THE PAST, PRESENT, AND FUTURE OF CLOUD GAMING

ABIresearch
THE TECH INTELLIGENCE EXPERTS™

TABLE OF CONTENTS

INTRODUCTION

Cloud gaming, like other hyped technologies such as Extended Reality (XR) and long-term visions like the metaverse, has experienced highs and lows. At its peak, cloud gaming was billed as the future of gaming—it even served as a key argument against the Microsoft and Activision merger, granting the former too much control over cloud gaming's future and, by association, the broader gaming industry. Despite these lofty visions, the cloud gaming market has also been hampered by technical and market conditions that have resulted in some notable failures.

This whitepaper explores cloud gaming's past, identifying the advancements and changes that have resulted in a resurgence in cloud gaming, and discusses the necessary balancing act between managing technologies and market needs that must differentiate cloud gaming, rather than merely serving as a substitute or complement to existing platforms.

RETROSPECTIVE ON CLOUD GAMING'S PAST

Cloud gaming has a storied history that dates to at least the early 2000s with Finnish company G-cluster (founded in 2000), before reaching broader market awareness nearly a decade later with OnLive[2] (announced in 2009 and launched in 2010) and Gaikai[1] (beta stage in 2010 and commercial launch in 2011). Despite offering innovative (for the time) business models and features, both Gaikai and OnLive failed to resonate with consumers and became acquisition targets of Sony in 2012 and 2015, respectively.

EARLY CLOUD GAMING EXPERIENCES AND CHALLENGES

Even with adequate marketing and exposure, cloud gaming failed to resonate with end users during this early cycle, in large part due to a combination of technological shortcomings and market conditions that created a challenging environment.

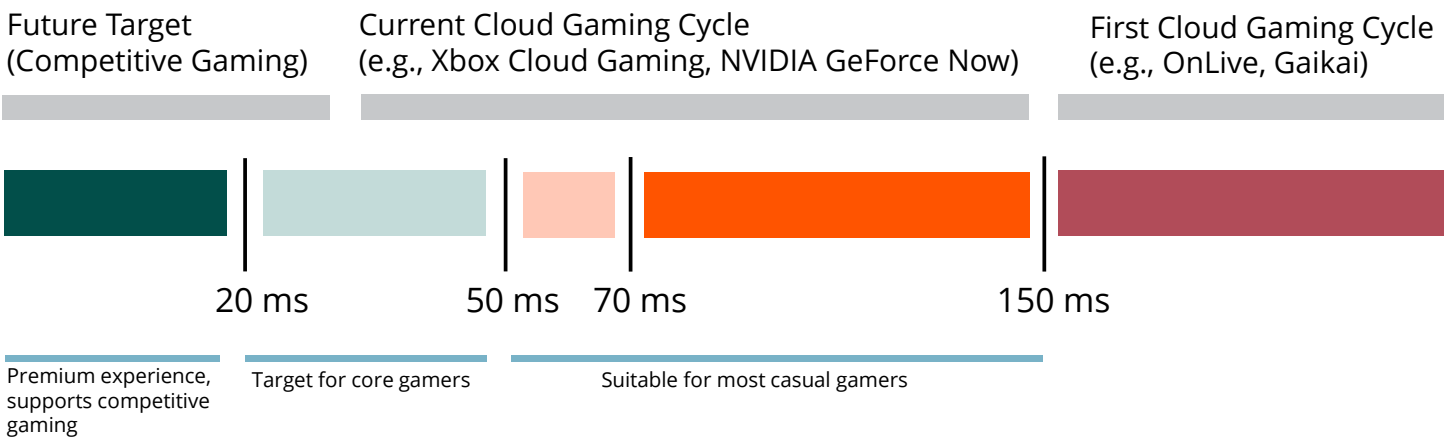
These platforms struggled to meet customer expectations due to both technology gaps and business models that did not align with prevailing market demands during this time. Earlier broadband was not fast or consistent enough to hit the necessary latency thresholds for smooth real-time gaming.

LATENCY AND PERFORMANCE ISSUES

Latency requirements can vary by game type and user perceptual variances, but there are several thresholds that are commonly identified as illustrated in Figure 1.

Figure 1: Latency Spectrum for Cloud Gaming

(Source: ABI Research)



Total input latency (from player input to rendered on screen) ideally hits 40 Milliseconds (ms) to 50 ms or less, but 150 ms is often cited as the absolute maximum latency users are willing to accept for more casual type gaming. Latency from early cloud gaming services, under some conditions did satisfy the 150 ms minimum requirement, but generally latencies fell between 150 ms and 250 ms, yielding an unsatisfactory gaming experience for most users [3,4,5,6,7].

Latency plays an important role for cloud gaming experiences, but it is important to understand when low latency is crucial. For non-interactive content that can utilize a buffer such as video streaming, including 360° video, network (cloud + connectivity) latency is generally not an issue as long as the buffer remains available; thus, this content is more tolerant to higher and inconsistent latency (except for initial application launch). Interactive content such as content-rich cloud gaming, on the other hand, is more susceptible to network latency. The response corresponding to certain user interactions, such as shooting an opponent, playing virtual football, or participating to any interactive gaming session requires low latency. In these scenarios, the visuals require a full Three-Dimensional (3D) rendering of the scene based on the additional interactions. If the network is at all involved with the 3D rendering or transporting the interaction information, then a reduction in latency is paramount to the experience.

In the case of Augmented Reality (AR)/Virtual Reality (VR) gaming experiences, latency is even more important. Head movement while wearing a Head Mounted Display (HMD) with head tracking is a specific interaction with distinct latency requirements that warrants more discussion. Motion to Photon (M2P) latency is the time between an action (e.g., head movement) and reaction (i.e., the display updated based on the movement). When a user moves their head, the brain expects an instantaneous visual and aural update, so delaying that even minutely can be problematic. M2P latency below 20 ms is currently targeted for many VR user experiences, but studies have shown that achieving M2P latency under 15 ms makes the delay imperceptible to nearly all of users. Considering the strict M2P requirements associated with HMD usage, ABI Research expects that AR and VR applications will keep M2P processing on the device. There are hybrid scenarios where 3D rendering happens off the device (the Personal Computer (PC) and console currently do this with tethered VR HMDs, while network edge or cloud rendering are expected to emerge in the future), but M2P processing remains on the device for high-quality visuals at a fixed low latency. For these hybrid scenarios, low network latency for 3D rendering will be crucial.

In addition to high latency these early cloud gaming experiences were further marred by suboptimal user interfaces and inconsistent connected devices support.

A granular analysis of end-to-end latency for cloud gaming experiences is analyzed in Section 4 of this paper.

DEVICE LANDSCAPE NOT READY

Portability of content, a key selling point for cloud gaming, had limited value in a market where mobile gaming skewed toward casual players and was still considerably smaller than the core PC and console segments. Limited peripheral and controller support also meant mobile cloud gaming was primarily done with touchscreen controls, reducing the range of suitable game types, and keeping it off most core gamers' areas of interest. Mobile data was also considerably more expensive during the earlier cloud gaming years, greatly limiting gaming time in public or incurring high costs to the user (with relatively poor service performance).

In addition to the technical challenges, cloud gaming also faced a challenging gaming market with limited game catalogs and expensive pricing structures that failed to match gamers' expectations.

CONTENT ISSUES

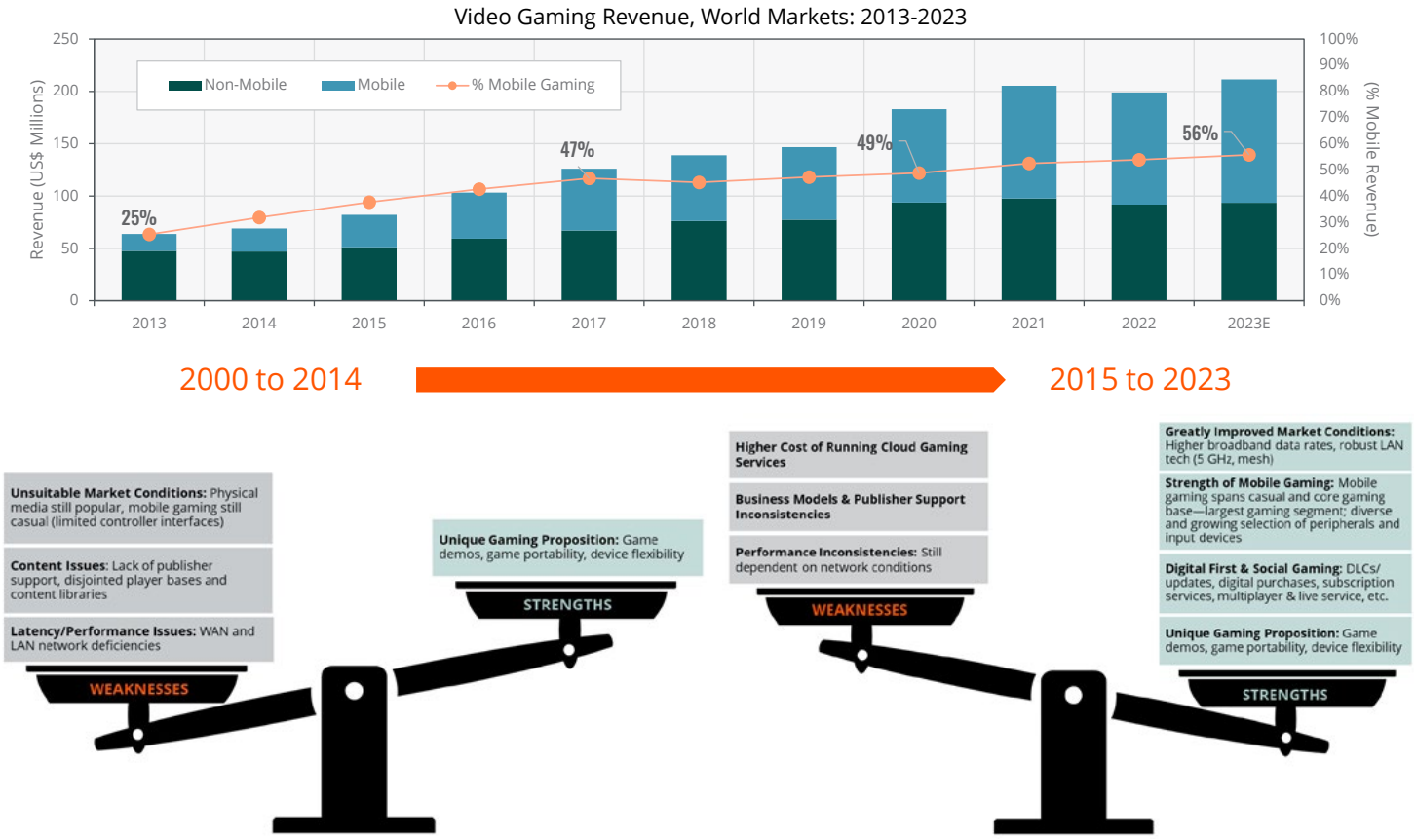
Publisher support was mixed (leaning conservative), leaving many games unavailable to cloud gaming services, particularly new launch titles. Relatedly, high reliance on first-party titles and platform exclusivity deals negatively impacted cloud gaming libraries. Furthermore, consumers could not transfer pre-existing content libraries to the cloud gaming platforms, necessitating duplicate purchases of titles. Around the time that OnLive ended its operations, however, the gaming market was already starting to shift in favor of cloud gaming, a transition that was driven by mobility.

MOBILE TECHNOLOGIES EASING THE MIGRATION TO CLOUD GAMING ENVIRONMENT

Mobile gaming, in particular, rose to prominence, and by ABI Research's estimates, shifted market share from 25% in 2013 to capturing over 50% of the market by 2021. Mobile gaming's rise is emblematic of many of the key factors that are creating a stronger foundation, shifting the scales in favor of cloud gaming.

Figure 2: Shifting the Scales in Favor of Cloud Gaming

(Source: ABI Research)



Mobile gaming is a fundamental gateway to cloud gaming adoption. The booming popularity of smartphones and tablets coupled with significant advances in hardware performance has driven

demand for easily accessible games independent of devices such as PCs and game consoles. Advances in mobile broadband services, enabled by 4G and 5G, now more reliably deliver cloud gaming at acceptable latency levels with the opportunity to exceed typical in-home installations.

Additionally, mobile chipsets are reaching console-quality fidelity, spurring demand for cross-platform gaming among core gamers. With technology barriers falling, mobile usage habits centered around app store monetization and digital distribution more broadly priming gamers for migration to seamless cross-device experiences are fueling the resurgence of cloud streaming platforms to potentially thrive where previous attempts faltered.

RESURGENCE OF CLOUD GAMING

In the decade since OnLive launched, both the video gaming market and technology landscape (beyond mobility) have drastically improved in cloud gaming's favor. Consumer behavior, business models, and key enabling technologies can now adequately, if not ideally support cloud gaming deployments.

The confluence of digital distribution, rise in mobile gaming across casual and core players, dramatically improved connectivity, and spread of connected devices have all played into cloud gaming's favor. Gamers are now more likely to identify as a consumer of a particular game type, franchise, or set of games, rather than a platform (e.g., mobile, console: Xbox/PlayStation/Nintendo, PC). Further, a growing portion of the game libraries now support cross-play and, in some cases, cross-buy opportunities (buy for one platform and gain access to other platform(s)).

Individual game titles with downloadable and seasonal content have shifted the landscape to gaming experiences and increased longevity with player engagement. At a minimum, cloud gaming services must match competitive platforms for social elements, content creation/gaming capture and sharing, multiplayer gaming, and in-game communication across platforms—this can be accomplished within a platform or through third-party support/partnerships. The broader gaming experience is also an area where cloud gaming can better differentiate. The scalability of cloud infrastructure, for example, could offer differentiable features such as advanced Artificial Intelligence (AI)-based video editing tools, cloud storage, and content management tools to easily access, modify, share, and post to video and social media services.

CURRENT MARKET ENVIRONMENT

Cloud gaming has now extended across the console, the PC, and the mobile markets. Media & entertainment leaders like Netflix are also currently in trials. Cloud companies Huawei, Tencent, Amazon Web Services (AWS), Microsoft, and Google also offer solutions catering to cloud gaming and/or launched or trialed game services themselves.

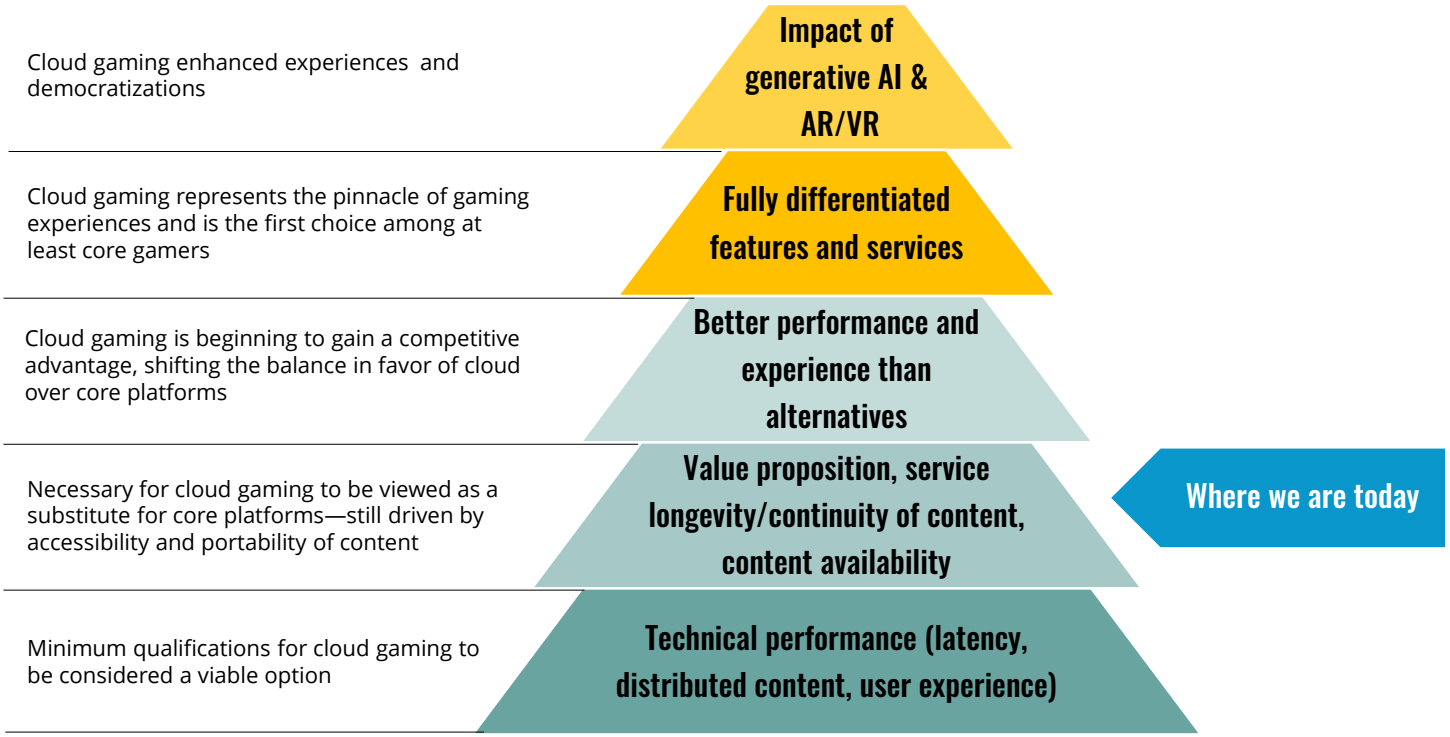
Expanding coverage, however, belies some of the challenges, pushing several cloud gaming services to shut down. Hatch (owned by Rovio), Jump, Google Stadia, and Deutsche Telekom's Magenta are among the most notable services affected.

Google Stadia's closure is perhaps the biggest indictment against cloud gaming's resurgence, but a deeper assessment suggests that its failure is tied closer to the service's shortcomings and failure to adequately address the market side of the equation beyond broader industry barriers. Hyperscalers have an advantage on the technology and infrastructure front, but are equally subject to the market needs and demands, which is where Stadia struggled. While AWS has similarly found it difficult to compete directly within the gaming industry, the synergies with its other services (e.g., Prime and Twitch) and products (e.g., Fire TV) enhance the value proposition; these considerations have also guided its regional rollouts of Luna. Similarly, Microsoft being among the early cloud gaming leaders is less about Azure than it is about its strong role within the gaming industry.

Despite setbacks and challenges, the cloud gaming industry has established a foundation to build upon. The opportunities within the next couple of years still reside within the second tier (with enhancements to the first level), where fundamental elements such as service and content longevity remain a critical need, with accessibility and portability as key drivers for value. Looking forward, the emergence of technologies such AI, generative AI, and AR/VR are likely to amplify the momentum of cloud gaming by enabling developers to create more contextualized, adaptive, immersive, personalized, and interactive experiences likely to drive adoption across a wide audience of users regardless of their demographic or social background, language they speak, or ethnic background.

Figure 3: Hierarchy of Needs for Cloud Gaming

(Source: ABI Research)



OPPORTUNITIES TODAY AND THE PATHWAY FORWARD

In today's market, core gamers are drawn to cloud gaming for mobility, the portability and accessibility of content. Accessibility, however, does not refer to substituting console/PC/mobile platforms for cloud services (e.g., to save initial hardware investments), but rather instant access to games without waiting to download and install. Microsoft reported that a majority of its Game Pass Ultimate subscribers on consoles use cloud gaming to access new purchases while the game downloads/installs; once the game is installed, the game is played locally on the machine. In this case, cloud gaming is a value-added feature, but like game demos, it is not a driver for the standalone cloud gaming services. This is further evidence of cloud gaming's need to provide value beyond the core gaming experience alone.

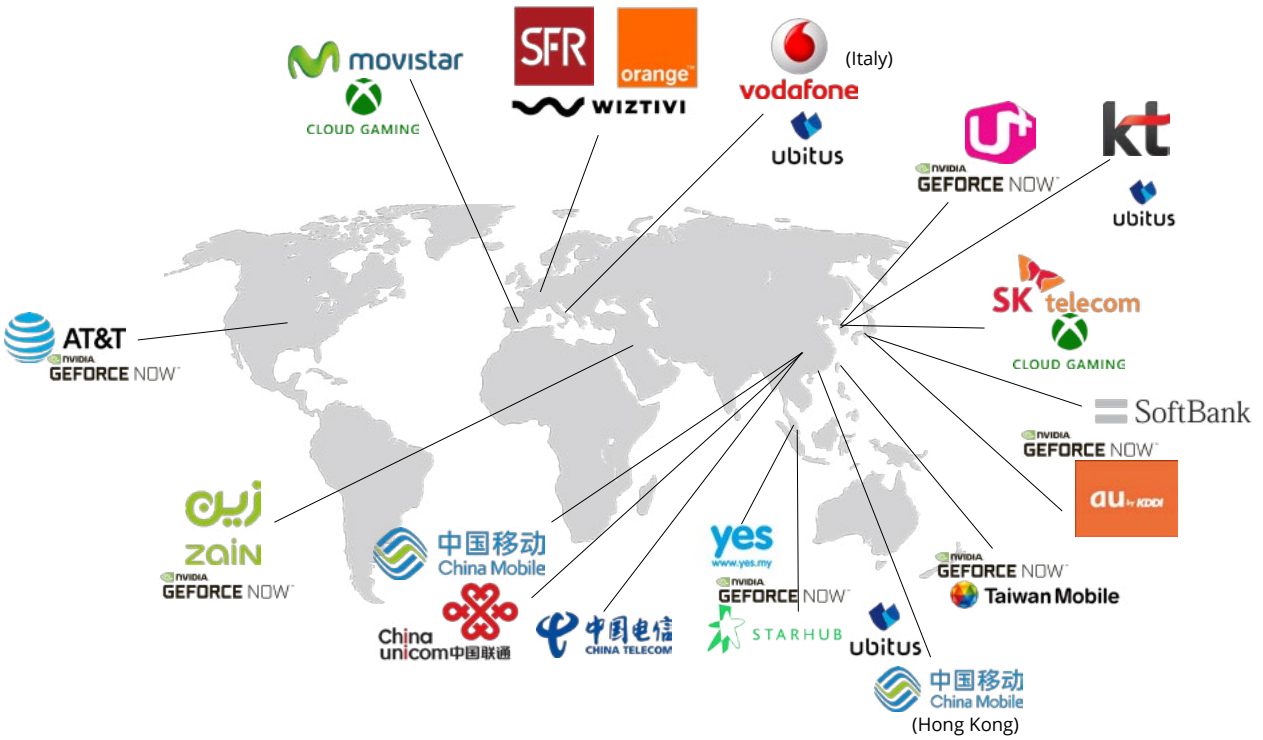
Cloud gaming offered as part of a larger bundle of content/services such as Microsoft's Game Pass or Sony's PlayStation Plus spreads the value of the subscription across multiple areas. While cloud gaming is not the main driver, it can still be financially viable because the cost structure of these packages is lower due to lower peaks of concurrent users. Even with high peaks, customers having to wait is less detrimental when cloud gaming is perceptually only part of the value proposition. Portability of content remains the key differentiator for cloud gaming at this juncture in time.

Mobility and Service Bundling

Operators have followed this trend, offering and/or bundling cloud gaming services, more recently targeting 5G deployments. Figure 4 illustrates some of the known operators that have launched or partnered to offer cloud gaming services.

Figure 4: Known Cloud Gaming Services and Partners by Operator

(Source: ABI Research)



European and Asian mobile operators, in particular, have paired 5G services with cloud gaming offerings—many of which have partnered with companies like NVIDIA, Ubitus, and Microsoft. While not all services have been successful, investments and support continue; KT, for example, is investing in its own cloud gaming technology to complement its work with Ubitus. The Asia-Pacific region has received additional attention because the competitive environment is favorable with a weaker console market and a stronger market for gaming on mobile devices.

Some mobile operators like Verizon have also trialed handheld gaming devices (e.g., the 5G-enabled Razer Edge). In Verizon's case, the Razer Edge was not a dedicated cloud gaming device (Android handheld), but with the arrival of other, lower priced (~US$300 versus US$600) handheld cloud gaming devices, adding a 5G modem could be another opportunity to package cloud gaming with 5G services. The cloud gaming service and the portable device would need to adequately differentiate themselves to justify both the upfront cost of the device and any monthly connectivity fees.

Cloud gaming companies may also be compelled to partner with Internet Service Providers (ISPs) to lower expenses. In South Korea, for example, ISPs can charge networking fees to services with high data traffic. Streaming services like Netflix and Twitch have struggled in South Korea due to the high interconnection fees—despite reducing the quality of the video streaming to cut costs, Twitch ultimately decided to leave the market. Cloud gaming services operating in this environment would similarly experience a higher cost structure, and while it appears the European Union (EU) has avoided instituting similar fees, other countries could follow suit.

While portability of content is important, if the cloud gaming market is to advance and prosper, the industry will clearly need to move up the pyramid to better differentiate these services.

TECHNOLOGY AND MARKET DEMANDS AND OPPORTUNITIES

As companies seek to move up the pyramid to better compete against incumbent gaming platforms, there is still room for improvements in core technologies and markets—latency remains a key target.

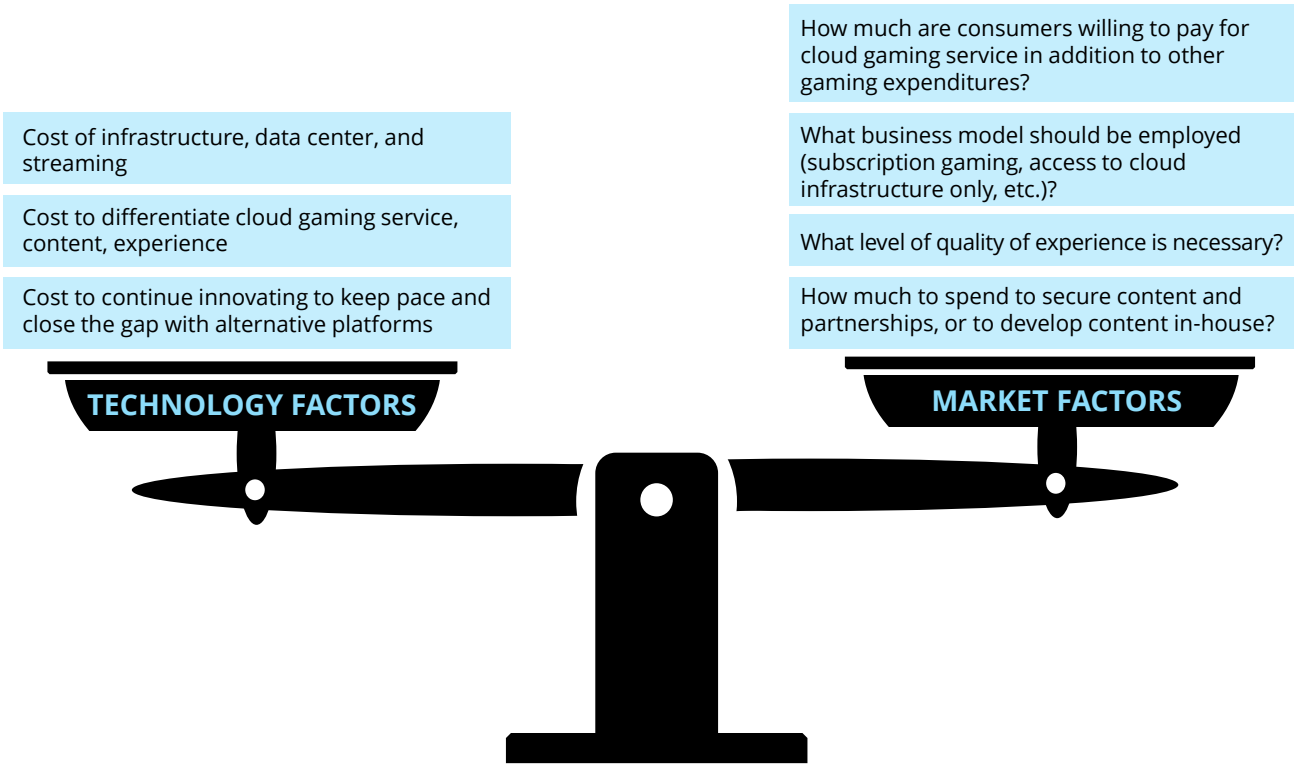
ACHIEVING LOWER LATENCY STREAMING

The perception and tolerance of latency can vary between users and the type of game being played, but cloud gaming services have been able to reduce latency levels to well within acceptable range for many users, setting (and hitting) minimum thresholds at 80 ms or less. While cloud gaming companies cannot directly control all sources of latency, efforts continue to address it within the network, processing/servers, and end devices.

As attention shifts to technology factors, it is important to keep in mind that every cloud gaming service and platform needs to evaluate and balance technology considerations against the market opportunities.

Figure 5: Cloud Gaming—Balancing Technology and Market Factors

(Source: ABI Research)



Expending significant resources to match PC or console platforms may still not result in some consumers viewing cloud gaming as a viable substitute for these platforms. There are other market factors that need to be weighed equally when assessing where to invest and what service areas to improve.

NETWORK LATENCY

At the service level, cloud gaming companies or services can partner with operators (fixed and mobile) to package services (or run operator-branded services) with higher-tier broadband subscriptions, targeting higher data rates, larger caps/unlimited, and, if available, lower latency gaming (e.g., 5G). These arrangements can increase Average Revenue per User (ARPU) and reduce churn, especially if cloud gaming is included as part of an operator's efforts to manage customers' other digital services subscriptions.

Network Slicing

Network slicing could play a role in maintaining desired levels of Quality of Service (QoS). China Unicom, for example, conducted a technical trial for network slicing and cloud gaming in Shenzhen with ZTE and Tencent. In the trial, two slices were created, a shared slice and a Very Important Person (VIP) guarantee slice that was reserved for VIP players who were able to maintain the desired QoS as per a premium Service Level Agreement (SLA).

Ericsson and Vodafone also conducted a network slicing trial for cloud gaming at Coventry University. Cloud gaming users experienced a 270% increase in throughput, 25% decrease in

latency, and 57% less jitter, resulting in participants scoring the scenario with network slicing significantly higher; 88% of trialists with network slicing gave a satisfaction score of 8-10, while just 13% (with 63% between 0 and 5) gave a similarly high score without network slicing. While these results are strongly positive, it should be noted that the sample size was relatively small at only 15 gamers. While network slicing can deliver on premium cloud gaming experiences, it will be some time before most consumers are willing to pay a premium for these types of SLAs. Distance from the data center, therefore, remains the primary way cloud gaming companies are working to reduce latency within the network.

Proximity to Data Centers

Networks typically account for a quarter to one-third of overall latency. Most efforts, to date, have worked to bring servers into ISP networks and regional data centers, particularly in densely populated regions. Meta, for example, in its Facebook cloud gaming efforts has leveraged edge computing to extend beyond its core data centers, deploying servers at edges within metropolitan areas that have access to larger populations.

The distance between data center and end users may, however, be a lower priority when target latencies have a wider range (e.g., less than 80 ms)—in contrast to a cloud gaming company that is offering a premium tier at less than 50 ms (or 20 ms) or targeting specific use cases. Multiplayer location-based gaming or immersive cloud-based experiences, for example, would require lower latencies (sub-20 ms) and could benefit from the network edge (e.g., 5G Multi-access Edge Computing (MEC)) to offer the lowest network latency possible.

The arrival of mainstream smart glasses will create opportunities to market premium services and features that leverage 5G-Advanced and, later, 6G features such as edge computing, Ultra-Reliable Low Latency Communications (URLLC), and high data rates. These devices will also serve as a catalyst for location-based gaming and a new interface for cloud gaming, allowing the user to create and size a virtual display that best suits the situation, adding only a wirelessly connected controller. VR devices (e.g., Oculus Quest 3) are already supporting cloud gaming within these virtual settings.

XR devices add additional sensors for user inputs, such as facial, body, and eye tracking that could enhance user experiences and place additional pressures on low-latency cloud streaming services. Cloud XR, more generally speaking, requires low M2P latency (sub-20 ms), which would require optimizations at all three stages, URLLC, and likely edge computing when in public spaces.

Mobility and in-public use cases push these experiences outside of the home where PCs and consoles are less viable, bringing additional opportunity for cloud gaming to capitalize.

DATA CENTER

In addition to data center location, optimizing rendering and encoding processes can also lower latency. Microsoft's cloud gaming service, for example, had comparably higher latency when

matching up against alternatives like NVIDIA's GeForce Now. The primary source of the additional latency stemmed from the rendering and encoding process. Microsoft had initially implemented a display pipeline based on the Xbox console hardware, which went through the High-Definition Multimedia Interface (HDMI) display hardware before encoding; this pathway resulted in 8 ms to 74 ms of latency. Microsoft shifted to direct capture, bypassing the HDMI pathway, and greatly reducing the latency at this stage to 2 ms to 12 ms. Microsoft made some tradeoffs moving to direct capture, such as lower supported resolution (1440p versus 4K) and lack of High Dynamic Range (HDR), reinforcing the importance of delivering latencies below 70 ms. This is a key example where a cloud gaming company properly balanced technical performance with meeting market needs.

NVIDIA is similarly employing its Reflex technology, which monitors and optimizes latency within supported games and an NVIDIA pipeline (NVIDIA Graphics Processing Unit (GPU), G-Sync display). Developers using the Reflex SDK can integrate Reflex into their games to reduce rendering latency—on the display side, select G-Sync displays that support Reflex Analyzer can measure the total latency from click to display.

DEVICES AND HOME NETWORKS

Video players can be optimized to reduce latency, along with partnering with device manufacturers to ensure displays and connected platforms are both suitable and optimized for the cloud gaming service. Partnering with a narrower group of Connected Television (CTV) vendors limits the potential target audience, but it ensures a better user experience to bolster word of mouth marketing.

Network management, Wi-Fi extenders, and mesh networking has also vastly improved the performance of Local Area Networks (LANs) throughout the home. Services like cloud gaming can also serve as a marketing tool to bolster the value of these network features.

Low latency alone, however, will not create a significant shift among many core gamers. Cloud gaming, by nature, will always have more latency compared to local gaming systems set up by the most devoted gaming hobbyists. Therefore, it is essential for the cloud gaming market to differentiate and continue to reduce operational costs.

THE IMPACT OF AI

AI will play a key role in enhancing the overall cloud gaming experience by enabling these platforms to be more responsive and adaptive with better streaming and low latency capabilities. For example, by predicting the gaming behavior and likely next step, AI can optimize and calibrate latency and QoS as a function of the game context. AI models can also fine-tune rendering and tune-up or down the streaming and graphic quality on the fly to adapt to the traffic and infrastructure conditions. AI could also facilitate personalized gameplay experiences, recommendations, and tailored difficulty settings to cloud gamers using analytics and gaming patterns to train the game. AI could also help in detecting anomalies in gameplay, stats, or activities that might suggest cheating through aimbots, wall hacks, or other techniques.

Looking forward, with the emergence of generative AI, cloud gamers will enjoy fresh and contextualized gaming content. For example, generative models can automatically create game maps with 3D objects such as trees, buildings, vehicles, and avatars that are customized to a player's preferences. This will enable more realistic and personalized gaming experiences through the cloud. Innovative, dynamic soundtracks and improved graphics using AI upscaling can further boost immersion in these cloud-streamed personalized worlds by reactively enhancing visuals and audio within games in real time.

LEVERAGING TECHNOLOGIES TO REDUCE COSTS

Computational requirements can make the cloud gaming service cost prohibitive depending on user load and construction of the service and platform. The number of concurrent users to a GPU is dependent on the type of game being rendered. In the previous cloud gaming cycle, a GPU was typically only able to only support one to two concurrent users for a core PC/console game. Today, with GPU virtualization, this has at least doubled, reaching two to four (or more) concurrent users. NVIDIA's RTX Server for cloud gaming, for example, consists of 10 twin blades, 20 CPU nodes, and 40 GPUs with GRID vGaming software to support up to 160 PC games running concurrently.

For mobile or casual games, the number of supported concurrent users was always higher, with virtualization and containerization able to push concurrent user support well above 10 users. Intel's GPUs have reportedly been well optimized for mobile gaming, allowing for values approaching 20 concurrent users per GPU for more casual games. For ,encoding cost savings can be achieved by moving to optimized hardware like Application-Specific Integrated Circuits (ASICs), such as NETINT's Smart VPU, which can support more users, lower latencies by offering direct memory access between the GPU and Vision Processing Unit (VPU), while typically operating at lower power levels.

REDUCING COST WITH AI

Real-time Deep Learning (DL) image enhancement and upscaling technologies, such as NVIDIA's Deep Learning Super Sampling (DLSS), can provide better visual experiences by improving the resolution (through upscaling) and increasing the Frames per Second (fps) of the rendered game than would be possible with the hardware alone. AMD and Intel, also active in the cloud gaming market, implemented similar technologies, FidelityFX Super Resolution and Xe Super Sampling, respectively. These technologies can provide better experiences at lower costs or allow servers to support a larger number of concurrent users.

Encoding

Efficiencies on the video encoding front can also cut costs by reducing data traffic. In the near term, new standards such as AV1 and MPEG-5 Part 2 Low Complexity Enhancement Video Coding (LCEVC), in conjunction (working as an enhancement) with a separate codec (e.g., H.264/AVC, H.265/HEVC, H.266/VCC, AV1, VP9, etc.) can reduce bitrates (up to 40%) and reduce energy costs. Looking further out on the time horizon, AI-driven efficiencies will further decrease data traffic.

Content Aware Encoding (CAE), which uses AI to identify areas of the image that are less critical to human viewing to reduce data, can yield bandwidth savings up to 50% or more. Video specialist Harmonic was an early leader in video CAE with its EyeQ technology, Apple has also generated interest following its acquisition of digital media company WaveOne in March 2023. Unlike most CAE examples, WaveOne is working on full neural network video compression, rather than implementing CAE or other AI-based encoding with pre-existing codecs (e.g., HEVC, AV1, etc.). Implementing full Neural Network (NN) video compression has the potential to deliver higher perceived quality of video over more conventional methods; a potential tradeoff is incompatibility with existing codec standards, which would require additional device/player support. To date (for video), full NN video compression has not been able to outperform conventional compression techniques, especially when AI is used with standard codecs.

While full NN video compression, with its limited support for pre-existing devices, would initially be antithetical to cloud gaming's goal of accessibility, as market support expands, it could become a viable option in the future. In the interim, CAE implemented with pre-existing codecs represents an opportunity that can still offer bandwidth savings without changing pre-existing encoding workflows, infrastructure, or end user devices. Video scaling is another option to reduce bandwidth, so long as the process does not add significant latency to the workflow. For example, a 4K video feed could be resized down to 1080p and then upscaled at the viewer's end back to 4K using metadata or encoding information contained within the stream. Samsung Research is exploring Video ScaleNets (VSNs), which implement an AI down-scaler and AI up-scaler to reduce bandwidth and deliver better results than conventional upscaling techniques alone. Samsung Research has explored VSN for video conferencing to reduce data traffic, but it is possible that techniques like VSN may not be suitable for cloud gaming applications targeting the lowest possible latencies. Like NN video compression, VSN would also require new end-device/player support.

In addition to these near-term targets and areas of improvement, there are more forward-looking opportunities that could further differentiate cloud gaming from the competitive field.

IMPLICATIONS OF NEW TECHNOLOGIES AND LONGER-TERM MARKET OPPORTUNITIES

As the cloud gaming industry moves forward, new technologies must improve the experience and correspondingly improve bottom-line performance. Pushing cloud gaming toward improving the gaming experience and providing a premium experience (in addition to accessibility) is likely to become the primary path forward.

IMPACT OF MOBILE DEVICES AND ON-DEVICE GAMING

As mobile device hardware advances and new standards such as WebGPU become established, competition at the lower tiers of gaming will become increasingly challenging.

INCREASINGLY ROBUST MOBILE HARDWARE AND WEB-BASED GAMING

Mobile devices continue to close the gap with fixed gaming hardware platforms, not only in terms of computing power, but also bringing in features like ray tracing and HDR-enabled gaming. Embedded AI (e.g., AI accelerators and Neural Processing Units (NPUs)) will also play a stronger role in mobile device use cases and, in time, mobile gaming. Mobile device platforms are already working to enable on-device Large Language Models (LLMs) that could be leveraged to enhance local gaming experiences and content personalization.

The buildup to the metaverse will place more data controls and protections with the users over centralized data centers and corporate entities. The need to support increasingly stringent privacy rules and regulations will also create stronger roles for federated learning, allowing users' personal data to remain on-device. The net effect of these hardware advancements makes earlier key value propositions such as bringing console-level gaming to mobile devices potentially less salient.

Further, the WebGPU standard (web API for GPU-accelerate graphics), was made available by default in Chrome 113 (for ChromeOS with Vulkan support, macOS, and Windows with Direct3D 12 support), and will come to Linux and Android-based devices in the near-future. While web/browser-based gaming using WebGPU will still require local hardware, it will provide greater accessibility to content without having to download and install games.

There will remain significant segments of the market where these higher-end features will not be common, but will trickle down to lower tier devices, making it increasingly difficult to compete 1:1 with cloud console and PC gaming on mobile devices. This creates a need for cloud gaming to differentiate in other ways, but it also creates opportunities for hybrid cloud/edge computing.

HYBRID EDGE/CLOUD COMPUTING

Cloud gaming workflows optimized for low-latency streaming could support hybrid computing applications by offering excess server capacity to third parties through an interoperable network of cloud servers like Open Caching for Content Delivery Networks (CDNs). Alternatively, cloud gaming services could offer browser-based gaming (that at least, in part, leverages a device's hardware) as a lower-cost option or part of the subscription (same as offering game downloads). Cloud GPU/CPU servers could also support other hybrid computing applications beyond web-based gaming.

Qualcomm, in this regard, has highlighted the potential for hybrid cloud XR in 5G environments, calling it "Boundless XR." Latency-sensitive processing would be handled on-device, while other computational tasks could be offloaded to the edge (for XR, at least) or cloud. While Qualcomm's example speaks to longer-term potential for XR and immersive-related hybrid computing environments, hybrid computing could serve other gaming applications as well. Microsoft, for example, continues exploring hybrid computing for gaming (leveraging Azure cloud infrastructure), an endeavor that was discussed as early as 2013. Hybrid gaming in this capacity could, theoretically, provide grander gaming experiences than would be possible with local hardware alone.

It is possible that a combination of web-based gaming and hybrid computing could reduce full cloud gaming's value proposition; however, it will not render the market obsolete. Both alternatives still rely on local computing, and cloud gaming will still be key for accessibility. Cloud gaming would still, at a minimum, appeal to users who do not meet system requirements and could offer the enhanced user experience offered by hybrid computing through the cloud gaming service to any device.

ENHANCING THE USER EXPERIENCE

Cloud gaming services need to make more of an effort on the markets side of the equation by providing increasingly unique experiences compared to conventional platforms. Cloud gaming services already charge premiums to access better hardware and features such as HDR and ray tracing, and while these features and performance could match higher-end tiers of gaming, it does not push cloud gaming ahead of alternatives. Rather than simply replicating the gaming experience, cloud gaming needs to provide a better, if not unique experience.

Cloud gaming companies could enable new gaming experiences that would only be possible with the scalability of the cloud. Taking this proactive step would require partnering with developers to create unique experiences and/or bring content development in-house. This strategic initiative is not novel, Google had secured developers to support Stadia and had an in-house development team (Stadia Games and Entertainment (SG&E)), but the service shut down before meaningful content could arrive, and there was little indication the games in development leveraged the cloud architecture in new and novel ways.

Cloud gaming could take social gaming and live services to new heights, creating a far more seamless experience, and leveraging additional computing capabilities and generative AI to make virtual environments more dynamic and grander in scale. Creating procedurally-generated spaces with cloud infrastructure, for example, could far outclass what is typically possible on most end devices. As the user base grows, cloud gaming services could bring social gaming to the forefront.

SOCIAL GAMING AND SPACES

Even absent a large installed base, the social element is an area worthy of development. While past platforms have attempted to build social spaces (e.g., Sony PlayStation Home, Meta Horizon Worlds, etc.), these experiences have typically been hampered by the processing power of the local hardware and, in many cases, asynchronous game libraries of its users. A key driver for gaming social spaces is the ability to meet people, socialize, and then spontaneously jump into a co-op or multiplayer game. On paper, this use case is promising, but in practice, it has been a far from seamless experience. Often, other members of the party do not have the game or at the very least have not yet downloaded and installed it. With a cloud gaming subscription service, all users would have access to the entire library and immediately jump into the game, fulfilling this vision of social gaming.

Social spaces would also engender new revenue opportunities, including marketing/advertising, sales of virtual goods and spaces, and in the case of operators, work in conjunction with their

role as a hub of their users' digital footprint. MNOs are already offering services to help their customers manage streaming services, which could extend—and be integrated—into a cloud-based social and gaming platform. Operators could also move some of their customer service efforts into these virtual spaces, leveraging digital humans and AI. As the metaverse further develops, these platforms would also serve as an optimal gateway to other metaverse spaces. Targeting these types of opportunities is important to appeal to younger audiences, who are viewing social media and content through a different lens—one that is increasingly focused on 3D experiences.

CLOUD FOR OTHER VIRTUAL EXPERIENCES

Cloud gaming infrastructure could one day be a driving force behind the metaverse and the buildup to this future. Accessibility is essential to the development of the metaverse and critical if it does, in fact, represent the future of the Internet. By necessity, computing will become more distributed, supporting applications by requirements, availability, and needs (e.g., latency). While a seamless metaverse environment would require "portaling" between virtual spaces/worlds that do not currently exist, cloud gaming infrastructure can be used to power some, if not many, of these experiences.

Virtual events platforms are already exploring some of these early use cases with cloud gaming platforms. Virbela's Campus Stream, for example, leverages NVIDIA's GeForce NOW platform to stream interactive virtual events. As more use cases and metaverse deployments occur within enterprise and industrial markets, these constitute additional opportunities for cloud gaming platforms to support additional end users.

Hyperscalers are a natural fit to power these experiences, but cloud gaming platforms, which are optimized for 3D rendering and low latency experiences, are well suited to play a key role here as well. The spread of smart glasses users will also bring these opportunities to edge computing and mobile networks for in-public mixed reality experiences. A great deal of work is needed on business models, server and user management, and interoperability, but in the interim, there are opportunities to leverage cloud gaming infrastructure for specific events or applications.

CONCLUSIONS AND RECOMMENDATIONS

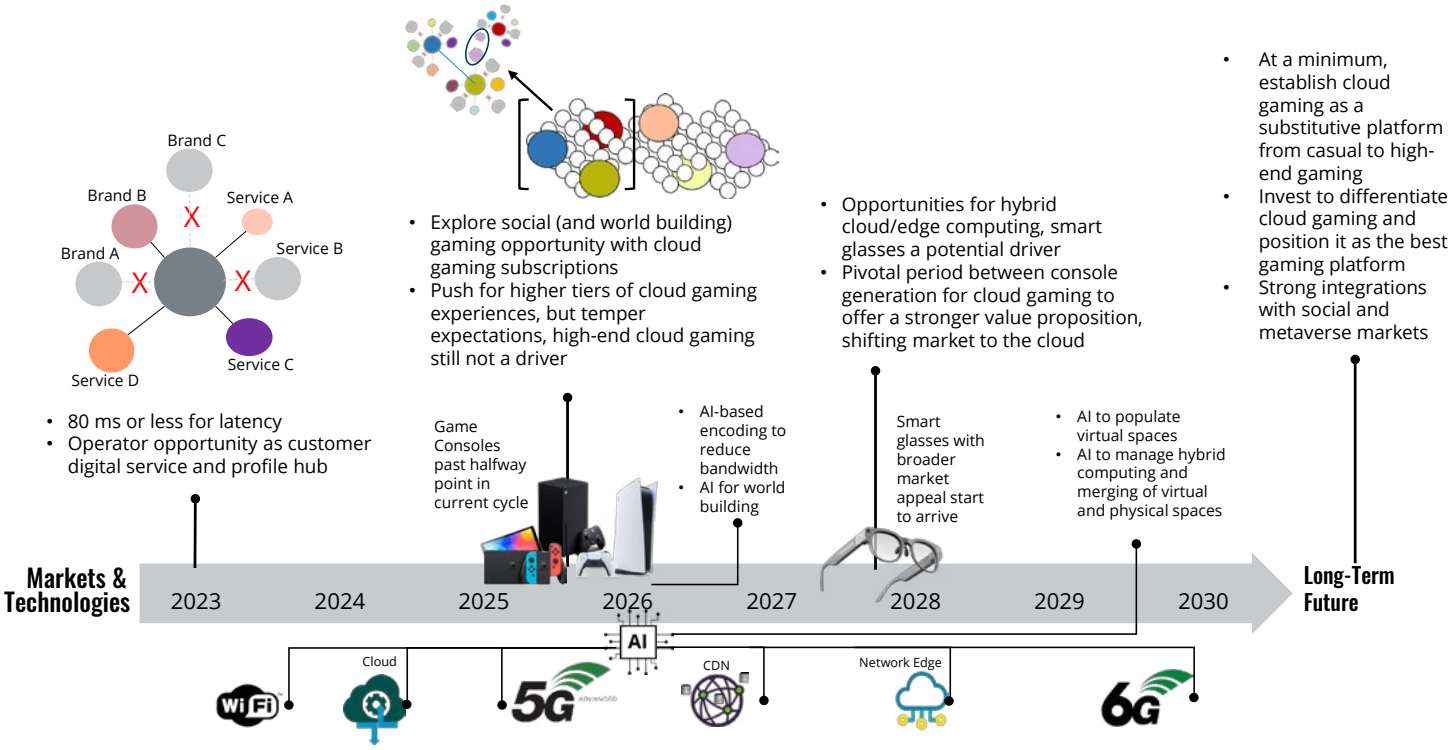
A significant amount of attention is devoted to the technical demands of cloud gaming, but equally important are the market factors that continue to prevent cloud gaming from reaching its fullest potential. While many of the technical aspects still require further development, it is not enough to simply provide an equitable gaming experience with other platforms (mobile, PC, console) without offering additional content, a better value proposition compared to conventional gaming platforms, or a differentiated experience.

Latency has vastly improved, but it remains a hurdle for some core gamers who play competitive multiplayer games and still call out latency as a limiting factor. Reflecting on the balance between technologies and markets, however, companies should question if these hardcore gamers should even be the end targets. Parallel advancements in local hardware make it difficult for cloud gaming to compete with these platforms without heavier investments. New technologies like network slicing, edge (and hybrid) computing, and AI can all contribute to closing the gap, but latency will always be higher than high-end local-based gaming.

Referring to Figure 3 and the hierarchy of needs for cloud gaming, it is possible to develop a potential timeline for when and how growth in cloud gaming can accelerate.

Figure 6: Timeline for the Development of Cloud Gaming

(Source: ABI Research)



In summary, the following are necessary for growth in the cloud gaming market to accelerate:

- Achieve a baseline target of 80 ms or less, targeting 20 ms to 70 ms for most users.
- Bring in new social elements to engage users beyond the core gaming experience. Explore ways to integrate digital content and service management into cloud gaming platforms and services.
- Focus on multiplayer and communal elements to increase word of mouth marketing, create networking effects, and increase switching costs.
- As cloud gaming populations grow, invest in first-party titles and, more importantly, live service games and gaming as a service that increases participation and user commitment to the cloud gaming platform.
- Leverage AI and new technologies to improve experiences, but more importantly reduce cost and increase efficiencies to allow further investments in upgrades and innovation.
- Track technological and market developments at the network edge and make investments when the opportunities present themselves. Smart glasses when they arrive in more mainstream-friendly form factors will jumpstart interest in hybrid edge computing and investments in cloud computing and application enablement at the edge.
- Above all else, continuously balance the technological considerations with what the market is able and willing to bear.

Momentum behind cloud gaming has picked up, but like the buildup to the metaverse, the runway for cloud gaming is extended. The length of time and necessary investments create challenges, but as is true for the metaverse, the opportunities are immense for those who properly navigate and influence the buildup to this future.

CAPITALIZING ON CLOUD GAMING OPPORTUNITIES

Beyond the obvious opportunities for game companies, the value chain from cloud to edge is poised to benefit from growth in cloud gaming, regardless of companies' direct activities within the gaming industry. NVIDIA, for example, would come out ahead if cloud gaming takes off, even if GeForce Now is not a leading platform. The same is true for hyperscalers. While many cloud services do not use the public cloud for their primary cloud gaming infrastructure, it is used to supplement and scale the service as needed. Hyperscalers like Huawei Cloud and Tencent Cloud, which offer solutions to power cloud gaming, will still have a significant role in building out this infrastructure. Hyperscalers will also play a larger role in developing video games and, more broadly, the metaverse once greater accessibility to content creation extends to more creatives in these markets.

Cloud based no-code/low-code generative AI tools, for example, will bring better parity between indie and large gaming studios. Cloud-based workstations will further help companies onboard and manage a more diverse base of developers and contributors. Other efficiency gains, such as AI-based Quality Assurance (QA) and playtesting will engender efficiencies and contribute to efforts to reduce costs. Game developers and publishers would also benefit from the growth in cloud gaming by making game development easier by reducing the number of platforms that need to be supported. These elements are necessary to both accelerate and reduce the cost to bring content to cloud gaming and, more broadly speaking, to a future metaverse.

On the market side of the equation, cloud gaming services will need to better engage gamers beyond closing the technical gap with alternative gaming platforms like the PC and console. As previously discussed, the likely importance and value of integrating social and community elements to transcend gaming will create significant opportunities for social media companies and network operators positioning themselves to play an active role in their customers' digital lives.

While the social networking opportunity has a clear relationship between gaming, social, and content creation/consumption, the opportunities for mobile operators are more nuanced, but no less important. MNOs can serve as their customers' central point of trust and service/content management hub for their digital footprints. Consumers are struggling to manage the growing number of digital subscriptions and digital footprints. To address these problems, operators like SK Telecom and Verizon are already helping their customers manage digital streaming subscriptions and other subscription services. This model can be extended to managing permissions, data collection/storage, recommendations, marketing, etc.

This would not mirror past efforts to bring together ad tech companies and operators to create new synergies, but rather a means to build trust among customers and help them both navigate and manage the changing privacy landscape. By gaining this trust, operators would have a solid foundation to bring customers into both metaverse and cloud gaming environments that require the same needs to manage content, services, and data privacy. These social spaces would also open the doorway for more cloud gaming, as users come to value the seamless transitions from social spaces into gaming environments.

Ultimately, the cloud gaming industry needs to better balance the market considerations and avenues for differentiation now that the technical elements are within satisfactory range for many users. Chasing the high end or setting premium gaming targets with the same gaming experiences creates an imbalance between spending and market opportunity. Differentiable gaming and user experiences is a better pathway to increasing revenue and extracting as much of the consumer surplus as possible.

Cloud gaming needs to market itself beyond portability of content and bring new value that positions the industry as a truly unique alternative, rather than just a complement (or simple substitute) to other platforms and services. New opportunities stemming from the buildup to the metaverse, hybrid and edge computing, and AI are potential catalysts and should be viewed as key drivers for growth. Cloud gaming has significant upside, but it will take time for additional pieces to fall into place and for consumer behavior to shift. Once this happens, cloud gaming will become a significant part, if not the future of gaming.

References:

[1] reference to Gaika: https://en.wikipedia.org/wiki/Gaikai

[2] Reference to Onlive: https://en.wikipedia.org/wiki/OnLive

[3] Kuan-Ta Chen, Yu-Chun Chang, Po-Han Tseng, Chun-Ying Huang, Chin-Laung Lei. (2011). Measuring the latency of cloud gaming systems. MM '11: Proceedings of the 19th ACM. international conference on Multimedia, Pages 1269–1272.

[4] Jiawei Sun, M. Claypool. Worcester Polytechnic Institute. 28 July 2019, Evaluating Streaming and Latency Compensation in a Cloud-based Game. Computer Science, https://web.cs.wpi.edu/~claypool/papers/drizzle/paper.pdf

[5] Ulrich Lampe , Qiong Wu, Sheip Dargutev , Ronny Hans , André Miede , and Ralf Steinmetz. Multimedia Communications Lab (KOM) Germany. September 2014. Cloud Computing and Services Science, Vol CCNS, 453 Chpate 4, p. 52-68.

[6] McManus, J. P., Day, T. G., & Mailloux, Z. J. (2019).The Effects of Latency, Bandwidth, and Packet Loss on Cloud-Based Gaming Services. https://digitalcommons.wpi.edu/iqp-all/5327

[7] M. Claypool and D. Finkel, "The effects of latency on player performance in cloud-based games," 2014 13th Annual Workshop on Network and Systems Support for Games, Nagoya, Japan, 2014, pp. 1-6, doi: 10.1109/NetGames.2014.7008964.

Published February 2024

157 Columbus Avenue, 4th Floor

New York, NY 10023

+1.516.624.2500

We Empower Technology Innovation and Strategic Implementation.

ABI Research is uniquely positioned at the intersection of end-market companies and technology solution providers, serving as the bridge that seamlessly connects these two segments by driving successful technology implementations and delivering strategies that are proven to attract and retain customers.