

Skripte und Lehrinhalte MEDT

[Startseite](#) / [Meine Kurse](#) / [SkMEDT](#) / [MEDT4](#) / [MEDT/VWSP/WAEN-Skript MEDT4](#)

MEDT/VWSP/WAEN-Skript MEDT4

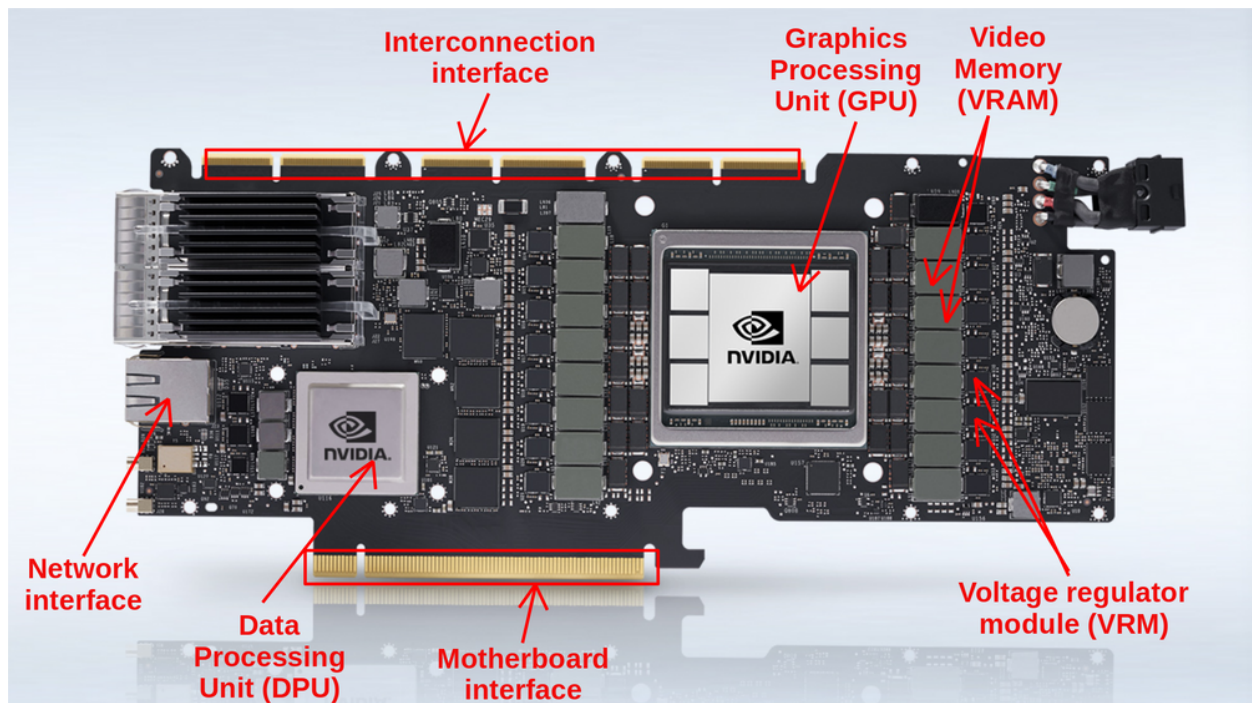
2. Grafikhardware-Technik

Zeit: 4-6 Einheiten

Aufbau einer Grafikkarte

Eine Grafikkarte besteht aus folgenden **mechanischen und elektronischen Teilen**:

- Graphics Processing Unit (GPU)
- Data Processing Unit (DPU)
- Video Memory (VRAM)
- Video BIOS (VBIOS)
- Voltage regulator module (VRM)
- Motherboard interface
- Interconnection interface
- Network interface & controller
- Ausgabe Interfaces
- Kühlung



Quelle Abbildung: Adil Lheureux

GPU

Das **Herzstück** jeder Grafikkarte ist der **Grafikprozessor (GPU)**, der die primären **Leistungsmerkmale** einer Grafikkarte definiert. Die **GPU** übernimmt die **komplexen Berechnungen** zur Darstellung von **Bildern, Videos und 3D-Grafiken** in Echtzeit. Einige High-End-Grafikkarten verfügen über mehrere GPUs, die im sogenannten **SLI- (NVIDIA) oder CrossFire- (AMD) Verbund** arbeiten, um die Rechenleistung weiter zu steigern. Diese **Multi-GPU-Systeme ermöglichen eine höhere Auflösung und bessere Frameraten**, insbesondere in grafikintensiven Anwendungen wie Spielen und professionellen Rendering-Aufgaben.

Die **GPU** besteht üblicherweise aus **4 Komponenten**:

- **Streaming Multiprocessors (SMs)**
- **Load/Store (LD/ST) units**
- **Special Function Units (SFU)**
- **Texture Mapping Unit (TMU)**

Ein **Streaming Multiprozessor (SM)** ist eine **Ausführungseinheit**, die aus einer **Sammlung von Kernen** besteht. Mehrere **Threads** können gleichzeitig in einem SM ausgeführt werden. Wir unterscheiden 2 Technologien:

- **Compute Unified Device Architecture (CUDA)** oder **Tensor Kerne** von **NVIDIA**
- **Stream-Prozessoren** von **AMD**

NVIDIA **CUDA Cores und Tensor Cores** gelten im Allgemeinen als **stabiler und optimierter**, insbesondere für **maschinelle Lernanwendungen**. **CUDA-Kern** waren auf jeder Nvidia GPU präsent, die in den letzten zehn Jahren veröffentlicht wurde, aber **Tensor Cores** sind eine **neuere Ergänzung**. **Tensor-Kerne** sind **viel schneller** als **CUDA-Kern** bei der Datenverarbeitung. In der Tat können **CUDA-Kerne** nur eine Operation in jedem Taktzyklus durchführen, aber **Tensor-Kerne** können mehrere Operationen pro Zyklus ausführen.

Load/Store (LD/ST)-Einheiten ermöglichen es, dass **Threads mehrere Daten- und Speicheroperationen pro Taktzyklus** ausführen.

Spezielle Funktionseinheiten (SFUs) berechnen **effizient strukturierte arithmetische oder mathematische Funktionen** von **Vektordaten** – zum Beispiel Sinus, Cosinus, Wurzel...

Texture Mapping Unit (TMU) übernimmt **anwendungsspezifische Aufgaben** wie **Bildrotation, Skalierung, Hinzufügen von Verzerrungen und Rauschen und Bewegen von 3D-Ebenen**.

Kühlung und Wärmeableitung

Aufgrund der hohen Rechenleistung erzeugen GPUs erhebliche Mengen an Wärme. Daher werden Grafikkarten entweder **passiv oder aktiv gekühlt**. Passiv gekühlte Karten verwenden **große Kühlkörper**, um die Hitze ohne den Einsatz von Lüftern abzuleiten, was sie lautlos, aber weniger effizient macht. **Aktive Kühlsysteme kombinieren Kühlkörper mit Lüftern**, die die Wärme **schneller und effektiver abführen können**. **Hochleistungsmodelle** setzen zudem oft auf eine **Flüssigkeitskühlung**, die besonders effektiv ist, aber auch höhere Kosten und Komplexität mit sich bringt.

Grafikspeicher (VRAM)

Um die enormen Datenmengen zu verarbeiten, die für hochauflösende und komplexe Grafiken erforderlich sind, verfügen Grafikkarten über einen **dedizierten Grafikspeicher (VRAM)**. Dieser ist über eine hochleistungsfähige Speicheranbindung, wie etwa **GDDR6 oder GDDR6X**, mit der GPU verbunden. Der **VRAM** arbeitet mit hohen **Taktraten** und breiten **Speicherbussen** (z.B. 256-Bit oder 384-Bit), um die benötigten **Daten schnell und effizient bereitzustellen**. Die Größe des **Grafikspeichers** kann bis zu **24 GB oder mehr betragen**, was besonders für Anwendungen wie **3D-Modellierung, Videoschnitt oder Spiele in 4K-Auflösung** von Vorteil ist. Der Zugriff auf den Grafikspeicher wird durch einen **integrierten Speichercontroller** gesteuert, der für eine optimale Datenübertragung sorgt.

RAM-DAC

Der **RAM-DAC** (Random Access Memory Digital-to-Analog Converter) war **früher** ein wesentlicher Bestandteil von Grafikkarten und diente dazu, **digitale Bildinformationen in analoge Signale für ältere VGA-Monitore** umzuwandeln. Die Leistungsfähigkeit des RAM-DAC wird in MHz angegeben, wobei höhere Werte höhere Auflösungen und Farbtiefen ermöglichen. Mit dem Aufkommen von digitalen Schnittstellen wie DVI, HDMI und DisplayPort hat der RAM-DAC jedoch an Bedeutung verloren. Moderne Grafikkarten besitzen dennoch oft einen RAM-DAC, um Kompatibilität mit älteren VGA-Anschlüssen zu gewährleisten.

Bus-Interface DPU

Das **Bus-Interface** verbindet die Grafikkarte mit dem restlichen Computersystem und ermöglicht den **Datenaustausch zwischen GPU und CPU**. Die meisten modernen Grafikkarten nutzen den **PCI Express (PCIe)** Standard, der in Versionen von PCIe 3.0 bis PCIe 5.0 verfügbar ist. PCIe 4.0 bietet beispielsweise eine Datenübertragungsrate von bis zu 16 GT/s (Gigatransfers pro Sekunde) pro Lane, was insgesamt eine Bandbreite von 32 GB/s (bidirektional) bei einer x16-Verbindung ermöglicht. Ältere Schnittstellen wie **AGP oder PCI** werden in aktuellen Systemen kaum noch genutzt.

GPU und Parallelisierung

Die **enorme Rechenleistung** einer **GPU resultiert** hauptsächlich aus ihrer Fähigkeit zur **Parallelverarbeitung**. Im Gegensatz zur CPU, die für die Ausführung allgemeiner Berechnungen optimiert ist, besteht eine GPU aus Tausenden **von Recheneinheiten (Kernen, SMs)**, die gleichzeitig arbeiten können. Diese Parallelität ist besonders bei der Verarbeitung von **Grafiken und Video-Rendering effektiv**, wo viele ähnliche Berechnungen gleichzeitig durchgeführt werden können. Typische Anwendungen wie **HD-Video-Transcoding oder Echtzeit-3D-Rendering** profitieren stark von dieser Architektur. Moderne GPUs integrieren zudem spezielle Hardware für Aufgaben wie **Raytracing oder KI-Berechnungen**, was sie noch leistungsfähiger und vielseitiger macht.

Shader

Hardware-Shader (Shadereinheiten) sind **kleine Recheneinheiten in Grafikchips**, die seit DirectX 8 und OpenGL 2.0 für **3D-Effekte** genutzt werden. **Fragment-Shader** berechnen die **Pixelfarbe**, während **Vertex-Shader** für **geometrische Berechnungen** zuständig sind. Neuere Shader-Typen sind der **Geometry-Shader** (DirectX 10/OpenGL 3.2), der zusätzliche Geometrie **erzeugen** kann, und der **Tessellation-Shader** (DirectX 11/OpenGL 4.0), der die Geometrie weiter **verfeinert**. Shader sind **integrale Bestandteile des Grafikchips** und werden durch den Grafiktreiber an die Standards DirectX und OpenGL angepasst. Eine typische GPU verfügt über mehrere **Tausend Shader**, die entweder als Unified Shader oder als spezialisierte Shader agieren. Vor der Einführung von DirectX 10 wurden Vertex-Shader und Pixel-Shader getrennt behandelt, was oft zu einer unausgewogenen Auslastung führte. Unified Shader können hingegen flexibel je nach Bedarf als Vertex- oder Pixel-Shader eingesetzt werden. Dies sorgt für eine gleichmäßigere Auslastung und bessere Leistung.

Texturen und Speicherverwaltung

Texturen sind wesentliche Elemente in der 3D-Grafik, da sie **Oberflächen mit Details versehen** und **realistischer erscheinen** lassen. In modernen Spielen und Anwendungen nehmen **Texturen den größten Teil des VRAM** ein. Die GPU lädt die **benötigten Texturen in den Grafikspeicher**, um sie schnell zugänglich zu machen. Wenn der VRAM nicht ausreicht, müssen Texturen häufig von der **Festplatte nachgeladen** werden, was zu Leistungseinbußen und Ruckeln (sogenanntes „Stuttering“) führen kann. Deshalb ist die Größe des **VRAMs ein entscheidender Faktor für die flüssige Darstellung in hochauflösenden Spielen und Anwendungen**. Je größer der Speicher, desto mehr Texturen können vorab geladen und effizient genutzt werden, was die Gesamtleistung der Grafikkarte erhöht.

Bildverbesserung

Um 3D-Darstellungen realistischer erscheinen zu lassen, bieten moderne Grafikprozessoren (GPUs) eine Vielzahl von **Bildverbesserungsfunktionen**. Eine dieser Funktionen ist die **anisotrope Filterung**, die Texturen auf schräg liegenden Flächen **schärfer und detailreicher darstellt**. Dies verbessert die **visuelle Qualität**, insbesondere bei weit entfernten oder schräg betrachteten Oberflächen, die ansonsten unscharf erscheinen würden.

Antialiasing ist eine weitere wichtige Technologie, die dazu dient, das sogenannte **Treppeneffekt-Problem (Aliasing)** an den Kanten von Objekten zu reduzieren. Beim Rendern von Bildern mit hoher Auflösung berechnet die GPU das Bild intern in einer höheren Auflösung und skaliert es dann auf die Zielauflösung herunter. Verschiedene Antialiasing-Techniken wie MSAA (Multisample Anti-Aliasing), SSAA (Supersample Anti-Aliasing) oder FXAA (Fast Approximate Anti-Aliasing) glätten die Kanten, indem sie die Farbinformationen der angrenzenden Pixel berücksichtigen. So wirkt das Bild insgesamt **weicher und realistischer**.

Eine weitere fortschrittliche Technik ist die **Tessellation**. Tessellation erhöht die **Detailgenauigkeit von 3D-Modellen**, indem **zusätzliche Geometrie** (Punkte und Dreiecke) **dynamisch zu den Objekten hinzugefügt** wird. Diese Methode wird häufig verwendet, um Oberflächen realistischer zu gestalten, beispielsweise bei der Darstellung von **Falten in Kleidung** oder **Unebenheiten in Geländetexturen**. Tessellation wird von modernen Grafik-APIs wie DirectX 11 und höher unterstützt und ist ein Schlüsselement für die Darstellung detailreicher 3D-Szenen.

HDR - High Dynamic Range

High Dynamic Range (HDR) ist eine Technologie, die einen erweiterten Kontrastumfang und eine intensivere Farbdarstellung ermöglicht. HDR-fähige Inhalte bieten eine breitere Palette an Helligkeitsstufen, sodass sowohl sehr helle als auch sehr dunkle Bereiche eines Bildes detaillierter und realistischer dargestellt werden können. Dies führt zu Bildern mit lebendigeren Farben, tiefschwarzen Schatten und strahlend weißen Highlights. HDR wird häufig in modernen Spielen, Filmen und Fotografien verwendet und erfordert sowohl eine HDR-kompatible Grafikkarte als auch ein HDR-fähiges Display. Es ist wichtig, echtes HDR von einfachen 3D-Effekten zu unterscheiden, die fälschlicherweise als "HDR" bezeichnet werden.

3D-Beschleunigung

3D-Beschleunigung wird durch **Grafik-APIs (Application Programming Interfaces) wie DirectX von Microsoft und OpenGL** ermöglicht. Diese Schnittstellen abstrahieren die Hardware-Funktionalitäten der GPU, sodass Entwickler plattformübergreifend arbeiten können, ohne spezifischen Code für jede einzelne Grafikkartenarchitektur schreiben zu müssen. DirectX bietet eine umfassende Sammlung von APIs für verschiedene Aufgaben, einschließlich der Grafik- und Spielprogrammierung, und ist vor allem in Windows-Umgebungen weit verbreitet. OpenGL hingegen ist ein offener Standard, der auf verschiedenen Plattformen, einschließlich Linux und macOS, unterstützt wird.

Vulkan ist eine moderne 3D-Schnittstelle, die eine Low-Level-API bietet. Dies ermöglicht Entwicklern einen direkten Zugriff auf die Hardware, was eine effizientere Nutzung von CPU und GPU ermöglicht. Vulkan ist plattformunabhängig und eignet sich sowohl für Desktop-PCs als auch für mobile Geräte. Im Vergleich zu DirectX 12 und OpenGL bietet Vulkan eine bessere Kontrolle über die Hardware, was zu einer höheren Leistung und besseren Skalierbarkeit führt, insbesondere bei Anwendungen, die intensive Grafikberechnungen erfordern.

VR - Virtual Reality

Virtual Reality (VR) ist eine Technologie, die es Benutzern ermöglicht, in eine vollständig immersive, computergenerierte Umgebung einzutauchen. Die aktuellen Fortschritte in der VR-Technologie, insbesondere durch VR-Headsets wie Oculus Rift, HTC Vive und PlayStation VR, haben die Immersion und die Qualität der VR-Erfahrungen erheblich verbessert. Für eine reibungslose VR-Erfahrung sind jedoch leistungsstarke Grafikkarten erforderlich, die hohe Bildwiederholraten und niedrige Latenzzeiten bieten, um Bewegungsunschärfe und Übelkeit zu vermeiden.

Obwohl VR bereits beeindruckende visuelle und interaktive Erlebnisse bietet, stehen der breiten Adoption noch einige Herausforderungen im Weg. Dazu gehören die hohen Kosten der Hardware, die Notwendigkeit leistungsstarker Computer und die begrenzte Verfügbarkeit von Inhalten, die das volle Potenzial der VR-Technologie ausschöpfen. Trotzdem findet VR zunehmend Anwendungen in Bereichen wie Bildung, Medizin, Architektur und Unterhaltung.

Adaptive-Sync/G-Sync/Freesync

Adaptive-Sync ist ein von der VESA (Video Electronics Standards Association) entwickelter Standard, der Teil des DisplayPort-Protokolls ab Version 1.2a ist. Diese Technologie passt die Bildwiederholfrequenz des Monitors dynamisch an die Bildausgabe der Grafikkarte an, um Ruckeln (Stuttering) und Bildschirmzerreißen (Tearing) zu verhindern, die auftreten können, wenn die Bildrate der Grafikkarte nicht mit der festen Bildwiederholrate des Monitors übereinstimmt. AMD FreeSync basiert auf diesem Standard und wird von vielen Monitoren und Grafikkarten unterstützt. NVIDIA G-Sync ist eine proprietäre Technologie, die ähnlich wie Adaptive-Sync funktioniert, aber speziell für NVIDIA-Grafikkarten entwickelt wurde. G-Sync-Monitore enthalten spezielle Hardwaremodule, die die Synchronisierung noch präziser steuern und dadurch eine optimale Bildqualität gewährleisten. Beide Technologien sind besonders bei Gamern beliebt, da sie eine flüssigere und reaktionsschnellere Spielerfahrung ermöglichen.

Ähnlich wie G-Sync ist FreeSync eine proprietäre Technologie von AMD.

	FreeSync	FreeSync Premium	FreeSync Premium Pro
Verzerrungsfreies Erlebnis	✓	✓	✓
Niedrige Latenz	✓	✓	✓
Kompensation niedriger Bildraten (LFC)	✓	✓	✓
< 3.440 horizontale Auflösung: maximale Aktualisierungsrate	≥ 144 Hz	≥ 200 Hz	≥ 200 Hz
≥ 3.440 horizontale Auflösung: maximale Aktualisierungsrate		≥ 120 Hz	≥ 120 Hz
AMD FreeSync HDR			✓

Übertakten/Untertakten/Undervolting einer Grafikkarte

Das Übertakten einer Grafikkarte erhöht die Taktraten der GPU und des Speichers über die vom Hersteller vorgegebenen Werte hinaus. Dies kann zu einer **Steigerung der Grafikleistung** führen, allerdings nur um einen begrenzten Prozentsatz. Tools wie **MSI Afterburner** oder **ASUS GPU Tweak** ermöglichen eine einfache Anpassung von Taktfrequenz und Spannung. Allerdings birgt das Übertakten Risiken, wie eine erhöhte Wärmeentwicklung, die zu Instabilität, Bildfehlern und im schlimmsten Fall zur Beschädigung der Grafikkarte führen kann.

Werksübertaktete Grafikkarten sind eine sichere Alternative, da sie vom Hersteller getestet und für den Betrieb mit höheren Taktraten freigegeben sind. Diese Karten bieten eine höhere Leistung, ohne die Risiken des manuellen Übertaktens.

Undervolting ist der **absichtliche Betrieb der GPU mit niedrigerer Spannung**. Dies führt zu **geringerer Leistung**, aber auch zu **weniger Leistungsaufnahme und geringeren Kerntemperaturen**.

Schnittstellen für Grafikkarten

Grafikkarten wurden ursprünglich über den internen Systembus verbunden, was jedoch mit der wachsenden Nachfrage nach höherer Bandbreite für Multimedia- und 3D-Anwendungen nicht mehr ausreichte. **AGP** (Accelerated Graphics Port) wurde von Intel als dedizierter Anschluss für Grafikkarten eingeführt, um die Bandbreite für Grafikdaten zu erhöhen. AGP wurde schließlich durch **PCI Express (PCIe)** ersetzt, einen **universellen Steckplatz**, der heute als **Standard** für die Verbindung von Grafikkarten mit dem Mainboard dient. PCIe bietet nicht nur höhere Bandbreiten, sondern auch eine bessere Skalierbarkeit mit verschiedenen Versionen (z.B. PCIe 3.0, 4.0, 5.0), die jeweils höhere Datenübertragungsraten bieten.

Anschlüsse

Der **DisplayPort** wurde entwickelt, um ältere Standards wie **DVI** (Digital Visual Interface) und **HDMI** (High Definition Multimedia Interface) abzulösen. Er unterstützt **höhere Auflösungen, schnellere Bildwiederholraten und erweiterte Farbtiefen**, was ihn ideal für moderne Displays macht. DisplayPort ist auch in der Lage, mehrere Monitore über eine einzige Schnittstelle anzusteuern (Multi-Stream Transport, MST). Zusätzlich unterstützt er **HDCP** (High-bandwidth Digital Content Protection) sowie das eigene **Kopierschutzprotokoll DPCP** (DisplayPort Content Protection), um die Sicherheit digitaler Inhalte zu gewährleisten. HDMI bleibt jedoch weiterhin weit verbreitet, insbesondere in der Unterhaltungselektronik, während DisplayPort eher im PC-Bereich dominiert.

◀ [Medientechnik, Wahrnehmung und Gestaltung Skript](#)

Direkt zu:

[Einführung DAW Reaper ▶](#)

✉ [Website-Support](#)

Sie sind angemeldet als [Markus Meyerhofer](#) ([Logout](#))

[SkMEDT](#)

[Laden Sie die mobile App](#)