FutureHorizons
The Global Semiconductor Industry Analysts

FH MONDAY

TSMC to Build an Advanced Semiconductor Fab in US

TSMC today announced its intention to build and operate an advanced semiconductor fab in the US...
TSMC today announced its intention to build and operate an advanced semiconductor fab in the United States with the mutual understanding and commitment to support from the U.S. federal government and the State of Arizona.

read more

Nvidia Reinvents GPU For AI and Data Centers

Jensen Huang's much-anticipated keynote speech earlier, postponed from Nvidia's GPU Technology Conference (GTC) in March, will unveil the company's eighth-generation GPU architecture. Emerging three years after the debut of the previous generation Volta architecture, Ampere is said to be the biggest generational leap in the company's history.

read more

LeapMind Enters IP Business with AI Accelerator

LeapMind (Tokyo, Japan) announced its entry into the processor IP business with Efficiera, an ultra-low-power AI inference accelerator IP product.Efficiera is optimized for models that have been heavily quantized using LeapMind's 'extremely low-bit quantization' software techniques

read more

FutureHorizons

TALK TO US

CEA-Leti Explores mmWave Bands for 6G

France-based technology research institute CEA-Leti has demonstrated a 140GHz 100 Gbps transmission using a simple mixed signal RF architecture as it explores a technology roadmap to address 'beyond-5G' applications and 6G in the D-band spectrum. D-band covers frequencies from 110GHz to 170GHz.

read more

EVENTS

[Silicon Chip Industry Seminar](#)

-15 June 2020– London UK

[Industry Forecast Briefing](#)

– 15 Sept 2020 – London UK

DON'T MISS OUT.- BOOK NOW BY CALLING

+44 1732 740440

OR EMAIL

mail@futurehorizons.com

Entering the Era of Real-Time AI

Computing infrastructure is undergoing a major shift as a wave of real-time services grow to become part of our everyday life. From intelligent personal assistants providing instantaneous information using natural language to retailers generating information on customer shopping behaviour through in-store analytics, these real-time services present a huge market opportunity for service

read more

TSMC To Build An Advanced Semiconductor Fab In US

TSMC today announced its intention to build and operate an advanced semiconductor fab in the United States with the mutual understanding and commitment to support from the U.S. federal government and the State of Arizona.

This facility, which will be built in Arizona, will utilize TSMC's 5-nanometer technology for semiconductor wafer fabrication, have a 20,000 semiconductor wafer per month capacity, create over 1,600 high-tech professional jobs directly, and thousands of indirect jobs in the semiconductor ecosystem. Construction is planned to start in 2021 with production targeted to begin in 2024. TSMC's total spending on this project, including capital expenditure, will be approximately US$12 billion from 2021 to 2029.

Nvidia Reinvents GPU For AI and Data Centers

Jensen Huang's much-anticipated keynote speech earlier, postponed from Nvidia's GPU Technology Conference (GTC) in March, will unveil the company's eighth-generation GPU architecture. Emerging three years after the debut of the previous generation Volta architecture, Ampere is said to be the biggest generational leap in the company's history.

Ampere is built to accelerate both AI training and inference, as well as data analytics, scientific computing and cloud graphics.

The first chip built on Ampere, the A100, has some pretty impressive vital statistics. Powered by 54 billion transistors, it's the world's largest 7nm chip, according to Nvidia, delivering more than one Peta-operations per second. Nvidia claims the A100 has 20x the performance of the equivalent Volta device for both AI training (single precision, 32-bit floating point numbers) and AI inference (8-bit integer numbers). The same device used for high-performance scientific computing can beat Volta's performance by 2.5x (for double precision, 64-bit numbers).

LeapMind Enters IP Business With AI Accelerator

LeapMind (Tokyo, Japan) announced its entry into the processor IP business with Efficiera, an ultra-low-power AI inference accelerator IP product.

Efficiera is optimized for models that have been heavily quantized using LeapMind's 'extremely low-bit quantization' software techniques. It is designed for convolutional neural networks (CNNs), the type of network typically used for image processing and analysis tasks today.

"This is the company's first hardware IP product. But we are working on the development of a core technology called extreme quantization technology that operates at both software and hardware-IP levels with a network optimized for practical applications and a dedicated compiler," a LeapMind spokesperson told EE Times.

CEA-Leti Explores mmWave Bands For 6G

France-based technology research institute CEA-Leti has demonstrated a 140GHz 100 Gbps transmission using a simple mixed signal RF architecture as it explores a technology roadmap to address 'beyond-5G' applications and 6G in the D-band spectrum. D-band covers frequencies from 110GHz to 170GHz.

Wireless communication in millimeter wave (mmWave) bands, which range from 20 GHz to 300 GHz, is expected to be a key enabling technology for 6G wireless systems, because the huge available bandwidth can accommodate ultra-high data-rate communications. Within that range of mmWave bands CEA-Leti's research is investigating D-band, a spectrum at 140 GHz that may play a major role for 6G wireless communication.

Entering the Era of Real-Time AI

Computing infrastructure is undergoing a major shift as a wave of real-time services grow to become part of our everyday life. From intelligent personal assistants providing instantaneous information using natural language to retailers generating information on customer shopping behaviour through in-store analytics, these real-time services present a huge market opportunity for service providers.

To derive value from these services, data and insights need to be instantly accessible and will largely be driven through AI-enabled services. In response, cloud giants like Amazon Web Services (AWS), Microsoft, Alibaba and SK Telecom are developing the computing infrastructure to deliver those services.