MediaTek Advanced Research Center

Call for Research

(MARC-CFR)

Research Needs

April. 2023

MediaTek

Strategic Technology Exploration Platform (STEP)

Research Needs

1. 6G Communication Systems

Key technology explorations for 6G Communication System Design

- **Research Needs Label: [6GSys]**
- **Motivation**
  1) 6G cellular communications network is expected to support higher spectral efficiency, higher data throughput, lower end-to-end latency, lower power consumption, more robust waveform and more secure than 5G.
  2) Evolutionary and/or revolutionary techniques are required to mitigate current NR technology gap in terms of capacity, coverage and efficiency for eMBB/vertical use cases and new application drivers such as cloud gaming and XR in FR1, FR2 and even higher frequency spectrum [1].
  3) In order to achieve these objectives of 6G, several potential research areas of interest are listed below for reference.
- **Potential areas of interest but not limited to**
  1) New MIMO and multiplexing techniques
     i. Channel prediction for outdated CSI due to UE mobility
     ii. Design a communications system with BS/UE-controlled RIS (Reconfigurable Intelligent Surface) for developing RIS channel model used in simulations and evaluating system coverage and throughput improvement especially for the deployment in high frequency band
        - Proof of concept prototyping of above system to validate the system gain
     iii. Design a communication system with UE-side full duplex, including single-frequency full duplex and sub-band full duplex to enhance system latency and spectral efficiency with form factor, cost and power constraint.
        - Proof of concept prototyping of above system to validate the system gain
     iv. MIMO system design (incl. transmission schemes and CSI reporting, etc.) and its evaluation, taking into account the scalability of the dimension of antenna arrays, for both centralized and distributed deployment topology. For the latter case, also consider practical impairments such as timing/frequency synchronization error among the geographically separated arrays.
     v. Efficient beam management and CSI acquisition framework for (centralized

and distributed) massive MIMO for RS/feedback/latency overhead reduction.

vi. Radio propagation channel measurement and modeling for new frequency bands such as 7 to 15 GHz band and sub-THz band.

2) Fusion of communications, computing and sensing (including high-precision positioning)

i. Sensing assisted communication: identify typical use cases (especially UE sides or UE assisted use cases), evaluate feasibility, summarize challenges and potential solutions. For example, UE estimates positions of obstacles, and combine the position information of BS and UE, to assist beam management, cell reselection/handover operations.

- Proof of concept prototyping of above system to validate the system gain

ii. Sensing of device-to-device: design system architecture and signal processing procedure, summarize challenges and solutions. For example, modify sidelink system to support integrated sensing and communication (ISAC).

- Proof of concept prototyping of above system to validate the system gain

iii. UE cooperative sensing: multiple UEs use their positioning information and sensing results to cooperatively identify objects and estimate locations, and further use the information to assist some sensing applications, or to assist communications.

- Proof of concept prototyping of above system to validate the system gain

iv. AI assisted sensing: apply ML-based methods sensing algorithm, or enhance sensing performance with AI, for example, motion or gesture recognition.

- Proof of concept prototyping of above system to validate the system gain

3) Non-terrestrial network

I. Design the satellite/cellular spectrum sharing mechanism to improve overall spectral efficiency while managing the interference

II. Design new waveform applicable for 6G non-terrestrial networks which can well coexist with terrestrial network and optimized for satellite channels.

4) Artificial intelligence for communications

I. Design and evaluate solutions for AI-enhanced physical layer performance/ robustness or AI-enabled features to achieve system optimization, targeting

for 3GPP Rel-19 AI study item or beyond

II. Study new AI system architecture or framework integrated with 3GPP network for distributed intelligence using, for example, federated learning and intelligence plane for an AI application, to protect user's consent and privacy

5) Cross-layer optimizations for future applications
   I. Develop split algorithm between edge server/device or device/device for split rendering and SLAM to improve network capacity and E2E latency
   II. Develop cross-layer optimization algorithm for streaming adaptation to improve user experience in mobile network and standardize cross-layer API for B5G/6G
   III. XR applications proof of concept. To verify user experience enhancement for cross-layer optimization.
6) New network architecture for Integrated communication and computing (ICC)
   I. Explore architecture options to integrate cloud platform, e.g. kubernetes, with 3GPP service architecture
   II. ICC proof of concept to demonstrate UE triggered computing offload.
7) [Sub-]Tera-Hz communication and sensing system design
   I. Design and implement electronics-based RF beamforming transceiver circuit in sub-THz to evaluate system performance and feasibility of ultra-fast beam tracking in analog domain with fine resolution and blockage resistance
   II. Design and implement electronics-based RF prototype of sub-THz transceiver for indoor/outdoor communication and sensing
   III. Provide sub-THz device capability and impairment models which can be used in link-level and system-level computer simulations
8) Energy efficiency and harvesting
   I. Design ultra-low-power radio system based on radio technology of <= 1 uW active power consumption with at least -80 dBm receiver sensitivity [4]. The system should achieve tunable frequency, mitigation of interference from/to legacy co-channel users, multiple access functionality and robustness against time/ frequency uncertainty.
      - Develop proof of concept platform to verify performance
   II. Design of distributive optimization for system energy efficiency, where joint optimization is over the usage of the transmission power and processing power of all base-stations and UEs, to resolve the excessive complexity for a centralized optimization (computation burden, communications overhead). Given 6G network evolution includes distributed network and more

intermediate network nodes, the distributive optimization scheme should be applicable to networks featuring distributed MIMO and/or UE collaboration operations.

III. Design of radio-signal based energy harvesting technique based on either or both of signal transmitted from base-station(s) or peer UEs. The design targets to power a communication device with 1-10 mW average power consumption at a distance between 10-100 meters. For the case of energy harvesting based on peer-UE signal, design of distributive payment/rewarding should be included.

- Develop proof of concept platform to verify feasibility and performance

9) Next generation security system

I. Design information-theoretic/physical layer based mechanism and secret key distribution protocol, such as QKD (Quantum Key Distribution), to integrate with asymmetric security using novel PQC (Post Quantum Cryptography) algorithm to guarantee E2E security level even after Quantum computer is available.

II. Analyze and evaluate 3GPP standard and implementation impacts based on the lightweight cryptography algorithm to be defined by NIST.

10) 6G modem system architecture

I. Architecture exploration of 6G modem IP with the key directions identified, including processors, platform, and HW/FW/SW partitioning, pursuing leading position in performance, low-power and cost effectiveness as a product.

II. Methodology and tools for supporting evaluation, simulation and profiling of the IP architecture design and implementation.

■ Appendix : (please see page 26)

2. Wireless RF

■ Research Needs Label: [RF]

■ Motivation

New communication standards such as 5G beyond and WiFi7 increases throughput, reduces latency, while for commercialization, transceivers need to consume low power and have smaller form factors. To fulfill these demands, advancement of the following technologies is required.

First is the power amplifier which is usually the most power consuming circuitry in a transceiver. Both 5G and WiFi7 adopts OFDM whose signal peak-to-average-power-reduction ratio (PAPR) is typically more than 6dB. Therefore, maintaining high efficiency at both peak output power and >=6dB power backoff is desirable. Also, for 5G and beyond, new mmWave frequency bands are being opened up. Multi-band mmWave power amplifiers are needed to reduce phased-array module and system sizes.
Second is receiver architecture and components for multi-mode operations. Compared to the conventional architecture, a direct sampling RF receiver offers greater flexibility, easier for integration and occupies smaller area in advanced process nodes. By removing mixers and using a wide-bandwidth ADC to digitize RF waveforms directly, signals can be processed in the digital domain. ADC with wide bandwidth and high sampling rate is the essential component for such a receiver architecture. If signals of interested RF bands can be sampled and digitized in ADC's first Nyquist zone, complicated filtering, signal processing and frequency planning can be greatly simplified.

Finally, because a higher-order modulation is required, for example, from 1024QAM to 4096QAM for WiFi6 and WiFi7, respectively, multi-mode, wide tuning range, high resolution, and high-quality signal sources, such as crystal oscillator and voltage-controlled-oscillator (VCO) are necessary.
Low power consumption, wide bandwidth, high performance and small form factor are generally required for all circuits and systems.

■ Specific areas of interest

1) High efficiency sub-6GHz power amplifier simultaneously achieving the following targets:
   i. Switch-cap PA with >15% 3dB fractional bandwidth; center frequency

between 2-6GHz using 1.8V supply.

ii. QFDM-64QAM average power > 20dBm, average PAE > 25%, while passing FCC emission requirement. If necessary, develop pre-distortion tailored for this specific PA and apply.

2) Multi-band mmWave power amplifier simultaneously achieving the following targets:

I. > 30% fractional 3dB bandwidth, center frequency between 20-250GHz, using 1.2V or lower supply voltage.

II. OP1dB >20dBm (if between 20-50GHz), >15dBm (if between 50GHz-100GHz), or > 10dBm (if >100GHz). Linear (small-signal) gain higher than 20dB.

III. Antenna integration in module, in package, or on chip is optional but will be a significant plus.

3) Wide-bandwidth Nyquist rate ADC:

I. Class 1: Sampling rate >3GS/s, over-sampling ratio between 2-4, dynamic range >57dB, interleaved paths <=2.

II. Class 2: Signal bandwidth >6GHz (preferably >13GHz), SNR/SFDR 55-60dB, Nyquist sampling preferred but the second Nyquist zone is possible. Emphasis on power efficiency.

III. Clocking, and driving and reference buffers for the ADC need to be included.

IV. Specific interest in architectures that employ digital calibration/compensation e.g. AI/machine learning to improve performance in advanced process technologies and overcome bottlenecks in traditional architectures

4) Direct IF bandpass receiver:

i. Sampling rate >3GS/s; IF signal bandwidth > 400MHz; dynamic range > 57dB.

ii. The receiver needs to deal with anti-aliasing without using bandpass filter at its input, at least up to 5th harmonics of the sampling clock.

5) VCOs (5-80GHz), DCOs (5-80GHz) and Crystal oscillators (<150MHz) exploring the following:

I. Wide tuning range (continuous or banded operation), high-performance and low-power.

II. Phase noise suppression techniques for 10kHz~1MHz (preferably 5MHz) frequency offset away from the carrier frequency

III. Switched-cap array with >20000ppm tuning range, <0.05ppm resolution, and DNL < 0.5LSB with sufficient quality factor compared to those of other tank elements

IV. Low power techniques to trade power with performance while satisfying key communication system requirements at respective operating modes of interest

6) Frequency synthesis
   I. Power efficient frequency synthesizers with <40 fs integrated jitter
   II. Focus on mmW frequency generation 28-150GHz and/or at 5-7GHz
7) Novel architectures for mmW / Sub-THz applications employing low resolution ADC / DAC
8) Process technology supporting ft/fmax higher than those from CMOS may be considered to design high mmWave and sub-mmWave (sub-THz) frequency bands for low power, high PA Pout and efficiency, and small form factors. Availability (commercialization viability) should be considered.
9) Ultra-broadband RF Phase shifter for Future RF beamformers:
   Main part of this project is the design of a broadband quadrature signal generator and then using vector sum to generate the phase shift with the specific accuracy and resolution. In recent years, there has been a come back to lumped couplers for mmWave frequencies but more ideas/inventions are needed to make them compact and broadband. Spec[negotiable]:
   I. Vector modulator based on phase shifter
   II. Compact (less than 2 differential inductor area consumption or similar)
   III. Ultra broadband covering n257:n262 [24.2GHz 48.2GHz]
   IV. Gain variation< 0.5dB across the band
   V. Phase accuracy >2 degrees
   VI. Phase resolution 5 bit
   VII. Loss < 3dB
10) Ultra-broadband, compact True Time Delays Future RF beamformers:
   This project is the design of a compact time delay with the specific accuracy and resolution in the band of interest. There have been attempts to use switched transmission lines for such delays but there is a limitation such as huge area consumption (inversely proportional to the frequency) and lossy switching systems.
   I. Spec[negotiable]:
      - Compact: this will be used within a transceiver beamformer IC, hence it should be as compact as possible
      - Ultra broadband covering n257:n262 [24.2GHz 48.2GHz]
      - Gain variation< 0.5dB across the band
      - Time accuracy > 2 degrees equivalent

i. Time resolution 5 bit

ii. Loss < 4dB

■ Special information:

if specific process is required to achieve required circuit performance, the research proposal needs to explicitly request and provide sufficient justifications. Access to such process can be discussed with corresponding MediaTek owners.

■ Reference : (mmWave power amplifiers)

1) https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/samsung-5g-fwa/white-paper_samsung-5g-fixed-wireless-access.pdf

2) https://www.techplayon.com/5g-nr-ue-power-classes/

3) S. N. Ali, et al. “A 25–35 GHz Neutralized Continuous Class-F CMOS Power Amplifier for 5G Mobile Communications Achieving 26% Modulation PAE at 1.5 Gb/s and 46.4% Peak PAE,” IEEE Trans, Circuits Syst. I, Reg. Papers, vol. 66, no. 2, pp. 834-847, Feb. 2019.

3. Analog Circuits

- **Research Needs Label: [Analog]**
- **Motivation**
  1) High performance, high bandwidth, power efficient analog circuit continue to play important role for wireless, wireline communications, smart home and AIoT applications. The key areas include power management, data converters and high speed Serdes. The focus includes innovations in architectures, circuits, and systems.
  2) Power management: Explore integrated circuits and/or application circuits that could improve power conversion efficiency for application processors, RF power amplifiers, mobile devices, IoT and wearable applications
  3) Data converter: Analog-to-digital converters and Digital-to-analog converters are fundamental and enabling building blocks for a wide range of applications from meter, audio to communications and beyond. The techniques to improve resolution, dynamic range, sampling rate, and energy efficiency (FoM) are highly demanded.
  4) High speed interface (e.g. serdes): Techniques to support high data rate, power efficient data links and also high density I/O system over advanced 2.5D/3D package are of interest.
- **Specific areas of interest**
  1) High speed interface Serdes: Power and area efficient circuits including but not limited to AGCs, equalizers, high-bandwidth amplifiers, analog and ADC/DAC-based front ends, TX drivers and low jitter clocking, clock recovery, etc. with state of the art performance (upon normalization over e.g. process technology if needed). Optical communication circuits and systems are also of interest.
  2) 2.5/3D (INFO/CoWoS) interconnect with data rate 32+ Gb/s/wire. Power efficiency <=0.3 pJ/bit with N4 process. (Could have normalization over e.g., process technology if needed)
  3) Chip-to-chip single-ended communications on substrate, with data rate 16+ Gb/s/wire. Innovative architecture to achieve best power efficiency is highly interested.
  4) High dynamic range, low power data converters and analog front end for audio and sensor applications. (>120dB, preferably >140dB)
  5) High sampling rate, power efficient data converters for WiFi, 5G and base station applications. (>=10bits, >1Gs/s/channel) [1]

6) Time-interleaved analog-to-digital converter calibration techniques for sampling rate >10Gs/s. (>=10bits)
7) IVR (Integrated Voltage Regulator) for SoC
   I. May include hybrid SC, LDO, etc. Vin=1.2V, Vout=0.3V ~ 1V, Iout > 2A [2]
8) XPU Power Delivery
   I. Multi-phase fast transient (>2A/0.1us) area efficient Buck converter with >90% efficiency @4-to-0.8V, Iout_max > 10A
   II. Multi-phase fast transient (>2A/0.1us) area efficient Buck converter with >90% efficiency @1.8-to-0.8V, Iout_max > 10A, and inductor <10nH
   III. ZCS/ZVS or resonant
9) RF PA Power Delivery / Modulator
   i. >100MHz (200MHz is preferred) ETM with efficiency >90% and low noise (e.g. spur noise -49dBm/MHz)[3]
10) Ultra-low voltage, low power analog circuits for bandgap, temperature sensor, oscillators and clocking with high stability, etc. (<=0.5V, nW)
11) Circuits and systems for analog AI, CIM, etc. that support AI computing acceleration and non-conventional computing.

■ Reference for Analog Circuit Research Needs: (please see page 27)

4. AI Systems and Hardware

■ Research Needs Label: [AI]

■ Motivation

We are seeking research proposals to address the challenges posed by AI applications, systems, and hardware. Advances in algorithms, big data, and computing capability have allowed AI to make significant strides in improving user experience while also enabling many new applications. Existing popular applications such as social media, video, and games continue to push these technologies forward, while the recent boom in AI-generated content (AIGC) has successfully taken AI applications to another level, allowing people to use AI to work more efficiently. However, the complexity of System-on-Chip (SoC) designs is increasing at a faster rate than Moore's Law, and devices are often resource-constrained, particularly in terms of memory bandwidth and thermal budget.

We invite proposals covering a variety of topics, from the application level (e.g. AIGC on Edge devices), to the machine-learning core (e.g. system AI), to the hardware level (e.g. RISC-V architecture design). Cross-area research and creative, original ideas are especially encouraged. We value contributions with high innovation value and originality that bring us closer to solving the challenges posed by AI systems and hardware.

■ Specific areas of interest

Application

1) AIGC on Edge devices (stable diffusion, multi-modality model, prompt, fine-tuning)
2) Autonomous driving, ADAS, PAS
3) Edge AI platform technology
4) Embedded vision and computational photography
5) Analog AI and CIM algorithms (TinyML)
6) Graph neural networks (GNN) applications

Machine-Learning Core

1) System AI, AI for systems (Power/DVFS/Memory)
2) Android system optimization on RISC-V
3) AI Benchmark

4) AutoML and platform-aware optimization
5) Lower-bit quantization for edge intelligence
6) AI compiler optimization (TVM/MLIR)

Hardware

1) RISC-V architecture design
2) Compute architecture
3) Design automation
4) Benchmark, analysis, profiling, modeling of hardware

- Reference for AI Systems and Hardware Research Needs: (please see page 28)

5. Multimedia

■ Research Needs Label: [MM]

■ Motivation

1) Image Processing and Computer Vision (CV) have been utilized to facilitate human's daily life through various applications, such as consumer electronics, advanced driver-assistance system (ADAS), surveillance cameras. In addition, edge devices (e.g., mobile devices, TV, tablets, etc.) become more and more popular, the needs of applying CV techniques to edge devices are getting higher. In recent years, Artificial Intelligence (AI) has shown significant progress, demonstrating improvement over traditional CV techniques. However, there are several drawbacks of current AI approaches. For example, AI methods usually have much higher computation and memory cost, increasing the difficulty for product deployment, which makes the methods impractical. Moreover, AI methods are data-driven and usually require large-scale data for training, leading to great difficulty for data preparation, including data collection and annotation. Therefore, developing efficient AI approaches that can practically benefit product deployment with the consideration of performance-efficiency balance is necessary.
2) We are soliciting research proposals for our needs as mentioned above. The purpose of the research is to develop practical AI methods that can benefit our life via various applications, such as smart phone cameras, ADAS, surveillance systems, edge devices, etc. The topics can be (but not limited to) application-, algorithm-, or methodology-oriented (please see the key words below). We also strongly encourage to propose research on under-explored areas with high **innovation values.**

■ Specific areas of interest

1) Real-world RAW image restoration
   i. Image restoration (denoising, demosaic, remosaic, super-resolution, deblur, etc.)
   ii. Real-world RAW images from smartphone or surveillance cameras
   iii. Unsupervised or semi-supervised approach is preferred
   iv. Joint training or multi-task is encouraged

2) Real-world RAW video restoration or enhancement

i. Video restoration (denoising, demosaic, remosaic, super-resolution, deblur, etc.)
ii. Video stabilization (using gyro sensor or video)
iii. Video frame interpolation
iv. With complexity/power consumption considerations
v. Joint training or multi-task is encouraged

3) Vision applications and scene/intention analysis for surveillance and ADAS system
   i. Joint training of visual perception systems (Detection, segmentation, …)
   ii. Scene/intention analysis for surveillance and ADAS system
   iii. Domain adaptation approaches. (Unsupervised or semi-supervised domain adaptation is preferred)
   iv. A simulator or a real platform for validating the proposed ADAS approach

4) Visual attention and transformers for low level image processing and visual recognition
   i. Practical vision applications with visual attention or transformers
   ii. Feasible complexity for edge devices
   iii. Domain adaptation consideration
   iv. Self-/semi-supervised learning is encouraged

5) AI video compression
   i. AI loop filtering [1][2][3][4]
   ii. AI intra prediction [5][6][7]
   iii. AI super resolution [8][9][10]
   iv. Other AI video coding tool(s) [11][12]
   v. End-to-end AI video coding [13][14][15]

6) Extended Reality (XR)
   i. Simultaneous localization and mapping (SLAM)
   ii. Object/scene 3D reconstruction
   iii. Natural user interface

■ Reference for Multimedia Research Needs: (please see page 29)

6. Modern GPU

■ Research Needs Label: [GPU]

■ Motivation

1) Nowadays as ecosystem grows, plenty of new applications drive GPUs to new limits in different ways. Mediatek is seeking research proposals to optimize those state-of-the-art rendering techniques. As first step, bottleneck analysis is fundamental for developing better architecture, algorithm, and user experience. Mediatek urges for methods or cross comparison results of modem GPU architecture on new graphics/computing technologies.

2) As examples, Mediatek is interested in the following rising topics. Raytracing technology is become popular on mobile, and Mediatek has the hardware support to speed up the calculation. However, the bottleneck of performance and power consumption are going to the requirement of memory bandwidth. There are some opportunities to improve them at different system levels: algorithms, data compression of raytracing geometry and cache policy. On the other hand, complexity reduction is another solution for the hungriness of user experience. Frame interpolation can greatly reduce the power consumption which is crucial for mobile devices. However, to achieve good quality for complex game situation is still quite challenging. Topics like occlusion handling, game engine integration and super resolution are often mentioned.

■ Specific areas of interest

Modem GPU architecture bottleneck analysis

Interest

1) Defining standard flows to efficiently analyze GPU architecture bottlenecks
2) Finding specific GPU architecture bottlenecks on modern scenario
   I. Using new technologies like: Ray tracing, DLSS, GPU-computing, mesh shader, tile shader, VRS, HDR, global illumination, AR/VR/XR, ..., etc.
3) Comparing across different GPU architectures about their bottlenecks
4) Developing GPU architecture bottleneck analysis tools
5) Developing GPU benchmarks to compare bottlenecks cross different GPU architectures
6) Developing GPU benchmarks to analyze GPU architecture bottlenecks with new rendering technologies

More information

1) It is welcome for further academic research on better solutions after the GPU architecture bottleneck analysis
   I. GPU architecture bottleneck analysis is the basic requirement and the starting point of further research
2) It would be better using mobile or automotive GPUs. But modern desktop GPU is acceptable
   I. Not limited to MTK platforms, but MTK can provide some tool chain support on MTK platforms
3) Comparing architecture bottleneck with GPU to other system (e.g., AI engine, ASIC, ... or even a brand-new GPU architecture) is fine
4) Not limited to real platform. Theoretical analysis, analytical estimation, or modeling are all acceptable

Raytracing on mobile

Interest：

1) The raytraced-based importance sampling / guiding method for faster lighting converge or denoising friendly.
2) Data compression of raytracing data at different level, including AS layout, AS depth reduction by geometry representation change.
3) Bandwidth reduction by smart caching policy or design.
4) Evolution of ray traversal and acceleration structure to handle game scene and animation smartly.

Some opportunities：

1) Graphics Algorithms:
   I. The developer can reduce the ray jobs by the importance sampling for the advance effects, such as indirect lighting and color bleeding and etc …, this may also reduce the cost of denoiser (such as the size of filters and number of filtering).
   II. The better initialization of ray jobs with a cache or guiding algorithms, such as ReSTIR, path guiding, radiance caching and so on, those methods may keep coherence between frames and increase the quality of 1st iteration.
2) Data compression of raytracing geometry:
   I. Raytracing algorithms are reported as the bandwidth bound problem. How to reduce the bandwidth for different data:
   II. Data size of geometry in Acceleration structure, including the depth of AS, or the size of geometry data at the leaf nodes. (triangles data layout

optimization)

III. Data in AS with non-lossless compression.

3) Cache policy for each memory hierarchy to reduce bandwidth:

I. Memory footprint is a big problem for raytracing job, we may need a new cache policy for each level of memory cache system.

4) Evolution of ray traversal and acceleration structure to handle game scene and animation smartly.

II. Avoid unnecessary rebuild and update of AS structure

III. Skin and skeleton support

IV. Level of detail support

Game Frame Interpolation using GPU

Interest:

1) Occlusion handling for complex scene
2) Game integration
3) Super resolution integration

Some opportunities:

1) Occlusion detection and handling

I. This is the most common challenging, the edge of the object covers background or is covered by another object. For games, semi-transparent objects are widely used for special effects, such as damage text, HP bar, and NPC icon. A robust method is needed to handle them.

2) Cooperation with in-game information

I. Unlike static video, game can provide additional information such as depth, opacity, object label, or even in-game motion.

3) Real-time segmentation or object tracking

I. Fast and small moving objects tend to disappear or get ignored by frame interpolation. Tracking these objects may solve this kind of problem.

4) Super Resolution Integration

I. Both frame interpolation and super resolution can reduce the power for mobile devices. It is possible to integrate them into an advance system.

7. 3D-IC Chip and Package

Novel Material, Architecture, Interconnection, Reliability, and Thermal Management for 3D-IC & Chiplet Package Applications

- **Research Needs Label: [3DIC] [Chiplet]**
- **Motivation**

As the size of chips continues to shrink, modern integrated circuit (IC) design is facing many challenges. One of the challenges is how to effectively integrate multiple chips in terms of power, performance, area, cost, and reliability (PPACR). 3D-IC and chiplet technologies are two promising solutions, but there are also some unique challenges, especially in packaging.

3D IC technology is a technique that stacks multiple chips in three-dimensional space. Each chip can contain different functions, such as processors, memory, and sensors. Since the chips are very close to each other, communication speed and energy efficiency can be greatly improved. However, stacking multiple chips together also brings some challenges. Here are some challenges that may be encountered in 3D IC packaging:

1) Heat dissipation issues: When multiple chips are stacked together, the heat they generate will also accumulate. This may cause excessive heat buildup, resulting in system crashes or performance degradation. To address this issue, more efficient heat dissipation solutions (such cooling strategies) and novel thermal interface material (TIM) need to be developed.
2) Power supply issues: When multiple chips are stacked together, they require higher power supply. This may result in unstable power supply, leading to system performance degradation. To address this issue, more efficient power management technology and power supply solutions need to be developed.
3) Signal interference issues: When multiple chips are close to each other, signal interference issues may arise. This may cause signal distortion or system crashes. To address this issue, more effective signal paths and signal shielding technology need to be developed.

On the other hand, chiplet technology combines individual chips together to achieve a complete system. Each chip can contain different functions, such as processors,

memory, and sensors. The chips can be connected through high-speed interfaces, such as advanced substrate, silicon interposers, or through-silicon vias (TSVs), etc.. Here are some challenges that may be encountered in chiplet packaging:

1) Heterogeneous integration issues: Chiplets may be produced by different manufacturers using different technologies and materials. This may lead to heterogeneous integration issues, such as thermal expansion mismatch and different mechanical properties. To address this issue, more effective bonding and interconnect technologies need to be developed.
2) Interconnect density issues: Since chiplets are smaller in size than traditional chips, they may require higher interconnect density. This may lead to interconnect density issues, such as signal crosstalk and power supply noise. To address this issue, more efficient interconnect design and signal shielding technology need to be developed.
3) Test and debug issues: Since chiplets are produced separately and then combined, testing and debugging may be more challenging. To address this issue, more effective test and debug technologies need to be developed to ensure the reliability and quality of the final product.

■ Specific areas of interest

1) Effective thermal management, innovative cooling strategy, optimal thermal design
2) Novel anisotropic thermal interface material (TIM)
3) High thermal conductivity molding compound
4) Backside power via for PDN layout application
5) Hybrid OX bonding scheme development for bonding interface strength and thermal performance optimization.
6) The thermal-mechanical stress evaluation of 3D-IC stacking chip/monolithic SoC in advancing packaging
7) Die-to-Die interconnect design
8) Innovative decoupling capacitor solutions in Packaging to meet ultra-high di/dt request

8. Special Topics

8.1 Generative Artificial Intelligence (GAI)

- **Research Needs Label: [GAI]**
- **Motivation**

Generative Artificial Intelligence (GAI) has shown significant progress in recent years, due to the emergence of massive pre-trained models such as DALL-E, Midjourney, and ChatGPT, which can unlock potential for a wide range of applications. These pre-trained models have already shown promising results in producing fluent text, producing images, and performing few-shot learning. As leading company in IC design, we foresee GAI as future tool for accelerating electronic design workflow. These GAI services can involve in hardware code writing, netlist graph generation, test cases generation for verification and more tasks in IC design flows. Although these models offer huge potential, they have proven difficult to train, control and comprehend, giving rise to scalability, grounding and interpretation challenges. Therefore, we are soliciting research proposals for addressing the followings areas:

- **Fundamentals**

Large model training system and theory

Application
Generative AI in IC design

- **Specific areas of interest**

1) Self-supervised Learning methodology (unimodal, multi-modal) (theory/application)
   I. Unimodal methodology:
   II. Text: for example, GPT-like models for generating hardware code
   III. Vision: for example, Vision transformer pretraining for thermal simulation
   IV. Graph: generative models in large graphs, especially for IC circuit graph generation
   V. Multi-modal methodology:
   VI. Text + Graph: Graph-Text Multi-Modal Pre-training [1]
2) Efficient large model training technologies

i. 3D parallelism (data parallel, tensor parallel, pipeline parallel) [2, 3]
ii. Compiler technology for deep learning [4]
iii. Communication volume reduction for distributed training [5,6]

3) Data efficient learning methods
   I. Active learning/ Curriculum learning methodology for generative model training [7]
4) Language models for data generation and augmentation [8]
   I. Goal: generating more and higher quality data
5) Grounded language model
   I. Goal: making language model ground on facts, physical law and symbolic operations
   II. Research in tool augmented neuro system [9] , neurosymbolic system [10]
   III. interplay betwen LLM and knowlege graphs [11]
   IV. Reinforcement Learning for aligning human intent [12]
      - Goal: making models following human intent/ instructions

Reference for Generative AI System Research Needs: (please see page 31)

8.2 AI for IC Design

- **Research Needs Label: [EDA]**
- **Motivation**

People continue to discover how to apply AI and ML to IC design. This leads to higher productivity and higher product quality. Certain areas of IC design, of high interest to MediaTek, are under-served by commercial tool vendors. Physical design plays a crucial role in various aspects of integrated circuit (IC) design. However, current approaches still heavily rely on manual tuning, and institutions and companies are investing more resources in solutions and academic articles to address this challenge. Nevertheless, there are still some missing pieces that need to be included in the reality IC design flow, such as design rule handling and data transmission timing minimization. Therefore, there is a need for efficient continuous and combinatorial optimization methodologies to handle the increasingly extreme design complexity and design rules. MediaTek seeks to develop in-house capability to address them.

- **Potential areas of interest (but not limited to)**

1) Eliminate redundant sign-off concerns.
2) Multi-factor design closure techniques.
3) multi-objective constrained optimization with low optimization budget (model-based optimization and Bayesian optimization are welcome)
4) multi-objective constrained combinatorial optimization
5) distributional searching, and critical indices approximation.

8.3 Health Monitoring Wearable

- **Research Needs Label: [HMW]**
- **Motivation**

Wearable devices are prosperously developed in recent year due to the advance of the VLSI technology. As the increase in ageing population poses a challenge to the medical and social care systems globally, wearable devices enabling long-term health monitoring provide an alternative solution for physiological assessment. Because cardiovascular diseases are the major cause of mortality, non-invasive sensing of cardiovascular signals, such as heart rates, blood oxygen and blood pressures, becomes a trend in biomedical consumer products. Also, wearable enable portable, continuous monitoring of sleep at home, overcoming the limitations of PSG in hospital. We seek for the topics of interest listed below:

- **Specific areas of interest**

1) Cardiovascular risk assessment, vascular stiffness monitoring, continuous blood pressure estimation, etc.
2) Circadian rhythm monitoring, sleep quality assessment, obstructed sleep apnea detection, etc.

Appendix: 6G Communication Systems

■ Reference:

1) 6G White Paper - MediaTek's vision for the next-generation of cellular mobile technologies: https://www.mediatek.com/blog/6g-whitepaper
2) 3GPP SA1 TR 22.837, “Study on Integrated Sensing and Communication”
3) IMT-2030 研究报告：通信感知一体化技术报告 （第二版）
4) D. D. Wentzloff, A. Alghaihab and J. Im, "Ultra-Low Power Receivers for IoT Applications: A Review," 2020 IEEE Custom Integrated Circuits Conference (CICC), Boston, MA, USA, 2020, pp. 1-8.

Appendix: Analog Circuits

■ Ref [1]

DAC		ADC		PLL	
Parameter	Specification	Parameter	Specification	Parameter	Specification
Resolution	14 bits	Resolution	12 bits	Ref. frequency	491.52MHz
Clock rate	16GHz	Clock rate	16GHz	o/p frequency	3.93 ~ 15.72GHz
o/p impedance	100ohm(diff.)	i/p impedance	100ohm(diff.)	R.M.S. jitter	100fs(10k~100M)
o/p bandwidth	8GHz	i/p bandwidth	8GHz		
o/p power	2dBm	i/p swing	$1.2V_{dpp}$		
IM3	-62dBc@7GHz	IM3	-62dBc@7GHz		
NSD	-156dBm/Hz	NSD	-153dBFS/Hz		

■ Ref[2]

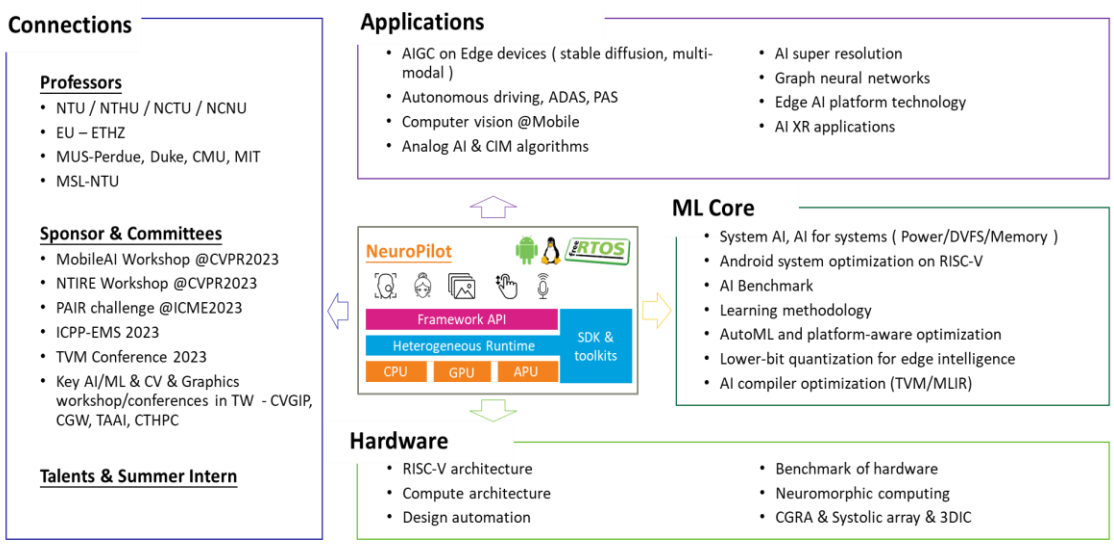
S. T. Kim, et al., “Enabling wide autonomous DVFS in a 22nm graphics execution core using a digitally controlled hybrid LDO/switched-capacitor VR with fast droop mitigation,” ISSCC, pp. 154-155, 2015

■ Ref[3]

J. -S. Paek et al., "A − 137 dBm/Hz Noise, 82% Efficiency AC-Coupled Hybrid Supply Modulator With Integrated Buck-Boost Converter," in IEEE Journal of Solid-State Circuits, vol. 51, no. 11, pp. 2757-2768, Nov. 2016, doi: 10.1109/JSSC.2016.2604296

Appendix: AI Systems and Hardware

■ NeuroPilot – Mediatek Edge AI Platform



Appendix: Multimedia

■ Reference:

1) Y.-H. Lam, A. Zare, F. Cricri, J. Lainema, and M. M. Hannuksela. 2020. Efficient Adaptation of Neural Network Filter for Video Compression. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20). Association for Computing Machinery, New York, NY, USA, 358–366. DOI: https://doi.org/10.1145/3394171.3413536
2) Y.-H. Lam, M. Santamaria, J. Lainema, F. Cricri, R. Ghaznavi-Youvalari, A. Zare, H. Zhang, H. R. Tavakoli, and M. Hannuksela. AHG11: Content-adaptive neural network post-processing filter. Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 Document JVET-V0075. April 2021.
3) H. Wang, J. Chen, K. Reuze, A. M. Kotra, and M. Karczewicz. EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer. Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 Document JVET-W0130. July 2021.
4) Y. Li, K. Zhang, and L. Zhang. AHG11: Deep In-Loop Filter with Adaptive Model Selection and External Attention. Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 Document JVET-W0100. July 2021.
5) M. Meyer, J. Wiesner, and C. Rohlfing, “Optimized convolutional neural networks for video intra prediction,” in Proc. of IEEE International Conference on Image Processing ICIP '20, IEEE, Piscataway, Oct. 2020
6) M. Meyer, J. Wiesner, J. Schneider, and C. Rohlfing, “Convolutional neural networks for video intra prediction using cross-component adaptation,” in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP '19, pp. 1607–1611, IEEE, Piscataway, May 2019
7) Y. Hu, W. Yang, M. Li, and J. Liu, “Progressive spatial recurrent neural network for intra prediction,” Computing Research Repository (CoRR), 2018
8) B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 1132-1140, doi: 10.1109/CVPRW.2017.151.
9) C. Lin, L. Zhang, K. Zhang, and Y. Li. AHG11: CNN-based Super Resolution for Video Coding Using Decoded Information. Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 Document JVET-W0099. July 2021.
10) X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” in Proceedings of the

European Conference on Computer Vision (ECCV) workshops, 2018.

(Appendix: Multimedia, cont'd)

11) F. Galpin, P. Bordes, T. Dumas, A. Robert, P. Nikitin, and F. Le Leannec. AHG11: Deep-learning based inter prediction blending. Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 Document JVET-V0076. April 2021.
12) Huo, D. Liu, F. Wu and H. Li, "Convolutional neural network-based motion compensation refinement for video coding", Proc. IEEE ISCAS, pp. 1-4, May 2018.
13) D. Minnen, J. Ballé, and G. Toderici, 'Joint Autoregressive and Hierarchical Priors for Learned Image Compression', arXiv:1809.02736.
14) Yoojin Choi, Mostafa El-Khamy, Jungwon Lee, 'Variable Rate Deep Image CompressionWith a Conditional Autoencoder', Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3146-3154
15) Fei Yang, Luis Herranz, Joost van de Weijer, José A. Iglesias Guitián, Antonio López, Mikhail Mozerov, "Variable Rate Deep Image Compression With Modulated Autoencoder", arXiv: 1912.05526.
16) Horgan et al, "Vision-based Driver Assistance Systems: Survey, Taxonomy, and Advances, " in 2015 IEEE 18th international conference on Intelligent Transportation Systems
17) Yurtsever et al, "A survey of Autonomous Driving: Common practices and emerging technologies, " in IEEE Access, Mar., 2020
18) Xu et al, "Dynamic video segmentation network, " in the IEEE conference on computer vision and pattern recognition 2018
19) Hong et al, "Virtual-to-real: Learning to control in visual semantic segmentation, " in International Joint Conferences on Artificial Intelligence (IJCAI) 2018

Appendix: Special Topics -

Generative Artificial Intelligence (GAI)

1) [1]: Park, S., Bae, S., Kim, J., Kim, T., & Choi, E. (2022, April). Graph-Text Multi-Modal Pre-training for Medical Representation Learning. In _Conference on Health, Inference, and Learning_ (pp. 261-281). PMLR.
2) [2]: DeepSpeed: Extreme-scale model training for everyone - Microsoft Research(https://www.microsoft.com/en-us/research/blog/deepspeed-extreme-scale-model-training-for-everyone/)
3) [3]: Li, S., Fang, J., Bian, Z., Liu, H., Liu, Y., Huang, H., ... & You, Y. (2021). Colossal-AI: A unified deep learning system for large-scale parallel training. arXiv preprint arXiv:2110.14883.
4) [4]: Dao, T., Fu, D. Y., Ermon, S., Rudra, A., & Ré, C. (2022). Flashattention: Fast and memory-efficient exact attention with io-awareness. _arXiv preprint arXiv:2205.14135_.
5) [5]: Ryabinin, M., Dettmers, T., Diskin, M., & Borzunov, A. (2023). SWARM Parallelism: Training Large Models Can Be Surprisingly Communication-Efficient. _arXiv preprint arXiv:2301.11913_.
6) [6]: Gan, S., Lian, X., Wang, R., Chang, J., Liu, C., Shi, H., ... & Zhang, C. (2021). Bagua: scaling up distributed learning with system relaxations. _arXiv preprint arXiv:2107.01499_.
7) [7]: Li, C., Yao, Z., Wu, X., Zhang, M., & He, Y. (2022). DeepSpeed Data Efficiency: Improving Deep Learning Model Quality and Training Efficiency via Efficient Data Sampling and Routing. _arXiv preprint arXiv:2212.03597_.
8) [8]: Wang, Y., Kordi, Y., Mishra, S., Liu, A., Smith, N. A., Khashabi, D., & Hajishirzi, H. (2022). Self-Instruct: Aligning Language Model with Self Generated Instructions. _arXiv preprint arXiv:2212.10560_.
9) [9]: Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., ... & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. _arXiv preprint arXiv:2302.04761_.
10) [10]: Karpas, E., Abend, O., Belinkov, Y., Lenz, B., Lieber, O., Ratner, N., ... & Tenenholtz, M. (2022). MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning. _arXiv preprint arXiv:2205.00445_.
11) [11]: West, P., Bhagavatula, C., Hessel, J., Hwang, J. D., Jiang, L., Bras, R. L., ... &

Choi, Y. (2021). Symbolic knowledge distillation: from general language models to commonsense models. _arXiv preprint arXiv:2110.07178_

12) [12]: Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., ... & Irving, G. (2019). Fine-tuning language models from human preferences. _arXiv preprint arXiv:1909.08593_.