A Communication Infrastructure for Smart Environments – A Position Article[1]

Zygmunt J. Haas
Cornell University, School of Electrical and Computer Engineering
Wireless Networks Laboratory
323 Frank Rhodes Hall, Ithaca, NY 14853
e-mail*: haas@ee.cornell.edu*
http://www.ee.cornell.edu/~haas/wnl.html

Abstract: With the today's state-of-the-art technology, miniature sensors and actuators can be manufactured and integrated with electronics onto extremely small footprint devices. Such devices could be embedded into different platforms, turning the ambience into a highly controllable, supportive, and cooperative environment. Applications of this technology can be easily envision in nearly every aspect of our life; in a workplace, at home, in a supermarket, in a departmental store, in a car, on the battlefield, etc. Of course, to realize this vision, a network is necessary that enables communication among these devices. Since the main characteristic of these devices is that their power supply is extremely restricted and that the span of the network is limited, conventional algorithms and protocols need to be adapted fto this "micro" networking environment. This position paper addresses some key issues in creation of such a micro-network.

Keywords: Ad Hoc Networks, Smart Environment, Zone Routing Protocol,

1. Introduction

The technological advances in the area of miniaturization allow now to produce extremely small devices with volume on the order of tens millimeters cube. Moreover, it is anticipated that further reduction in size will be possible in the near future [Akin98] and [Asada98]. Small footprint devices manufactured with the Micro-Electro-Mechanical Systems (MEMS) technology can accommodate a variety of sensing functions, such as temperature, pressure, or acceleration, and actuation functions, such as rotation, linear displacement, or torque. These mechanical operations can be coupled with electronic functions, which may include on-board communication means. Inclusion of communications would allow these miniature devices not only to exchange and share information among the platforms on which these devices are mounted (e.g., people, animals, rotating machinery parts), but to form a network that implements distributed execution of algorithms. Algorithms executed in such a network interact with and within the environment in which the network is embedded, allowing environment-dependent processing and information sharing. In other words, with inclusion of even limited amount of memory and processing power on these micro-devices, an ad hoc environment can be devised, in which the participating entities may both contribute and receive services and information. The far-fetched vision is that of "intelligent powder," in which these miniature and lightweight devices can be instantaneously deployed and can operate autonomously in a concerted manner to carry out certain actions.

The micro-sensors/actuators, to which we refer to in this article, posses the raw ability to transmit a signal between two relatively close devices. To create a network, one must provide the nodes with the following basic functions: a medium access control, an addressing scheme, and a routing algorithm. Of course, these are the traditional issues in the design of a communication network. What is new here is that the "traditional" algorithms and protocols may not be applicable to the problem at hand, mainly because of the extremely limited computing, storage, and communication resources available in the micro-sensor nodes. Extremely low-battery power and limited amount of non-volatile memory lead to the requirement of very low complexity algorithms and protocols. In other words, "the name of the game" is to design a system with as simple protocols as possible that require as little communication among the network nodes. Additional characteristics of our network of miniature devices are the very large number of network nodes

[1] *This work has been sponsored in part by ONR contract no. N00014-00-1-0564 and AFRL contract no. F360602-97-C-0133*

(in the thousands) and the ability of the network nodes to move relative to the rest of the network. However, probably the most challenging of all is the requirement that the network should be deployable without pre-planning and should continue to operate as the carrying platform moves among different environments. Thus, another design attribute is that the protocols need to be versatile enough to adapt to the both, different and differing operational conditions of the network.

Interestingly, the problem of creating a network and supporting connectivity in this micro-networking environment is related to a problem that has been a subject of extensive research in the past 30 years - the *multi-hop* networks (e.g., the early 1970's effort of DARPA on Packet Radio NETwork technology and the more recent GLOMO program [GLOMO]). Under the name of *ad-hoc networks*, research of this type of communication has recently received renewed interest. The issues related to ad-hoc networks are being extensively addressed in the Internet Engineering Task Force, MANET Working Group [IETF-MANET]. Because of the similarities of the ad-hoc and the micro-sensor networks, MANET protocols could be adapted with some changes to create such a *smart communication environment.*

In this paper, we concentrate mainly on the issue of routing in a network that will support communication among a large collection of small-sized, possible mobile nodes, operating in an environment with continuously changing conditions. Because of its very small coverage area, we term such a network a *micro-network*.

2. What is an Ad-Hoc Network?

An ad-hoc network is a self-organizing wireless network made up of mobile nodes and requiring no fixed infrastructure. The limitations on power consumption imposed by portable wireless radios result in a nodal transmission range that is typically small relative to the span of the network. To provide communication throughout the entire network, nodes are designed to serve as relays if needed. The result is a distributed multi-hop network with a time-varying topology.

Because ad-hoc networks do not rely on existing infrastructure and are self-organizing, they can be rapidly deployed to provide robust communication in a variety of hostile environments. This makes ad-hoc networks very appropriate for providing tactical communication for military, law enforcement and emergency response efforts. Ad-hoc networks can also play a role in civilian forums, such as the electronic classroom, convention centers, and construction sites. Finally, ad hoc networks could also be used for creating communications among distributed collection of devices, such as sensors or actuators. With such a broad scope of applications, it is not difficult to envision ad-hoc networks operating over a wide range of coverage areas, nodal density and nodal velocity.

This potentially wide range of ad-hoc network operating configurations poses a challenge for developing efficient routing protocols. On one hand, the effectiveness of a routing protocol increases as network topology information becomes more detailed and up-to-date. On the other hand, in an ad-hoc network, the topology may change quite often, requiring large and frequent exchanges of data among the network nodes. This is in contradiction with the fact that all updates in a wireless communication environment travel over the air and are costly in resources, especially with energy-limited devices, as is the case in our micro-networking environment.

Existing routing protocols can be classified either as *proactive* or as *reactive*. Proactive protocols attempt to continuously evaluate the routes within the network, so that when a packet needs to be forwarded, the route is already known and can be immediately used. In contrast, reactive protocols invoke a route determination procedure on an on-demand basis. The reactive route discovery is usually based on a query-reply exchange, where the route query is flooded through the network to reach the desired destination. The advantage of the proactive schemes is that route information is readily available when needed, resulting in little delay prior to data transmission. In contrast, reactive schemes may produce significant delay in order to determine a route when routing information is needed, but not available.

Routing schemes, whether proactive or reactive, require some exchange of control traffic. This overhead can be quite large in ad-hoc networks, if the topology frequently changes. Reactive protocols produce a large amount of traffic by effectively flooding the entire network with route queries. The combination of excessive control traffic and long route query response time rule out pure reactive routing protocols for real-time communication applications. Pure proactive schemes are likewise not appropriate for ad-hoc networks, as they continuously use a large portion of the network capacity to keep the routing information current. Proactive protocols tend to distribute topological changes widely in the network, even though the creation/destruction of a new link at one end of the network may not be a significant piece of information at the other end of the network. Furthermore, since ad-hoc network nodes may move quite fast, and as the changes may be more frequent than the route requests, most of this maintained routing information is never used! This results in unacceptable waste of the network capacity. In our approach, which we term the *Zone Routing Protocol,* we create a hybrid, proactive-reactive scheme, in which the route discovery process is both fast and efficient in the amount of control traffic.

3. The Sensor Network Architecture

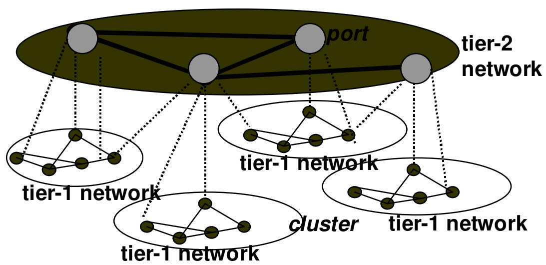

Fig. 1: A Two-Tier Hierarchical Network

The overall network architecture discussed here is that of a two-tiered network, as shown in Figure 1. The lower tier, tier-1, is composed of clusters of sensor devices, which can communicate among themselves in a peer-to-peer manner. In addition, the sensor devices can communicate with larger transceivers, which we term here *ports.* The main function of a port is to provide an interface between the "cloud" of the miniature sensor devices - the tier-1 network - and an external network or, among the "clouds" of sensor networks themselves. That is, a port interconnects two tier-1 networks if they cannot be directly connected due to transmission power limitations. A tier-1 node may not be able to communicate with any port or may be in communication with a single or multiple ports. Ports communicate among themselves by creating the tier-2 network. Tier-2 network can interface to an external network, with some of the ports serving as gateways.

Because it is assumed that tier 2 nodes are not battery power limited, the tier 2 network uses "conventional" wireless transmitter/receiver design. In contrast, tier-1 networks are transmission power restricted. Furthermore, communication among tier-1 nodes may be also based on passive communication, such as reflection of a modulated RF carrier. Thus, the tier 1 nodes communicate through multi-hop routing. Consequently, the issues related to the design of the network protocols in the two tiers are quite different. Here, we focus on the more challenging problem of communication within the tier 1 network, with the first issue that we discuss being the nodal addressing scheme.

4. Hardware-based Addressing

Traditional addressing function is done by assigning an address to every network node from a *logical* address space. However, logical addressing requires a fair amount of processing. An alternative is to use *physical addressing*, which could be implemented by using some characteristic of the transmitted signal that will allow selection of the intended receiver. As one example, we have considered to associate each node with a particular offset from the central frequency of the transmitted signal; this offset identifies the node. As it is possible to build very sharp miniaturized filters [Young98] and [Pehlke97], it is possible to create a large "address" space of thousands of nodes. Thus, to communicate between adjacent nodes, the source device needs to tune to the particular offset of its neighbor; i.e., to transmit to a node i, the transmitter tunes to frequency f_i, where

$$f_i = f_c - \left(\left\lfloor n/2 \right\rfloor - i \right) \cdot f_o; \qquad 0 \le i \le n,$$

and where f_c is the central frequency, f_o is the frequency offset between adjacent "addresses," which depends on the selectivity of the RF filter, and n is the size of the address space. Of course, this is just one example of physical addressing and there are many other possibilities as well.
Once the address scheme is in place, the next step in creating a network is for each node to learn about the existence of the other nodes around it.

5. On-demand Neighbor Discovery

In wireless packet data networks, neighbor discovery is usually done by some sort of a *beaconing* process, in which the nodes listen to periodically transmitted signals from their neighbors. Such an operation is rather wasteful in transmission power. A number of alternative solutions to the periodic beaconing process are possible. For instance, one approach would be to perform an *on-demand* neighbor discovery by a simple query, which is initiated by a node that needs to learn who are its neighbors. Of course, the design of such a neighbor discovery process has to be tightly coupled with the addressing scheme used.

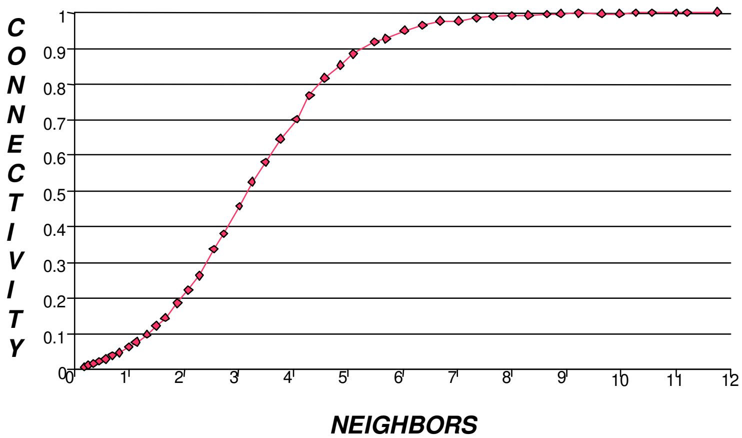

Fig 2: Connectivity vs.the number of neighbors

Another issue that needs to be addressed in this context is the number of neighbors that a node is designed to see. The number of neighbors is controlled by the transmission power of the nodes. On one hand, large number of neighbors is advantageous, since there is a smaller chance of loosing network connectivity; i.e., of creating a network partition. However, many neighbors also lead to larger transmission power of a node, because, for example, of the neighbor discovery process. Additionally, it also results in a larger number of transmissions and, thus, more transmission power. Furthermore, it also reduces the network throughput due to more contention among the accessing nodes. Our study of the on-demand neighbor discovery protocol suggests that 5, 6, 7, or 8 neighbors correspond to 90%, 95%, 98%, and nearly 100% connectivity, respectively (see Fig. 2). The actual choice of the number of neighbors is dependent on factors such as: the transmission power capabilities of the sensor devices, addressing and medium access control schemes used, and the required richness of routes between source and destination devices.

Once the communication between adjacent nodes is implemented, an end-to-end network routing scheme is required, as the distance between the two communicating nodes will be typically much larger than the transmission range of a single node. We propose to use a simple routing scheme, the *Zone Routing Protocol (ZRP)* that allows discovering routes within the network of a large number of micro-sensors nodes with little communication overhead and with little nodal storage requirements.

6. The Zone Routing Protocol – a Short Description

The wired Internet uses routing protocols based on topological broadcast, such as the OSPF [Moy94]. These protocols are not suitable for the micro-networking environment due to the relatively large bandwidth required for topological update messages.

Routing in multi-hop packet radio networks was based in the past on shortest-path routing algorithm [Leiner87], such as Distributed Bellman-Ford (DBF) algorithm [Bertsekas92]. These algorithms suffer from slow convergence rate (the "counting to infinity" problem). Besides, DBF-like algorithms incur large update message penalty. New multi-hop routing protocols were proposed ([Cheng89], [Johnson96], [Murthy95], [Park97], [Perkins94], and [Perkins99]). However, synchronization requirements, large storage, and extra processing overhead are common in these protocols, which may make them inappropriate for the micro-networking environment..

In contrast, our routing protocol [Pearlman99], based on the notion of *Routing Zones* incurs very low overhead in route determination. It requires maintaining minimal amount of routing information in each node and the cost in wireless resources for maintaining routing information of inactive routes is very small. The protocol identifies multiple routes to the destination (increasing reliability and performance), with no looping problems. However, the most appealing feature of the protocol is that its behavior is adaptive, based on the mobility and communication patterns of the micro-sensor devices. The *Zone Routing Protocol* limits the propagation of such information to the neighborhood of the change only, thus limiting the cost of topological updates. This feature is of particular interest in a large network, such as the micro-sensor network.

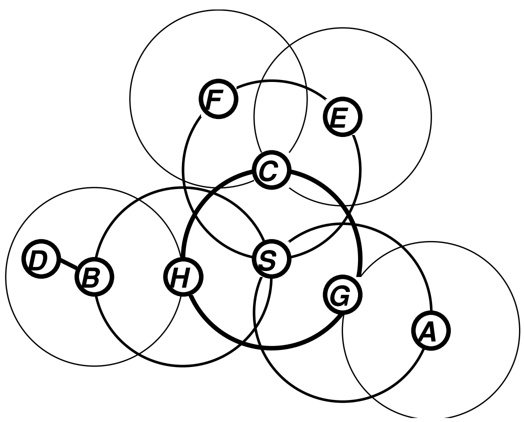

Fig 3: An example of ZRP operation

ZRP is based on the notion of a *routing zone*, which is defined for each node and includes the nodes whose distance in hops is at most some predefined number, which is referred to here as the *zone radius.* Each node is required to know the topology of the network within its routing zone only. Thus, nodes are updated about topological changes only within their routing zone. Therefore, in spite of the fact that a network can be quite large, the updates are only locally propagated. Since for radius greater than one the routing zones heavily overlap, the routing tends to be extremely robust. The routes within the network are specified as a sequence of nodes, separated by approximately the zone radius. We illustrate the operation of the *route discovery* procedure by an example in Fig. 3. The source sensor *S* attempts to send a packet to the destination sensor node *D.* To find a route within the network, *S* first checks whether *D* is within its routing zone.[2] If so, *S* knows the route to node *D.* Otherwise, *S* sends a query to all the nodes on the periphery of its zone[3]; that is, to nodes *C, G,* and *H.* Now, in turn, each one of these nodes, after verifying that *D* is not in its routing micro-zone forwards the query to its "peripheral" nodes. In particular, *H* sends the query to *B,* which recognizes *D* as being in its routing zone and responds to the query, indicating the forwarding path: *S-H-B-D.* The mechanism by which *B* learns about the forwarding path is the simple r*oute accumulation* protocol, by which each node that forwards the query appends its identity to the query message.

Because the routing zones heavily overlap, the route query will be forwarded to many network nodes, multiple times. In fact, it is very possible that the query will be forwarded to all the network nodes, effectively flooding the network. To prevent this from happening, the query termination and query forwarding strategies used in traditional flooding algorithms need to be properly extended for use in a routing zone architecture.

In order to understand the need for query control mechanism in ZRP, it is important to stress one of the key features of the routing zone: A node's response to a route query contains information about that node's entire routing zone. From this perspective, excess route query traffic can be regarded as a result of overlapping query threads (i.e. overlapping queried routing zones). Thus, the design objective of query control mechanisms should be to reduce the amount of route query traffic by steering threads outward from the source's routing zone and away from each other. This problem is addressed in ZRP from two different perspectives: thread overlap detection/termination and thread overlap prevention. We omit

[2] Recall that each node knows all the nodes and the routings within its routing zone.

[3] I.e., nodes that are zone-radius away

further discussion of these important mechanisms due to space limitations and refer the reader to [Pearlman99] and [Haas00].

The main advantage of the *ZRP* is in the fact that the number of "flood" messages to discover a route is significantly reduced from those of the other reactive-type protocols. This is so, because the route discovery queries leap between the "peripheral" zone nodes in quantum of the zone radius. An important aspect of the Zone Routing is that it discovers multiple routes to the destination. Finally, the route discovery process can be made even more efficient in resources, at the expense of longer latency. This could be done by sequentially, rather than simultaneously, querying the peripheral zone nodes, either one-by-one or in-groups. Thus, there is a tradeoff between the cost and the latency of the route discovery procedure.

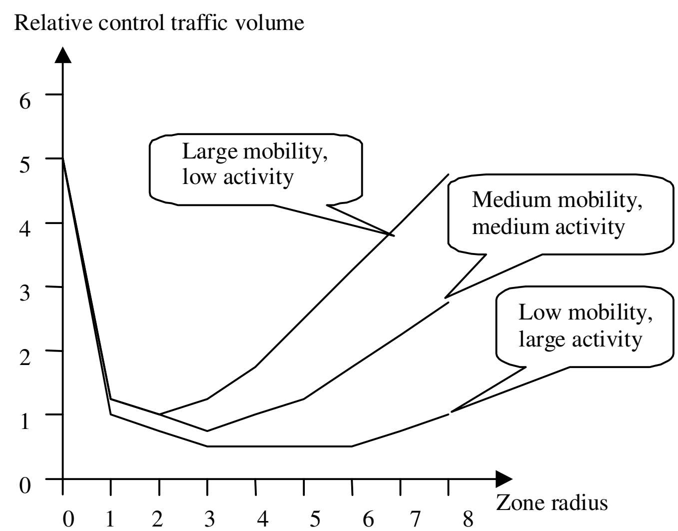

Fig. 4. Overhead of ZRP control traffic

The performance of the ZRP has been evaluated as a function of the network parameters, such as the size of the zone radius, the density of the micro-sensors, the transmission radius of the micro-sensors, migration patterns of the micro-sensors, and the size and shape of the micro-network coverage. A sample of the ZRP performance is shown in Fig. 4. The three representative curves in the figure show the relative volume of the control traffic as a function of the size of the zone radius. For highly mobile network with low traffic activity, smaller zone radius results in minimal control traffic, and vice versa. For extremely high mobility, the network degenerates to flooding (zone radius of 1) and for stationary network, the whole network becomes a single zone (large zone radius). The figure clearly demonstrates the drastic reduction in the amount of the control traffic, as compared with the purely proactive approach (large zone radius) and the purely reactive approaches (zone radius of 1). We omit here additional performance evaluation results due to space limitation; those could be found in [Haas00] or [Pearlman99].

7. Processing vs. Communication Costs

After the network has been established, the nodes are in position to communicate. One can imagine a wide spectrum of applications being implemented on the proposed micro-sensor network. In particular, some applications may require processing at the sensor nodes. Since the processing capability of the nodes is quite limited, a natural possibility is to off-load some of the processing from the network nodes to be processed on the ports, or even in the "backbone" fixed network. However, shipping of processing out of the sensor nodes that collect the information to other network components requires transmission, which is also limited, both in the network capacity and in the energy expenditure.

Therefore, there exists a trade-off between the distribution of processing on the network nodes vs. other network components. As another aspect of this trade-off, the information collected by the sensor nodes themselves can be locally processed (e.g., compressed), with less information to be transmitted out of the sensor nodes. Again, we see that there is a tradeoff between local processing and off-node processing with transmission.

Another tradeoff to consider comes about due to limited storage of the network nodes. Processing of some algorithms may require availability of storage. Even the routing protocol itself needs some limited storage for execution. The question is what is the minimal amount of storage that is needed to perform some network operations.

Thus, there is a three-dimensional space that one may try to optimize: computation vs. storage vs. transmission. The challenge is to determine what are the permissible boundaries in this three-dimensional space, given the device and circuit technology with which the sensor nodes are designed. As the processing, storage, and transmission capabilities of the network nodes increase, more complex network control protocols can be deployed. The challenge is to optimize the performance of the network protocols such as the network capacity and delay, subject to the amount of the above three resources in the miniaturized network nodes.

Yet another aspect worth mentioning is the tradeoff between the size of the devices and its computation power, its storage capability, and its transmission capacity. In particular, if one would size the device to reduce its dimension by some factor, how would this affect the above three parameters. Although such results are very much technology-dependent, we believe that they can be extrapolated over a spectrum of technologies.

8. Summary and Concluding Remarks

In this position paper, we have introduced the vision of micro networking - creating a network of miniature devices (sensors and actuators), which could possibly be manufactured using the MEMS technology. These devices include, in addition to limited communication capabilities, also restricted memory and computing resources. Thus, the challenge is to design a set of protocols that, on one hand, would allow sufficient coupling between the devices, but, on the other hand, be extremely efficient in its use of computing and communication. In particular, we have introduced the notion of physical addressing, on-demand neighbor discovery, and the Zone Routing Protocol, which, working together, may consist of a basic set of schemes to implement the smart communication environment of the future. Our Wireless Networking Laboratory at Cornell University has been pursuing these issues in collaboration with other research groups at Cornell University.

9. References

[Akin98] T. Akin, K. Najafi, R. Bradley, “A Wireless Implantable Multichannel Digital Neural Recording Systems for a Micromachined Sieve Electrode,” IEEE Journal of Solid-State Circuits, vol. 33, no. 1, 1998

[Asada98] G. Asada, M. Dong, T.S. Lin, F. Newberg, G. Pottie, W.J. Kaiser, and H.O. Marcy, “Wireless Integrated Network Sensors: Low Power Systems on a Chip,” Proceedings of the 1998 European Solid State Circuits Conference.

[Bertsekas92] D. Bertsekas and R. Gallager, *Data Networks,* Second Edition, Prentice Hall, Inc., 1992.

[Cheng89] C. Cheng, R. Reley, S.P.R. Kumar, and J.J. Garcia-Luna-Aceves, “A Loop-Free Extended Bellman-Ford Routing Protocol without Bouncing Effect,” *ACM Computer Communications Review,* 19(4), 1989, pp.224-236.

[GLOMO] *http://www.darpa.mil/ito/research/glomo/index.html*

[Haas00] Z.J. Haas and M.R. Pearlman, “The Performance of Query Control Schemes for the Zone Routing Protocol," in Ad-Hoc Networks, edited by Charlie Perkins, Addison Wesley, 2000

[IETF-MANET] *http://www.cs-ipv6.lancs.ac.uk/ipv6/documents/standards/general-comms/ietf/manet/*

[Johnson96] D.B. Johnson and D.A. Maltz, “Dynamic Source Routing in Ad-Hoc Wireless Networking,” in *Mobile Computing,* T. Imielinski and H. Korth, editors, Kluwer Academic Publishing, 1996.

[Leiner87] B.M. Leiner, D.L. Nielson, and F.A. Tobagi, “Issues in Packet Radio Network Design,” *Proceedings of the IEEE,* vol.75, pp.6-20, January 1987.

[Moy94] J. Moy, “OSPF Version 2], *RFC 1583,* March 1994.

[Murthy95] S. Murthy and J.J. Garcia-Luna-Aceves, “A Routing Protocol for Packet Radio Networks,” Proc. of ACM Mobile Computing and Networking Conference, *MOBICOM’95,* Nov. 14-15, 1995.

[Park97] V.D. Park and M.S. Corson, “A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks,” *IEEE INFOCOM’97,* Kobe, Japan, 1997.

[Pearlman99] M.R.Pearlman and Z.J. Haas, "Determining the Optimal Configuration for the Zone Routing Protocol" *IEEE Journal on Selected Areas in Communications (JSAC)*, issue on *Ad-Hoc Networks*, vol.17, no.8, August 1999

[Perkins94] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," *ACM SIGCOMM,* vol.24, no.4, Oct. 1994, pp.234-244.

[Perkins99] C.E. Perkins and E.M. Royer, "Ad Hoc On-Demand Distance Vector Routing," *IEEE WMCSA'99,* New Orleans, LA, Feb. 1999.

[Pehlke97] D. R. Pehlke, et al, "Extremely Hight-Q Tunable Inductor for Si-Based RF Integrated Circuit Applications", IEDM'97 International Electron Devices Meeting Technical Digest, 1997.

[Young98] D. J. Young, V. Malba, J. Ou, A. Bernhardt, B. E. Boser, " A Low-noise RF VCO using on-chip high-Q 3-D coil inductor and micromachined variable capacitor," Solid-State Sensor and Actuator Workshop, Hilton Head, SC, 1998.