Procesadores Específicos de Dominio

Enrique Reyes González

January 2024

Universidad de Castilla la Mancha

Contents

1 Resumen libro

Esta guía proporciona pautas para desarrollar procesadores de dominio, no solo para el beneficio de las aplicaciones de usuario, sino también para simplificar el diseño y ahorrar costos al fabricante. Las pautas incluyen:

1. Usar memorias dedicadas para minimizar la distancia en la que los datos se mueven.
2. Aprovechar los recursos ahorrados por no hacer optimizaciones de microarquitecturas más avanzadas y llevarlos a más unidades aritméticas o memorias más grandes.
3. Usar la forma más simple de paralelismo para nuestro dominio.
4. Reducir los datos y usar el tipo de datos más simple para nuestro dominio.
5. Usar lenguajes de programación específicos de dominio para llevar código a estos procesadores.

1.1 Ejemplos de Dominios

Esta sección explora ejemplos de procesadores específicos de dominio aplicados a redes neuronales, incluyendo:

1. Redes Neuronales Profundas (DNN).
2. Redes Neuronales Convolucionales (CNN).
3. Redes Neuronales Recurrentes (RNN).

1.2 Unidad de Procesamiento Tensorial (TPU) de Google

Examina diferentes enfoques de TPUs, desde sus orígenes en Google hasta su arquitectura, set de instrucciones, microarquitectura, implementación y software.

1.3 Microsoft Catapult

Descripción del proyecto de Microsoft que implementa FPGAs en una placa bus PCI para servidores de datos. El objetivo es utilizar la flexibilidad de las FPGAs para adaptar su uso a diferentes aplicaciones a lo largo del tiempo.

1.4 CPUs vs GPUs vs Aceleradores DNN

Realización de benchmarks a diferentes tipos de DNN y roofline de estos benchmarks.

1.5 Falacias y Trampas

Exploración de conceptos relacionados con falacias y demostración de su falsedad.

1.6 Conclusiones

Resumen del capítulo, destacando el renacimiento de la arquitectura de computadores después de Moore y la velocidad con la que se avanza.

2 System On Chip

System On Chip elegido: Snapdragon 865.

2.1 Diagrama de Bloque Simple sobre del Snapdragon 865.

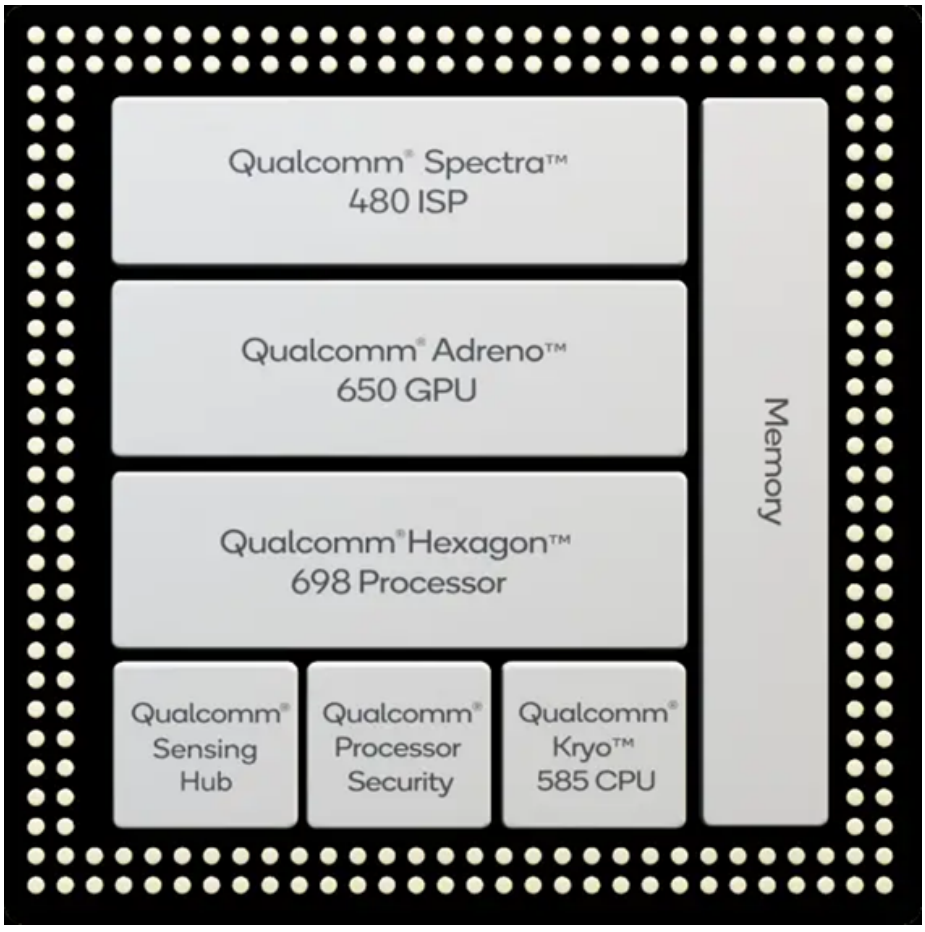

Figure 1: Composicion del Snapdragon 865

2.2 Componentes

2.2.1 Qualcomm Spectra 480 ISP

ISP (Image Signal Processor) es un procesador de imágenes que puede procesar hasta dos gigapíxeles por segundo. Ofrece captura de video hasta 8k, captura de video Dolby Vision, grabación a 60 fps en cámara lenta sin límite de tiempo, captura de fotos a 200 megapíxeles y grabación de video a 4K HDR con 64 MP de captura de fotos. También incluye Snapdragon LLV (Low Light Video) para adaptarse a entornos oscuros y capturar fotos más visibles, así como Touch to Track para enfocarse en objetos durante la grabación.

2.2.2 Qualcomm Adreno 650 GPU

La GPU integrada en el chip es Adreno 650, que tiene un 50% más de unidades de cómputo (ALU) que su predecesora, Adreno 640. La velocidad de los píxeles shaders es un 50% más rápida que la Adreno 640. En 2021, solo la GPU Apple A13 pudo superar a la Adreno 650 en gráficos para móviles y tabletas.

2.2.3 Qualcomm Hexagon 698 Processor

Hexagon 698 Processor es parte de la familia de procesadores de señales digitales (DSP) llamados Hexagon. Estos son microprocesadores especializados con una arquitectura optimizada para operaciones en procesamiento de señales digitales (DSP), que involucran numerosas operaciones aritméticas.

2.2.4 Microarquitectura

Las instrucciones de Hexagon operan con ints, floats y coma flotante. La unidad de procesamiento puede dispatchar hasta 4 instrucciones a la vez y llevarlas a 4 unidades de ejecución cada ciclo. La arquitectura es multithreading y soporta very long instruction words (VLIW), diseñada para un rendimiento óptimo con mínima potencia. Utiliza dynamic multithreading, permitiendo cambiar dinámicamente entre diferentes hilos de ejecución.

2.2.5 Capacidad de AI Engine

El Snapdragon 865 incorpora la 5ª generación de on-device AI engine basada en el DSP Hexagon 698, con una capacidad de 15 trillones de operaciones por segundo (TOPS).

2.2.6 Software

El procesador funciona sobre un puerto de Linux en una capa de máquina virtual llamada "Hexagon Virtual Machine". Originalmente cerrada, en abril de 2013 se implementó una versión semiabierta llamada "Hexagon MiniVM". El compilador es Hexagon/HVX V66 ISA con soporte para la versión 8.0.0 de LLVM.

versión Hexagon	Tamaño de nodo (nm)	Año de salida	Numero de hilos simultáneos	Encontrado en
698	7	2020	15 TOPS	Snapdragon 865/865+/870

Figure 2: Enter Caption

2.2.7 Codec

Modelo	MPEG-4	H263	VC-1	H.264	H.264 10-bit	VP8	H.265 10-bit	VP9	VP9 10-bit	H.265	VVC	AV1
865	D/E	D/E	D/E	D/E	D/E	D/E	D/E	D/E	D/E	D/E	NO	NO

Figure 3: Enter Caption

2.2.8 Qualcomm Sensing Hub

Se basa en una IA que genera respuestas y acciones personalizadas dependiendo de varios factores del usuario, información que va recolectando el móvil como tu edad, salud, actividades favoritas, localización actual e información Wi-Fi o Bluetooth.

2.2.9 Qualcomm Processor Security

Es la capa hardware de seguridad, que protege el dispositivo en forma de multicapa. Qualcomm Mobile Security también incluye seguridad biométrica y detección de malware en tiempo real. Como nota aparte, Qualcomm también hace uso de otras partes del móvil para proveer seguridad adicional, como la cámara para reconocer el iris y la cara al verificar tu identidad durante un pago, el antivirus junto con la IA para tener protección en tiempo real y tokens hardware.

2.2.10 Qualcomm Kryo 585 CPU

Según Qualcomm, tiene un rendimiento y eficiencia energética un 25% superior al Kryo 485, el "hermano" de la pasada generación de CPUs de Qualcomm.

2.3 Arquitectura y Rendimiento

- **1x Kryo 585 Prime @ hasta 2.84 GHz:** Núcleo de alto rendimiento diseñado para tareas intensivas que requieren un alto rendimiento, con una frecuencia máxima de 2.84 GHz.
- **3x Kryo 585 Gold @ 2.42 GHz:** Tres núcleos de rendimiento medio con una frecuencia máxima de 2.42 GHz cada uno, ofreciendo un equilibrio entre rendimiento y eficiencia energética.
- **4x Kryo 585 Silver @ 1.80 GHz:** Cuatro núcleos de bajo consumo de energía con una frecuencia máxima de 1.80 GHz cada uno, destinados a tareas menos exigentes y contribuyendo a la eficiencia energética general del sistema.

2.4 Cachés

- **1x 512 KB pL2 cache para Prime:** Caché de nivel 2 (L2) de 512 KB para el núcleo Prime, utilizada para almacenar datos temporales cercanos al núcleo y mejorar la velocidad de acceso a la memoria.
- **3x 256 KB pL2 cache para Gold:** Cada núcleo Gold tiene una caché L2 de 256 KB.
- **4x 128 KB pL2 cache para Silver:** Cada núcleo Silver tiene una caché L2 de 128 KB, más pequeñas en comparación con las de los núcleos Prime y Gold, destinadas a tareas menos intensivas.
- **4 MB sL3 cache y 3 MB system level cache:** Caché de nivel 3 (L3) compartida de 4 MB para almacenar datos compartidos entre todos los núcleos y una caché de nivel de sistema (system level cache) de 3 MB.

2.5 Proceso de Fabricación

TSMC 2nd generation 7 nm (N7P) Process: Se refiere al proceso de fabricación utilizado para construir el chip. En este caso, se utiliza el proceso de 7 nanómetros de segunda generación (N7P) de TSMC (Taiwan Semiconductor Manufacturing Company), conocido por su eficiencia energética y densidad de transistores.

3 Coprocesador y Procesadores de Dominio

Hemos hablado ya de varios coprocesadores.

3.1 Qualcomm Hexagon 698 Processor

Se clasifica como un procesador de señales digitales (DSP), que generalmente se considera un tipo de coprocesador. Está diseñado para operaciones especializadas en procesamiento de señales digitales y tiene capacidades específicas para tareas como inteligencia artificial (IA). Puede realizar hasta 15 trillones de operaciones por segundo (TOPS).

3.2 Qualcomm Spectra 480 ISP

Sí, el Sensing Hub incorpora inteligencia artificial (IA) y realiza funciones específicas basadas en datos del usuario, como la edad, la salud, la ubicación y otras. Puede considerarse un coprocesador de inteligencia artificial.

3.3 Qualcomm Sensing Hub

El Sensing Hub incorpora inteligencia artificial (IA) y realiza funciones específicas basadas en datos del usuario, como la edad, la salud, la ubicación y otras. Puede considerarse un coprocesador de inteligencia artificial.

3.4 Qualcomm Processor Security

Aunque no está especificado detalladamente, la capa de seguridad del procesador se puede considerar un componente de seguridad específico que podría clasificarse como un coprocesador.