# DiNAMAC: A disruption tolerant, reinforcement learning-based Mac protocol for implantable body sensor networks

Vignesh Raja Karuppiah Ramachandran, Duc V. Le, Nirvana Meratnia and Paul J.M Havinga
Pervasive Systems, Dept. of Computer Science, University of Twente, Enschede, The Netherlands

***Abstract*— Ongoing advancements in Body Sensor Networks (BSN) have enabled continuous health monitoring of chronically ill patients, with the use of implantable and body worn sensor nodes. Inevitable day-to-day activities such as walking, running, and sleeping cause severe disruptions in the wireless link among these sensor nodes, resulting in temporary shadowing of wireless signals. These disruptions in the wireless link not only reduce the reliability of the network but also increase the power consumption. Both signal disruption and power consumption must be reduced in order to achieve long term monitoring of physiological signals in chronic patients. In this paper we propose a MAC protocol called DiNAMAC (Disruption tolerant reiNforcement leArning-based MAC), which is not only aware of the wireless link quality but also is aware of network resource availability and application requirements. DiNAMAC uses reinforcement learning to adapt the scheduling based on channel conditions and to prioritize data transmission and availability according to the application requirements. In addition, we design DiNAMAC based on a model-free learning technique to make it more practical in real-world applications. Our simulation results show that DiNAMAC performs better than conventional MAC protocols in terms of latency and throughput even with when the wireless link quality is challenged by large temporal variations.**

## I. Introduction

Health-care technologies are continuously being evolved, aiming to improve the quality of life and to reduce the health-care costs. The medical costs of patients are much lower when the health of the patient is monitored continuously at their own home rather than at hospitals [1]. Many high precision body worn and implantable sensors are being developed to continuously monitor health condition of the patients in order to ensure that the recorded vital physiological signals clearly exhibit all relevant symptoms of a given disease. A network of these implantable medical devices and sensors forms an implantable sensor networks (IBSN).

To ensure availability and quality of required data for continuous monitoring, a Medium Access Control (MAC) is needed which ensures the network reliability and energy efficiency simultaneously. Although there exist a number of MAC protocols developed for various WSN applications, the specific requirements of IBSN make them not completely suitable for IBSN.

It is a well established fact that the link quality between the sensor nodes of IBSNs is heavily disrupted even by normal human activities such as walking and running [2] [3]. This is mainly due to the very low isotropically radiated power of $25\mu W$ set by the spectrum regulations. Additionally at such low power transmission, fading and shadowing of RF power is very frequent due to the conductivity of human body [4].

Therefore, one of the important requirements for the MAC protocol is the capability to deal with unreliable wireless link and to handle the enormous temporal variations in the wireless link quality caused due to the day-to-day activities such as walking, running, sleeping, and sitting of a patient [5].

These temporal variations are very unpredictable since the patient can carry out random activities at any time [2]. Moreover, the pattern of activity by itself is dynamic, for example, patients walk in different patterns resulting in very different RF disruption models for the same activity. Modelling such highly dynamic RF disruption, if not impossible, is a complex process since it is both patient and environment centric.

In [2], we investigated the possibility to detect simple activities such as walking, running, sleeping, sitting and standing based on the RF signal strength of the sensor nodes in IBSN and their disturbances. Although we were able to define an RF disruption pattern for these simple activities in a controlled environment, we found that in real-world, there can be a number of disruption patterns for the same activity performed by the same person at different points of time and places. This will lead to inability to anticipate the correct influence and disruption of the activity on the RF signal and will consequently greatly decrease the efficiency of the MAC protocol to adapt its duty-cycle according to the disruption pattern. This will result in increased re-transmission and thereby reduce the throughput and energy-efficiency of the network. Therefore, we define adapting the duty-cycle according to the disruption pattern, without compromising on the throughput, latency and energy-efficiency, as the optimization problem of our MAC protocol.

In this paper, we will explore reinforcement learning [6] to optimize the MAC protocol by defining the disruption as a stochastic process with Markov property. In this respect, the main contributions of this paper are:

- Definition of dynamic disruption as a stochastic process with Markov property, thereby defining the problem of achieving disruption tolerance as a MDP
- Developing a reinforcement learning-based optimization framework
- Designing a near-optimal disruption tolerant MAC protocol called DiNAMAC based on our reinforcement learning-based framework.
- Performance evaluation of DiNAMAC in simulations.

Vignesh raja is with Pervasive Systems Research Group of EWI faculty at The University of Twente, Netherlands e-mail: (v.r.karuppiahramachandran@utwente.nl).

The rest of this paper is organized as follows. In Section II, we list the existing MAC algorithms and explain their potential drawbacks which make them unsuitable for disruptive wireless channel. In Section III, we explain the disruptive wireless channel and the choice of link quality metric, followed by Section IV which explains the reinforcement learning framework. In Section V, optimization framework of the DiNAMAC is explained. The simulation setup and the evaluation methods are explained in Section VI, followed by the discussion of the results in Section VII. Finally we conclude the paper in Section VIII

## II. Related work

The optimization problem of the MAC protocol is a critical issue in wireless sensor networks. Majority of existing work deals with optimizing the duty-cycle to ensure maximum throughput, minimum latency, and maximum energy-efficiency [7].

There are also a number of adaptive MAC protocols specifically designed for medical applications [8], [9], [10].

The MEDMAC protocol presented in [8], is an adaptive TDMA-based MAC protocol which incorporates an adaptive TDMA synchronization mechanism in which only a multi-superframe beacon has to be listened to by the nodes. The main aim of the MEDMAC was to reduce power consumption in the network of heterogeneous medical devices, by an adaptive guard band algorithm. The guard band enable the nodes to sleep longer by missing few synchronization beacons. This longer sleep cycle reduces power consumption by eliminating the need for the nodes to wake-up in every synchronization beacon period.

The HDMAC protocol presented in [9] is a TDMA protocol, which synchronized the ON time of different nodes based on the Heartbeat rhythm. Authors of [9] show that adapting the synchronization beacons increased the network lifetime by 15%–300% more than other pure TDMA MAC.

The Medical emergency MAC [10] operates in a hybrid Time Division Multiple Access (TDMA) scheme and aims to optimize the duty-cycle based on the application requirements. This MAC adapts the duty cycle based on the priority of the medical devices.

It is a well established fact that the power consumption of the wireless radio is decreased by reducing the duty cycle of the radio in MAC protocol. Adapting duty cycle to ensure high throughput and low power consumption is a common optimization problem of MAC protocols.

Use of machine learning framework to solve the optimization problem of duty-cycle have also been investigated earlier. One such approach is presented in [11], where authors developed a near-optimal transmission strategy using Reinforcement Learning (RL) technique that chooses the optimal modulation level and transmit power while adapting to the incoming traffic rate, buffer condition, and the channel condition. In case of IBSN, the modulation level and transmission power are heavily constrained by the regulations in order to ensure the safety of the patient by preventing the overheating of tissue around the antenna, and data integrity in the wireless communication of medical implants [12].

Authors of [13] introduced RL-MAC, which uses RL mechanism to achieve high throughput and low power consumption for a wide range of traffic conditions, by inferring the current state of other nodes and adapting the duty cycle based on the data generation rate of the sensor node.

Similarly, authors of [14], introduced QL-MAC which uses RL mechanism to find an efficient radio schedule on the basis of the node's own traffic and the traffic load of its neighbours. The simulation results show that the adaptive behaviour of QL-MAC guaranteed better network performances with respect to both the packet delivery ratio and the average energy consumption.

Both, RL-MAC and QL-MAC are not disruption tolerant, since the model of the RL technique is based only on the data throughput and energy efficiency and it had not included the link quality information in the RL optimization framework.

Most of the MAC protocols do not consider the disruptions in the wireless channel, which causes temporal variations in the link quality between the sensor nodes. Since temporal fluctuations in the RF link quality was not considered by these protocols, when they are exposed to the dynamic temporal link quality, the optimization of duty-cycle significantly drops [15]. Moreover, there is no mechanism to handle the increased Packet Error Rate (PER), which is caused due to the fluctuations in RF link quality.

To the best of our knowledge, DiNAMAC is the first to explore reinforcement learning technique at the MAC layer to handle the link disruptions caused due to human activities in Body Sensor Networks (BSN).

## III. Disruptive wireless channel model

Let us assume an IBSN network of $N$ nodes is deployed in a stochastic wireless channel that has Markov property. The physical interpretation of the communication link between two nodes in such wireless channel is given by the complement of the information outage probability between any pair of nodes for a given data rate $x$ in bits [16], and is represented as,

$$P_{out}(SNR, x) = P_r(\log_2(1 + SNR \times |h|^2) > x) \quad (1)$$

where $h$ is the channel transfer coefficient and the average signal to noise ratio SNR is related to the path loss component $\eta$ such that $SNR \propto r_{ij}^{-\eta}$, where $r_{ij}$ is the relative distance between the nodes i and j. Typically, $\eta > 4$ in the case of implant communication, as recorded in our previous studies [17]. We consider Rayleigh fading model and that all channels are statistically independent. Since number of nodes in IBSN is less than conventional sensor networks and due to the fact that the IEEE 802.15.6 standard for medical implant communication requires channel to be accessed at node's destined time frames, we assume that the intra-network interference is not present in IBSN. Hence the connection probability between node i and node j separated by a relative distance $r_{ij}$ is given by

$$H(r_{ij}) = e^{-\beta r_{ij}^{\eta}} \quad (2)$$

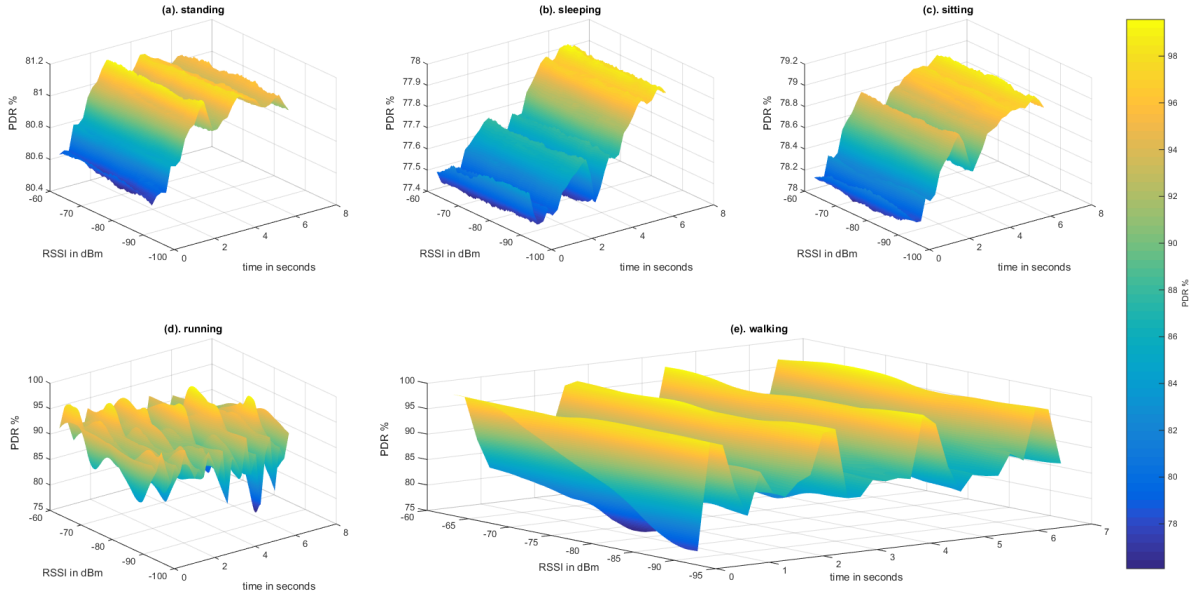


Fig. 1: Time vs RSSI. The color-map of the surf-plot shows the distribution of PDR in different activities

where, $\beta$ is responsible for the characteristic connection distance between two nodes such that $r = \beta^{-1/\eta}$. (This and other notation used throughput this paper is gathered in Table I.)

According to [16] this connection probability has two sources of randomness: random node positions, and random link formation based on the channel fading model controlled here by $\eta$. In the limit of $[\eta \to \infty]$, the link between two nodes is no longer probabilistic. We will later make use of this randomness to model the reward function of MAC design in the reinforcement learning based framework.

### A. Link quality metric in stochastic wireless channel

In highly random wireless channels, received signal strength do not completely represent quality of the link, as it does not include noise information in the channels. Another commonly used metric is Signal to noise ratio $E_b/N_0$, where $E_b$ is the signal energy of each transmitted bit and $N_0$ is the noise spectral density in the channel. Although SNR contains the information of noise density, direct instantaneous measurement of SNR on IBSN is not practical. Because, the approximation of noise floor of the human body channel is not only dependant on the bandwidth, but it is also dependant on internal body temperature of the human, composition of the tissues and bones and location of the sensor nodes on the human body [18], [19].

Moreover SNR of the wireless channel does not directly represent the successful packet delivery, as the use of intelligent carrier modulation techniques will result in a higher Packet Delivery Ratio (PDR) even in a lower SNR.

To this end, we will use packet error rate $\mathcal{P}_P$ as a link quality metric, since it carries information about bit error rate $\mathcal{P}_B$, which carries information about the signal to noise ratio $SNR = E_b/N_0$, i.e.

$$\mathcal{P}_P = 1 - (1 - \mathcal{P}_B)^{N_{bits}} = \mathcal{N}\,\mathrm{erfc}(\sqrt{E_b/N_0}) \qquad (3)$$

where, $Nbits$ is the number of bits transmitted in a packet, $\mathcal{N}$ is the modulation coefficient which depends on the type of modulation used in transmission.

There are two main advantages in choosing $\mathcal{P}_P$ as a link quality metric.

Firstly, in IBSN, the client nodes always carry out uplink communication and central controller always carries out the downlink communication. An exception to this is for control packets which are a small fraction of the sensor data. Since, central controller is responsible for optimization of MAC protocol, need for client nodes to be aware of link quality is non-existent. Most of the up-time, the central controller, which is also the receiver, will be carrying out error detection mechanisms to verify the integrity of the data. This procedure implicitly enables the central controller to measure the link quality after each packet is received, eliminating the need for any additional calculation or measurement schemes.

Secondly, $\mathcal{P}_P$ also contains accumulated information of SNR for a time duration at which the packet was in the physical medium. This will include even the minimal variations of temporal noise of the channel and will enhance the Markov property of the wireless channel.

## IV. Markov Decision Process and RL Techniques

The objective of our work is to decrease $\mathcal{P}_P$ in a highly disruptive and stochastic wireless channel, with an optimal latency and energy-efficiency. To achieve this, we design a MAC agent, which predicts the $\mathcal{P}_P$ in the consecutive time-frame, and allocates a time slot to a node, in which the node

and the controller is expected to have a good link quality. We define this allocation of an optimal time-slot to the node, without compromising on latency and energy efficiency as a Markov Decision Process (MDP) of the Reinforcement Learning (RL) model.

### A. Markov decision process of RL model

A MDP of the RL model is defined as a tuple $(S, A, \{P_{sa}\}, \gamma, R)$, where $S$ represents the *state* space $\{s_1, s_2, ..., s_n\}$, $A$ represents the *action* set $\{a_1, a_2, ..., a_n\}$, $P_{sa}$ represents the transition probability function. For each state $s \in S$ and action $a \in A$, $P_{sa}$ is a distribution over the state space, $\gamma \in [0, 1)$ is called the *discount factor*, and $R$ is the *reward function* such that $R: \ S \times A \mapsto R$. Also, the decision policy is defined as $\pi$, which maps the state set to an action set such that $\pi : S \mapsto A$. At an event $i$, the RL agent is at state $s_i \in S$, it selects an action $a_i \in A$, according to the defined policy $\pi(a_i)$. This action-state policy will interact with its environment, and state $s_i$ will now transit to state $s_{i+1} = s' \in S$ with a transition probability of $P_{s_i a_i}$. The environment now provides the RL agent with a feedback reward of $r_i(s, a)$. This whole process is repeated for all event iterations. The goal of the RL agent is to maximize the policy or the value function $\mathcal{V}^{\pi}(s)$ [6], where,

$$\mathcal{V}^{\pi}(s) = E\left[\sum_{i=0}^{\infty} \gamma^i r_i(s_i, a_i) | s_0 = s, \pi)\right] \tag{4}$$

where, $E[.]$ represents the expected return when starting from a state $s$ and following the policy $\pi$.

Given a fixed policy $\pi$, the equation 4 can be rewritten as the Bellman equations,

$$\mathcal{V}^{\pi}(s) = R(s, a) + \gamma \sum_{s' \in S} P_{sa}(s') \mathcal{V}^{\pi}(s') \tag{5}$$

where, $R(s, a) = E[r(s, a)]$ is the mean value of the reward $r(s, a)$, the immediate reward that we get rightaway simply for starting in state $s$. The second term is the expected sum of future discounted rewards.

In our framework, we define the $\mathcal{P}_P$ as the state, and allocation of time-slot $T_s$ and deploying a relay node between controller and client node $R_n$ as the actions such that, $[\mathcal{P}_P] \in S$ and $[T_s, R_n] \in A$. In addition, transition from state $\mathcal{P}_P$ to state $\mathcal{P}_P^{'}$ has a probability of $P_{sa}$, such that $P_{sa} \in \{P_{sa}\}$.

In practical scenarios, the transition probabilities $\{P_{sa}\}$, reward function $R$, and the model of the environment are unknown. In such cases, Temporal Difference (TD) based RL is commonly used [6], [11], [13]. TD methods can learn directly from raw experience without a model of the environment's dynamics. In our case the environment is stochastic and dynamics of the channel is unknown. TD learning will adapt the duty-cycle and allocate the time slots based on the predicted outcome of the environment based on the previous knowledge of the environment.

### B. Q-learning

Since our goal is to train DiNAMAC in the absence of transition probability and reward function, we propose to use an off-policy TD based control algorithm known as Q-learning, which is a model-free technique learning from delayed reinforcement to determine an optimal policy [6].

In Q-learning, policies and the value function are represented by a look-up table indexed by state-action pairs. Formally, for each state s and action a, we define the Q value under policy $\pi(a)$ to be:

$$Q^{\pi}(s, a) = R(s, a) + \gamma \sum_{s' \in S} P_{sa}(s') \mathcal{V}^{\pi}(s') \tag{6}$$

From 6, it is known that the expected discounted reward starting from s, will take an action a from $A$ and thereafter follow the policy $\pi$. Therefore, the value function of an optimal policy $\pi^*$, denoted by $V^*$, can be defined as:

$$\mathcal{V}^*(s) = \mathcal{V}^{\pi^*}(s) = \max_{\pi} \mathcal{V}^{\pi}(s) \tag{7a}$$

$$\mathcal{V}^*(s) = \max_{a \in A} \left( R(s, a) + \gamma \sum_{s' \in S} P_{sa}(s') \mathcal{V}^*(s') \right) \tag{7b}$$

Let us assume that $Q^*(s, a)$ be the optimal action function under the $\pi^*$, then

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in S} P_{sa}(s') \max_{a \in A} Q^*(s', a) \tag{8}$$

The Q-value is updated by a learned action value function, by iterating the value of $Q^*$. The updating rule is defined based on the temporal difference method, which is given by,

$$Q_{i+1}(s, a) = \begin{cases} Q_i(s, a) + \alpha\Delta & \text{if } s_i = s, \ a_i = a \\ Q_i(s, a) & \text{otherwise} \end{cases} \tag{9}$$

where, $\alpha$ is the learning rate and $\Delta$ is the temporal difference defined by,

$$\Delta = r_i(s, a) + \gamma \max_{a \in A} Q_i(s', a) - Q_i(s, a) \tag{10}$$

From Equation 8 and 10, it is clear that the state-value function will make the agent to always learn to maximize its reward. To make the agent achieve our objective, it is important that the reward function must truly indicate the objective to be accomplished. In other words, reward function should not directly define the duty-cycle or enable a relay node, rather it should encourage the MAC agent to learn from the environment and derive value-policies for different states of packet error ratio ($\mathcal{P}_P$).

In the following section, we explain the reward function and optimization policies to achieve our goal of increasing the throughput and decreasing the latency, thereby increasing the energy efficiency in a wireless channel with a high $\mathcal{P}_P$.

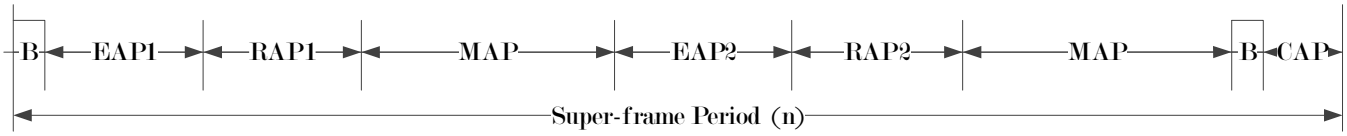


Fig. 2: Superframe of IEEE 802.15.6

Fig. 3: Location of IBSN nodes in simulation set-up

## V. DiNAMAC design

DiNAMAC is an hybrid access mechanism based on the IEEE 802.15.6 MAC layer specifications and is intended to operate in MICS band communication [12]. We assume that the client nodes (BN) of IBSN are resource-constraint and are implanted subcutaneously in the human body with one external body-worn coordinator node (BNC). An example of IBSN is shown in Figure 3, where the red node is BNC and coffee-colored nodes are implanted subcutaneously in the body. The BN nodes can also act as a relay node (BNR) to hop information with the same mechanism as described in the IEEE 802.15.6 standard [12].

The super-frame structure as defined by IEEE 802.15.6 standard for medical implants has three different phases, exclusive access phase (EAP) in which the nodes can only send emergency medical data in contention free fashion, random access phase (RAP) in which all nodes can contend to send medical-data, and a managed access phase (MAP) in which the BNC allocates either carrier sense multiple access with collision avoidance (CSMA/CA), slotted aloha, polled allocation, or scheduled time division multiple access mechanism (TDMA) for each BN node to send the medical data. The super-frame structure is shown in Figure 2.

DiNAMAC is operated in beacon mode, in which a beacon frame $B$ is transmitted from BNC to the BN nodes at the start of each super-frame period to facilitate the network management, including the coordination of medium access phases, duty-cycle of the BN nodes connected the BNC, and clock synchronization information. DiNAMAC uses TDMA during the MAP phase for transmitting the non-emergency data which applies to continuous monitoring of medical symptoms. It uses CSMA/CA during the EAP phase for sending emergency data, which can be an anomaly in the medical data.

In DiNAMAC, BNC performs minimalistic operations to adapt the super-frame for reliable communication. BNC does not require additional packets to carry out these operations because the required information, $\mathcal{P}_P$, is derived from the data-packets transmitted previously. BNC propagates the adaptive super-frame structure to the BN nodes through the beacons, which makes DiNAMAC simple yet powerful. The BN nodes determine their position, link quality, relay mechanism, and periodicity, through the beacon frame of the BNC, based on which the BN node determines the access mechanism, time slot and the relay to transmit the information to BNC.

| Notation | Description |
|---|---|
| $\mathcal{P}_P$ | Packet error ratio |
| $n_{rx}$ | No. of successfully received packets |
| $n_{frx}$ | No. of corrupted packets |
| $\mathcal{N}_{eff}$ | $n_{rx}/(n_{rx}+n_{frx})$ |
| $\tau_{slot}$ | Duration of one time slot |
| $\tau_{on}$ | Duration of ON time within a time slot |
| $\tau_d$ | Duration of disruption window |
| $\tau_{fad}$ | Duration of fading window |
| $\tau_{Tx}$ | Duration of transmit time |
| $\tau_{Rx}$ | Duration of receive time |
| $\alpha$ | Learning rate |
| $\gamma$ | Discount factor |
| $\eta$ | Weighing factor |
| $\Delta$ | Temporal difference of $Q_i(s,a)$ and $Q_{i+1}(s,a)$ |
| $N_{bits}$ | Number of bits in a packet |
| $\mathcal{R}$ | Data rate in bits per second |
| $S$ | State space |
| $A(s)$ | Action space |
| $r_i(s,a)$ | Reward policy in $i^{th}$ iterative index |
| $\epsilon$ | Probability with which the agent executes the action with highest Q value |

TABLE I: RL notations used in text and derivations

The goal of DiNAMAC is to maximize highly reliable communication by minimizing the $\mathcal{P}_P$ in an disruptive environment. The unique feature of DiNAMAC is that there is no need for a priori knowledge of the environmental changes which cause the disruption of wireless channel, rather it is predicted at each iteration of the super-frame duration. This posteriori knowledge of the wireless channel will enable DiNAMAC to handle a wide range of dynamic disruptions, resulting in a disruption tolerant communication.

### A. Action Policy

In order to achieve the goal of DiNAMAC we define our action policy $\pi$ with two main actions that has to be taken in different instances of disruption, namely the ON time allocation and Relay node establishment.

*Definition 5.1:* **ON time** $\tau_{\mathbf{on}}$ is the duration in which

a node is actively transmitting within the time slot $\tau_{\text{slot}}$ allocated to it in the $i^{th}$ super-frame.

As shown in Figure 1(d) and Figure 1(e), the periodic activities performed by human beings have sufficient time periods of favorable link quality, which is expressed by the PDR (%). If the duration and position of $\tau_{on}$ is perfectly allocated to match with when the signal quality is high, represented by PDR crests in Figure 1(d) and Figure 1(e), optimal throughput and energy-efficiency will be achieved. Our DiNAMAC aims at optimizing $\tau_{on}$ allocation in terms of location and duration to archive optimal throughput and energy-efficiency. By doing so, DiNAMAC is able to yield high PDR even when the link quality varies over time because the movement of the person.

*Definition 5.2:* **Relay node BNR** is a temporary node placed between the shadowed BNC and the BN to deal with severe shadowing of the RF link.

If the RF link between the BN and the BNC is shadowed for a prolonged period of time as shown in Figures 1(a), (b) and (c), the BN increases the possibility of missing life-critical medical events and also will increase the latency. As the transmission power cannot be increased in the MICS RF band, DiNAMAC temporarily uses relay nodes to overcome this problem. By doing so, DiNAMAC is able to assure high PDR even when the link between the BN and the BNC nodes is constantly shadowed by the human body.

### B. RL reward function for DiNAMAC

Designing the reward function is crucial for the RL learning process of the MAC agent. Since our goal is to maximize the throughput and minimize the energy consumption, we design our reward function based on these values.

For the energy efficiency, the node is supposed to accurately predict the disruption windows $\{\tau_d^i, \tau_d^{i+1}, ..., \tau_d^n = \mathcal{T}_d\}$ in the wireless channel, and allocates ON time $\tau_{on}$, as an exclusion function of $\mathcal{T}_d$, i.e. $\{\tau_{on} \cap \mathcal{T}_d = \emptyset\}$. In IBSN it is not desirable to decrease the energy consumption without considering the expected throughput, i.e. a node should not be allocated longer ON time to increase the throughput. Therefore, the reward of ON time allocation has to be a function of effective package receiving times ratio, denoted by $\mathcal{N}_{eff} = n_{rx}/(n_{rx} + n_{frx})$, where $n_{rx}$ is the number of successfully received packets in a receive time $\tau_{rx}$, and $n_{frx}$ is the number of corrupted packets in the transmission time $\tau_{tx}$.

In addition, the fading between BNC and BN is constant and strong over a period of time $\tau_{fad}$, which is greater than the slot time $\tau_{slot}$. In such cases, MAC agent establishes a relay node BNR between BNC and shadowed BN node, in the next time slot allocated for the BN within the super-frame duration. Since the node position in fixed in IBSN for any duration of time, the selection of BNR is predetermined and the one-hop route information from any BN to BNC through an optimal BNR node is already known to the BNC. This establishment of BNR at continued shadowing will further improve the $\mathcal{N}_{eff}$ in addition to the ON time allocation.

For the throughput, the packet error ratio at the start of the next frame, denoted by $\mathcal{P}_P^{'}$, is a valid indicator of the effectiveness of the allocated ON time in the previous frame. Hence, the effectiveness of the ON time window is expressed as a function of the packet error rates $\mathcal{P}_P$, $P_P^{'}$, and the throughput. We refer the term throughput as the data rate $\mathcal{R}$ at which the node is transmitting is in bit/s. To this end, we incorporate these two rewards in a reward function defined by

$$r_i(\mathcal{P}_P) = \begin{cases} \frac{(n_{\text{rx}}+n_{\text{frx}})(\mathcal{N}_{\text{eff}})^{n_{\text{hop}}}}{\tau_{\text{slot}}-\tau_{\text{on}}} - \eta\frac{\mathcal{P}_P^{'}-\mathcal{P}_P}{\log_{N_{bits}}(\mathcal{R})} & \text{if } \mathcal{P}_P \neq 0,\ \mathcal{P}_P^{'} > \mathcal{P}_P \\ \frac{(n_{\text{rx}}+n_{\text{frx}})(\mathcal{N}_{\text{eff}})^{n_{\text{hop}}}}{\tau_{\text{slot}}-\tau_{\text{on}}} & \text{if } \mathcal{P}_P \neq 0,\ \mathcal{P}_P^{'} \leq \mathcal{P}_P \\ -\eta\frac{\mathcal{P}_P^{'}-\mathcal{P}_P}{\log_{N_{bits}}(\mathcal{R})} & \text{if } \mathcal{P}_P = 0,\ \mathcal{P}_P^{'} \neq 0 \\ 1 & \text{if } \mathcal{P}_P = 0,\ \mathcal{P}_P^{'} = 0 \end{cases} \quad (11)$$

where, $n_{\text{hop}}$ is the number of hops that is made in the last frame (in IBSN it is usually either 0 or 1) and $\eta$ is the weighing factor.

### C. Q-learning algorithm

In our learning process, at the end of each frame, the RL agent evaluates the temporal difference $\Delta$, updates the Q-value and selects the next action according to the $\epsilon$-greedy method [6]. Using this approach, with probability 1 - $\epsilon$, the agent executes the action with the highest Q value, and with probability $\epsilon$ the agent randomly chooses an alternative action. This is done to balance exploitation of presumed optimal state-action pairs and exploration of novel policy modifications. Naturally, when $n_{frx}$ increases due to high $\mathcal{P}_P$, one would expect the allocation of ON time to have an overlap with the disruption window $\tau_d$. Also, since the traffic load and the networking condition vary in our case, we adopt a constant learning rate $\alpha$ = 0.1 as recommended in [6] in order to adapt to the non-stationary environment. We further note that if the shadowing due to fading effect is constant over a relatively long period of time, the $\mathcal{P}_P$ will always not change abruptly, hence the learning process is accelerated. This accelerated learning will further improve the energy-efficiency. The learning algorithm is given in Figure 4.

### D. Operation of BN and BNR

The BN node receives the beacons and identifies the duration of each phase and the time slot allocated for it. If there is no data to be sent, the BN node transmits the no-data frame and goes to sleep, otherwise it transmits on the ON time slot allocated for it. In case of a emergency event or an anomaly in the sensed medical data, the BN node contends to send data in the consecutive EAP phase with reference to the latest received beacon. BN node does not have to send any additional packets for the operation of DiNAMAC.

**procedure** RL˙OPTIMIZE˙MODULE($\alpha, \gamma, \epsilon$)
  Init$[A(s)]$ $\forall$ $s \in S$;
  BNC $\leftarrow$ Assign $(\alpha, \gamma, \epsilon)$
  Init [IBSN] $\leftarrow$ $\langle Tx_{\text{power}}, \mathcal{R}, (\text{BNC} \leftarrow \text{BNR} \leftarrow \text{BN})\rangle$
    ▷ PHY & Topology fixed
  Init [802.15.6 MAC] $\leftarrow$ $\langle \mathcal{B}, \tau_{on}, \tau_{slot}, \mathcal{N}_{eff}, \mathcal{P}_P\rangle$
  Init˙$Q(s,a) = 0 \; \forall \; s \in S, \forall \; a \in A(s)$
  **for** Every $i^{th}$ SuperFrame_BNC **do**
    Observe $\mathcal{P}_P$
    $\mathcal{N}_{eff} \leftarrow (n_{rx}/(n_{rx} + n_{frx}))$
    Perform $a_i = (t_{on} \| BNR)$
      ▷ in $s_i = \mathcal{P}_P$ based on $[\epsilon \; \{Q(s,a)\}]$
    $i++$;
  **end for**
  Reward [MAC] $\leftarrow$ $r_i(\mathcal{P}_P)$
  $\mathcal{P}'_P \leftarrow$ Update $\mathcal{P}_P$
  $\Delta \leftarrow r_i(s,a) + \gamma \max_{a \in A} Q_i(s',a) - Q_i(s,a)$
  **if** $s_i == s, \; a_i == a$ **then**
    $Q_{i+1}(s,a) \; = \{ \; Q_i(s,a) + \alpha\Delta\}$
  **else**
    $Q_{i+1}(s,a) = Q_i(s,a)$
  **end if**
  **go to** *7:*
**end procedure**

Fig. 4: RL optimization framework

In case of the relay initiation, BN node which acts as the BNR will encapsulate the data in its frame and transmit it to the BNC in the additional time-slot allocated for it. This process is similar to the relay mechanism as described in IEEE 802.15.6 standard [20]. The choice of selecting a BN node as a relay node is pre-determined based on the location of the node. For example, in Figure3, the BN node in the ankle, uses BN node in the thigh as the relay node. Since the location of these two nodes is fixed, the choice of relay node is easily pre-determined and the relay-topology for each BN node is pre-stored in the BNC.

## VI. Evaluation of DiNAMAC

For the purpose of evaluation, we use simulations in an OMNeT++ based simulator called Castalia v3.2 [21]. Castalia simulator has a good implementation of temporal variations of the physical channel and is modifiable in the simulator. We measured the RSSI values between each BN node and the BNC in a round-robin fashion, during each activity in an indoor environment. The BN nodes were sending a dummy packet of 40 bytes to the BNC every 10 ms (optimal for measuring the impact of human activities in RSSI [3]), with which the BNC calculates the RSSI and PDR, while the human was performing different activities. These RSSI values are translated to temporal variations as mentioned in [21]. In addition to our measured data, we also used the data from [3] and [22] to simulate the temporal variations of the physical channel caused due to the human activities. The packet error rate is calculated after the packet is received at the receiver.

For the simulation of MAC protocols we used the PHY parameters as listed in Table II. We simulated 5 BN nodes which are located in two limbs, thigh, chest and back-shoulder, and one BNC connected in a star topology, as shown in Figure 3, which is suitable for most of the medical applications [17]. Note that the thigh node is a potential relay node between the ankle BN node and the BNC. We vary the data generation rate for each node in order to simulate the heterogeneity of the IBSN nodes that generate the medical data at different rates, for example a blood-glucose sensor and a heart rate sensor. In order to compare the performance DiNAMAC with existing MAC protocols, we simulate RL-MAC which has no relay mechanism but uses RL framework to adapt to the dynamic traffic load [13], and HACMAC [2], which uses a priori knowledge of the disruption window, adapts the duty cycle and establishes relay nodes for different disruption patterns as the need be for a given disruption. All the MAC are evaluated in the same physical channel described in [15]. We use specifications of the Microsemi ZL70102 radio [23] to evaluate the power consumption of the BNC, BN, and BNR nodes. However, it has to be noted that the actual power consumption depends on the hardware design of the sensor nodes and the overheads caused by the additional hardware to support the functioning of the wireless radio.

| Parameter | Value | |
|---|---|---|
| Channel bandwidth | 300 | KHz |
| Reception current | 7 | mA |
| Transmission current (main) | 11 | mA |
| Idle current (main) | 3 | mA |
| Sleep current | 3 | $\mu$A |
| Slot time (TDMA) | 500 | $\mu$s |
| Number of retries (CSMA) | 10 | – |
| $N_{bits}$ Packet size | 40 | bytes |
| $\mathcal{R}$ Data rate | 20 | Kbps |
| $\alpha$ Learning rate | 0.1 | – |
| $\gamma$ Discount factor | 0.5 | – |
| $\eta$ Weighing factor | 5.5 | – |

TABLE II: Simulation parameters

## VII. Results

Simulation results related to reliability and power consumption are shown in Figure 5-9. Figure 5 illustrates the average PDR of the network calculated on the BNC node. The RL-MAC without relay mechanism has the lowest success in terms of packet-delivery ratio because of lack of any mechanism to compensate for the dynamic disruption in the link quality. Optimization framework of RLMAC focuses on improving the throughput by rewarding the mac agent based on the number of packets queued at the transmitter. Figure 5 proves that rewarding based on the packets queued will not work efficiently in a highly disruptive environment. At high data generation rates, the RL-MAC has around 20% packet loss.

HACMAC protocol predicts ON time based on a priori knowledge of the disruption to compensate for the periodic activities like walking and running. Although it outperforms the RL-MAC at lower data generation rates, as it goes higher, the HACMAC is under performing than the RL-MAC. This is due to the lack of any mechanism to cope up with higher data generation rates and limitation of priori knowledge of the disruption patterns. However, it still has $15\%$ packet loss at higher data rates, which is not desirable for IBSN.

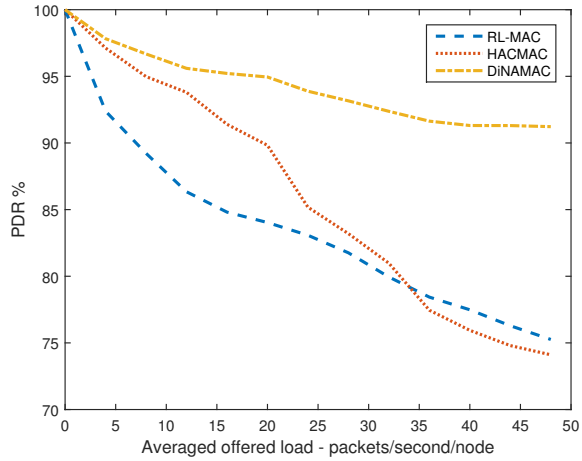


Fig. 5: Offered load vs Packet delivery ratio

In contrary, DiNAMAC has the lowest packet loss of less than $10\%$ even at higher data generation rates, due to the fact that it compensates for the disruption in link quality by an active learning mechanism, which includes packet error ratio as the observable metric. Since packet error rate is a function of modulation technique and bandwidth used by the transmitter, the higher data generation rates are implicitly accounted for in the learning mechanism.

Application level latency is the difference in time at which the packet is generated and the time at which packet is received at the receiver. HACMAC has no learning mechanism, which fails to handle the latency at higher data generation rate in a disruptive environment. In addition, if the disruption patterns are not known to HACMAC, adaptability of the MAC protocol fails. This results in lot of re-transmission before the packet is received. Such re-transmission tremendously increases the latency, and in cases of higher data generation rates, the latency is worst and clearly visible in Figure 6. In addition, it under performs RL-MAC when higher data generation rates. RL-MAC has constant increase in latency for higher data generation rates as the disruption in link quality changes dynamically. On average, the latency is lower than HACMAC for higher data generation rates and higher for smaller generation rates. It is clear from the results that the reward function of DiNAMAC is highly optimizing the MAC agent such that it achieves the lowest latency when compared with HACMAC and RL-MAC. Also, from Figure 6, it is observed that the prediction of disruption and learning rate is optimal, as the latency stays constant, after considerable packets are generated.

The power consumption evaluation results of the MAC protocols are demonstrated in Figure 7- 9. The power consumed is calculated by measuring the periods at which the BNC, BN and BNR are transmitting, receiving, idly listening (radio is turned on but not in receive mode) or sleeping (radio is turned off). Note that, each activity has different power consumption as mentioned in Table II. The power consumption of the BNC is greatly reduced in DiNAMAC due to the fact that it does not carry out additional transmission to support the optimization framework. DiNAMAC

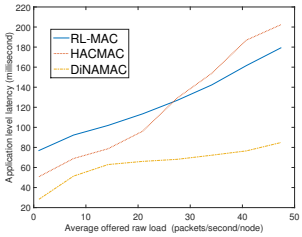


Fig. 6: Offered load vs Application level latency of IBSN

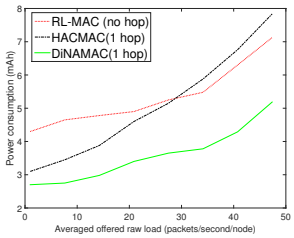


Fig. 7: Power consumption - BNC

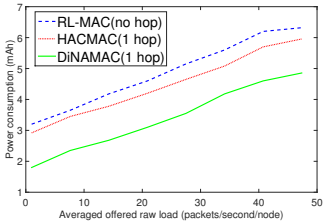


Fig. 8: Power consumption - BN

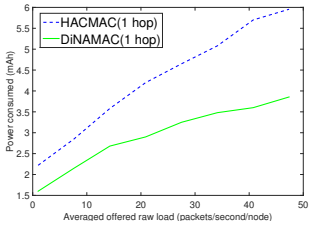


Fig. 9: Power consumption - BNR

benefits from the continuous and rigorous learning rate of the Q-learning, which is enables the BNC of the DiNAMAC framework to have the lowest power consumption when compared to HACMAC and RL-MAC.

Moreover, HACMAC requires the BNC extra receiving periods to effectively decide between adapting the duty cycle or establishing a relay node. This gets worse if the disruption features are not known and data generation rates of the BN are higher. Also, at lower data generation rates, the power consumption is less, because HACMAC adapts faster to the changing time. Although the complexity of the HACMAC grows at higher data generation rates and causes high energy consumption, in lower data generation rates and HACMAC is energy efficient than RL-MAC.

In RL-MAC there is no mechanism to compensate for the disruption in link quality at all, which results in a higher number of retries in prolonged attenuation periods, than the rest of the MAC protocols. This results in higher power consumption. However, in average, power consumption is reduced at higher data generation rates, because the optimization framework learns to improve the throughput and energy-efficiency.

The power consumption of BN node in RL-MAC is higher because of the additional compensation that the MAC agent has to undergo for the failed transmission. The client node is partly responsible for handling the congestion in network traffic, which requires the BN node to be in the receiving mode and monitor the data flow from neighbouring nodes. In HACMAC the BN nodes do not perform additional compensations; however, the BN nodes have to perform a number of re-transmissions in the case of unknown disruption patterns. The difference between power saving mechanisms in HACMAC and RL-MAC on the BN nodes is quite small, because both of them require directly or indirectly to be aware of the network conditions. In the case of RL-MAC, BNC is solely responsible for handling the network traffic

and disruptive link quality. As seen from previous results, the robust learning framework eliminates the need for BN to re-transmit the data packets. This enables the BN node to carry out the normal transmission, which makes the BN node of DiNAMAC have the least power consumption than HACMAC and RL-MAC.

We used HACMAC and DiNAMAC with 1 hop extended star-topology to compare the BNR power consumption. RL-MAC has no relay mechanism so that we cannot compare the power consumption of the relay nodes. In DiNAMAC the BNC selects the BNR nodes by calculating the link quality after each super-frame duration which enables the BNR nodes to be in idle mode in most of the super-frame duration, and BNR enters the receive mode only in the scheduled time slot. However, allocated time window of BNR can overlap with disruption window resulting in a large number of failed transmitted packets. This will require BNR to re-transmit not only their own packets, but also the relayed packets. Thus power consumption of BNR increases largely in higher data generation rates. In DiNAMAC, the intuitive learning process creates state-action policy to be efficient to considerably reduce the number of re-transmissions required by the BNR to achieve the maximum throughput. It is evident from the results, the BNR of DiNAMAC has a lower power consumption.

## VIII. CONCLUSION

We presented the DiNAMAC protocol that is able to handle the disruption in link quality caused due to different human activities. In addition, our DiNAMAC protocol is model-free so that it is practical for most real-world applications. DiNAMAC also complies with the IEEE 802.15.6 regulations. In particular, DiNAMAC optimizes the MAC protocol by using a reinforcement learning technique to incrementally learn the optimal action policy with regards to dynamic environment, through updating the long-term value of packet delivery ratio. Hence, DiNAMAC guarantees a high reliability of wireless communication in the presence of shadowing effect. We compared the performance of baseline MAC protocols such as HACMAC in terms of packet loss through simulation, and demonstrated the need to handle the dynamic disruption in link quality caused by different kinds of human-activities. The power consumption of the wireless radio decreases greatly with DiNAMAC when compared with that of RL-MAC, and HACMAC.

## ACKNOWLEDGMENT

This research is funded by STW project Cyber Physical Systems (CPS).

## REFERENCES

[1] Biotronik, "Home monitoring," Accessed: 2017-03-18. [Online]. Available: https://www.biotronik.com/en-us/products/services/home-monitoring

[2] V. R. Karuppiah Ramachandran, P. J. M. Havinga, and N. Meratnia, "Hacmac: A reliable human activity-based medium access control for implantable body sensor networks," in *Proceedings of the 13th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2016, San Francisco, CA, U.S.A.* IEEE Computer Society, 2016, pp. 383–389.

[3] F. Di Franco, I. Tinnirello, and Y. Ge, "1 hop or 2 hops: Topology analysis in body area network," in *Networks and Communications (EuCNC), 2014 European Conference on*. IEEE, 2014, pp. 1–5.

[4] N. E. Roberts, S. Oh, and D. D. Wentzloff, "Exploiting channel periodicity in body sensor networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 1, pp. 4–13, March 2012.

[5] V. R. Karuppiah Ramachandran, E. D. Ayele, N. Meratnia, and P. J. M. Havinga, "Potential of wake-up radio-based mac protocols for implantable body sensor networks (ibsn)—a survey," *Sensors (Switserland)*, vol. 16, 2016.

[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 2012, vol. 1, no. 1.

[7] W. Ye, J. Heidemann, and D. Estrin, "An Energy-Efficient MAC Protocol for Wireless Sensor Networks," vol. 00, no. c, pp. 1567–1576, 2002.

[8] N. F. Timmons and W. G. Scanlon, "An adaptive energy efficient mac protocol for the medical body area network," in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on*. IEEE, 2009, pp. 587–593.

[9] H. Li and J. Tan, "Heartbeat-driven medium-access control for body sensor networks." *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 14, no. 1, pp. 44–51, Jan. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19726272

[10] M. a. Huq, E. Dutkiewicz, and R. Vesilo, "MEB MAC: Improved channel access scheme for medical emergency traffic in WBAN," *2012 International Symposium on Communications and Information Technologies (ISCIT)*, pp. 371–376, Oct. 2012.

[11] C. Pandana and K. J. R. Liu, "Near-optimal reinforcement learning framework for energy-aware sensor communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 788–797, April 2005.

[12] "Ieee standard for local and metropolitan area networks - part 15.6: Wireless body area networks," Feb 2012, pp. 1–271.

[13] Z. Liu and I. Elhanany, "Rl-mac: A qos-aware reinforcement learning based mac protocol for wireless sensor networks," in *2006 IEEE International Conference on Networking, Sensing and Control*, 2006, pp. 768–773.

[14] S. Galzarano, A. Liotta, and G. Fortino, *QL-MAC: A Q-Learning Based MAC for Wireless Sensor Networks*. Cham: Springer International Publishing, 2013, pp. 267–275.

[15] V. R. Karuppiah Ramachandran, N. Meratnia, K. Zhang, and P. J. M. Havinga, "Towards implantable body sensor networks - performance of mics band radio communication in animal tissue," in *Proceedings of the 10th EAI International Conference on Body Area Networks, BODYNETS 2015, Sydney, Australia*. New York: ACM, September 2015.

[16] O. Georgiou, C. P. Dettmann, and J. P. Coon, "Network connectivity: Stochastic vs. deterministic wireless channels," in *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 77–82.

[17] V. R. K. Ramachandran, B. J. van der Zwaag, N. Meratnia, and P. Havinga, "Implantable body sensor network mac protocols using wake-up radio x2014; evaluation in animal tissue," in *2015 9th International Symposium on Medical Information and Communication Technology (ISMICT)*, March 2015, pp. 88–92.

[18] M. Vallejo, J. Recas, P. G. del Valle, and J. L. Ayala, "Accurate human tissue characterization for energy-efficient wireless on-body communications," *Sensors*, vol. 13, no. 6, pp. 7546–7569, 2013. [Online]. Available: http://www.mdpi.com/1424-8220/13/6/7546

[19] M. N. Islam and M. R. Yuce, "Review of medical implant communication system (mics) band and network," {*ICT*} *Express*, vol. 2, no. 4, pp. 188 – 194, 2016, special Issue on Emerging Technologies for Medical Diagnostics.

[20] "Ieee standard for local and metropolitan area networks - part 15.6: Wireless body area networks," Feb 2012, pp. 1–271.

[21] NICTA, "Castalia simulator," Accessed February 2016. [Online]. Available: http://castalia.npc.nicta.com.au

[22] R. D'Errico and L. Ouvry, "Time-variant ban channel characterization," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*. IEEE, 2009, pp. 3000–3004.

[23] Microsemi, "Zl70102 full datasheet," Accessed February 2016. [Online]. Available: http://www.microsemi.com