Title
Joint Power Management and Adaptive Modulation and Coding for Wireless Communications Systems With Unreliable Buffering Memories

Permalink
https://escholarship.org/uc/item/5886t2tc

Authors
Khairy, MS
Khajeh, A
Eltawil, AM
et al.

Publication Date
2014-04-22

DOI
10.1109/TCSI.2014.2309791

Peer reviewed
Joint Power Management and Adaptive Modulation and Coding for Wireless Communications Systems With Unreliable Buffering Memories

Muhammad S. Khairy, Member, IEEE, Amin Khajeh, Member, IEEE, Ahmed M. Eltawil, Member, IEEE, and Fadi J. Kurdahi, Fellow, IEEE

Abstract—To guard against process variability in advanced semiconductor nodes, especially for high-density memories, designers resort to overdesigning policies resulting in increased power consumption. A promising approach to save power is to utilize Voltage over-Scaling (VoS). However VoS results into unreliable buffering memories where a predictable statistically amount of errors are introduced to memories. The goal is to trade off channel dependent SNR slack versus hardware induced errors, to achieve predetermined quality metrics, at reduced power consumption. By design, modern communication systems attempt to minimize channel-dependent SNR slack via adaptive modulation and coding (AMC) schemes, thus reducing the gains of on-chip power management. This paper investigates the interaction between on-chip power management via VoS on embedded memories versus network based AMC techniques. A novel mathematical approach that analytically describes the system packet error rate (PER) performance under the VoS induced noise is presented. Based on this model, different AMC and power management algorithms are presented that utilize the received SNR estimates to find the best AMC mode and memory voltage that achieves performance goals at reduced power consumption. Simulation results show that the proposed algorithms can achieve up to 58% energy efficiency for the memory-subsystems compared to conventional AMC algorithm with perfect memories.

Index Terms—Adaptive modulation and coding, dynamic power management, embedded memories, energy efficient systems, low power, voltage over scaling.

I. INTRODUCTION

ADVANCES IN scaling CMOS components facilitated the integration of large embedded buffering memories into system on chip (SoC). Furthermore, the trend to integrate as many processing layers as possible on one SoC led to an explosion in embedded storage capacity as reported by the International Technology Roadmap for Semiconductors (ITRS) report [1]. With the growing share of embedded memories in terms of both area and power metrics, managing their power consumption will have a direct and significant impact on the overall efficiency of the system. However, system designers traditionally refrain from applying conventional power management schemes to embedded memory due to the inherent susceptibility to noise and the possibility of introducing errors to the data flow. In other words, designers are forced to treat memories differently than logic and usually provide a separate higher (than logic) voltage rail. While maintaining perfectly functional memories under all anticipated operating conditions, was both acceptable and achievable in older technologies, it is now becoming excessively difficult and expensive due to process variations which lead to excessive design marging.

Recently, aggressive voltage over scaling (VoS) [2]–[4] has been proposed as an effective technique to significantly reduce both the dynamic and leakage components of power consumption for buffering memories, and thus achieving energy efficient systems [5], [6]. However, reducing the supply voltage of buffering memories via VoS results in spatially uniform random errors [7]–[9]. The amount and rate of these errors are controlled by the reduction of the supply voltage. Thus, when the wireless receiver experiences a high signal to noise (SNR) channel, a power management unit can trade off hardware errors (noise) versus channel SNR, to minimize power consumption [10], [11]. On the other hand, in most current wireless communication systems such as LTE and 802.11ac, adaptive modulation and coding (AMC) is employed in which the suitable channel code and modulation are chosen based on the channel conditions to maximize the system throughput [12], [13] and minimize SNR slack. Thus both loops are competing for the same resource, namely SNR slack. The outer loop (AMC loop) utilizes SNR slack to increase throughput, while the inner loop (on-chip power management) utilizes slack to reduce power.

This paper presents novel algorithms that jointly handle the selection of the AMC modes and the on-chip power management of VoS buffering memories. The statistical correlation of the Rayleigh fading channel is exploited to model the system as a Markov decision process (MDP) with the objective of finding the appropriate AMC mode and suitable supply voltage for the buffering memories such that the required packet error rate (PER) performance is guaranteed, while maximizing the power savings. To achieve these goals, 1) an equivalent model of a VoS memory (composed of an error-free memory followed by a statistical additive error injector) is utilized to develop a mathematical model that combines the channel noise and the buffering memory errors into an equivalent error distribution; 2) The proposed framework is used to analytically model the PER of the system with the VoS memory for the different modulation and coding modes and all VoS levels of the buffering.
memory; 3) Three novel power management techniques that control the supply voltage of the buffering memory and the selection of the AMC mode based on the proposed framework are presented.

Some relevant work in literature [14], [15] considered energy efficient multimedia wireless communications. However, the adopted power management is a system level power management which employs either simple on/off management for blocks or applying dynamic voltage and frequency scaling (DVFS) to a processor unit. To the authors’ knowledge, this work is the first to jointly address embedded memory VoS reliability versus network level AMC selection to enable higher energy efficiency.

The remainder of the paper is organized as follows: Section II illustrates the system model where a mathematical formula of the system performance in terms of bit error rate (BER) and packet error rate (PER) under VoS buffering memories are derived. The problem formulation and the proposed power management policies are presented in Section III and the simulation results are discussed in Section IV. Discussions and comparison of the performance of the proposed algorithms are presented in Section V. Finally, conclusions are drawn in Section VI.

II. SYSTEM MODEL

A typical OFDM system is shown in Fig. 1 in which a large buffering memory is used to buffer several OFDM symbols. The wireless channel is assumed to be Rayleigh fading and is modeled as a finite state Markov channel (FSMC). At the receiver side, a power manager tracks the received SNR to opportunistically reduce the power consumption of the system by aggressively scaling down the supply voltage of the buffering memory when the receiver experiences a high SNR channel. Meanwhile based on the channel conditions, the AMC unit will select an appropriate channel code and modulation scheme and feed it back to the transmitter.

A. FSMC for the Wireless Fading Channel

The received signal in a rich multipath wireless fading channel is commonly modeled as a Rayleigh distribution. For such Rayleigh channel with slow variation, finite state Markov channel (FSMC) is a very useful and popular model that considers the correlation of the fading between the channel samples [16], [17]. In FSMC, the range of the channel signal-to-noise (SNR) is portioned into $K$ non-overlapping intervals denoted by $[\Gamma_k, \Gamma_{k+1})$, $k = 0, 1, \ldots, K - 1$, where $\Gamma_0 = 0$, $\Gamma_{K+1} = \infty$. Let $\mathcal{H} = \{H_0, H_1 \ldots H_{K-1}\}$ denote the state space of the FSMC, then the channel is said to be in state $H_k$ if the received SNR is in the interval $[\Gamma_k, \Gamma_{k+1})$. The channel SNR is assumed to be the same for the whole packet duration, but it varies from one packet to another based on a time-correlated process. The received SNR has an exponential distribution as described in (1) where $\bar{\gamma}$ represents the average received SNR. In this work, data aided (DA) techniques based on the known preamble samples are used to estimate the CSI and hence the SNR [18]–[20].

$$P(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}}, \gamma \geq 0; \quad (1)$$

B. Unreliable Embedded Buffering Memory

The buffering memory state is defined by the value of its supply voltage. We define $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ as the set of memory states, where $n$ represents the number of different states of the memory or equivalently the number of the different supply voltages. In our model, we assumed that the memory can operate at four different distinct voltages. Accordingly, the memory has four different states corresponding to the four values of the supply voltage. Embedded buffering memories under supply voltage scaling introduce errors in the stored words in the form of spatially random uniform bit flips. The rate of these bit flips $P_e(V_{dd})$ is characterized by the memory supply voltage [8]. Table I shows the buffering memory voltage states, the corresponding error rate in the memory and the normalized power consumption based on 6T SRAM in 65 nm process technology [21].

An equivalent model of the faulty memory is shown in Fig. 2, where the faulty memory is replaced by an ideal error-free memory followed by a virtual error injector. Hence,
the word $Y$ read from memory could be expressed as a function of the stored word $X$ as:

$$Y = X + \text{err}(P_e(v_j))$$  \hfill (2)$$

We assume that data is stored in memory as standard two’s complement binary numbers which are composed of $N$ bits, with $d$ and $r$ representing the decimal and fractional parts of the word respectively, such that $N = d + r$ bits. Since the bit flips are spatially independent, a retrieved data word from memory could be either error free or erroneous, as shown in (3).

$$\text{err} = \begin{cases} 0, & \text{No bit flip} \\ e_k, & \text{k bit flips (}1 \leq k \leq N) \end{cases} \hfill (3)$$

In case of one bit flip at any bit location $i$ where $0 \leq i \leq N-1$, the magnitude of the error $e_1$ could be expressed as

$$e_1 = \pm 2^{1-r}, \quad 0 \leq i \leq N-1$$  \hfill (4)$$

In the case of two simultaneous bit flips, the error magnitude $e_2$ is similarly expressed as

$$e_2 = \pm (2^{1-r} \pm 2^{l-r})$$  \hfill (5)$$

where $0 \leq i \leq N-1; i+1 \leq j \leq N-1$

In general, the error magnitude $e_k$ due to $k$ simultaneous bit flips is given by

$$e_k = \pm (2^{1-r} \pm 2^{l-r} \pm \ldots \pm 2^{j-r} \pm 2^{r})$$  \hfill (6)$$

where $0 \leq i \leq N-1; i+1 \leq j \leq N-1; \ldots; q+1 \leq l \leq N-1$

Hence, assuming that each bit has equal probability of being one or zero, the probability of having $k$ simultaneously bit flips resulting into error magnitude $e_k$, is given by:

$$P(e_k) = \frac{1}{2^k} P_e(v_j)^k (1 - P_e(v_j))^{N-k}$$  \hfill (7)$$

By assuming a 1% error rate or less, and ignoring high order bit flips (more than 1 flip per word), the error probability mass function (PMF) could be approximated as follows:

$$f_{\text{err}}(\text{err}) \approx \begin{cases} (1 - P_e(v_j))^N, & \text{err} = 0 \\ \frac{1}{2} P_e(v_j) (1 - P_e(v_j))^{N-1}, & \text{err} = e_1 \end{cases} \hfill (8)$$

This approximation is even more valid for more aggressive memory error rates of 10% or less, since the contribution of higher order bit flips to the error PMF is extremely small. For example, the probability to have 2 or more bits flip under 10% memory error rate is less than 0.13% which is extremely small and can be ignored.

Because of the symmetry of the error distribution around zero, it has a zero mean, while the variance is expressed as:

$$\sigma_e^2(v_j) = \sum \text{err}^2 \times f_{\text{err}}(\text{err})$$  \hfill (9)$$

Using the approximate distribution of the error given in (8), the variance could be written as:

$$\sigma_e^2(v_j) = \sum e_1^2 P_e(v_j) \left(1 - P_e(v_j)\right)^{N-1}$$  \hfill (10)$$

Finally,

$$\sigma_e^2(v_j) = \frac{P_e(v_j) \left(1 - P_e(v_j)\right)^{N-1}}{2} \sum_{i=0}^{N-1} (2^i - r)^2$$  \hfill (11)$$

C. Error Propagation Through the FFT Stage

Fast Fourier transform (FFT) is a fundamental block in all orthogonal frequency division multiplexing (OFDM) based systems. Based on the equivalent model of VoS memories presented in the previous subsection, the erroneous retrieved data from the memory is expressed as the sum of the original data plus the error as in (2). Since the FFT is a linear operation, the output of the FFT could be expressed as:

$$\text{FFT}(x + \text{err}) = \text{FFT}(x) + \text{FFT}(\text{err})$$  \hfill (13)$$

Then, the error after the FFT is expressed as:

$$\text{FFT}(\text{err}) = \frac{1}{\sqrt{N_{\text{FFT}}}} \sum_{k=0}^{N_{\text{FFT}}-1} (e_r(k) + j e_i(k)) e^{-j \frac{2 \pi k n_{\text{FFT}}}{N_{\text{FFT}}}}$$  \hfill (14)$$

Based on the central limit theory [22], the addition of a large number of random variables, approaches an asymptotic Gaussian distribution. Therefore, the distribution of the error after the FFT can be modeled as a Gaussian. In [23], a detailed derivation of the mean and variance of the resulting Gaussian distribution has been proposed. Based on the fact that the error distribution before the FFT has zero mean, the real and imaginary parts of the error after the FFT are approximated as Gaussian with mean and variance as shown in (15) [23].

$$E_{r,\text{FFT}} \sim \mathcal{N}(0, \sigma_e^2(v_j)), \quad E_{i,\text{FFT}} \sim \mathcal{N}(0, \sigma_e^2(v_j))$$  \hfill (15)$$

D. Equivalent Gaussian (Channel Noise and Memory Error)

The received signal for subcarrier $k$ in a Rayleigh fading channel is expressed as:

$$y_k = h_k s_k + n_k + E_{r,\text{FFT},k}$$  \hfill (16)$$

where $n_k$ is the complex Gaussian noise of zero mean and variance $\sigma_n^2$ and $E_{r,\text{FFT},k}$ is the complex memory error after the FFT which is approximated as a complex Gaussian of zero mean and variance $\sigma_e^2 = 2\sigma_e^2(v_j)$. Since the channel Gaussian noise and the Gaussian memory errors are independent, we can combine them into an equivalent Gaussian $n_k = n_k + E_{r,\text{FFT},k}$ of zero mean and variance $E_{r,\text{FFT},k}$ given by:

$$\sigma^2(v_j) = \sigma_n^2 + \sigma_e^2(v_j)$$  \hfill (17)$$
E. Least Square Equalization (Zero Forcing)

Based on the combined equivalent noise $\tilde{n}_k$, the received signal for each subcarrier after the FFT can be expressed as:

$$y_k = h_k s_k + \tilde{n}_k$$  \hspace{1cm} (18)

Any channel equalization technique such as ZF or MMSE equalizer can be used within the framework of the proposed power management algorithm to populate look-up tables. The objective is to find the impact of voltage scaling on the system BER performance either in a derived closed form or numerically based on simulations. Least square equalizer is used in this section to derive a closed-formula of BER for different voltage levels. The same analysis presented here can be extended to other equalization methods such as the MMSE [29]. While ZF results in noise amplification at low SNR, it achieves very close performance to the MMSE equalizer at higher SNR, which is the region of operation of the proposed algorithms. Thus assuming least squares equalization, one can express the equalized signal $\hat{y}_k$ as:

$$\hat{y}_k = s_k + \tilde{n}_k / h_k$$  \hspace{1cm} (19)

The expression for the BER of a square MQAM with Gray bit mapping as a function of received SNR $\gamma$ and constellation size $M$ is expressed in (20) [24]

$$\text{BER}(\gamma, v_j) \approx \frac{2}{\log_2 M} \left( 1 - \frac{1}{\sqrt{M}} \right) \text{erfc} \left( \sqrt{\frac{1.5\gamma}{\sigma^2(M - 1)}} \right)$$  \hspace{1cm} (20)

Hence, the average BER given the channel state $H_k$ and the buffering memory supply voltage $v_j$ can be expressed by (21).

$$\text{BER}(H_k, v_j) = \frac{1}{p_k} \int_{r_k}^{r_{k+1}} \text{BER}(\gamma, v_j) \times f(\gamma)d\gamma$$  \hspace{1cm} (21)

where $p_k$ representing the steady state probability of being in a state $k$ is given by (22).

$$p_k = \int_{r_k}^{r_{k+1}} \text{P}(\gamma)d\gamma = e^{-r_k} - e^{-r_{k+1}}$$  \hspace{1cm} (22)

Therefore, based on the BER expression in (20) and the exponential distribution of the channel SNR in (1), the average BER in (21) could be derived in a similar way to [17] and can be expressed as

$$\text{BER}(H_k, v_j) = \frac{f_k - f_{k+1}}{p_k}$$  \hspace{1cm} (23)

where

$$f_k = a_M \times \text{erfc} \left( \sqrt{\frac{r_k}{\sigma^2 / b_M}} \right) \times e^{-r_k/\gamma_0}$$

$$- a_M \sqrt{\frac{\gamma_0}{\sigma^2 / b_M + \gamma_0}} \text{erfc} \left( \sqrt{\frac{\gamma_0 r_k}{\sigma^2 / b_M + \gamma_0}} \right)$$

and

$$a_M = \frac{2}{\log_2 M} \times \left( 1 - \frac{1}{\sqrt{M}} \right), b_M = \frac{1.5}{M - 1}$$

III. PROBLEM FORMULATION

A. Independent DPM and AMC

The objective of the AMC is to utilize the knowledge of the channel state information to maximize the data rate by adapting
the transmission modulation and coding scheme such that a certain packet error rate performance $P_{b_0}$ is achieved. In the first power management algorithm (Algorithm 1), we consider independent on-chip dynamic power management (DPM) and AMC loops. The SNR range is divided into $N + 1$ non-overlapping intervals where $N$ denotes the number of AMC modes. The SNR thresholds $\{\Gamma_j\}_{j=0}^N$ are obtained by solving (25) iteratively such that the target packet error rate for each AMC mode is set to $P_{b_j}$ [26].

\[
\text{PER}_n = \frac{1}{p_k} \int_{\gamma_n}^{\Gamma_{n+1}} \text{PER} (\gamma) d\gamma = P_{b_0}, \quad n = 1, 2, \ldots, N \tag{25}
\]

Then, based on the received SNR, an AMC mode $n$ is chosen when $\gamma \in [\Gamma_n, \Gamma_{n+1})$. It is worth mentioning that since the selection of the AMC mode and buffering memory power management are independent, the choice of the SNR thresholds are obtained assuming perfect buffering memories (i.e., operating at the nominal supply voltages).

The problem of buffering memory DPM is modeled as unconstrained Markov decision process (MDP) with the composite states space $S = \mathcal{H} \times \mathcal{V} \times \mathcal{AMC}$ and actions space $A = A_v$. The objective is to find the optimal policy $\pi^*$ that minimizes the average cost function given by (26).

\[
J(\pi; \lambda) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E [c(s_i, \pi(s_i))] \tag{26}
\]

At each time step $i$, the power manager observes the channel state, the AMC mode and the buffering memory supply voltage. Then, based on that composite state, it chooses a control action $a_i = \pi(s_i)$ which incurs the cost of

\[
c(s_i, \pi(s_i)) = P(s_i, \pi(s_i)) + \lambda \times \text{PER}(s_i, \pi(s_i)) \tag{27}
\]

where the power cost $P(s_i, \pi(s_i))$ represents the sum of the memory power consumption and the memory switching power which is given by (28). The share of the switching power can be considered negligible since the switching frequency of the supply voltage is relatively small [27].

\[
P(s_i, \pi(s_i)) = P_{\text{Mem}}(\pi(s_i)) + P_{\text{switching}}(\pi(s_i)) \tag{28}
\]

Solving the unconstrained MDP is equivalent to solving the dynamic programming (DP) equation in (29) either by value iteration or policy iteration [28].

\[
J^{\pi^*}(s) = \min_{\pi \in \Phi} \left[ c(s, \pi(s)) + \sum_{s' \in S} P(s' | s, \pi(s)) \times J^{\pi^*}(s') \right], \forall s \tag{29}
\]

In this model of the MDP, the state transition probability can be expressed:

\[
P(\left\{ s' | s, \pi(s) \right\}) = P(h | h) \times P(\mathcal{AMC}' | h) \times P\left( v' | v, \pi(s) \right) \tag{30}
\]

where

\[
P(v' | v, \pi(s)) = \begin{cases} 1, & v' = \pi(s) \\ 0, & \text{otherwise} \end{cases}
\]

and

\[
P(\mathcal{AMC}' | n) = \begin{cases} 1, & \gamma \in [\Gamma_n, \Gamma_{n+1}) \\ 0, & \text{otherwise} \end{cases}
\]

Based on the first order FSMC where channel state transition occurs only between consecutive states [16], the channel transition probability is given by (31).

\[
P_{H_{k-1} \rightarrow H_{k+1}} = \frac{N(\Gamma_{k+1}) \times T_{\text{pkt}}}{p_k}, \quad k = 0, 1, \ldots, K - 1
\]

\[
P_{H_{k} \rightarrow H_{k-1}} = \frac{N(\Gamma_k) \times T_{\text{pkt}}}{p_k}, \quad k = 1, 2, \ldots, K \tag{31}
\]

where $T_{\text{pkt}}$ is the packet time in second and $N(1')$ is the level cross rate which can be written as in (32) given the Doppler frequency $f_D$ [16].

\[
N(\Gamma_k) = \sqrt{\frac{2\pi \Gamma_k f_D e^{-\frac{\Gamma_k}{\gamma_0}}}{\gamma_0}} \tag{32}
\]

The details of the algorithm are summarized in Algorithm 1 where the algorithm is divided into two parts; offline and online parts. All the computations of both the buffering memory power management and the AMC algorithms are performed offline. Look-up tables (LUTs) are used to store power policy $a^* = \pi^*(s_i)$ which is the solution of the DPM problem, as well as the SNR thresholds. During the online phase of the algorithm, the LUT is accessed with the current system state index $s_i = \{h_i, v_i, \mathcal{AMC}_i\}$ to find the optimal action which is the next buffering memory supply voltage.

B. Joint DPM and AMC

In this section, we present a novel AMC algorithm which is aware of both the wireless channel conditions and the buffering memory status. The details of the proposed algorithm are presented in Algorithm 2. The conventional AMC chooses the AMC mode based only on the wireless channel. However, the proposed algorithm finds the appropriate AMC mode based on the combined effect of the wireless channel and hardware errors. Fig. 5 illustrates this idea where the effective SNR is based on the combination of the wireless channel and the VoS buffering memory.
In the previous section, an equivalent Gaussian noise model was derived which combined both the channel noise and hardware errors. This model enabled the system designer to mathematically characterize the packet error performance of the system under different supply voltages of the buffering memory. Thus, given a certain supply voltage \( v_i \) and a target PER \( P_{\text{e}} \), the SNR thresholds \( \{\Gamma_k(v_i)\}_{k=1}^N \) can be derived in a similar fashion such that a packet error rate of \( P_{\text{e}} \) is achieved for each AMC model. In this scenario, different sets of SNR thresholds are defined such that each set corresponds to a certain supply voltage. The SNR threshold sets are obtained during the offline portion of the AMC algorithm. Based on these SNR thresholds, the joint optimization of the AMC and buffering memory power management is similarly modeled as an unconstrained MDP with the composite states space \( S = \mathcal{H} \times \mathcal{V} \times \mathcal{A} \) and action space \( A = A_0 \). The main difference in this scenario is the formulation of the state transition probability (33) in which the choice rule of the AMC is dependent on the both the channel state and the buffering memory supply voltage

\[
P(s'|s,a) = P(h|h) \times P(AMC'|h,v) \times P(v'|v,a)
\]

where

\[
P(AMC') = \begin{cases} 
1, & \gamma(v) \in [\Gamma^*_n(v) \Gamma^*_{n+1}(v')] \\
0, & \text{otherwise}
\end{cases}
\]

Similarly, the results of the MDP problem are stored in LUTs which are accessed online to find the proper action of the buffering memory supply voltage (steps 1–2 in the online part of the algorithm). Based on the received channel SNR and the next supply voltage of buffering memory \( v_{i+1} \), an AMC mode \( n \) is chosen when the effective SNR \( \gamma(v) \in [\Gamma^*_n(v_{i+1}) \Gamma^*_{n+1}(v_{i+1})] \).

In this scenario, because of the extra knowledge of the buffering memory status, it is expected that the system will satisfy the target PER requirement of the system at the cost of slight throughput degradation while achieving considerable amount of power savings. Simulations results in the next section illustrate the trade-off between different power management policies in achieving such joint optimization.

C. Aggressive DPM and AMC

The objective of buffering memory power management is to minimize power consumption by utilizing the available SNR slack to aggressively reduce the supply voltage. However, in the presence of SNR slack, the AMC technique at the transmitter adapts to support a higher modulation and coding scheme leading to a conflict with the on-chip power manager. In this approach, we aim to jointly optimize for maximizing the power savings with the minimum degradation in system throughput. The main idea is to artificially control the AMC loop via reporting a lower AMC mode, even if the receiver is experiencing a good channel. This is achieved by combining the wireless channel and buffering memory in a similar way to the joint PM and AMC explained in Fig. 5. However, in this case, the effective SNR is calculated always based on the worst case supply voltage of the buffering memory \( v_{\text{agg}} \). This will result into lower reported SNR and accordingly, the base station will lower the modulation and a high SNR slack will be available at the receiver. The power manager can utilize this slack to aggressively reduce the supply voltages of the buffering memories. In other words, the available slack becomes part of the optimization problem. The details of the algorithm are described in Algorithm 3.

The main difference in the calculation of the SNR thresholds \( \{\Gamma_k\}_{k=1}^N \) is the assumption of the most aggressive memory
supply voltage which will result into only one set of SNR thresholds. The buffering memory power management problem is formulated in a similar fashion to the joint DPM and AMC as an unconstrained MDP with the composite states space \( S = H \times V \times \text{AMC} \) and action space \( A = A_v \). The state transition probability can be written as

\[
P \left( s' | s, a \right) = P \left( h' | h \right) \times P \left( \text{AMC}' | h \right) \times P \left( v', a \right)
\]

where

\[
P \left( \text{AMC}' = n | h \right) = \begin{cases} 1, & \gamma \left( v_{agg} \right) \in \Gamma_n \Gamma_{n+1} \\ 0, & \text{otherwise} \end{cases}
\]

After solving the MDP problem, results are stored into LUTs and then utilized in the online part of the algorithm. The choice of the AMC mode for the next packet is based on the effective SNR \( \gamma \left( v_{agg} \right) \) where an AMC mode \( n \) is chosen when the channel SNR \( \gamma \left( v_{agg} \right) \in \Gamma_n \Gamma_{n+1} \). This problem formulation will force the transmitter to utilize a lower AMC mode, thus creating a higher SNR slack at the receiver side, allowing the power manager to choose the appropriate supply voltage for the buffering memory.

The size of the LUT required to store the policy actions of any of the previous algorithms depends mainly on the number of the states in the space \( S \) and the possible values of the received SNRs where there exists a policy for each SNR. Generally, considering \( N \) AMC modes, \( K \) channel states and \( m \) voltage states, the size of the state \( S \) will be \( m \times K \times N \). Thus, considering \( N_{\text{hit}} \) for quantizing the received SNR, the number of entries of the LUT is \( 2^{N_{\text{hit}} \times m} \times K \times N \) where each entry requires \( \log_2 m \) bits to stores the policy action. Therefore, as an example of 4 AMC modes, 5 channel states, 4 supply voltages states and 10 bits quantization of the SNR, the size of the required LUT is 20 Kbytes.

IV. SIMULATION RESULTS

To justify the benefits and compare the performance of the proposed algorithms 1) Independent DPM and AMC 2) Joint DPM and AMC and 3) Aggressive DPM and AMC discussed in Section III, a simulation framework based on Fig. 1 was setup in a WI-FI environment. The AMC modes of the OFDM system are chosen to be BPSK, QPSK, 16 QAM and 64 QAM modulation with rate 1/2 convolutional codes and a constraint length of 7. The OFDM symbol has 128 subcarriers and the packet size is set to 1500 Bytes. The wireless channel is based on a Rayleigh channel model with a maximum Doppler frequency \( f_d \) of 100 Hz. The average received SNR is set to 15 dB (\( \gamma = 15 \) dB). The memory supply voltage can take one value of the four discrete levels shown in Table I.

A. Independent DPM and AMC

In this scenario whenever a slack in the SNR exists, the AMC algorithm selects the appropriate modulation scheme of the next transmitted packet such that 10% PER is statistically achieved with the correlated Rayleigh fading channel. Thus, the objective of the power manager is to utilize the residual slack in SNR to save power consumption by lowering the supply voltage of the buffering memory. In order not to drastically degrade system performance, the power manager should be conservative in selecting the buffering memory supply voltage. Thus, it is expected that power savings will not be substantial. In reality, the value of the Lagrangian multiplier in the cost function in (27) controls the trade-off between system throughput and power consumption. Fig. 6 illustrates simulation results of the uncoded BER, normalized power consumption and system throughput (Mbps) for different values of the Lagrangian multiplier. As expected, a large values of \( \lambda \) results in optimizing system performance (PER) at the expense of no power savings, while a low value of \( \lambda \) results in a more aggressive power reduction strategy (at the expense of higher BER and/or degraded throughput). As shown in Fig. 6, one can find two separate ranges of \( \lambda \) that result into considerable power savings at the cost of slight throughput degradation. For instance, setting \( \lambda = 100 \) will create a power policy which results into almost 20% power savings without noticeable performance degradation (less than 1% degradation) since the amount of introduced errors are extremely small \( P_e \sim 1.69 \times 10^{-15} \sim 5.21 \times 10^{-12} \). A more aggressive power policy \( \{ \lambda = 2 \} \) will result into higher power savings 35% at the cost of degraded PER performance and 20% lower throughput. The trade-off between system throughput and PER versus power

Fig. 6. Lagrangian multiplier trade-off over system performance and average power consumption for independent AMC and DPM.

Fig. 7. Trade-off between PER, system throughput and power consumption for independent AMC and DPM.
consumption is illustrated in Fig. 7 where each point represents the performance metrics of the power management policy for a given value of $\lambda$.

B. Joint DPM and AMC

In this scenario, the required packet error rate is achieved by adapting the AMC mode according to the joint effect of wireless channel and the status of the buffering memory. As compared to the independent AMC and DMP, a higher power savings can be achieved with an improved system performance. Fig. 8 shows the performance metrics (BER and throughput) and the average power consumption for different values of $\lambda$. As expected, the system achieves a constant bit error rate independent of the value of $\lambda$. Very high values of $\lambda$ guarantee achieving the highest throughput without any power savings. However, as the value of $\lambda$ is reduced, more power savings are achieved at the cost of lower AMC modes which in turns reduces the system throughput.

It is very important to point that for extreme small value of $\lambda$, the system can still function and achieve some throughput while for the independent AMC and DPM the system is starving with 100% PER and no throughput. The trends of the PER and throughput versus the power consumption for different power policies are shown in Fig. 9.

C. Aggressive PM With AMC

This mode is very efficient when it is more important to save power while maintaining a reasonable throughput. Simulations results are depicted in Figs. 10 and 11 in which the BER, PER, average power consumption and throughput for different values of $\lambda$ are shown.

In this algorithm, the choice of SNR thresholds is based on the most aggressive supply voltage. Thus, for very small values of $\lambda$ where most aggressive VoS is applied, the target PER (10%) is achieved. Higher values of $\lambda$ will result in a better BER and PER performance. However, the improvement of the PER from 10% to 0.1% will only lead into slight throughput enhancement of 9.9%. Thus, in this mode, it is very efficient to set $\lambda$ to a small value to maximize the power savings at the cost of negligible throughput degradation.

V. DISCUSSION

In this section, we compare the three proposed algorithms in Section III in terms of throughput and power consumption for different values of SNRs. Although, the proposed joint AMC and PM algorithms save a considerable amount of power at the receiver side, it comes at the cost of throughput degradation. For a fair comparison between these algorithms, we introduce a figure of merit (FoM) that links both quantities. This metric
is defined in (35) as the ratio of the average throughput per the normalized power consumption of the buffering memory. High values of this metric reflect an energy efficient system that can achieve good performance (throughput) with lower power consumption.

\[ \Upsilon = \frac{\eta}{P_c} \]  

(35)

Fig. 12 depicts the simulation results of the average throughput and the normalized power consumption of buffering memories at different SNR. The proposed algorithms in Section III are compared to the ideal case of AMC with perfect memory, which achieves the best performance at the cost of highest power consumption. At low SNRs, all the proposed algorithms are operated at the lowest voltage level (most aggressive VoS level). This is due to the fact that the introduced errors in the memory are much lower in power as compared to the channel noise power. Thus, large power savings could be achieved. However, with the increase of the SNR, the effect of the introduced hardware noise due to VoS becomes comparable to the channel noise. Thus, the buffering memories supply voltage is increased with the SNR.

For the independent AMC and PM, two different scenarios are considered: a conservative and a moderate power savings approach. In the conservative mode, a considerable amount of power savings is achieved with almost negligible throughput degradation while the moderate mode delivers a higher amount of power savings at the cost of slight throughput degradation. Both scenarios are illustrated in Fig. 12. The joint AMC and PM algorithm has the same behavior of the independent approach with moderate \( \lambda \) in terms of normalized power consumption but outperforms significantly in terms of average throughput since it considers the lumped effect of both channel noise and hardware errors. At high SNR, the proposed joint algorithm delivers higher throughput than the independent approach with the conservative \( \lambda \). The aggressive DPM with AMC always delivers the highest power savings since the buffering memory is always operated at the lowest supply voltage. As a result, it has the lowest throughput among the proposed algorithms.

Finally, Fig. 13 shows a comparison of the FoM of the proposed algorithms versus the AMC with perfect buffering memory. As expected the joint AMC with PM outperforms all the other algorithms up to an SNR of 23 dB (based on the system setup and parameters presented in Section IV). Beyond this value, a large slack in the SNR exists, which in turn can be utilized by aggressively reducing the supply voltage of the buffering memory to maximize the power savings and achieve a better energy efficient system. In other words, at very high SNR, it is more efficient to utilize the aggressive PM approach than the joint one.

The proposed joint PM and AMC achieves up to 32% and 58% improvement in energy efficiency as compared to the independent AMC and DPM and the conventional AMC with perfect memory respectively. Furthermore, the independent AMC and PM approach outperforms the conventional AMC algorithm with perfect buffering memories and delivers up to 26% enhancement in energy efficiency.

It is important to note that although the proposed PM algorithm enhances the receiver energy-efficiency, it comes at the cost of lower energy-efficiency at the transmitter side. The evaluation of the overall system energy-efficiency is a future research direction where network level simulation will be utilized to quantify the system energy efficiency.

VI. CONCLUSION

In this paper, a novel mathematical model that characterizes system performance in terms of PER under unreliable VoS buffering memories has been derived. Based on that model, three different algorithms that control the buffering memory supply voltage for communication systems with adaptive modulation and coding have been proposed. Simulations results showed that by jointly adapting the AMC mode and the supply voltage of the unreliable buffering memory, 58% and 32% performance gain is achieved as compared to AMC with perfect buffering memories and the independent AMC and buffering memory power management.

REFERENCES


[22] M. S. Khairy (S’08–M’14) received the B.Sc. and M.Sc. degrees with honors from the Electronics and Communications Department, Cairo University, Egypt in 2005 and 2009, respectively, and the Ph.D. degree from the University of California, Irvine, CA, USA, in 2013. He joined Qualcomm Research, San Diego, CA, USA, in 2013 where his research involves architectures and DSP algorithms for advanced radio technology. Dr. Khairy’s research interests include wireless communication, low power design of digital communication, digital signal processing, algorithm design, and VLSI architecture for wireless communication systems.

[23] Fadi Kurdahi (F’05) received his Ph.D. from the University of Southern California, Los Angeles, CA, USA, in 1987. Since then, he has been a faculty at the Department of Electrical and Computer Engineering at University of California, Irvine, CA, USA (UCI), where he conducts research in the areas of system level simulation, reconfigurable computing and cyber-physical systems, and serves as the Director for the Center for Embedded Computer Systems (CECS). His research was disseminated in about 200 publications and has received numerous accolades including the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS best paper award in 2001 and the ISQED best paper award in 2006. He served on several journal boards and conference committees. He received the Distinguished Alumnus award from his alma mater, the American University of Beirut in 2008. He is a Fellow of the AAAS.