

Placas Gráficas em Portáteis

Afonso Sampaio e Maria Beatriz Rocha

8 de novembro de 2023

Conteúdo

I	Introdução às Placas Gráficas de Portáteis	2
1	Definição e História	3
1.1	Definição e função das placas gráficas em portáteis	3
1.1.1	Definição	3
1.1.2	Funções	3
1.2	Evolução histórica das placas gráficas para portáteis	4
1.2.1	Início dos portáteis (Década de 1980)	4
1.2.2	Década de 1990	4
1.2.3	Década de 2000	4
1.2.4	Década de 2010	4
1.2.5	Década Atual	4
II	GPUs	5
2	Arquitetura de uma GPU	6
2.1	Primórdios de GPUs	6
2.2	GPUs Atuais	6
2.2.1	Arquiteturas da NVIDIA	6
2.2.2	Arquiteturas da AMD	8
3	Tipos de GPUs	9
3.1	GPUs Integradas vs GPUs Dedicadas	9

Parte I

Introdução às Placas Gráficas de Portáteis

Capítulo 1

Definição e História

1.1 Definição e função das placas gráficas em portáteis

As placas gráficas, também conhecidas como GPUs (Unidades de Processamento Gráfico) ou placas de vídeo, desempenham um papel fundamental em portáteis, desktops e outros dispositivos de computação quando se trata de processamento de gráficos e de vídeos.

1.1.1 Definição

Uma GPU é um componente de hardware dedicado, projetado especificamente para lidar com o processamento de gráficos e de vídeos em um computador.

1.1.2 Funções

Renderização de Gráficos

A função principal de uma GPU de um portátil é renderizar¹ gráficos. Isso inclui a exibição de elementos visuais na tela, como imagens, texto, vídeos e animações. A GPU alivia a carga da CPU (Unidade Central de Processamento) ao processar esses gráficos, melhorando o desempenho geral do sistema.

¹Obter o produto final de um processamento digital

Gráficos 3D e Jogos

São essenciais em portáteis que sejam usados em jogos ou tarefas que envolvem gráficos 3D, como modelagem 3D ou renderização de vídeos, uma vez que é responsável pelo processamento de cálculos complexos necessários para renderizar gráficos 3D em tempo real, proporcionando uma experiência de jogo suave e uma renderização rápida de projetos 3D.

Aceleração de Vídeo

Oferecem aceleração de vídeo, o que significa que elas podem decodificar e codificar vídeos de maneira mais eficiente do que a CPU. É importante para a visualização de vídeos em altas resoluções e para edição de vídeos, visto que ajuda a economizar tempo e energia e melhora a qualidade da reprodução.

Processamento Paralelo

São, também, altamente eficientes em realizar cálculos em paralelo, o que as torna ideais para tarefas que exigem grande capacidade de processamento, como Machine Learning², criptomining³, entre outros. Muitos portáteis modernos utilizam GPUs dedicadas para acelerar estas tarefas.

1.2 Evolução histórica das placas gráficas para portáteis

1.2.1 Início dos portáteis (Década de 1980)

Nos primórdios dos portáteis, as placas gráficas eram geralmente integradas na motherboard e ofereciam um desempenho muito limitado. A maioria usava telas monocromáticas ou gráficos básicos e não havia muita capacidade gráfica dedicada.

1.2.2 Década de 1990

À medida que os portáteis se foram tornando mais populares, houve uma demanda crescente por melhores capacidades gráficas. A NVIDIA e a ATI (mais tarde adquirida pela AMD) começaram a produzir GPUs dedicadas, para portáteis. No entanto, ainda eram relativamente fracas em comparação com as versões de desktop.

1.2.3 Década de 2000

Na década de 2000 houve um aumento significativo no poder gráfico dos portáteis. As GPUs dedicadas tornaram-se mais poderosas e capazes de executar jogos e aplicações gráficas exigentes. Tecnologias como o SLI (Scalable Link Interface) da NVIDIA e o CrossFire da ATI permitiam aos portáteis serem equipados com múltiplas GPUs.

1.2.4 Década de 2010

Nesta década, houve uma revolução na indústria dos portáteis devido ao aumento do interesse em jogos de computadores. Fabricantes como a NVIDIA e a AMD começaram a lançar GPUs de alto desempenho especificamente projetadas para portáteis gaming. Estes tornaram-se populares, com telas de alta taxa de atualização e resoluções mais altas.

1.2.5 Década Atual

Hoje em dia a evolução de GPUs em portáteis continua. A NVIDIA recentemente lançou a sua última geração de GPUs, as RTX 40 para portáteis, que são baseadas na arquitetura Ada Lovelace e que oferecem um desempenho bastante competitivo superando os seus rivais da AMD com as suas GPUs Radeon RDNA3.

²Dá aos computadores a capacidade de identificar padrões em dados massivos e fazer previsões (análise preditiva)

³Usar recursos de um sistema para resolver grandes cálculos matemáticos que resultam em alguma quantidade de criptomoeda sendo concedida aos solucionadores.

Parte II

GPUs

Capítulo 2

Arquitetura de uma GPU

A arquitetura de uma GPU refere-se ao design e à organização interna das unidades de processamento gráfico em uma GPU. Esta arquitetura determina como esta vai processar dados gráficos, realizar cálculos e exibir imagens na tela do computador. Ao longo dos anos, várias arquiteturas de placas gráficas foram desenvolvidas por diferentes fabricantes, com avanços significativos em termos de desempenho e eficiência.

2.1 Primórdios de GPUs

As GPUs passaram por uma evolução desde a Arquitetura de Barramento Fixo, que era baseada em funções fixas sem programação; passando pela Arquitetura de Transformação e Iluminação(T&L) que consistia em separar estas duas operações do processamento dos pixels, fornecendo assim um melhor desempenho; até chegarmos à Arquitetura de Shader Unificado, que permitiu a execução de tarefas programáveis, incluindo o uso de sombreadores para efeitos visuais mais avançados.

2.2 GPUs Atuais

2.2.1 Arquiteturas da NVIDIA

A NVIDIA lançou várias arquiteturas de GPUs, mas iremos apenas citar as mais conhecidas e as que iremos utilizar nos testes de desempenho:

- **NVIDIA Maxwell (2014):** A grande mudança começou aqui. As GPUs tornaram-se mais eficientes; houve melhorias de shaders¹; foi a primeira a ter suporte total para DirectX² 12; permitia o controlo total sobre a frequência e voltagem da GPU; o aparecimento do HEVC(H.265)³, o que resultou em melhor qualidade de vídeo e menor uso da CPU; a primeira a ter suporte para G-SYNC⁴, eliminando assim o tearing⁵.
- **NVIDIA Pascal (2016):** É a arquitetura mais comum de todas, visto que possui uma boa eficiência energética juntamente com um desempenho sólido. Apesar de não ser das arquiteturas mais recentes, continua a ter suporte para quase todas as tecnologias modernas lançadas hoje em dia, para além disso, podem ser compradas a preços acessíveis.

¹Programa de computador usado para fazer shading: a produção de níveis de cor apropriadas para uma imagem e produzir efeitos especiais ou pós-processamento de vídeo.

²Conjunto de APIs que permite aos softwares dar instruções diretas para os componentes de hardware de áudio e de vídeo, melhorando assim o desempenho das aplicações na execução de recursos multimédia

³Formato de vídeo com alta eficiência de compressão para vídeos de alta resolução

⁴Tecnologia da Nvidia que elimina a "quebra" de quadros em jogos, presentes nos monitores gamer

⁵Fenómeno que cria uma imagem distorcida

- **NVIDIA Turing (2018)**: Usada nas graficas da série RTX20 onde se deu uma grande evolução: foram incluídos núcleos de Ray Tracing⁶, bem como núcleos de inteligência artificial; introdução da tecnologia DLSS⁷ (Deep Learning Super Sampling); uma maior performance e eficiência global bem como um aumento muito significativo do desempenho em RV; a primeira a ter suporte para DirectX 12 Ultimate; e a primeira a ter suporte para novas tecnologias de resolução, conseguindo suportar resoluções até 4K 144HZ.

⁶Técnica avançada de renderização gráfica usada para simular o comportamento realista da luz em ambientes virtuais, como jogos, animações, filmes e aplicações de modelagem 3D.

⁷Tecnologia desenvolvida pela NVIDIA para melhorar o desempenho gráfico em jogos e aplicativos, ao mesmo tempo em que mantém ou até mesmo melhora a qualidade da imagem, utilizando a inteligência artificial para atingir estes mesmos objetivos.

- **NVIDIA Ampere (2020):** Usada nas gráficas da série RTX30 e nas séries para profissionais Quadro, que tinha as seguintes vantagens: um desempenho muito superior em relação à geração anterior, tendo também uma maior eficiência energética; uma grande melhoria nos núcleos de Ray Tracing tornando este método bastante fluido e responsivo; melhoria nos núcleos de inteligência artificial; a capacidade de executar computação de alto desempenho sendo capaz de fazer a análise de dados e computação de alto desempenho; suporte para resoluções de 8K 60Hz
- **NVIDIA Ada Lovelace (2022):** Usada nas gráficas da série RTX40 que trouxe: grandes aumentos no desempenho e na eficiência das GPUs; melhoria dos núcleos de Ray Tracing bem como nos núcleos de inteligência artificial; suporte para tecnologia DLSS 3; suporte para o codec AV1⁸.

2.2.2 Arquiteturas da AMD

Do outro lado da concorrência temos a AMD que ao longo dos anos também lançou arquiteturas bastante diferentes e com um desempenho escalar, apesar de andar sempre atrás dos calcanhares da NVIDIA não criando, assim, muita tecnologia nova e que as distinguisse das da NVIDIA. Aqui seguem as mais conhecidas e que têm uma rivalidade direta com as mencionadas da NVIDIA:

- **AMD Vega (2017):** Uma arquitetura com características muito semelhantes às da NVIDIA Pascal tendo a vantagem de esta se encontrar a um preço mais acessível e oferecer um desempenho similar; usa a tecnologia HBM2 que tem uma largura de banda de memória muito superior ao tipo GDDR5 e GDDR5X encontrado em placas NVIDIA.
- **AMD RDNA(2019):** A AMD quase conseguiu igualar a NVIDIA nesta geração em GPUs a preços mais acessíveis; continuavam a ter uma largura de banda de memória maior mas em contrapartida estas ainda não possuíam núcleos de Ray Tracing nem núcleos de inteligência artificial como a NVIDIA além do que tinham uma melhor eficiência energética.
- **AMD RDNA 2 (2020):** A segunda geração da RDNA trouxe algumas melhorias bastante consideráveis em relação à geração anterior: o aparecimento da tecnologia Infinity Cache⁹ que foi e continua a ser um dos grandes pontos fortes da AMD; o aparecimento de núcleos de Ray Tracing mas que infelizmente não conseguem sequer superar os da primeira geração da NVIDIA para além de que continuam a não possuir núcleos de inteligência artificial; a banda de memória continua a ser maior no lado da AMD; e continuam a ter um preço bem mais acessível do que as da NVIDIA e também são mais eficientes, no geral.
- **AMD RDNA 3 (2022):** A terceira geração da tecnologia RDNA veio melhorar as coisas introduzidas na geração anterior: a segunda geração da tecnologia Infinity Cache surgiu e veio aumentar ainda mais a banda da memória; melhoria dos núcleos de Ray Tracing que continua a não ser o ponto forte da AMD conseguindo apenas igualar a geração arquitetura Ampere; implementação de núcleos de AI; continua a estar a um preço muito mais acessível do que a concorrência mas esta arquitetura é também menos eficiente em relação à arquitetura Ada Lovelace; suporte a displayports 2.1 capazes de suportar resoluções 16K 60Hz.

⁸Codec de compressão de vídeo que oferece licenciamento sem royalties e que pode atingir um desempenho até 30% melhor que o H.265

⁹Tecnologia que permite que a GPU acesse dados frequentemente utilizados mais rapidamente tornando-a assim mais eficiente

Capítulo 3

Tipos de GPUs

3.1 GPUs Integradas vs GPUs Dedicadas

As placas gráficas integradas e dedicadas são dois tipos de GPUs que são encontradas em portáteis.

Paramêtros	GPU Integrada	GPU Dedicada
Defenição	Está incorporada na CPU do portátil e compartilha a memória do sistema (RAM) com o processador principal	É uma unidade gráfica separada que tem a sua própria memória dedicada (VRAM) e é projetada especificamente para tarefas gráficas mais intensivas, como jogos, edição de vídeo e modelagem 3D
Desempenho	Têm um desempenho gráfico muito em comparação com as dedicadas, sendo mais adequadas para tarefas cotidianas, como navegação na web, processamento de texto e reprodução de vídeos	Têm um desempenho gráfico muito em comparação com as integradas, permitindo ao utilizador executar jogos e aplicações mais exigentes com maior fluidez e qualidade gráfica