

# The Hyperscale Evolution of 5G



Hyperscalers have significantly ramped up their 5G activity over the past 2 years. Microsoft, Google and Amazon have each been progressing their cloud-based support for mobile network operators' core network functionality within their respective public clouds. And they've also been maturing their offerings for on-premise deployments of their edge-cloud managed services in an effort to serve the edge computing needs of 5G in enterprise use-cases. This progress is well documented in the public domain.

But there's a particular lack of clarity around the hyperscalers' ambitions for the Radio Access Network (RAN) itself and how they each intend to integrate RAN functionality into their offerings. Will it be enough to mimic the RANs provided by the current crop of MNOs? Or, will the hyperscalers need to push much, much further if they're to realise the full scope of potential benefits of marrying the cloud and 5G? While the direction of travel is clear, the final destination is not.

## 3 Stages of Evolution

Perhaps the clearest public articulation of a hyperscaler's intentions comes from Microsoft in their [submission](#) to the FCC's notice of inquiry on [Promoting the Deployment of 5G Open Radio Access Networks](#) last April. Microsoft laid out their vision for next generation wireless networks in terms of a 3-stage migration. The **first stage**, already well established, has seen operators exploit network function virtualization by deploying virtualized core network solutions on their own private cloud infrastructure. Microsoft claims to provide more than 100 commercial mobile operators with virtualized packet core implementations and more than 400 with virtualized voice service infrastructure.

The **second stage** sees operators, now comfortable with the reality of virtualized core functionality, migrate these functions from their own private cloud infrastructure to the public cloud. Microsoft has been making strides in this area. Their 2020 acquisitions of Metaswitch and Affirmed Networks allowed them to continue building their telecoms offering by bringing VNF and CNF engineering in-house and developing their own brand, Azure for Operators. Last June saw the [announcement](#) of their acquisition of AT&T's Network Cloud platform; both the IP and the technical expertise underpinning it. AT&T operated their own core network since their launch of 5G in 2018 and this agreement with Microsoft will see their carrier-grade Network Cloud platform subsumed into Azure for Operators. This step-change in the maturity and hardness of Azure for Operators will make this AT&T-proven IP and expertise available globally. For AT&T, and for other operators, the focus for them is now on differentiating themselves and their services by making best use of Azure for Operator's hyperscaling capabilities, rather than also trying to be experts in cloud provisioning. So far, so good.

The **third stage** of the journey is the one that's still in flux. Operators are adopting more cloud technology and it's moving towards their network edge. At the same time the hyperscalers are developing solutions to make their cloud offerings available closer and closer to the customer's premises, whether it's a telco customer or a non-telco enterprise customer. Microsoft's Azure Stack, Google's Anthos and Amazon's AWS Outposts are each being pitched as the hyperscalers' infrastructure solution for on-premise deployment.

Bringing the edge public cloud and 5G together makes a compelling case on paper. To achieve the kinds of low-latencies that many speculative 5G use-cases demand, heavy data processing - much of it AI/ML-based - and data storage needs to happen at the edge, near the radio. Each of the hyperscalers is pitching their cloud compute platform as being the right mix of compute hardware, storage and managed services.

The question they're grappling with is how to handle the RAN; go native, back the disaggregated Open RAN approach, or partner with a closed and integrated legacy provider? And in order to answer that question the hyperscalers are on a steep learning curve.

## Current State of Play

Microsoft, Google and Amazon have each announced key 5G partnerships and initiatives with mobile operators, legacy equipment providers and new Open RAN players.

Earlier this year, Microsoft announced a [partnership with Nokia](#) to integrate Nokia's Cloud RAN, Open RAN, RAN Intelligent Controller (RIC) and multi-access edge cloud (MEC) solutions with Azure. But we can also see, from their public comments to the FCC on Open RAN, that they want to go further. While Microsoft is supportive of the general objectives underpinning the development of Open RAN, their computer science-based heritage and history pushes their ambition further in a way that traditional RAN vendors may not be yet ready to embrace.

Microsoft calls for full virtualization of the RAN, e.g vCUs and vDUs, and for the creation of clean developer-ready APIs and the creation of an ecosystem of diverse RAN component vendors which would avoid vendor lock-in. In moving to a fully virtualized RAN Microsoft sees potential for their expertise to help scale RAN solutions by allowing functionality to move between the near- and far-edge depending on requirements such as latency, cost and energy. Whether Microsoft's goals are more idealistic, more ambitious and market-disruptive than Nokia's is unknown, as yet.

Google set out their [telecoms strategy](#) in 2020. In addition to providing core networking solutions through Google Cloud and providing ML/AI-driven business insights to operators, Google is also targeting 5G edge computing solutions. They've [partnered](#) with AT&T to start testing a portfolio of 5G edge computing solutions for verticals such as retail, manufacturing, transportation by bringing AT&T's network together with their technologies, including AI/ML, Kubernetes, and edge computing. And Google has also introduced the edge-targeted Anthos for Telecom, their competitor to Azure for Operators.

On the radio side, Google [announced](#) last February that they were partnering with Intel to grow their cloud-native telco solutions. Specifically, they're working with Intel to "*develop reference architectures and integrated solutions for communications service providers to accelerate their deployment of 5G and edge network solutions.*" This will see Google and Intel bring together Google Cloud and Intel's FlexRAN, along with Intel's MEC SDK - OpenNESS - to further develop Google's Anthos for Telecom offering. FlexRAN provides an L1 solution for the RAN stack however it's as yet unclear which partner Intel and Google are working with to provide L2/L3 in this venture. They also announced a joint Network Functions Validation Lab to accelerate the testing of core network functions running on the Anthos for Telecom platform. They aim to extend this lab to allow customers to test edge application strategies.

In the more conventional telco space, last June Google announced a partnership with Ericsson and the Italian operator [TIM](#) to pilot cloud solutions for edge enterprise use cases such as automotive and transportation. And, similar to the partnership between Microsoft and Nokia, Google has also partnered with [Nokia](#) to "*develop 5G solutions combining Nokia's RAN, Open RAN, and Cloud RAN, with Google's edge computing platform.*". This project will see the integration of Nokia's 5G vDU and vCU with Google's edge platform.

Amazon's MWC 2021 [virtual showcase](#) demonstrated the wide variety of partners they've been working with to progress their telco presence and build AWS support for a variety of users and use cases. At MWC, Bill Vass, AWS VP of Engineering at AWS, described the 3 laws that drive the business case for integrating 5G and the cloud; the Law of Physics (demanding processing near to the action to meet latency requirements), the Law of Economics (driving the need to store large datasets near the site of their collection for reasons of cost), and the Law of the Land (that demands data residency) as driving the design of the AWS architecture to facilitate flexible choices by their customers as to where they process data; at the edge or at the core, or in between.

While AWS has been facilitating third-parties to build on its platform, e.g. Mavenir building out an Open RAN cloud-native solution for DISH, AWS has also been involved in its own RAN efforts. Similar to Google, AWS is working with TIM in Italy. In this case AWS and TIM,

who are using Mavenir's RAN stack, are investigating the use of edge computing and 5G for a number of industrial verticals. And, just like Microsoft and Google, AWS and Nokia also [announced](#) a partnership earlier this year to explore "*how the combination of Nokia's RAN (Radio Access Network), Open RAN, Cloud RAN and edge solutions can operate seamlessly with AWS Outposts*".

From what's in the public domain it's clear to see that each of these three hyperscalers are well advanced in progressing what Microsoft have called **the third stage**. The operators' network edges and the cloud vendors' edge offerings are being brought together in various fashions, pushing the envelope of what's possible.

Whether it's Microsoft's partnership with Nokia, Amazon's work with DISH and Mavenir, or Google's work with Ericsson, a cynic could view these partnerships with existing vendors and operators as intellectually and experientially parasitic relationships; the hyperscalers are learning by partnering with those who already can. The hyperscalers know that the telecoms space, particularly at the edge, is challenging in ways that their experience in the cloud does not address. But building in-house RAN know-how is critical for the next stage.

They haven't yet, it seems, answered the question as to whether they should go native, or push for fully Open RAN-based solutions, or partner with a closed legacy provider. Indeed, it's most likely they'll choose not to choose for a while as there is still much to learn about the viability of 5G for enterprise at the edge.

## Stage 4 - Pushing the Envelope

The third stage can not be the final stage. Successfully transitioning the conventional network operations and operators to cloud-native solutions was, and is, challenging, but solving this transition doesn't mean the hyperscalers' journey has finished. Once they've reached a steady-state in their initial cloudification of 5G, they need to ask, what next? And they need to ask how do they get there?

Microsoft claimed in their filing with the FCC last April that it's "*committed to working with, not disintermediating, operators.*" But whether Microsoft or the other hyperscalers can or should proceed on this basis of not taking out the middle-man, i.e. the operators or existing vendors, may not be their choice.

The search space for lucrative 5G applications is huge. Every vertical and every environment is different. With Industry 4.0, we've already seen that it takes a specialised knowledge of the operating technologies in a given industry to understand how, or indeed if, they can benefit from new IT infrastructure such as 5G. And whatever the hyperscalers do now is only setting the groundwork for how they'll support 5G-Advanced and whatever comes after that.

In order for hyperscalers to facilitate a search of this space their platforms, whether Azure, Anthos or AWS, need to offer the engineers, developers and innovators as much insight into the systems' performance, clarity over how it achieves that performance and as much control over how it achieves that as possible. For 5G and its successors to flourish in the

cloud there needs to be a transformation in skills and know-how and the general approach to handling the RAN.

For now, the hyperscalers seem content with the basic Open RAN approach to disaggregating the RAN. The CU and DU will be accessible through well-defined APIs facilitating AI/ML-driven RICs, but they will be black boxes. And while the hyperscalers may offer a choice of interoperable black-boxes to their operator customers, right now it appears to be a choice between Nokia and Ericsson components, between black-box A or black-box B.

The ideals of clean APIs and softwarization are long accepted and adopted by tech companies. However, they are something of a minimum. In order for the hyperscalers to fully push the envelope of what's possible with 5G they may need to move beyond black boxes and work with alternative RAN solutions which are fully open, explorable and customizable.

More specifically, we'd argue that the hyperscaler ecosystem needs to start working with accessible open source RAN solutions that readily enable innovation. Furthermore, that RAN codebase needs to be demonstrably transparent; in order to lower barriers to speedy innovation, and to enable a new generation of RAN developers, the RAN codebase must be readable and understandable. The purpose of each line, each class and each component should be clear and mappable to the specifications. Beyond that, the code should embrace a modular architecture that actively supports and anticipates the evolution of functionality of the virtualized RAN by a diverse community. And lastly, such a code base must also be fit for commercial use; it has to be deployable, reliable and performant on the kinds of systems the hyperscalers are developing.

Some efforts are underway to create this kind of 5G platform software; both [US GOV OPS](#) and the [O-RAN Software Community](#) are working in this area. However, neither effort is, as yet, yielding a viable open source virtualized 5G RAN. And while we at SRS are not there either, we are rapidly moving towards it. Our open source [srsRAN project](#) showcases transparency and reliability for 4G and we are building upon those foundations to address this key 5G technology gap.

Not every 5G network, or 5G RAN, is going to be supported by a hyperscaler's infrastructure. The case for the cloud is not universal. But for those 5G networks where it does make sense to marry the hyperscaler's edge cloud with the RAN, then developers need to be able to make use of every tool available. As the hyperscalers provide the tools for themselves and their customers to explore different verticals and environments in search of viable, productive and lucrative business cases they need to know exactly how those tools work and how they can be modified and optimised.

To move beyond the **third stage**, the hyperscalers will need to develop platforms that robustly and transparently allow for the exploration and demonstration of what's possible.

## About SRS

[SRS](#) is a European software radio company, founded in 2014, providing a suite of products covering 4G and 5G technologies for both UE and e/gNodeB, as well as a range of test & measurement instruments. As the world's leading provider of open-source software for 4G & 5G systems, SRS gives researchers and developers the radio technologies to build better and faster through the [srsRAN](#) project. We work with commercial partners to develop custom solutions for bespoke production networks across challenging terrestrial, air-to-ground and satellite deployments.