

# Eine neue High-End-AMD-Grafikkarte auf Basis der uDNA-Architektur: Eine Prognose für Ende 2026

## Executive Summary

Dieser Bericht beleuchtet die potenzielle Ausgestaltung einer neuen High-End-Grafikkarte von AMD, die voraussichtlich Ende 2026 auf der Grundlage der kommenden uDNA-Architektur auf den Markt kommen wird. AMD hat mit der Veröffentlichung der RDNA 4-Architektur (RX 9070 XT) bereits einen bedeutenden architektonischen Wandel vollzogen und vollzieht nun mit uDNA einen weiteren strategischen Schritt hin zu einer vereinheitlichten GPU-Architektur. Diese zielt darauf ab, die Leistung in Bereichen wie Raytracing und KI-Beschleunigung deutlich zu verbessern, wo AMD seine Position weiter stärken möchte.

Eine Flaggschiff-GPU auf Basis der uDNA-Architektur wird voraussichtlich die Leistung der aktuellen RDNA 4-Generation, insbesondere bei Raytracing- und KI-Workloads, erheblich übertreffen. Dies würde AMD als starken Konkurrenten zu NVIDIAs kommenden Angeboten positionieren. Die Karte dürfte auf einem hochmodernen Fertigungsprozess (wahrscheinlich TSMC N2P oder A16) basieren und die nächste Generation der Speichertechnologie (GDDR7 oder potenziell HBM4 für Top-Modelle) nutzen.

## 1. Einleitung: Der Beginn der uDNA-Ära

AMD hat mit der RDNA 4-Architektur einen bedeutenden architektonischen Wandel vollzogen, der bereits erhebliche Verbesserungen bei Raytracing und KI-Beschleunigung mit sich brachte. Nun steht AMD mit der uDNA (Unified DNA)-Architektur an der Schwelle zu einem weiteren strategischen Schritt, indem es seine auf Gaming ausgerichtete RDNA-Architektur und seine auf Rechenzentren spezialisierte CDNA-Architektur zu einer einzigen, vereinheitlichten Plattform zusammenführt. Diese strategische Entscheidung, die von Jack Huynh, Senior Vice President von AMD, auf der IFA 2024 bekannt gegeben wurde, zielt darauf ab, die Softwareentwicklung zu vereinfachen, die Leistung zu steigern und die Kompatibilität über verschiedene Anwendungen hinweg zu verbessern. Dieser Schritt spiegelt NVIDIAs erfolgreichen Ansatz wider, sowohl Gaming- als auch Compute-Aufgaben innerhalb eines einzigen Ökosystems zu unterstützen, indem eine einzige grundlegende Architektur genutzt wird. Die Entscheidung von AMD, RDNA und CDNA zu uDNA zu vereinheitlichen, ist nicht nur eine technische Konsolidierung; sie stellt eine direkte strategische Antwort auf NVIDIAs langjährige Dominanz dar, insbesondere in den KI- und professionellen Visualisierungsbereichen, die stark von NVIDIAs vereinheitlichter CUDA-Plattform profitieren. Durch die Vereinfachung der Entwicklung und die Verbesserung der Kompatibilität versucht AMD, die Hürden für Entwickler zu senken, um ihre Hardware für vielfältige Workloads zu adaptieren und so ihre Marktreichweite über das traditionelle Gaming hinaus zu erweitern. Dies deutet auf eine langfristige Vision hin, einen größeren Anteil an den aufstrebenden KI- und HPC-Märkten zu gewinnen und gleichzeitig ihre Gaming-GPU-Angebote zu stärken. Die AMD Radeon RX 9070 XT, basierend auf der RDNA 4-Architektur (Navi 48), wurde am 6.

März 2025 offiziell auf den Markt gebracht. Sie wird im TSMC N4P-Prozess gefertigt und verfügt über 53,9 Milliarden Transistoren. Die RX 9070 XT bietet bereits eine verbesserte Raytracing-Leistung und eine deutlich leistungsfähigere FSR 4-Upscaling-Technologie im Vergleich zur RX 7900 XTX. In reiner Rasterisierungsleistung ist sie der RX 7900 XT überlegen und kann es sogar mit der RX 7900 XTX aufnehmen.

Das Ziel dieses Berichts ist es, eine detaillierte Prognose einer hypothetischen neuen High-End-Grafikkarte von AMD auf Basis der uDNA-Architektur zu erstellen, die voraussichtlich Ende nächsten Jahres, also Ende 2026, erscheinen wird. Der Bericht wird alle verfügbaren Informationen, einschließlich Gerüchte, Leaks und Branchen-Roadmaps, synthetisieren, um die erwarteten technischen Spezifikationen und Leistungsfähigkeiten zu prognostizieren. Der genaue Zeitrahmen von "Ende nächsten Jahres" (Ende 2026) ist entscheidend. Die vorliegenden Informationen deuten darauf hin, dass die ersten uDNA-GPUs im zweiten Quartal 2026 in die Massenproduktion gehen sollen, wobei der Fokus zunächst auf dem High-End-Segment liegt. Gleichzeitig ist die Massenproduktion von TSMCs N2 (2nm)-Prozess für die zweite Hälfte des Jahres 2025 geplant, N2P für 2026 und A16 (1.6nm) für die Hochvolumenfertigung (HVM) Ende 2026. Diese zeitliche Übereinstimmung legt nahe, dass eine High-End-uDNA-GPU, die Ende 2026 auf den Markt kommt, mit größter Wahrscheinlichkeit TSMCs N2P- oder sogar den hochmodernen A16-Prozess für ihre Haupt-Compute-Dies nutzen wird. Die Wahl dieses fortschrittlichen Knotens ist absolut entscheidend, um die prognostizierten Leistungs- und Effizienzsteigerungen zu erzielen, die notwendig sind, um effektiv mit NVIDIAs Wettbewerbs-Roadmap (z.B. Rubin im Jahr 2027, Blackwell im Jahr 2025 für HPC) zu konkurrieren. Der "anfängliche Fokus auf das High-End-Segment" unterstreicht zudem AMDs Absicht, NVIDIAs Top-Tier-Angebote von Beginn der uDNA-Generation an direkt herauszufordern, anstatt einer schrittweisen Einführung.

## **2. uDNA-Architektur: Eine vereinheitlichte und fortschrittliche Grundlage**

### **Architektonische Philosophie: Verschmelzung von RDNA (Gaming) und CDNA (Rechenzentrum) für Vielseitigkeit**

uDNA zielt darauf ab, die unterschiedlichen Merkmale von RDNA (Gaming) und CDNA (Rechenzentrum) in einer einzigen, kohärenten Plattform zu integrieren. Diese Vereinheitlichung ist darauf ausgelegt, die Softwareentwicklung und -optimierung zu vereinfachen, was es Entwicklern von Natur aus erleichtert, zu arbeiten und die Kompatibilität zwischen verschiedenen Anwendungen zu verbessern. Das ultimative Ziel ist es, dass ein einziges GPU-Design eine breite Palette von Workloads effizient bewältigen kann, einschließlich traditionellem Gaming, anspruchsvollen KI-Workloads und Hochleistungsrechnen (HPC). Die Vereinheitlichung von RDNA und CDNA geht weit über ein bloßes Hardware-Redesign hinaus; sie stellt einen tiefgreifenden Vorstoß in das Software-Ökosystem dar. Durch die Vereinfachung der Entwicklung und Optimierung begegnet AMD einer historischen Herausforderung: Entwickler für ihre Plattformen zu gewinnen und zu halten, insbesondere für Nicht-Gaming-Workloads, wo NVIDIAs CUDA ein Quasi-Monopol etabliert hat. Eine wirklich vereinheitlichte Architektur bedeutet, dass Entwickler Code schreiben können, der effizient über AMDs gesamte GPU-Palette läuft, von Consumer-Grafikkarten bis hin zu Rechenzentrums-Beschleunigern, was potenziell die Entwicklungskosten und -zeit reduziert.

Dieser strategische Schritt könnte zu einer breiteren Akzeptanz von AMD-GPUs in professionellen Anwendungen führen und eine reichhaltigere Softwarebibliothek fördern, was wiederum ihre Hardware für Endbenutzer in verschiedenen Segmenten attraktiver macht.

## **Fertigungsprozess: Nutzung von TSMCs N3E-Knoten und seinen erwarteten Vorteilen (Leistung, Effizienz, Dichte)**

Die RDNA 4-Architektur (Navi 48), die in der RX 9070 XT zum Einsatz kommt, wird im TSMC N4P-Prozess gefertigt. Frühe Berichte deuteten darauf hin, dass uDNA-basierte GPUs TSMCs N3E-Fertigungsprozess nutzen würden, der eine Leistungssteigerung von 18 % und eine Reduzierung des Stromverbrauchs um 32 % im Vergleich zum N5-Knoten verspricht. Eine genauere Betrachtung von TSMCs öffentlicher Roadmap zeigt jedoch, dass die N2 (2nm)-Massenproduktion für die zweite Hälfte des Jahres 2025 geplant ist, N2P für 2026 und A16 (1.6nm) für die Hochvolumenfertigung Ende 2026 bereitsteht. Für eine High-End-GPU, die Ende 2026 auf den Markt kommt, ist es sehr wahrscheinlich, dass AMD den fortschrittlichsten verfügbaren Prozessknoten für seinen Haupt-Compute-Die nutzen wird, wahrscheinlich N2P oder sogar A16, insbesondere für ein Flaggschiffprodukt, das darauf ausgelegt ist, an der Spitze zu konkurrieren. A16 bietet insbesondere eine Geschwindigkeitsverbesserung von 8-10 % gegenüber N2P und ist für Rechenzentrums-KI-Prozessoren mit Backside Power Delivery optimiert, was perfekt zur vereinheitlichten Vision von uDNA passt. Die ursprüngliche Erwähnung von N3E für uDNA könnte eine frühe oder konservative Prognose gewesen sein. Angesichts des "Ende nächsten Jahres" (Ende 2026) Startzeitpunkts für eine *High-End*-Karte und TSMCs bestätigter Roadmap, die N2P im Jahr 2026 und A16 (1.6nm) bereit für HVM Ende 2026 zeigt, wird es sehr wahrscheinlich, dass AMDs Flaggschiff-uDNA-GPU auf N2P oder sogar A16 abzielt. A16s "Backside Power Delivery" (BSPDN) wird ausdrücklich als "beste Lösung für Rechenzentrums-KI-Prozessoren" bezeichnet, was perfekt zu uDNAs Kernziel passt, RDNA (Gaming) und CDNA (Rechenzentrum) für verbesserte KI-Beschleunigung zu verschmelzen. Diese aggressive Einführung eines führenden Knotens würde einen erheblichen Vorteil in Bezug auf Transistordichte, Energieeffizienz und Rohleistung bieten, was absolut entscheidend ist, um direkt mit NVIDIAs Architekturen der nächsten Generation wie Blackwell (gefertigt auf TSMC 4NP, mit 208 Milliarden Transistoren) und zukünftigen Rubin-Architekturen zu konkurrieren. Diese strategische Wahl signalisiert AMDs Absicht, einen substanziellen Generationssprung sowohl in Effizienz als auch in roher Verarbeitungsleistung zu erzielen.

## **Kernarchitektonische Innovationen**

### **Neu gestaltete Compute Units und das "Wavegroup"-Ausführungsmodell**

Aufbauend auf RDNA 4, das bereits Out-of-Order-Warteschlangen für Speicheranfragen und eine verbesserte Skalareinheit mit FP32-Unterstützung zur Steigerung der realen Leistung pro CU eingeführt hat, wird uDNA (oder RDNA 5, seine Gaming-Variante) dies mit einem überarbeiteten Threading-System namens "Wavegroup"-Modell weiterentwickeln. Diese Innovation ist speziell darauf ausgelegt, die Thread-Verwaltung zu optimieren und die Parallelität zu verbessern, insbesondere für hochparallele Aufgaben wie Raytracing und KI-Berechnungen. Die praktische Auswirkung sind effizientere GPU-Kerne, die zu einer besseren Gesamtleistung bei hohen Workloads und potenziell weniger CPU-Engpässen in komplexen Szenen führen. Die Einführung des "Wavegroup"-Modells und die früheren RDNA

4-Verbesserungen wie Out-of-Order-Speicheranfragen unterstreichen gemeinsam einen starken mikroarchitektonischen Fokus auf die Verbesserung der Effizienz innerhalb der Compute Units. Diese Strategie zielt darauf ab, mehr Arbeit pro Taktzyklus und pro CU zu extrahieren, anstatt sich ausschließlich auf eine brutale Erhöhung der CU-Anzahl zu verlassen. Eine verbesserte Thread-Verwaltung und Parallelität sind für moderne, hochparallele und oft speichergebundene Workloads, wie komplexe Raytracing- und groß angelegte KI-Berechnungen, von größter Bedeutung. Dies deutet darauf hin, dass AMD Leistungsengpässe auf einer grundlegenden architektonischen Ebene angeht, was zu konsistenteren und breiteren Leistungssteigerungen über verschiedene Anwendungen hinweg führen kann, die über bloße synthetische Benchmarks hinausgehen.

## **Deutliche Fortschritte bei den Raytracing-Fähigkeiten**

Die RDNA 4-Architektur hat bereits signifikante Fortschritte bei den Raytracing-Fähigkeiten gemacht, indem sie die Leistung der Ray Accelerators bei der Analyse von Schnittpunkten verdoppelt hat (8 Schnittpunkte eines Strahls mit der Bounding-Box-Struktur (BVH) pro Zyklus) und Verbesserungen für BVH-Strukturen wie Oriented Bounding Boxes (OBB) eingeführt hat, die die Raytracing-Leistung um bis zu 10 % verbessern können. Es wird gemunkelt, dass uDNA die **Raytracing-Leistung im Vergleich zur aktuellen RDNA 4-Hardware-Generation verdoppeln** wird. Es ist wichtig zu präzisieren, dass dies die Recheneffizienz der Lichtsimulation betrifft, was bedeutet, dass die für Raytracing-Berechnungen in jedem Frame benötigte Zeit halbiert wird, anstatt die Gesamtbildraten in Spielen direkt zu verdoppeln. AMDs historischer Kampf, NVIDIAs Raytracing-Leistung zu erreichen, wird mit uDNA aggressiv angegangen. RDNA 4s grundlegende Verbesserungen, wie die verdoppelte Schnittpunktanalyse und die Einführung von OBBs, sind entscheidende Vorläufer. Die gemunkelte "Verdoppelung" der RT-Leistung für uDNA zielt direkt auf diese langjährige Schwäche ab. Während die Klarstellung, dass dies die "Recheneffizienz" betrifft und nicht eine direkte Verdoppelung der FPS, wichtig ist, ist dieser Effizienzgewinn von größter Bedeutung. Er bedeutet, dass für eine gegebene Raytracing-Workload die uDNA-GPU die RT-spezifischen Berechnungen in der Hälfte der Zeit abschließt, wodurch GPU-Ressourcen für andere Aufgaben freigegeben werden oder die Implementierung weitaus komplexerer und visuell reichhaltigerer Raytracing-Effekte ermöglicht wird. Dies ist ein entscheidender Schritt, um hochauflösendes Raytracing auf AMD-Hardware praktikabler und leistungsfähiger zu machen, und positioniert es, um direkt mit NVIDIAs führenden Lösungen zu konkurrieren, die in diesem Bereich historisch führend waren (z.B. Ada Lovelaces RT-Fähigkeiten). Die erwartete Integration von uDNA in Konsolen der nächsten Generation deutet zudem darauf hin, dass Entwickler eine vereinheitlichte, leistungsstarke RT-Plattform über PC und Konsole hinweg haben werden, was eine breitere und anspruchsvollere Einführung von Raytracing in zukünftigen Spieliteln vorantreiben könnte.

## **Verbesserte KI-Beschleunigung und ihre Rolle bei Funktionen wie Upscaling und Frame Generation**

Die RDNA 4-Architektur bringt bereits eine verbesserte KI-Beschleunigung mit sich, und die FSR 4-Technologie ist bereits "way more capable" und "often beating DLSS CNN". Es wird erwartet, dass uDNA die **KI-Leistung im Vergleich zur RDNA 4-Generation verdoppelt** und neue Hardware speziell für KI- und Machine-Learning-Aufgaben integrieren wird. Diese signifikante Verbesserung unterstützt und optimiert direkt KI-gesteuerte Funktionen wie

Bild-Upscaling (z.B. AMD FidelityFX Super Resolution, FSR) und Frame-Generation-Technologien. Das übergeordnete Ziel ist es, AMDs Produkte im Vergleich zu NVIDIAs etablierten KI-Beschleunigungsfähigkeiten, insbesondere deren DLSS- und Frame-Generation-Technologien, wettbewerbsfähiger zu machen. Zum Vergleich: NVIDIAs Blackwell-Architektur (exemplifiziert durch den B200 für Rechenzentren) verfügt über einen Second-Generation Transformer Engine, neue Präzisionsoptionen (einschließlich FP4) und Hardware-Unterstützung für strukturierte Sparsity, wobei bis zu 15-mal höhere Inferenz-Durchsätze für große Modelle im Vergleich zu seinem Vorgänger, dem H100, beansprucht werden. Obwohl Blackwell derzeit auf HPC ausgerichtet ist, werden seine architektonischen Fortschritte unweigerlich auf NVIDIAs Consumer-GPUs übertragen und eine hohe Messlatte für die KI-Leistung setzen. Die starke Betonung der "Verdoppelung der KI-Leistung" und die Aufnahme neuer dedizierter KI-Hardware signalisieren AMDs klare Erkenntnis, dass KI keine Nischenfunktion mehr ist, sondern eine grundlegende Komponente der modernen GPU-Leistung, sowohl für Gaming (über Upscaling und Frame Generation) als auch für professionelle Anwendungen. Hierbei geht es nicht nur um reine Rechenleistung; es geht darum, ein robustes Ökosystem von KI-gesteuerten Funktionen effizienter und qualitativ hochwertiger zu gestalten. Durch die signifikante Verbesserung der KI-Fähigkeiten kann AMD wirklich wettbewerbsfähige Alternativen zu NVIDIAs DLSS und Frame Generation anbieten, die zu kritischen Leistungs-Multiplikatoren in anspruchsvollen modernen Spielen geworden sind. Darüber hinaus positioniert dies AMD strategisch für die breitere Integration von KI in Anwendungen jenseits des Gamings, wie z.B. Inhaltserstellung, wissenschaftliche Forschung und sogar lokale Large Language Model (LLM)-Inferenz, Bereiche, in denen NVIDIA derzeit eine dominante Marktposition innehat.

### **Verbesserungen der Rasterisierungs-IPC und der Path-Tracing-Leistung**

Die RDNA 4-Architektur hat bereits eine verbesserte Rasterisierungsleistung von etwa **20 % pro Compute Unit** im Vergleich zu RDNA 3 gezeigt. uDNA (oder RDNA 5) wird voraussichtlich eine weitere Verbesserung der Rasterisierungs-IPC von etwa **20 % pro Compute Unit im Vergleich zu RDNA 4** liefern. Dies stellt einen signifikanten Gewinn an Instruktionen pro Takt (IPC)-Effizienz für die konventionelle Rasterisierungsdarstellung dar. Noch auffälliger sind die Path-Tracing-Ergebnisse: RDNA 4 lieferte bereits fast doppelt so viele FPS, eine **100 %ige Steigerung** gegenüber seinem Vorgänger (RDNA 3). uDNA wird voraussichtlich **ähnliche Zuwächse bei Path Tracing im Vergleich zu RDNA 4** aufweisen. Dieser substantielle Anstieg markiert AMDs bemerkenswertesten IPC-Gewinn seit der Einführung der ursprünglichen RDNA-Architektur. Während AMD offensichtlich erhebliche Investitionen in die Verbesserung der Raytracing- und KI-Fähigkeiten tätigt, zeigt der prognostizierte substantielle 20 %ige Rasterisierungs-IPC-Gewinn unmissverständlich eine ausgewogene und umfassende Leistungsstrategie. Rasterisierung bleibt die grundlegende Rendering-Technik für die überwiegende Mehrheit der Gaming- und Content-Erstellungs-Workloads. Durch die grundlegende Verbesserung der IPC zusammen mit potenziellen Erhöhungen der CU-Anzahl und höheren Taktraten stellt AMD sicher, dass seine uDNA-Architektur kein "One-Trick-Pony" ist, das sich ausschließlich auf Nischenfunktionen konzentriert, sondern ein robuster und starker Performer über das gesamte Spektrum der Grafikanwendungen hinweg. Diese breite Verbesserung bietet überzeugende Gründe für Benutzer, ein Upgrade durchzuführen, selbst wenn ihre primären Anwendungsfälle nicht stark auf Raytracing angewiesen sind.

### 3. Prognostizierte technische Spezifikationen der High-End uDNA-GPU

#### Geschätzte Kernkonfiguration: Compute Units, Stream Processors, Ray Accelerators

Das aktuelle Flaggschiff der RDNA 3-Generation, die Radeon RX 7900 XTX, verfügt über 96 Compute Units (CUs), 6144 Shading Units (Stream Processors) und 96 Ray Accelerators. Die kürzlich veröffentlichte Radeon RX 9070 XT (RDNA 4) verfügt über 64 Compute Units, 4096 Shading Units und 64 Ray Accelerators. Gerüchte deuten darauf hin, dass uDNA eine "Verdoppelung der CUs von 64 auf 128" beinhalten könnte. Da die RX 9070 XT 64 CUs hat, würde dies für ein uDNA-Flaggschiff eine direkte Verdoppelung der CUs im Vergleich zur aktuellen RDNA 4-High-End-Karte bedeuten. Angesichts des angegebenen "anfänglichen Fokus auf das High-End-Segment" für uDNA und der Notwendigkeit, effektiv mit NVIDIAs Flaggschiff (die RTX 4090 hat 16384 CUDA-Kerne) zu konkurrieren, könnte ein uDNA-Flaggschiff realistisch **128 bis 144 Compute Units** anstreben. Dies würde sich in etwa **8192 bis 9216 Stream Processors** (unter Annahme von 64 SPs pro CU, konsistent mit RDNA 3 und RDNA 4) und **128 bis 144 Ray Accelerators** niederschlagen. Die Benutzeranfrage fragt explizit nach einer *High-End*-Karte. Um jedoch wirklich mit NVIDIAs RTX 4090 (16384 CUDA-Kerne) und den erwarteten Blackwell-Consumer-GPUs (die zweifellos von den massiven 208 Milliarden Transistoren des B200 herunter skaliert werden) zu konkurrieren, würde AMD zusätzlich zu den prognostizierten IPC-Gewinnen eine substanzielle Erhöhung der Rohkernanzahl benötigen. Ein Sprung von 64 CUs (RX 9070 XT) auf 128-144 CUs, kombiniert mit dem prognostizierten 20 %igen Raster-IPC-Gewinn *über RDNA 4*, würde AMD in Bezug auf den reinen Compute-Durchsatz und die Gesamtleistung in eine sehr starke Wettbewerbsposition bringen.

#### Speichersubsystem: Einführung der GDDR7-Technologie (Geschwindigkeit, Bandbreite, VRAM-Kapazität)

Der JEDEC GDDR7-Standard wurde am 5. März 2024 offiziell veröffentlicht, und Micron hat GDDR7 ebenfalls im Jahr 2024 offiziell eingeführt. GDDR7 bietet signifikante Verbesserungen, darunter bis zu 32 Gbit/s pro Pin (was eine 100 % höhere Bandbreite pro Pin im Vergleich zu GDDR6 darstellt) und eine insgesamt 40 % höhere Bandbreite (1,5 TB/s) im Vergleich zu GDDR6 (1,1 TB/s), und das alles bei 20 % höherer Energieeffizienz. Zum Vergleich: Die aktuelle RX 7900 XTX verwendet 24 GB GDDR6-Speicher an einem 384-Bit-Speicherbus und erreicht eine Bandbreite von 960 GB/s. Die NVIDIA RTX 4090 verwendet 24 GB GDDR6X-Speicher an einem 384-Bit-Bus und bietet eine Bandbreite von 1008 GB/s. Die AMD Radeon RX 9070 XT verwendet 16 GB GDDR6-Speicher an einem 256-Bit-Speicherbus und erreicht eine Bandbreite von 644,6 GB/s. Für eine High-End-uDNA-GPU, die Ende 2026 auf den Markt kommt, ist die Einführung von GDDR7 sehr wahrscheinlich und ein logischer Schritt. Wenn AMD einen 384-Bit-Speicherbus beibehält und 32 Gbit/s GDDR7-Chips verwendet, würde dies eine theoretische Bandbreite von **1536 GB/s** (berechnet als  $384 \text{ Bit} * 32 \text{ Gbit/s} / 8 \text{ Bit/Byte}$ ) ergeben. Bezüglich der VRAM-Kapazität erscheint die Beibehaltung von 24 GB oder eine Erhöhung auf 32 GB plausibel. Dies könnte durch die Verwendung von 16-Gbit-GDDR7-Chips für 24 GB oder 24-Gbit-GDDR7-Chips für 36 GB an einem 384-Bit-Bus erreicht werden. Obwohl

32-Gbit-GDDR7-Chips existieren, könnten sie für die ersten Consumer-Karten weniger üblich sein. Angesichts der steigenden VRAM-Anforderungen für hochauflösendes Gaming, fortschrittliche Texturen und insbesondere KI-Workloads wäre eine Kapazität von **24 GB bis 32 GB GDDR7** eine vernünftige und wettbewerbsfähige Prognose für ein Flaggschiff. Die erhebliche Bandbreitenerhöhung von GDDR7 ist absolut entscheidend, um die deutlich gesteigerten Rechenfähigkeiten der uDNA-Architektur effektiv zu versorgen. Eine Verdoppelung der Pro-Pin-Bandbreite von GDDR6 bedeutet, dass selbst wenn AMD eine ähnliche Speicherbusbreite wie RDNA 3 beibehält (z.B. 384-Bit), die effektive Speicherbandbreite einen dramatischen Sprung erleben wird (von 960 GB/s auf etwa 1,5 TB/s). Dieser massive Bandbreitenzuwachs ist entscheidend für anspruchsvolle Workloads wie hochauflösendes Gaming, komplexes Raytracing und insbesondere speicherbandbreitenintensive KI-Anwendungen. Dies stellt sicher, dass die verstärkten Compute Units und spezialisierten RT/AI-Beschleuniger nicht durch unzureichenden Speicherzugriff ausgebremst werden, wodurch sie näher an ihrer theoretischen Spitzenleistung arbeiten und das volle Potenzial der neuen Architektur entfalten können.

## Cache-Hierarchie: Evolution des Infinity Cache

Die RX 7900 XTX von RDNA 3 verfügt über 96 MB Infinity Cache. Die RX 9070 XT (RDNA 4) verfügt über 64 MB L3 Cache. Bemerkenswerterweise ist der Infinity Cache von RDNA 3 in seiner Rohkapazität kleiner als der von RDNA 2 (96 MB vs. 128 MB), gleicht dies jedoch durch eine deutlich höhere effektive Speicherbandbreite (3500 GB/s vs. 1024 GB/s) aus. Die Architektur nutzt auch Techniken wie Work Tiling, um die Cache-Trefferquoten zu erhöhen und die Ausführung von Workloads zu beschleunigen. Für uDNA ist eine fortgesetzte Evolution des Infinity Cache zu erwarten. Während die Rohgröße aufgrund von Kosten- und Komplexitätsüberlegungen bei fortschrittlichen Prozessknoten möglicherweise nicht drastisch ansteigt, werden seine Effizienz und effektive Bandbreite höchstwahrscheinlich weiter verbessert. Dies könnte die Nutzung von Fortschritten aus dem "Wavegroup"-Modell für bessere Datenzugriffsmuster und eine intelligentere Cache-Verwaltung beinhalten. Eine leichte Erhöhung der Größe (z.B. **128 MB oder 192 MB**) in Kombination mit einer noch höheren effektiven Bandbreite und ausgefeilteren Cache-Verwaltungsalgorithmen würde eine logische und wirkungsvolle Weiterentwicklung der uDNA-Architektur darstellen. Infinity Cache ist ein Eckpfeiler von AMDs RDNA-Architektur, entwickelt, um die Auswirkungen schmalere Speicherbusse im Vergleich zu NVIDIAs Ansatz mit höherer Bandbreite GDDR6X/HBM zu mildern. Während RDNA 3 eine Verschiebung hin zur Optimierung der effektiven Bandbreite gegenüber der Rohkapazität zeigte, wird uDNA diesen Trend wahrscheinlich fortsetzen, indem es sich auf die Maximierung der Cache-Trefferquoten und der Gesamteffizienz konzentriert. Das "Wavegroup"-Modell und Out-of-Order-Speicheranfragen sind direkt synergetisch mit diesem Ziel und tragen zu einer besseren Cache-Auslastung bei, indem sie sicherstellen, dass Daten effizienter angefordert und verarbeitet werden. Dies ist entscheidend, um eine hohe Leistung bei latenzempfindlichen Workloads wie Gaming aufrechtzuerhalten und die massiven Datenströme, die in modernen KI-Anwendungen inhärent sind, effektiv zu verwalten. Der primäre Fokus liegt hier nicht nur auf der Erhöhung der Cache-Kapazität, sondern auf einem intelligenten Datenmanagement zur Maximierung der *effektiven* Bandbreite und zur Reduzierung der Latenz.

## Taktraten und Stromverbrauch (TDP)

Die RX 7900 XTX arbeitet mit einem Boost-Takt, der typischerweise zwischen 2498 MHz und

2680 MHz liegt. Ihre Referenz-Gesamtplattenleistung (TDP) ist mit 355 W angegeben, wobei kundenspezifische Designs bis zu 420 W erreichen. Die NVIDIA RTX 4090 hat einen Boost-Takt von ca. 2520 MHz und eine höhere TDP von 450 W. Die RX 9070 XT (RDNA 4) hat einen Boost-Takt von bis zu 2970 MHz und eine TDP von 304 W. TSMCs N3E-Prozess verspricht eine Leistungssteigerung von 18 % und eine Reduzierung des Stromverbrauchs um 32 % im Vergleich zu N5. Der N2-Prozess bietet eine um 10-15 % höhere Leistung bei gleicher Leistungsaufnahme oder eine um 20-30 % geringere Leistungsaufnahme bei gleicher Leistung im Vergleich zu N3E. Der A16-Prozess, noch fortschrittlicher, bietet eine Geschwindigkeitsverbesserung von 8-10 % gegenüber N2P. Angesichts dieser signifikanten Fortschritte bei den Prozessknoten sollte uDNA von Natur aus höhere Taktraten bei ähnlichen oder sogar geringeren Leistungsbudgets *pro Transistor* erreichen. Daher könnte eine Flaggschiff-uDNA-GPU realistisch **Boost-Taktraten im Bereich von 3,0-3,5 GHz** anstreben. Für die Gesamtleistungsaufnahme (TDP) gilt, dass, obwohl die Effizienz *pro Transistor* verbessert wird, eine High-End-Karte, die auf absolute Spitzenleistung abzielt, die Leistungsgrenzen wahrscheinlich ausreizen wird, um den Durchsatz des fortschrittlichen Knotens und der erhöhten Kernanzahl zu maximieren. Eine TDP im Bereich von **400-500 W** ist plausibel, ähnlich oder leicht über den aktuellen Flaggschiffen, um Spitzenleistung zu erzielen. Der Übergang zu hochmodernen Prozessknoten wie N3E, N2P oder A16 bietet von Natur aus erhebliche Energieeffizienzgewinne (z.B. 32 % Energieeinsparung bei gleicher Leistung für N3E vs. N5, 20-30 % geringerer Stromverbrauch für N2 vs. N3E). Auf dem High-End-GPU-Markt nutzen Hersteller diese Effizienzgewinne jedoch typischerweise nicht primär zur Reduzierung des absoluten Stromverbrauchs, sondern zur *Steigerung* der Leistung, indem sie mehr Transistoren einpacken und höhere Taktraten innerhalb ähnlicher oder leicht erweiterter Leistungsbudgets anstreben. Dies bedeutet, dass, obwohl die Effizienz *pro Transistor* dramatisch verbessert wird, der Gesamtstromverbrauch einer *Flaggschiff*-GPU hoch bleiben (400-500 W) oder sogar leicht ansteigen könnte, da AMD darauf abzielt, die absolute Leistung zu maximieren und die Wettbewerbsfähigkeit gegenüber NVIDIA aufrechtzuerhalten. Dies ist ein etablierter Branchentrend, bei dem Leistungsbudgets oft beibehalten oder erhöht werden, um die maximal mögliche Leistung aus neuen Prozessknoten freizuschalten.

## Konnektivität und Display-Ausgänge (z.B. HDMI 2.2)

Die aktuelle RX 7900 XTX unterstützt bereits HDMI 2.1a und DisplayPort 2.1 und die RX 9070 XT unterstützt HDMI 2.1b und DisplayPort 2.1a, was eine hohe Bandbreitenkonnektivität für moderne Displays bietet. uDNA-GPUs werden explizit mit **HDMI 2.2** ausgeliefert, das eine Bandbreite von bis zu 80 Gbit/s bietet. Dies reicht aus, um echtes 8K-Gaming bei sehr hohen Bildwiederholraten zu ermöglichen oder die immensen Datenanforderungen von hochauflösenden Virtual-Reality (VR)-Headsets zu unterstützen. DisplayPort 2.1, bereits vorhanden, würde sicherlich weiterhin unterstützt. Die bestätigte Aufnahme von HDMI 2.2 ist eine direkte und notwendige Antwort auf die schnell steigenden Anforderungen an hochauflösende Displays mit hohen Bildwiederholraten und immersive VR-Erlebnisse der nächsten Generation. Dies stellt sicher, dass die uDNA-GPU nicht nur leistungsstark genug ist, um anspruchsvolle Inhalte zu rendern, sondern auch in der Lage ist, diese Inhalte ohne Engpässe an die neuesten Display-Technologien zu übertragen. Diese Funktion ist entscheidend für die Zukunftssicherung der High-End-GPU und bietet eine nahtlose, hochbandbreite Verbindung für die fortschrittlichsten Monitore und Headsets, was perfekt zu ihrer Positionierung im "High-End"-Segment passt.



**Tabelle 1: Vergleichende Spezifikationen (Aktuelle RX 7900 XTX, RX 9070 XT, NVIDIA RTX 4090, Prognostizierte High-End uDNA-GPU)**

Spezifikation	AMD Radeon RX 7900 XTX	AMD Radeon RX 9070 XT	NVIDIA GeForce RTX 4090	Prognostizierte AMD uDNA Flaggschiff (z.B. Radeon RX 10900 XTX)
<b>Architektur</b>	RDNA 3	RDNA 4	Ada Lovelace	uDNA (RDNA 5)
<b>Fertigungsprozesse</b>	TSMC 5nm (N5)	TSMC 4nm (N4P)	TSMC 4NP	TSMC N2P/A16
<b>Compute Units (CUs)</b>	96	64	N/A (CUDA-Kerne)	128-144
<b>Stream Processors / CUDA Cores</b>	6144	4096	16384	8192-9216
<b>Ray Accelerators / RT Cores</b>	96	64	Dedizierte RT-Kerne	128-144
<b>Boost-Takt (MHz)</b>	~2498-2680	~2970-3100	~2520	~3000-3500
<b>Speichertyp</b>	GDDR6	GDDR6	GDDR6X	GDDR7
<b>Speichergröße (GB)</b>	24	16	24	24-32
<b>Speicherbus (Bit)</b>	384	256	384	384
<b>Speichergeschwindigkeit (Gbps effektiv)</b>	20	20.1	21	32
<b>Speicherbandbreite (GB/s)</b>	960	644.6	1008	~1536
<b>Infinity Cache (MB)</b>	96	64 (L3 Cache)	72 (L2-Cache)	128-192
<b>Gesamtleistungsaufnahme (TDP) (W)</b>	355-420	304-360	450	400-500
<b>Wichtiger Display-Ausgang</b>	DisplayPort 2.1, HDMI 2.1a	DisplayPort 2.1a, HDMI 2.1b	DisplayPort 1.4a, HDMI 2.1	DisplayPort 2.1, HDMI 2.2

Diese Tabelle ist von unschätzbarem Wert, da sie einen klaren, prägnanten und direkten Vergleich der wichtigsten technischen Spezifikationen bietet und es den Lesern ermöglicht, den prognostizierten Generationssprung und die Wettbewerbspositionierung des uDNA-Flaggschiffs schnell zu erfassen. Für ein technisch versiertes Publikum sind diese numerischen Spezifikationen die grundlegende Basis für das Verständnis des Leistungspotenzials einer GPU und ihrer zugrunde liegenden Fähigkeiten. Durch die Gegenüberstellung der aktuellen AMD- und NVIDIA-Flaggschiffe mit der prognostizierten uDNA-Karte werden die erwarteten Generationsverbesserungen sofort hervorgehoben und klar veranschaulicht, wie AMD sich im Wettbewerb mit seinem Hauptkonkurrenten positionieren will. Dieses Format ermöglicht eine schnelle Referenz und untermauert die Darstellung eines signifikanten Leistungssprungs.

## 4. Leistungsanalyse: Wie stark wird sie sein?

### Rasterisierungs-Leistungsprognosen: Quantifizierung von IPC-Gewinnen und Gesamtdurchsatz

Die RDNA 4-Architektur hat bereits eine um etwa **20 % bessere Rasterisierungsleistung pro Compute Unit** im Vergleich zu RDNA 3 geliefert. Die uDNA-Architektur wird voraussichtlich eine weitere Verbesserung der Rasterisierungs-IPC von etwa **20 % pro Compute Unit im Vergleich zu RDNA 4** liefern. Dies stellt einen erheblichen Gewinn an Instruktionen pro Takt (IPC)-Effizienz für die traditionelle Rasterisierungsdarstellung dar. Angesichts der prognostizierten Zunahme der Compute Units (z.B. von 64 in der RX 9070 XT auf 128-144 im uDNA-Flaggschiff) und der erwarteten höheren Taktraten (z.B. 3,0-3,5 GHz) wäre der gesamte Rasterisierungsleistungszuwachs außergewöhnlich signifikant. Wenn die RX 9070 XT (64 CUs) eine Basis-Rasterisierungsleistung darstellt, hätte eine uDNA-GPU mit 128 CUs theoretisch  $(128 \text{ CUs} / 64 \text{ CUs}) * 1,20 \text{ (IPC-Gewinn)} = 2 * 1,20 = \sim 2,4x$  die Rasterisierungsleistung der RX 9070 XT bei gleicher Taktrate. Unter Berücksichtigung höherer Taktraten (z.B. 3,2 GHz vs. 2,9 GHz für die 9070 XT) könnte der gesamte Leistungszuwachs realistisch näher bei **2,5x - 3,0x gegenüber der RX 9070 XT** im reinen Rasterisierungsdurchsatz liegen. Während AMD offensichtlich erhebliche strategische Investitionen in die Verbesserung der Raytracing- und KI-Fähigkeiten tätigt, zeigt der prognostizierte substanzielle 20 %ige Rasterisierungs-IPC-Gewinn unmissverständlich eine ausgewogene und umfassende Leistungsstrategie. Rasterisierung bleibt die grundlegende Rendering-Technik für die überwiegende Mehrheit der Gaming- und Content-Erstellungs-Workloads. Durch die grundlegende Verbesserung der IPC zusammen mit potenziellen Erhöhungen der CU-Anzahl und höheren Taktraten stellt AMD sicher, dass seine uDNA-Architektur kein "One-Trick-Pony" ist, das sich ausschließlich auf Nischenfunktionen konzentriert, sondern ein robuster und starker Performer über das gesamte Spektrum der Grafikanwendungen hinweg. Diese breite Verbesserung bietet überzeugende Gründe für Benutzer, ein Upgrade durchzuführen, selbst wenn ihre primären Anwendungsfälle nicht stark auf Raytracing angewiesen sind.

### Raytracing-Leistung: Detaillierte Diskussion der "verdoppelten" Leistung und ihrer Auswirkungen in der Praxis

Die RDNA 4-Architektur hat bereits signifikante Fortschritte bei den Raytracing-Fähigkeiten gemacht, indem sie die Leistung der Ray Accelerators bei der Analyse von Schnittpunkten verdoppelt hat (8 Schnittpunkte eines Strahls mit der Bounding-Box-Struktur (BVH) pro Zyklus) und Verbesserungen für BVH-Strukturen wie Oriented Bounding Boxes (OBB) eingeführt hat, die die Raytracing-Leistung um bis zu 10 % verbessern können. Es wird durchweg gemunkelt, dass uDNA die **Raytracing-Leistung im Vergleich zur aktuellen RDNA 4-Hardware-Generation verdoppeln** wird. Es ist entscheidend zu verstehen, dass dies die Recheneffizienz der Lichtsimulation betrifft, was bedeutet, dass die für Raytracing-Berechnungen in jedem Frame benötigte Zeit halbiert wird, anstatt die Gesamtbildraten direkt zu verdoppeln.

**Auswirkungen in der Praxis:** Diese signifikante Steigerung der Raytracing-Effizienz bedeutet, dass komplexere und visuell anspruchsvollere Raytracing-Effekte (z.B. globale Beleuchtung, präzise Reflexionen, realistische Brechungen) mit wesentlich geringerem Leistungsaufwand

gerendert werden können. Obwohl dies nicht immer zu einer direkten Verdoppelung der FPS in jedem Spiel führt, ermöglicht es Entwicklern, Raytracing-Erlebnisse mit höherer Wiedergabetreue bei spielbaren Bildraten zu implementieren oder umfangreichere und allgegenwärtigere Raytracing-Effekte zu integrieren, ohne die Gesamtleistung zu beeinträchtigen. Dies behebt direkt AMDs historische Raytracing-Leistungslücke gegenüber NVIDIAs führenden Lösungen. AMDs konsequenter und aggressiver Fokus auf die Verbesserung der Raytracing-Fähigkeiten, der sich von RDNA 2 über RDNA 3, RDNA 4 und nun die gemunkelte Verdoppelung von uDNA erstreckt, ist ein klares strategisches Gebot. Raytracing wird schnell zu einem unverzichtbaren und zunehmend wichtigen Merkmal im modernen Gaming, das die visuelle Wiedergabetreue vorantreibt. Durch die signifikante Steigerung der RT-Effizienz will AMD nicht nur zur aktuellen Generation von NVIDIA aufschließen, sondern seine Architektur auch proaktiv für kommende Spiele zukunftssicher machen, die zunehmend auf diese fortschrittlichen Rendering-Techniken angewiesen sein werden. Die kritische Unterscheidung zwischen "Recheneffizienz" und "Gesamtbildrate" ist von entscheidender Bedeutung: Sie bedeutet, dass die *Kosten* von RT-Berechnungen effektiv halbiert werden, was Entwicklern deutlich mehr Leistungsspielraum verschafft, was indirekt entweder zu besseren Bildraten oder zur Möglichkeit führt, visuell beeindruckendere und komplexere Effekte zu implementieren. Diese strategische Ausrichtung wird durch die erwartete Integration von uDNA in Konsolen der nächsten Generation weiter verstärkt, da Konsolen oft als Katalysator für eine breitere Einführung neuer Rendering-Funktionen in der gesamten Branche dienen.

## **KI- und Upscaling-Leistung: Wettbewerbsfähigkeit bei KI-gesteuerten Funktionen**

Die RDNA 4-Architektur bringt bereits eine verbesserte KI-Beschleunigung mit sich, und die FSR 4-Technologie ist bereits "way more capable" und "often beating DLSS CNN". Es wird explizit erwartet, dass uDNA die **KI-Leistung im Vergleich zur RDNA 4-Generation verdoppelt** und neue, dedizierte Hardware speziell für KI- und Machine-Learning-Aufgaben integrieren wird. Diese substanzielle Verbesserung unterstützt und optimiert direkt KI-gesteuerte Funktionen, die für modernes Gaming entscheidend sind, wie Bild-Upscaling (wie AMDs FSR) und Frame-Generation-Technologien. Das übergeordnete Ziel ist es, AMDs Produkte im Vergleich zu NVIDIAs etablierten und dominanten KI-Beschleunigungsfähigkeiten, insbesondere deren DLSS- und Frame-Generation-Lösungen, deutlich wettbewerbsfähiger zu machen. Zum Vergleich: NVIDIAs Blackwell-Architektur (exemplifiziert durch den B200 für Rechenzentren) verfügt über 208 Milliarden Transistoren, einen Second-Generation Transformer Engine, neue Präzisionsoptionen (einschließlich FP4) und beansprucht bis zu 15-mal höhere Inferenz-Durchsätze für große Modelle im Vergleich zu seinem Vorgänger, dem H100. Obwohl Blackwell derzeit auf HPC ausgerichtet ist, werden seine architektonischen Fortschritte unweigerlich auf NVIDIAs Consumer-GPUs übertragen und eine hohe Messlatte für die KI-Leistung setzen. Die prognostizierte Verdoppelung der KI-Leistung geht nicht nur darum, höhere reine Rechenwerte zu erreichen; es geht im Grunde darum, ein robustes und wettbewerbsfähiges Ökosystem von KI-gesteuerten Funktionen zu ermöglichen. Für das Gaming bedeutet dies direkt effektivere und qualitativ hochwertigere FSR- und Frame-Generation-Technologien, die entscheidend sind, um hohe Bildraten bei anspruchsvollen Auflösungen und Einstellungen aufrechtzuerhalten. Über das Gaming hinaus positionieren die deutlich verbesserten KI-Fähigkeiten uDNA strategisch für eine breitere Akzeptanz in kreativen

Anwendungen (z.B. KI-Kunstgenerierung, Videobearbeitung mit KI-Verbesserungen, 3D-Content-Erstellung) und potenziell sogar für eine effiziente lokale Large Language Model (LLM)-Inferenz, Bereiche, in denen NVIDIA derzeit einen dominanten Marktanteil innehat. Dieser strategische Schritt soll AMD-GPUs vielseitiger und für eine größere Bandbreite von Benutzern und Entwicklern attraktiver machen, wodurch ihr adressierbarer Markt erweitert wird.

## **Path Tracing: Erwarteter signifikanter Aufschwung**

RDNA 4 hat bereits eine außergewöhnliche **100 %ige Steigerung (fast doppelt so viele FPS)** bei hoch anspruchsvollen Path-Tracing-Workloads im Vergleich zu RDNA 3 gezeigt. uDNA (und damit RDNA 5, angesichts seiner evolutionären Natur) wird voraussichtlich **ähnliche Zuwächse bei Path Tracing im Vergleich zu RDNA 4** aufweisen. Dies stellt eine wirklich dramatische Verbesserung dar, die stark auf ein grundlegendes architektonisches Redesign oder eine tiefgreifende Optimierung speziell für diese extrem rechenintensive Rendering-Technik hindeutet. Path Tracing gilt weithin als die rechenintensivste Form des Raytracings und bietet eine unvergleichliche visuelle Realität durch hypergenaue globale Beleuchtung und Lichtsimulation. Die berichtete Leistungssteigerung von 100 % für RDNA 4 und ähnliche Zuwächse für uDNA ist ein wichtiger Indikator für AMDs zukunftsorientierte Strategie. Während volles Path Tracing derzeit auf eine Handvoll hochmoderner Titel beschränkt ist und immense Rechenleistung erfordert, positioniert dieser massive Aufschwung uDNA als außergewöhnlich fähig, wenn Path Tracing in zukünftigen Spielen und professionellen Visualisierungsanwendungen immer häufiger eingesetzt wird. Dies deutet stark darauf hin, dass AMD nicht nur beim grundlegenden Raytracing aufholt, sondern seine Architektur aktiv auf die nächste Grenze des Echtzeit-Renderings vorbereitet und die Führung bei den fortschrittlichsten visuellen Erlebnissen anstrebt.

## **Gesamte Leistungsstaffelung: Positionierung gegenüber aktuellen und gemunkelten NVIDIA-Angeboten der nächsten Generation**

Die Radeon RX 7900 XTX hat derzeit einen durchschnittlichen PassMark G3D Mark von 31.202. Die Radeon RX 9070 XT ist in der Rasterisierungsleistung "at worst on par" mit der RX 7900 XTX, aber "most often faster", und bietet eine "10% better rasterization performance" als die RX 7900 XT. Sie hat auch "better ray tracing" und "much much better upscaling" (FSR 4). Zum Vergleich: Die NVIDIA RTX 4090 hat einen durchschnittlichen G3D Mark von 38.233, was einen Vorteil von +22,5 % gegenüber der 7900 XTX darstellt. Die gemunkelte NVIDIA RTX 5090 wird mit 39.774 (+27,5 % gegenüber der 7900 XTX) und die RTX 5090 D mit 42.858 (+37,4 % gegenüber der 7900 XTX) prognostiziert. Wenn das uDNA-Flaggschiff eine 2,5- bis 3,0-fache Rasterisierungsleistung der RX 9070 XT erreicht (wie oben prognostiziert), würde dies seinen G3D Mark in den geschätzten Bereich von **70.000 - 85.000+** bringen (basierend auf einer Schätzung, dass die 9070 XT etwa 30.000-35.000 G3D Mark erreichen könnte, da sie der 7900 XTX ebenbürtig oder überlegen ist). Diese prognostizierte Leistung würde die gemunkelte NVIDIA RTX 5090/5090 D deutlich übertreffen und potenziell sogar NVIDIAs nächstes Consumer-Flaggschiff, abhängig von dessen endgültigen Spezifikationen und dem Ausmaß, in dem Blackwells Rechenzentrums-Fortschritte für den Consumer-Markt adaptiert werden. Die kombinierten prognostizierten Leistungssteigerungen für uDNA (2,5- bis 3,0-fache Rasterisierung, 2-faches Raytracing, 2-fache KI und ähnliche Zuwächse bei Path Tracing gegenüber RDNA 4) deuten auf AMDs klare und aggressive Ambition hin, im

High-End-Consumer-GPU-Markt nicht nur zu konkurrieren, sondern potenziell *zu führen*. Seit vielen Jahren hält NVIDIA weitgehend die unangefochtene Leistungsspitze im obersten Segment. Ein uDNA-Flaggschiff, das NVIDIAs Consumer-Angebote der nächsten Generation (wie die gemunkelte RTX 5090/5090 D ) deutlich übertrifft, würde NVIDIA dazu zwingen, mit extremer Aggression zu reagieren, was möglicherweise zu einem viel dynamischeren und intensiveren Wettbewerb auf dem Markt führt, der letztendlich den Verbrauchern durch Innovation und potenziell günstigere Preise zugutekommt. Dies würde einen monumentalen strategischen Sieg für AMD darstellen, der ihren vereinheitlichten Architekturansatz bestätigt und die etablierte Leistungshierarchie grundlegend in Frage stellt.

**Tabelle 2: Geschätzter Leistungszuwachs (Relativ zur RX 7900 XTX, RX 9070 XT und RTX 4090 über verschiedene Workloads)**

Workload-Kategorie	Basislinie (AMD Radeon RX 7900 XTX)	AMD Radeon RX 9070 XT (RDNA 4)	Aktueller Wettbewerber (NVIDIA GeForce RTX 4090)	Prognostizierte s AMD uDNA Flaggschiff	Prognostizierte s NVIDIA Next-Gen Flaggschiff (z.B. RTX 5090)
Traditionelle Rasterisierung	1.0x	~1.0x - 1.1x	~1.2x	~2.5x - 3.0x (vs. RX 9070 XT)	~1.3x - 1.4x (vs. RTX 4090)
Raytracing	1.0x	~1.2x - 1.5x	~1.5x - 2.0x	~2.0x (vs. RX 9070 XT)	~1.8x - 2.2x (vs. RTX 4090)
KI/Upscaling	1.0x	~1.5x - 2.0x	~1.5x - 2.0x	~2.0x (vs. RX 9070 XT)	~2.0x - 3.0x (vs. RTX 4090)
Path Tracing	1.0x	~2.0x	~1.5x - 2.5x	~2.0x (vs. RX 9070 XT)	~1.8x - 2.8x (vs. RTX 4090)
<b>Gesamtleistung (G3D Mark)</b>	31.202	~30.000 - 35.000	38.233 (+22.5%)	70.000 - 85.000+	39.774 - 42.858 (+27.5% - +37.4% vs. RX 7900 XTX)

Diese Tabelle ist absolut entscheidend, um den Aspekt "wie stark wird sie sein" der Benutzeranfrage direkt und wirkungsvoll zu quantifizieren. Sie geht über reine Spezifikationen hinaus, um die praktischen Leistungsvorteile zu veranschaulichen, die Benutzer erwarten können. Durch die Einbeziehung sowohl aktueller AMD- und NVIDIA-Karten als auch prognostizierter NVIDIA-Angebote der nächsten Generation bietet sie eine umfassende Wettbewerbslandschaft. Dies ermöglicht es dem Leser, sofort zu visualisieren, wo AMD erhebliche Fortschritte erwartet und wie es sich gegenüber seinem Hauptkonkurrenten positionieren könnte, und bietet eine klare und überzeugende visuelle Zusammenfassung der gesamten Leistungsdarstellung.

## 5. Marktauswirkungen und strategische Aussichten

**Wettbewerbslandschaft: AMDs potenzielle Position auf dem**

## High-End-GPU-Markt

Die prognostizierten substanziellen Leistungssteigerungen, insbesondere in historisch schwächeren Bereichen wie Raytracing und KI, sind explizit darauf ausgelegt, AMDs langjährige Wettbewerbsnachteile zu beheben. Wenn sich diese Prognosen bewahrheiten, könnte uDNA AMD in die Lage versetzen, NVIDIAs etablierte Dominanz im High-End-Consumer-Segment direkt herauszufordern, was potenziell zu einem erhöhten Marktanteil und einem wettbewerbsintensiveren Preiseumfeld in der gesamten Branche führen könnte. NVIDIAs Blackwell-Architektur (B200 für Rechenzentren) verfügt über erstaunliche 208 Milliarden Transistoren auf TSMCs 4NP-Prozess und liefert eine beispiellose KI/HPC-Leistung. Während spezifische Blackwell-Consumer-Details noch spärlich sind, wird NVIDIAs nächstes Consumer-Flaggschiff (RTX 5090) voraussichtlich etwa 27,5-37,4 % schneller sein als die RX 7900 XTX. Wenn uDNA sein Versprechen von einer 2,5- bis 3,0-fachen Rasterisierungsleistung gegenüber der RX 9070 XT einhält, würde es diese gemunkelten NVIDIA-Zahlen deutlich übertreffen und die Bühne für einen extrem intensiven Wettbewerb an der Spitze des Marktes bereiten. Wenn AMDs uDNA-Architektur ihre prognostizierten Leistungssteigerungen (insbesondere die 2,5- bis 3,0-fache Rasterisierung, 2-faches Raytracing, 2-fache KI und ähnliche Zuwächse bei Path Tracing gegenüber RDNA 4) wirklich liefert, birgt sie das tiefgreifende Potenzial, den High-End-GPU-Markt grundlegend neu zu gestalten. Seit vielen Jahren hält NVIDIA weitgehend die unangefochtene Leistungsspitze im obersten Segment. Ein uDNA-Flaggschiff, das NVIDIAs Consumer-Angebote der nächsten Generation (wie die gemunkelte RTX 5090/5090 D ) deutlich übertrifft, würde NVIDIA unweigerlich dazu zwingen, mit extremer Aggression und Innovation zu reagieren, was möglicherweise zu einem dynamischeren und intensiveren Wettbewerb auf dem Markt führt, der letztendlich den Verbrauchern durch beschleunigten technologischen Fortschritt und potenziell günstigere Preise zugutekommt. Dies würde einen monumentalen strategischen Sieg für AMD darstellen, der ihren vereinheitlichten Architekturansatz bestätigt und die lang etablierte Leistungshierarchie direkt in Frage stellt.

## Synergien mit der Konsolenentwicklung (PS6, Xbox)

Die uDNA-Architektur wird stark als die grundlegende Technologie gehandelt, die die nächste Generation der PlayStation (PS6) und Xbox-Konsolen antreiben wird. Diese tiefe Konsolenintegration bietet erhebliche Vorteile: Sie ermöglicht reichhaltigere Lichteffekte, detailliertere Grafiken und verbesserte KI-gesteuerte Funktionen in Spielen über alle Plattformen hinweg. Entscheidend ist auch, dass sie zu einer besseren plattformübergreifenden Spieloptimierung führt, die Portierungsbemühungen zwischen PC und Konsole vereinfacht und dazu führen könnte, dass vollwertiges Raytracing und fortschrittliche KI-basierte Frame-Generation zu Standardfunktionen in Konsolenspielen werden, nicht nur zu verwässerten Versionen. Die tiefe und erwartete Integration von uDNA in kommende Konsolengenerationen bietet einen signifikanten und oft unterschätzten Vorteil für AMDs PC-GPUs. Spielentwickler, die ihre Titel für PS6 und Xbox optimieren, werden ihre Titel von Natur aus für die uDNA-Architektur optimieren. Dies bedeutet, dass PC-Spieler, die mit uDNA-Karten ausgestattet sind, direkt von besser optimierten Titeln profitieren werden, potenziell weniger "Shader-Stottern"-Probleme erleben (ein häufiges Problem auf dem PC, das Konsolen aufgrund fester Hardware oft mildern) , und eine konsistentere Leistung im gesamten AMD-Ökosystem genießen. Diese starke plattformübergreifende Synergie kann die Einführung

der fortschrittlichen Funktionen von uDNA wie Raytracing und KI in Spielen erheblich beschleunigen, die gesamte Branche vorantreiben und AMDs Gesamtposition auf dem Gaming-Markt stärken.

## **Implikationen für Gaming- und professionelle Workloads**

Für das Gaming verspricht uDNA nicht nur höhere Bildraten, sondern auch flüssigeres Gameplay, deutlich bessere visuelle Qualität und verbesserte KI-gesteuerte Funktionen. Die prognostizierten substanziellen Gewinne bei Raytracing und Path Tracing werden die Grenzen der visuellen Wiedergabetreue verschieben und immersivere und realistischere Spielerlebnisse ermöglichen. Für professionelle Workloads machen die vereinheitlichte Architektur und die dramatisch verbesserten KI-Fähigkeiten uDNA-GPUs weitaus vielseitiger und wettbewerbsfähiger für anspruchsvolle Aufgaben wie die Inhaltserstellung (z.B. 3D-Rendering, Videobearbeitung), wissenschaftliches Rechnen sowie KI-Entwicklung und -Inferenz. Das "Wavegroup"-Modell und die verbesserte Effizienz der Compute Units werden auch hochparallelen Aufgaben zugutekommen, die in diesen professionellen Anwendungen häufig vorkommen. Durch die strategische Vereinheitlichung von RDNA und CDNA erweitert AMD explizit den Anwendungsbereich und den Zielmarkt seiner GPUs über das traditionelle Gaming hinaus. Die deutlich verbesserten KI- und Rechenfähigkeiten bedeuten, dass uDNA-Karten für professionelle Anwendungen, die derzeit stark auf NVIDIAs CUDA-Ökosystem angewiesen sind, weitaus wettbewerbsfähiger sein werden. Dieser strategische Schritt zielt darauf ab, einen größeren Anteil am schnell wachsenden Markt für KI-beschleunigte Workloads zu erobern, wodurch AMD-GPUs eine viel attraktivere und praktikablere Option für eine breitere Benutzergruppe, einschließlich Kreativer, Forscher und Entwickler, werden, nicht ausschließlich für Gamer.

## **6. Fazit: Ein Blick in AMDs High-End-Zukunft**

Die hypothetische High-End-uDNA-GPU, die voraussichtlich Ende 2026 auf den Markt kommen wird, stellt AMDs ehrgeizigsten und bedeutendsten Architektursprung seit Jahren dar. Aufbauend auf hochmodernen TSMC-Prozessknoten (höchstwahrscheinlich N2P oder A16) und mit GDDR7-Speicher der nächsten Generation verspricht sie wirklich substanzielle generationsübergreifende Leistungsverbesserungen in allen kritischen Metriken: Rasterisierung, Raytracing, KI-Beschleunigung und Path Tracing.

Diese neue GPU hat das tiefgreifende Potenzial, AMDs Position im High-End-Segment neu zu definieren, indem sie eine überzeugende Kombination aus Leistung pro Watt und einem umfassenden Funktionsumfang bietet, der NVIDIAs langjährige Führung direkt herausfordert. Ihre vereinheitlichte Architektur und tiefe Konsolenintegration werden ein stärkeres, kohärenteres Entwickler-Ökosystem fördern, das sowohl PC- als auch Konsolenspielern zugutekommt. Das Aufkommen der uDNA-Ära signalisiert AMDs klare Absicht, ein Full-Spectrum-GPU-Führer zu werden, der in der Lage ist, die sich entwickelnden und zunehmend anspruchsvollen Bedürfnisse des modernen Gamings, fortschrittlicher KI-Workloads und hochleistungsfähiger professioneller Berechnungen vollständig zu erfüllen.

### **Quellenangaben**

1. GPU IPC Showdown: NVIDIA Blackwell vs Ada Lovelace; AMD ...,

<https://www.techpowerup.com/338264/gpu-ipc-showdown-nvidia-blackwell-vs-ada-lovelace-amd-rdna-4-vs-rdna-3> 2. Better, more capable than expected: RDNA 4 architecture deep dive - HWCooling.net, <https://www.hwcooling.net/en/better-more-capable-than-expected-rdna-4-architecture-deep-dive/> 3. AMD's next-gen GPU architecture: UDNA and RDNA 5 - Jon Peddie Research, <https://www.jonpeddie.com/news/amds-next-generation-gpu-architecture-udna-and-rdna-5/> 4. AMD UDNA: The New Unified Architecture for GPUs and AI of the Future, <https://www.turtlesai.com/en/pages-2821/amd-udna-the-new-unified-architecture-for-gpus-and> 5. AMD's UDNA Architecture Explained: One GPU to Rule Them All? - Gamertech, <https://gamertech.org/singlenews/UDNA> 6. 2 nm process - Wikipedia, [https://en.wikipedia.org/wiki/2\\_nm\\_process](https://en.wikipedia.org/wiki/2_nm_process) 7. TSMC N2 NANOSHEET | Taiwan Semiconductor Manufacturing Company Limited, <https://n2.tsmc.com/english/dedicatedFoundry/technology/N2.htm> 8. TSMC's says 1.6nm node to be production ready in late 2026 — roadmap remains on track, <https://www.tomshardware.com/tech-industry/tsmcs-1-6nm-node-to-be-production-ready-in-late-2026-roadmap-remains-on-track> 9. TSMC says first 1.6nm chips coming in 2026 - The Register, [https://www.theregister.com/2024/04/25/tsmc\\_first\\_16nm\\_chips/](https://www.theregister.com/2024/04/25/tsmc_first_16nm_chips/) 10. TSMC's 1.6nm node to be production ready in late 2026 — roadmap remains on track : r/hardware - Reddit, [https://www.reddit.com/r/hardware/comments/1gx8p2n/tsmcs\\_16nm\\_node\\_to\\_be\\_production\\_ready\\_in\\_late/](https://www.reddit.com/r/hardware/comments/1gx8p2n/tsmcs_16nm_node_to_be_production_ready_in_late/) 11. NVIDIA RUBIN Series - VideoCardz.com, <https://videocardz.com/nvidia/data-center/nvidia-rubin> 12. NVIDIA Blackwell GPUs: Architecture, Features, Specs - NexGen Cloud, <https://www.nexgencloud.com/blog/performance-benchmarks/nvidia-blackwell-gpus-architecture-features-specs> 13. NVIDIA Blackwell GPU architecture: Unleashing next-gen AI performance | GenAI Research, <https://wandb.ai/onlineinference/genai-research/reports/NVIDIA-Blackwell-GPU-architecture-Unleashing-next-gen-AI-performance--VmlldzoxMjgwODI4Mw> 14. RDNA 5 2X Ray Tracing Performance: PS6, PC & Xbox Leak Analysis - YouTube, <https://www.youtube.com/watch?v=AE7MHR3L7ss> 15. rdna™ 3: beyond the current gen - GPUOpen, [https://gpuopen.com/download/RDNA3\\_Beyond-the-current-gen-v4.pdf](https://gpuopen.com/download/RDNA3_Beyond-the-current-gen-v4.pdf) 16. AMD Radeon RX 7900 XTX Specs - GPU Database - TechPowerUp, <https://www.techpowerup.com/gpu-specs/radeon-rx-7900-xtx.c3941> 17. NVIDIA - GeForce RTX 4090 24GB GDDR6X Graphics Card - Titanium/Black - Best Buy, <https://www.bestbuy.com/site/nvidia-geforce-rtx-4090-24gb-gddr6x-graphics-card-titanium-black/6521430.p?skuld=6521430> 18. NVIDIA GeForce RTX 4090 GPU - Benchmarks and Specs - NotebookCheck.net Tech, <https://www.notebookcheck.net/NVIDIA-GeForce-RTX-4090-GPU-Benchmarks-and-Specs.674574.0.html> 19. GDDR7 SDRAM - Wikipedia, [https://en.wikipedia.org/wiki/GDDR7\\_SDRAM](https://en.wikipedia.org/wiki/GDDR7_SDRAM) 20. The evolution of GDDR: From GDDR1 to GDDR7 | Micron Technology Inc., <https://www.micron.com/about/blog/memory/dram/the-evolution-of-gddr-from-gddr1-to-gddr7> 21. ADVANCED GAMING WITH AMD RDNA™ 3 ARCHITECTURE, <https://fichastecnicas.pchmayoreo.com/11322-01-40G.pdf> 22. Radeon RX 7900 XTX - Price performance comparison - Video Card Benchmarks, <https://www.videocardbenchmark.net/gpu.php?gpu=Radeon+RX+7900+XTX&id=4644> 23. AMD UDNA architecture rumored to power PS6 and next Xbox with big ray tracing and AI gains - Reddit, [https://www.reddit.com/r/Amd/comments/1lhk2xn/amd\\_udna\\_architecture\\_rumored\\_to\\_power\\_ps6\\_and/](https://www.reddit.com/r/Amd/comments/1lhk2xn/amd_udna_architecture_rumored_to_power_ps6_and/)