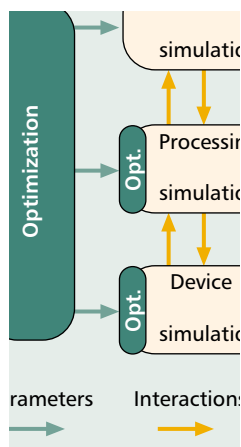


LOW-ENERGY WIRELESS COMMUNICATION NETWORK DESIGN

WAYNE STARK, HUA WANG, ANDREW WORTHEN, STÉPHANE LAFORTUNE, AND DEMOSTHENIS TENEKETZIS, UNIVERSITY OF MICHIGAN



Energy-efficient wireless communication network design is an important and challenging problem. To optimize performance one must account for the coupling among several subsystems and simultaneously optimize their operation under an energy constraint.

ABSTRACT

Energy-efficient wireless communication network design is an important and challenging problem. Its difficulty lies in the fact that the overall performance depends, in a coupled way, on the following subsystems: antenna, power amplifier, modulation, error control coding, and network protocols. In addition, given an energy constraint, improved operation of one of the aforementioned subsystems may not yield better overall performance. Thus, to optimize performance one must account for the coupling among the above subsystems and simultaneously optimize their operation under an energy constraint. In this article we present a generic integrated design methodology that is suitable for many kinds of mobile systems and achieves global optimization under an energy constraint. By pointing out some important connections among different layers in the design procedure, we explain why our integrated design methodology is better than traditional design methodologies. We present numerical results of the application of our design methodology to a situational awareness scenario in a mobile wireless network with different mobility models. These results illustrate the improvement in performance that our integrated design methodology achieves over traditional design methodologies, and the trade-off between energy consumption and performance.

INTRODUCTION

Energy-efficient wireless communication network design is an important and challenging problem. It is important because mobile units operate on batteries with limited energy supply. It is challenging because there are many different issues that must be dealt with when designing a low-energy wireless communication system (e.g., amplifier design, coding, modulation design, resource allocation, and routing strate-

gies), and these issues are coupled with one another. Furthermore, the design and operation of each component of a wireless communication system present trade-offs between performance and energy consumption. *The key observation is that constraining the energy of the nodes in a wireless network imposes a coupling among the design components that cannot be ignored in performing system optimization.* Therefore, the challenge is to exploit the coupling among the various components of a wireless communication system and understand the trade-off between performance and energy consumption in each individual component/subsystem in order to come up with an overall integrated system design that satisfies an energy constraint and has optimal performance with respect to some performance metric. Traditional design methodologies that optimize each layer separately may not be appropriate in terms of overall system optimality. The purpose of this article is to present a methodology for the design, simulation, and optimization of wireless communication networks that achieves maximum performance under an energy constraint. The presentation of our methodology also gives some insight as to why traditional design methodologies may not achieve overall system optimality. The integrated design methodology is applied to two scenarios of mobile ad hoc networks. The results show that significant gains are possible with an integrated design approach over traditional designs.

Before we proceed, we illustrate through simple examples the coupling among the different components of a wireless communication system, and highlight the trade-off between performance and energy consumption at individual components of the system. To illustrate the coupling among different components of a wireless communication system, we first need to describe some key features of the amplifier's operation.

Consider the design and operation of an amplifier. The amplifier boosts the power of the intended transmitted signal so that the antenna can radiate sufficient power for reliable communication. However, typical power amplifiers have maximum efficiency in converting DC power into RF power when the amplifier is driven into saturation. In this region of operation, the amplifier voltage transfer function is nonlinear. Because of this nonlinearity,

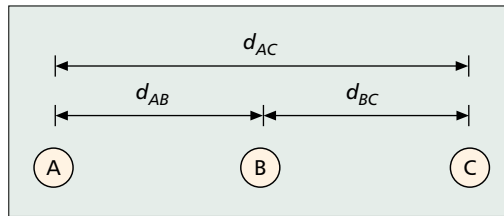
This research was supported by the Department of Defense under the Multidisciplinary University Research Initiative (MURI) on "Low-Energy Electronic Design for Mobile Platforms" and managed by the Army Research Office (ARO) under grant DAAH04-96-1-0377.

the amplifier generates unwanted signals (so called intermodulation products) in the band of the desired signal and in adjacent bands. When the amplifier drive level is reduced significantly (large backoff), the amplifier voltage transfer characteristic becomes approximately linear. In this case it does not generate intermodulation products. However, with large backoff the amplifier is not able to efficiently convert DC power into RF power. Thus, there is considerable wasting of power at low drive levels, whereas at high drive levels the amplifier generates more interfering signals.

We can now illustrate the coupling among individual components arising in the design of a wireless system. Consider packet routing in a wireless network that contains no base stations (i.e., an ad hoc network). For simplicity, consider a network with nodes A, B, and C, as shown in Fig. 1. If node A wants to transmit a message to node C, it has two options: Transmit with power sufficient to reach node C in a single transmission, or transmit first from A to B with smaller power, and then from B to C. Since the received signal power typically decays with distance as d^α , for α between 2 and 4, there is significantly smaller power loss due to propagation in the second option because $d_{AC}^\alpha > d_{AB}^\alpha + d_{BC}^\alpha$. However, even though node A transmits with smaller output power, it does not necessarily proportionally decrease the amount of power actually consumed because of the amplifier's effect discussed above. Furthermore, besides the energy required for packet transmission, there are energy requirements for packet reception and information decoding. The probability of packet error that is achieved depends on the energy allocated to the receiver. Thus the optimal network protocol (direct transmission from A to C or routing from A to B to C) depends on the amplifier characteristics as well as the energy needed to demodulate and decode a packet. Consequently, there is a coupling among amplifier design, coding and modulation design, decoding design, and routing protocols.

To highlight the tradeoff between energy consumption and system-wide performance, consider the *situational awareness problem* in a mobile wireless network. In this problem, the objective of each node is to be aware of the position of every other node during a given time period. If energy consumption is ignored, and the overall performance metric is the average (over all nodes and over time) position estimation error, this error is minimized when all nodes continuously communicate their positions with one another. Such a strategy requires significant energy. If, on the other hand, the objective is to minimize the average position estimation error under an energy constraint, the nodes will have to jointly decide when to communicate and whom to communicate to during the given period, since a continuous communication strategy would use all the available energy too quickly and could lead to large average position estimation error subsequent to the energy depletion of the battery.

Traditional design methodologies for wireless communication systems that attempt to optimize each layer separately may achieve



■ Figure 1. Three nodes in a network.

global system optimality only by coincidence. However, through an understanding of the interactions and coupling among the functions at the different layers, it is possible to design a wireless communication system in a manner that truly integrates the functions of all layers. Therefore, we propose a methodology that decomposes the system into coupled layers and exploits the interactions among them to come up with an energy-efficient design. The goal of the decomposition, besides better understanding of the design procedure for global system optimality, is to obtain a computationally tractable approach to quantifying system performance with respect to different optimization criteria. Tackling such a problem can be a formidable task. We are not aware of any previous design and optimization attempts that encompass all the layers. Most previous research [1–6] on low-energy ad hoc mobile wireless networks focuses on the optimization at the component/subsystem level. We hope our work will provide some guidelines for further research in global system optimization.

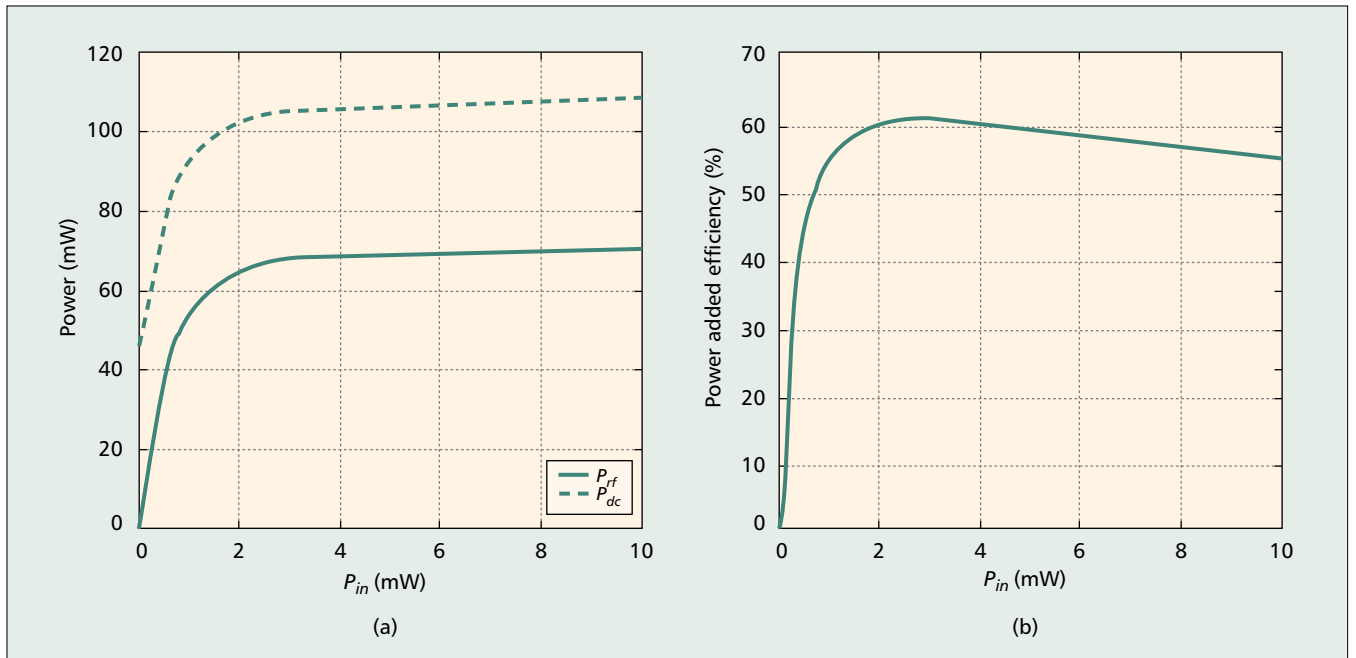
The remainder of the article is organized as follows. We first present a methodology for system design that incorporates the effect of the different layers on system performance. This methodology is fairly general and can be applied to many different applications besides the situational awareness scenario we consider later in this article. We next describe the component models for the amplifier, propagation, coding, modulation, and network protocols for the system under investigation. After that we explain how global optimization works together with each system layer and present optimization results for the situational awareness application. We conclude the article with a summary and discussion about potential applications.

SYSTEM DESIGN METHODOLOGY

We first describe the system decomposition and optimization, both of which form the constituent parts of our design methodology. We then comment on the decomposition and optimization. We consider a wireless network consisting of mobile nodes that need to communicate with one another in order to take some action or to share information, such as their respective positions. The overall goal is to characterize and optimize some performance metric under an energy constraint.

As pointed out in the first section, in order to develop a systematic and computationally tractable design methodology, we divide the problem into interacting design layers as shown in Fig. 2. The system decomposition consists of

The systemwide objective is to optimize a performance metric that reflects the collective operation of the mobile units under an energy constraint. This is achieved by simultaneously optimizing over a number of parameters that characterize the objective.



■ Figure 3. Characteristics of the power amplifier: a) radiated and DC power; b) power added efficiency.

A DESIGN EXAMPLE: MODELS FOR SYSTEM DECOMPOSITION

We apply our integrated design methodology to a particular network design problem, namely, the situational awareness problem in wireless mobile networks. In the situational awareness problem a number of mobile nodes desire to keep track of the location of each other over some time duration. The nodes operate with batteries and thus have a finite energy constraint. The transmission of information by a node requires a certain amount of energy as does the processing of any received signal. The goal of the design is to minimize the mean absolute error of the position estimates. There is a plethora of parameters that could be considered for optimization. We focus on a small set of parameters to illustrate the design and simulation methodology. In addition, we describe the system decomposition and justify our choice of the coupling parameters among different layers. We present the system optimization in a later section. We proceed to describe each layer in a bottom-up manner.

THE DEVICE LAYER

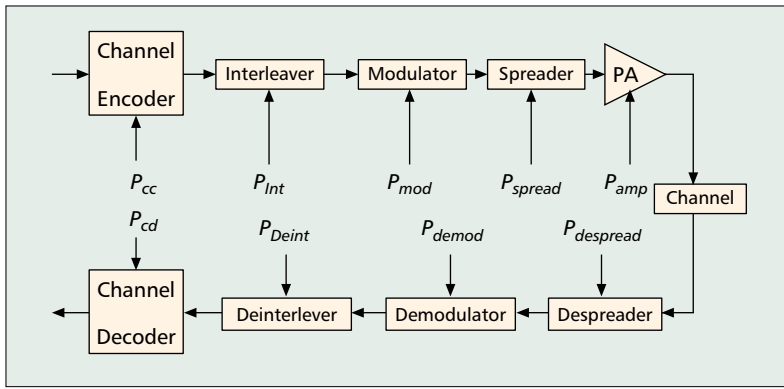
We present the model for the device layer and the coupling parameters between the device and processing layers. We justify why the coupling parameters we choose are appropriate for the wireless communication systems under investigation. While not all the components of a transmitter and a receiver have been considered in the model of this article, we have chosen a few parameters that have an important effect on the coupling among different layers and illustrate the trade-off between performance and energy consumption.

At the device layer, we assume each node has an omnidirectional dipole antenna and a small power amplifier. Because the power amplifier is a major source of energy consumption and our global objective is to achieve high-precision situ-

ational awareness (i.e., low estimation errors) for every node under an energy constraint, it is important to understand the role of the amplifier power added efficiency in the overall optimization problem. Let P_{in} denote the input power, P_{rf} the radiated power, and P_{dc} the consumed DC power. The power added efficiency is defined as

$$\text{Power added efficiency} = \frac{P_{rf} - P_{in}}{P_{dc}}. \quad (1)$$

The characteristics of the power amplifier [7] shown in Fig. 3 are tabulated for use at the processing layer. From Fig. 3 we see that the input-output relation of the power amplifier is fairly linear when the input power is small; however, the amplifier operates at a very low efficiency. When the input power is large, the amplifier operates at higher efficiency, but with large input-output nonlinearity. This nonlinearity generates in-band and out-of-band signals called *intermodulation signals*, which adversely affect the performance of the processing layer, which in turn indirectly affects the design of the network layer protocols. In the current literature [2, 6] on energy-efficient routing protocols, the transmitted energy (power) is usually chosen as the routing metric. Unfortunately, this does not correspond to the actual consumed energy with most amplifiers. Within the context of our problem, to capture the operation of the amplifier and the coupling among the device layer and other higher layers, we consider the following parameters associated with an amplifier's operation: the total consumed power P_{tot} , the output power P_{out} , and the AM-to-AM voltage characteristics [7]. We note here that the intermodulation interference also depends on the modulation scheme chosen and consider a nonconstant modulation scheme. For such a modulation scheme, we characterize the rela-



■ Figure 4. A processing layer block diagram.

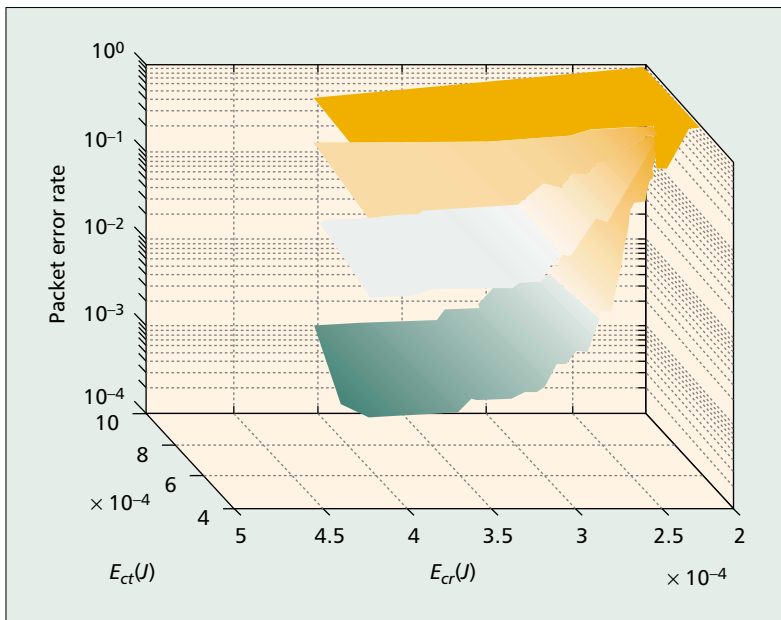
tion between the average amplifier output power and the energy constraint E_{ct} for transmitting a packet by

$$P_{out} = g_1(E_{ct}). \quad (2)$$

This relation is tabulated for use by higher layers. In certain situations it is possible that the actual consumed energy at the transmitter, E_{ta} , is less than the constraint on the consumed energy at the transmitter. In this case we define a function $E_{ta} = g_2(E_{ct})$ that maps the energy constraint to the actual energy.

THE PROCESSING LAYER

We describe the model used at the processing layer, discuss the coupling between the device and processing layers, and between the processing and network layers. More details about the specific design choices can be found in [8]. Our goal is not to model every part of this system but to understand the coupling between layers and the trade-off between performance (global and local) and energy. We first describe the basic block diagram of the processing layer, and discuss the performance-energy trade-offs of several elements in the block diagram. Finally, we



■ Figure 5. P_e as a function of E_{ct} and E_{cr} for convolutional code.

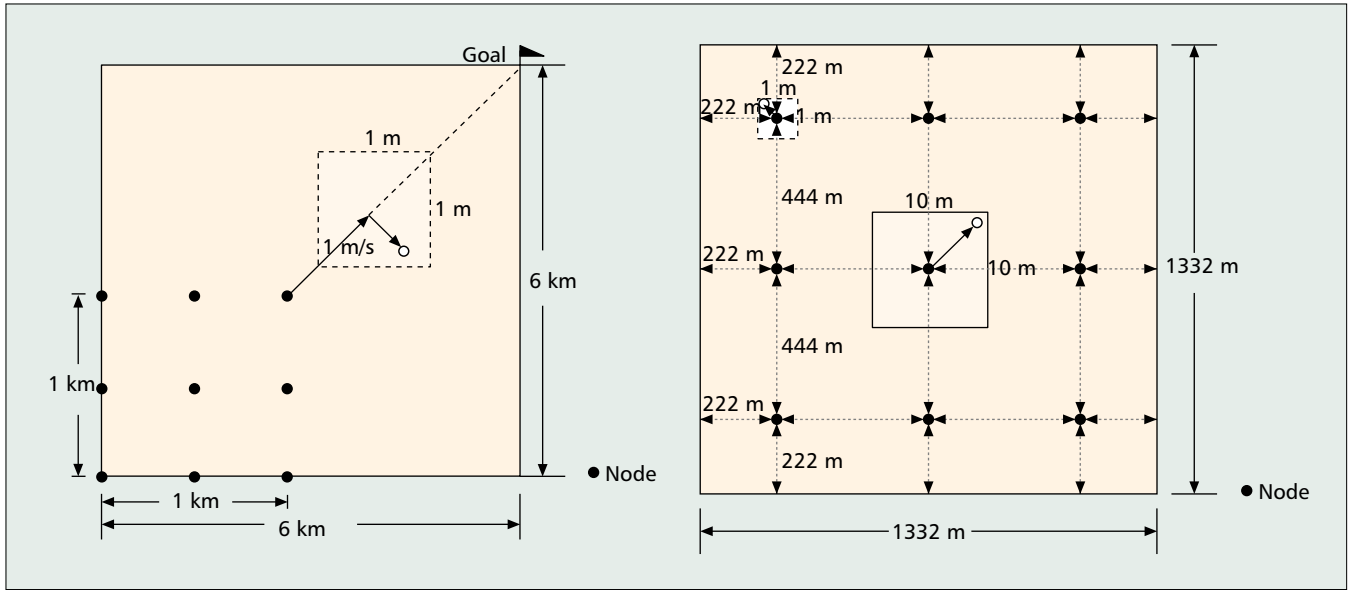
describe precisely the optimization at the processing layer, and the coupling between the processing and network layers.

The basic block diagram of the processing layer is shown in Fig. 4. A block of information is presented to the channel encoder. The channel encoder adds redundancy to protect against channel errors. The output of the channel encoder is interleaved, modulated, and spread in bandwidth. The resulting signal is amplified by a power amplifier (PA) and transmitted. At the receiver the inverse operations are needed to accurately recover the block of information. While each of these operation described consumes power in order to process the data or signal, we focus on the energy being consumed by the power amplifier, demodulator, and channel decoder. We focus on these elements since generally they consume much more energy than other elements in the system. The performance-energy trade-off of the amplifier was discussed previously. Thus, we focus on the performance-energy trade-off of the demodulator and decoder.

In the system we consider the modulation is binary phase shift keying with raised-cosine pulses. In order to demodulate the signal, filtering of the received signal is necessary. Typically, this filtering is done digitally with an oversampled signal. The architecture of the filter is a finite-impulse response filter implemented with a tapped delay line. The taps are multiplied by coefficients and then summed. The amount of power consumed while doing this operation depends on the number of bits used to represent the signals and the coefficients in the multiplication. As the number of bits used increases, the performance becomes closer to the ideal case of analog operation while consuming more energy. We let N_E denote the number of bits used in the filter for data bits and quantization bits, and optimize the performance over N_E .

The error control coding technique we consider is convolutional encoding with maximum likelihood (Viterbi) decoding. Error control coding can significantly reduce the required transmitted energy needed for reliable communication at the expense of decreased bandwidth efficiency and energy needed for decoding. In attempting to recover the transmitted data from the demodulator output, the channel decoder must perform certain correlations and do comparisons. The performance of the decoder (like the demodulator) depends on the number of bits used to represent the signal at the output of the demodulator (N_D). The larger number of bits used, the better the performance. However, the more bits used in the representation, the more energy consumed by the decoder. We optimize the performance over N_D .

To calculate the energy needed to process (demodulate and decode) signals we model the individual digital circuits used in the algorithm. We assume complementary metal oxide semiconductor (CMOS) circuits in which the main contribution to energy is attributed to the charging and discharging of parasitic capacitors that occur during logical transitions [9]. Circuit models for the critical parts of the demodulator and decoder have been developed and energy con-



■ Figure 6. Mobility models.

sumption calculated as a function of the number of levels of quantization. We let E_{ra} denote the energy consumed by the receiver in order to process a packet. This is a function of the number of bits of quantization used by the demodulator, N_E , and the decoder N_D [10, 11]:

$$E_{ra} = g(N_D, N_E).$$

The performance measure that couples the processing layer with the network layer is the packet error probability, P_e . In general, for any choice of coding and demodulation schemes, P_e depends on the energy constraint E_{ct} for the transmitter to send a packet, the energy constraint E_{cr} for the receiver to process a packet, the received signal-to-noise ratio (SNR), SNR , the number N_E bits of quantization for equalizer data input and equalizer coefficients, and the number N_D bits of quantization for decoding. The parameters N_E and N_D affect only the performance of the processing layer. Consequently, in accordance with our methodology:

- For given E_{ct} , E_{cr} , and SNR , we locally optimize P_e and BER with respect to N_E and N_D

$$\begin{aligned} P_e &= \min_{N_D, N_E} f(N_D, N_E, SNR, E_{ct}, E_{cr}) \\ &= g_3(SNR, E_{ct}, E_{cr}). \end{aligned} \quad (3)$$

- We generate parameterized version of P_e with respect to E_{ct} , E_{cr} , and SNR , and build a performance table for these parameterized versions of P_e . In addition, because of the integer constraint on quantization levels in the demodulator and decoder, the actual energy consumed by the receiver, E_{ra} may be less than the constraint on energy E_{cr} . Let N_E^* (SNR, E_{ct}, E_{cr}) and N_D^* (SNR, E_{ct}, E_{cr}) be the optimum number of quantization levels in the demodulator and decoder, respectively. Thus, the actual energy consumed in the receiver is a function of the constraint on the energy and SNR,

$$E_{ra} = g(N_E^*, N_D^*) = g_4(SNR, E_{ct}, E_{cr}). \quad (4)$$

The network layer (and global optimization) utilizes the table of P_e as a function of E_{ct} , E_{cr} , and SNR for calculating its global performance.

Figure 5 shows the locally optimized packet error rate as a function of the transmitter energy constraint and receiver processing energy constraint. The surfaces are, from top down, $SNR = 1$ dB, 2 dB, 3 dB, and 4 dB.

In summary, at the processing layer a table is generated from simulation that contains the performance as a function of the energy allowed at the transmitter, the energy allowed at the receiver, and the SNR at the receiver. This table is then used at the network layer to determine the overall performance of the system.

NETWORK LAYER

We describe the model for the network layer and the parameters of the network protocols that affect global performance. We explain why these parameters are appropriate for the wireless communication systems under investigation.

We consider a network of nine nodes. All nodes move according to a specific mobility model. In the situational awareness problem, each node attempts to keep track of the positions of all the other nodes. This is accomplished by communication and estimation. All nodes share their respective position information according to a specific communication protocol. In the rest of this section we present the mobility models, propagation models, communication protocols, and estimation schemes used by the nodes. We refer the reader to [8] for further technical details regarding these various models.

Mobility Models — We describe two mobility models that we use in the various optimization problems we consider. In both mobility models, each node in the network moves to a new location at the end of every T_m s. In the examples considered in this article $T_m = 1$.

In mobility model 1, we consider a region of

Variable	Meaning	Value	Unit
λ_c	Carrier wavelength	10	m
h_t	Height of transmitter antenna	1	m
h_r	Height of receiver antenna	1	m
G_t	Gain of transmitter antenna	1	–
G_r	Gain of receiver antenna	1	–
d	Propagation distance	(0, 9000)	m
P_t	Transmitted power	$[10^{-7}, 1]$	W
P_r	Received power	(Eq. 5)	W

■ **Table 1.** Variables used in wireless system modeling.

size 6 km \times 6 km and a group of nine nodes initially deployed in an area as shown in Fig. 6a. All nodes travel toward the same destination which is located at $\underline{G} = (6000, 6000)$ m. Each node travels at average speed v , where $v = 1\text{m/s}$ (not drawn to scale in Fig. 6a). At each step, each node's motion is subject to a random disturbance in x and y coordinates.

In mobility model 2, all nodes are initially deployed in a region of size 1332 \times 1332 m and move within this region as shown in Fig. 6b. The mobility of each node is characterized by a two-state discrete-time Markov chain as shown in Fig. 7, where the two states are labeled *stay* and *move*. In each of these states, each node's motion is subject to random disturbance in the x and y coordinates, where the disturbance is parameterized by a scaling factor of 1 m in the stay state and 10 m in the move state.

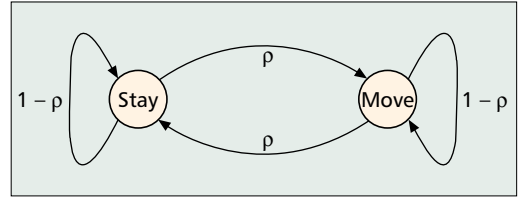
Propagation Model — The transmitted signal from each node experiences propagation loss and fading. In the results that follow, we assume a two-path propagation model. Table 1 shows the variables and their typical values needed to explain the model.

The two-path propagation model consists of a direct path and a path reflected off the ground with 180° phase change at the reflection point from the transmitter to the receiver. The cumulative effect of both paths determines an attenuation A between received power and transmitted power, which is given by [12]

$$A = \frac{P_r}{P_t} = 4 \left(\frac{\lambda_c}{4\pi d} \right)^2 G_t G_r \sin^2 \left(\frac{2\pi h_t h_r}{\lambda_c d} \right) \approx G_t G_r \frac{h_t^2 h_r^2}{d^4}, \quad (5)$$

where the approximation is valid when $d \gg \max\{h_t, h_r\}$. The two-path propagation model characterizes the large-scale propagation loss of many fading channels reasonably well, which is why we adopt it in our network layer simulation.

Communication Protocols — We now describe the medium access control (MAC) strategy as well as the routing protocols. We consider two communication protocols: a single-hop transmission



■ **Figure 7.** The Markov chain for mobility model 2, $\rho = 0.05$.

protocol (which may be considered a single-hop routing protocol) and a multihop routing protocol. For both communication protocols, the omnidirectionality of the antenna at each node makes the potential connections among nodes point-to-multipoint, that is, if a node sends out a packet, the electromagnetic wave will propagate in all directions and may be received by many other nodes. Therefore, in the design of wireless communication protocols, communication occurs in a broadcast medium, which is very different from traditional wired networks, where the connections are mostly point-to-point. This omnidirectionality property of the antenna can be exploited in the situational awareness scenario under consideration.

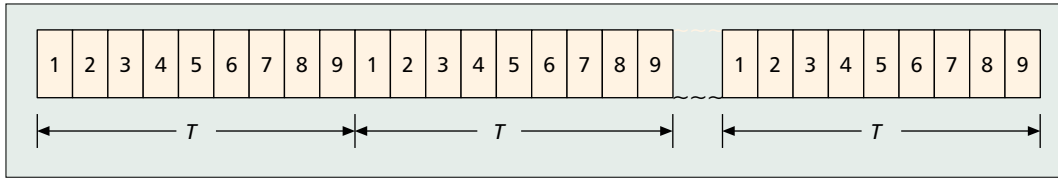
Single-Hop Transmission Protocol — In the single-hop transmission protocol, each node transmits its position information packets every T s, where T is a design parameter. The MAC is time-division multiple access (TDMA), where each node is assigned a transmission slot of duration T/N , where $N = 9$ in our case, as shown in Fig. 8. The slot duration is much larger than a packet duration. In a given slot, each packet transmission is followed, with probability q , by a retransmission, and so forth, until the slot ends. The retransmission probability q is considered a design parameter.¹ Since each node operates on a battery with limited capacity, we constrain the energy used for each packet transmission or retransmission to be upper bounded by E_{ct} , which we consider a design parameter. The position information packets may be received by many other nodes, each of which consumes a certain amount of energy to process the packets. Again, due to the limited capacity of the battery, we constrain the energy used to process the incoming packets to be upper bounded by E_{cr} , which we consider a design parameter. When a node receives a packet, it does not send back any acknowledgment, nor does it forward the packet it receives to other nodes. As a consequence, every packet in the single-hop transmission protocol travels only one hop.

In summary, we choose T , q , E_{ct} , and E_{cr} as the design parameters at the network layer that affect global performance.

Multihop Routing Protocol — In the multihop routing protocol, there are mainly two issues:

- Setting up a routing path
 - Transmitting position information packets
- For routing, we adapt the Open Shortest Path First (OSPF) protocol [13, 14], which is a link state routing protocol, to our situational awareness scenario and take advantage of the omnidi-

¹ This simple retransmission strategy was chosen because more complex automatic repeat request (ARQ) schemes are not well suited to the broadcast environment under consideration.



■ Figure 8. TDMA for single-hop transmission protocol.

rectionality of the antenna. The objective of our routing algorithm is to determine the first-hop nodes a node should reach, and the power this node should use to reach these first-hop nodes when it transmits position information packets originated by itself or forwards position information packets received from other nodes. Each node uses a fixed power to update the link state of the network every T_r s, which is chosen as a design parameter. Based on the link state information of the network, the routing algorithm calculates the required transmission power for the position information packets. Let d_{max} be the largest distance among the distances between a node and any of its first-hop nodes. Since the antenna is omnidirectional and the source node knows its farthest first-hop node, it can use this knowledge by choosing the radiated power for position information packets to be Cd_{max}^A , where C is chosen as a design parameter. For transmitting position information packets, each node transmits position information packets every T_p s using power Cd_{max}^A , where T_p is chosen as a design parameter. In order to receive either a routing protocol packet or a position information packet, a node consumes energy E_{cr} to process it. A packet is decoded incorrectly with a probability that depends on C , E_{cr} , and SNR .

In summary, we choose T_r , T_p , C , and E_{cr} as the design parameters at the network layer that affect global performance.

Estimation Schemes — We consider two estimation schemes corresponding to the two mobility models we described earlier. In both estimation schemes, a node, say node i , updates its estimate $\hat{w}_k^{(i,j)}$ of node j 's position $w_k^{(j)}$ at time kT_e , where $k = 1, 2, \dots$. In the examples considered in this article $T_e = 2$.

For mobility model 1, the estimates are based on the following strategy. Node i knows the mobility model for all the other nodes. Since according to the mobility model the nodes move toward the goal in a straight line subject to noise, the new estimate is the extrapolation toward the goal of the position contained in the packet that was last received correctly, by an amount proportional to the product of velocity and time.

For mobility model 2, we use the following estimation scheme. Node i 's estimate of node j 's position at time kT_e is the position of node j contained in the last correctly received packet by node i . Let τ denote the time of that reception. Since the mean of the increment of node j 's position at any time is zero, if node i does not correctly receive any packet between time τ and kT_e , the estimate of node j 's position at time kT_e is node j 's position at time τ .

PERFORMANCE METRIC

For the purpose of optimization, we use mean absolute error as the performance metric. For both estimation schemes, the estimation error of node j 's position made by node i at time kT_e is defined as

$$e_k^{(i,j)} = w_k^{(j)} - \hat{w}_k^{(i,j)}, \quad (6)$$

where $w_k^{(j)}$ is the actual position of node j at time kT_e . The performance metric $J(i)$ at node i is defined to be

$$J(i) = E \left[\frac{1}{K(I-1)} \sum_{j=1, j \neq i}^K \sum_{k=1}^K \|e_k^{(i,j)}\| \right], \quad (7)$$

where KT_e is the time horizon under consideration. In the above equation, the expectation is with respect to the mobility, the noise in the receiver, and the randomness in retransmission (in the case of single-hop transmission protocol only).

The overall network performance measure is given by the average of the position estimation error contributed by all the nodes in the network:

$$J = \frac{1}{I} \sum_{i=1}^I J(i). \quad (8)$$

The goal is to minimize J over the parameters that affect global performance subject to a constraint on the energy used by each node. Let $E(i)$ denote the energy used by node i over the time horizon KT_e . The constraint on energy is

$$\max_{1 \leq i \leq I} E(i) \leq E. \quad (9)$$

Based on the discussion in earlier sections, we have the following objectives: For the single-hop transmission protocol the goal is to determine the design parameters

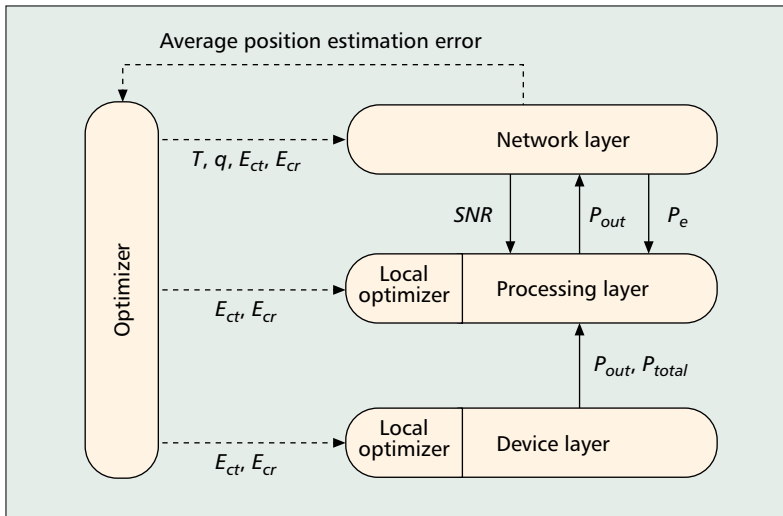
$$\begin{aligned} & [T^*, q^*, E_{ct}^*, E_{cr}^*] \\ & = \arg \min_{\substack{[T, q, E_{ct}, E_{cr}] \\ \max E^{(i)} \leq E}} J(T, q, E_{ct}, E_{cr}), \end{aligned} \quad (10)$$

and the corresponding performance $J^* = J(T^*, q^*, E_{ct}^*, E_{cr}^*)$. For the multihop routing protocol the goal is to determine the design parameters

$$\begin{aligned} & [T_r^*, T_p^*, C^*, E_{cr}^*] \\ & = \arg \min_{\substack{[T_r, T_p, C, E_{cr}] \\ \max E^{(i)} \leq E}} J(T_r, T_p, C, E_{cr}) \end{aligned} \quad (11)$$

and the corresponding performance $J^* = J(T_r^*, T_p^*, C^*, E_{cr}^*)$.

The two-path propagation model characterizes the large-scale propagation loss of many fading channels reasonably well and this is the reason why we adopt it in our network layer simulation.



■ Figure 9. Coupling of different layers.

DESIGN EXAMPLE: GLOBAL OPTIMIZATION

We illustrate how our methodology, described earlier, applies to a wireless system in the situational awareness scenario. We consider a single-hop transmission protocol and mobility model 1. Similar procedures apply to the other settings. The parameters that describe the coupling among the layers and must be shared by the different layers in the setting under consideration are shown in Fig. 9.

The integrated design methodology of an earlier section is applied in an iterative fashion. At each iteration the optimization is, in part, simulation-based because we do not have precise analytical expressions for the local and global optimization criteria we employ. The optimization program attempts to find the global minimum of the objective function J in Eq. 8. The global optimization and simulation modules perform the following steps in attempting to find the globally optimal solution:

- Step 1: The optimizer module determines the (new) parameters $[T, q, E_{ct}, E_{cr}]$ for which the network performance is to be evaluated.
- Step 2: The network simulator module approximates the objective function in Eq. 8 for the given $[T, q, E_{ct}, E_{cr}]$ using Monte Carlo simulation techniques. It returns the average position estimation error to the optimizer module.
- Step 3: Steps 1 and 2 are repeated until a terminating condition is reached.

The optimizer module used in step 1 is a type of *simulated annealing* algorithm. The method of *simulated annealing* is a technique that has attracted significant attention since it is suitable for optimization problems with a large number of parameters, especially ones where a desired global extremum is hidden among many local extrema. The simulated annealing algorithm used in our integrated design is called *Hide-and-Seek* and was developed by Romeijn and Smith [15]. The steps of the Hide-and-Seek algorithm are given in [8].

We have implemented the network simulator module in OPNET, a widely used network development and analysis tool [16]. Our network simulator module has the following steps that

involve the interactions among the three layers for the above global optimization of step 2:

- Step 2.1: Given the parameters SNR, E_{ct} , and E_{cr} selected by the optimizer module, the device layer determines the amplifier output power P_{out} , the actual consumed energy E_{ta} for transmitting a packet, and the actual consumed energy E_{ra} for receiving a packet:

$$P_{out} = g_1(E_{ct}), \quad (12)$$

$$E_{ta} = g_2(E_{ct}), \quad (13)$$

$$E_{ra} = g_4(SNR, E_{ct}, E_{cr}). \quad (14)$$

The function g_1 captures the result of the local optimization of the amplifier model at the device layer, which is parameterized in terms of E_{ct} , as explained earlier. The resulting P_{out} is used at the network layer (see step 2.2 below). We assume that the energy used by the transmitter is always all of E_{ct} , and therefore $g_2(E_{ct}) = E_{ct}$. Due to the quantization at the demodulator and the decoder at the receiver, the actual consumed energy E_{ra} may be smaller than the energy constraint E_{cr} ; their relation is determined by the function g_4 that is well defined and known.

- Step 2.2: For each transmission scheduled at the network layer and for each receiver, the network layer determines the SNR at the output of the receiving antenna as follows. The transmitted power P_t is the product of the amplifier output power and transmitting antenna efficiency

$$P_t = P_{out}\eta_t. \quad (15)$$

The power received at the output of the receiving antenna is

$$P_r = \eta_r A P_t, \quad (16)$$

where η_r is the receiving antenna efficiency and A is the channel attenuation given by Eq. 5. The SNR is

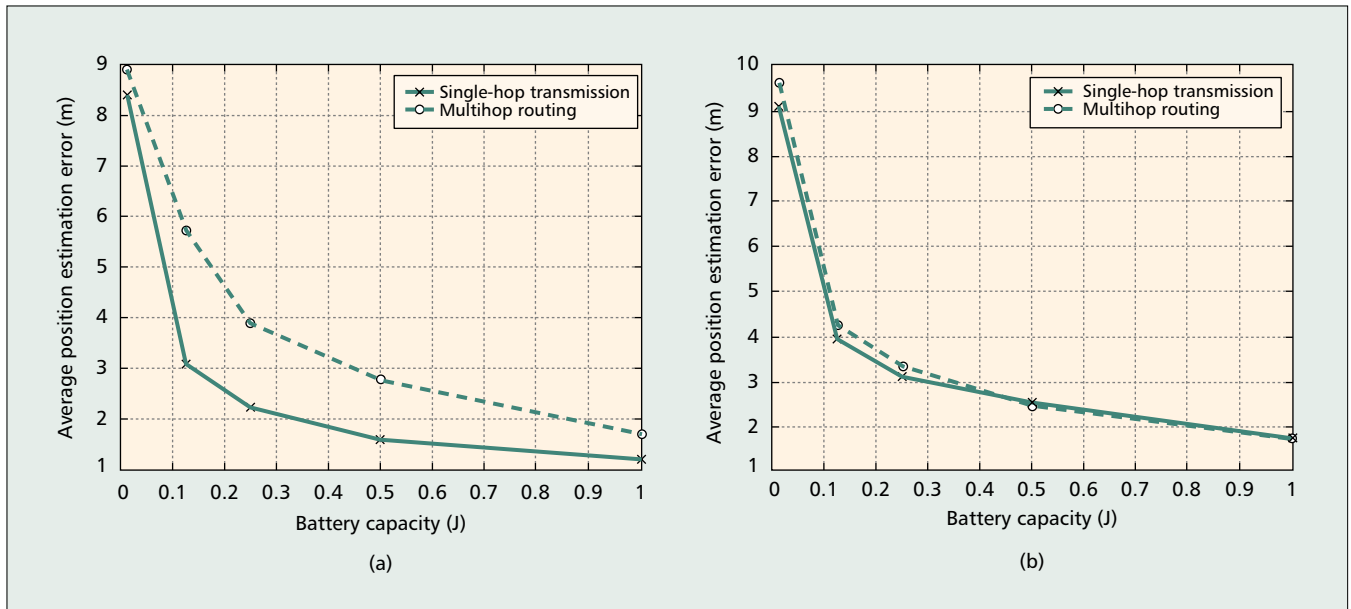
$$SNR = \frac{P_r \cdot T_s}{N_0}, \quad (17)$$

where T_s is the symbol duration and N_0 is the power spectral density of the thermal noise at the receiver. The remaining variables in the above equations are defined in Table 1.

- Step 2.3: For each transmission scheduled at the network layer and for each receiver, the processing layer determines the packet error rate P_e using SNR obtained from the network layer (step 2.2) and using E_{ct} and E_{cr} obtained from the optimizer module. P_e is obtained at the processing layer from E_{ct}, E_{cr} , and SNR by solving an optimization problem with respect to the parameters N_E and N_D as described in a previous section. For the sake of computational efficiency, this optimization at the processing layer is done offline and its results are summarized using the function g_3 whose arguments are E_{ct}, E_{cr} , and SNR :

$$P_e = g_3(SNR, E_{ct}, E_{cr}). \quad (18)$$

Therefore, the network simulator module implements this step by table lookup. For a given P_e , the network layer flips a biased coin



■ Figure 10. Performance comparison of different algorithms with different drop areas: a) 1 km × 1 km; b) 2 km × 2 km.

to determine if each packet is correctly received.

- Step 2.4: Each node uses the estimation model described earlier to update its estimates of the position of the other nodes. The network simulator module accumulates the position estimation errors of all nodes in the network. When the simulation terminates, the network layer averages the accumulated value over all nodes and over time. Thirty network simulation runs are produced and their average is the approximation of the objective function in Eq. 8 that is returned to the optimizer module.

In Step 3, the termination condition that we chose for our experiments was to stop after 200 iterations of Steps 1 and 2.

It is critical to understand that the simulation and optimization effort has been carefully divided into a device layer simulation, a processing layer simulation, and a network layer simulation. More importantly, the interactions between layers have been identified and incorporated into the performance evaluation.

PERFORMANCE RESULTS

We present numerical results for two situational awareness scenarios that highlight our integrated design methodology and compare its performance to that of traditional design methodologies. In particular, within the context of mobility model 1, we use our design methodology to compare the performance of a single-hop transmission protocol with that of a multihop routing protocol. Furthermore, within the context of mobility model 2 and single-hop transmission, we compare our design methodology with traditional design methodologies and illustrate the improvement in performance achieved by our approach.

INTEGRATED DESIGN FOR MOBILITY MODEL 1

First we consider mobility model 1 where the nodes are initially located in either a 1 km × 1 km or 2 km × 2 km area and move toward a sin-

gle goal. We evaluate and optimize the objective function given in Eq. 8 by Monte Carlo simulation using the steps described earlier. The resulting performance is shown in Fig. 10. As can be seen in Fig. 10, the single-hop transmission protocol does better than the multihop routing protocol when the nodes are initially dropped in a 1 km × 1 km area. When the nodes are dropped in a 2 km × 2 km area, the multihop routing protocol does almost the same as the single-hop transmission protocol. We expect that the multihop routing protocol outperforms the single-hop transmission protocol for larger drop areas when the battery capacities are large.

Many factors come into play to explain why single-hop transmission does better than multihop routing in the case of the smaller drop area. Among these, we mention propagation, amplifier efficiency, and routing overhead. We briefly discuss each of these factors.

Propagation: Propagation loss becomes much less significant at small distance due to our propagation model, where the signal strength degrades proportionally to the fourth power of the propagation distance. In particular, if the distance between a pair of nodes is increased from 1 to 2 km, the signal attenuation will be increased by a factor of 16. Thus, as the distances between nodes becomes larger, the effect of propagation loss on performance grows dramatically.

Amplifier: For small distances between nodes it is possible for the transmitter to reach all nodes in the network. Transmitting to nodes very close does not save much energy compared to transmitting to the farthest node since the amplifier will be operating in the region of low efficiency. Thus, the energy consumed does not decrease proportionally to the decrease in desired output energy, as discussed in the example in the first section. So by routing messages through very close nodes, the amount of energy saved does not increase proportionally to the decrease in power loss from transmission.

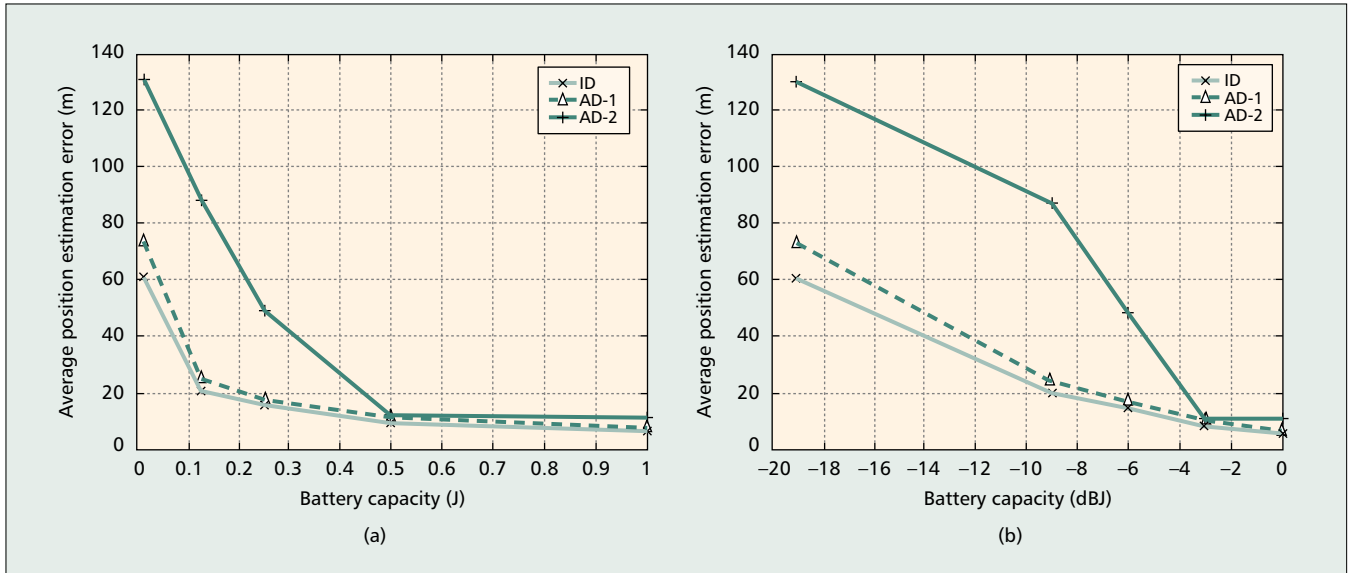


Figure 11. Performance comparison of different design methodologies: a) normal scale; b) log scale in battery capacity.

Routing overhead: In multihop routing a certain amount of energy is needed to update routing tables, which is not needed in single-hop transmission.

In summary, all these effects play a role in determining the overall system performance shown in Fig. 10. While we have made these conclusions for a very specific set of parameters and models, we believe that when the nodes are close to each other single-hop transmission is generally more efficient than multihop routing. What changes is the threshold (in terms of distance) where one strategy is better than the other.

Finally, it is of interest to understand where energy is being consumed in these systems. Within the context of the single-hop transmission protocol, about 25 percent of the total consumed energy is used for transmitting packets in both cases of drop areas ($1 \text{ km} \times 1 \text{ km}$ and $2 \text{ km} \times 2 \text{ km}$), and the other 75 percent is used for receiving packets. Within the context of the multihop routing protocol, about 20 percent of the total consumed energy is used for transmitting both position information packets *and* routing protocol packets when the drop area is $1 \text{ km} \times 1 \text{ km}$; this percentage changes to about 30 percent when the drop area is $2 \text{ km} \times 2 \text{ km}$. When the multihop routing protocol is in use, depending on the different energy constraints, about 60–80 percent of the total consumed energy is spent on transmitting *and* receiving packets that are only used for updating routing tables. The remaining energy is used for transmitting and receiving position information packets.

MERITS OF INTEGRATED DESIGN

To determine the merits of the integrated design (ID) methodology, we compare its performance with two different traditional designs applied to the single-hop transmission protocol. In the first design, called AD-1, the optimization of the processing layer is done independent of the optimization at the network layer, but the network layer uses the results of the optimization at the processing layer. We can view this as a one-way

coupling between the processing and network layers. In the second design, called AD-2, the two layers are designed totally independently.

Alternative Design 1 — In alternative design 1 (AD-1), we partially decouple the optimization by imposing a constraint on the packet error probability for the transmission between the two most distant nodes. Let SNR_f be the SNR between the two most distant nodes. In our scenarios, this is when the nodes are at the two diagonally opposite corners of the region under consideration. We first consider the following optimization problem at the processing layer:

$$\begin{aligned} & \left[\hat{E}_{ct}, \hat{E}_{cr} \right] \\ & = \arg \min_{[E_{ct}, E_{cr}]} E_{ct} + E_{cr} \quad (19) \\ & \quad P_e(E_{ct}, E_{cr}, SNR_f) \leq 10^{-2} \end{aligned}$$

The goal in this first step is to minimize the total energy (transmitter and receiver) needed for the longest possible transmission distance in order to maintain a packet error probability of 0.01. The next step is to optimize the performance metric given by Eq. 8 over the network parameters using the results of the optimization at the processing layer. Specifically, the goal is to determine

$$\left[\tilde{T}, \tilde{q} \right] = \arg \min_{[T, q]} J(T, q, E_{ct}, E_{cr}) \quad (20)$$

$$\max E^{(i)} \leq E$$

In order to compare AD-1 with the integrated design approach, recall Eq. 10 for the integrated design, which we restate here for convenience:

$$\begin{aligned} & \left[T^*, q^*, E_{ct}^*, E_{cr}^* \right] \\ & = \arg \min_{[T, q, E_{ct}, E_{cr}]} J(T, q, E_{ct}, E_{cr}) \quad (21) \\ & \quad \max E^{(i)} \leq E \end{aligned}$$

Comparing Eq. 21 with Eqs. 20 and 19 reveals that

$$J(T^*, q^*, E_{ct}^*, E_{cr}^*) \leq J(\tilde{T}, \tilde{q}, \hat{E}_{ct}, \hat{E}_{cr}); \quad (22)$$

thus, the performance of the integrated design is at least as good as AD-1.

Alternative Design 2 — In AD-2, we completely decouple the optimization at the network and processing layers. In AD-2, we proceed as follows:

- We optimize E_{ct} and E_{cr} as in AD-1.
- We select the parameter values $T^o = 60$ s and $q^o = 0.01$ at the network layer, without doing any optimization.

Design AD-2 is consistent with many traditional design methodologies where network layer parameters are selected based on engineering and application considerations, without any explicit optimization that would account for the processing and device layers. Therefore, the performance of AD-2 is $J(T^o, q^o, E_{ct}, E_{cr})$, and it is clear that

$$\begin{aligned} J(T^*, q^*, E_{ct}^*, E_{cr}^*) \\ \leq J(\tilde{T}, \tilde{q}, \hat{E}_{ct}, \hat{E}_{cr}) \leq J(T^o, q^o, \hat{E}_{ct}, \hat{E}_{cr}). \end{aligned} \quad (22)$$

Numerical Results — The position error for designs AD-1, AD-2, and ID are shown in Fig. 11, within the context of mobility model 2. These results show a degradation in performance when AD-1 is used compared to ID. In particular, we see from Fig. 11b that there is a degradation of about 2 dB in energy when the average position estimation error is about 60 m. The performance of AD-2 is also shown in Fig. 11. The degradation in performance of AD-2 compared to ID is quite significant. This is evident when we plot the energy in dB units (with a reference of 1 J) in Fig. 11b. In particular, about 12 dB more energy is required to obtain the same average position estimation error when the average position estimation error is about 60 m.

By comparing the plots for the different designs, it is clear that at relatively high position estimation error (on the order of 60 m) it is critical to take into account the battery capacity when optimizing the network protocol design. That is, doing the optimization of the network protocol parameters T , q and letting these depend on the battery capacity in the global optimization results in significant gain over fixed network protocol parameters. For low battery capacity, low duty cycle should be used in transmitting position information packets.

On the other hand, the optimization of the physical layer subject to a constraint on error probability that is independent of battery capacity, is close to the optimal design. However, this may not always be the case in other scenarios or if a different error probability constraint is chosen. On the other hand, the optimal design does not require significantly more computation than the one-way coupling design, AD-1. Essentially, AD-1 and AD-2 perform some type of intuitive guessing of design parameters independent of battery capacity, while the ID approach optimizes the parameters dependent on battery capacity.

We caution that the above observations about the performance differences of different designs are for the specific scenario we consid-

ered. Clearly, the performance gain achieved with the integrated approach compared to other approaches will vary from scenario to scenario. More simulation experiments, for a range of scenarios, mobility models, and channel models, are needed to quantify more completely the benefits of an integrated design and optimization strategy.

CONCLUSION

We propose an integrated design methodology and applied it to the optimization of the situational awareness problem in ad hoc mobile wireless networks. We give evidence of why the integrated design methodology outperforms other design methodologies that do not account for or exploit coupling among layers. This evidence is supported by simulation experiments. Since the optimization and simulation at the processing and device layer are done offline, the complexity and scalability of integrated design are almost the same as those of network layer design. Therefore, integrated design can be applied to any network layer design as long as such network layer design is feasible. In future research, it would be of interest to classify other cases where an integrated design approach leads to large performance gains over traditional approaches.

ACKNOWLEDGMENTS

The results presented in this article were obtained in the course of a comprehensive research project on Low-Energy Electronics Design for Mobile Platforms that started at the University of Michigan in 1996 and was supported by the Department of Defense (DoD) MURI managed by ARO. Many individuals at the University of Michigan have contributed to this project and to numerous discussions that led to the elaboration of the design methodology and associated simulation experiments presented in this article. The authors wish to extend special thanks to the following individuals for their help at different stages of this work: Olivier Boivineau, Vuk Borich, Jack East, Riten Gupta, Al Hero, Sangjin Hong, Tara Javidi, Tingfang Ji, Linda Katehi, Seksan Kiatsupaibul, Il-Suek Koh, Paul Liang, JongTae Lim, Christopher Lott, Pinaki Mazumder, David Neuhoff, Troy Nolan, Kamal Sarabandi, John Shumpert, Robert Smith, Kimberly Wasserman, and Do-Sik Yoo. In addition, the authors acknowledge the helpful comments of Jim Freebersyser, Jim Harvey, Jim Mink, and Bill Sander.

REFERENCES

- [1] J. H. Chang and L. Tassiulas, "Energy Conserving Routing in Wireless Ad-hoc Networks," *INFOCOM*, 2000, pp. 22–31.
- [2] R. Ramanathan and R. Rosales-Hain, "Topology Control of Multihop Wireless Networks using Transmit Power Adjustment," *Proc. IEEE INFOCOM*, 2000.
- [3] V. Rodoplu and T. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE JSAC*, vol 17, no. 8, Aug. 1999.
- [4] K. Scott and N. Bambos, "Routing and Channel Assignment for Low Power Transmission in PCS," *IEEE ICUPC*, 1996, pp. 498–502.
- [5] S. Singh, M. Woo, and C. S. Raghavendra, "Power-Aware Routing in Mobile Ad Hoc Networks," *Proc. ACM/IEEE MOBICOM*, Oct. 1998, pp. 181–90.
- [6] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the Construction of Energy-Efficient Broadcast and Multicast

When the multihop routing protocol is in use, depending on the different energy constraints, about 60–80 percent of the total consumed energy is spent on transmitting and receiving packets that are only used for updating routing tables. The remaining energy is used for transmitting and receiving position information packets.

In future research, it would be of interest to classify other cases where an integrated design approach leads to large performance gains over traditional approaches.

- Trees in Wireless Networks," *IEEE INFOCOM*, Mar. 2000.
- [7] V. Borich et al., "Nonlinear Effects of Power Amplification on Multicarrier Spread Spectrum Systems," *Int'l. Microwave Symp.*, vol. 1, 1998, pp. 323–26.
- [8] W. Stark et al., "Low Energy Wireless Communication Network Design," Univ. of MI EECS CSPL tech. rep., 2001.
- [9] J. M. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Prentice Hall PTR, 1995.
- [10] S. Hong and W. E. Stark, "Power Consumption vs. Decoding Performance Relationship of VLSI Decoders for Low-Energy Wireless Communication System Design," *Proc. IEEE Int'l. Conf. Elec., Cir. Sys.*, 1999, pp. 1593–96.
- [11] S. Hong and W. E. Stark, "Design and Implementation of a Low Complexity VLSI Turbo-Code Decoder Architecture for Low Energy Mobile Wireless Communications," *J. VLSI Signal Processing*, Feb. 2000, pp. 43–57.
- [12] T. Rappaport, *Wireless Communications: Principles and Practice*, Englewood Cliffs, NJ: Prentice Hall, 1996, pp. 69–122, 139–96.
- [13] J. Moy, "OSPF Version 2," Internet RFC 2328, Apr. 1998.
- [14] J. Moy, "Multicast Extensions to OSPF," Internet RFC 1584, Mar. 1994.
- [15] H. E. Romeijn and R. L. Smith, "Simulated Annealing for Constrained Global Optimization," *J. Global Optimization*, vol. 5, 1994, pp. 101–26.
- [16] OPNET Technologies, "OPNET Modeler Manual."

ADDITIONAL READING

- [1] D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice Hall, 1992.
- [2] S. Hong, S. Kim, and W. E. Stark, "Low Power Parallel-Pipeline Multiplier Design for DSP Applications through Coefficient Optimization," accepted to *Int'l. J. Custom-Chip Des., Sim., and Testing*, Nov. 2002.
- [3] J. G. Proakis, *Digital Communications*, McGraw-Hill, 2000.
- [4] E. M. Royer and C.-K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Pers. Commun.*, vol. 6, no. 2, Apr. 1999, pp. 46–55.
- [5] A. Worthen et al., "Performance Optimization of VLSI Transceivers for Low-Energy Communications Systems," *Proc. IEEE MILCOM*, vol. 2, Oct. 1999, pp. 1434–38.

BIOGRAPHIES

WAYNE STARK (stark@eecs.umich.edu) received B.S. (with highest honors), M.S., and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana. Since September 1982 he has been a faculty member in the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, where he is currently a professor. From 1984 to 1989 he was an associate editor of *IEEE Transactions on Communications*. He was selected by the National Science Foundation as a 1985 Presidential Young Investigator. His research interests are in the areas of coding and modulation for wireless communication systems.

HUA WANG received his B.E. and M.E. degrees in electronic engineering from Tsinghua University, Beijing, China. He is currently studying toward his Ph.D. degree in the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. His research at Tsinghua University focused on source coding algorithms and their implementations on DSP chips. His current research interests include channel coding and wireless communication networks.

ANDREW WORTHEN graduated with his B.S. (with highest honors), M.S., and Ph.D. degrees from the University of Michigan, Ann Arbor, where he held a National Science Foundation Graduate Research Fellowship and an AFCEA Postgraduate Fellowship. In 2001 he joined MIT Lincoln Laboratory, Lexington, Massachusetts. His research interests include coding theory, software-defined radios, and satellite communication systems.

STÉPHANE LAFORTUNE received a B.Eng. degree from Ecole Polytechnique de Montréal in 1980, an M.Eng. degree from McGill University in 1982, and a Ph.D. degree from the University of California at Berkeley in 1986, all in electrical engineering. Since September 1986 he has been with the University of Michigan, Ann Arbor, where he is a professor of electrical engineering and computer science. His research interests are in discrete event systems and communication networks. He co-authored, with C. Cassandras, the textbook *Introduction to Discrete Event Systems* (Kluwer, 1999).

STÉPHANE LAFORTUNE received a B.Eng. degree from Ecole Polytechnique de Montréal in 1980, an M.Eng. degree from McGill University in 1982, and a Ph.D. degree from the University of California at Berkeley in 1986, all in electrical engineering. Since September 1986 he has been with the University of Michigan, Ann Arbor, where he is a professor of electrical engineering and computer science. His research interests are in discrete event systems and communication networks. He co-authored, with C. Cassandras, the textbook *Introduction to Discrete Event Systems* (Kluwer, 1999).

DEMOSTHENIS TENEKETZIS is a professor of electrical engineering and computer science at the University of Michigan, Ann Arbor. In winter and spring 1992 he was a visiting professor at the Swiss Federal Institute of Technology, Zurich, Switzerland. Prior to joining the University of Michigan he worked for Systems Control Inc., Palo Alto, California, and Alphatech Inc., Burlington, Massachusetts. He received his diploma in electrical engineering from the University of Patras, Greece, and his M.S., E.E., and Ph.D. degrees, all in electrical engineering, from the Massachusetts Institute of Technology, Cambridge. His research interests are in stochastic control, decentralized systems, queuing and communication networks, stochastic scheduling and resource allocation problems, mathematical economics, and discrete event systems.