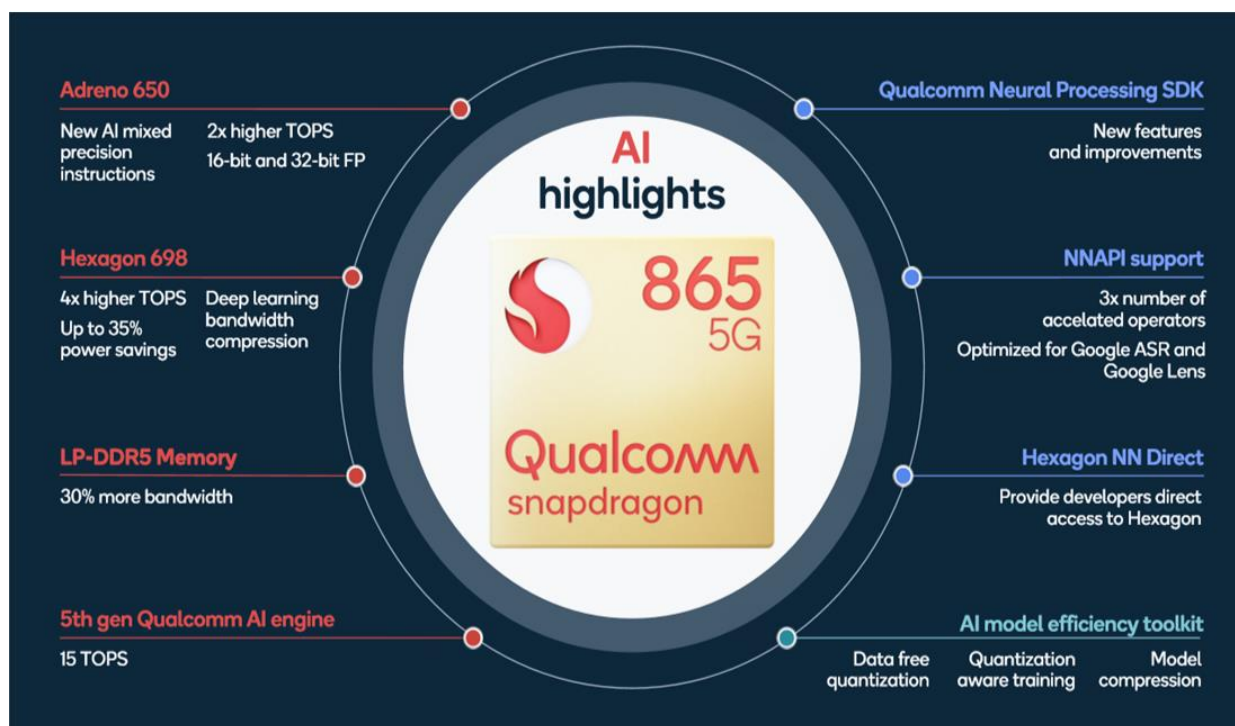
So, one might wonder why so many different types of processors are needed to do AI. The answer lies in optimizing efficiency to preserve battery life, while keeping the devices cool. By having domain-specific architectures like GPU, tensors, etc., the platform can deliver higher performance while consuming less power; a mobile phone churning out 15 TOPS must be power-efficient. Benchmark testing of the new Snapdragon platform conducted by Anandtech at the recent Snapdragon Tech Summit demonstrated excellent performance: <https://www.anandtech.com/show/15207/the-snapdragon-865-performance-preview-setting-the-stage-for-flagship-android-2020/4>

## SOFTWARE TO MAKE IT ALL WORK

Qualcomm offers three primary avenues for AI software developers. First, the Qualcomm Neural Processing software development kit (SDK) is OS independent and offers access to the chip's features through high-level programming constructs, and it supports a broad range of network models. For Android developers, the company also supports the Google Neural Networks Application Programming Interface (NNAPI) library, optimized for Google Advanced Search Recognition (ASR) and Google Lens. For developers needing maximum performance, Qualcomm offers Hexagon NN Direct access to the chip's instruction sets and memory.

Furthermore, the company also is expanding support for operators, an essential building block of a neural network such as convolution or pooling, to increase access even further. In addition to some 160 existing operators, Qualcomm is now adding “User Defined Operators,” providing even more flexibility and access to the platform. Finally, Qualcomm offers an AI model efficiency toolkit that improves power and performance through quantization (using lower precision math) and compression. These software solutions provide access to all major AI frameworks, including TensorFlow and PyTorch.

**FIGURE 4: QUALCOMM’S HARDWARE AND SOFTWARE DEVELOPMENT SUITE**



Source: Qualcomm

## A FEW INTERESTING EXAMPLES

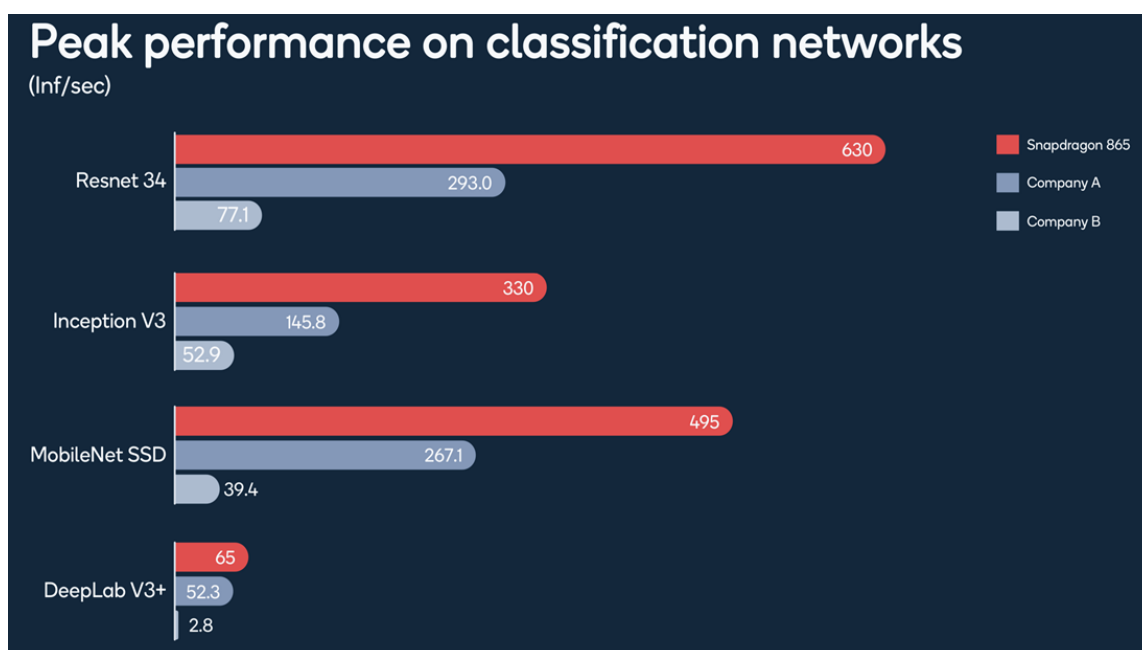
At the recent Snapdragon Tech Summit, Qualcomm showcased dozens of partners who are developing and deploying applications on Snapdragon. A few demos demonstrated the power of the AI engine on these new platforms. First, the low power required by the AI engine enables always-on environment processing—a function of the new Sensing Hub. For example, identifying music now requires one to open an application, such as Shazam, and have it start listening to a song from which it extracts information and passes that to the cloud, where Shazam’s software finds a match. With always-on, you



can avoid missed opportunities knowing that the system is always listening and capturing content for uploads.

Another good example of the power of the AI engine is how it can improve latency in voice recognition. Instead of sending voice commands to a cloud service for processing, the new Snapdragon products have enough processing power and memory to recognize most voice input and process the audio clip into text for messaging, emails, etc. — without relying on the availability of cloud AI services. Of course, users still need to ship that resulting text off to the cloud if they want to perform a web search, but we have all encountered situations with minimal or absent cloud services that would preclude the use of voice to text dictation.

**FIGURE 5: SNAPDRAGON 865 AI PERFORMANCE BENCHMARKS**



Industry benchmarks validate the competitive leadership of the Snapdragon 865 AI engine's performance.

*Source: Qualcomm*

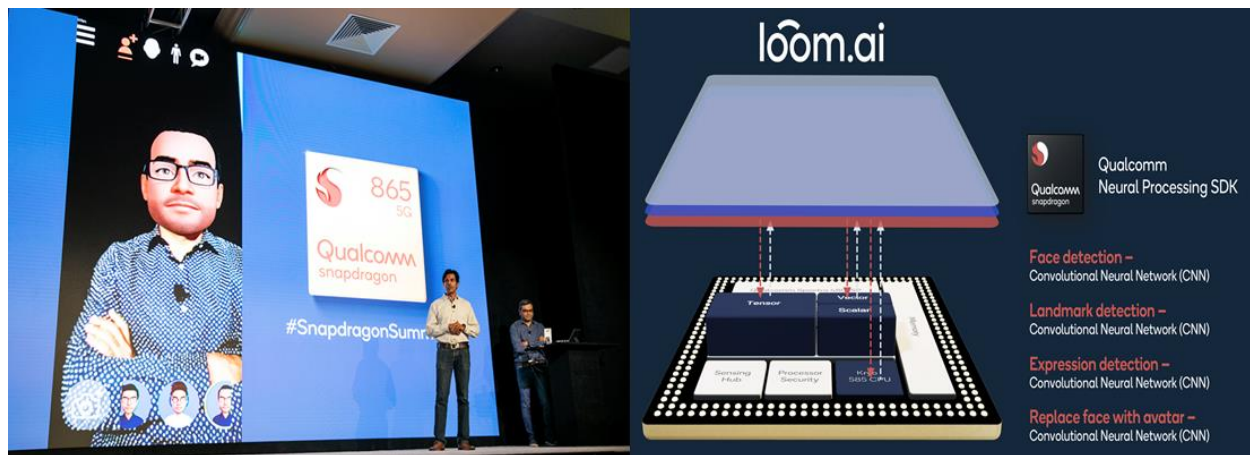
A more compelling case was a demonstration of voice to voice language translation, all performed on the device. Cloud services are available today from Google and others, but this was the first on-device translation from English to Mandarin that Moor Insights & Strategy has seen.

Another great example was Qualcomm's collaboration with Snap Inc., the creators of the Snapchat app. The company demonstrated how acceleration on the Hexagon

processor was able to deliver smooth 30fps video capture for the app, which is 4 times the speed compared to CPU, and 2.2 times compared to a GPU. The Snapchat application detects the Snapdragon hardware and utilizes a specialized Qualcomm library to achieve that capability. All of these examples are the result of the Snapdragon’s improved performance, as the benchmarks in Figure 5 attest.

A final example of advanced AI and graphics capabilities of the new platform comes from the startup Loom.ai, which uses Hollywood-derived technologies to create realistic 3D animated avatars for corporate and social media use. The phone app can create, render, and display a 3D avatar to be used, for example, in a corporate teleconference platform, running all of the graphics and AI models on the Snapdragon 865. The facial expressions of the avatar accurately track the user and render the resulting life-like image in augmented reality. With the additional horsepower of the 865, the phone can overlay multiple participants on a virtual setting.

**FIGURE 6: LOOM.AI AVATAR DEMONSTRATION**



Loom.ai uses the tensor, vector, and CPU processors on the Snapdragon platform simultaneously to run four distinct neural networks on the Snapdragon 865.

Source: Qualcomm

## DISTRIBUTED INTELLIGENCE BECOMES A REALITY

With so much performance and power efficiency on the device, will consumers still need access to cloud or edge cloud processing services? The answer lies in the amount of data and processing power for AI models, which is increasing at an exponential rate. In fact, it is doubling every three and a half months, according to open.ai. So, while many “simple” tasks such as dictation are becoming feasible on a standalone mobile device, far more complex tasks are constantly being created that will demand a tightly

integrated link between the edge and the data center. 5G will provide the ideal platform to provide that tight linkage with low latency and high bandwidth connectivity.

At the Consumer Electronics Show this year, Qualcomm announced new Snapdragon technologies that demonstrate the value of distributed intelligence. The Cloud AI 100, pre-announced last year, is working in-house and is expected to be sampled to leading customers within a few weeks. New opportunities being evaluated for the AI 100 include intelligent edge servers, Advanced Driver Assistance Systems (ADAS), and radio access network (RAN) infrastructure. In addition, provision of 5G edge services will be enabled by the new X55 5G modem. Qualcomm expects the Cloud AI 100 to be launched commercially in the second half of 2020.

Qualcomm also announced the Snapdragon Ride platform at CES, which is intended to provide ADAS and Autonomous Driving solutions to automakers. Snapdragon Ride is a comprehensive platform comprised of a family of high-performance safety SoCs, safety accelerators and a fully functional autonomous driving stack. Qualcomm believes Ride offers a combination of performance, thermal efficiency and robust software stack to allow scalability across L1 through L5 ADAS solutions. Snapdragon Ride will be available to Qualcomm's lead customers in the first half of 2020.

## CONCLUSIONS

Qualcomm has already established a leadership position in 5G and AI technologies, and is extending that lead in mobile and adjacent markets. The performance enhancements of the new Snapdragon 865 and 765 are impressive, typically ranging from 25 to 100 percent over the existing 855, which by most measures is the fastest platform in the market today. Combined with the company's second-generation 5G modem, and a growing ecosystem of over 230 design wins, we think it is safe to say that Qualcomm has won the first round of 5G silicon, worldwide, in both time-to-market and feature/function enablement. Also, the Qualcomm AI Engine forms a foundation that can scale up to create the Cloud AI100 data center platform, which Qualcomm plans to launch in 2020, allowing the company to fully realize its distributed intelligence vision and strategy. It is this vision and the execution of the distributed intelligence strategy that will continue to set Qualcomm apart from its competitors.



## IMPORTANT INFORMATION ABOUT THIS PAPER

### *CONTRIBUTOR*

[Karl Freund](#), Senior Analyst at [Moor Insights & Strategy](#)

### *PUBLISHER*

[Patrick Moorhead](#), Founder, President, & Principal Analyst at [Moor Insights & Strategy](#)

### *INQUIRIES*

[Contact us](#) if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

### *CITATIONS*

This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

### *LICENSING*

This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

### *DISCLOSURES*

This paper was commissioned by Qualcomm Technologies, Inc. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

### *DISCLAIMER*

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.

©2020 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.