Exploiting Temporal Correlation in Wireless Channel for Energy-Efficient Communication

Priyadarshi Mukherjee, Deepak Mishra, and Swades De

Abstract—Widespread adoption of Internet-of-Things (IoT) technology depends on cost-affordability of the devices and convenience of their usage in terms of long life. Energy-efficiency of the miniaturized wireless IoT devices is a key to the cost and convenience factors. While there have been a few prior studies on wireless channel adaptive communication strategies, in this work we relook at the problem aiming at effectively characterizing the wireless fading channel and devising energy-efficient link-layer retransmission strategy suitable for IoT devices. In particular, we propose a new channel-adaptive strategy that depends on the rate of variation of wireless channel irrespective of the underlying channel fading distribution. We further show that the concept of average fade duration dependent retransmission is a special case of the proposed strategy. Extensive simulations show that the proposed scheme most effectively characterizes the temporal variations of wireless channel in comparison with the other existing schemes. Performance of the proposed retransmission strategy is compared with the competitive protocols using the Markov model. Numerical results demonstrate that the proposed scheme offers a gain of about 9% in terms of data throughput and about 12% in terms of energy efficiency in comparison to its nearest existing benchmark scheme.

Index Terms—Fading channel, energy-aware communication, automatic repeat request, high probability analysis, optimal waiting interval, energy-efficiency optimization

I. INTRODUCTION AND MOTIVATION

In the recent years, there have been explosive growth of high throughput intensive and energy hungry applications in wireless communication. As estimated by Cisco, global Internet traffic will reach 21 GB per capita by 2020 from 7 GB per capita in 2015 [1]. With this meteoric increase of Internet traffic in the coming years, energy consumption of the wireless networks is also expected to increase sharply. But sometimes this large demand for energy is unaffordable in energy-constrained scenarios such as wireless sensor network (WSN) and Internet of Things (IoT) applications. As a result, energy-efficient and green communication [2], [3] have attracted a lot of recent attention in industry as well as academia. In the context of IoT communications,

1) 3GPP LTE-advanced (LTE-A) is one of the most promising technologies. It has discontinuous transmission/reception mechanism to allow devices turn off their radio interfaces and go to sleep [4].
2) Wake-up receiver [5] is a promising technology where the wake-up signal can be effectively used in optimally timing field data transmission, thereby further increasing the energy efficiency.

All types of battery-constrained applications, including the two possible scenarios described above, require energy to be used judiciously by taking into account the statistical knowledge of temporal channel variations. Various link layer and medium access control protocols have also been proposed for reduced energy consumption. However to the best of our knowledge, the aspect of temporal characterization of wireless channel has not been sufficiently studied yet in this context. Judicious channel characterization can help in estimating fading state of the channel, which is in turn helpful in deciding the right instants of data transmission. the proposed scheme optimally estimates the ‘sleep duration’ of the radio nodes and hence leads to enhancement of the energy-efficiency of the system.

With the fourth-generation wireless access (4G) technologies currently being deployed, the world is moving towards 5G wireless communication technologies that aim at simultaneously obtaining energy-efficiency as well as spectral efficiency [6]. Temporal characterization of wireless channel can be used to enhance spectral efficiency as well as energy efficiency.

We perform wireless channel characterization and apply it on our newly proposed link layer strategy that estimates the frame retransmission timings based on the current channel condition. The proposed strategy bridges the gap between fully-blind [7] and fully-instantaneous [8] schemes for determining optimal retransmission intervals. The adaptive retransmission strategy based on our developed temporal characterization results in appreciable energy saving as well as throughput enhancement. The proposed strategy is of high relevance in applications such as battery-limited WSNs and IoT communications, energy harvesting networks, and wireless powered communication networks [9], where channel state information (CSI) acquisition incurs a high cost.

A. Related Works

To deal with channel errors, classical link layer error control schemes can be divided into two broad categories, namely, automatic repeat request (ARQ) and forward error correction (FEC). In communication scenarios, such as in WSN-based telemetry and IoT-based routine monitoring applications, where delay requirements are not strict, ARQ scheme appears to be a better option. Conventional ARQ schemes, namely, Stop-and-Wait (SW), Go-Back-N (GBN), and Selective-Repeat (SR), are independent of wireless channel variation.

Lu and Chang [10] studied the above-mentioned basic ARQ protocols in the context of two types of correlated
channel errors: Markovian-
 channel errors: Markovian-
 k errors and gap errors. Li and Zhou [11] modeled the transmission mechanism of SR protocol by a discrete-time Geom/G/1/∞ queue model. Chakraborty et al. studied adaptive GBN protocol in [12], where temporal variation of wireless channel was characterized by Gilbert-Elliott model. Performance of GBN as well as SR is superior compared to SW, however they also have their limitations. SW protocol is useful in short-range communication application, such as, telemetry and IoT networks, that look for protocol simplicity and high energy efficiency.

De Vuyst and Bruneel studied SW protocol in [13], where the errors occurring in the channel were taken to be correlated in time. The channel errors were modeled using a two-state Markov chain. In order to obtain higher throughput, Benelli and Garzelli [14] proposed certain modifications in the classical SW protocol to reduce the time that a transmitter spends in waiting state. However, all these link layer schemes waste energy in retransmitting blindly even when the channel condition is poor.

To obtain high link layer throughput even in presence of wireless channel fading, adaptive modulation and coding (AMC) has been a well-investigated topic [15]–[17]. In AMC, there exists a condition for not sending data payload (mode 0) when the channel is in deep fading state. Still, mode 0 transmission wastes energy when the channel is in deep fade. In order to obtain high system reliability as well as high throughput, a combination of ARQ and FEC (hybrid ARQ) is commonly used [18]–[20]. However even in these schemes, energy is wasted in blind transmissions when the channel is in deep fade.

There have been some research on channel-aware link layer retransmissions [21]–[25]. Guth and Ha [21] presented a feedback-dependent adaptive policy, in which the frame size is reduced to the next smaller level after a certain number of retransmissions, and it is increased to the next higher level after a certain number of consecutive successes. Minn et al. [22] proposed the idea of determining the channel state to be ‘good’ or ‘bad’ depending on the count of consecutive ACKs (acknowledgements) and NAKs (negative ACKs); the system state is accordingly switched between ‘low’ and ‘high’ error rate modes, respectively. Zorzi and Rao [23] introduced a probing-based retransmission scheme in which the transmitter sends short probing frames upon reception of a NAK until a probing ACK is received. However, the channel parameters did not have any direct impact in determining the probing interval. De et al. [24] proposed certain channel adaptive variants of retransmission, which reduce the chance of blind retransmissions when the channel is in deep fade state. It was shown that, with only binary feedback, the average fade duration (AFD) as the probing interval offers a higher throughput as well as energy saving, and further higher performance gain is possible by proper adjustment of probing interval with received signal strength feedback. Moon [25] proposed the idea of sensing and transmission over the channel at predetermined intervals. If the channel state is found below a threshold, it is sensed again after the channel coherence time interval.

Overall we observe that, although there have been a few prior studies on wireless channel adaptive coding and link-layer transmission strategies, effective characterization of wireless channel for its energy-efficient usage is still missing in the literature. In particular, the existing schemes do not take a generic approach of accounting the channel state dependence in link layer strategies that would work irrespective of the fading distribution. We argue that, accounting temporal variation of the channel in order to decide the optimum channel usage would help increase channel utilization as well as energy efficiency. Our current work is intended to explore this generic approach of channel state prediction for improved system performance.

B. Contributions

In this work, with an aim towards improved channel utilization and energy efficiency, we relook at the channel adaptive SW-ARQ protocol that was studied in [23], [24], and [25]. We first characterize the temporal variation of wireless fading channel. Based on this characterization, we propose a new link layer frame retransmission strategy that aims to maximize data throughput and energy efficiency. The key contributions of our work are as follows:

1) Based on current strength and long-term channel statistics, an analytical model is developed for calculating probability of the signal strength staying below a given threshold over a given time duration.

2) An optimization problem is formulated for calculating the waiting interval before next transmission. Accordingly, a fast algorithm is proposed to find the optimal interval.

3) Numerical and simulation results demonstrate that the estimated optimum waiting interval efficiently characterizes temporal variation of the wireless channel. It is also shown that the optimum waiting interval reduces to AFD as a special case.

4) Based on the optimum waiting estimate, a variant of the SW-ARQ protocol is proposed. The simulation results demonstrate that the data throughput and energy efficiency of the proposed scheme is respectively about 9% and 12% higher in comparison to its nearest approach.

C. Paper Organization

The remaining paper is organized as follows. In Section II the system model and proposed protocol operations are described. Section III first introduces the concept of time derivative of the signal, which is followed by the probabilistic analysis and its generalization. In Section IV the optimum waiting interval for next transmission is determined, and convergence with AFD as a special case is studied. Section V deals with the proposed protocol performance analysis by formulating an optimization problem that maximizes the energy efficiency. Section VI contains the numerical results, and finally Section VII concludes the paper.

II. SYSTEM MODEL AND PROTOCOL DESCRIPTION

A. System Model

We consider link layer communication between a node pair in a mobile environment. Table I shows the list of variables
TABLE I
LIST OF VARIABLES USED AND THEIR RESPECTIVE DESCRIPTIONS

<table>
<thead>
<tr>
<th>$T_p$</th>
<th>Slot duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Upsilon_f$</td>
<td>Frame duration</td>
</tr>
<tr>
<td>$\Upsilon_{A/N}$</td>
<td>ACK/NAK frame duration, $\Upsilon_{A/N} = q\Upsilon_f(q \ll 1)$</td>
</tr>
<tr>
<td>$f_D$</td>
<td>Doppler frequency</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Tx sends data frames in every $\beta$ slot when channel is in good state</td>
</tr>
<tr>
<td>$Z$</td>
<td>Received signal envelope or signal strength indicator (SSI)</td>
</tr>
<tr>
<td>$Z_0$</td>
<td>$Z$ at time $t$</td>
</tr>
<tr>
<td>$Z_{th}$</td>
<td>SSI threshold separating good and bad channel condition</td>
</tr>
<tr>
<td>$N_g$</td>
<td>Estimated waiting time interval in slots</td>
</tr>
<tr>
<td>$T_{gs}$</td>
<td>Estimated waiting time interval in sec, $T_{gs} = N_g T_p$ sec</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Level crossing probability</td>
</tr>
<tr>
<td>$\dot{Z}$</td>
<td>Rate of temporal variation of $Z$, $\dot{Z} \sim N(0, \sigma)$</td>
</tr>
<tr>
<td>$Z_{N}$</td>
<td>Temporal variation of $Z$ in $N^{th}$ time slot from time instant $t$</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>Number of transmissions between two consecutive estimated waiting intervals</td>
</tr>
<tr>
<td>$\nu_f$</td>
<td>Transmit and receive energy per data frame</td>
</tr>
<tr>
<td>$\nu_{ACK/NAK}$</td>
<td>Transmit and receive energy per ACK/NAK frame</td>
</tr>
<tr>
<td>$\nu_{idle}$</td>
<td>Idle energy consumption per frame</td>
</tr>
</tbody>
</table>

used, along with the descriptions. The system is considered slotted, with slot duration $T_p$ sec. Data frames are assumed always available at the transmitter (Tx) for transmission, and the channel remains invariant within a frame duration $\Upsilon_f$ but may vary from frame to frame [15].

Depending on received signal quality, the receiver (Rx) sends ACK or NAK for each received data frame. The Rx node sends a NAK if the signal strength indicator (SSI) at Rx is below a predetermined threshold $Z_{th}$. Thus, the decision of the Rx to send ACK or NAK depends on fading envelope of the underlying wireless channel.

In case of NAK, Rx also sends the useful channel information, such as Doppler frequency $f_D$; its estimation from the received signal is known from [26], [27]. There exists various ways to estimate $f_D$ at the receiver. Tepedelenlioglu et al. [26] has provided a detailed summary of all the existing $f_D$ estimation schemes present in literature, some of which are:

1) node velocity (proportional to $f_D$),
2) extracting $f_D$ from the level crossings,
3) performing autocorrelation analysis to infer $f_D$, and
4) squared deviations of the logarithmically compressed signal envelope (developed by Holtzman and Sampath [27]).

$f_D$ corresponding to relative velocity $v$ of the Rx is $f_D \approx \frac{v f_c}{c}$, where $f_c$ is the carrier frequency and $c$ is the velocity of light in vacuum. The product $f_D T_p$ signifies the temporal variation rate of the wireless channel. From [28] we know that small value of $f_D T_p \left( < 0.1 \right)$ indicates correlated ‘slow’ fading channel, whereas a large value of $f_D T_p \left( > 0.2 \right)$ implies that the two samples of the channel are almost independent, i.e., ‘fast’ fading.

Even if Tx and Rx are fixed, the received signal can still experience the Doppler effect due to the presence of scatterers [29]. This notion of the Tx, Rx, and scatterers has been equivalently summed up by the term ‘node velocity’.

Node velocity does not only mean that the Tx or Rx node is moving. It also means that both Tx and Rx may be stationary, but the scatterers present along the path are in motion, which is a relevant scenario in today’s and upcoming wireless IoT communication scenarios.

It can be noted that, in case of applications like IoT, with statically placed nodes these scatterers are the main sources of Doppler effect. Even in presence of very slow moving scatterers, like foliage movement or walking pedestrians, the proposed $T_{gs}$ estimation model works. In fact, system performance benefit from the proposed estimation approach is even higher in these scenarios, as they actually represent highly correlated channel conditions. This makes the proposed $T_{gs}$ estimation scheme even more useful in applications like IoT.

B. Protocol Description

Based on the information obtained from NAK, we propose a channel adaptive SW-ARQ protocol, which we call dynamic SW (D-SW). The Tx node sends data frames in every $\beta$ slot ($\beta \geq 1$) when the channel is in ‘good’ state. Upon receipt of a NAK, the Tx chooses to wait for an interval before next transmission. The waiting interval is determined based on $f_D$ and channel state estimate received over the NAK.

Clearly, in the proposed D-SW protocol there is no periodicity associated with the estimated waiting interval; it is solely based on the current state of the channel. For the optimal waiting interval estimation we characterize temporal variation of the signal received at Rx, which is discussed next.

D-SW estimates the waiting interval more accurately compared to the improved SW-ARQ, as demonstrated in Section VI. Hence, D-SW does not require any probing signal as in the improved SW-ARQ [24] to check the channel status after a waiting interval; it directly resumes data transmission after the waiting interval. This results in significant energy saving per successful data frame transmitted, which is eventually reflected
in terms of higher energy efficiency of D-SW in comparison to the improved SW-ARQ protocol.

Without loss of generality, we take data frame length \( \Upsilon_f \) to be of single slot, i.e., \( \Upsilon_f = T_p \). This assumption is based on the block fading scenario considered in this work [15], [23]. ACK/NACK frame durations \( \Upsilon_{\text{ACK/N}} \) are assumed small compared to \( \Upsilon_f \), i.e., \( \Upsilon_{\text{ACK/N}} = \varrho \Upsilon_f \), where \( \varrho \ll 1 \). This assumption is based on the standard practice that ACK/NACK frames are frames with very little payload [23], [24]. We also note that ACK/NACK frames consume energy and hence reduction in their frequency is desirable [23].

III. PROBABILISTIC ANALYSIS OF SSI

We introduce the concept of time derivative of fading envelope of a wireless channel in this section, where we calculate the probability of SSI staying below a given threshold.

A. Time Derivative of the Fading Envelope

We consider the communication system with simple receiver structure, that uses noncoherent demodulation. The complex received signal at Rx is \( S(t) = S_j(t) + jS_Q(t) \), where \( j = \sqrt{-1} \) and \( S_j(t) \) and \( S_Q(t) \) are the mutually independent quadrature components with identical variance. At a time instant \( t \), the received signal envelope or SSI at Rx is \( Z(t) = \sqrt{S_j^2(t) + S_Q^2(t)} \). The distribution of \( Z(t) \) depends on the underlying fading distribution. With \( f_Z(z) \) denoting the probability distribution of \( Z \), we use the time derivative of \( Z \) in our analysis, i.e., \( \dot{Z}(t) = \frac{dZ(t)}{dt} \) at \( Z(t + \Delta t) - Z(t) \). \( \dot{Z}(t) \) signifies the rate of change of \( Z(t) \) with time.

To the best of our knowledge, till date the information about \( \dot{Z} \) is used for stochastically characterizing the wireless channel [31] to obtain its two key second-order statistics, namely, level crossing rate (LCR) and average fade duration (AFD). Mathematically, \( \text{LCR} = \int_0^\infty \dot{Z} f_Z(Z, \dot{Z}) d\dot{Z} \) and \( \text{AFD} = \frac{\text{Pr}\{Z < Z_{th}\}}{\text{LCR}} \), where \( Z_{th} \) is the target threshold and \( f_Z, Z(z, \dot{Z}) \) is the joint probability density function of \( Z(t) \) and \( \dot{Z}(t) \) at the same time instant.

From [32], it is noted that \( Z \) (index \( t \) is removed for brevity) is a zero mean Gaussian random variable (RV) irrespective of the underlying fading distribution of the wireless channel, i.e., \( Z \sim \mathcal{N}(0, \sigma) \).

Since our time derivative analysis and subsequent waiting interval estimation only requires \( f_Z(\dot{Z}) \), this approach is general, i.e., valid for any fading distribution.

B. Probability of SSI Staying below a Threshold

Suppose SSI at the Tx at time \( t \) is \( Z(t) = Z_0 \) and it is less than the threshold \( Z_{th} \). The waiting interval \( T_{gs} \) is estimated using the value of \( Z_0 \). If the estimated interval is \( N_g \) slots, then the actual interval is \( T_{gs} = N_g T_p \) sec. It may be noted that, \( T_{gs} \) thus calculated is neither an instantaneous nor a mean, but a probabilistic value, calculated on the basis of a level crossing probability \( \epsilon \) which the underlying application demands. \( T_{gs} \) being probabilistic in nature, \( \epsilon \) plays a pivotal role in its calculation. The lesser the value of \( \epsilon \), the more sure we are about the estimated \( T_{gs} \) for a given SSI.

Conversely, it can be stated that, given \( Z_0 < Z_{th} \) we are interested in finding out for how long the channel will continue to remain in ‘bad’ state (\( Z < Z_{th} \)) before reaching ‘good’ state (\( Z \geq Z_{th} \)).

Using first order approximation of the Taylor series expansion of \( Z(t + T_p) \), where \( T_p \ll 1 \), we get

\[
Z(t + T_p) = Z(t) + \dot{Z} \cdot T_p + \ddot{Z} \cdot T_p^2 \cdot \frac{1}{2!} + \cdots
\]

\[
= Z(t) + \dot{Z} \cdot T_p + O(T_p^2)
\]

\[
\approx Z(t) + \dot{Z} \cdot T_p.
\]

(1)

Here \( \dot{Z} \) (a RV) denotes the rate of temporal variation of the signal envelope. Hence, the probability that \( Z \) will not reach \( Z_{th} \) in the next slot is:

\[
\text{Pr}\{Z(t + T_p) < Z_{th}\} = \text{Pr}\{Z(t) + \dot{Z} \cdot T_p < Z_{th}\}
\]

\[
= \text{Pr}\{Z_0 + Z_1 < Z_{th}\}
\]

(2)

Accordingly, \( Z_1 = \dot{Z} \cdot T_p \) (also a RV) denotes the temporal variation of signal envelope in next one slot. Thus we have \( Z_1 \sim \mathcal{N}(0, \sigma_1) \), where \( \sigma_1 = T_p \sigma \) (from Section III-A). \( Z_1 \sim \mathcal{N}(0, \sigma_1) \) theoretically means \( Z_1 \in (-\infty, \infty) \). However, as the SSI \( Z_0 + Z_1 \geq 0 \), \( Z_1 \in (-Z_0, \infty) \), i.e., \( Z_1 \) does not truly follow \( \mathcal{N}(0, \sigma_1) \); rather it is a truncated Gaussian distribution, with distribution function expressed as:

\[
f_{Z_1}(z_1) = \begin{cases} \frac{1}{\sqrt{2\pi} \sigma_1} e^{-\frac{z_1^2}{2\sigma_1^2}} & -Z_0 \leq z_1 \\ 0 & \text{elsewhere} \end{cases}
\]

(3)

Here \( \Phi_1(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \) is the cumulative distribution function of standard univariate normal distribution. Value of \( \sigma_1 \) depends upon the fading distribution model.

Using (3), (2) can be rewritten as:

\[
\text{Pr}\{Z_0 + Z_1 < Z_{th}\} = \int_{-\infty}^{z_{th} - Z_0} f_{Z_1}(x) dx
\]

\[
= \frac{\Phi_1 \left( \frac{Z_{th} - Z_0}{\sigma_1} \right) - \Phi_1 \left( -\frac{Z_0}{\sigma_1} \right)}{1 - \Phi_1 \left( -\frac{Z_0}{\sigma_1} \right)}.
\]

(5)

Similarly, probability that \( Z \) will continue to remain below \( Z_{th} \) in the next two time slots is:

\[
\text{Pr}\{Z(t + T_p) < Z_{th}, Z(t + 2T_p) < Z_{th}\} \approx \text{Pr}\{Z_0 + Z_1 < Z_{th}, Z_0 + Z_2 < Z_{th}\}
\]

(6)

where \( Z_2 \) is also a zero mean truncated Gaussian RV like \( Z_1 \) with \( \sigma_2^2 = 2\sigma_1^2 \), just the difference being that \( Z_2 \) denotes the temporal variation of the fading envelope over the next two slots, whereas \( Z_1 \) denotes temporal variation over next one slot only.

In order to obtain (6), we require the joint probability
distribution $f_{Z_1, Z_2}(z_1, z_2)$ to be integrated over suitable limits, i.e.,

$$\Pr \{ Z_0 + Z_1 < Z_{th}, Z_0 + Z_2 < Z_{th} \}$$

using (6) first probability that
Hence proceeding in the same way as in previous subsection,

$$\Pr \{ Z_0 + Z_1 < Z_{th}, Z_0 + Z_2 < Z_{th} \} = \int_{z_1} \int_{z_2} f_{Z_1, Z_2}(z_1, z_2)dz_1dz_2$$

where $f_{Z_1, Z_2}(z_1, z_2)$ is the truncated bivariate Gaussian distribution (4) given by [33].

Here $\rho_{12}$ is the correlation between $Z_1$ and $Z_2$, which we assume known from the feedback sent to the controller by the ARQ module, $\Theta = 1 - \Phi_2(\infty, -\frac{Z_0}{\sigma_1}) - \Phi_2(\frac{Z_0}{\sigma_1}, \infty) + \Phi_2(\frac{Z_0}{\sigma_1} - \frac{Z_0}{\sigma_2}),$ and $\Phi_2(a, b) = \frac{1}{2\pi\rho_{12}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(t_1^2+2\rho_{12}t_1t_2+t_2^2)} dt_1dt_2$ is the joint cumulative distribution function of standard bivariate normal distribution.

We obtain (7) using the algorithm defined in [34] for computation of multivariate truncated Gaussian probability.

C. Generalization of Probability of SSI Staying below a Threshold for Next N Slots

Generalizing the concept over next $N$ slots, we have, $\hat{\sigma}_N^2 = N\sigma_t^2$. In general, $Z_1 \neq Z_2$ and hence

$$\Pr \{ Z_0 + Z_1 < Z_{th}, Z_0 + Z_2 < Z_{th} \} \neq \Pr \{ Z_0 + Z_1 < Z_{th} \} \Pr \{ Z_0 + Z_2 < Z_{th} \}.$$  

(8)

Hence proceeding in the same way as in previous subsection, probability that $Z$ will not reach $Z_{th}$ or above in none of the first $N$ slots is:

$$\Pr \{ Z(t+T_p) < Z_{th}, \cdots, Z(t+NT_p) < Z_{th} \} \approx \prod_{i=0}^{N-1} \Pr \{ Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th} \}$$

using (6) with $f_{Z_R}(z_R, \Sigma, Z_{R0})$ being the $N$-variate truncated Gaussian distribution [35] given by:

$$f_{Z_R}(z_R, \Sigma, Z_{R0}) = \begin{cases} \frac{1}{\sqrt{2\pi\Sigma_{12}^2}} e^{-\frac{1}{2}(z_R^2 - 2\rho_{12}z_Rz_{R0} + z_{R0}^2)} & \text{if } Z_R \in \mathbb{R}_{\geq 0}^N \setminus \{ z_R \} \\ 0 & \text{elsewhere} \end{cases}$$

(10)

where $Z_R = [Z_1, Z_2, \cdots, Z_N]^T$ is an $N$-dimensional random variable, $\Sigma$ is the $N \times N$ positive definite covariance matrix of $Z_R$, $Z_{R0} = -Z_0[1, 1, \cdots, 1]^T$, $\mathbb{R}_{\geq 0}^N \setminus \{ z_R \}$ is the rate of temporal variation of the channel, desired threshold $Z_{th}$ is reached in finite time. In other words, irrespective of velocity $v$ we always have

$$\lim_{N \to \infty} \Pr \{ Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th} \} = 0.$$  

(11)

Proof: By the chain rule of probability,

$$\lim_{N \to \infty} \Pr \{ Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th} \} = \lim_{N \to \infty} \Pr \{ Z_0 + Z_1 < Z_{th}, Z_0 + Z_2 < Z_{th}, \cdots, Z_0 + Z_{N-1} < Z_{th}, Z_0 + Z_N < Z_{th} \}$$

(12)

$$= \lim_{N \to \infty} \Pr \{ Z_0 + Z_N < Z_{th} \} \times \Pr \{ Z_0 + Z_{N-1} < Z_{th} | Z_0 + Z_N < Z_{th} \} \times \Pr \{ Z_0 + Z_{N-2} < Z_{th} | Z_0 + Z_{N-1} < Z_{th}, \cdots, Z_0 + Z_N < Z_{th} \}.$$  

(13)

Now, $\lim_{N \to \infty} \Pr \{ Z_0 + Z_N < Z_{th} \} < 1$ as $Z_N \sim N(0, \sigma_N^2)$ and $\lim_{N \to \infty} \hat{\sigma}_N^2 = \lim_{N \to \infty} N\sigma_t^2 = \infty$. Likewise,

$$\lim_{N \to \infty} \Pr \{ Z_0 + Z_{N-1} < Z_{th} | Z_0 + Z_N < Z_{th} \} \approx \lim_{N \to \infty} \Pr \{ Z_0 + Z_N < Z_{th} \} < 1.$$  

(14)

Proceeding similarly we can write,

$$\lim_{N \to \infty} \Pr \{ Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th} \} = \lim_{N \to \infty} \prod_{i=1}^{N} a_i = 0,$$

where

$$a_i < 1 \text{ and } x \triangleq \min \{ k | \Pr \{ Z_0 + Z_k < Z_{th} \} < 1 \wedge k \in \mathbb{Z}^+ \}.$$  

Hence the proof.

This proof of concept has been demonstrated graphically in Fig. 1. In this figure, we show the variation of the probability with increasing value of $N$ when an unit power signal is transmitted over a Rayleigh fading channel with $Z_{th} = \ln 2$, $Z_0 = 0.01$, and $T_p = 500$ $\mu$s. The rate of convergence of the probability is a function of velocity $v$. For instance, the rate of convergence at $v = 50$ kmph is faster than that at $v = 10$ kmph. We use the above observations in the next section, where we propose a novel scheme for estimating waiting interval $N_g$. 

□
We define the lower bound of $Z$ for a given set of $N$ elements described in Section III-C. The smaller the value of $N$, the more slots that can be avoided for sensing consideration, because $Z$ is very much application specific. Therefore, we intend to find the maximum number of sensing. Hence, for a given $N$, we are interested in estimating the maximum possible value of $N_g$.

Our main focus is on the aspect of energy efficiency, i.e., we are interested in saving the energy spent in unnecessary sensing. Therefore, we intend to find the maximum number of slots that can be avoided for sensing consideration, because with very high probability $1 - \epsilon$ the channel will continue to be in the $Z < Z_{th}$ region in each of the next $N$ slots given that it is currently in the $Z < Z_{th}$ region. Hence for a given set of system parameters ($f_D$ and $T_p$), $Z_{th}$, $Z$, and $\epsilon$, we calculate $N_{g^*}$ (optimal value of $N_g$) by solving the optimization problem P1 in (15).

\[
\Pr \{ Z_0 + Z_1 < Z_{th}, \ldots , Z_0 + Z_{N_g} < Z_{th} \} \geq 1 - \epsilon.
\]

(P1) : maximize $N_g$ \[N_g \geq 0\] 
subject to \[\Pr \{ Z_0 + Z_1 < Z_{th}, \ldots , Z_0 + Z_{N_g} < Z_{th} \} \geq 1 - \epsilon.\]

Fig. 1. Variation of not reaching probability versus $N$. 

IV. PROPOSED SCHEME

In this section, we use high probability analysis [36] as a tool and propose a waiting interval estimation scheme based on the analysis presented in Section III. The estimated optimum waiting interval (OWI) is used in D-SW protocol operation.

A. Waiting Interval Estimation Scheme

Our main motivation is to estimate the waiting interval $N_g$ from the SSI $Z$ by taking into account the rate of temporal variation of the channel. As noted in Fig. 1, $\Pr \{ Z_0 + Z_1 < Z_{th}, \ldots , Z_0 + Z_{N_g} < Z_{th} \} \geq 1 - \epsilon$, is a strictly decreasing function of $N_g$. Hence, for a given level crossing probability $\epsilon$, we are interested in estimating the maximum possible value of $N_g$.

Our main focus is on the aspect of energy efficiency, i.e., we are interested in saving the energy spent in unnecessary sensing. Therefore, we intend to find the maximum number of slots that can be avoided for sensing consideration, because with very high probability $1 - \epsilon$ the channel will continue to be in the $Z < Z_{th}$ region in each of the next $N$ slots given that it is currently in the $Z < Z_{th}$ region. Hence for a given set of system parameters ($f_D$ and $T_p$), $Z_{th}$, $Z$, and $\epsilon$, we calculate $N_{g^*}$ (optimal value of $N_g$) by solving the optimization problem P1 in (15).

\[
\Pr \{ Z_0 + Z_1 < Z_{th}, \ldots , Z_0 + Z_{N_g} < Z_{th} \} \geq 1 - \epsilon.
\]

(P2) : maximize $N_g$ \[N_g \geq 0\] 
subject to \[\prod_{i=1}^{N_g} \frac{\Phi_1 \left( \frac{Z_{th} - Z_i}{\sigma_i} \right)}{1 - \Phi_1 \left( -\frac{Z_i}{\sigma_i} \right)} \geq 1 - \epsilon.\]

2) $N_{g^{ub}}$. Given $Z_{th}$, $N_{g^{ub}}$ is calculated by estimating the probability that $Z$ crosses $Z_{th}$ in the $N_g$th slot irrespective of whether it had crossed $Z_{th}$ before or not. Accordingly we obtain $N_{g^{ub}}$ by solving the following optimization problem:

(P3) : maximize $N_g$ \[N_g \geq 0\] 
subject to \[\Phi_1 \left( \frac{Z_{th} - Z_i}{\sigma_i} \right) \geq 1 - \epsilon.\]

The importance of $N_{g^{lb}}$ and $N_{g^{ub}}$ is that they allow us to reformulate (15) into (19), an optimization problem with an unimodal objective function. Utilizing this unimodal nature of the objective function we propose Algorithm 1 of minimal complexity to estimate $N_{g^*}$, whose convergence is discussed next.

With $N_{g^{lb}}$ and $N_{g^{ub}}$ we reformulate (15) to obtain $N_{g^*}$ as:

(P4) : $N_{g^*} = \left\{ N_g \right\} \arg \min_{N_{g^{lb}} \leq N_g \leq N_{g^{ub}}} \left\{ \Pr \{ Z_0 + Z_1 < Z_{th}, \ldots , Z_0 + Z_{N_g} < Z_{th} \} - (1 - \epsilon)^2 \right\}.$

We propose Algorithm 1 based on golden section based line search [37] to solve (19).

Solutions of problems P2 and P3 are obtained by solving their respective constraints at equality for $N_g$. P2 involves product of multiple $\Phi_1(\cdot)$, whereas P3 requires a single $\Phi_1(\cdot)$ to be calculated. It is important to note that P1 can also be solved likewise without the use of any algorithm. However, the computational complexity will be huge as it requires integration over arbitrary $N$-variate truncated Gaussian distributions, with $N$ increasing in each step.

On the other hand, solving P2 or P3 do not require such
Algorithm 1 Algorithm to find $N_g^*$

Require: $Z_{th}$, $Z_0$, $f_D$, $T_p$, $\epsilon$, and $\xi > 0$
Ensure: $N_g^*$

1: Calculate $N_g^{lb}$ using (17) and $N_g^{ub}$ using (18)
2: Define $r(N_g) = |\text{Pr}\{Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th}\} - (1 - \epsilon)|^2$
3: Set $k = 0$
4: Calculate $N_g = N_g^{ub} - 0.618 \times (N_g^{ub} - N_g^{lb})$
5: Calculate $N_g = N_g^{lb} + 0.618 \times (N_g^{ub} - N_g^{lb})$
6: Calculate $r(N_p)$ and $r(N_q)$
7: Set $\Delta_g = N_g^{ub} - N_g^{lb}$
8: while $\Delta_g > \xi$ do
9: if $r(N_p) \leq r(N_q)$ then
10: Set $N_g^{lb} = N_q, N_g = N_p$, and $N_g = N_g^{ub} - 0.618 \times (N_g^{lb} - N_g^{ub})$
11: else
12: Set $N_g^{ub} = N_p, N_g = N_q$, and $N_g = N_g^{lb} + 0.618 \times (N_g^{ub} - N_g^{lb})$
13: end if
14: Set $k = k + 1$
15: end while
16: Calculate $r(N_g)$ and $r(N_q)$
17: Set $\Delta_g = N_g^{ub} - N_g^{lb}$
18: if $r(N_p) < r(N_q)$ then
19: Set $N_g^* = N_p$
20: else
21: Set $N_g^* = N_q$
22: end if

Fig. 2. Effect of variation of system parameters on $N_g^*$.

complex integration. Specifically, for a given set of system parameters, P2 and P3 need to be solved just once to obtain the two bounds that are required for the algorithm. Once they are calculated, we obtain $N_g^*$ using Algorithm 1.

Fig. 2 shows the effect of system parameters on $N_g^{ub}, N_g^*$,

and $N_g^{lb}$. Taking $T_p = 500 \mu s$, and a suitable $Z_{th} = \ln 2$, we consider the transmission of a unit power signal in Rayleigh fading environment. Fig. 2(a) shows the joint effect of $Z_0$ and node velocity on $N_g^*$.

We can see that for some $Z_0, N_g^*$ acquires large value for lower node velocity and vice-versa; for $Z_0 = -100$ dBm, $N_g^* = 68$ slots when $v = 5$ kmph compared to $N_g^* = 6$ slots when $v = 60$ kmph. On the other hand, Fig. 2(b) shows the effect of node velocity on $N_g^{ub}, N_g^*$, and $N_g^{lb}$. We can observe from Fig. 2(b) that the lower bound $N_g^{lb}$ is relatively a tighter bound compared to $N_g^{ub}$.

Increasing $v$ implies increasing rate of variation of the channel, which is inversely proportional to $N_g^*$. Also, from [32], irrespective of the underlying fading distribution, $Z \sim \mathcal{N}(0, \sigma)$, where $\sigma \propto f_D$. Thus, increase in $v$ is reflected in $f_D$, which results in a proportional increase of $\sigma$. This can also be observed from Fig. 2(a) and Fig. 2(b), which indicate that for a given $Z_0, N_g^*$ decreases with increase in node velocity $v$.

Hence we conclude that, for a given ‘bad’ channel state the corresponding estimated waiting interval is larger for scenarios with low $f_D T_p$, and it decreases with increasing $f_D T_p$. While Fig. 2(a) shows this observation as $v$ and $Z_0$ are varied simultaneously, Fig. 2(b) shows that this observation holds true even for the two bounds $N_g^{lb}$ and $N_g^{ub}$.

B. Convergence of Algorithm 1

Here we discuss convergence of the proposed algorithm as follows:

We know from Lemma 1 that $\Pr\{Z(t + T_p) < Z_{th}, \cdots, Z(t + N_g T_p) < Z_{th}\}$ is a monotonically decreasing function of $N_g$. Thus, noting that $\epsilon$ is a non-negative quantity, the objective function of P4, i.e., $r(N_g) = [\Pr\{Z_0 + Z_1 < Z_{th}, \cdots, Z_0 + Z_N < Z_{th}\} - (1 - \epsilon)]^2$ is unimodal in nature with the optimum achieved when $1 - \epsilon = \Pr\{Z(t + T_p) < Z_{th}, \cdots, Z(t + N_g T_p) < Z_{th}\}$. Hence in this context, the local minima is itself the global minima, and as a result the golden section based line search [37] can be used.

Since $r(N_g)$ is unimodal in $[N_g^{lb}, N_g^{ub}]$, the algorithm based on golden section search converges in finite number of iterations with the search space reducing in every iteration in such a way that it always contains the optimal $N_g^*$. This allows to obtain an upper bound [37] on the required number of iterations till convergence.

The algorithm terminates after $N$ iterations if

$\left(\frac{N_g^{ub} - N_g^{lb}}{0.618^N}\right) \leq \xi$, where $\xi$ is the tolerance level. In other words, it can be stated that for a given set of $N_g^{ub}, N_g^{lb}$, and $\xi$,

$$N \leq \frac{\ln(\frac{\xi}{N_g^{lb} - N_g^{ub}})}{\ln(0.618)} \quad (20)$$

$$= \frac{-1}{\ln(0.618)} \ln \left(\frac{N_g^{ub} - N_g^{lb}}{\xi}\right) \approx 2 \ln \left(\frac{N_g^{ub} - N_g^{lb}}{\xi}\right).$$

Hence, fast convergence of the proposed algorithm can be guaranteed. Moreover it can be further observed from the
above inequality that, besides $\xi$, convergence of the algorithm also depends on $(N^u_g - N^l_g)$.

Taking $\epsilon = 0.1$, $T_p = 500 \mu$s, a suitable $Z_{th} = \ln 2$, and tolerance level $\xi = 0.01$, we consider the following case in a unit power signal transmission scenario in Rayleigh fading environment: $Z_0 \ll Z_{th}$ (deep fading state) and $v = 5$ kmph.

For the stated system parameters, we can see from Fig. 3 that the number of iterations is 14, which is a small value. It can be further observed that the optimal solution is very close to $N^l_g$ as compared to $N^u_g$, which signifies that the lower bound $N^l_g$ is tight. It can also be noted from Fig. 3 that convergence is faster in the initial stages as the gap between $N_g$ and $N^u_g$ reduces very sharply in the first few iterations.

C. Convergence with AFD

In this sub-section we prove that AFD is a special case of OWI.

Lemma 2. OWI reduces to AFD with $\epsilon = 0.5$.

Proof: From (19) we know that, if $Z_0$ is the initial SSI for a given set of system parameters, then SSI $Z$ crosses the threshold $Z_{th}$ in $N^u_g$ slots with probability $\epsilon$. This implies that $Z$ stays below $Z_{th}$ for the next $N^u_g$ time slots with probability $(1-\epsilon)$, i.e., it crosses $Z_{th}$ in the next $N^l_g$ time slots with a finite probability of $\epsilon$. AFD on the other hand defines the average time that $Z$ stays below a certain threshold $Z_{th}$, irrespective of the constant instantaneous SSI value $Z_0$. AFD $\gamma_{Z_{th}}$ is defined [31] as the average time $Z$ stays below $Z_{th}$. It is expressed as:

$$\gamma_{Z_{th}} = \Pr\left\{Z < Z_{th}\right\} \int_0^{Z_{th}} f_{Z,Z}(z, \dot{z}) dz,$$

where $f_{Z,Z}(z, \dot{z})$ is the joint probability distribution of $Z$ and $\dot{Z}$. AFD defined in terms of slots is $\gamma_{\text{slots}} = \frac{\gamma_{Z_{th}}}{T_p}$.

To relate OWI with AFD, we first define the notion of $Z_0$ in the context of AFD. As the definition of AFD is about average time, we need to consider all the initial starting points $Z_0$ that are below $Z_{th}$ and the time duration over which the system stays below $Z_{th}$. We define a RV $Z_A$ as, $Z_A = \{Z : 0 \leq Z < Z_{th}\}$ and substitute $Z_0 = \mathbb{E}[Z_A]$. In this way we take the average value of all the possible initial starting points in the region $0 \leq Z < Z_{th}$. It is important to note that $Z_A$ does not have the same probability distribution as $Z$, but a truncated version of $f_Z(z)$ in the range $[0, Z_{th})$, i.e.,

$$f_{Z_A}(z) = \frac{f_Z(z)}{\int_0^{Z_{th}} f_Z(z) dz}, \text{ where } 0 \leq z < Z_{th}. \quad (22)$$

Suppose we have $N^*_g = N^*_g$ if $\epsilon = 0.5$ and $Z_0$ as defined above. This implies that, in 50% of the next $N^*_g$ slots $Z$ will remain below $Z_{th}$ and 50% of the time above it. From Section III-B we know that $Z_0 + i \cdot Z_{th} \forall i = 1(1)N^*_g$ are Gaussian RVs. This implies as if $Z_0$ is taking many steps, but each of small magnitude in a time interval of $N^*_g T_p$ sec. As $T_p \to 0$, these small steps become a continuum and this entire process converges to a process with continuous sample path, i.e., “Brownian motion” (Donsker’s theorem, [38]). According to the property of Brownian motion, it follows a normal distribution [38].

With $\epsilon = 0.5$, i.e., level crossing probability of 0.5, $N^*_g = N^*_g$ is the median of the process stated above. By the property of normal distribution, mean and median are the same. Hence, the mean of the above process with $\epsilon = 0.5$ is $N^*_g$ slots, i.e., the average duration that $Z$ stays below $Z_{th}$ is $N^*_g T_p$ sec. Thus, for $\epsilon = 0.5$, estimated waiting interval converges to AFD, i.e., $N^*_g T_p = \text{AFD}$.

To reinforce the above claim, we reformulate (19) as:

$$(\text{P5}) : \epsilon_{\text{AFD}} = \left\{ \epsilon \mid \arg \min_{0 < z < 1} \left\{ \mathbb{P}\{Z_0 + Z_1 < Z_{th}, \ldots, Z_0 + \gamma_{\text{slots}} < Z_{th}\} - (1 - \epsilon)^2 \right\} \right\},$$

where $\gamma_{\text{slots}} = \frac{\epsilon_{\text{AFD}} T_p}{T_p}$. It can be observed that the objective function of (23) is unimodal in nature, with the minimum achieved when $\mathbb{P}\{Z_0 + Z_1 < Z_{th}, \ldots, Z_0 + \gamma_{\text{slots}} < Z_{th}\} = 1 - \epsilon$. Accordingly we use the golden section based line search [37] technique to obtain $\epsilon_{\text{AFD}}$.

Though the claim holds for all fading distributions, results corresponding to Rayleigh fading scenario are presented here. If $Z$ is Rayleigh distributed, then $Z_A$ is a truncated Rayleigh RV. Accordingly using (22) we get:

$$Z_0 = \mathbb{E}[Z_A] = \frac{1}{1 - e^{-\frac{Z_{th}^2}{2\kappa}}} \left[ \sqrt{\frac{\pi}{2\kappa}} \text{erf}\left(\frac{Z_{th}}{\sqrt{2\kappa}}\right) - Z_{th} e^{-\frac{Z_{th}^2}{2\kappa}} \right]. \quad (24)$$

Here $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, $\kappa$ is the scale parameter of the Rayleigh distribution, and $2\kappa^2$ denotes the average received power $\Omega$ at the Rx. From Section III-A we know that, in case of Rayleigh fading scenario $\hat{\sigma} = \pi f_D \sqrt{\Omega}$. Hence we can conclude that for Rayleigh fading channel $\hat{\sigma} = \sqrt{2\pi} f_D \kappa$, as $\Omega = \mathbb{E}[Z^2] = 2\kappa^2$.

Fig. 4(a) shows the variation of $\left(\epsilon_{\text{AFD}} - 0.5\right)$ versus node velocity when a unit power signal is transmitted over a Rayleigh fading channel with slot duration $T_p = 100 \mu$s and $Z_{th} = 0.52$. Similarly Fig. 4(b) shows the effect of variation of $Z_{th}$ on $(\epsilon_{\text{AFD}} - 0.5)$ for slot duration $T_p = 100 \mu$s and velocity $= 30$ kmph. In both these figures it is observed that
\[ \text{Residuals } \epsilon_0 = 0.5 \]

\[ \text{transition matrix probabilities are expressed as (40):} \]

\[ \epsilon_0 \text{ problem that maximizes the energy efficiency for a given set and energy efficiency. We then formulate an optimization optimum next transmission instant. The performance of D-SW slots, where } Z \text{ currently below outage duration estimation scheme. Hence OWI estimation can also be seen as a probabilistic estimation for which the channel will continue to remain in outage. }

Remark 1. OWI estimation scheme probabilistically estimates the time that } Z \text{ will continue to stay below } Z_{th} \text{ given that it is currently below } Z_{th}. \text{ In other words, given the channel is in outage, OWI estimation probabilistically estimates the duration for which the channel will continue to remain in outage. Hence OWI estimation can also be seen as a probabilistic outage duration estimation scheme.}

V. PERFORMANCE ANALYSIS OF D-SW

The estimated OWI is used in D-SW to determine the optimum next transmission instant. The performance of D-SW is quantified in terms of data throughput, energy consumption, and energy efficiency. We then formulate an optimization problem that maximizes the energy efficiency for a given set of parameters by calculating the corresponding optimal \( \epsilon. \)

In order to determine the performance, the channel is modeled as a two state Markov process (Fig. 5) with 1-step transition matrix \( M = \begin{bmatrix} p_{11} & p_{10} \\ p_{01} & p_{00} \end{bmatrix} \) [28], [39].

Tx sending data frames every \( \beta \) slots, the \( \beta \)-step transition probabilities are expressed as [40]:

\[ p_{11}(\beta) = \frac{p_{01} + (1 - \delta)^\beta}{\delta} p_{10} \quad \text{and} \quad (25) \]

\[ p_{01}(\beta) = \frac{p_{01}[1 - (1 - \delta)^\beta]}{\delta}, \text{ where } \delta = p_{01} + p_{10}. \]

If there are \( \zeta \) consecutive data transmissions in every \( \beta \) slots, where \( \zeta \) is a RV, first \( \zeta - 1 \) frame transmissions are successful and one is unsuccessful, followed by the estimated duration \( N_g^+ \). The average number of transmissions between two consecutive estimated waiting durations is obtained as:

\[ E[\zeta] = \frac{1 + p_{10}(\beta)}{p_{10}(\beta)} \quad (26) \]

A. Performance Measures

1) Data throughput: Data rate \( D_R \) is defined as the average number of data frames delivered successfully per second.

In between two consecutive waiting phases, there occurs a single unsuccessful data transmission out of \( E[\zeta] \). Thus we get \( D_R \) as obtained in (27).

Lemma 3. \( D_R \) is a strictly decreasing function of \( N_g \) for a given \( Z_{th}, f_D, \text{ and } T_p. \)

**Proof:** By relaxing the integer constraint on \( N_g \) [37], we obtain \( \frac{\partial D_R(N_g)}{\partial N_g} = -AC(\frac{N_g}{D_N} \ln D - \frac{D_N}{N_g} + 1) \) from (27). In order to prove the lemma, we need to show that \( \frac{\partial D_R(N_g)}{\partial N_g} < 0 \forall Z_{th}, f_D, \text{ and } T_p. \) By definition we have both \( A, C > 0. \) Hence we need to prove that

\[ (N_g D_N \ln D - D_N + 1) > 0 \forall Z_{th}, f_D, \text{ and } T_p. \] \( (28) \)

By definition, we have \( N_g \geq 1 \) as the number of slots cannot be less than 1. The motivation behind this work of estimating the optimal waiting interval is to use the temporal correlation present in the channel for estimating \( N_g \) for a given \( Z_{th}, f_D, \text{ and } T_p. \) But a considerable amount of temporal correlation is present in the channel only when it is slow fading in nature, i.e., \( f_D T_p < 0.1. \) When \( f_D T_p < 0.1 \), we have \( 0 < \delta < 0.5 \) and accordingly we can observe that under this condition, \( D \notin [0,1] \) but \( D \in (0.5, 1). \) Finally with \( N_g \geq 1 \) and \( D \in (0.5, 1), \) (28) always holds and this completes the proof.

The above observation is intuitive also; increasing value of \( N_g \) implies that we are accessing the channel less frequently to send data and as a result the overall data throughput decreases. In other words, \( N_g^{rate} = 1. \)

2) Energy consumption per successfully delivered data frame, \( E_{EB} \): Let \( \nu_f, \nu_{ACK/NAK}, \text{ and } \nu_{idle} \) denote transmit and receive energy per data frame, transmit and receive energy per ACK/NACK frame, and per slot idling energy, respectively. These parameters are as follows:

\[ \nu_f = \Upsilon_f (\theta_f^2 + \theta_r^2) \phi, \quad \nu_{ACK/NAK} = \Upsilon_{A/N} (\theta_f^2 + \theta_r^2) \phi, \quad (29) \]

and \( \nu_{idle} = 2\theta_f^2 \Upsilon_f \phi. \)
Here \( \theta_t, \theta_r, \text{ and } \theta_w \) are the current consumptions in transmission, reception, and waiting modes respectively with \( \varphi \) being the circuit resistance. Accordingly, \( \mathcal{E}_B \) is obtained as in (30).

3) Energy Efficiency: It is obvious that a higher \( \mathcal{E}_B \) will result in a higher \( \mathcal{D}_R \), which in turn results in low battery life. This tradeoff is captured by taking the utility function as [41]:

\[
\mathcal{E}_{\text{Eff}} = \frac{\mathcal{D}_R}{\mathcal{E}_B}.
\]

(31)

From (27) and (30) we can see that both \( \mathcal{D}_R \) and \( \mathcal{E}_B \) are functions of \( N_g \), which also gets reflected in \( \mathcal{E}_{\text{Eff}} \). After we obtain \( N_g^* \) from Algorithm 1 for a given set of \( Z_0, Z_{th}, f_D, T_p \), and \( \epsilon \), we replace \( N_g \) by \( N_g^* \) in (27), (30) and (31) to obtain \( \mathcal{D}_R, \mathcal{E}_B \) and \( \mathcal{E}_{\text{Eff}} \) respectively.

**Lemma 4.** For a given set of \( Z_{th}, f_D, \text{ and } T_p \), there exists only one \( N_g = N_g^{\text{eff}} \) that maximizes \( \mathcal{E}_{\text{Eff}} \).

**Proof:** Using (27) and (30), we get:

\[
\mathcal{E}_{\text{Eff}} = \frac{A^2}{\left\{ B + \frac{CN_g}{1-D\alpha_g} \right\}} \cdot \left\{ E + \frac{F(\lambda+G+N_g)}{1-D\alpha_g} \right\} ^{-1}.
\]

(32)

Rewriting (32),

\[
\frac{1}{f(N_g)} = \frac{1}{f_1(N_g)} \frac{1}{f_2(N_g)}.
\]

(33)

where \( f_1(N_g) = \frac{1}{C_{\theta} \lambda^2}, f_2(N_g) = \frac{1}{C_{\theta} \lambda}, \) and \( f_2(N_g) = E + \frac{F(\lambda+G+N_g)}{1-D\alpha_g} \). As in Lemma 3, here also we relax the integer constraint on \( N_g \) to find the critical point of \( f(N_g) \). Accordingly we put \( \frac{\partial f(N_g)}{\partial N_g} = 0 \) to obtain

\[
T(N_g D_N^N \ln D + 1 - D_N^N) + S + 2RN_g
\]

\[
\frac{A^2}{\left\{ B + \frac{CN_g}{1-D\alpha_g} \right\}} \cdot \left\{ E + \frac{F(\lambda+G+N_g)}{1-D\alpha_g} \right\} ^{-1} = 0,
\]

\[
\frac{2D_N^N (RN_g^2 + SN_g) \ln D}{(1 - D_N^N)^3}
\]

(34)

where \( R = \frac{C_{\text{FE}}}{A^2}, S = \frac{C_{\text{FA}}}{A^2}, \text{ and } T = \frac{B + E_{\text{CC}}}{A^2} \). Simplifying (34), we obtain:

\[
(UN_g - T)D_N^N - (XN_g^2 + \eta N_g - \tau)D_N^N - (V + 2RN_g) = 0,
\]

(35)

where \( U = T \ln D, X = 2R \ln D, \eta = (T + 2S) \ln D - 2R, \text{ and } \tau = 2T + S \). It can be clearly seen that (35) is a quadratic equation of \( D_N^N \) the solution of which is (36).

We have: \( X < 0, \eta < 0 \) (: \( D < 1 \) and hence \( \ln D < 0 \), \( D_N^N > 0 \), and \( \tau > 0 \). Hence the only feasible solution of (36) is (37). Taking natural logarithm on both sides for further simplification, we get (38). The left hand side of (38) is a linearly increasing function of \( N_g \) and the right hand side is a logarithmic function of \( N_g \). Therefore, they intersect at a single point. This proves that there exists one and only one critical point \( N_g = N_g^{\text{eff}} \). It has been observed numerically that \( \frac{\partial f(N_g^{\text{eff}})}{\partial N_g} \) changes its sign from negative to positive at \( N_g = N_g^{\text{eff}} \), i.e., \( f(N_g) \) has a minima at \( N_g^{\text{eff}} \). From (33) we know that \( \mathcal{E}_{\text{Eff}} = \frac{1}{f(N_g^{\text{eff}})} \). Hence it is concluded that \( \mathcal{E}_{\text{Eff}} \) has a maxima at \( N_g = N_g^{\text{eff}} \).

\( \square \)

**B. Optimal \( \epsilon \) Estimation for Maximizing Energy Efficiency**

We observe that various objectives such as energy minimization and energy efficiency maximization lead us to various optimal solutions. But when we incorporate practical channel state behavior, unconstrained nature of the above-stated objectives changes. This happens because when various practical aspects are incorporated, the size of the feasible set of solutions decreases. To be specific,

1) By Lemma 3 we have \( N_g^{\text{safe}} = 1 \). This practically means transmitting data frames in every slot irrespective of the channel condition, which may not always be a feasible solution. For instance, there is no point in transmitting data frames in every slot when we know that the channel is slow fading in nature and is presently in deep fade \( (Z_0 < Z_{th}) \).

2) In Lemma 4, current SSI \( Z_0 \) has no role in estimating \( N_g^{\text{eff}} \). Moreover this \( N_g^{\text{eff}} \) may not always be a practically feasible solution. For instance irrespective of the current state of channel, we have \( N_g^{\text{eff}} = 7 \) for transmission of unit energy signal in a Rayleigh fading environment with system parameters \( Z_{th} = \ln 2, \text{ velocity } = 5 \text{ kmph} \), \( (f_D = 4.167 \text{ Hz}), \text{ and } T_p = 0.0005 \text{ sec} \). 

Thus we can see that \( Z_0 \) does play an important role in obtaining a practically feasible \( N_g^{\text{eff}} \), which has been overlooked in Lemma 4. When \( Z_0 \) is incorporated in Lemma 4, the optimization problem stands as follows:

\[
(P6) \quad \epsilon : \text{ maximize } \mathcal{E}_{\text{Eff}} \quad \epsilon \quad \text{ subject to } \quad C1 : \epsilon \geq 0, \quad C2 : \epsilon \leq \epsilon_u, \quad C3 : N_g \geq 1, \quad \text{ and } \quad C4 : g(Z_0, Z_{th}, f_D, T_p, \epsilon) = N_g,
\]

where Algorithm 1 is represented as a function \( g(Z_0, Z_{th}, f_D, T_p, \epsilon) \) that gives an output \( N_g \) for a given set of \( Z_0, Z_{th}, f_D, \text{ and } T_p \). C1 is intuitive from (15) but C2 is not. Theoretically \( \epsilon \in [0, 1] \), but it is the application at hand that specifies the acceptable range of \( \epsilon \). Accordingly, we introduce an user-defined upper limit \( \epsilon_u \), and thus we have \( \epsilon \in [0, \epsilon_u] \).

The boundaries of the feasible set \( \{ \epsilon : 0 \leq \epsilon \leq \epsilon_u \} \) is dependent on the fact that, for a given set of \( Z_0, Z_{th}, f_D, \text{ and } T_p \), there exists a one-to-one mapping between \( \epsilon \) and \( N_g \). If \( \epsilon \)
corresponding to $N_g^{\text{eff}}$ belongs to $[0, \epsilon_u]$, then that particular $\epsilon$ is termed as the optimal $\epsilon^*$ for a given set of $Z_0, Z_{th}, f_D,$ and $T_p$. Otherwise we take the corner point solution as $\epsilon^*$, i.e.,

$$
\epsilon^* = \min \{\epsilon_{opt}, \epsilon_u\}, \quad \text{where} \quad (40)
$$

where $\epsilon_{opt} = \{\epsilon_{opt} : g(Z_0, Z_{th}, f_D, T_p, \epsilon_{opt}) = N_g^{\text{eff}}\}.$

After obtaining $\epsilon^*$, we calculate $D_R$ and $E_B$ corresponding to D-SW protocol using (27) and (30). As D-SW directly resumes data transmission after the waiting period, we put $\lambda = 0$ in the obtained $E_B$.

We conclude this section by briefly discussing the various approaches, namely, adaptive probing variants (AP1 and AP2), and coherence time based waiting (CT):

1) API [24]: This scheme proposes to take AFD as the waiting interval $T_{gs}$.
2) AP2 [24]: Using (21), this scheme proposes $T_{gs}$ to be calculated as $T_{gs} = \frac{0.5 \times (\gamma_{Z_{th}}Z_{th} - Z_n)}{\gamma_{Z_{th}}}$, where $Z_n = Z_n + \frac{Z_{th}}{2L}$ is the quantized SSI lying in $\{Z_n, Z_{n+1}\}$ if the entire $\{Z | Z < Z_{th}\}$ range is sub-divided into $L$ number of levels with quantization step size $\frac{Z_{th}}{2L}$.
3) CT [25]: It takes coherence time $T_c$ as default $T_{gs}$ irrespective of the value of $Z_0$. As a rule of thumb, we take the relation $T_c = \frac{0.423}{T_D}$ [42]. Hence in this case, we have $T_{gs} = \frac{T_c}{T_D}$.

It is important to note that in all of these existing approaches, Tx sends a probing packet after time $T_{gs}$. Data transmission resumes if ACK is received, and the probing process is repeated if a NAK is received again. Accordingly $D_R, E_B,$ and $E_{\text{Eff}}$ are calculated. We compare D-SW with these schemes in Section VI in terms of $D_R, E_B,$ and $E_{\text{Eff}}$.

VI. NUMERICAL RESULTS

In this section, we evaluate the performance of D-SW protocol via extensive simulations and also compare its performance in terms of $D_R, E_B,$ and $E_{\text{Eff}}$ with the existing competitive schemes: API [24], AP2 [24], and CT [25].

The results are obtained by considering transmission of an unit power signal over Rayleigh fading channel. The acceptable SSI threshold is $Z_{th} = \ln 2$. The user-defined parameter $\epsilon_u$ is taken as $0.5$, i.e., the considered scenario has $50\%$ error as an acceptable limit and hence $\epsilon \in [0, 0.5]$. For propagation delay is ignored in context of short-range links [23], such as in IoT networks. Also for the numerical results, we consider $\beta = 1$, i.e., queuing and processing delays at the transmitter are ignored [23], [24].

A. Verification of D-SW Performance through Simulation

We first compare time $T_{gs} = N_g^{\text{eff}}T_p$ sec to reach above the acceptable signal threshold $Z_{th}$ in D-SW with Monte Carlo simulations. As observed in Fig. 6(a), SSI $Z_0$ does play a key role in determining $T_{gs}$ in D-SW. It can be further observed that $Z_0 \ll Z_{th}$ (i.e., channel being in deep fade state) is not the same as $Z_0$ being just less than $Z_{th}$. Also it is noted that, as $Z_0$ is increased from $-65$ dBm to $-20$ dBm, the corresponding $T_{gs}$ is more or less the same irrespective of mobility speed $v$.

With Fig. 6(a) showing that estimated waiting interval matches with the results obtained through Monte Carlo simulations, it validates D-SW and justifies the role of $Z_0$ in determining $T_{gs}$. Another interesting observation is that $T_{gs}$ for a particular $Z_0$ increases with decreasing value of $v$. This is because, at a smaller $v$ the channel fading is slower, and hence more time is taken in changing channel state. Lastly, for any particular $v$, $\epsilon_1 > \epsilon_2$ implies that the same relation holds for the corresponding $T_{gs}$’s also, i.e., $T_{gs}$ for $\epsilon = \epsilon_1$ is more than the $T_{gs}$ for $\epsilon = \epsilon_2$.

The accuracy of the proposed estimation scheme is explained in Fig. 6(b), where the error in $T_{gs}$ estimation against the actual simulation values for the various schemes
are plotted. Error of a scheme is defined as: Error(%) = \left[ \frac{T_{est} - T_{sim}}{T_{est}} \right] \times 100, where \( T_{est} \) and \( T_{sim} \) respectively represent the analytically estimated and simulation values for \( T_{gs} \). For generating this result we have considered \( Z_0 = -54 \) dBm as the input parameter. Fig. 6(b) shows that the error in case of D-SW is just 5% compared to 40% of its nearest competitor AP2.

Fig. 6(c) on the other hand shows the comparison of the average number of NAK frames per cycle in the various \( T_{gs} \) estimating schemes. From the figure we can see that the estimated waiting time in D-SW matches very closely with the simulated value. Unlike in the other approaches, the average number of NAK frames per cycle in case of D-SW is approximately 1. This implies that D-SW surpasses all the other existing competitive approaches by a significant margin in terms of efficiently characterizing the temporal variation of the wireless channel.

Moreover, Fig. 6(c) also justifies the usefulness of the proposed D-SW. As the average number of NAK frames per cycle is approximately 1, we can conclude that the proposed interval estimation is reasonably accurate and hence there is no need to send probing packets at all.

B. Effect of NAK frame size \( L_{NAK} \) on Energy efficiency \( \varepsilon_{Eff} \)

D-SW requires extra overhead cost due to inclusion of some additional number of bits \( L_{NAK} \) in the NAK frame. To study the effect of \( L_{NAK} \) on performance of D-SW, \( \varepsilon_{Eff} \) versus \( L_{NAK} \) is plotted in Fig. 7. It is observed that performance of D-SW initially increases with increase in \( L_{NAK} \), and then saturates beyond \( L_{NAK} \geq 5 \). This observation suggests that the optimal upper limit of \( L_{NAK} \) can be safely considered 5, because further increase in \( L_{NAK} \) does not offer any additional energy efficiency benefit.

C. Performance Comparison

In this subsection, we compare the performance of D-SW in terms of \( D_R \), \( \varepsilon_B \), and \( \varepsilon_{Eff} \), with the ones proposed in [24] and [25] that have the closest system settings.

1) Data throughput \( D_R \) comparison: Irrespective of the underlying scheme, Fig. 8(a) exhibits an overall decreasing trend of \( D_R \) with node velocity \( v \). It can be observed that increase in \( v \) results in high \( f_{DT} \), which implies that the channel is now less correlated than it was with lower \( f_{DT} \). Hence the channel should be sensed more frequently in order to estimate its state, which ultimately leads to the decrease of \( D_R \) and increase of \( \varepsilon_B \), respectively.

The figure shows that AP1 [24] and CT [25] achieve very low \( D_R \) in comparison to the others. They respectively achieve \( D_R \) in the range of 650 to 700 and 600 to 650 frames/sec, which is approximately 32% and 42% lesser with respect to D-SW. This implies that some intelligence is required at the Tx to obtain high \( D_R \). Unlike AP1 and CT, AP2 [24] and D-SW incorporates the additional knowledge of \( Z_0 \). As a result it is able to achieve higher \( D_R \). We can observe from Fig. 8(a) that D-SW performs consistently better than AP2 [24] in terms of obtaining higher \( D_R \) with a performance gain; the gain margin being approximately 9%.

2) Energy consumption \( \varepsilon_B \) comparison: For comparison of \( \varepsilon_B \), the values of \( \theta_t, \theta_r, \) and \( \theta_w \) are taken from Chipcon CC1000 data sheet [43] as: \( \theta_t = 17.4 \) mA, \( \theta_r = 19.7 \) mA, and \( \theta_w = 426 \) \( \mu \)A with \( \varphi = 1 \) \( \Omega \).

Irrespective of the scheme considered, we observe an overall increasing trend in \( \varepsilon_B \) with velocity \( v \). From Fig. 8(b), it can be seen that \( \varepsilon_B \) of AP1 and CT are almost identical. \( \varepsilon_B \) of AP2 is higher than the other two schemes, though by a very small margin. It is interesting to observe that \( \varepsilon_B \) of D-SW is the least compared to the existing protocols; the maximum difference between \( \varepsilon_B \) of D-SW and AP2 is approximately 4%. Hence we can see that though \( D_R \) of D-SW is approximately 9% more than that of AP2, its corresponding \( \varepsilon_B \) is less compared to
AP2. In other words, D-SW is more energy-efficient compared to AP2, which is discussed next.

3) Energy efficiency $E_{\text{Eff}}$ comparison: From Section V-A3, we know that higher the value of $E_{\text{Eff}}$, the better is the scheme. Fig. 8(c) shows the variation of $E_{\text{Eff}}$ against node velocity, where we can observe that D-SW performs better compared to both CT and AP1 by a significant margin of more than 30%. Even when D-SW is compared to its nearest competitor AP2, there exists a consistent gap of approximately 12% with $E_{\text{Eff}}$ of AP2. This shows the advantage of the proposed D-SW protocol. We can also conclude from Fig. 8(c) that when the optimum waiting interval estimation is accurate, Tx can directly resume data after waiting for the estimated time interval without sending any probing packet further to check the condition of the channel. This results in increasing the energy-efficiency of the system.

Moreover, we have also seen in Fig. 6 that in terms of characterizing the temporal variation of the wireless channel, D-SW is the most efficient one. The error corresponding to D-SW is approximately 5%, whereas it was in the range of 40% in case of its nearest competitor AP2. Similarly, the average number of NAK frames per cycle is almost unity in D-SW compared to its nearest competitor AP2. When we combine this fact along with approximately 12% additional energy saving that D-SW offers over its nearest competitive approach AP2, it clearly brings out the advantage of the proposed protocol.

VII. CONCLUSION

In this paper we have proposed a new channel-adaptive, energy-efficient link-layer data frame retransmission strategy, called D-SW, for IoT applications over short-range wireless links. D-SW optimally estimates the next waiting interval based on the current state of the channel. This optimal waiting interval estimation helps to avoid unnecessary probing of the channel without missing out on an opportunistic chance of data transmission whenever the channel condition is suitable. The analysis of the proposed D-SW strategy in general is valid irrespective of the underlying fading channel distribution. We have also demonstrated that the well-known concept of AFD is a special case of the estimated waiting interval. Extensive simulations have been done which validate the proposed analysis. Numerical results exhibit the improved performance of D-SW in terms of efficiently characterizing the temporal variation of the wireless channel as well as obtaining high data throughput and energy efficiency. We have also estimated an optimal level crossing probability $p$ that maximizes the energy efficiency for a given set of operating parameters, $Z_0$, $Z_{th}$, $f_D$, and $T_p$. The overall performance suggests that for short-range wireless links D-SW is a good choice that offers significant increase in data throughput as well as energy-efficiency.

ACKNOWLEDGMENT

This work has been supported in parts by the ITRA Media Lab Asia project under Grant no. ITRA/15(63)/Mobile/MBSSCRN/01 and the Department of Science and Technology under Grant no. SB/S3/ESEE/0248/2014.

REFERENCES


