5G and edge computing: Why does 5G need edge?

5G and edge computing are symbiotic technologies: they are both poised to significantly improve the performance of applications and enable huge amounts of data to be processed in real-time. 5G increases speeds by up to ten times that of 4G, edge computing reduces latency by bringing compute capabilities into the network, closer to the end user.

Anna Boyle, Senior Consultant

March 2024

5G and edge computing relationship: Powering real-time applications with speed and reduced latency

We argue that 5G needs mobile edge computing for two reasons:

1. It is inherent to 5G standards as it is the only way to meet the latency targets that have been set (1ms network latency). While telecoms operators have reported that 5G in the lab can deliver network speeds that are more than twenty times faster than LTE1, this will not reflect the experience of the average user. There are still major unknowns in how 5G will achieve these speeds – ultra-low latency standards will only be revealed in 3GPP's Release 16 later this year. We feel it is likely 5G will rely upon edge computing to reach the targets that are being set.

2. The gradual approach operators are taking to deploy 5G – the 5G go slow cycle – will mean coverage of "full 5G" will be insufficient to cultivate an ecosystem of new applications. However, edge could seed a 5G market even before widespread coverage.

What is the 5G go slow cycle?

Operators in different markets have diverse approaches towards 5G. research shows operators in the US, China, Korea and Japan are chasing 5G most enthusiastically and back that enthusiasm with real investment.

This is characterised by T-Mobile's CTO Neville Ray: *"We are all in on 5G. Every dollar we spend is a 5G dollar, and our agreement with Nokia underscores the kind of investment we're making to bring customers a mobile, nationwide 5G network."*

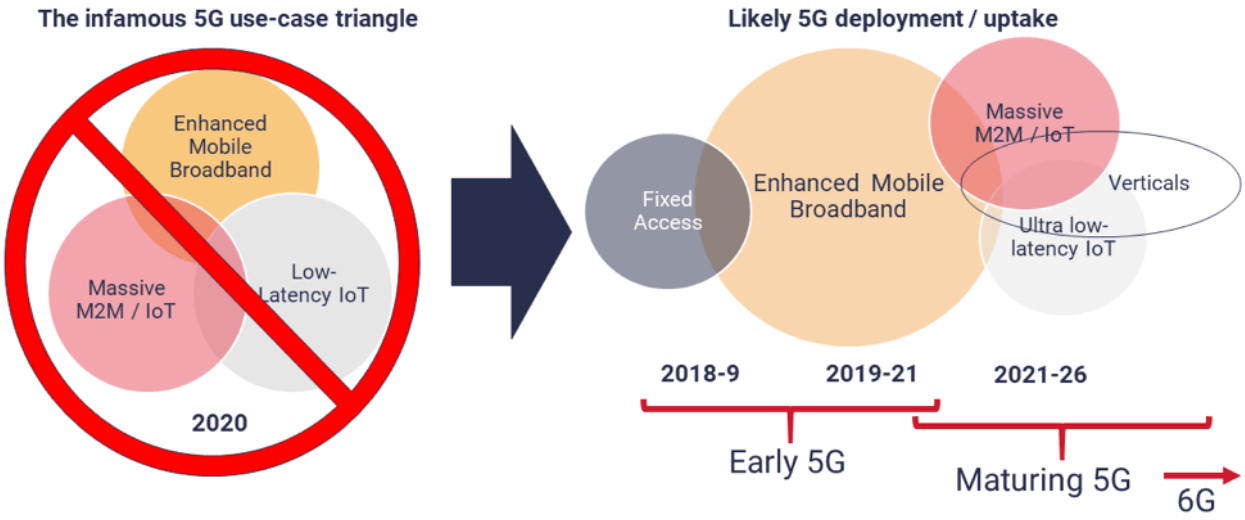
Initially, operators in Europe, such as Orange, Telefonica, Veon, Deutsche Telekom, BT, Telenor and Vodafone, did not hold back from taking part in the 5G jamboree but were more muted in their approach. Most importantly, they continue to invest heavily in their LTE networks (for example in LTE-A Pro) in their pursuit of a "glide path" to 5G. More recently, some have announced early 5G deployments in major cities, such as BT (EE) and Vodafone.

BT's Consumer Division CEO, Marc Allera, talks wisely about learning lessons from the past

"3G technology was overhyped, the price was overhyped and the consumer was underwhelmed by the experience. That said, we learnt some valuable lessons, and opened people's eyes to doing more on their mobile. 3G taught us a lesson in financial prudence while 4G taught us how to deploy rapidly at scale."

Operators in South East Asia, the Middle East and Europe are (privately) cautious about the prospects for rapid roll-out of broad coverage. They think 5G offers compelling cost advantages over LTE for enhanced mobile broadband (eMBB) in denser urban locations and potentially fixed wireless access (FWA) in certain suburban ones (where fibre has not been widely adopted). We set this out in our recent report: **5G: 'Just another G' – yet a catalyst of change**.

Figure 1: Many 5G benefits will not been delivered until 2025 and beyond >



The timeline for 5G standards, development and deployment is staggered. Not all use cases will be commercially available at the same time. This creates a predicament over how to expand beyond these 5G islands. Without wider coverage, how can operators catalyse ecosystems to invest in new services (low-latency, immersive or Massive MIMO)?

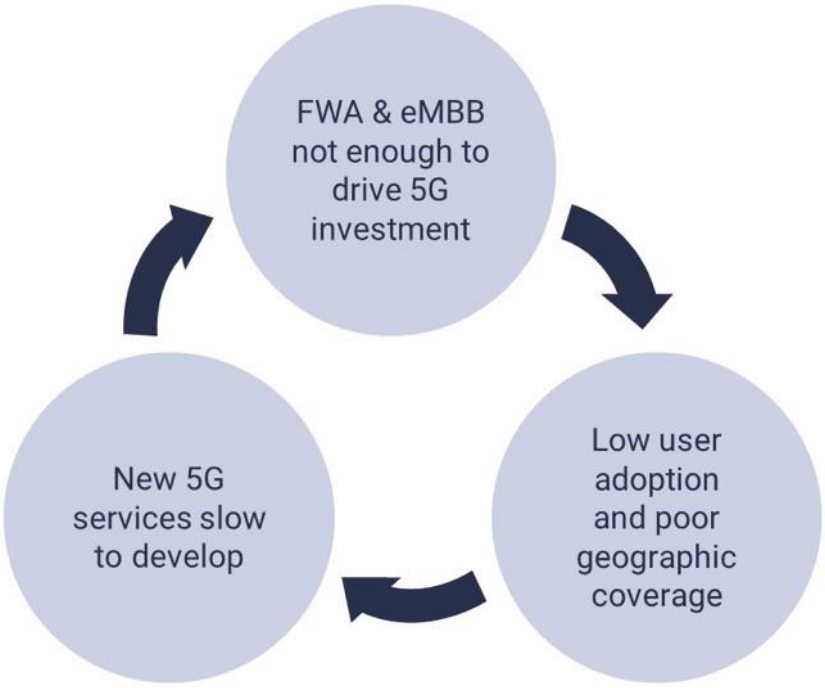
Plus, much of the promise of 5G rests on being able to enable new use cases that are not just a step up from LTE mobile broadband, for example those that are characterised as ultra-low latency and massive M2M. Enterprises are already demanding these capabilities, yet the telecoms industry is still trying to agree standards for these later releases of 5G.

The role of 5G in the manufacturing industry – $740Bn opportunity by 2030

At the same time, without coverage of 5G in these new use case domains and end-users with 5G devices, application owners will not be incentivised to develop applications that use the technology.

These questions, particularly when considering the massive CAPEX investment that 5G requires, have only be exacerbated by the COVID-19 pandemic. At a critical time, 5G spectrum auctions in countries such as Spain and Austria have been delayed because of health and financial concern. Operators are trying to navigate uncertainty in how our global economy will recover from the virus, and many will not accelerate 5G investment and innovation until there's more certainty on how their business will be affected.

Figure 2: The 5G go slow cycle has only be exacerbated by COVID-19



How do 5G and edge computing work together?

5G and edge computing are key complementary technologies for delivering data-intensive consumer and enterprise applications like real-time inferencing for AI, cloud gaming, autonomous drones or remote telesurgery. This is because these applications require a shorter, faster pipe to transfer data

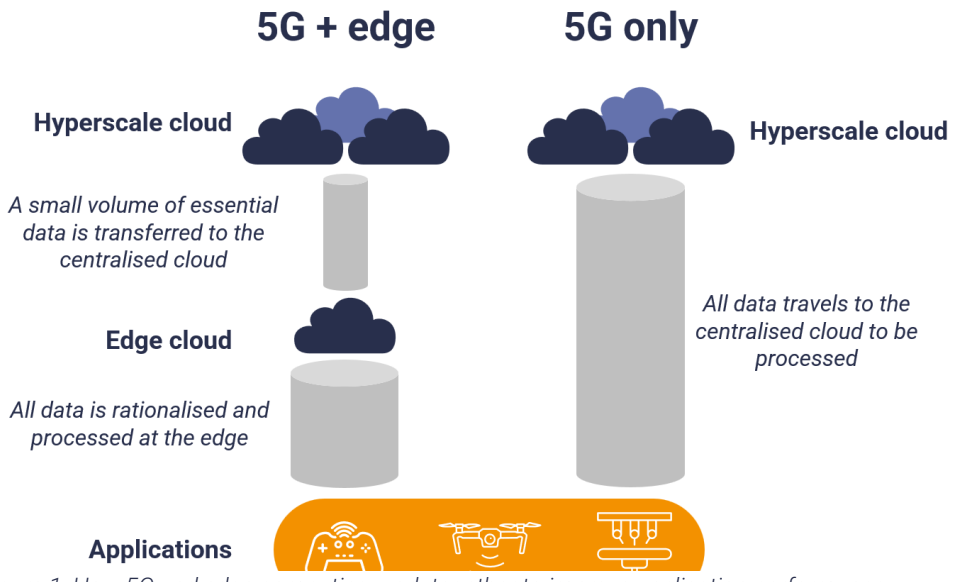


Figure 1: How 5G and edge computing work together to improve application performance

from the end-user to where data is processed to reduce latency and maintain good user experience. 5G increases the speed the data travels at, and edge computing reduces the distance it travels before it is processed. In short, edge enhances the performance of 5G.

5G and edge computing combine to support AI and GenAI applications in the most performant way

Much of the value of AI and GenAI applications is reliant on them being able to 'think' in real time. For many applications, 5G and edge computing are the best combination of technologies to enable the lowest latency to deliver real-time inferencing. Rather than hosting an AI model in the hyperscale cloud, models can be trained in the hyperscale cloud and then run at the edge, with 5G delivering fast data rates between the edge node and the end user. Read more about this in STL's report – How AI is sparking the adoption of edge computing.

Figure 3 outlines some of the enterprise AI applications where this relationship will be important.

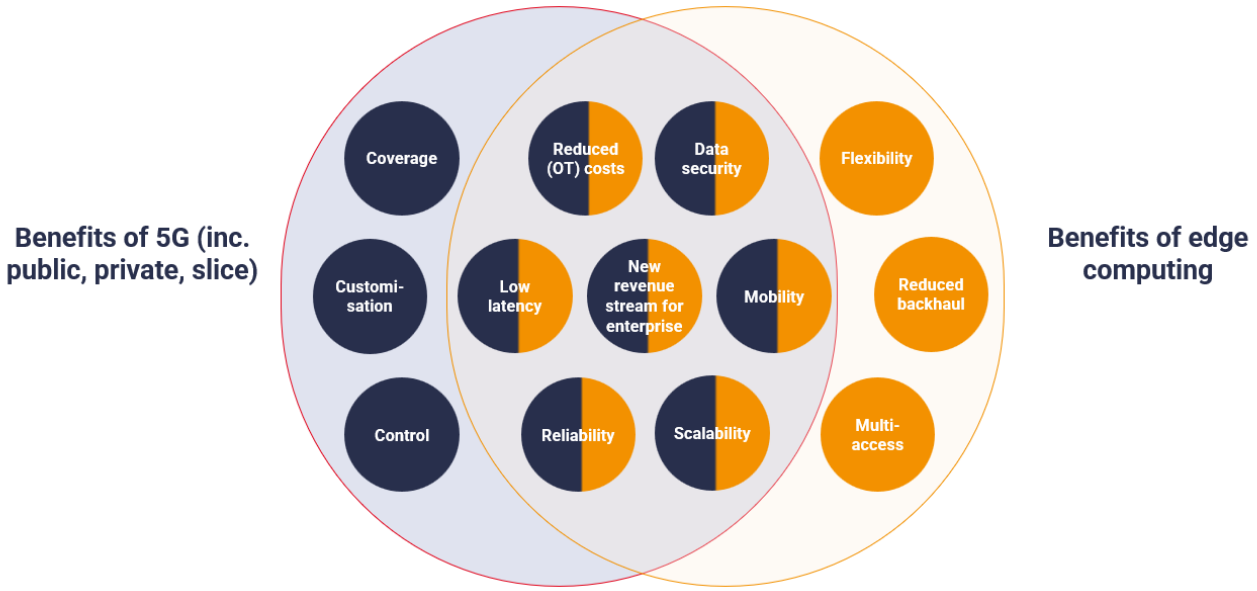


Figure 2: Benefits of 5G and edge computing

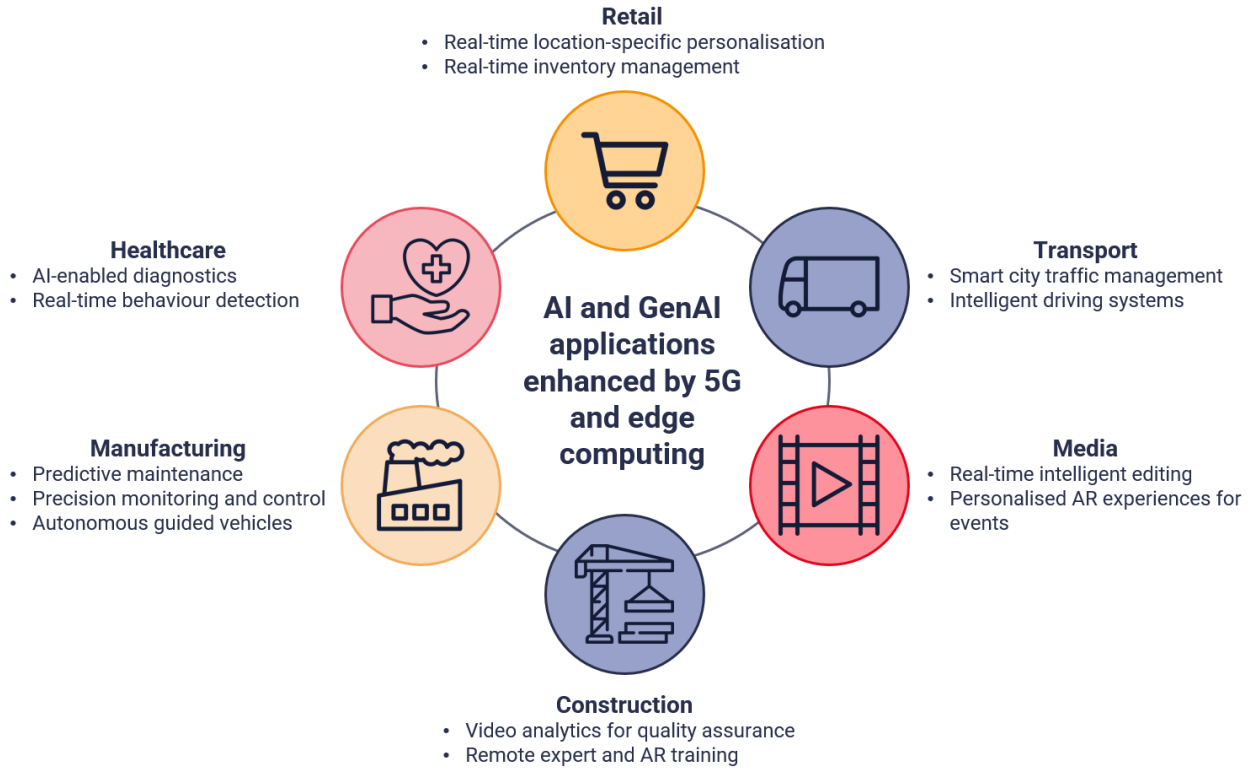


Figure 3: AI and GenAI applications enabled or enhanced by 5G and edge computing

Edge computing can attenuate the carbon impact of 5G

5G networks consume less energy than 4G or 3G networks, while performing the same functions and being faster and more reliable. For example, a 5G cell site takes just 15% of the energy of a 4G cell site to transmit the same data. For this reason, a faster transition to 5G globally could save 0.5 billion tonnes of CO_2 by 2030.

However, in the coming years there is expected to be a substantial expansion in the number of connected devices and the widespread adoption of new data-intensive applications (like AI) that 5G can facilitate, likely leading to a rise in network energy consumption.

Edge computing can attenuate the carbon impact of the huge volumes of data travelling through the network. For data-intensive applications, such as those requiring high-definition video or extensive data analysis, sending data constantly back to the cloud will be energy intensive, expensive and deteriorate the customer experience. Instead, data can be filtered out, with the full stream travelling only as far as a local edge site, before being analysed, rationalised, cleaned and only what is necessary being streamed and stored in the centralised cloud. Figure 1 demonstrates this.

Plus, smaller data centres can be easier to power with renewable energy or are more likely to be located in places where waste heat energy from servers can be reused. Companies like WindCORES, who host edge data centres of up to 30 kW inside wind turbines and Deep Green, who extract heat

from on-site edge micro data centres to heat swimming pools, have seen the commercial opportunity in sustainable edge

Edge computing reduces the cost to deliver 5G-enabled applications

Minimising the volume of data that needs to travel to the central cloud for processing directly translates into lower energy expenses for network providers. Since energy costs constitute a substantial portion of the operational expenses for these providers, the savings garnered through efficient data management can be substantial. As these operational efficiencies are realised with edge, network providers are in a position to pass on some of these cost benefits to end users. This not only enhances the competitiveness of service providers but also makes the adoption of new technologies more financially accessible for consumers and businesses alike.

Anna Boyle is a Senior Consultant at STL Partners. She has supported Tier-1 telecoms operators with their edge computing and 5G strategies. Anna sits in STL's Edge Practice covering topics including the global edge computing market, investment trends and adoption of enterprise and consumer edge computing applications.

Get in touch with the author to learn more

anna.boyle@stlpartners.com

Or visit STL Partners' Edge Hub

www.stlpartners.com/edge-computing