AECMAGAZINE

HP ZBook Ultra G1a / AMD Ryzen AI Max Pro

It's not often a mobile workstation comes along that truly rewrites the rulebook, but the HP ZBook Ultra G1a does just that, with an integrated graphics processor that has fast access to more memory than even the most powerful discrete laptop GPU, writes Greg Corke

ntel has dominated the mobile workstation sector for decades. Its processors have powered everything from ultra-compact 14-inch models to high-performance 17-inch behemoths. Today, Intel Core chips are almost always paired with discrete Nvidia laptop GPUs. AMD hardly gets a look in.

But with the new 14-inch HP ZBook Ultra Gla, and its powerful AMD Ryzen AI Max Pro 'Strix Halo' processor, that balance could be starting to shift.

The AMD Ryzen AI Max Pro comes with an integrated RDNA 3.5-based AMD Radeon GPU with professional graphics driver. But unlike integrated GPUs of the past, this one flexes some serious muscle. Together with 16 highperformance 'Zen 5' CPU cores, this means the HP ZBook Ultra Gla can deliver unprecedented levels of performance in a highly portable 14-inch form factor.

But there's more. Unlike a discrete GPU with a fixed pool of memory, the integrated AMD Radeon GPU can tap directly into up to 96 GB of high-speed system memory—far exceeding what any other laptop GPU provides. As long as the laptop is properly configured you shouldn't have to worry about datasets exceeding the

GPU memory limit, which can happen with discrete laptop GPUs with 8 GB or even 16 GB of memory. When that limit is breached, workflows can grind to a halt or worse, cause the software to crash.

The 14-inch mobile powerhouse

'14-inch' and 'powerhouse' are two words not often used together, but the HP ZBook Ultra G1a delivers the kind of performance you'd usually expect from a much larger 15-inch or 17-inch laptop.

For architects, engineers and product

designers, it's the first 14-inch mobile workstation that can truly be used for GPU-accelerated 3D visualisation. Most other 14-inch models are largely limited to 2D/3D CAD and BIM workflows.

The AMD Radeon 806OS GPU in the HP ZBook Ultra G1a—integrated into the top-tier AMD Ryzen AI Max+ Pro 395 processor—is significantly more powerful than those typically used in other

14-inch mobile workstations. Its performance is comparable to the Nvidia RTX 3000 Ada (8 GB), a discrete GPU usually reserved for larger laptops with higher power demands.

Most 14-inch laptops are limited to Nvidia RTX 500 Ada (4 GB), Nvidia RTX 1000 Ada (6 GB), or Nvidia RTX 2000 Ada (8 GB) GPUs. These entry-level GPUs are not only less powerful, on paper, but have access to far less memory - limitations that can impact demanding workflows (more on this later).

The HP ZBook Ultra G1a also marks a massive leap forward from HP's previous AMD-based mobile workstation, the HP ZBook Firefly G11A (*read our review - www.tinyurl.com/AEC-Firefly*), which comes

with 2024's AMD Ryzen Pro 8000 Series processor and integrated AMD Radeon 780M GPU. In most GPU intensive tasks, the HP ZBook Ultra G1a is around two to three times faster.

D Sallouial C D P IIII

While all eyes are on the GPU, the HP ZBook Ultra G1a also comes with a very powerful 'Zen 5' CPU with 16-cores and 32-threads. In multi-threaded workflows like rendering and simulation, this puts it head and shoulders above most 14-inch mobile workstations, which tend to come with Intel processors such as the Intel Core Ultra 9 185H, with far fewer highperformance cores. It's even on par with more powerful laptop processors like the Intel Core i9 13950HX that are only available in 15-inch and 17-inch models. However, in single threaded or lightly threaded tasks such as CAD and BIM, Intel still appears to have the lead.

AMD Ryzen AI Max Pro has twice the number of cores as the older AMD Ryzen Pro 8000 Series processor which comes with eight 'Zen 4' cores. This gives it a massive uplift in highly multi-threaded workflows. The AMD Ryzen AI Max Pro platform also supports much faster memory — 8,000MT/s LPDDR5X compared to

DDR5-5600 on the AMD Ryzen Pro 8000 Series — so memory-intensive workflows like simulation and reality modelling should get an additional boost.

Like most new laptop chips, the AMD Ryzen AI Max Pro also comes with a Neural Processing Unit (NPU), capable of dishing out 50 TOPS of AI performance, meeting Microsoft's requirements for a CoPilot+ PC.

While 50 TOPS NPUs are not uncommon, it's the amount of memory that the NPU and GPU can address that makes the AMD Ryzen AI Max Pro particularly interesting for AI. In theory, having access to large amounts of memory should allow the processor to handle large AI workloads, such as multi-billion parameter large language models (LLMs), which would not fit into the fixed memory of a discrete GPU.

AECMAGAZINE

On a more practical level for architects and designers, the chip's ability to handle large amounts of memory could offer an interesting proposition for text-to-image generators like Stable Diffusion, which are increasingly used for ideation for early stage design. It should be able to output higher resolution images without the massive slow-down that typically happens when GPU memory becomes full (*read our GPUs for Stable Diffusion article* - www.tinyurl.com/AEC-GPU-SD).

A massive pool of memory

Discrete GPUs, such as Nvidia RTX, have a fixed amount of on-board memory. In a 14-inch mobile workstation this is 4 GB, 6 GB or 8 GB. In contrast, the AMD Radeon

GPU, built into the AMD Ryzen AI Max Pro processor, has direct and fast access to a massive, unified pool of system memory. It can use up to 75% of the system's total RAM, allowing for up to 96 GB of GPU mem-

ory when the HP ZBook Ultra G1a is configured with its maximum 128 GB.

This means the mobile workstation can handle certain workloads that simply aren't possible with other laptop GPUs.

When a discrete GPU runs out of memory it must 'borrow' some from system memory, but as data must be transferred over the PCIe bus, this is highly inefficient.

Performance can drop dramatically depending on how much memory the GPU needs to borrow. Render times increase and frame rates can fall from double digits to low single digits—making it nearly impossible to navigate models or scenes. In some cases, the software may crash entirely.

The HP ZBook Ultra G1a allows users to control how much memory is allocated to the GPU. In the BIOS, simply choose a profile - from 512 MB, 4 GB, 8 GB, all the way up to 96 GB (should you have 128 GB of RAM to play with). Of course, the larger the profile, the more it eats into your system memory, so it's important to strike a balance.

The amazing thing about AMD's technology is that should the GPU run out of its ringfenced memory, then it can seamlessly borrow more from system memory, if available, temporarily expanding its capacity. Since this memory resides in the same physical location, access remains very fast. Even with the smallest 512 MB profile, when borrowing 10 GB, we found 3D performance only dropped by a few frames per second, maintaining that all-important smooth experience within the viewport.

This means that if system memory is in short supply, opting for a smaller GPU memory profile can offer more flexibility by freeing up RAM for other tasks.

Premium portability

As a laptop, the HP ZBook Ultra G1a is a very impressive piece of kit —solid, exceptionally well-built, lightweight, and slim. Weighing just 1.50 kg and 18.5mm in profile, it's the thinnest ZBook ever. Considering the sheer processing power packed inside, that's nothing short of remarkable.

14-inch mobile workstation this is 4 GB, 6 The slender chassis is made possible by GB or 8 GB. In contrast, the AMD Radeon the power efficient AMD processor,

66

'14-inch' and 'powerhouse' are two words not often used together, but the HP ZBook Ultra G1a delivers the kind of performance you'd usually expect from a much larger 15-inch or 17-inch laptop

"

which is rated at 50W. Under heavy loads it draws 70W at peak, regardless of whether you're hammering CPU or GPU. For extreme multi-tasking, power is shared between both processors, resulting in lower clock speeds.

The HP ZBook Ultra Gla comes with a 140W power supply, chunkier than most other 14-inch mobile workstations. But it's still USB-C, so you get flexibility.

For cooling there's an improved HP Vaporforce thermal system that incorporates a vapour chamber with large dual turbo fans, along with expanded rear ventilation. This keeps the system running cool and relatively quiet. Fan noise was perfectly acceptable and consistent, even when rendering for hours.

With CPU renderers - KeyShot, V-Ray, Corona Render and Cinebench - all-core frequencies ranged from 3.5 GHz to 3.7GHz. In GPU renderers - Twinmotion, D5 Render, KeyShot, and Lumion - GPU clock speeds fell between 2,600 MHz and 2,850 MHz.

Running off battery, the system's power consumption drops to 40W, leading to reduced all-core CPU frequencies of around 2.5 GHz and GPU frequencies of 1,700 MHz. At these levels, fan noise was virtually silent. The HP XL-Long Life 4-cell, 74.5 Wh polymer battery delivered 95 minutes of runtime under full GPU load in Twinmotion, and 125 minutes in the SolidWorks SPECapc benchmark, which combines demanding graphics with intensive single- and multi-threaded CPU tasks. For typical day-to-day usewhere modelling happens in short bursts and many editing tasks are relatively lightweight-you can expect significantly longer battery life. The good news is that the laptop charges quickly, reaching 50% in just 29 minutes and 80% in 55 minutes.

There are two display options: a standard FHD (1,920 × 1,080) panel, and one of the standout features of our test machine—a stunning 2,880 × 1,800 OLED

> touchscreen with a 120Hz refresh rate, 400 nits of brightness, and 100% DCI-P3 colour gamut. There's only a single NVMe TLC SSD (512 GB - 4 TB), which is hardly surprisingly given the size of the machine.

The HP ZBook Ultra Gla is well equipped with ports -two USB-C (40 Gbps) Thunderbolt 4 with DisplayPort 2.1, one USB-C (10 Gbps) with DisplayPort 2.1- all supporting power delivery - and one USB Type-A (10 Gbps) with charging support. While there's no RJ-45 Ethernet port, wired connectivity is still possible via a USB adapter, and you can also get fast data transfer from built in Wi-Fi 7. In use, the keyboard feels solid and responsive, while the large trackpad is smooth and natural. That said, serious 3D modelling is always best handled with an external mouse.

Finally, for Teams and Zoom calls, there's an impressive 5 MP IR camera with Poly Camera Pro software. Advanced features like AutoFrame, Spotlight, Background Blur, and virtual backgrounds are powered by the 50 TOPS NPU, optimising power efficiency and helping maximise battery life. The camera shutter's black-and-white stripes is a nice touch, making it immediately obvious when it's closed.

The elephant in the room is the price. If you're used to budget-friendly 14-inch mobile workstations — or assumed an integrated GPU would save you money — this might come as a shock. Our fully loaded review unit — featuring an AMD Ryzen AI Max+ Pro 395, 128 GB of RAM, 2 TB SSD, and a 14" 2.8K touch display comes in at a hefty \$4,049.

There are savings to be had, though. Dropping to a 1 TB SSD knocks off a significant \$490 (bringing it down to \$3,559), and opting for 64 GB of RAM instead of 128 GB saves another \$460 (down to \$3,099).

It certainly pays to shop around — prices vary wildly. We've even seen our exact review configuration listed on hp. com for a jaw-dropping \$9,060.

On test

We put the HP ZBook Ultra G1a to work

in a variety of real-world CAD, visualisation, simulation and reality modelling applications. Our test machine was fully loaded with the top-end AMD Ryzen AI Max+ Pro 395 processor and 128 GB of system memory, of which 64 GB was allo-

cated to the AMD Radeon 8060S GPU. All testing was done at the OLED display's maximum 2,880 x 1,800 resolution.

We compared the laptop to the previous generation HP ZBook Firefly G11A, as well as a range of desktop workstation CPUs and desktop workstation GPUs with 8 GB of memory. While we didn't have access to recent Intel-based mobile workstations for testing, the comparisons still offer valuable context—particularly in terms of GPU memory limitations.

Computer Aided Design (CAD)

In 3D CAD software Solidworks 2025 the HP ZBook Ultra G1a easily handled everything we threw at it. While the previous generation HP ZBook Firefly G11A with AMD Ryzen Pro 8000 Series processor stuttered with very large models, the HP ZBook Ultra G1a took everything in its stride. It even made light work of the colossal Maunakea Spectroscopic Explorer telescope assembly with 8,149 parts and 58.9 million triangles, delivering silky smooth model navigation. As Solidworks is largely single threaded or lightly threaded, there wasn't a huge difference between the ZBook Ultra G1a and ZBook Firefly G11A in terms of computational performance.

Visualisation

Real-time GPU rendering / visualisation is where you start to see the true benefits of the HP ZBook Ultra Gla. In workloads that require a lot of GPU memory, having access to a large pool gives it a significant performance edge over systems with memory-constrained discrete GPUs.

In many viz tools, memory usage rises significantly with output resolution of final renders. In Solidworks Visualize, for example, a small suspension model uses 1.2 GB to load but then a considerable 7.4 GB to render at 4K and 12.8 GB to render at 8K. The much larger snow bike model with 32 million polygons uses 12.4 GB to render at 4K and 19.1 GB to render at 8K.

Of course, if you load the equivalent CAD model in Solidworks at the same time, as most designers would, that adds an additional 5.0 GB to the overall GPU memory footprint. In multi-application workflows like these, when there's not enough GPU memory to go around, each app must release memory before the

66

In the HP ZBook Ultra G1a, it feels like we're witnessing a genuine shift in the landscape for 14-inch mobile workstations

"

other app can claim what it needs. This transition can cause a stutter as memory is freed and allocated. And if you're trying to render in the background while continuing to model in Solidworks, insufficient memory will likely result in much longer render times.

We found Twinmotion by Epic Games demands even more from GPU memory. The Snowdon Tower scene, for example, which is fairly typical of a mainstream arch viz dataset, uses 8.1 GB of GPU memory simply to load, 9.8 GB to render at FHD resolution, 14.2 GB to render at 4K and a whopping 35.1 GB to render at 8K!

Of course, with 64 GB of GPU memory to work with, the HP ZBook Ultra Gla handled the task with ease. But what happens when a discrete GPU runs out of memory? To render six 4K images, the 8 GB desktop GPUs were around 6 GB short, resulting in significantly longer render times.

The HP ZBook Ultra Gla completed the job in just 183 seconds, while the AMD Radeon Pro W7500 (8 GB) took 659 seconds, the Radeon Pro W7600 (8 GB) 688 seconds, and the Nvidia RTX A1000 (8 GB) 799 seconds.

The important thing to note here is that while these desktop GPUs may only appear slightly less powerful on paper, they shouldn't perform dramatically slower. However, once GPU memory is maxed out and the discrete GPU starts swapping to system RAM, performance drops sharply and render times increase significantly.

We saw even more dramatic behaviour in Lumion 2024, where our test scene uses 12 GB of GPU memory when rendering at 4K. With both 8 GB Radeon Pro desktop GPUs, the software simply couldn't cope with running out of memory and subsequently crashed.

Of course, for lighter workloads, which require less than 8 GB, render times are much more comparable. In the D5 Render benchmark, for example, which uses well under 8 GB of GPU memory, the HP ZBook Ultra Gla completed the task in 266 seconds. By comparison, the AMD Radeon Pro W7500 (8 GB) took 386 seconds, the Radeon Pro W7600 (8 GB) 261 seconds, and the Nvidia RTX A1000 (8 GB) 294 seconds.

we're scape With 64 GB of GPU memory at its disposal, the HP ZBook Ultra Gla never approached its limit—even with our most demanding models. However, as datasets grow larger, although it can continue rendering, navigating and working within the 3D environment can become increasingly challenging.

In D5 Render, for example, the HP ZBook Ultra G1a took 795 seconds to render a colossal lakeside model with 2,177,739,558 faces at 4K resolution, using 21 GB of GPU memory. However, viewport performance dropped to just 4.3 frames per second, making it difficult to navigate the scene-a serious productivity killer. For very large models like this, a more powerful GPU such as the Nvidia RTX 5000 Ada (16 GB) would probably be a better choice and you'd also be able to take advantage of Nvidia DLSS to boost real time performance using AI. However, that would require stepping up to a 16-inch or 18-inch mobile workstation.







1 Rendering this Twinmotion Snowdon Tower scene requires 14.2 GB of GPU memory at 4K resolution and a whopping 35.1 GB at 8K resolution

2 Running out of GPU memory can have serious consequences. In Lumion, our 12 GB test scene renders without issue on the HP ZBook Ultra G1a, thanks to the AMD Radeon 8060S having access to 64 GB of shared memory. However, on a desktop workstation with an 8 GB AMD Radeon Pro GPU, the same scene causes the software to crash

3 GPU memory usage is not limited to a single application. Here Solidworks CAD uses 5.0 GB while Solidworks Visualize uses 12.4 GB to render at 4K

4 With access to 64 GB of GPU memory, the HP ZBook G1a can render scenes that would typically need to be handled by a CPU renderer. In KeyShot 2025.2 beta we manged to GPU render this colossal multi room supermarket model from of Kesseböhmer Ladenbau that features 447 million triangles and uses up 18.1 GB of GPU memory simply on loading

It's also important to note that AMD's RDNA 3.5-based GPUs generally trail Nvidia when it comes to hardware ray tracing performance. For example, in Lumion 2024, the HP ZBook Ultra G1a is only 1.44 times slower than a desktop Nvidia RTX 2000 Ada GPU in raster rendering, but that gap widens to 2.61 times slower during ray-trace rendering. More critically, enabling path tracing in Twinmotion on the HP ZBook Ultra G1a caused the software to crash.

The GPU software shift

Over the past decade, AMD GPUs have become incompatible with certain professional 3D software tools. While AMD GPUs work directly with open standards like DirectX 12, the graphics API which forms the backbone to Unreal Engine, Lumion, Twinmotion, Revit, D5 Render and others, any software developed with Nvidia CUDA has not been able to run on AMD Radeon (Pro) GPUs.

The good news is this is starting to change. As Independent Software Vendors (ISVs) cotton on to the huge potential of AMD's new integrated graphics technology, there appears to be real appetite to support it. AMD has collaborated with several ISVs to port Nvidia CUDA code over to AMD's HIP framework. For visualisation, software includes Luxion KeyShot, Maxon Cinema 4D and Redshift, and Rhino with Cycles. For simulation, there's Altair Inspire and Ansys Discovery.

Of course, there's still a lot of work to do. Some AEC-focused ISVs depend on Nvidia GPUs to accelerate specific features. In reality modelling software Leica Cyclone 3DR, for example, AI classification is built around Nvidia CUDA (*read our workstations for reality modelling article www.tinyurl.com/WS-reality*).

One of the most interesting developments is from Luxion, which currently supports AMD GPUs in a beta version of KeyShot 2025.2.

Luxion was a long-time advocate of CPU rendering, but in 2020 finally embraced the GPU, adding render support for Nvidia RTX through OptiX. Today, many of its customers use GPUs to benefit from

AECMAGAZINE

faster results. However, CPU rendering remains essential for more complex projects. When a discrete GPU runs out of memory and must swap out to system RAM, rendering performance can slow significantly and warning messages appear. If memory demands then increase further, the system can eventually crash.

With the HP ZBook Ultra G1a KeyShot customers have real choice - render on CPU or GPU, it's doesn't matter as they can both have access to the same memory.

We tested the hardware in KeyShot 2025.2 beta, using an absolutely colossal multi room supermarket model courtesy of Kesseböhmer Ladenbau that uses 18.1 GB of GPU memory simply to load. The scene features 447 million triangles (nearly 900 times more than KeyShot's sample head-phone scene) and includes 2,228 physical lights and 237 382 parts with incredible detail. There are chiller cabinets, cereal boxes, and 3D fruit and vegetables!

In fairness, this model is probably too big to use on a laptop like this, day in day out. While FHD renders took under 10 minutes, we found it very hard to move around the scene.

But the key point here, is that it would have been previously unthinkable for a scene this complex to be GPU-rendered on a 14-inch laptop.

Don't forget the CPU

It's easy to forget that the HP ZBook Ultra Gla also has an incredibly powerful CPU. In our CPU rendering benchmarks, its scores in V-Ray 6,0 (33,285), Cinebench 2024 (1,596), KeyShot 11,.3 (4.15) and Corona Render 10 (9,923,886) are almost exactly twice that of the HP ZBook Firefly G11A - although this is hardly surprising given it has twice the number of cores.

What might surprise you is that compared to the desktop AMD Ryzen 9 9950X processor, which boasts the same number of 'Zen 5' cores but draws 230 watts vs 70 watts at peak, there's not that big a difference.

The desktop AMD Ryzen 9 9950X is only 45% faster in V-Ray, 39% faster in Cinebench, 44% faster in KeyShot and 50% faster in Corona Render 10.

The competition

Mobile workstation manufacturers including Dell, Lenovo, and HP—are currently in a transitional phase. New models featuring Intel and Nvidia processors have been partially announced, and by July we can expect a full rollout of workstation-class laptops powered by Intel Core Ultra (Series 2) CPUs and Nvidia Blackwell laptop GPUs.



There are indirect limits to model size. This colossal D5 Render scene needs more than

20 GB of GPU memory but still renders

fine on the HP ZBook Ultra G1a. However.

because of its complexity, viewport

performance drops to just 4.3 frames per

second, making it difficult to navigate the

scene - a real productivity killer

It's important to note that Nvidia is increasing memory across its professional GPU lineup. However, this upgrade is modest on the models most likely to appear in 14-inch laptops. The

RTX Pro 500 Blackwell (6 GB), RTX Pro 1000 Blackwell (8 GB), and RTX Pro 2000 Blackwell (8 GB) all remain at 8 GB or below.

In contrast, the higher-end models— RTX Pro 3000 Blackwell (12 GB), RTX Pro 4000 Blackwell (16 GB), and RTX Pro 5000 Blackwell (24 GB)—offer significantly more memory than their Ada Generation predecessors. This capacity lift could make them much better equipped to handle more demanding datasets. And let's not forget that the RTX Pro 4000 Blackwell and RTX Pro 5000 Blackwell will likely offer significantly more performance than the integrated AMD Radeon 8060S GPU in the HP ZBook Ultra G1a.

Conclusion

In the HP ZBook Ultra G1a, it feels like we're witnessing a genuine shift in the landscape for 14-inch mobile workstations. This isn't just about cramming more power into a device that slips easily into a backpack — it's about redefining what's possible with integrated GPUs. And, with one eye on the future, perhaps even workstation GPUs in general.

HP is currently the only major workstation OEM to take on the processor, but Dell and Lenovo will certainly be paying close attention.

Naturally, there's a practical limit to how large models the laptop can handle. An architect or designer might tolerate longer render times with a GPU that's not in the high-end league. But if that trade-off means only being able to navigate a model at a few frames per second, it can quickly lead to frustration. In some cases, full interactivity from a more powerful GPU will outweigh the benefits of ultimate portability. It's all about finding the right balance.In saying that, not all pro applications need a fully interactive 3D viewport, and the HP ZBook Ultra Gla could

be a good fit for GPU-accelerated simulation and reality modelling - not forgetting CPU-accelerated workflows too.

Of course, there are still some software compatibility hurdles to overcome, particularly in CUDA-only tools. But from the software developers we have spoken with, there appears to be genuine excitement about the new technology. For the first time in years, it feels like the playing field is beginning to level. To sustain this momentum, AMD will need to continue to invest in software development.

Perhaps the most exciting thing is where this technology could be heading — both in the short term and further down the line.

HP has already announced the HP Z2 Mini G1a, a micro desktop workstation powered by the same AMD Ryzen AI Max Pro processor. But with a bigger 300W power supply, it's likely to deliver significantly better performance.

Could we also see AMD Ryzen AI Max Pro make its way into larger mobile workstations, which typically come with 230W power supplies? It certainly seems possible.

Looking further ahead, we're eager to see what the rumoured Zen 6-based 'Medusa Halo' processor — expected to launch next year — might bring to the table. According to the "Moore's Law is Dead" a respected tech-focused YouTube channel, it could offer a 30% to 50% GPU performance boost over the current AMD Ryzen AI Max Pro chip.

If these predictions hold true, the AMD chip could deliver a substantial edge when handling larger models, potentially redefining the capabilities of both mobile and desktop workstations.