Chapter 1

Introduction to Wireless Sensor Networks

1.1 Overview

With the popularity of laptops, cell phones, PDAs, GPS devices, RFID, and intelligent electronics in the post-PC era, computing devices have become cheaper, more mobile, more distributed, and more pervasive in daily life. It is now possible to construct, from commercial off-the-shelf (COTS) components, a wallet size embedded system with the equivalent capability of a 90's PC. Such embedded systems can be supported with scaled down Windows or Linux operating systems. From this perspective, the emergence of wireless sensor networks (WSNs) is essentially the latest trend of Moore's Law toward the miniaturization and ubiquity of computing devices.

Typically, a wireless sensor node (or simply sensor node) consists of sensing, computing, communication, actuation, and power components. These components are integrated on a single or multiple boards, and packaged in a few cubic inches. With state-of-the-art, low-power circuit and networking technologies, a sensor node powered by 2 AA batteries can last for up to three years with a 1% low duty cycle working mode. A WSN usually consists of tens to thousands of such nodes that communicate through wireless channels for information sharing and cooperative processing. WSNs can be deployed on a global scale for environmental monitoring and habitat study, over a battle field for military surveillance and reconnaissance, in emergent environments for search and rescue, in factories for condition based maintenance, in buildings for infrastructure health monitoring, in homes to realize smart homes, or even in bodies for patient monitoring [60; 76; 124; 142].

After the initial deployment (typically ad hoc), sensor nodes are responsible for self-organizing an appropriate network infrastructure, often with multi-hop connections between sensor nodes. The onboard sensors then start collecting acoustic, seismic, infrared or magnetic information about the environment, using either continuous or event driven working modes. Location and positioning information can also be obtained through the global positioning system (GPS) or local positioning algorithms. This information can be gathered from across the network and appropriately processed to construct a global view of the monitoring phenomena or objects. The basic philosophy behind WSNs is that, while the capability of each individual sensor node is limited, the aggregate power of the entire network is sufficient for the required mission.

In a typical scenario, users can retrieve information of interest from a WSN by injecting queries and gathering results from the so-called base stations (or sink nodes), which behave as an interface between users and the network. In this way, WSNs can be considered as a distributed database [45; 184]. It is also envisioned that sensor networks will ultimately be connected to the Internet, through which global information sharing becomes feasible (Figure 1.1).

The era of WSNs is highly anticipated in the near future. In September 1999, WSNs were identified by *Business Week* as one of the most important and impactive technologies for the 21st century [31]. Also, in January 2003, the *MIT's Technology Review* stated that WSNs are one of the top ten emerging technologies [125]. It is also estimated that WSNs generated less than $150 million in sales in 2004, but would top $7 billion by 2010 [133]. In December 2004, a WSN with more than 1000 nodes was launched in Florida by the ExScal team [61], which is the largest deployed WSN to date.

1.2 Enabling Technologies

1.2.1 *Hardware*

The hardware basis of WSNs is driven by advances in several technologies. First, System-on-Chip (SoC) technology is capable of integrating complete systems on a single chip. Commercial SoC based embedded processors from Atmel, Intel, and Texas Instruments have been used for sensor nodes such as UC Berkeley's motes [48; 173], UCLA's Medusa [120] and WINS [197], and MIT's μAMPS-1 [187]. Several research groups, such as the PicoRadio team from UC Berkeley [139], have been trying to integrate prototype sensor nodes (PicoNode I) onto a few chips (PicoNode II). Many interesting SoC

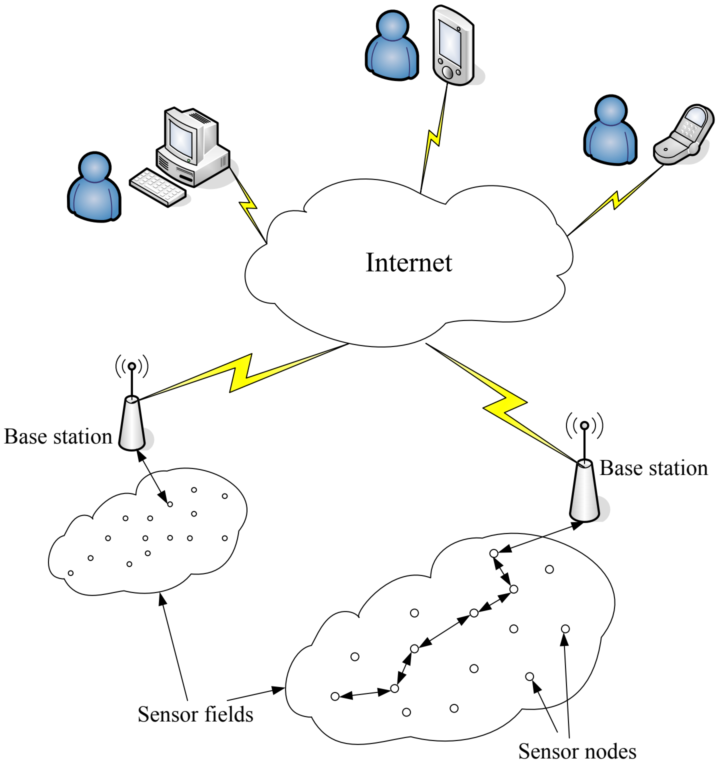


Fig. 1.1 Accessing WSNs through Internet.

designs related to wireless communication and sensor nodes can also be found at the SoC Design Challenge, 2004-2006 [174].

Second, commercial RF circuits enable short distance wireless communication with extremely low power consumption. Commercial products from RF Monolithics, Chipcon, Conexant Systems, and National Semiconductor have been used on various sensor nodes, including motes, Medusa, WINS, and μAMPS. A SoC based ZigBee radio is also available from Ember Cooperation [58]. These commercial radios can usually achieve a data rate of tens to hundreds of Kbps, while consuming less than 20 mW of power for both packet transmission and receiving [140]. With wideband technology, enhanced modulation schemes and error detection mechanisms are employed to provide increased robustness.

Third, Micro-Electro-Mechanical Systems (MEMS) technology [122] is now available to integrate a rich set of sensors onto the same CMOS chip. Commercially available sensors now include thermal, acoustic/ultrasound, and seismic sensors, magnetic and electromagnetic sensors, optical transducers, chemical and biological transducers, accelerometers, solar radiation detectors, photosynthetically active radiation detectors, and barometric pressure detectors [105]. These sensors can be used in a broad range of applications, including acoustic ranging, motion tracking, vibration detection, and environmental sensing.

The above technologies, along with advanced packaging techniques, have made it possible to integrate sensing, computing, communication, and power components into a miniaturized sensor node.

1.2.2 *Wireless Networking*

Besides hardware technologies, the development of WSNs also relies on wireless networking technologies. The 802.11 protocol, the first standard for wireless local area networks (WLANs), was introduced in 1997. It was upgraded to 802.11b with an increased data rate and CSMA/CA mechanisms for medium access control (MAC). Although designed for wireless LANs that usually consist of laptops and PDAs, the 802.11 protocols are also assumed by many early efforts on WSNs. However, the high power consumption and excessively high data rate of 802.11 protocols are not suitable for WSNs. This fact has motivated several research efforts to design energy efficient MAC protocols [109; 145; 189; 206]. Recently, the 802.15.4-based ZigBee protocol was released, which was specifically designed for short range and low data rate wireless personal area networks (WPAN). Its applicability to WSNs was soon supported by several commercial sensor node products, including MicaZ [48], Telos [140], and Ember products [58].

Above the physical and MAC layers, routing techniques in wireless networks are another important research direction for WSNs. Some early routing protocols in WSNs are actually existing routing protocols for wireless ad hoc networks or wireless mobile networks. These protocols, including DSR [88] and AODV [138], are hardly applicable to WSNs due to their high power consumption. They are also designed to support general routing requests in wireless networks, without considering specific communication patterns in WSNs. Nevertheless, the customization of these protocols for WSNs and the development of new routing techniques have become hot research topics [26; 51; 66; 73; 85; 95; 107; 160;

202]. The main idea behind these research efforts is to enable energy efficient and robust routing by exploiting link and path diversity.

1.2.3 *Collaborative Signal Processing*

Collaborative signal processing algorithms are another enabling technology for WSNs. While raw data from the environment are collected by sensor nodes, only useful information is of importance. Hence, raw data need to be properly processed locally at sensing nodes, and only processed data is sent back to the end users. Since computation is much more energy efficient than wireless communication, this avoids wasting energy on sending large volumes of raw data. Such signal processing is often required to be performed by a set of sensor nodes in proximity, due to the weak sensing and processing capabilities of each individual node.

Information fusion is an important topic for collaborative signal processing. Since sensor readings are usually imprecise due to strong variations of the monitoring entity or interference from the environment, information fusion can be used to process data from multiple sensors in order to filter noise measurements and provide more accurate interpretations of the information generated by a large number of sensor nodes. A rich set of techniques is applicable in this context, including Kalman filtering, Bayesian inference, neural networks, and fuzzy logic [7; 52; 91; 113; 165; 198].

Other signal processing techniques that have been developed for WSNs include time synchronization [57; 65; 179], localization [131; 154; 155], target tracking [50; 108; 214], edge and boundary detection [38; 101; 132], calibration [83; 194], adaptive sampling [137; 195], and distributed source coding [86; 153].

1.3 Evolution of Sensor Nodes

There has been a long history for (remote) sensing as a means for humans to observe the physical world. For example, the telescope invented in the 16th century is simply a device for viewing distant objects. As with many technologies, the development of sensor networks has been largely driven by defense applications.

1.3.1 *Military Networks of Sensors*

Since the early 1950s, a system of long-range acoustic sensors (hydrophones), called the Sound Surveillance System (SOSUS), has been deployed in the deep basins of the Atlantic and Pacific oceans for submarine surveillance. Beams from multiple hydrophone arrays are used to detect and locate underwater threats. Recently, SOSUS has been replaced by the more sophisticated Integrated Undersea Surveillance System.

Networks of air defense radars can be regarded as an example of networked large scale sensors. Both ground-based radar systems and Airborne Warning and Control System (AWACS) planes are integrated into such networks to provide all-weather surveillance, command, control, and communications. The radar dome on AWACS planes is 30 feet in diameter and six feet thick. It can detect flying targets in a range of more than 200 miles. In the 1980s and 1990s, the Cooperative Engagement Capability (CEC) [33] was developed as a military sensor network, in which information gathered by multiple radars was shared across the entire system, to provide a consistent view of the battle field.

Another early example of sensing with wireless devices is the Air Delivered Seismic Intrusion Detector (ADSID) system, used by US Air Force in the Vietnam war. Each ADSID node was about 48 inches in length, nine inches in diameter, and weighted 38 pounds. Equipped with a sensitive seismometer, these ADSID nodes were planted along the Ho Chi Minh Trail to detect vibrations from moving personnel and vehicles. The sensed data were transmitted from each node directly to an airplane, over a channel with unique frequency.

Although the ADSID nodes were large, and the high energy cost of direct communication limited the lifetime of nodes to only a few weeks, they successfully demonstrated the concept of wirelessly networked sensors. With the success of digital packet radios for wireless networking by the ALOHAnet Project [2] at Hawaii and DARPA's Packet Radio Project [90] in 1970s, wireless communication within the same frequency band using MAC techniques and packet-based multihop communication became possible.

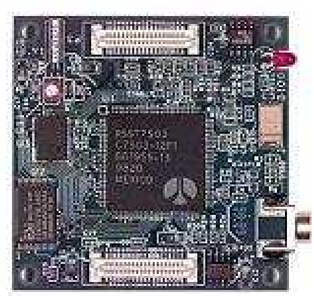
1.3.2 *Next Generation Wireless Sensor Nodes*

1.3.2.1 *WINS from UCLA*

In 1996, the Low Power Wireless Integrated Microsensors (LWIMs) [28] were produced by UCLA and the Rockwell Science Center. By using commercial, low cost CMOS fabrication, LWIMs demonstrated the ability to integrate multiple sensors, electronic interfaces, control, and communication on a single device. LWIM supported over 100 Kbps wireless communication at a range of 10 meters using a 1 mW transmitter.

In 1998, The same team built a second generation sensor node — the Wireless Integrated Network Sensors (WINS) [11]. Commercial WINS from Rockwell Science Center [197] each consists of a processor board with an Intel StrongARM SA1100 32-bit embedded processor (1 MB SRAM and 4 MB flash memory), a radio board that supports 100 Kbps with adjustable power consumption from 1 to 100 mW, a power supply board, and a sensor board. These boards are packaged in a 3.5"x3.5"x3" enclosure (Figure 1.2). The processor consumes 200 mW in the active state and 0.8 mW when sleeping.

(a) The WINS processor board

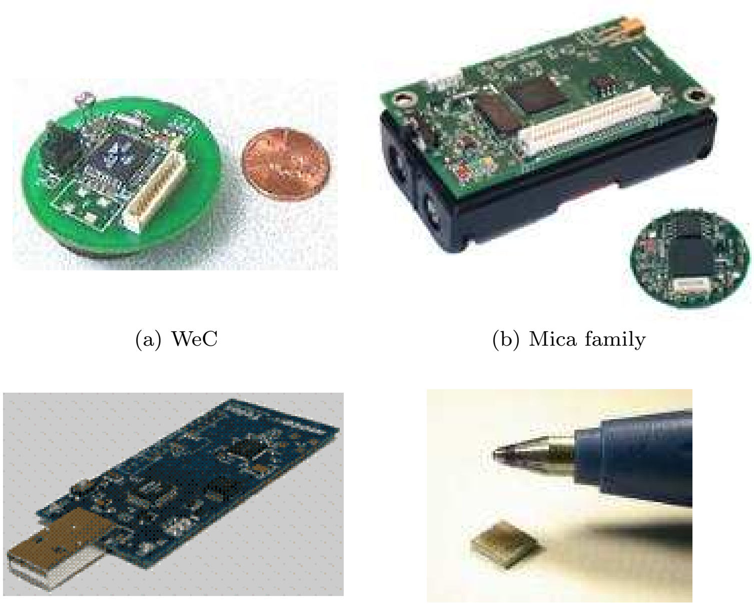


(b) The WINS radio board

Fig. 1.2 WINS node from Rockwell Science Center.

1.3.2.2 *Motes from UC Berkeley*

While WINS offer relatively powerful processing and communication capabilities, other research efforts have been developing smaller and cheaper nodes with less power consumption. In 1999, the Smart Dust project [173] at UC Berkeley released the first node, WeC, in their product family of *motes* (Figure 1.3(a)). WeC was built with a small 8-bit, 4 MHz Atmel microcontroller (512 bytes RAM and 8 KB flash memory), which consumed 15 mW active power and 45 μW sleeping power. WeC also had a simple radio supporting a data rate up to 10 Kbps, with 36 mW transmitting power and 9 mW receiving power. Later on, René and Dot were built in 1999 and 2000, respectively, with upgraded microcontrollers.



(a) WeC (b) Mica family

(c) Telos (d) Spec prototype

Fig. 1.3 Motes from UC Berkeley.

Along this line, the Mica family was released in 2001, including Mica [75], Mica2, Mica2Dot, and MicaZ [48]. While Mica still used an 8-bit 4 MHz microcontroller (ATmega103L), it offered enhanced capabilities in terms of memory and radio, compared with preceding products. Specifically, Mica was designed with 4 KB Ram, 128 KB flash, and a simple bit-level radio using RFM TR1000 that supported up to 40 Kbps with almost the same power consumption as the radio module on WeC. Mote architecture allowed several different sensor boards, or a data acquisition board, or a network interface board to be stacked on top of the main processor/radio board. These boards supported various sensors, most of which are listed in Section 1.2.1. The basic processor/radio board was approximately one inch by two inches in size (Figure 1.3(b)).

The follow-ups to Mica, Mica2 and Mica2Dot were built in 2002 with an ATmega128L microcontroller that reduced standby current (33 mW active power and 75 μW sleep power). They also had improved radio modules (Chipcon CC1000) with more options for frequency range, and increased resilience to noise by using FSK modulation. One year later, MicaZ was produced with a Chipcon CC2420 wideband radio module that supported 802.15.4 and ZigBee protocols, with a data rate up to 250 Kbps. This radio module also supported on-chip data encryption and authentication.

The latest member in the family, Telos [140], was released in 2004 (Figure 1.3(c)). Telos offered a set of new features: (1) a microcontroller from Texas Instruments with 3 mW active power and 15 μW sleep power, (2) an internal antenna built into the printed circuit board to reduce cost, (3) an on-board USB for easier interface with PCs, (4) integrated humidity, temperature, and light sensors, and (5) a 64-bit MAC address for unique node identification.

An interesting research testbed is the Spec platform [74], which integrated the functionality of Mica onto a single 5 mm^2 chip (Figure 1.3(d)). Spec was built with a micro-radio, an analog-to-digital converter, and a temperature sensor on a single chip, which lead to a 30-fold reduction in total power consumption. This single-chip integration also opened the path to low cost sensor nodes.

The integrated RAM and flash memory architecture has greatly simplified the design of the mote family. However, the tiny footprint also requires a specialized operating system, which was developed by UC Berkeley, called TinyOS [185]. TinyOS features a component-based architecture and event-driven model that are suitable for programming with small embedded devices, such as motes. The combination of Motes and TinyOS is gradually becoming a popular experimental platform for many research efforts in the field of WSNs.

1.3.2.3 *Medusa from UCLA*

The design philosophy and operational space of motes are quite different from those of WINS. On one hand, motes are designed for simple sensing and signal processing applications, where the demand for computation and communication capabilities is low. On the other hand, WINS are essentially an embedded version of PDAs, for more advanced computationally intensive applications with large memory space requirements. To bridge the gap between the two extremes, the Medusa MK-2 sensor node was developed by the Center for Embedded Networked Sensing (CENS) at UCLA in 2002 (Figure 1.4).

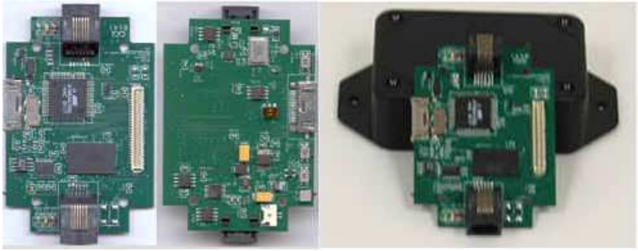


Fig. 1.4 Medusa node from UCLA.

One distinguishing feature of Medusa MK-2 is that it integrates two microcontrollers. The first one, ATmega128, is dedicated to less computationally demanding tasks, including radio base band processing and sensor sampling. The second one, AT91FR4081, is a more powerful microcontroller (40 MHz, 1 MB flash, 136 KB RAM) that can be used to handle more sophisticated, but less frequent signal processing tasks (e.g., the Kalman filter). The combination of these two microcontrollers provides more flexibility in WSN development and deployment, especially for applications that require both high computation capabilities and long lifetime.

1.3.2.4 *PicoRadio from UC Berkeley*

All the aforementioned sensor architectures are based on batteries. Due to the slow advancement in battery capacity, techniques for energy scavenging from the environment have been an attractive research field. In 2003, the Berkeley Wireless Research Center (BWRC) presented the first radio transmitter, PicoBeacon (Figure 1.5), purely powered by solar and vibrational energy sources. With a custom RF integrated circuitry that was developed for power consumption less than 400 μW, the beacon was able to achieve duty cycles up to 100% for high light conditions and 2.6% for typical ambient vibrational conditions. It is anticipated that an integrated wireless transceiver with < 100 μW power consumption is feasible in the near future.

Fig. 1.5 PicoBeacon from UC Berkeley.

The BWRC also produced SoC based sensor nodes instead of using COTS components. In 2002, PicoNode II was built using two ASIC chips that implemented the entire digital portion of the protocol stack. Together, the chip set consumed an average of 13 mW when three nodes were connected. The team is also building PicoNode III, which will integrate a complete PicoNode into a single small aspect-ratio package.

1.3.2.5 *μAMPS from MIT*

The same ASIC based approach is being taken by the μAMPS group from MIT. Following its first testbed, μAMPS-I (Figure 1.6), the team is now trying to build a highly integrated sensor node comprised of a digital and an analog/RF ASIC, μAMPS-II. The interesting feature of μAMPS-II is that the node will be able to operate in several modes. It can operate as either a low-end stand-alone guarding node, a fully functional node for middle-end sensor networks, or a companion component in a more powerful high-end sensor systems. Thus, it favors a network with heterogeneous sensor nodes for a more efficient utilization of resources.

Besides the above sensor nodes, other commercial products and testbeds for WSNs include Ember products [58], Sensoria WINS [161], Pluto mote [40], PC104 testbed [136], and Gnome testbed [193].

Fig. 1.6 μAMPS-I from MIT.

1.3.3 *Why Microscopic Sensor Nodes?*

The transition from large to small scale sensor nodes has several advantages.

(1) Small sensor nodes are easy to manufacture with much lower cost than large scale sensors. They are even disposable if the envisioned US$1 target price can be realized in the future.
(2) With a mass volume of such low cost and tiny sensor nodes, they can be deployed very closely to the target phenomena or sensing field at an extremely high density. Therefore, the shorter sensing range and lower sensing accuracy of each individual node are compensated for by the shorter sensing distance and large number of sensors around the target objects, which generates a high signal to noise ratio (SNR).
(3) Since computing and communication devices can be integrated with sensors, large-sample in-network and intelligent information fusion becomes feasible. The intelligence of sensor nodes and the availability of multiple onboard sensors also enhances the flexibility of the entire system.
(4) Due to their small size and self-contained power supply, sensor nodes can be easily deployed into regions where replenishing energy is not available, including hostile or dangerous environments. The survivability of nodes also increases with reduced size.

(5) The high node density enables system-level fault tolerance through node redundancy.

These advantages are illustrated by the microclimate monitoring of coastal redwood trees [150]. It is known that the movement of water from the ground to the canopy through the trunk is caused by the difference in water vapor pressure in the leaf and water vapor pressure in the air. To understand precisely the effects of microclimate variables, such as temperature and humidity, it is necessary to gather such information at different locations on the tree.

Because of their coarse resolution, it is difficult for large scale sensors, such as weather stations, to perform this task. However, by mounting a sufficient number of small sensor nodes along the tree trunk, it is possible to gather the desired information with a relatively low cost. These sensor nodes are able to collect both spatially and temporally dense sampling to enable a comprehensive view of the microclimate around the redwood tree. Because of wireless networking, it is easy to add more sensor nodes or move mounted nodes for better coverage. It is also possible to place redundant sensor nodes in order to enable local information fusion for better sensing accuracy.

Once deployed, the long lifetime of the network allows data collection over several years. The in-network storage capacity makes it possible to transfer intermittently the gathered data to a laptop. Also, these autonomous and intelligent sensor nodes are able to self-organize and self-heal the wireless network should node or link failures occur. This untethered operation avoids costly human management and maintenance.

1.4 Applications of Interest

An outline of the envisioned applications for WSNs is given in [11]. Descriptions of general applications for WSNs can also be found in [39] and [199]. For the purpose of this book, we categorize the applications into two classes. The first class, data gathering applications, focuses on entity monitoring with limited signal processing requirements. The primary goal of these applications is to gather information of a relatively simple form, such as temperature and humidity, from the operating environment. Some environmental monitoring and habitat study applications also belong to this class.

The second class of applications require the processing and transportation of large volumes of complex data. This class includes heavy industrial monitoring and video surveillance, where complicated signal processing algorithms are usually employed. We refer to these applications as computationally intensive applications.

In the following sections, we describe several academic and industrial efforts based on the above categorization. While both classes of applications are important for realizing the potential of WSNs, the involved techniques can be quite different due to their varying computation and communication demands. In Section 1.6, we discuss these differences in the context of this book.

1.4.1 *Data Gathering Applications*

1.4.1.1 *Habitat Study*

Habitat study is one of the driving applications for WSNs [34]. Such applications usually require the sensing and gathering of bio-physical or biochemical information from the entities under study, such as Redwoods [150], Storm Petrels [116], Zebras [89], and Oysters [84]. In many scenarios, habitat study requires relatively simple signal processing, such as data aggregation using minimum, maximum, or average operations. Hence, motes are ideal platforms for such applications.

The famous Great Duck Island project was initiated in the Spring of 2002 by Intel Research and UC Berkeley, to monitor the microclimates in and around Storm Petrel nesting burrows [116]. Thirty two motes were deployed on the island, each equipped with sensors for temperature, humidity, barometric pressure, and mid-range infrared. The network was designed to have a tiered structure. The motes were grouped into patches so that data collected in each patch could be relayed via a gateway to a base station, where data logging was performed. Within one year of monitoring, the system gathered approximately 1 million readings. In 2003, a second generation network, with more than 100 nodes, was also deployed.

Cape Breton University and the National Research Council of Canada are conducting an on-going bio-physical monitoring effort in the bras d'Or Lakes. Their goal is to study the life cycle of an oyster parasite (MSX), requiring the gathering of temperature and salinity parameters [84]. COTS sensor nodes will be deployed in the shallow shoreline of the lakes, which is preferred by oysters and easily accessible for biological and oceanographic monitoring.

1.4.1.2 *Environmental Monitoring*

Environmental monitoring is another application for WSNs. The vast spaces involved in such applications require large volumes of low cost sensor nodes that can be easily dispersed throughout the region. For instance, WSNs have been studied for forest fire alarm [99], landscape flooding alarm [8], soil moisture monitoring [32], microclimate and solar radiation mapping [141], and environmental observation and forecasting in rivers [43].

Researchers at University of West Australia are developing a prototype WSN for outdoor, fine-grained environmental monitoring of soil water [32]. Such a network can be used to assist salinity management strategies, or to monitor irrigated crops, urban irrigation, and water movement in forest soils. In January 2005, a prototype network was built, which included 15 Mica2 nodes integrated with soil moisture sensors and other gateway and routing nodes. The system distinguishes itself by using a reactive data gathering strategy — frequent soil moisture readings are collected during rain, while less frequent readings are collected otherwise. This strategy helps increase the system lifetime.

1.4.2 *Computation-Intensive Applications*

1.4.2.1 *Structural Health Monitoring*

Health monitoring for civil structures has long been a research topic for industry and academia. Traditional methods include visual inspection, acoustic emission, ultrasonic testing, and radar tomography. The emergence of WSNs has prompted new, non-destructive, and cheap methods for many tasks related to structural health monitoring [114; 178; 200].

The volume of raw data to be gathered and transported for such applications is on the order of 1-10 Mbps [37]. Thus, transmitting only useful information obtained from local signal processing becomes imperative for sustaining a long system lifetime. Many sophisticated and computationally intensive signal processing algorithms have been studied, including the Fast Fourier Transformation (FFT), Wavelet Transform, Autoregressive Models [175], and AR-ARX Damage Detection Pattern Recognition Method [175]. To serve the large computation demand from these algorithms, while maximizing energy savings, a dual-core design method has been employed. For instance, with the aforementioned Medusa node and the sensor node developed by Lynch [114], while a low-end microcontroller is responsible for frequent sensing and communication tasks, a high-end embedded processor is occasionly utilized when heavy signal processing is required.

An on-going Structural Health Monitoring (SHM) project by University of Southern California [164] has developed two software systems, Wisden and NET-SHM. These systems facilitate continuous data acquisition over a self-configuring multi-hop WSN, with high data rate and reliable communication requirements. Moreover, a full-scale testbed ceiling of 28×48 feet has been built with actuators to deliver deterministic excitations. Currently, the team is constructing robotic actuators that can be remotely controlled to move above the ceiling. The team is also investigating the use of other modalities, such as images, to enhance the fidelity of the system.

1.4.2.2 *Heavy Industrial Monitoring*

Sensors have already been widely used in industrial applications, such as the monitoring of automated assembly lines. Integrating wireless technology with these sensors enables condition based maintenance (CBM) to reduce downtime and enhance safety, with low installation and maintenance cost. Condition based maintenance can replace traditional high-cost, schedule-driven, manual maintenance for various industrial entities, including power plants, oil pipelines, transportation systems and vehicles, engineering facilities, and industrial equipment.

Industrial applications are unique in their requirement of highly reliable operation in harsh environments. For example, the electromagnetic radiation of machines may cause microcontroller malfunction or wireless communication interference. Also, the large variation in temperature and humidity demands reliable hardware components. Moreover, industrial applications often require the processing of large volumes of data with sophisticated signal processing algorithms. Thus, computation demand is usually high for these applications.

Intel Research has deployed a network with 160 Mica2 motes on a ship to measure the vibrations in the ship's pumps, compressors, and engines as an indicator of potential failure [29; 54]. These motes were organized into clusters, with Stargate gateways [48] forming the backbone of the network. Without operator intervention, the deployed network operated for 4 months without major failures. This experiment was still preliminary since the diagnosis of the ship equipments was performed in a centralized way at the base station, instead of distributed within the network. However, it paved the path for WSNs to a broad range of applications in industrial environments.

1.5 Research Topics and Challenges

Due to potentially harsh, uncertain, and dynamic environments, WSNs are envisioned to operate in an autonomous and untethered fashion. This poses considerable challenges ranging through network organization, topology discovery, communication scheduling, routing control, and signal processing. Also, tight energy budgets enforce energy efficient designs for hardware components, network stacks, and application algorithms.

In this section, we briefly describe a list of research challenges for WSNs. For the purpose of this book, we are particularly interested in the first three challenges. In Chapter 2, we discuss them in detail.

(1) **Data-centric paradigm**: The operating paradigm of WSNs is centered around information retrieval from the underlying network, usually referred to as a *data-centric* paradigm. Compared to the *address-centric* paradigm exhibited by traditional networks, the data-centric paradigm is unique in several ways. New communication patterns resemble a reversed multicast tree. In-network processing extracts information from raw data and removes redundancy among multiple source data. Also, cooperative strategies among sensor nodes are used to replace the non-cooperative strategies for most Internet applications. The development of appropriate routing strategies that take the above factors into consideration is challenging.
(2) **Collaborative information processing and routing**: The data-centric paradigm involves two fundamental operations in WSNs: information processing and information routing. Many research efforts are motivated by the fact that information processing and routing are mutually beneficial. While information processing helps reduce the data volume to be routed, information routing facilitates joint information compression (or data aggregation) by bringing together data from multiple sources. However, it is often non-trivial to model and analyze the inter-relationship between information processing and routing. In many situations, the problem of finding a routing scheme in conjunction with joint compression for energy minimization turns out to be NP-hard.

(3) **Energy-efficient design**: Once deployed, it is often infeasible or undesirable to re-charge sensor nodes or replace their batteries. Thus, energy conservation becomes crucial for sustaining a sufficiently long network lifetime. Among the various techniques proposed for improving energy-efficiency, cross-layer optimization has been realized as an effective approach. Due to the nature of wireless communication, one performance metric of the network can be affected by various factors across layers. Hence, a holistic approach that simultaneously considers the optimization at multiple layers enables a larger design space within which cross-layer tradeoffs can be effectively explored.

(4) **Network discovery and organization**: Due to the large scale of WSNs, each sensor node behaves based on its local view of the entire network, including topology and resource distribution. Here, resources include battery energy and sensing, computation, and communication capabilities. To establish such a local view, techniques such as localization and time synchronization are often involved. A local view depends on the initial deployment of sensor nodes, which is itself a challenging topic. The network is usually organized using either a flat or hierarchical structure, above which topology control, MAC, and routing protocols can be applied accordingly.

    One key challenge is to handle network dynamics during the process of network discovery and organization. These dynamics include fluctuation in channel quality, failure of sensor nodes, variations in sensor node capabilities, and mobility or diffusion of the monitored entity. Autonomous adaptation of network discovery and organization protocols, in light of such dynamics, is the key to deliver proper system functionality.

(5) **Security**: Since WSNs may operate in a hostile environment, security is crucial to ensure the integrity and confidentiality of sensitive information. To do so, the network needs to be well protected from intrusion and spoofing. The constrained computation and communication capability of sensor nodes make it unsuitable to use conventional encryption techniques. Lightweight and application-specific architectures are preferred instead.

1.6 Focus of This Book

The focus of this book is on algorithm development and performance analysis for cross-layer optimization for energy-efficient information processing and routing in WSNs.

While our research efforts stem from the general concept of information processing and routing, this book covers the following three specific topics:

(1) information processing within a cluster of sensor nodes (or in-cluster information processing)
(2) information transportation over a given multi-hop tree structure (referred to as data gathering tree)
(3) information routing for computationally intensive applications over a general graph.

Each of these three topics is important and challenging in itself. Together, they cover a complete operating flow, from raw data sensing and processing at local clusters to information gathering and routing across the network. This is the major motivation to choose these three topics.

To facilitate cross-layer optimization in these topics, we study a set of fundamental techniques, referred to as *system knobs*. These system knobs are parameters that are exposed at certain levels, and can be tuned to adjust the performance of the system. In this book, we are particularly interested in three of them: voltage scaling, rate adaptation, and tunable compression. These techniques address the energy issue from computation, communication, and joint compression perspectives, respectively. Specifically, voltage scaling and rate adaptation achieve energy savings by trading computation/communication delay for energy, while tunable compression explores the tradeoffs between computation and communication energy cost. We illustrate these tradeoffs in Figure 1.7.

These three system knobs are applied in the aforementioned research topics. For the first topic, we investigate the application of voltage scaling and rate adaptation to maximize the system lifetime for in-cluster processing. For the second topic, we study rate adaptation for minimizing the energy cost for information transporting over an existing tree. For the last topic, we show that tunable compression can be incorporated into routing tree construction for minimizing the overall computation and communication energy in information routing.

One scenario for our research efforts is the cluster-based network scheme [72; 167; 208; 207]. In this scheme, the whole network is parti-

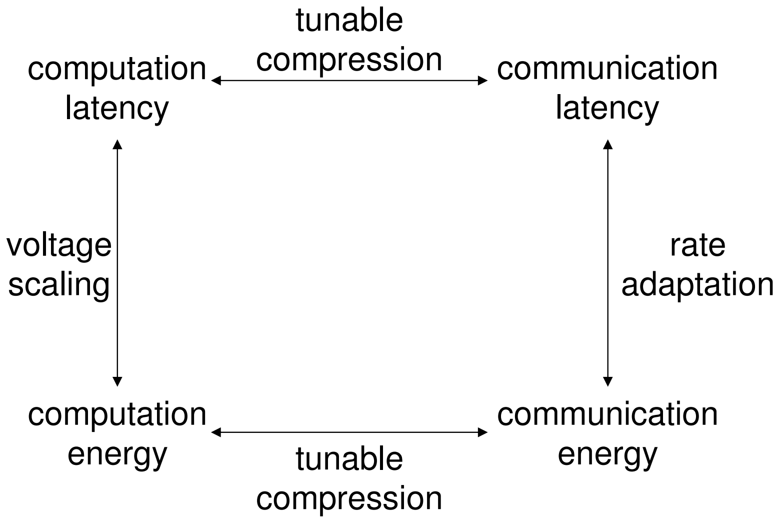


Fig. 1.7 Tradeoffs explored by three system knobs: voltage scaling, rate adaptation, and tunable compression.

tioned into either static or dynamic clusters, with one sensor node per cluster designated as a cluster head. We assume that each cluster behaves as a basic function unit, where in-cluster processing is responsible for converting raw data into useful information. The processed information is then transported back to the base station through either direct communication from cluster heads [72], a multi-hop tree that consists of only cluster heads [167], or a general multi-hop tree consisting of any sensor nodes in the network [208]. While the construction of a cluster-based infrastructure is beyond the scope of this book, we can see that our three research topics fit well into this scheme. Moreover, the proposed techniques are applicable to other scenarios as well.

Note that the research efforts presented in the book by no means provide a complete solution to information processing and routing. Our works are based on a relatively high model of the system. We are not concerned with the details of specific hardware to realize the system knobs, protocols for MAC layer scheduling and networking layer communication, or techniques for signal processing and data compression. Our focus is to improve the energy-efficiency of the systems by assuming that all such techniques are available. From a cross-layer optimization perspective, our work sits between the hardware and application layers when voltage scaling is employed, MAC and application layers for rate adaptation, and routing and application layers for tunable compression.