IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 21, 2017, accepted August 15, 2017, date of publication August 22, 2017, date of current version September 6, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2741671

Energy-Efficient Stabilized Automatic Control for Multicore Baseband in Millimeter-Wave Systems

JOONGHEON KIM[1], (Member, IEEE), JAE-JIN LEE[2], JONG-KOOK KIM[3], (Senior Member, IEEE), AND WOOJOO LEE[4]

[1]School of Computer Science and Engineering, Chung-Ang University, Seoul 156-756, South Korea
[2]Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea
[3]School of Electrical Engineering, Korea University, Seoul 136-701, South Korea
[4]Department of Electronic Engineering, Myongji University, Yongin 449-728, South Korea

Corresponding authors: Jong-Kook Kim (e-mail: jongkook@korea.ac.kr) and Woojoo Lee (e-mail: woojoole@usc.edu)

This work was supported in part by the National Research Foundation of Korea under Grant 2017R1A4A1015675 and in part by ETRI grant (NZV μ-grain Architecture for Ultra-Low Energy Processor) through the Korean Government under Grant 17ZB1600.

ABSTRACT The fifth generation (5G) cellular network is upon us. Academia and Industry have intensively collaborated together to bring the power of 5G cellular networks to the masses, and now the 5G millimeter-wave (mmWave) platforms come into being in the market. One of the most popular 5GmmWave platforms mounts the massive mmWave phased antenna arrays in order to transfer a huge number of bits in a second (e.g., more than ten gigabits-per-second) to the baseband in the platform. While exploiting chip multicore processors (CMPs) may be the best solution to process such huge data in the mmWave baseband platform, power dissipate by the CMPs should become critical. Starting from an intuition that utilizing all processors in every single time introduces inefficient energy consumption, this paper proposes an energy-aware queue-stable control (EQC) algorithm to control the activation/deactivation of individual processors and antenna arrays for pursuing time average energy consumption minimization subject to the stability of queues in the 5G-mmWave baseband. Results from intensive simulations based on realistic experimental setups demonstrate the efficacy of the proposed EQC that achieves significant energy savings while queue stability is maintained.

INDEX TERMS Millimeter-wave, multicore baseband, chip multicore processor (CMP), dynamic control.

I. INTRODUCTION

5G denotes a new era in which connectivity will become amazingly responsive, extraordinarily dense, blazingly fast, highly reponsive and very low power. Academia and Industry have intensively collaborated together on new demonstrations and lab tests designed to bring the power of 5G cellular networks to the masses. As a result, the 5G millimeter-wave (mmWave) platforms now come into being in the market. Samsung and SK have announced that they successfully tested handover between 5G base stations at 28GHz in the outdoor environment [1], and Intel and Ericsson have demonstrated that their 5G radio prototype system works fairly well with autonomous driving and virtual reality technology use cases [2].

To support more than ten gigabits-per second (Gbps) transmission, most 5G base stations exploit the massive mmWave phased antenna arrays which is a kind of massive multi-input-multi-output (MIMO) systems. Samsung's and Intel's 5G base stations are the representative examples of such mmWave systems. Owing to the mmWave phase antenna, the existing base stations prove that ultra-high data rate can be achieved. However, another innovation that the 5G network should ultimately achieve, energy efficiency [3], has not been brought yet. For example, the mmWave baseband employs chip multicore processors (CMPs) to process huge data, little attention has been paid to the power consumption of the CMPs, which may be critical, and the question of how to minimize it.

This paper starts from the intuition of adopting the power-gating (PG) technique and core-consolidation technique to the CMPs in the mmWave baseband. Instead of utilizing all processors in every single time, which introduces inefficient energy consumption, using the minimum number of cores and turning off the unnecessary cores should result in power saving while processing performance is maintained. However, applying the well-known PG and core-consolidation techniques to the mmWAVE baseband is not trivial. That is because the queue stability should be taken into consideration

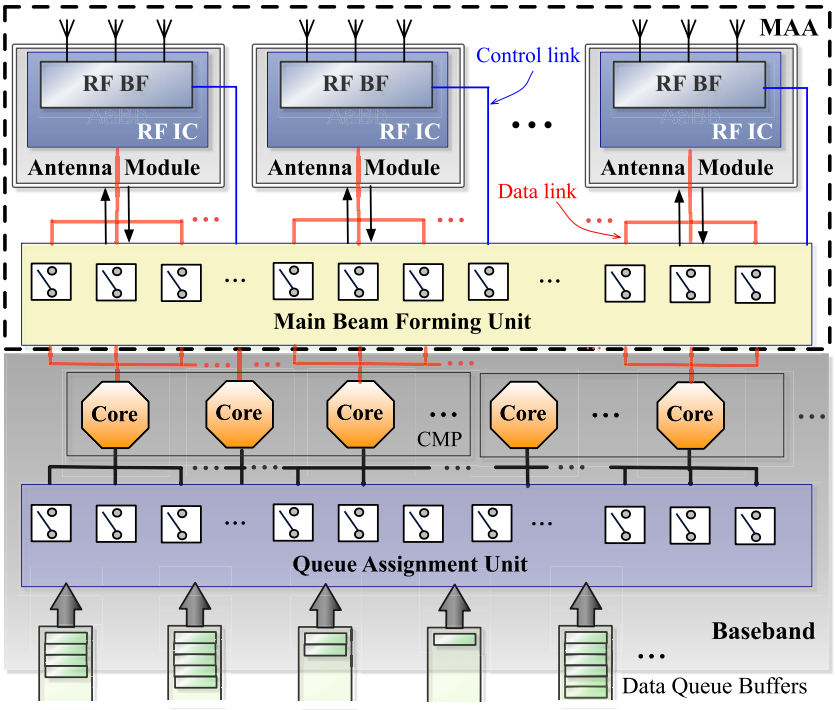

FIGURE 1. A diagram of the target mmWave base station architecture composed of MAA and baseband blocks.

in the mmWave baseband, namely any energy minimization strategy should be performed only when the queue stability is guaranteed. Therefore, decision to allocate data queue to the cores can be done by considering not only how many cores can be power gated, but also how to stabilize the remaining data queues in the system. In this paper, we propose an energy-aware queue-stable control (EQC) algorithm that realize the energy minimization while maintaining the queue stability in the system.

Meanwhile, the phase array antenna is another heavy power consumer in the mmWave base station. Therefore, we also propose an algorithm to minimize power consumption of the array antenna. Among different phase array antenna architectures, we choose the Intel's modular array antenna (MAA) architecture, due to its power and performance superiority [4]–[6]. Then, we exploit the PG technique along with using the minimum number of antenna modules. The precise algorithm (dubbed to adaptive MAA allocation algorithm) will be described in Section III.

Fig. 1 presents our target mmWave base station. To apply the proposed EQC algorithm and adaptive MAA allocation algorithm to the base station, we adopt reconfigurable networks between data queue buffers and processor cores, and between processor cores and antenna array modules (Details will be introduced in the following section). With the target base station architecture, we perform intensive simulations to evaluate the proposed algorithms. The simulation results demonstrate that the proposed EQC algorithm guarantees the queue stability, while it improve the energy efficiency of the baseband. In addition, the efficacy of the adaptive MAA allocation is proven in that energy saving is clearly achieved in the mmWave base station.

II. MILIMETER-WAVE BASE STATION ARCHITECTURE

A. mmWave MODULAR ANTENNA ARRAY ARCHITECTURE

In general, conventional antenna system architectures used in mmWave band are inadequate to combine wide-angles with high directionality. Existing reflective, parabolic dishes and lens antennas can create narrow beam, thus delivering the needed 30–40 dB antenna gain, but they lack the flexibility to cover wide angle coverage and are relatively bulky. Phased patch antenna arrays allows steering the beam to a desired direction. However, to achieve the necessary directivity, the array must consist of a large number of elements (several hundred to thousands).

Phase antenna arrays composed of a large number of antenna elements have been proposed to achieve the necessity of the wide directionality. The Phase antenna array architectures currently used for mass production employ a single module, containing a radio frequency integrated circuits (RFIC) chip that includes controlled analogue phase shifters capable of providing several discrete phase shifting levels. The antenna elements are connected to the RFIC chip via feed lines. However, due to the loss inherent in the feed lines, this approach reduces antenna gain and efficiency, and becomes a severe problem when the number of antenna elements and RF increase [7].

More recently, modular antenna array (MAA) architectures have been proposed to over come the limitation of the single module array architecture [4], [5]. The presented

MMA architectures provide flexibility in form factor choice, beam steering, and array gain in a cost effective manner. Instead of exploiting an individual antenna module, the MAA is constructed using modular, composite mmWave antenna arrays. Each module is implemented with a dedicated RFIC chip serving several antenna elements and an RF beam forming (RF-BF) unit with discrete phase shifters. The upper side in Fig. 1 (as indicated by the dashed line) shows a block diagram of the MAA architecture. Owing to the modularity, the length of the feed lines in the MAA architecture can be kept much shorter, thereby the feed line loss decrease significantly. This makes the MAA architecture much more flexible and efficient.

This paper targets such MAA architectures. Especially, we explore an MAA architecture composed of multiple sub-arrays, each of which has 8-by-2 elements. The measured transmit power and transmit antenna gain of one MAA in the target architecture are 10 dBm and 15 dBi at 60 GHz, respectively. Note that the aperture of the MAA and total transmitted power may exceed that of an individual sub-array module proportionally to the number of the sub-array modules used in a linear scale. Therefore, much narrower beams may be created and, thus, much greater antenna gains may be achieved with the MAA rather than individual sub-arrays. It is also possible that sectors of different sub-arrays may be configured in such a way as to vary the coverage angle of the composite array, thereby creating several coverage angles.

The target MAA architecture employs the main BF unit, as described in Fig. 1. This main BF unit conceptually consists of a switching network and a switching controller, so that it controls the links between antenna modules in the MAA and cores in the baseband. According to the amount of data from the cores, the switching controller in the unit dynamically assigns the data to the antenna modules by turning on/off the switches in the switching network. In this paper, we propose an MAA allocation algorithm (in Section III-B.2) for the switching controller, which aims to minimize the power consumption of the antenna modules while satisfying the data transmission quality to users.

B. mmWave BASEBAND ARCHITECTURES

The proposed baseband architecture consists of CMPs, a queue assignment unit, and data ports as shown in Fig. 1. The cores (ARM Cortex A15 in this paper) in the CMPs process data from data queue buffers inputted from the data port. In Fig. 1, the queue assignment unit conceptually includes a switching network and its controller, so that it can dynamically assigns data to the cores.

This paper focuses on power consumption of the CMPs in the baseband architecture and various data queue conditions whereby some conditions may offer chances to turn off some unused cores to save power. In fact, energy aware task allocation is a traditional research topic and have studied in-depth. Diverse task allocation strategies have been presented to minimize the energy consumption in multiprocessor System-on-Chip (MPSoC), including stochastic and energy-gradient techniques [8], bin-packing algorithms with task migration techniques [9], and energy-reliability joint optimization methods [10]. More recently, the energy-aware task allocation strategies have extended to GPU [11], real-time embedded system [12], and energy-harvesting system [13]. According to the specific characteristics and constrains of the target systems, the presented strategies have successfully carried out the energy savings.

Unfortunately, applying the previous task allocation strategies directly to the mmWave base station may be impossible. That is because the mmWave base station is different to the other platforms, in that the queue stabilization in the base station is a top-priority. Namely, any energy minimization technique should be performed only when the queue stabilization is guaranteed. Therefore, taking into consideration the queue stabilization, we propose a tailor-made task allocation strategy for the mmWave baseband.

Especially for energy minimization, we exploit the power gating (PG) technique [14], whereby all the used cores are fully utilized but the unused cores are power-gated to save energy. Note that we do not consider the dynamic voltage and (frequency) scaling technique (DV(F)S) in this paper, which may need multiple dedicated DC-DC converters whose power consumption should be accounted for (i.e., it may be significant [15], [16].) Furthermore, it can make our objective problem of the time average energy consumption minimization subject to the stability of queue to be too complicated. Instead, we uses the simple but powerful PG technique to save the CMP energy consumption, while guarantying the queue stabilization in the proposed baseband architecture. The detailed algorithm will be introduced in the following Section III-B.1.

III. ENERGY-AWARE QUEUE-STABLE CONTROL (EQC) ALGORITHM

A. ALGORITHM DESIGN RATIONALE

The proposed EQC algorithm aims at the minimization of selected (i.e., activated) cores for low-power system operation. As the power-gated cores increases, the queue-backlog will increase because the processing from the queue becomes slow. Therefore, the queue stability condition should be considered at the same time.

With the fundamental concept of stochastic network optimization [17], we design a core selection algorithm for multiple users in terms of time-average power consumption minimization subject to queue stability (details are in Sec. III-B.1). After that, the MAA module selection for each user is additionally considered for more power saving in the target base station architectures (details are in Sec. III-B.2).

B. ADAPTIVE CONTROL FOR CMP-BASED mmWave BASEBAND

The proposed adaptive control for CMP-based mmWave baseband algorithm consists of two phases, i.e., (i) adaptive core allocation for each queue in order to minimize

time-average power expenditure under queue stabilization condition; and (ii) MAA allocation for each user.

1) ADAPTIVE CORE ALLOCATION FOR EACH QUEUE

The baseband queue dynamics for a CMP-based mmWave base station in $t \in \{0, 1, \cdots\}$ can be formulated as follows [17]:

$$Q_i[t+1] \triangleq \{Q_i[t] - \mu_i[t], 0\} + \lambda_i[t] \tag{1}$$

for all associated users $u_i \in \mathcal{U}$ where $Q_i[t]$ is the baseband queue-backlog size for user $u_i \in \mathcal{U}$ (i.e., data queue buffers in Fig. 1), where $\mathcal{U}$ is the set of associated users (unit: Million Whetstones Instructions Per Second (MWIPS)), $Q_i[0] = 0$, $\mu_i[t]$ is the departure process that is depending on the capabilities of multi-core CMP for $Q_i[t]$, and $\lambda_i[t]$ is the arrival process into the queue $Q_i[t]$ from backbone networks (i.i.d. random), respectively. Note that MWIPS is one of major performance metrics in embedded computer systems. The departure process in (1) denoted by $\mu_i[t]$ depends the number of selected cores, i.e.,

$$\mu_i[t] \triangleq f_{\text{MWIPS}}\left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]\right) \tag{2}$$

for all $u_i \in \mathcal{U}$ where $f_{\text{MWIPS}}(\cdot)$ stands for the number of processed MWIPS (i.e., processing capabilities) where the input variable is the number of scheduled cores, c_j stands for the core with the index of j, and the $\mathcal{C}_k$ stands for the set of cores in a CMP-based processor p_k (i.e., CMP in Fig. 1). Note that $\mathcal{K}$ number of CMP-based processors exist in the system. In addition, $x^i_{(j,k)}[t]$ is a boolean scheduling variable, i.e., $x^i_{(j,k)}[t] \in \{0, 1\}$, for queue $Q_i[t]$ and core c_j where $x^i_{(j,k)}[t] = 1$ when the core c_j in processor p_k is selected for the processing instructions from queue $Q_i[t]$ (vice versa).

The power expenditure that is depending on the number of scheduled cores can be obviously expressed as follows:

$$\mathcal{E}_i[t] \triangleq f_{\text{power}}\left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]\right), \forall u_i \in \mathcal{U}. \tag{3}$$

The mathematical optimization program for the minimization of the time-average expected power expenditure in a queue $Q[t]$ can be formulated as follows:

$$\min : \lim_{t\to\infty} \frac{1}{t} \sum_{t'=0}^{t-1} \mathcal{E}\left[t'\right], \tag{4}$$

where $\mathcal{E}\left[t'\right]$ can be obtained by (3); and this objective function has a queue rate stability constraint:

$$\lim_{t\to\infty} \frac{1}{t} \sum_{t'=0}^{t-1} \mathbb{E}\left[Q\left[t'\right]\right] < \infty. \tag{5}$$

Then, a new variable $\vec{Q}[t]$ is defined which denotes the vector of all queues, i.e., $\vec{Q}[t] \triangleq \left[Q_1[t], \cdots, Q_{|\mathcal{U}|}[t]\right]^{\mathsf{T}}$ where T denotes the transpose of the given vector. In addition, $L[t]$ is also defined which presents the quadratic Lyapunov function $L[t] = \frac{1}{2}\vec{Q}^{\mathsf{T}}[t]\vec{Q}[t]$. Then, let $\Delta[t]$ be a conditional quadratic Lyapunov function that can be formulated as $\mathbb{E}[L(t+1) - L[t]|\vec{Q}[t]]$, i.e., the drift on time t. The dynamic policy is designed in order to solve the given optimization formulation by observing queue backlog sizes $Q[t]$ and to determine the amount of power expenditure to maximize a bound on $\mathbb{E}[\vec{\mathcal{E}}[t]|\vec{Q}[t]] - \gamma\Delta[t]$ where $\vec{\mathcal{E}}$ is the column vector of $\mathcal{E}_i[t]$, and γ is a positive constant control parameter of the policy. The proposed algorithm conducts minimizing a bound on the $\mathbb{E}[\vec{\mathcal{E}}[t]|\vec{Q}[t]] - \gamma\Delta[t]$ expression, and this gives the following dynamic algorithm, which minimizes

$$\mathcal{E}_i[t] + \gamma Q_i[t]\left(\lambda_i[t] - \mu_i[t]\right) \tag{6}$$

and this can be re-formulated as follows by (2) and (3):

$$f_{\text{power}}\left(k_i\right) + \gamma Q_i[t]\lambda_i[t] - \gamma Q_i[t] f_{\text{MWIPS}}\left(k_i\right) \tag{7}$$

where

$$k_i = \sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]. \tag{8}$$

Since $\lambda_i[t]$ is not controllable, the associated term is eliminated from (7), i.e.,

$$f_{\text{power}}\left(k_i\right) - \gamma Q_i[t] f_{\text{MWIPS}}\left(k_i\right), \tag{9}$$

where $k_i = \sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]$, and thus our final closed-form equation in order to compute the number of cores for minimizing time-average power expenditure under queue-stability in each queue $Q_i[t]$ should be presented as follows:

$$\begin{aligned}
&\tilde{x}^i_{(j,k)}[t] \\
&= \arg \min_{\forall x^i_{(j,k)}[t] \in \{0,1\}} \left\{ f_{\text{power}}\left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]\right) \right. \\
&\quad \left. - \gamma Q_i[t] f_{\text{MWIPS}}\left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t]\right) \right\}, \forall u_i \in \mathcal{U}
\end{aligned} \tag{10}$$

where $\tilde{x}^i_{(j,k)}[t]$ stands for the optimal core selection vector for queue $Q_i[t]$ under the stochastic network optimization in terms of time-average power expenditure minimization with queue stability condition. The pseudo-code of this proposed algorithm is presented in the [Phase 1] part of Algorithm 1.

2) MAA ALLOCATION FOR EACH USER

The link budget estimation procedure (depending on the number of MAA modules) is illustrated in Fig. 2. The procedure consists of two steps: (i) calculating the received signal strength; and (ii) finding the supportable modulation and coding scheme (MCS) set that is defined in IEEE 802.11ad specification [18]. For the first step, the transmitted signal from MAA modules (transmitter, Tx) to user (receiver, Rx) over 60 GHz mmWave channels will be attenuated by path-loss depending on the separation distance, as shown in Fig. 2.

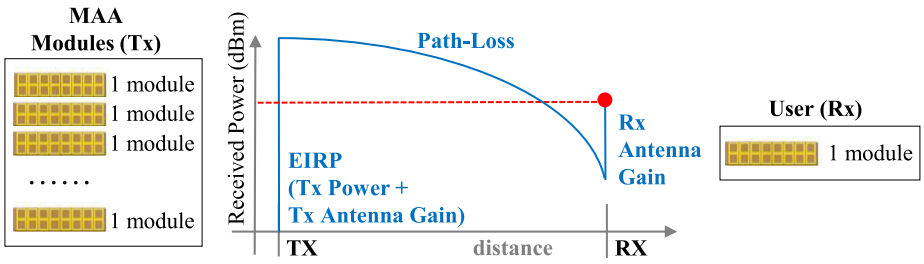

FIGURE 2. Link budget calculation procedure.

When the signal arrives at the user after experiencing attenuation effects, a receiver antenna gain will be added. This procedure can be formulated as follows:

$$P^{\mathrm{Rx}}_{\mathrm{dBm}}(n_i, d) = G^{\mathrm{Tx}}_{\mathrm{dBi}}(n_i) + P^{\mathrm{Tx}}_{\mathrm{dBm}}(n_i) - L(d) + G^{\mathrm{Rx}}_{\mathrm{dBi}} \quad (11)$$

where $P^{\mathrm{Rx}}_{\mathrm{dBm}}(n_i, d)$ is a received signal strength at user u_i where the base-station is equipped with n_i number of MAA modules in a dB scale, $G^{\mathrm{Tx}}_{\mathrm{dBi}}(n_i)$ stands for the transmit antenna gain with n_i number of MAA modules, $P^{\mathrm{Tx}}_{\mathrm{dBm}}(n_i)$ stands for the transmit power with n_i number of MAA modules, $L(d)$ is path-loss depending on the separation distance d, and $G^{\mathrm{Rx}}_{\mathrm{dBi}}$ is the receive antenna gain at user u_i, respectively. In (11), the $P^{\mathrm{Tx}}_{\mathrm{dBm}}(n_i)$ and $G^{\mathrm{Tx}}_{\mathrm{dBi}}(n_i)$ of one MAA module (i.e., $n_i = 1$) are measured to 10 dBm and 15 dBi. For n_i number of modules where $n_i \geq 1$, following formulation can be used:

$$P^{\mathrm{Tx}}_{\mathrm{dBm}}(n_i) = f_{\mathrm{dBm}}\left(f_{\mathrm{mW}}\left(P^{\mathrm{Tx}}_{\mathrm{dBm}}(1)\right) \times n_i\right), \quad (12)$$

$$G^{\mathrm{Tx}}_{\mathrm{dBi}}(n_i) = f_{\mathrm{dBm}}\left(f_{\mathrm{mW}}\left(G^{\mathrm{Tx}}_{\mathrm{dBm}}(1)\right) \times n_i\right), \quad (13)$$

where $f_{\mathrm{mW}}(x) \triangleq 10^{(x/10)}$ and $f_{\mathrm{dBm}}(x) \triangleq 10\log_{10}(x)$. In (11), the path-loss $L(d)$ can be obtained as follows [20]:

$$L(d) = A + 20\log_{10}(f) + 10n\log_{10}(d) \quad (14)$$

in a dB scale where $A = 32.5$ dB, which is a specific value for the selected type of antenna and beamforming algorithm relying on the antenna beamwidth. Further, a path-loss coefficient is set as $n = 2$, and f is a carrier frequency in GHz (i.e., $f = 60$). In (11), $G^{\mathrm{Rx}}_{\mathrm{dBi}}$ is assumed to be 15 dBi (i.e., one MAA module due to user mobile phone space limitation).

In this second step, the supportable MCS can be found using the IEEE 802.11ad based on the calculated received signal strength using (11). This is done by comparing the calculated values with receiver sensitivity values defined in the IEEE 802.11ad Table 21-3 [18]. For example, if the calculated received signal strength is −61.5 dBm, the supportable MCS is MCS7 and its associated achievable rate is 1925 Mbps. Finally, this obtained achievable rate is denoted by $\mathsf{R}_i(n_i, d)$, i.e., data rate between MAA base-station (with n_i number of MAA modules) and user u_i where the separation distance is d.

With these given equations and parameters, the achievable data rates between MAA base station and user u_i can be calculated depending on the number of allocated MAA modules. In the first phase, the numbers of cores for each associated user are determined, and thus the number of processed MWIPS for each user can be also calculated with $f_{\mathrm{MWIPS}}(\cdot)$. Then the MAA modules for each user can be allocated for processing the instructions as much as $f_{\mathrm{MWIPS}}(\cdot)$ while activating as small as possible for power expenditure minimization in radio architectures. The pseudo-code of this proposed algorithm is presented in the [Phase 2] part of Algorithm 1.

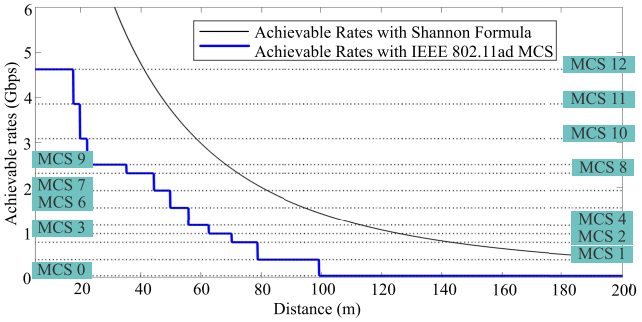

FIGURE 3. Achievable rate comparison between IEEE 802.11ad MCS-based method (in sec. III-B.2) and shannon formula based method (in appendix).

Note that this practical link budget calculation procedure with IEEE 802.11ad standard is doable only when the MCS set is provided. In theoretical contributions, the link capacity calculation with Shannon's capacity formula is widely used when the MCS set is not specified. The actual performance gap between the two methods are simulated as shown in Fig. 3; and more details about the link capacity calculation procedure with Shannon's formula is presented in Appendix.

IV. EVALUATION

In order to evaluate the efficacy of the proposed energy-aware queue-stable control (EQC) algorithm, the required parameters are measured from Juno ARM platforms in advance [19]. We measure the power consumption of each core in the big.Little processor (i.e., it contains two Cortex A15 cores and four Cortex A7 cores) equipped in the platform. Note that due to the fact that the processing at base-station baseband is heavy, we suppose that only A17 cores are used in our target baseband. The measured average power consumption

Algorithm 1 The Proposed Algorithm Including EQC Algorithm and Adaptive MAA Allocation Algorithm

Given Parameters;

- u_i: user with the index of i
- c_j: core with the index of j
- p_k: processor with the index of k
- $x^i_{(j,k)}[t]$: core scheduling vector for user u_i and the core c_j in processor p_k
- $\tilde{x}^i_{(j,k)}[t]$: optimal core selection for user u_i and the core c_j in processor p_k
- $\mathcal{U}$: the set of users associated with CMP-based baseband
- $\mathcal{C}_k$: the set of cores in CMP-based processor p_k
- $\mathcal{K}$: the set of processors
- γ: a tradeoff parameter for energy-delay
- $Q_i[t]$: queue-backlog at baseband for user u_i
- $f_{\text{MWIPS}}(k_i)$: processing capability when k_i cores are utilized for user u_i
- $f_{\text{power}}(k_i)$: power expenditure when k_i cores are utilized for user u_i
- d: distance between MAA and user
- $\mathbf{R}_i(n_i, d)$: achievable rates from MAA (with n_i number of MAA modules) to user u_i
- $\tilde{n}_i$: optimal number of MAA modules for user u_i

Proposed Algorithm;

// T: the number of discrete-time operation ;

$t \leftarrow 0$;

while $t \leq T$ **do**

For each $u_i \in \mathcal{U}$;

[Phase 1] Energy-Aware Queue-Stable Control;

- Observes $Q_i[t]$;
- Computes $\tilde{x}^i_{(j,k)}[t]$ as follows:

$$\arg \min_{\forall x^i_{(j,k)}[t] \in \{0,1\}} \left\{ f_{\text{power}} \left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t] \right) - \gamma Q_i[t] f_{\text{MWIPS}} \left(\sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t] \right) \right\}$$

[Phase 2] Adaptive MAA Allocation;

- $n^{\text{core}}_i[t] \leftarrow$ the number of selected cores for user $u_i \in \mathcal{U}$ at time t; and this can be calculated as;

$$n^{\text{core}}_i[t] = \sum_{\forall p_k \in \mathcal{K}} \sum_{\forall c_j \in \mathcal{C}_k} x^i_{(j,k)}[t].$$

- $\alpha_1 \leftarrow f_{\text{MWIPS}}\left(n^{\text{core}}_i[t]\right)$;
- Obtains d
- $n_i \leftarrow 1$; $\pi \leftarrow 0$;

while $(\pi == 0)$ **do**

- Calculates $\mathbf{R}_i(n_i, d)$ with (11), (12), (13), (14), and Table 1
- $\alpha_2 \leftarrow \mathbf{R}_i(n_i, d)$;
  if $\alpha_1 \leq \alpha_2$ **then**
  - $\tilde{n}_i \leftarrow n_i$; $\pi \leftarrow 1$;
- $n_i \leftarrow n_i + 1$;

$t \leftarrow t + 1$;

(unit: mW) and processing capabilities (unit: MWIPS) are as shown in Table 2.

TABLE 1. IEEE 802.11ad Rx sensitivity table [18].

Rx Sensitivity	Supportable MCS	Achievable Rates
−78 dBm	MCS0	27.5 Mbps
−68 dBm	MCS1	385 Mbps
−66 dBm	MCS2	770 Mbps
−65 dBm	MCS3	962.5 Mbps
−64 dBm	MCS4	1155 Mbps
−63 dBm	MCS6	1540 Mbps
−62 dBm	MCS7	1925 Mbps
−61 dBm	MCS8	2310 Mbps
−59 dBm	MCS9	2502.5 Mbps
−55 dBm	MCS10	3080 Mbps
−54 dBm	MCS11	3850 Mbps
−53 dBm	MCS12	4620 Mbps

TABLE 2. Measurement data with juno platforms.

# of cores	Average Power (Unit: milli-Watt)	MWIPS (Unit: 10^6)
0	0.00	0.0
1	104.00	683.0
2	168.88	1936.5

To use the measurement data in Table 2, the (2) and (3) should be alternatively updated as follows:

$$\mu_i[t] = f_{\text{MWIPS}}\left(c^*\right), \quad \forall u_i \in \mathcal{U}, \tag{15}$$

$$\mathcal{E}_i[t] = f_{\text{power}}\left(c^*\right), \quad \forall u_i \in \mathcal{U}, \tag{16}$$

where c^* is the number of cores in Table 2. For the performance evaluation, we assume that five CMPs are embedded in the proposed mmWave base-station baseband, where the CMP has dual A15 cores. Therefore, $c^* \in \{0, 1, \cdots, 10\}$.

The performance evaluation results of the EQC algorithm are presented in Fig. 4. In this figure, the buffer/queue-backlog dynamics are presented in terms of (i) all cores are deactivated (no processing from the queue) and (ii) the proposed energy-aware queue-stable control (EQC) algorithms with various γ settings (i.e., 1.5×10^{-20}, 3.5×10^{-20}, and 5.0×10^{-20}). As shown in the figure, it is verified that the proposed EQC algorithm guarantees the stability of the queues in the baseband. When the tradeoff constant γ is relatively large, the queue-stability is more important, whereas the system works for more energy-efficiency if γ is relatively small, as also mathematically expressed in (10). This theoretical argument is proved with the resulting performance evaluation. When $\gamma = 5.0 \times 10^{-20}$, which is the largest setting among the given three, the queue-backlog converges in a lower level ($\approx 1.8 \times 10^{12}$ MWIPS), i.e., more cores are utilized for more queue stability. On the other hand, the queue-backlog converges in a higher level ($\approx 5.8 \times 10^{12}$ MWIPS), when γ is the lowest setting as 1.5×10^{-20}.

In Fig. 5, it can be observed that the accumulated power consumption is high when $\gamma = 1.5 \times 10^{-20}$ due to the fact that the lowest γ setting works for energy-efficiency. Therefore, this theoretical formulation in Sec. III-B.1 is

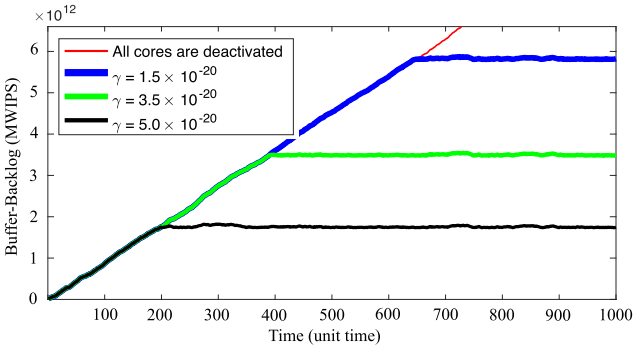

FIGURE 4. Queue dynamics.

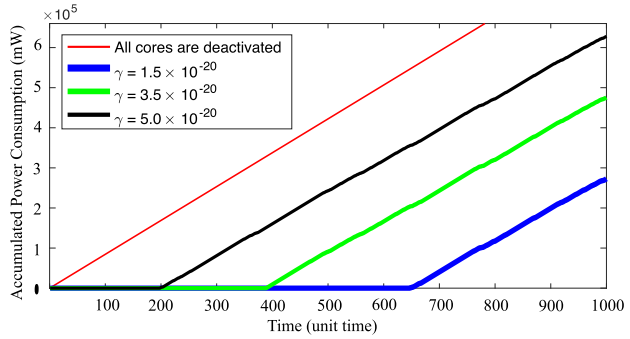

FIGURE 5. Accumulated power consumption.

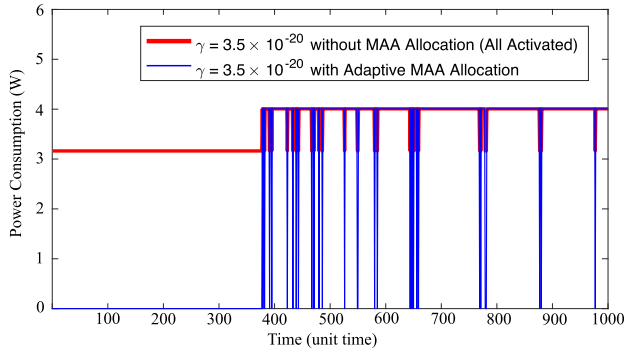

FIGURE 6. Power consumption comparison: when the adaptive MAA allocation algorithm is applied or not.

verified via this performance evaluation with measurement-based traces.

Fig. 6 presents the power consumption results, where the blue red line indicate the cases when the proposed MAA allocation algorithm is applied or not applied (thus all MAA are activated), respectively. The power consumption consists of the activated core and MAA power expenditure. Note that the EQC algorithm is applied to both cases. As presented in Fig. 6, the applying the adaptive MAA allocation algorithm clearly decreases the power consumption of the system.

Finally, Fig. 7 shows the accumulated power consumption results when the adaptive MAA allocation is applied or not applied. As seen in the figure, the adaptive MAA allocation

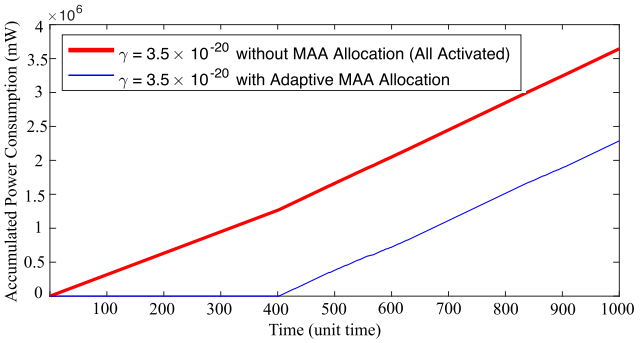

FIGURE 7. Accumulated power consumption comparison: when the adaptive MAA allocation algorithm is applied or not.

algorithm obviously improves energy efficiency, in that about 1.2×10^6 mW less power consumption is maintain after 400 unit time.

V. CONCLUSION

This paper have proposed the EQC algorithm to control the activation/deactivation of individual CMP core for pursuing time average energy consumption minimization subject to the stability of queues in the baseband. In addition, this paper also have proposed the adaptive MAA allocation algorithm to minimize power consumption of the modular antenna arrays in the base station. With the practical experimental setup, intensive simulations have been performed to demonstrate the efficacy of the proposed algorithms. The simulation results show that the EQC and adaptive MAA allocation algorithms achieve significant energy savings while maintaining the stability of the mmWave base station.

As a future research direction, baseband calculation load should be taken account for more precise latency consideration.

Appendix

60 GHz LINK CAPACITY WITH SHANNON'S FORMULA

Then the wireless link capacity (i.e., data rate) from MAA to the associated user $u_i \in \mathcal{U}$ can be obtained as follows [20]:

$$\mathsf{R}_i\left(n_i, d\right) = \mathsf{BW} \cdot \log_2\left(1 + \frac{P_{\mathrm{mW}}^{\mathrm{Rx}}\left(n_i, d\right)}{N_{\mathrm{mW}}}\right) \tag{17}$$

where $\mathsf{R}_i\left(n_i, d\right)$ stands for the achievable data rate between MAA base-station (with n_i number of MAA modules) and user u_i where the separation distance is d, BW is the channel bandwidth (2.16 GHz in IEEE 802.11ad [18]), and N_{mW} is background noise in a milli-Watt scale. In (17), $P_{\mathrm{mW}}^{\mathrm{Rx}}\left(n_i, d\right)$ (received signal power) can be as follows:

$$P_{\mathrm{mW}}^{\mathrm{Rx}}\left(n_i, d\right) = f_{\mathrm{mW}}\left(G_{\mathrm{dBi}}^{\mathrm{Tx}}\left(n_i\right) + P_{\mathrm{dBm}}^{\mathrm{Tx}}\left(n_i\right) - L(d) + G_{\mathrm{dBi}}^{\mathrm{Rx}}\right). \tag{18}$$

Due to the fact that the transmit antenna gains and transmit power are 15 dBi and 10 dBm in one MAA module [5], the total power at a transmitter is

25 dBm (=316.227766 mW). Therefore, the total power at the transmitter with n_i number of MAA modules is:

$$G_{\mathrm{dBi}}^{\mathrm{Tx}}(n_i) + P_{\mathrm{dBm}}^{\mathrm{Tx}}(n_i) = f_{\mathrm{dBm}}(316.227766 \cdot n_i). \quad (19)$$

In (17), N_{mW} is obtained as [21]:

$$N_{\mathrm{mW}} = f_{\mathrm{mW}}\left(k_B T_e + 10\log_{10}(\mathsf{BW}) + \mathcal{L} + F_N\right) \quad (20)$$

where $k_B T_e$ is a noise power spectral density (-174 dBm/Hz [21]), $\mathcal{L}$ is implementation loss (10 dB [20]), and F_N is a noise figure (5 dB [20]). Therefore, N_{mW} is 2.7193×10^{-7} mW.

REFERENCES

[1] Samsung Newsroom. *SK Telecom and Samsung Electronics Complete Field Trial of Handover between mmWave 5G Base Stations*. Accessed on Sep. 20, 2016. [Online]. Available: https://news.samsung.com/global/sk-telecom-and-samsung-electronics-complete-field-trial-of-handover-between-mmwave-5g-base-stations

[2] Intel Newsroom. *Intel Spotlights 5G Progress*. Accessed on Mar. 7, 2017. [Online]. Available: https://newsroom.intel.com/editorials/intel-spotlights-5g-progress

[3] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 106–112, Apr. 2014.

[4] R. J. Weiler *et al.*, "Enabling 5G backhaul and access with millimeter-waves," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Bologna, Italy, Jun. 2014, pp. 1–5.

[5] J. Kim, L. Xian, and A. S. Sadri, "Numerical simulation study for frequency sharing between micro-cellular systems and fixed service systems in millimeter-wave bands," *IEEE Access*, vol. 4, pp. 9847–9859, Dec. 2016.

[6] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[7] S. S. Holland and M. N. Vouvakis, "The planar ultrawideband modular antenna (PUMA) array," *IEEE Trans. Antennas Propag.*, vol. 60, no. 1, pp. 130–140, Jan. 2012.

[8] B. Gorjiara, N. Bagherzadeh, and P. H. Chou, "Ultra-fast and efficient algorithm for energy optimization by gradient-based stochastic voltage and task scheduling," *ACM Trans. Des. Autom. Electron. Syst. (TODAES)*, vol. 12, no. 4, Sep. 2007, Art. no. 39.

[9] E. W. Bríao, D. Barcelos, and F. R. Wagner, "Dynamic task allocation strategies in MPSoC for soft real-time applications," in *Proc. IEEE Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Munich, Germany, Mar. 2008, pp. 1386–1389.

[10] H. Lin and X. Qiang, "Energy-efficient task allocation and scheduling for multi-mode MPSoCs under lifetime reliability constraint," in *Proc. IEEE Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Dresden, Germany, Mar. 2010, pp. 1584–1589.

[11] J. Wang, A. Dhar, D. Chen, Y. Liang, Y. Wang, and B. Guo, "Workload-allocation and thread structure optimization for MapReduce on GPUs," in *Proc. SRC TECHCON*, Sep. 2014, pp. 1–3.

[12] Z. Wang, Z. Gu, and Z. Shao, "WCET-aware energy-efficient data allocation on scratchpad memory for real-time embedded systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 11, pp. 2700–2704, Nov. 2015.

[13] X. Lin, Y. Wang, M. Pedram, and N. Chang, "Concurrent task scheduling and dynamic voltage and frequency scaling in a real-time embedded system with energy harvesting," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 35, no. 11, pp. 1890–1902, Nov. 2016.

[14] L. Benini, A. Bogliolo, and G. De Micheli, "A survey of design techniques for system-level dynamic power management," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 3, pp. 299–316, Jun. 2000.

[15] A. Sinkar, H. Wang, and N. S. Kim, "Workload-aware voltage regulator optimization for power efficient multi-core processors," in *Proc. IEEE Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Dresden, Germany, Mar. 2012, pp. 1134–1137.

[16] W. Lee, Y. Wang, and M. Pedram, "Optimizing a reconfigurable power distribution network in a multicore platform," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 34, no. 7, pp. 1110–1123, Jul. 2015.

[17] J. Kim, G. Caire, and A. F. Molisch, "Quality-aware streaming and scheduling for device-to-device video delivery," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 2319–2331, Aug. 2016.

[18] *Specification*, IEEE Standard IEEE 802.11ad, Dec. 2012.

[19] S. Yoo, Y. Shim, S. Lee, S.-A. Lee, and J. Kim, "A case for bad big.LITTLE switching: How to scale power-performance in SI-HMP," in *Proc. ACM SOSP Workshop Power-Aware Comput. Syst. (HotPower)*, Monterey, CA, USA, Oct. 2015, pp. 1–5.

[20] J. Kim, S.-C. Kwon, and G. Choi, "Performance of video streaming in infrastructure-to-vehicle telematic platforms with 60-GHz radiation and IEEE 802.11ad baseband," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10111–10115, Dec. 2016.

[21] A. F. Molisch, *Wireless Communications*. Hoboken, NJ, USA: Wiley, 2011.

JOONGHEON KIM (S'05–M'06) received the B.S. and M.S. degrees in computer science from Korea University, Seoul, South Korea, in 2004 and 2006, respectively, and the Ph.D. degree from the University of Southern California (USC), Los Angeles, CA, USA, in 2014. Before joining USC, he was a Research Engineer with LG Electronics, Seoul, from 2006 to 2009. From 2013 to 2016, he was a Systems Engineer with the Intel Corporation, Santa Clara, CA, USA. Since 2016, he has been an Assistant Professor with Chung-Ang University, Seoul. He was the recipient of the Annenberg Graduate Fellowship with the Ph.D. admission from USC in 2009.

JAE-JIN LEE received the B.S., M.S., and Ph.D. degrees in computer engineering from Chungbuk National University, Cheongju, South Korea, in 2000, 2003, and 2007, respectively. He is currently a Group Leader with the SoC Design Research Group, Electronics and Telecommunications Research Institute, Daejeon, South Korea. His research interests include processor and compiler designs in ultra-low power embedded systems.

JONG-KOOK KIM (M'04–SM'14) received the B.S. degree in electronic engineering from Korea University, Seoul, South Korea, in 1998, and the M.S. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2000 and 2004, respectively. He is currently an Associate Professor with Korea University, Seoul. His research interests include heterogeneity in computing systems, real-time mobile computing, performance measures, resource management, evolutionary heuristics, energy-aware computing, efficient computing, shipboard computing, distributed systems, artificial neural networks, and deep learning. He is a Senior Member of the ACM.

WOOJOO LEE received the B.S. degree from the Department of Electrical Engineering, Seoul National University, Seoul, South Korea, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, CA, in 2010 and 2015, respectively. He was a Senior Researcher with the SoC Design Research Group, Electronics and Telecommunications Research Institute. He is currently an Assistant Professor with the Department of Electronic Engineering, Myongji University, Yongin, South Korea. His research interest includes ultra-low power VLSI designs, SoC designs, embedded system designs, and system-level power and thermal management.