# 3

# IoT Sensing, Mobile and Cognitive Systems

**CHAPTER OUTLINE**



## 3.1 Sensing Technologies for Internet of Things

Integrating the digital world and physical world is the ultimate goal of IoT. This could be regarded as the third evolution of information industry. First, the network scale becomes very large in order to interconnect the enormous number of things in the physical world. Second, network mobility increases rapidly due to the pervasive use of mobile and vehicular devices. Third, the fusion of heterogeneous networks becomes deeper with various types of devices connected to the Internet. Furthermore, mobile Internet, cloud

*Big-Data Analytics for Cloud, IoT and Cognitive Computing*, First Edition. Kai Hwang and Min Chen.

Companion Website: http://www.wiley.com/go/hwangIOT

computing, big data, software defined networking and 5G all have an impact on IoT development.

### 3.1.1 Enabling Technologies and Evolution of IoT

In Figure 3.1, we have identified many technologies that enable IoT infrastructure development for various applications. The supportive technologies are divided into two categories: i) the enabling technologies build up the foundations of IoT. Among the enabling technologies, tracking (RFID), sensor networks and GPS are critical; ii) the synergistic technologies play the supporting roles. For example, biometrics could be widely applied to personalize the interaction between humans and machine and objects. Artificial intelligence, computer vision, robotics and telepresence can make our life better automated in the future.

In 2005, the concept of IoT came into the limelight. The IoT should be designed to connect the world's objects in a sensory manner. The approach is to tag things through RFID, feel things through sensors and wireless networks, and think things by building embedded systems that interact with human activities. The IoT is now becoming a major thrust, not only in the research community but also in big industry like IBM and Google. The IoT is really enabled by many related technologies. To name just a few, pervasive computing, social-media clouds, wireless sensor networks, cloud computing, big data, machine to machine communications and wearable computing, etc.

In 2008, the US National Intelligence Council published a report on "Disruptive Civil Technologies", which also identified the IoT as a critical technology in US interests to 2025. Quantitatively speaking, the IoT should be designed to encode 50 to 100 trillion objects. Moreover, the IoT should be designed to follow the movement of those objects. With more than a 6 billion human population, means every person is surrounded by

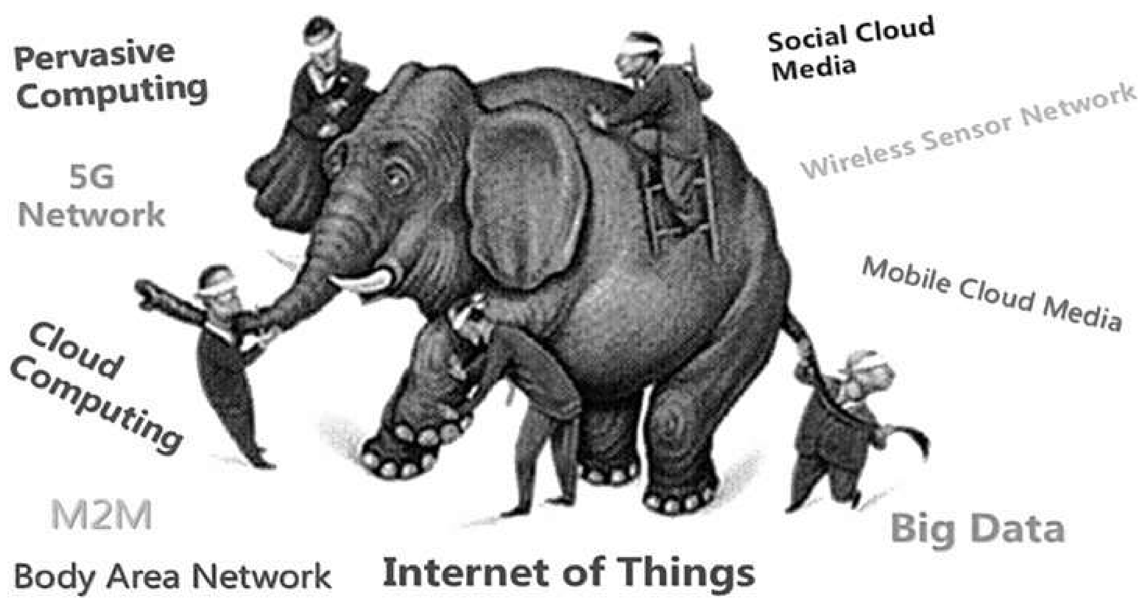



**Figure 3.1** IoT enabling and synergistic technologies.

1000 to 5000 objects on a daily basis. Imagine how IoT can improve our interactions or convenience with everything (object) surrounding us.

#### 3.1.1.1 Enabling and Synergistic Technologies

Over a period of 25 years, IoT development could become mature and more sophisticated. For example, supply chains could be more perfected in the first 10 years (2000–2010). Vertical market applications may be the next wave of advances. Ubiquitous positioning is expected to become a reality as we move toward 2020. Beyond that, a physical world web may appear to reach the ultimate goal of IoT. That goal is to achieve tremendous improvement in human abilities, societal outcomes, nation's productivity and the quality of life in general.

With an ever-increasing number of mobile devices and the resulting explosive mobile traffic, 5G networks call for various technology advances to transmit the traffic more effectively while changing the world by interconnecting a tremendous amount of mobile devices. However, mobile devices have limited communication and computation capabilities in terms of computation power, memory, storage and energy. In addition to the broadband bandwidth support from 5G, cloud computing needs to be utilized to enable mobile devices to obtain virtually unlimited dynamic resources for computation, storage and service provision that will overcome the constraints in the smart mobile devices. Thus, the combination of 5G and cloud computing technology are paving the way for other attractive applications.

With the support of mobile cloud computing (MCC), a mobile user basically has one more option to execute the computation of its application, i.e. offloading the computation to the cloud. Thus, one principal problem is under what conditions should a mobile user offload its computations to the cloud. The scenario of computation offloading at remote clouds requires a user to be covered by WiFi. Since the terminal device at the user end has limited resources, i.e. hardware, energy, bandwidth, etc., the cellphone itself cannot perform some compute-intensive tasks. Instead, the data related to the computation task can be offloaded to the remote cloud via WiFi or other high bandwidth channels.

Many IoT challenges are widely open, yet to be solved. Specific challenges include privacy, participatory sensing, data analytics, GIS (geographic information system) based visualization and cloud computing. Other areas are related to IoT architecture standardization, energy efficiency, security, protocols and Quality of Service. Standardization of frequency bands and protocols plays a pivotal role in accomplishing these goals. A roadmap of key developments for IoT research in the context of pervasive applications is shown in Figure 3.2. This diagram shows the growth of five key IoT application domains from 2010 to 2025.

In the early 2000s, IoT was primarily applied to expedite supply chain management. Vertical market and ubiquitous positioning applications have dominated IoT applications since 2010. Eventually, we will see the widespread use of IoT in a physical-world web, where teleoperations and telepresence will enable the monitoring and control of any remote objects. Ultimately, the IoT will enable the creation of a physical-world web that is connected to everything on Earth. This will make our daily lives convenient and well informed with the rest of the world. This will make smarter decisions, save human life, avoid disasters and reduce human burdens to a great extent. On the other hand, the

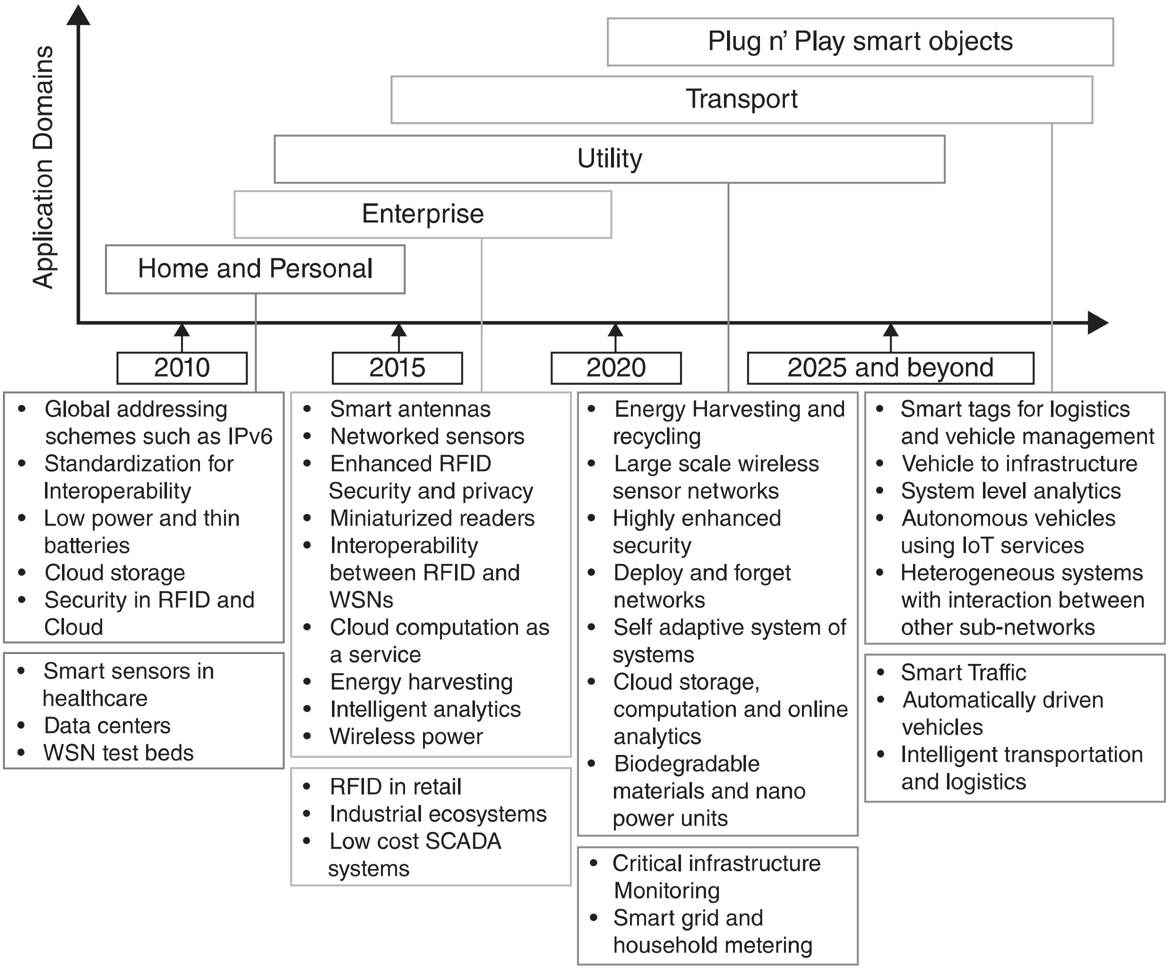



**Figure 3.2** Projected IoT upgrade in five IoT application domains from 2010 to 2015. (Courtesy of Gubbi et al. 2013. [10]) Reproduced with permission of Elsevier.

rise of IoT brings some negative impacts. For example, we may lose privacy. Criminals or enemy powers could use IoT to stage even more destruction. Legal systems need to be established to prevent or avoid these negative IoT impacts.

### 3.1.2 Introducing RFID and Sensor Technologies

This section briefly introduces Radio Frequency Identification (RFID) and sensors. More details are given in Section 3.3 and 3.4. With the rapid advances in electronics, electromechanics and nanotechnologies, ubiquitous devices grow rapidly in quantity and smaller in physical size. These objects are referred to as "things", such as computers, sensors, people, actuators, refrigerators, TVs, vehicles, mobile phones, clothes, food, medicines, books, passports, luggage, etc. They are expected to become active participants in business, information and social processes. These participants can react autonomously in the physical world. They influence or trigger actions and create services with or without direct human intervention. There are lots of sensor devices for sensing and data collection. Each senor node could combine the functions of sensing, communications and local processing.

#### 3.1.2.1 RFID Technology

The first step of enabling smart services is to collect contextual information about the environment, "things" and objects of interest. For example, sensors can be used to continuously monitor a human's physiological activities and actions such as health status and motion patterns. RFID technology can be utilized for collecting crucial personal information and storing it on a low-cost chip that is attached to an individual at all times. RFID is a radio-frequency (RF) electronic technology that allows automatic identification or location of objects, people and animals in a wide variety of deployment settings. In the past decade, RFID systems have been incorporated into a wide range of industrial and commercial systems, including manufacturing and logistics, retail, item tracking and tracing, inventory monitoring, asset management, anti-theft, electronic payment, anti-tampering, transport ticketing, supply-chain management, etc.

A typical RFID application consists of an RFID tag, an RFID reader and a backend system. With a simple RF chip and an antenna, an RFID tag can store information that identifies the object to which it is attached. There are three types of RFID tags, i.e. passive tags, active tags and semi-active tags. A passive tag obtains energy through RF signals from the reader, while an active tag is powered by an embedded battery, which enables larger memory or more functionality. Though a semi-active tag communicates with RFID readers like a passive tag, additional modules can be supported through an internal battery. When it comes within the proximity of an RFID reader, the information stored in the tag is transferred to the reader, and onto a backend system, which can be a computer employed for processing this information and controlling the operation of other sub-system(s).

#### 3.1.2.2 Sensors and Sensor Networks

In the last decade, we have witnessed a growing interest in deploying the sheer number of micro-sensors that collaborate in a distributed manner on data gathering and processing. Sensor nodes are expected to be inexpensive and can be deployed in various environments. A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. The sensor nodes form a multi-hop *ad hoc* wireless network. The biggest difference between WSN and cellular network is that WSN does not need a base station and each sensor node works as both transmitter and receiver. Due to limited resources at sensor nodes, routing in WSN is a challenging task, while minimizing energy consumption during data dissemination.

#### 3.1.2.3 Wireless Sensor Network

A WSN is a group of specialized transducers with a communications infrastructure intended to monitor and record conditions at diverse locations. Commonly monitored parameters are temperature, humidity, pressure, wind direction and speed, illumination intensity, vibration intensity, sound intensity, power-line voltage, chemical concentrations, pollutant levels and vital body functions. A sensor network consists of multiple detection stations called sensor nodes, each of which is small, lightweight and portable. Every sensor node is equipped with a transducer, microcomputer, transceiver and power source. The transducer generates electrical signals based on sensed data.

The sensor processor handles input signals and stores or transmits the output. The transceiver can be hard-wired or wireless. The power for each sensor node is derived from the electric utility or from a battery. A sensor node may vary in size from that of a shoebox down to the size of a dust grain. The cost of sensor nodes also varies widely, ranging from hundreds of dollars to a few pennies, depending on the size of the sensor network and the complexity required of individual sensor nodes. Size and cost constraints on sensor nodes are often decided by energy, memory, computational speed and bandwidth of the sensors used.

The widely-used sensor technologies are the Zigbee devices specified in the IEEE 802.15.4 Standard [1]. The radio frequency applied in Zigbee results in low data rate, long battery life and secure networking. They are used mainly in monitory and remote control IoT or mobile applications. Many supermarkets, department stores and hospitals are installed with Zigbee networks. The data rate ranges from 20–250 Kbps. They can operate up to 100 meters. However, Zigbee devices can be networked together to cover a much large area. The Zigbee network is highly scalable. Zigbee networks are used in wireless home-area networks (WHAN). This technology is simpler to use and less expensive than Bluetooth or WiFi.

### 3.1.3 IoT Architectural and Wireless Support

Basic IoT architecture is introduced below in three layers, namely the sensing, networking and application layers. The IoT system is likely to have an event-driven architecture. In Figure 3.3, the IoT development is shown with a three-layer architecture. The top layer is formed by the driven applications. IoT applications for healthcare will be presented in Chapter 8. The bottom layer consists of various types of sensing and automatic

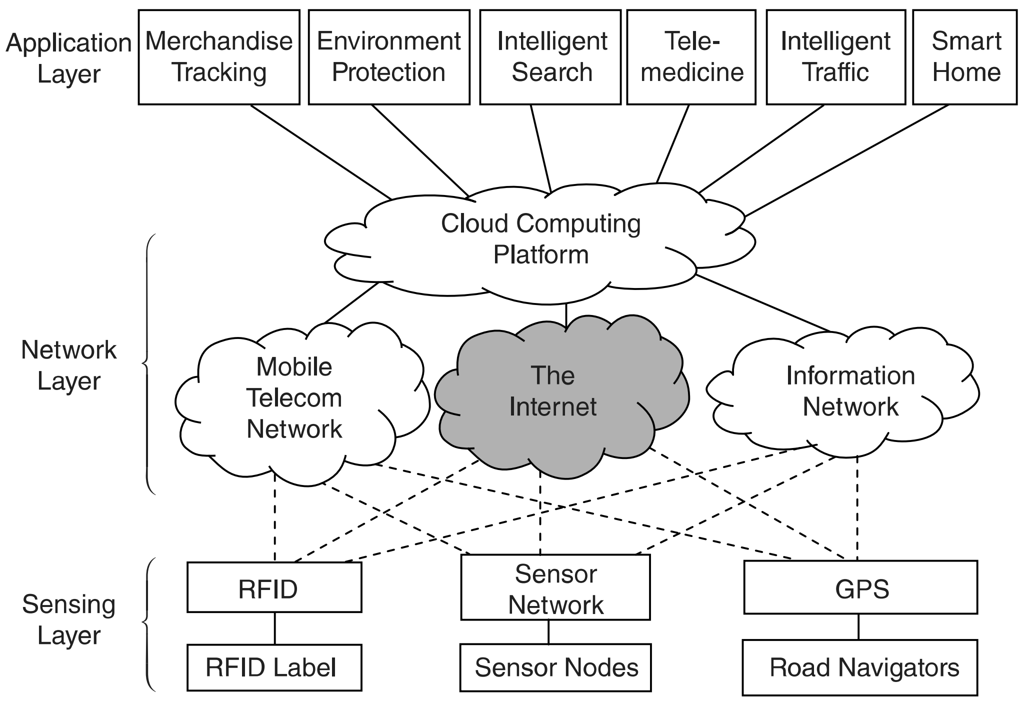



**Figure 3.3** The architecture of an Internet of Things (IoT) and its underlying technologies.

information generation devices: namely sensors, ZigBee devices, RFID tags and road-mapping GPS navigators, etc. The sensing devices are locally or wide-area connected in the form of sensor networks, RFID networks and GPS systems, etc. Signals or information collected at these sensing devices are linked to the applications through the cloud computing platforms at the middle layer.

The signal-processing clouds are built over the mobile networks, the Internet backbone and various information networks at the middle layer. In the IoT, the meaning of a sensing event does not follow a deterministic or syntactic model. In fact, the service-oriented architecture (SoA) model is adoptable here. Large number of sensors and filters are used to collect the raw data. Various compute and storage clouds and grids are used to process the data and transform them into information and knowledge formats. The sensory data is used to put together a decision-making system for intelligence applications. The middle layer is also considered as a semantic web or grid. Some actors (services, components, avatars) are self-referenced.

## 3.2 IoT Interactions with GPS, Clouds and Smart Machines

This covers the networking requirements, including the wireless, wireline and mobile core networks. We will examine the local and global positioning systems and cloud-based radio access networks possible for 5G mobile systems. Finally, we will study four frameworks for IoT interactions with the rest of the world.

### 3.2.1 Local versus Global Positioning Technologies

The requirement to integrate the cyber world and physical world is higher than ever. Localization becomes the bridge to interconnect these two worlds. Given WSN as an example, together with location information, the sensory data becomes meaningful. There is a real WSN project named ZebraNet, for the use of biologists who want to track and study animals. However, without locations, animals cannot be tracked, thus not enabling further studies. As another example of ubiquitous computing, the location is critical to differentiate various scenarios for providing personalized services for users.

According to the capabilities of diverse hardware, we classify the measuring techniques into six categories (from fine- to coarse-grained): location, distance, angle, area, hop count and neighborhood. Among them, the most powerful physical measurement is directly obtaining the position without any further computation. GPS is such an infrastructure. We discuss the other five measurements in this chapter, with emphasis on the basic principles of the measuring techniques. Basically, distance-related information can be obtained by radio signal strength or radio propagation time, angle information by antenna arrays, and area, hop count and neighborhood information by the fact that radios only exist for nodes in the vicinity.

#### 3.2.1.1 Local Positioning Technology

One method to determine the location of a device is through manual configuration, which is often infeasible for large-scale deployments or mobile systems. As a popular system, GPS is not suitable for indoor or underground environments and suffers from high hardware costs. Local positioning systems rely on high-density base stations being deployed, an expensive burden for most resource-constrained wireless *ad hoc* networks.

Limitations of the existing positioning systems motivate a novel scheme of network localization, in which some special nodes (a.k.a. anchors or beacons) know their global locations and the rest determine their locations by measuring the geographic information of their local neighboring nodes. Such a localization scheme for wireless multihop networks is alternatively described as "cooperative," "ad hoc," "in-network localization" or "self-localization."

Thus, network nodes cooperatively determine their locations by information sharing. The terms of "known" and "unknown" nodes refer to the nodes being aware and being unaware of their locations, respectively. Suppose a specific positioning process is one in which an unknown node determines its location based on the information provided by a number of known nodes. The unknown node is also known as a target node or a to-be-located node, while the known nodes as reference nodes.

Localization solutions consist of two basic stages: i) measuring geographic information from the ground truth of network deployment; and ii) computing node locations according to the measured data. Geographic information includes a variety of geometric relationships from coarse-grained neighbor-awareness to fine-grained internode rangings (e.g. distance or angle). Based on physical measurements, localization algorithms solve the problem: how to spread the location information from beacon nodes over a network-wide range.

Generally, the design of localization algorithms largely depends on a wide range of factors, including resource availability, accuracy requirements and deployment restrictions, and no particular algorithm is an absolute favorite across the spectrum. Due to hardware limitations, ranging is not always available for wireless devices. In such situations, range-free approaches are cost-effective alternatives, in which nodes merely know their neighbors.

Without direct distance ranging, the physical distance of a pair of nodes is estimated by the hop count or the proximity. The basic idea of hop count-based localization is to use hop-by-hop message delivery to calculate hop counts from nodes to anchors. The hop-count information is further converted to the distance estimates. Eventually, each node adopts trilateration or other methods to determine its location according to the estimated distances. Another possibility is to explore the relative proximity of nodes. When distance ranging is not available, the fact that one node is closer to some other node can aid the localization process.

#### 3.2.1.2 Satellite Technology for Global Positioning

Global positioning is done with multiple satellites deployed in outer space. The deployment of satellites is shown in Figure 3.4. Each satellite continually transmits messages that include the transmission time and satellite position. A GPS receiver calculates its position by precisely timing the signals sent by satellites. The receiver uses the messages it receives to determine the transit time and computes the distance to each satellite using light speed. Each of these distances and the satellites' locations define a signal sphere. The receiver is located at the intersection signal spheres from multiple satellites. More details of GPS and other positioning technologies are given in Section 3.4.4.

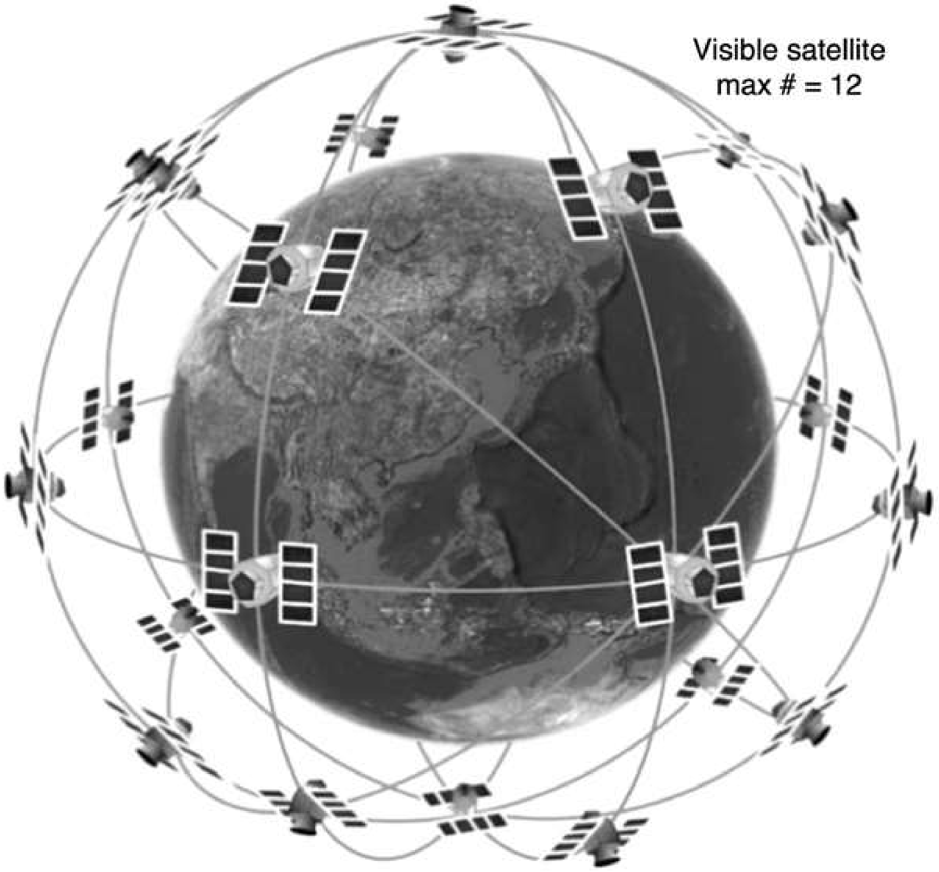



**Figure 3.4** The 24-satellite GPS architecture: the satellites circle the Earth twice per day in multiple layers of fixed orbits without interference to each other.

### Example 3.1 GPS System Developed in the USA

The US GPS is built with three segments. The space segments are the satellites circulating in outer space. The user segments include any moving or stationary objects such as airplanes, ships and moving vehicles on the Earth's surface. The control segment includes some ground antennas and master and monitor stations on the Earth's surface that are scattered globally. The uplink and downlink data types are different. The computed signal travel time is used for displaying the receiver location. A number of applications for GPS do make use of this cheap and highly accurate timing, including time transfer, traffic signal timing and synchronization with cell phone base stations.

The US Air Force develops, maintains and operates the space and control segments of the GPS system. There are 24 satellites deployed around the Earth (Figure 3.4) in fixed orbits. The satellites orbit at an altitude of approximately 20,200 km. GPS satellites broadcast signals from space, by which each GPS receiver calculates its three-dimensional location (latitude, longitude and altitude) plus the current time. The space segment is composed of 24 satellites in medium Earth orbit and also includes the boosters required to launch them into orbit.

GPS satellites circle the Earth twice a day in precise orbits and transmit signals to the Earth. GPS devices on the ground receive these signals and use triangulation to calculate the user's exact location. In general cases, four satellites are required to locate a single point on the Earth's surface. The system was initially developed for military use in 1975.

Now the system, under strict regulation, is open for civilian and commercial use, mainly in vehicle tracking and navigation applications. ■

### 3.2.2 Standalone versus Cloud-Centric IoT Applications

Typically, standalone IoT focuses on stable environments in which new applications would likely improve the quality of our lives: at home, while travelling, when sick, at work, when jogging and at the gym, just to cite a few. These environments are now equipped with objects with only primitive intelligence, mostly without any communication capabilities. Giving these objects the ability to communicate with each other and to elaborate the information perceived from their surroundings implies having different environments where a very wide range of applications can be deployed. These can be grouped into the following domains.

Transportation and logistics domain, healthcare domain, smart environment (home, office, plant) domain, personal and social domain; among the possible applications, we may distinguish between those either directly applicable or closer to our current living habits and those that are futuristic, which we can only imagine at the moment, since the technologies and/or our societies are not ready for their deployment. In the following subsections we provide a review of the short-medium term applications for each of these categories and a range of futuristic applications.

**Example 3.2 A Smart Power Grid supported by the Internet of Things**

A smart grid includes an intelligent monitoring system that keeps track of all electricity flowing in the system. Smart meters and sensors, a digital upgrade of current utility meters, track energy usage in real time so that both the customer and the utility company know how much is being used at any given time. Energy is paid for using "time of day" pricing, meaning electricity will cost more at peak times of use.

For example, when power is least expensive, the user can allow the smart grid to turn on selected home appliances such as washing machines or factory processes that can run at arbitrary hours. At peak times it could turn off selected appliances to reduce demand. More involved users will be able to use the smart meter to view energy usage remotely and make real-time decisions about energy consumption. A fridge or air conditioning system could be turned down remotely while residents are away.

With the development of WSNs, as well as low-power embedded systems and cloud computing, cloud-assisted IoT systems are gradually maturing to support smart and computation-intensive IoT applications involving a large amount of data. The home IoT applications can be upgraded by the emerging cloud computing environment. The scalable and elastic cloud-assisted framework shifts computation and storage into the network to reduce operational and maintenance costs. ■

#### 3.2.2.1 Cloud-Centric IoT System Applications

The information delivered from different domains (e.g. smart grid and healthcare) is difficult to understand and handle for the computer in the cloud service. With the support of a semantic model, an ontology-based approach can be used to implement information interaction and sharing in cloud-assisted home IoT. As shown in Figure 3.5, separate cloud systems can interoperate, with an additional root cloud providing different services for healthcare, energy management, convenience and entertainment, etc.

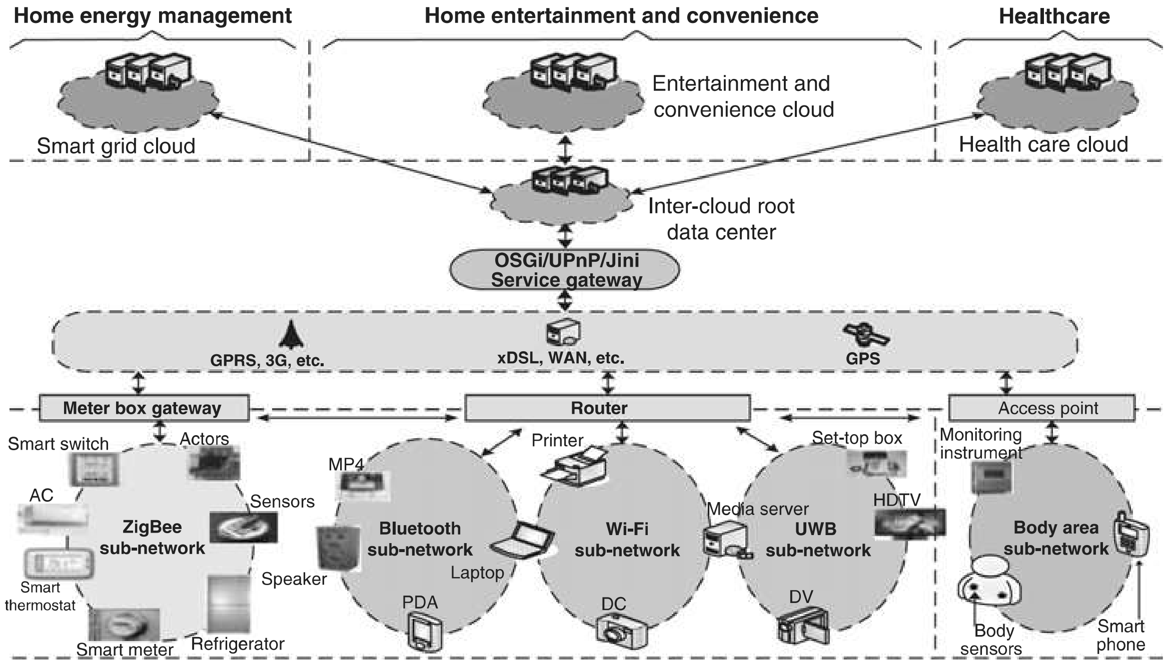



**Figure 3.5** Cloud-centric IoT system for smart home environment.

The service gateway implements various technologies, protocols, standards and services to diversify communications capabilities and integrate devices. Currently, most service gateways implement well-defined software modes and systems, such as Jini, UPnP and OSGi. In addition, the communication of heterogeneous objects in IoT is a major problem, because different objects provide different information in different formats for different purposes. The semantic web technologies and models may also be used to help solve this problem. The semantic web technologies can be applied to facilitate communication in home IoT applications.

In recent years, cloud computing has provided novel perspectives in cloud-assisted technologies for distinct purposes. A cloud-assisted communication system may include multiple cloud systems operating with different policies to share resources, so that end-to-end QoS to users can be maintained, even in the event of large fluctuations in computing load that cannot be handled by a single cloud system. It is known that the previous architectures for IoT have not taken into account this cloud-assisted capability. In our view, it is an important factor for IoT to achieve functionality completeness. Therefore, compared to the previous survey literatures, we propose a cloud-assisted layer for the advancement of IoT architecture. The following example demonstrates the joint effort between Intel and China Mobile towards the development of 5G mobile core networks.

**Example 3.3 Cloud-Based Radio Access Network (C-RAN) for 5G Mobile Systems**
A large number of base stations are used in current 3G or 4G mobile core networks. They are facing a series of problems: namely bulky in physical size, slow in data rate, air losses during handover between cells, and demand appreciable power to keep them running smoothly without interruption. The C-RAN is joint project between Intel and China Mobile toward an efficient solution to these problems. The idea is illustrated in Figure 3.6. Details of this C-RAN architecture can be found in the white paper by Chen and Ran, 2011. [4]

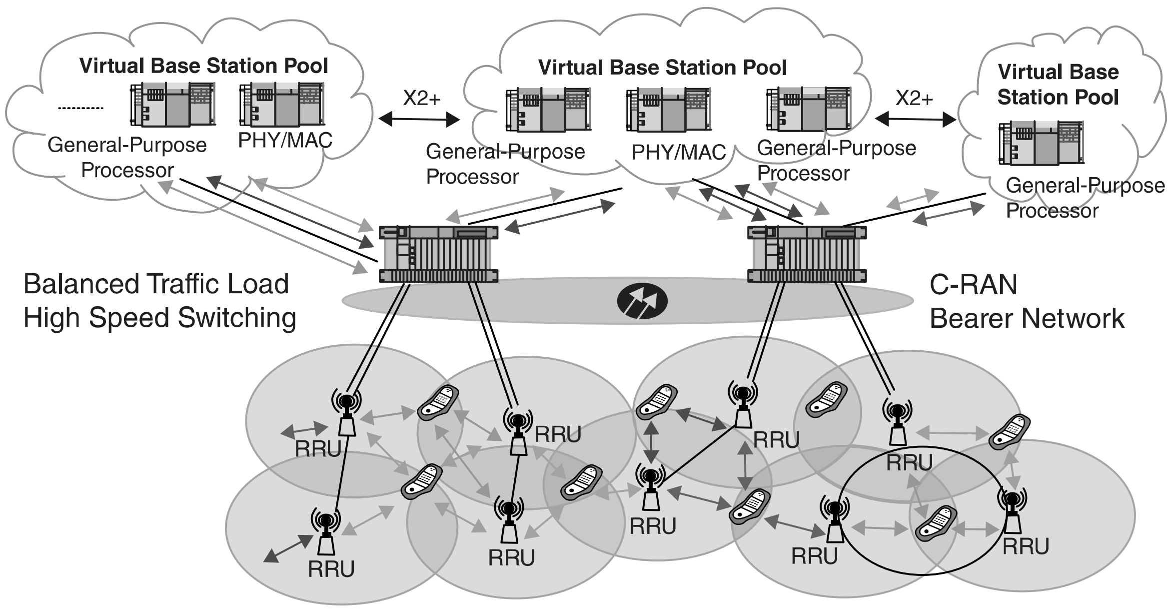



**Figure 3.6** Conceptual architecture of a cloud-based radio access network (C-RAN). (Courtesy of China Mobile Research Institute, 2009.)

The bulky antenna towers used in a conventional based station are replaced by a large number of small remote radio heads (RRH), which operate with little power (even solar energy can do the job) and get easily distributed with high density in populated user areas. The control and processing in physical-based stations are replaced by using virtual base station (VBS) pools housed in a hierarchy of cloud-based switching centers. Balanced traffic load between the RRHs and the VBS pools is enabled by using high-speed optical transport networks and switches with fiber cables and microwave links.

The advantages of using C-RAN are summarized in four aspects: i) centralized processing resource pool can support 10–1000 cells with high efficiency; ii) cooperative radios are used in multi-cell joint scheduling and processing, which solve the air loss and handover problems; iii) C-RAN offers real-time services by targeting to open IT platform, resources consolidation and flexible multi-standard operation and migration; and iv) a green and clean mobile telecommunication is realized with much less power assumption, lower operating expenses and fast system rolling out. Many other companies are also building similar C-RAN systems, including CISCO and Korean Telecommunication. ■

### 3.2.3 IoT Interaction Frameworks with Environments

As with any new kinds of "networks", IoT connects not only networked terminals such as mobile phones, computers and smart devices, but also daily life objects that until now have been to us just "un-networked things" or "inert objects". We first describe layer-based architecture for IoT, as shown in Figure 3.7.

- **Object sensing and information gathering**: The first step of enabling smart services is to collect contextual information about the environment, "things" and objects

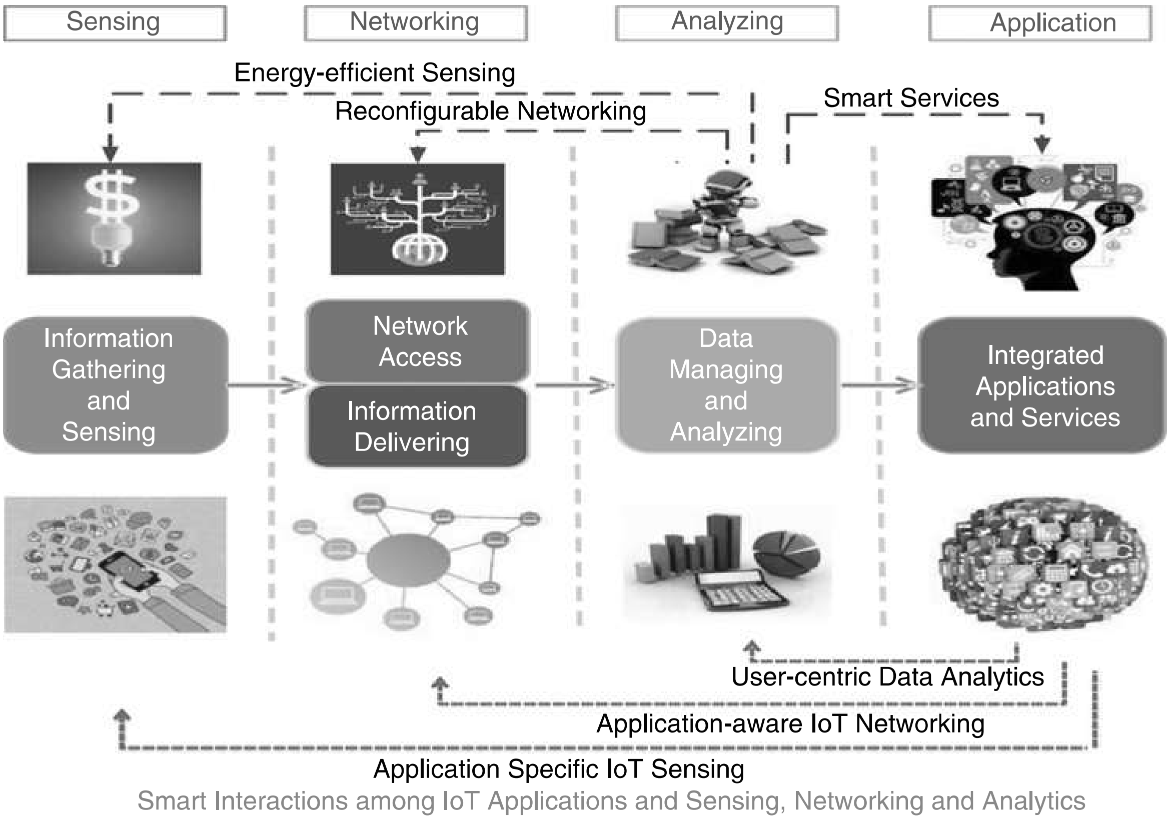



**Figure 3.7** Interactions among IoT sensing, mobile monitoring and cloud analytics.

of interest. For example, sensors can be used to continuously monitor a human's physiological activities and actions such as health status and motion patterns. RFID techniques can be utilized for collecting crucial personal information and storing them on a low-cost chip that is attached to an individual at all times.

- **Information delivering**: Various wireless technologies can be used for delivering the information, such as wireless sensor networks (WSNs), body area networks (BANs), WiFi, Bluetooth, Zigbee, GPRS, GSM, cellular and 3G, etc. Such diverse communication techniques can accommodate more applications into the system.
- **Information processing for smart services**: Ubiquitous machines must process information in both "autonomic" and "smart" ways in order to provide pervasive and autonomic services. For example, the meaningless information could be filtered out according to the users' interests in social networks.

IoT sensing can interact with many other cyber systems, as illustrated in Figure 3.7. For example, the user can require personalized analytics performance based on specific data. The profile of sensing data and networking strategy can be adjusted based on application-specific requirements. Based on the intelligence obtained by data analytics, energy-efficient sensing can be achieved through the interaction between the sensing layer and analyzing layer; the data analytics also can benefit smart services. As for the network layer, the management functions (e.g. network function virtualization together with software defined networks) realized in the analyzing layer also has the potential to help operators satisfy tight service level agreements, accurately monitor and manipulate network traffic, and minimize operating expenses.

**Table 3.1** Requirements of four IoT computing and communication frameworks.

| Framework | WSN | M2M | BAN | CPS |
|---|---|---|---|---|
| Sensing Requirement | XXXX | XX | XXX | XXX |
| Networking Demand | XX | XXXX | XX | XXXX |
| Analyzing Complexity | XX | XX | XXX | XXXX |
| Application Industrialization | XXXX | XXX | XX | X |
| Security Demand | X | XX | XXX | XXXX |

Through interfacing with WSNs, a wide range of information can be collected by sensors for M2M systems. Thus, in addition to M2M communications, machines also can act through the collected information by integrating with WSNs. With the capabilities of decision-making and autonomous control, M2M systems can be upgraded to CPS. Thus, CPS is an evolution of M2M by the introduction of more intelligent and interactive operations, under the architecture of IoT. Focusing on the different types of applications, IoT has different incarnations such as WSN, M2M, BAN and CPS.

In Table 3.1, we mark from one X to four X's (i.e. XXXX) to indicate the demand of different IoT frameworks on the relevant features listed in the row headings. More stars refer to higher demand of that particular feature under the column framework. CPS applications have the potential to benefit from massive wireless networks and smart devices, which would allow CPS applications to provide intelligent services based on knowledge from the surrounding physical world. We observe that WSNs are the very basic scenario of IoT. It is regarded as the supplement of M2M as the foundation of CPS. The CPS is evolved from M2M in intelligent information processing.

In what follows, we specify four wireless frameworks for the deployment of IoT applications. These appear as WSN, M2M, BAN and CPS, as briefed below:

- **Wireless Sensor Network (WSN)**: It consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, and to cooperatively pass their data through the network to a main location. WSNs, emphasizing the information perception through all kinds of sensor nodes, are the very basic scenario of IoT.
- **Machine to Machine (M2M) Communication**: Typically, M2M refers to data communications without or with limited human intervention, among various terminal devices such as computers, embedded processors, smart sensors/actuators and mobile devices, etc. The rationale behind M2M communications is based on three observations: i) a networked machine is more valuable than an isolated one; ii) when multiple machines are interconnected, more autonomous applications can be achieved; and iii) smart and ubiquitous services can be enabled by machine-type devices intelligently communicating with other devices at any time and anywhere.
- **Body-Area Network (BAN)**: A new type of network architecture inherited from sensor networks by the use of novel advances in lightweight, small-size, ultra-low-power and intelligent monitoring wearable sensors, which continuously monitor a human's physiological activities and actions, such as health status and motion patterns.
- **Cyber Physical System (CPS)**: It is a system of collaborating computational elements controlling physical entities.

## 3.3 Radio Frequency Identification (RFID)

RFID technology is a sort of non-contact information transfer mode realized by the radio frequency signals through space coupling (alternating magnetic field or electromagnetic field), and it achieves the purpose of automatic identification through the transferred information. In the late 1990s, an MIT group came up with the term of IoT. The progress of RFID has fueled up the IoT development. The US Auto-ID Center was the first to propose the concept of auto-ID tracking, which became one of the earliest forms of IoT deployment. In 2008, the US National Intelligence Council published a report on "Disruptive Civil Technologies", which identified the IoT as a crucial technology on national interest.

### 3.3.1 RFID Technology and Tagging Devices

RFID devices have varying sizes, power requirements, operating frequencies, amounts of rewriteable and nonvolatile storage, and software intelligence. They operate from a few centimeters to hundreds of meters. However, an internal power source is needed to enable large RFID devices to operate over large distances. Conversely, smaller RIFD devices do not need any power supply. RFID works through a combination of three functional components, i.e. RFID tag, RFID reader and reader antenna. Let us start with an example RFID.

#### 3.3.1.1 RFID Tags

It consists of a tiny silicon chip and a small antenna. The tag's components are enclosed within plastic, silicon or sometimes glass. Data stored in the microchip waits to be read. Typically, the tag's antenna receives electromagnetic energy from an RFID reader. Using power harvested from the reader's electromagnetic field, the tag returns radio signals back to the reader. The reader picks up the tag's radio signals and interprets the frequencies into meaningful data.

There are three types of RFID tags, i.e. active, semi-active and passive. An active RFID tag contains a battery and can transmit signals autonomously. A passive RFID tag has no battery and requires an external source to provoke signal transmission. A semi-active RFID tag is actually a battery-assisted passive RFID tag. However, the battery is only activated when an RFID reader sends an energizing signal. Based on the radio frequency used, the passive RFID tags operate from low frequency (LF) to high frequency (HF), and to ultra high frequency (UHF) domains.

#### 3.3.1.2 RFID Reader and Antenna

This is the device station that talks with the tags. A reader may support one or more antennae. Compared to the barcode, RFID has an important advantage, i.e. a reader device can detect objects without line of sight. Some RFID readers can identify multiple objects concurrently. An antenna is used to radiate the energy and then capture the return signal sent back from the tag. It can be integrated with a handheld reader device or connected to the reader by cable. The RFID reader antenna detects RFID tags similar to radar illuminating a target. However, RFID operates at shorter ranges.

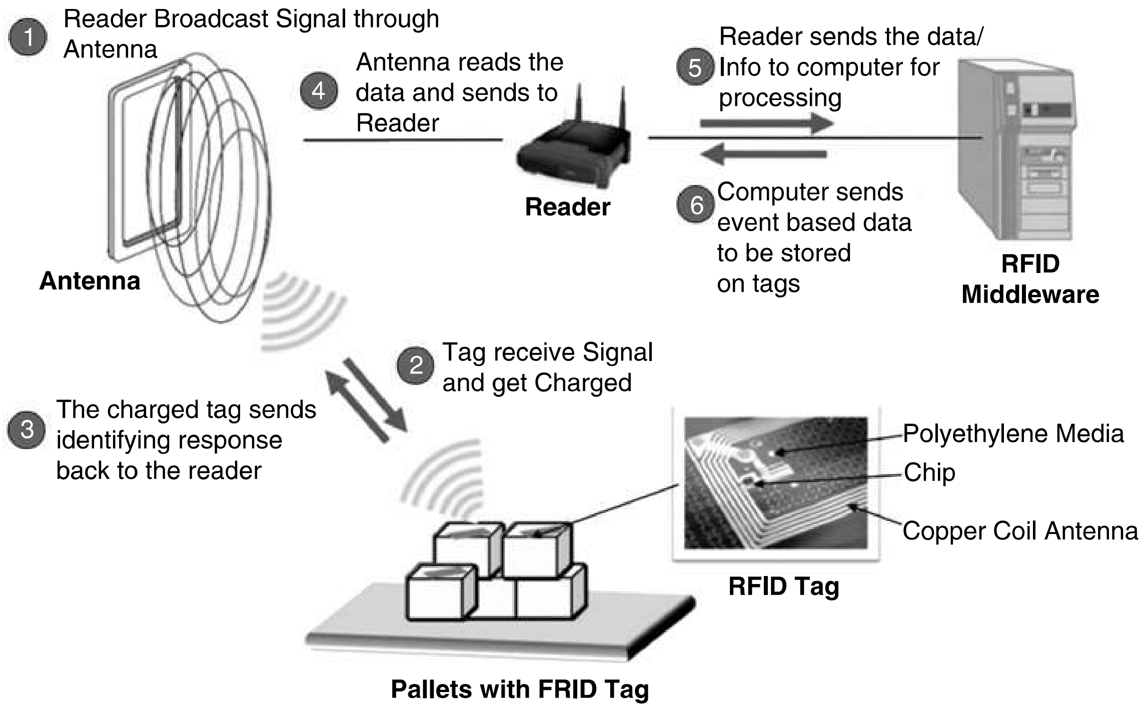



**Figure 3.8** RFID readers retrieve product data on e-labels (RFID tags) placed on package boxes.

**Example 3.4 RFID Technology for Merchandize Tagging or e-Labeling**
Electronic labels or RFID tags appear on merchandize or shipping boxes. The e-label is made from polyethylene media housing small IC chips and printed circuitry driven by a copper coil antenna. The tag itself has no power supply attached to it. The tag is energized by signal waves broadcast from the reader antenna. Figure 3.8 shows a sequence of six events. Events 1 to 3 show the energization and handshaking between the reader and the tag. Events 4 to 6 show how the antenna reads the data on the label to the backend computer for processing.

The computer sends updated event-based data to be stored on tags for future use. The RFID in the middle is executed by the backend computer to complete the reading and updating process. Of course, the idea can be modified to serve other remote identification purposes. For example, RFID labels are also used in department stores, supermarkets, inventory searches, the shipping industry, etc. ■

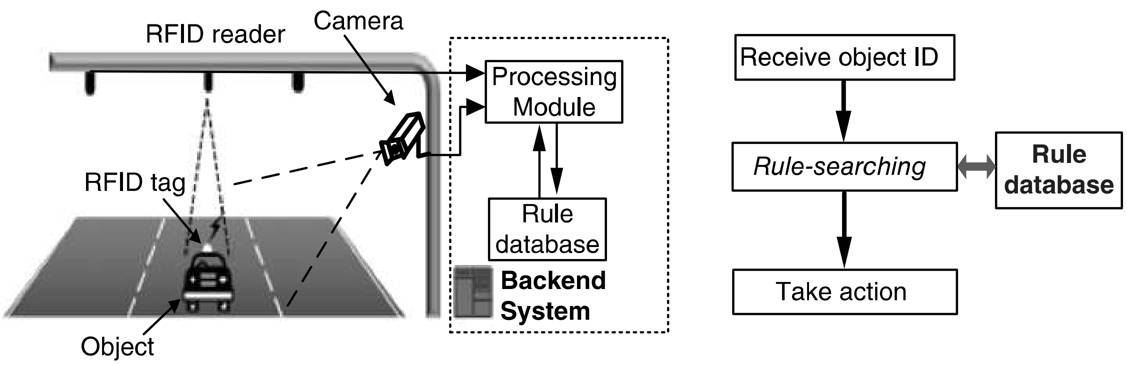
### 3.3.2 RFID System Architecture

RFID technology and products entered business applications in the 1980s. Now, the categories of RFID products increase considerably, various tags have been greatly developed with the cost constantly decreasing, and the industries with large-scale applications have started to expand. Derived from radar technology, RFID has an operating principle similar to that of radar. First, the reader sends out an electronic signal through the antenna, the tag emits the identification information stored internally after receiving the signal, then the reader receives and identifies the information sent back by the identification tag via the antenna, and finally the reader sends the identification result to the host.

A typical RFID system consists of an RFID tag, an RFID reader and a backend system. With a simple RF chip and an antenna, an RFID tag can store information that identifies the object to which it is attached. There are three types of RFID tags, i.e. passive tags, active tags and semi-active tags. A passive tag obtains energy through RF signals from the reader, while an active tag is powered by an embedded battery, which enables larger memory or more functionality. Though a semi-active tag communicates with RFID readers like a passive tag, additional modules can be supported through an internal battery. When it comes within the proximity of an RFID reader, the information stored in the tag is transferred to the reader, and onto a backend system, which can be a computer employed for processing this information and controlling the operation of other sub-system(s).

**Example 3.5 Automobile Speeding Watch in a Typical Rule-Search RFID System** When a vehicle carrying a RFID tag is speeding on a highway, it passes a check point equipped with an RFID reader, and the identification data of the vehicle is transmitted to the RFID reader and backend system. The vehicle's ID is checked against the database at the backend. The system checks the database to issue the citation. In Figure 3.9(a), the vehicle is being monitored by a camera to check its speed. If the speed limit is exceeded, based on the traffic monitory rule, some actions are triggered, such as issuing a citation ticket to the vehicle driver. Alternatively, the vehicle may be chased by a police car to avoid endangering other drivers sharing the road.

In Figure 3.9(b), the rule-searching process is executed by such a speeding-check RFID system. The basic format of a rule consists of a simple conditional statement and a series of action codes: if {condition (environmental parameters), then {<action1 (parameter1)>, <action2 (parameter2)>, where environmental parameters (e.g. temperature or humidity sensed by some sensors) are used to determine whether the condition of a rule is satisfied. An action represents the operation that the system has against the running vehicle. Given the example in Figure 3.9(a), the rule-searching result could be: if {Speed > 120 km/hr} then {notify the police()} ■




(a) Functional components in a RFID System

(b) Rule-searching process in RFID back-end system

**Figure 3.9** A typical example of an RFID system applied to automobile speeding check.

### 3.3.3 IoT Support of Supply Chain Management

RFID technology plays an important role in business and the marketplace. Many industrial, government and community services can benefit from these applications. These include activities or initiatives to promote the development of better and more efficient societies, cities and governments. The typical RFID IoT applications includes retailing and logistics services and supply chain management.

#### 3.3.3.1 Retailing and Logistics Services

Emergence of RFID applications depends strongly on adoption by retailers, logistics organizations and package-delivery companies. In particular, retailers may tag individual objects in order to solve a number of problems all at the same time: accurate inventories, loss control and ability to support unattended walk-through point of sale terminals (which promise to speed checkout, while reducing both shoplifting and labor costs). Cold-chain auditing and assurance could require tagging food and medicine with temperature-sensitive materials and/or electronics. Assuring or monitoring whether perishable materials are intact and/or need attention may entail communications among things such as refrigeration systems, automated data logging systems and human technicians.

For example, at the grocery store, you buy a carton of milk. The milk container will have an RFID tag that stores the milk's expiration date and price. When you lift the milk from the shelf, the shelf may display the milk's specific expiration date, or the information could be wirelessly sent to your personal digital assistant or cell phone. As you exit the store, you pass through doors with an embedded tag reader. This reader tabulates the cost of all the items in your shopping cart and sends the grocery bill to your bank. Product manufacturers know what you have bought and the store's computers know exactly how many of each product needs to be reordered.

Once you get home, you put your milk in the refrigerator, which is equipped with a tag reader. This smart refrigerator is capable of tracking all of the groceries stored in it. It can track the foods you use, how often you restock your refrigerator and can let you know when that milk and other foods spoil. Products are also tracked when they are thrown into a trash can or recycle bin. Based on the products you buy, your grocery store gets to know your unique preferences. Instead of receiving generic newsletters with weekly grocery specials, you might receive one created just for you.

#### 3.3.3.2 Supply Chain Management

Supply Chain Management can be aided by an RFID system. The idea is to manage a whole network of related businesses or partners involved in product manufacturing, delivery and services as required by the end customers. At any given time, market forces could demand changes from suppliers, logistics providers, locations and customers, and from any number of specialized participants in a supply chain. This variability has significant effects on the supply chain infrastructure, ranging from the foundation layers of establishing the electronic communication between the trading partners to the more complex configuration of the processes, and the arrangement of work flows that are essential to the fast production process.

A supply line combines the processes, methodologies, tools and delivery options to guide collaborative partners to work in a sequence to conduct business with high

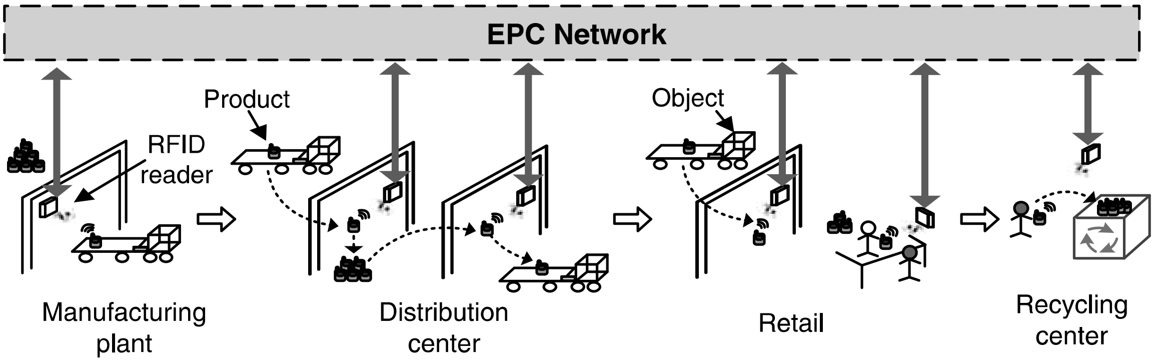



**Figure 3.10** Supply chain management in a multi-partner business pipeline.

efficiency and delivery speed. The cooperative companies must work in step quickly as the complexity and speed of the supply chain increases due to the effects of global competition, rapid price fluctuations, surging of oil prices, short product life cycles, expanded specialization and talent scarcity. A supply chain is an efficient network of facilities that procures materials, transforms these materials to finished products and finally distributes the finished products to customers. The following example could explain how supply chain can be aided by the IoT, which is particularly tailored to promote business efficiency and fast growth.

**Example 3.6 Supply Chain Management Aided by the Internet of Things**

Supply chain management is a process used by companies to ensure that their supply chain is efficient and cost-effective. In Figure 3.10, the supply chain for the production and sales of consumer products is illustrated. The supply chain involves material or component suppliers, distribution centers, communication links, cloud datacenters, large number of retail stores, corporate headquarters (like Wal-Mart) and bank payments, etc. These are business partners that are linked by satellite, Internet, wired and wireless networks, truck, train or shipping companies, electronic banking, cloud providers, etc.

Sensors, RFID tags and GPS devices could be placed everywhere along the supply chain. The whole idea is to promote on-line business, e-commerce or mobile transactions. Supply chain management is comprised of five major stages of operations:

1) **Planning and Coordination:** A plan or strategy must be developed to address how a good or service can satisfy the needs of customers.
2) **Material and Equipment Supplies:** This phase involves building a strong relation with the raw material suppliers and also planning methods for shipping, delivery and payment.
3) **Manufacturing and Testing:** The product is tested, manufactured and scheduled for delivery.
4) **Delivery of Products:** Customer orders are taken and delivery of the goods is planned.
5) **After Sale Service and Returns:** At this stage customers may return defective products and the company also addresses customers' demands. Supply chain software is used by many companies for efficient supply chain management.

■

## 3.4 Sensors, Wireless Sensor Networks and GPS Systems

With the development of circuit design, signal processing and Micro-Electro Mechanical Systems (MEMS), various kinds of sensors are produced, from simple optical sensors and temperature sensors to complicated sensors such as carbon dioxide sensors, etc. To use which kind of sensors is usually determined by the specific application requirements. In general, processors interact with sensors through either analog signals or digital signals. Analog signal-based sensors output physically measured analog quantity such as voltage. The analog quantities must be digitalized before being used. Therefore, these sensors need external analog-digital converters as well as additional calibration technique. In digital signal-based sensors, the interaction between processors and sensors is simplified, since digital quantities are provided by sensors directly.

### 3.4.1 Sensor Hardware and Operating Systems

Sensors bridge the physical world and electronic systems. Sensors yield data in the form of analog or digitized signals that are fed to the sensor's node for immediate processing. However, depending on the circumstances, some form of specialized pre-processing or filtering can also take place beforehand, either as part of an algorithm implemented in the sensor node, or as part of an intermediate hardware component, although the former case has become prevalent.

We consider two categories of sensors: one for environmental surveillance and the other for body sensing. Environmental surveillance sensors are used to collect environmental information. Body sensors are deployed to gather vital body data. As shown in Table 3.2, typical environmental surveillance sensors include sensors for visible light, temperature, humidity, pressure, magnetism, acceleration, gyroscopic, sound, smoke, passive RF-optics, structured light, soil moisture, carbon dioxide ($CO_2$) gas, etc.

#### 3.4.1.1 Inertial Motion Sensors

In this category, accelerometers and gyroscopes are by far the most common devices employed to estimate and monitor body posture, and miscellaneous human motion

**Table 3.2** Characteristics of environmental surveillance sensors.

| Manufacturer | Sensor | Voltage (V) | Power | Sampling Time |
|---|---|---|---|---|
| Taos | Visible light | 2.7–5.5 | 1.9 mA | 330 us |
| Dallas Semiconductor | Temperature | 2.5–5.5 | 1 mA | 400 ms |
| Sensirion | Humidity | 2.4–5.5 | 550 uA | 300 ms |
| Intersema | Pressure | 2.2–3.6 | 1 mA | 35 ms |
| Honeywell | Magnetism | Any | 4 mA | 30 us |
| Analog Devices | Acceleration | 2.5–3.3 | 2 mA | 10 ms |
| Panasonic | Sound | 2–10 | 0.5 mA | 1 ms |
| Motorola | Smoke | 6–12 | 5 uA | – |
| Melixis | RF-Opticals | Any | 0 mA | 1 ms |
| Li-Cor | Structured light | Any | 0 mA | 1 ms |
| Ech2o | Soil moisture | 2–5 | 2 mA | 10 ms |

patterns. This capability is indispensable for many types of applications, especially in the realm of healthcare, sports and console gaming. To this end, accelerometers measure gravitational pull and inclination, whereas gyroscopes measure angular displacement. In general, their combined use yields orientation information and diverse user motion patterns.

#### 3.4.1.2 Bioelectrical Sensors

These particular types of sensors are employed to measure electrical variations over the user/patient's skin that can be directly or indirectly correlated to the current activity or condition of a body organ. Electrocardiographic sensors are typical examples of these, which usually take the form of circular pads that are strategically placed around the human torso and extremities to monitor heart activity (ECG). Similar types of sensors placed over the skin are employed to measure the electrical activity of skeletal muscles (EMG) in order to help in the diagnosis of nerve and muscle disorders.

A body sensor node mainly consists of two parts: the physiological signal sensor(s) and the radio platform, to which multiple body sensors can be connected. The general functionality of body sensors is to collect analog signals that correspond to human's physiological activities or body actions. Such an analog signal can be acquired by the corresponding radio-equipped board in a wired fashion, where the analog signal is digitized. Finally, the digital signal is forwarded by the radio transceiver. The types of commercially available body sensors are listed as follows:

- **Electrochemical sensors:** These types of sensors generate an electrical output driven by a small chemical reaction between the sensor's chemical agent and bodily substance. A good example is the blood glucose sensor, which measures the amount of glucose circulating in the blood. Another example is the monitoring of carbon dioxide ($CO_2$) concentration levels in human respiration.
- **Optical sensors:** Devices that emit and receive light in both the visible and the infrared light bands are commonly employed in the non-invasive measurement of oxygen saturation in blood circulating in the human body. To this end, a pulse oximeter measures the degree of light absorption as light passes through the user/patient's blood vessels and arteries.
- **Temperature sensors:** This popular sensor type is placed over the skin in various places around the human body, and is routinely employed during physiological assessment of patients.

In a smart IoT environment, we may monitor and explore the physical world to the extreme. For example, humans can neither tolerate a temperature of over 1000 °C nor distinguish subtle changes of temperature. Therefore, the IoT for environmental protection has presented higher requirements on wide-range temperature measurement using heat-resistant sensors. Our daily life has been affected by extensive use of sensors, such as temperature and humidity sensors in an air conditioner, a temperature controller in a water heater, a sound controller for the lamp in a corridor, and a remote control for a TV set, etc.

Furthermore, sensors have been widely applied in fields such as environmental protection, medical health, industry and agriculture and military and national defense. A sensor is an apparatus or device that can sense the specified measure and convert it to a usable output signal according to a certain rule, and is generally constituted by sensing element, transduction element and basic circuit. The sensing element refers to the part in a sensor that can directly sense physical quantity; the transduction element converts the output of sensing element to circuit parameters (i.e. voltage and inductance); and finally the basic circuit converts the circuit parameters to electrical output.

#### 3.4.1.3 Sensor Architecture Design

Figure 3.11(a) shows a typical sensor node with sensor, radio and memory modules. The sensor module consists of a sensor, a filter and an analog-to-digital converter (ADC). The sensor converts some form of energy to analog electric signals, which are bandpass-filtered and digitized by the ADC for further processing. We will discuss the radio systems for BANs and WPANs used for transmissions of sensed data in the next section. A sensor is easily confused with several concepts, such as sensor node, wireless sensor node and wireless sensor network. The sensor acts here as a signal converter.

The sensor node is usually called a sensor plus microprocessor, the function of which is to further convert the analog signal to the digital signal. The wireless sensor node further integrates the wireless communication chip, etc. on the basis of the traditional

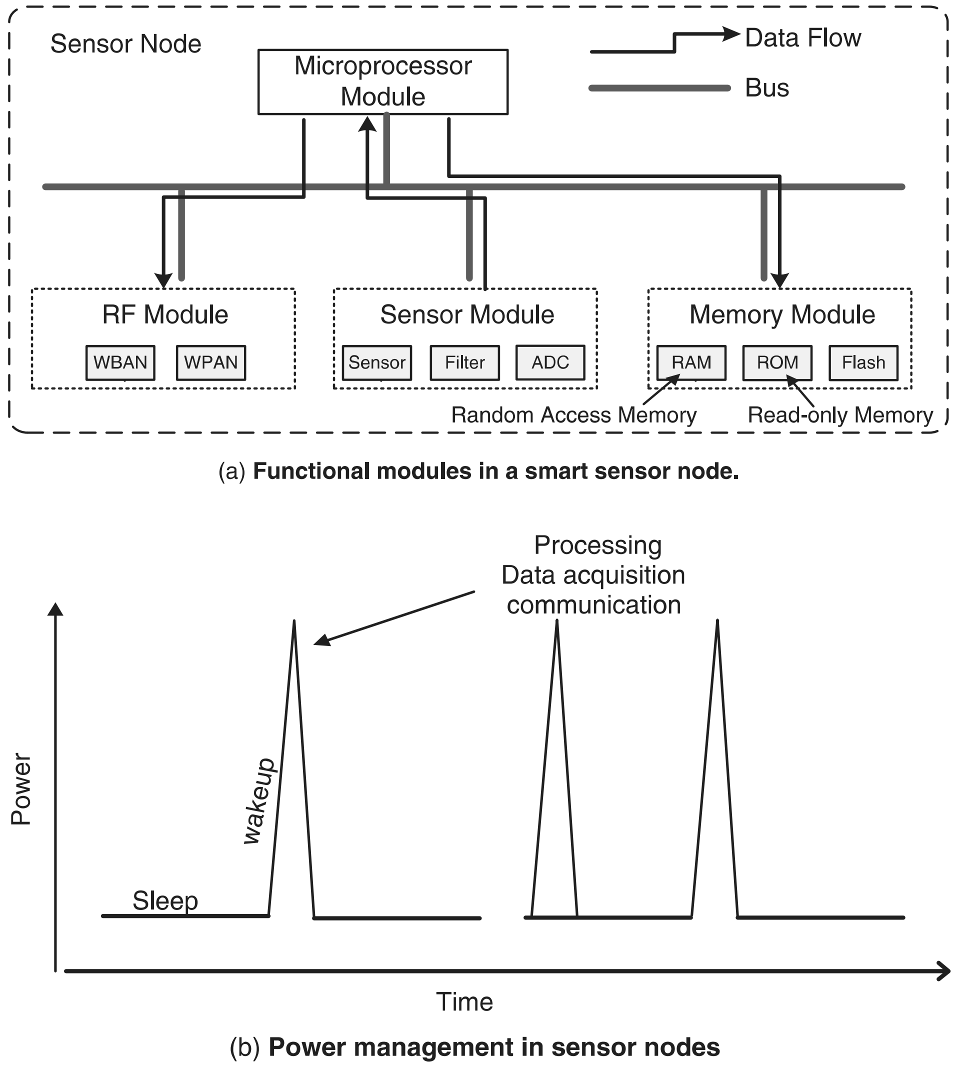



**Figure 3.11** Power management in typical sensor operations.

sensor node: micro-operation systems such as TinyOS and Contiki are installed with some embedded programs to analyze and process the information received and transmit it via the network. If multiple wireless sensor nodes are placed to make them interconnect and form one *ad hoc* network, such a network is called a sensor network or a wireless sensor network.

#### 3.4.1.4 Power Consumption

Wireless sensor nodes are generally deployed in the open air, so they cannot gain their power supply through wires. Therefore, their hardware design must consider energy saving as an important design objective. For example, under normal operating modes, the power of a typical sensor processor is between 3 and 15 mW. The periodic power management of a typical sensor is illustrated in Figure 3.11(b). The sleep, wake-up and processing spikes take the form of a periodic cycle, repeating within a fixed time interval. Most of time, the sensor is in sleep mode, with the device waking up periodically. The spike of power enables data collection and communication operations. Infrequently, triggering events are detected and they activate the sensor devices. The long life time of sensor operation may span over months to years, depending on whether the device is powered by solar energy or by other sustained energy sources.

#### 3.4.1.5 Price and Size

Generally, a larger-scale network deployment requires a larger amount of sensor nodes to complete a complicated task. There is a tradeoff between node price and number of sensor nodes with a fixed budget. Therefore, their hardware design must consider cost-effectiveness as its critical design objective. Typically, wireless sensor nodes should be easy to transport and deploy, so their hardware design must regard micro-miniaturization as an important design objective. However, the limitation of node size also restricts the functions of the sensor nodes.

#### 3.4.1.6 Flexibility and Expansibility

Sensor nodes are applied to different kinds of applications, so their hardware and software design must be flexible and expansible. In addition, flexibility and expansibility are important safeguards to realize the large-scale deployment of the sensor network. The hardware design of nodes should meet certain standard interfaces; for example, the interface of nodes and sensor plates make it beneficial to install sensors with different functions on nodes.

Moreover, the software design must be tailorable and able to install software modules with different functions according to the demands of different applications. Meanwhile, the design of software must also consider the extendibility of the system in time domains. For example, the sensor network should be able to add new nodes continuously, with this process not influencing the existing performance of the network. Another example is that node software should be able to update programs automatically through the network rather than redeploy each time after the deployed nodes are taken back and burned.

Table 3.3 compares the features of representative sensor platforms in terms of OS support, wireless standard and data rate, etc. We focus on operating system support, wireless standard used, maximum data rate, outdoor range and power level. These features of the system reveal the main characteristics of a sensor from the general application

**Table 3.3** Comparison of typical sensor nodes in daily life applications.

| Name | OS Support | Wireless Standard | Date Rate | Outdoor Range (m) |
|---|---|---|---|---|
| BAN node | TinyOS | IEEE 802.15.4 | 250 kbps | 50 |
| BTNode | TinyOS | Bluetooth | 24 Mbps | 100 |
| eyesIFX | TinyOS | TDA5250 | 64 kbps | – |
| iMote | TinyOS | Bluetooth | 720 kbps | 30 |
| iMote2 | TinyOS or .NET | IEEE 802.15.4 | 250 kbps | 30 |
| IRIS | TinyOS | IEEE 802.15.4 | 250 kbps | 300 |
| Micaz | TinyOS | IEEE 802.15.4 | 250 kbps | 75 to 100 |
| Mica2 | TinyOS | IEEE 802.15.4 | 38.4 kbps | >100 |
| Mulle | TCP/IP or TinyOS | Any | 250 kbps | >10 |
| TelOS | TinyOS | Bluetooth or IEEE 802.15.4 | 250 kbps | 75 to 100 |
| ZigBit | ZDK | IEEE 802.15.4 | 250 kbps | 3700 |

designer's perspective. We can see that all sensors achieve low power consumption, but possess low data rates ranging from 38.4 to 720 kbps, which is insufficient for large-scale body sensor networks or applications involving multimedia data traffic such as video streaming. The package running on the IEEE 802.15.4 ZigBee sensor has been widely adopted. Bluetooth turns out to be less energy inefficient. Interference from other radio devices sharing the 2.4 GHz ISM band may pose another problem when using them for building body-area networks.

#### 3.4.1.7 Robustness

Robustness is an important safeguard to realize the long-time deployment of sensor networks. For common computers, if the system crashes, people can reboot it to recover the system; however, this is not useful for sensor nodes. Therefore, the design of the node program must be robust to guarantee that the nodes can work efficiently for a long time. For example, if the cost of the hardware design permits, we can adopt multi-form sensors, so that even if one kind of sensor breaks down, the other one can be used for the whole system. When designing software, we usually need to modularize the functions and have a total test of each function module before system deployment.

The next section is devoted to the selection of hardware components, communication interfaces, power supplies and operating systems for sensor module design and applications.

#### 3.4.1.8 Energy-Supplying Devices

Generally, sensor nodes are battery-powered, which makes the nodes more easily deployed. In theory, a battery with 2000 mAh capacity can continue to output a 10 mA current for 200 hours. However, because of various factors such as voltage changes and environmental changes, the capacity of the battery cannot be utilized totally. Besides being battery-powered, nodes can also use renewable energy sources such as solar energy and wind energy. For example, under direct sunlight, a one-square-inch solar panel can provide 10 mW of electric energy; while under indoor lighting, the same panel can provide 10–100 uW electric energy.

The electric energy collected in the daytime can be used by the nodes at night. The key technology to utilize renewable energy is how to store it, with two kinds of technology presently in use. One is to use rechargeable batteries, their main advantages being that there is relatively less self-discharge and higher utilization rate of electric energy, and main disadvantages being that the charging efficiency is relatively lower and the charging times are limited. Another relatively new technology is to use ultra-capacitors, with their main advantages being that the charging efficiency is high and the charging times can reach 1 million. Also, they are not easily influenced by factors such as temperature and vibration.

**Example 3.7 Networked Sensor Applications for Environmental Protection:** Many important environmental protection schemes can be supported by wireless sensor networks. Natural environment protection scenarios may include: i) seismic structure response to damage assessment after earthquakes; ii) contamination transport for pollution control; iii) ocean pollution control by monitoring marine microorganisms; and iv) ecosystem bio-complexity analysis through sensor monitory. These environmental protection scenarios require the use of a large number of micro-sensors, on-board processing and wireless interfaces feasible at very small scale, which can monitor phenomena up-close. These schemes enable spatially and temporally dense environmental monitoring. Large-scale distributed embedded sensing can now reveal some phenomenon, which has not been easily observed previously. ■

#### 3.4.1.9 Microprocessors

The microprocessor is the core responsible for computing in wireless sensing nodes. The present microprocessor chips also integrate internal storage, flash memory, module converters, digital Io, etc. This kind of deeply integrated characteristic makes them particularly suitable for use in the wireless sensing network. We are now going to analyze several processor key characteristics, which influence the whole operating performance of the nodes.

#### 3.4.1.10 Communication Chips

Communication chips are important components of wireless sensing nodes. There are two main characteristics of the energy consumption of communication chips: one is that they consume the largest proportion of the energy in one wireless sensing node. For example, in the frequently-used TelosB Node, at present the current of CPU is only 50 0uA under normal circumstances, while the current reaches nearly 20 mA when the communication chips are sending and receiving data.

When communication chips with low power consumption are sending and receiving data, their energy consumption shows little difference. This means that as long as communication chips are open, they are consuming almost the same amount of energy whether they are sending and receiving data or not. Generally, the transmission distance of the communication chip is an important index for us to choose the sensing node. Their transmission distance is governed by the transmitted power of the chip.

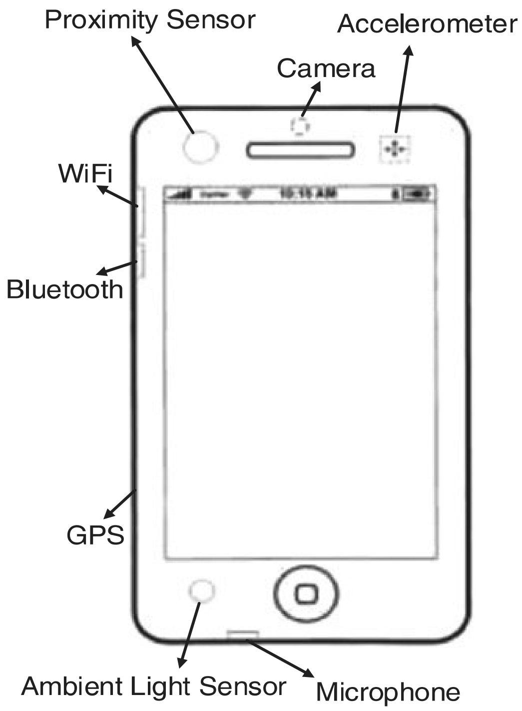



**Figure 3.12** Sensor devices built inside a typical smart phone in 2016.

#### 3.4.1.11 Operating System in Sensor Nodes

As the core of the sensor node software system, the node operation system provides the upper applications with hardware drives, resource management, task scheduling, programmatic interfaces, etc. The main characteristic that makes the node operation system different from the traditional one is that its hardware platform resource is really limited. Typical sensor node operation system includes TinyOS and MOS, etc. The sensor OS is extremely miniaturized and the TinyOS is the most widely-used in wireless sensing networks.

### 3.4.2 Sensing through Smart Phones

Regarding IoT sensing, a big change is happening around us, such as mobile phones now having directly integrated various sensors for sensing the physical world. As shown in Figure 3.12, more and more smart phones are equipped with devices such as GPS receiver, camera, sound recorder, thermometer, altimeter and barometer, etc. Due to the intrinsic capability of connection to the Internet in mobile environments, the smart phone enables sensor devices to be migrated to a global user group. Thus, the smart phone not only builds a bridge between sensor devices and the Internet, but also connects humans with social and cyber worlds, and so becomes an important element for IoT.

Nowadays, assessment of the user's health status is a promising research topic in the area of health IoT. There are various factors that can be used to make the assessment, such as mood, social interactions, sleeping habits, activity levels, perceived life satisfaction levels, etc. However, prior studies required face-to-face interactions between study coordinator and subject, thereby limiting the geographic reach and the scale of these studies.

**Table 3.4** Sensors and usage installed at a typical smart phone.

| Sensor/Data Name | Type | Sampling Period |
|---|---|---|
| Memory, CPU load, CPU utilization, Battery, Network traffic, Connectivity status | System | 1 s |
| Geo-location | Sensor | 10 s |
| Accelerometer, Magnetometer, Gyroscope | Sensor | 100 ms |
| Proximity, Pressure, Light, Humidity, Temperature | Sensor | 1 s |
| Phone Activity, SMS, MMS | User Activity | 1 s |
| Screen state, Bluetooth, WiFi | User Activity | 3 min |

Table 3.4 shows several examples of data that can be captured by a smart phone. The sensors built into modern smart phones enable various information collections, such as physical activity, location, mobility patterns, social interactions (e.g. using proximity sensors) and heart rate. In addition, phone usage patterns and trends can also provide highly valuable context information. Such data include browsing histories, communication habits (calls, texting), social networking activities and app usage.

**Example 3.8 Smart Phone or Smart Watch Features for Healthcare Applications**
Various sensory data, such as accelerometer, GPS, SMS, camera, sound recorder, thermometer, altimeter and barometer, etc., can provide additional contextual information that are essential for a better understanding of trends and outliers in survey responses (e.g. mood-related survey responses can be different when submitted from home, the workplace, when on vacation, etc.). Also, smart phones make it easier to collect data over extended periods of time.

With digitalization of medical information and rapid distribution of smart devices, currently healthcare services are actively planned and developed based on smart devices. By 2015, 500 million smart phone users were expected to apply mobile health application, especially for exercise, diet and chronic disease management. Unlike most other chronic diseases, diabetes can be managed by the patient. Therefore, smart mobile device can be a universal tool for self-diabetes management because of its high penetration and functions.

A mobile healthcare application for Android OS was developed to provide self-diabetes management. The application consists of Diabetes management, Weight management, Cardio-cerebrovascular risk evaluation, Stress and depression evaluation and Exercise management. With the support of a smart phone or smart watch, various healthcare related data can be collected, such as Heart rate, Breathing Rate, Skin temperature, Duration time of sleep, Activity level (e.g. Static, Walking, Running), Facial expression video, etc. ■

### 3.4.3 Wireless Sensor Networks and Body Area Networks

As shown in Table 3.5, WSNs have been classified into 3 generations over the past 30 years. The sensors used in the first generation were mainly vehicle-placed or air-dropped single sensors. They were bulky, like a shoe box, and weighed several

**Table 3.5** Three generations of wireless sensor networks.

| WSN Features | First Gen. (1990s) | Sec. Gen. (2000s) | Third Gen. (2010s) |
|---|---|---|---|
| Manufacturers | Custom constructors, e.g. for TRSS | Crossbow Technology, Inc. Sensoria Corp, Ember Corp. | Dust, Inc. and others |
| Physical Size | Large shoe box and up | Pack of cards to shoe box | Dust particle |
| Weight | Kilograms | Grams | Negligible |
| Node architecture | Separate sensing, processing and communication | Integrated sensing, processing and communication | Integrated sensing, processing and communication |
| Topology | Point-to-Point, star | Client server, peer-to-peer | Peer to peer |
| Power supply lifetime | Large batteries; hours, days and longer | AA batteries; days to weeks | Solar; months to years |
| Deployment | Vehicle-placed or air-drop single sensors | Hand-placed | Embedded, sprinkled, left behind |

kilograms. Networks assumed only-star or point-to-point topologies and were powered by large batteries that could last for hours or days. In the second generation, the sensors become smaller, like a pack of play cards, and weighed several grams, and the AA batteries lasted for days and weeks. They appeared in client-server or P2P configurations. The current generation are the size of dust particles, of negligible weight, and appear in P2P networks for embedded and remote applications.

Wireless *ad hoc* sensor networks apply a large number of (mostly stationary) sensors. Aside from the deployment of sensors at the ocean's surface or the use of mobile, unmanned, robotic sensors in military operations, most nodes in a smart sensor network are stationary. Networks of 10,000 or even 100,000 nodes are envisioned in the future and scalability becomes a demand. Low energy use is expected in the modern sensors. Since in many applications the sensor nodes will be placed in a remote area, service of a node may not be possible. In this case, the lifetime of a node is determined by the battery life, thereby requiring reduced energy expenditure or the use of solar energy to power the devices.

Advances in wireless communication technologies, such as wearable and implantable biosensors, along with recent developments in the embedded computing area, are enabling the design, development and implementation of body area networks. This class of networks is paving the way for the deployment of innovative healthcare monitoring applications. The differences between BAN and WSNs are listed as follows:

#### 3.4.3.1 Deployment and Density

The number of sensor/actuator nodes deployed by the user depends on different factors. Typically, BAN nodes are placed strategically on the human body, or are hidden under clothing. In addition, BANs do not employ redundant nodes to cope with diverse types of failures. An otherwise common design provision is the conventional WSNs. Consequently, BANs are not node-dense. WSNs, however, are often deployed in places that may not be easily accessed by the operators, which requires that more nodes be placed to compensate for any node failures.

#### 3.4.3.2 Data Rate

Most WSNs are employed for event-based monitoring, where events can happen at irregular intervals. By comparison, BANs are employed for registering a human's physiological activities and actions, which may occur in a more periodic manner, and may result in the applications' data streams exhibiting relatively stable rates.

#### 3.4.3.3 Latency

This requirement is dictated by the applications, and may be traded for improved reliability and energy consumption. However, while energy conservation is definitely beneficial, replacement of batteries in BAN nodes is much easier done than in WSNs, whose nodes can be physically unreachable after deployment. Therefore, it may be necessary to maximize the battery life-time in a WSN at the expense of higher latency.

#### 3.4.3.4 Mobility

BAN users may move around. Therefore, BAN nodes share the same mobility pattern, unlike WSN nodes that are usually considered stationary. BANs are commonly regarded as an enabling technology for a variety of applications, including health and fitness monitoring, emergency response and device control. Recent breakthroughs in solid-state electronics allow the creation of low-power, low-profile devices that can be modularly interconnected in order to create so-called sensor nodes comprised of one or more sensor devices, an MCU, and a radio transceiver that eliminates the need for wires to communicate with the node in order to transfer the collected data. BANs are merely at the beginning stage. Figure 3.13 illustrates a general architecture of a BAN-based health monitoring system.

In their most basic form, sensor devices operate by preloading MCUs with binary programs that access low-level hardware interfaces, which in turn obtain data from the actual sensor devices. Programs contain the necessary instructions for sensor devices

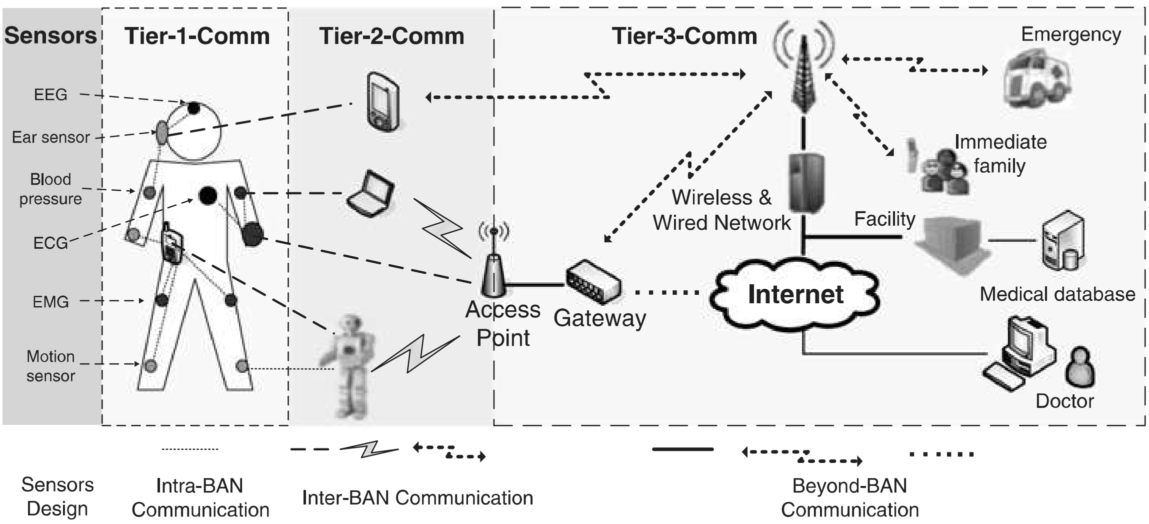



**Figure 3.13** A three-tier architecture based on a BAN communications system.

to collect one or more readings in a particular time period. Raw sensor data can be subsequently processed in order to convert it to meaningful information that can be interpreted after it has been transmitted by the radio chip to an external device or system for further analysis. As their name implies, sensor nodes are meant to be either worn around or implanted in the human body.

Moreover, two or more sensor devices in their immediate vicinity can establish wireless links in order to coordinate their joint operation, thus creating a networked system. Therefore, the existing literature often refers to BANs as wireless BAN (WBAN) or Wireless Body Area Sensor Network (WBASN). The rest of this section introduces some of the most relevant advances in BAN technology, followed by a description of important technical challenges that researchers must tackle in order to make BANs efficient, reliable and economical.

This system monitors ECG: (electroencephalography) EEG, (electromyography) EMG, motion sensors and blood pressure sensors send data to nearby personal server (PS) devices. Then, through a Bluetooth/WLAN connection, these data are streamed remotely to a medical doctor's site for real-time diagnosis, to a medical database for record keeping, or to the corresponding equipment that issues an emergency alert. In this article, we separate the BAN communications architecture into three components.

Tier-1-Comm design (i.e. intra-BAN communications), Tier-2-Comm design (i.e. inter-BAN communications) and Tier-3-Comm design (i.e. beyond-BAN communications), are as shown in Figure 3.13. These components cover multiple aspects that range from low-level to high-level design issues, and facilitate the creation of a component-based, efficient BAN system for a wide range of applications. By customizing each design component, for example cost, coverage, efficiency, bandwidth, QoS, etc., specific requirements can be achieved according to specific application contexts and market demands.

### 3.4.4 Global Positioning Systems

Location-based service (LBS) is a key enabling technology for IoT applications. Most of the "things" in the IoT domain are interconnected via various wireless communication networks. When sensory data collected by IoT sensing technologies are sent to clouds, the associated location data is important. Without location information, the basic knowledge of the circumstance and users cannot be obtained, which results in the failure of the IoT applications. Thus, localization is a critical process.

One popular method to determine the location of a device is through Global Positioning Systems (GPS). An active GPS (aGPS) receiver can receive satellite signals and transmit position information to an aGPS control center. The aGPS is becoming the standard for companies who wish to monitor fleet vehicles as well as other heavy equipment. Real-time GPS tracking is practical for acquiring immediate and detailed information about large numbers of vehicles or objects that are being tracked. This could be the car rental business that provides cars for many customers. Real-time vehicle tracking processes are divided into the following four steps:

1) GPS receiver in each car accesses receiving signals from a network of satellites.
2) The collected satellite information is sent to the GPS center via mobile networks.
3) The control center enters calculated location information over global maps.

4) The control center sends commands to each unit to trigger alarms, stop engines, change direction or some personal messages, etc. Autonomous localization of GPS receivers is essential, since location makes the sensory data geographically meaningful.

Many applications and services of wireless networks directly or indirectly rely on location information. For IoT applications, localization is important since abundant data sensed through IoT are meaningless without locations for location-based services. Though GPS is a straightforward solution, it has two limitations for practical uses. First, the GPS signal is highly dynamic and unstable in indoor or dynamic environments, resulting in poor location accuracy. Second, it is costly to equip each sensor node with a GPS receiver for such a large-scale IoT system. A combined scheme that uses GPS and network localization simultaneously is recommended.

Each satellite is visible only by GPS receivers on certain parts of the Earth's surface. At different times, each receiver can see only a subset (about 6) of the satellites. Four satellites are sufficient to locate the receive position accurately. At different times, different satellite subsets will become visible to the receiver. The control segment is composed of a master control station and a host of dedicated and shared ground antennas and monitor stations. The user segment is composed of hundreds of thousands of US and allied military users of the secure GPS precise positioning services.

The ground GPS receiver calculates the 3-D location from four or more satellites with the help of a few ground reference stations and a master station, as shown in Figure 3.14. Essentially, the GPS receiver compares the time a signal was transmitted by a satellite with the time it was received. The time difference tells the GPS receiver how far away the

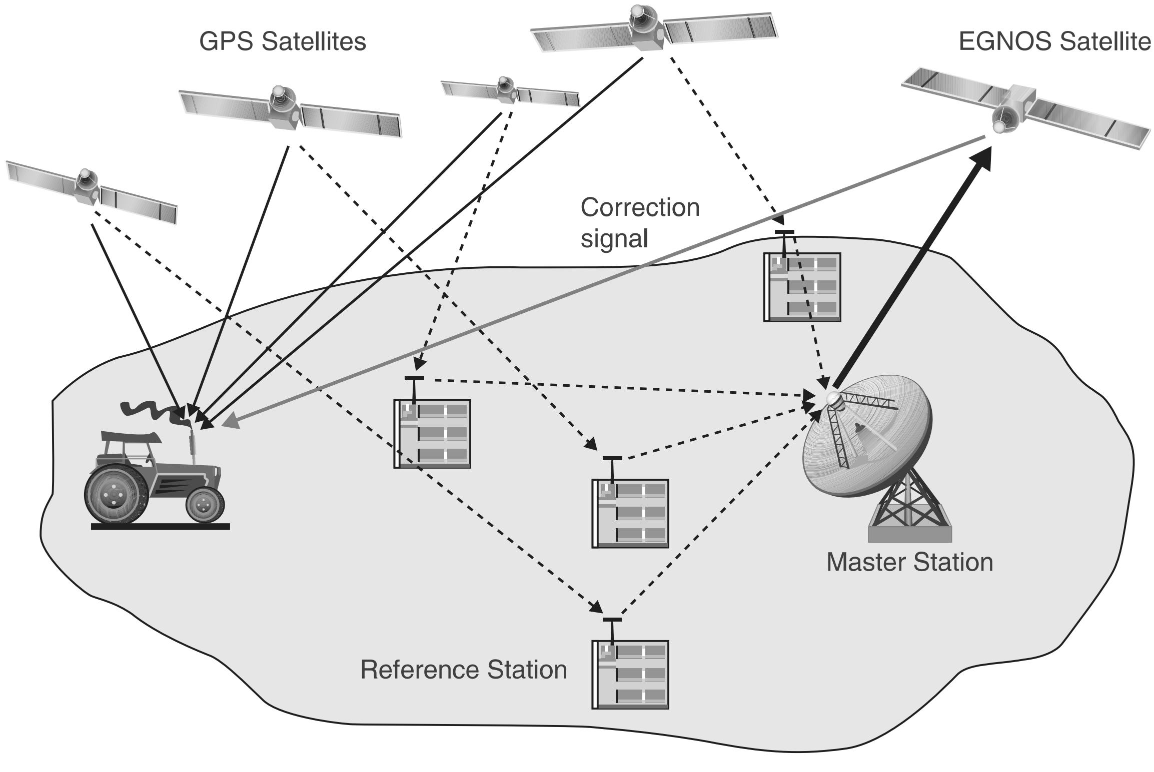



**Figure 3.14** The ground GPS receiver calculates the 3-D location from four or more satellites with the help from a few ground reference stations and a master station.

satellite is. With distance measured from several satellites, the receiver can determine the user's position and display it on the unit's electronic map. Tens of millions of civil, commercial and scientific users are only allowed to use a degraded functionality of the so-called standard positioning services that cannot be used in hostile attack purposes.

#### 3.4.4.1 Passive versus Active GPSs

The GPS tracking device makes it possible to track people, vehicles and other assets, anywhere on Earth. There are two types of GPS tracking systems, passive versus active. In passive tracking, the GPS is just a receiver, not a transmitter. Passive GPS tracking devices lack a transmission capability to send the GPS data from the vehicle. Therefore, the passive GPSs are also known as data loggers used primarily as recording devices. Active GPS tracking units incorporate a method to transmit the user information from a vehicle. Although satellite uplink of data is available, cellular data communication is the most common and cost effective. Automatic incremental updating provides a continuous source of tracking throughout a recording period. This provides current as well as historical logging positions.

Passive GPS tracking devices store GPS location data in their internal memory, which can then be downloaded to a computer for viewing at a later time, while active GPS tracking systems send the data at regular intervals to be viewed in real time. When real-time data is not required, passive GPS tracking devices tend to be favored more by individual consumers for their compact convenience and affordability. Concerned parents can install a GPS tracking unit just about anywhere in their teens' vehicles to monitor their driving habits and know where they have been going, and even law enforcement officials now rely on passive GPS tracking to trail criminal suspects and enhance civilian safety via electronic surveillance of parolees. Passive GPS tracking units also serve as a theft prevention and retrieval aid in consumer as well as commercial vehicles.

#### 3.4.4.2 Operating Principles of GPS

Knowing the distance from receiver to a fixed-position satellite implies that the receiver is on the surface of a sphere centered at the satellite. With four satellites, the receiver location is detected at the intersection of four sphere surfaces. The intersection of two satellite spheres is generally a circle. This circle could be reduced to a single point, if the two spheres merely touch on their surfaces. Having found the two intersecting sphere surfaces, now we consider how the intersecting circle intersects with a third satellite sphere. A circle and a sphere surface intersect at zero, one or two points. With the receiver on the surface of the Earth, the receiver needs to choose the point which is closest to the receiver from the two intersecting points.

Obviously, the above triangulation method may result in some error in narrowing down to exactly one point with minimum inaccuracy. To locate the point accurately, the receiver has to use a fourth satellite to home in more precisely. The fourth satellite sphere will come very close to the final two intersecting points of the three satellite spheres. The final receiver location is thus decided by noting the closest point calculated from the two final points to the sphere surface of the fourth satellite. In the case of no error, the precise position is located. Otherwise some offset, say 10 meters from the exact location, can result from the error introduced. To further reduce errors, more satellite could be involved, but this would be rather costly.

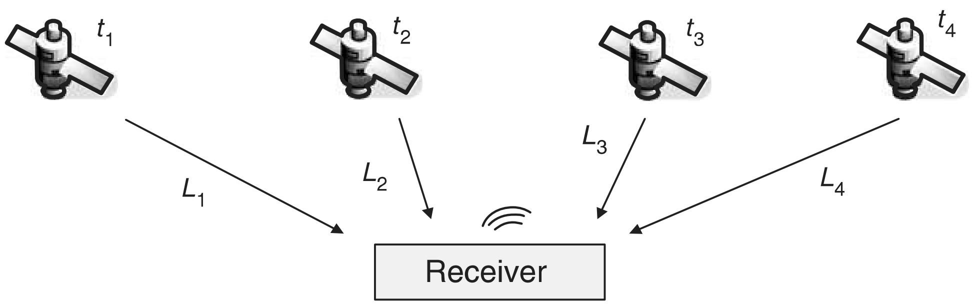
Each satellite continually transmits messages that include: i) the time the message was transmitted; ii) precise orbital information (the ephemeris); and iii) the general system health and rough orbits of all GPS satellites (the almanac). A GPS receiver calculates its position by precisely timing the signals sent by GPS satellites high above the Earth's surface. A GPS receiver must be locked on to the signal of at least three satellites to calculate a 2-D position (latitude and longitude) and track movement. With four or more satellites in view, the receiver can determine the user's 3-D position (latitude, longitude and altitude).

Once the user's position has been determined, the GPS unit can calculate other information, such as speed, bearing, track, trip distance, distance to destination, sunrise and sunset time, and more. The receiver utilizes the messages it receives to determine the transit time of each message and computes the distances to each satellite. These distances along with the satellites locations are used to compute the position of the receiver. This position is displayed perhaps with a moving map display or latitude, longitude and elevation information. Many GPS units show derived information such as direction and speed calculated from positional changes.

#### 3.4.4.3 Triangulation Location Calculation

This location calculation method is illustrated in Figure 3.15. The receiver uses messages received from four satellites to determine the satellite positions and time sent. The $x$, $y$ and $z$ components of position and the time sent are designated as $[\boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{z_i}, \boldsymbol{t_i}]$, where the index $i = 1, 2, 3$ or $4$ denotes the satellite. Knowing when the message was received $\boldsymbol{t_{ri}}$, the receiver computes the message's transit time as $\boldsymbol{t_{ri}} - \boldsymbol{t_i}$. Assuming the message traveled at the speed of light $c$, the distance traveled is calculated by $\boldsymbol{d_i} = (\boldsymbol{t_{ri}} - \boldsymbol{t_i}) \times \boldsymbol{c}$.

Having discussed how sphere surfaces intersect, we now formulate the equations for the case when errors are present. Let $\boldsymbol{b}$ denote the clock error or bias, the amount by which the receiver's clock is off. The receiver has four unknowns, the three components




$$L_1 = c\,(t - t_1) = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$$

$$L_2 = c\,(t - t_2) = \sqrt{(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2}$$

$$L_3 = c\,(t - t_3) = \sqrt{(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2}$$

$$L_4 = c\,(t - t_4) = \sqrt{(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2}$$

**Figure 3.15** Triangulation method to calculate delayed location signals from four satellites.

of GPS receiver position and the clock bias [$\boldsymbol{x, y, z, b}$]. The equation of the sphere surfaces is calculated by the following expression for i = 1, 2, 3, and 4:

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = ([tr_i + b - t_i]c)^2 \tag{3.1}$$

A multidimensional root finding method such as the Newton–Raphson method can be used. This approach is to linearize around an approximate solution, say [$\boldsymbol{x}^{(k)}$, $\mathbf{y}^{(k)}$, $\mathbf{z}^{(k)}$, $\boldsymbol{b}^{(k)}$] at iteration $k$, then solve four linear equations derived from the quadratic equations above to obtain the corresponding values at time instance of $k + 1$. The Newton–Raphson method converges faster than other position methods. When more than four satellites are available, the calculation can choose from four of the results shown in Figure 3.15. The error of position calculation is very sensitive to the clock error. Therefore, in the satellite-based navigation system, clock synchronization is critically important to minimize the location error.

Three satellites are sufficient to locate the receiver position, since the space has three dimensions and a position near the Earth's surface is always assumed. However, even a very small clock error multiplied by the light speed of satellite signals may result in a large positional error. Therefore, most receivers use four or more satellites to solve the receiver's location and time. The computed time is often hidden by most GPS applications, which use only the location information. A few specialized GPS applications do use the time for time transfer, traffic signal timing and synchronization of cell phone base stations.

Although four satellites are required for normal operation, if the 1-D variable is already known, a receiver can determine its position using only three satellites. For example, a ship or plane may have known elevation. Some GPS receivers may use additional clues or assumptions (i.e. reusing the last known altitude, dead reckoning, inertial navigation or including information from the vehicle computer) to give a less accurate (degraded) position when fewer than four satellites are available.

#### 3.4.4.4 Worldwide Deployment Status

Table 3.6 summarizes four global positioning systems used today. In addition to the GPS deployed by the US, which is now open for global civilian applications by many countries, the Russians have deployed a GLONASS (Global Navigation Satellite System) for Russian military use exclusively. In the European Union, there is the Galileo positioning

**Table 3.6** Four Global Positioning Systems in US, EU, Russia and China.

| Features | GPS | GLONASS | Beidou | Galileo |
|---|---|---|---|---|
| Political entity | United States | Russia | China | European Union |
| Coding | CDMA | FDMA/CDMA | CDMA | CDMA |
| Orbital height | 20,180 km (12,540 mi) | 19,130 km (11,890 mi) | 21,150 km (13,140 mi) | 23,220 km (14,430 mi) |
| Period | 11.97 hours (11 h 58m) | 11.26 hours (11 h 16m) | 12.63 hours (12 h 38m) | 14.08 hours (14 h 5m) |
| Number of satellites | At least 24 | 31 (24 operational) | 5 GEO, 30 MEO satellites | 22 operational satellites supported |

system. By 2015, China had launched 20 satellites toward a complete system, with 31 satellites planned for the 2020s.

## 3.5 Cognitive Computing Technologies and Prototype Systems

This section is devoted to the study of cognitive computing technologies and prototype systems. We check through the development processes of three experimental cognitive systems at the IBM Almaden Research Center, Google Brain Team Projects and the Chinese Academy of Sciences. Finally, we show how IoT contexts benefit cognitive services, and present the IoT context in cognition and recent cognitive devices such as AR glasses and VR headsets.

### 3.5.1 Cognitive Science and Neuroinformatics

Cognitive science is interdisciplinary in nature. It covers the areas of psychology, artificial intelligence, neuroscience and linguistics, etc. It spans many levels of analysis from low-level machine learning and decision mechanisms to high-level neural circuitry to build brain-modeled computers. In 2008, a fundamental concept of cognitive science was given by Paul Thagrad: "Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures." In general, three approaches are adopted in cognitive computing applications:

1) Apply software library on clouds or supercomputers for machine learning and neuroinformatics studies.
2) Use representation and algorithms to relate the inputs and outputs of artificial neural computers.
3) Use neural chips to implement brain-like computers for machine learning and intelligence.

Neuroinformatics attempts to combine informatics research and brain modeling to benefit both fields of science. The traditional computer-based informatics facilitate brain data processing and handling. Through hardware and software technologies, we can arrange databases, modeling and communication in brain research. Or conversely, enhanced discoveries in neuroscience may invoke the development of new models of brain-like computers.

In Table 1.12, we have identified related fields to cognitive computing and neuroinformatics technologies. A major concern of AI research is to find out how human learning, memory, language, perception, action and knowledge discovery are handled. We hope to apply machine intelligence to help human decision making or make our daily lives activities safer, effective, efficient and comfortable, etc.

**Example 3.9 Cognitive Science and Neuroinformatics at Academia and IBM Labs**
McCulloch and Pitts developed the very first artificial neural network (ANN) model of computation inspired by the structure of biological neural networks. The first instance of cognitive science experiments were performed at the MIT Social Psychology Department using computer memory as models for human cognition. Much of the research

efforts on cognitive science were supported by the National Institute of Health in the USA.

IBM founded the Blue Brain Project in May 2005. The project was carried out on an 8000-processor Blue Gene/L supercomputer built by IBM. At that time, this was one of the fastest supercomputers in the world. The mission of the Blue Brain Project is to understand mammalian brain function and dysfunction through detailed simulations. The IBM projects cover the following aspects:

- **Databases**: 3-D reconstructed model neurons, synapses, synaptic pathways, microcircuit statistics, computer model neurons and virtual neurons.
- **Visualization**: microcircuit builder and simulation results visualizator, 2-D, 3-D and immersive visualization systems are being developed.
- **Simulation environment**: a simulation environment for large-scale simulations of morphologically complex neurons on IBMl's Blue Gene supercomputer.
- **Simulations and experiments**: iterations between large-scale simulations of neocortical microcircuits and experiments.

■

Traced back further, the IBM Watson Center had a Deep Blue project that led to the chess playing competition that defeated World Chess Champion Garry Kasparov in 1997. That was the world's first cognitive system built to work with traditional computer hardware. Recently, IBM announced that they will push cognitive services as a major thrust of effort in the next decade. The goal is to develop a cognitive industry to serve human societies and to promote global economy. We will cover the IBM synaptic chip development in building a brain-like computer in Section 3.5.2. Then, we will examine Google's Brain Team efforts in Section 3.5.3. Clouds and IoT technologies play a vital role in these industrial transformations.

### 3.5.2 Brain-Inspired Computing Chips and Systems

In this section, we study some new processor chips, non-von Neumann architecture and ecosystem development at IBM, Nividia, Intel and the China's Institute of Computing Technology for cognitive computing. Even though these projects are still at the research stage, they represent the emerging technologies which combine computing with cognition to augment human capacity and understanding of all kinds of environments surrounding us.

#### 3.5.2.1 IBM SyNapse Program

IBM has a SyNapse research program, devoted to the development of new hardware and software for cognitive computing. This project has been supported by DARPA (Defense Advanced Research Projects Agency) in the US. In 2014, IBM unveiled a neurosynaptic computer chip design in *Science Magazine*, known as the TruthNorth processor. This processor can mimic the human brain's computing abilities and power efficiency. The chip design can enable wide-ranging applications, such as assisting vision-impaired people to navigate safely through their environment.

This chip could cram supercomputer-like powers into a microprocessor the size of a postage stamp. Rather than solving problems through brute-force mathematical calculations, the chip is designed to understand its environment, handle ambiguity and take action in real time and in context. It was estimated that an average human brain has 100 billion neurons and 100 to 150 trillion synapses. Modeled on the human brain, the TrueNorth chip incorporates 5.4 billion transistors, the most IBM has ever put on a chip [23]. The chip features 1 million programmable neurons and 256 million programmable synapses.

It was speculated that this synaptic chip could be applied to power small-and-rescue robots, automatically distinguish between voices in a meeting and create accurate transcripts for each speaker. It may even have the potential to issue tsunami alerts, monitor oil spills or enforce shipping lane rules. What is amazing is that the chip consumes only 70 milliwatts of power to perform the above functions, about the same level consumed by a hearing aid. The chip is still in the prototype stage.

It was announced at a conference that IBM may spend $3 billion to push for the future of such computer chips and explore its cognitive service potentials. It does not require the heavy computational loads for complex operations as in biological cognitive systems. For example, if a robot running with today's microprocessors is walking toward a pillar, it would depend on image processing and huge computing resources and power to avoid a collision. By comparison, a robot using a synaptic chip would steer clear of the danger by sensing the pillar, much as a person would do with little power consumption.

Experts believe an innovation like SyNapse's TrueNorth could help overcome the performance limits of the von Neumann architecture, the mathematics-based system at the core of almost every computer built since 1948. "It is a remarkable achievement in terms of scalability and low power consumption," said Horst Simon, the director of the US Department of Energy's Berkeley Lab and an expert on computer science. IBM expects the chip to help transform science, technology, business, government and society by enabling vision, audition and multi-sensory applications. This could be the first step in designing future computers based on the human brain model. Similarly, Nividia's graphics has leaned also in this direction to power supercomputer brains.

#### 3.5.2.2 China's Cambricon Project

This project started as a joint research program between Dr Tianshi Chen at the Institute of Computing Technology, Chinese Academy of Sciences and Prof Oliver Tenam at French INRIA. The joint research team has developed a series of hardware accelerators, known as *Cambricon,* for neural network and deep-learning applications. The chip was built as a synaptic processor to power artificial neural computing in machine learning operations. The Cambricon stays away from the classic von Neumann architecture in order to match the special kind of operations performed in artificial neural network computations.

Machine-learning tasks are becoming pervasive in a broad range of systems, from embedded systems to datacenters. For example, deep-learning algorithms, using convolutional and deep neural networks (CNNs and DNNs to be treated in Chapter 6), take a long learning cycle to be usefully trained on a conventional computer. The Cambricon accelerator was designed to focus on large-scale CNN and DNN solutions. The joint ICT-INRIA team proved that it is possible to build accelerators with high throughput, capable of performing 452 GOP/s (key neural network operations in synaptic weight multiplications).

The chip was fabricated on a small footprint of 3.02 $mm^2$ and 485 mW silicon technology. The team has also worked out a new instruction set architecture (ISA) for efficient use of such a brain-like processor. This new ISA is specially tailored for neural network or cognitive computing. Compared with a 128-bit 2 GHz SIMD GPU accelerator, the accelerator chip achieved a speed 117 faster with 21 times reduction in power consumption. With an extended 64-chip machine-learning architecture, the team has shown a speed of 450 times over an array of GPU chips with a power reduction of 150 times.

Through multi-national efforts, human-oriented cognitive computing comes closer to a reality. It is interesting to watch the development of special purpose neuron-based synaptic processors in multi-core or many-core multiprocessor chips and the use of a massive number of such chips to build future cognitive supercomputers. We will study machine learning and deep learning algorithms in Chapters 4 to 6. Although these ML/DL algorithms are applied to today's clouds or server clusters, as covered in Chapters 7 to 9, they could be targeted to guide the design of future cognitive computing systems.

### 3.5.3 Google's Brain Team Projects

In research and industry, speech recognition is always in great demand. It would be nice to have an intelligent recording machine that could listen to human speech and produce documented textual reports. Similarly, the United Nations need to have automatic language translation systems, not only translation between text documents, but also between speeches and documented reports in different languages.

**Example 3.10 Google's Brain Project on Developing New ML/DL Products**
Google Brain project started in 2011 as a joint effort between Jeff Dean, Greg Corrado and Andrew Ng. Ng is interested in using deep learning techniques to crack the AI problems. They have built a large-scale deep learning software system, called DistBelief on top of Google's cloud computing infrastructure. In 2012, the *New York Times* reported that a cluster of 16,000 computers were used to mimic some aspects of human brain activity and had successfully trained itself to recognize a cat based on 10 million digital images taken from YouTube videos.

In 2013, Geoffrey Hinton joined Google, a leading researcher in the deep learning field. Subsequently, Google merged DeepMind Technologies and released TensorFlow. Notable Google products developed out of the Brain Team include the Android speech recognition system, Google photo search and video recommendations in YouTube. Figure 3.16 lists some of the service products developed at Google. The team has also worked on developing mobile and embedded machine intelligence applications, and started with Android and then iOS services. In addition, the brain team has worked with Google X and the Quantum Artificial Intelligence Lab at NASA in joint space programs.

In May 2016, Google announced a custom ASIC (application Specific integrated circuit) chip they have built specifically for machine learning tailored for TensorFlow programming. They installed the TPUs inside their data centers for more than a year, and achieved ten times in speed gains in machine learning operations. The TPU is a small circuit board inserted inside a hard drive attached to Google servers. The TPU is a pretrained chip to support TensorFlow computing with a high volume of low-precision (e.g. 8 bit) arithmetic. The TPU is indeed programmable as TensorFlow program changes.

Multiple TPUs were used in the AlphaGo match to be discussed in Section 9.4. The TensorFlow offers a software platform for the deep learning applications to be covered in

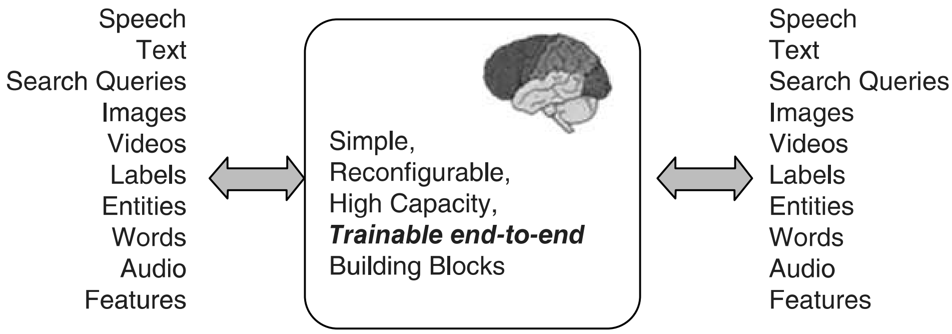



**Figure 3.16** The promise of deep learning at Google's Brain Project (Reprinted with permission from a public slide presentation by Jeff Dean, 2016).

Section 7.5. The TPU accelerates TensorFlow computations. Android support of TensorFlow is available for mobile execution. iOS support will come soon. Intel has also optimized their high-end server processor for neural computing. We will assess Google DeepMind projects in Chapter 9. ■

It is fair to say that deep neural networks play a crucial role in understanding speech, images, language and vision applications. Both pre-trained models or APIs must have low overheads and be easy to use in ML system development. In what follows, we check some of the deep learning systems developed by the Google Brain Team for various big data applications, as shown in Figure 3.17. Among these, we see DL applications in Android apps, drug discovery, gmail, image understanding, maps, natural language understanding, photos, robotics research, speech and YouTube among many others. Among the 50 internal product development teams at Google, the interest of using deep

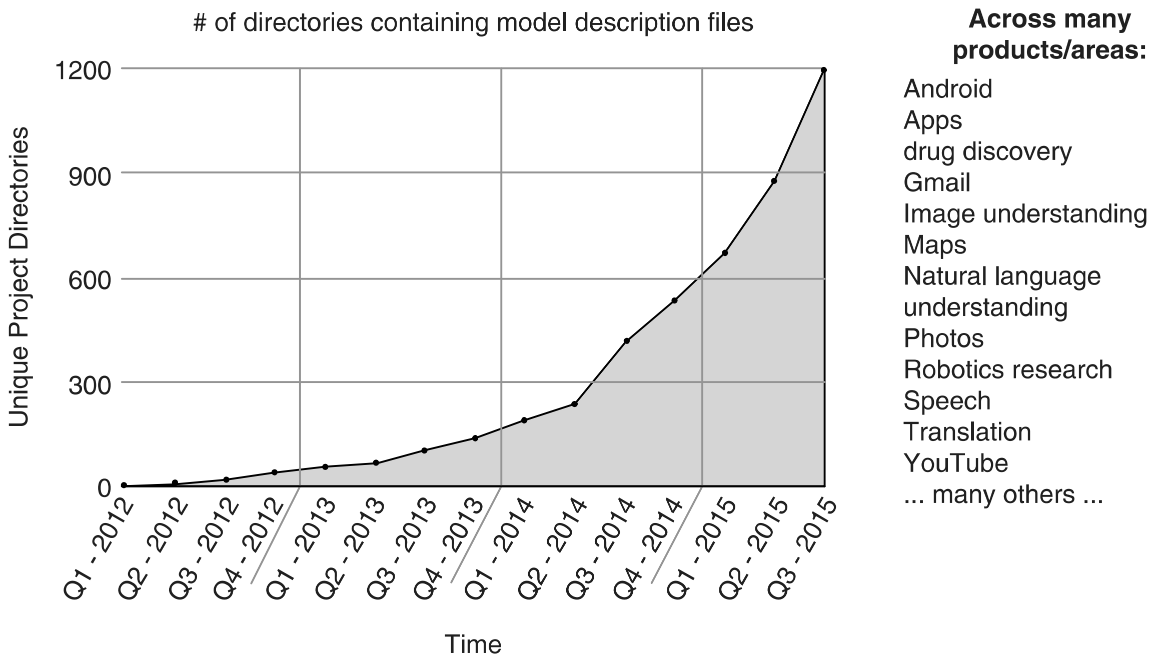



**Figure 3.17** Growing use of deep learning at Google teams (Reprinted with permission from public presentation by Jeff Dean, 2016).

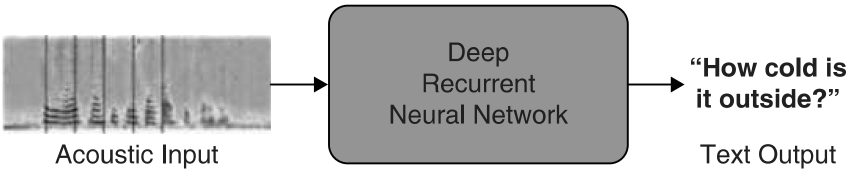



**Figure 3.18** The concept of a Google speech recognition system built with deep recurrent neural networks.

learning has been measured by the number of unique project directories containing model description files.

In general, three approaches are adopted in cognitive computing apps:

1) Apply software library on clouds or supercomputers for machine learning and neuroinformatics studies.
2) Use representation and algorithms to relate the inputs and outputs of artificial neural computers.
3) Design hardware neural chips to implement brain-like computers for machine learning and intelligence.

In Figure 3.18, we demonstrate the idea of building a Google speech recognition system with deep recurrent neural networks (DRNN).

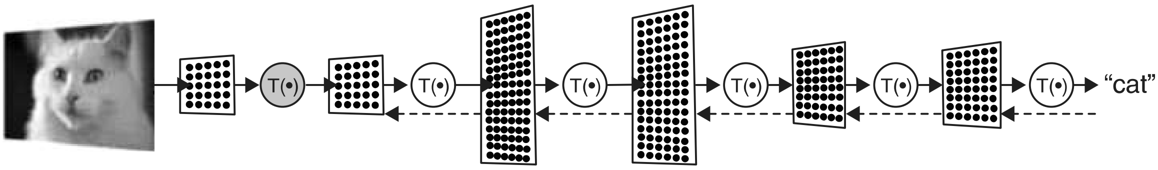
Here, acoustic speech signals are fed into the system as input. Through repeated learning from the DRNN system, text output as a question such as: "How cold is it outside?" is generated automatically. Apple's Siri system also has built such conversational capabilities. Deep convolutional neural networks have proven very useful for this purpose as well. In addition, object recognition and detection are of equal importance. This is part of the traditional field of pattern recognition and image processing domain. Deeper convolution and scalable object detection offer viable approaches to solve the problem on modern clouds.

Machine translation can be done by sequence-to-sequence learning processes with neural networks. Neural machine translation has been worked out at Google. Language modeling was also conducted with a 1 billion word benchmark. Another exciting area is automatic parsing grammar as a foreign language. The whole purpose of ANNs is to learn a complicated function from data. ANN has been a hot research field for at least 30 years.

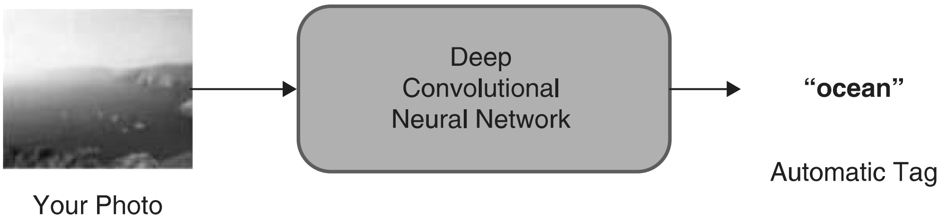
**Example 3.11 The Use of Google's ImageNet for Image Understanding**

The ANN models, especially the deep convolutional neural networks (DCNN), have been reincarnated. They offer a collection of simple, trainable mathematical functions, that are compatible with many variants of machine learning. Figure 3.19 shows the ideas behind using a DCNN to recognize a "cat" or an "ocean" from thousand of classes over millions of photo images. Image captioning has been in high demand in Google search requests.

Searching a personal photo without tags is equivalent to the task of identifying 1 image out of 1000 different classes. The work was carried out at Google using the ImageNet. Another project using GoogleNet also emphasizes a deeper convolution approach in




(a) **Recognizing a cat (not a dog)**




(b) **Distinguishing an ocean view from many other views**

**Figure 3.19** Using a deep convolution neural network to understand particular images out of millions of photos belong to different or similar classes.

the inception area. Neural networks have made rapid progress in image recognition. The ImageNet project challenges many classification tasks. The Inception team using GoogLeNet achieved only 6.66% error in 2014. ■

Another interesting area is to combine vision with translation or to combine vision with robotics intelligence. In other words, we want robots to go through a deep learning process by large-scale interactions. This is critical to the autonomous vehicle driving projects, which is actively pursued at Google, Baidu and other research centers. For example, we want the robots to learn hand-eye coordination through a deep learning system on board a car. Some progress has already been demonstrated in several ongoing projects in the US and China.

### 3.5.4 IoT Contexts for Cognitive Services

In practice, the deployment of cognitive services relies on different contexts, such as location information which was used by services offered over the Internet in order to provide location-aware customization to users. Once the mobile devices (phones and tablets) became a popular and integral part of everyday life, context information (gravity, rotation vector, orientation, geomagnetic field, proximity, light, pressure, humidity and temperature, etc.) collected from sensors built into the devices (e.g. accelerometer, gyroscope, GPS, and pulse oximetry, etc.) were used to provide context-aware functionality. For example, built-in sensors are used to determine user activities, environmental monitoring, health and well-being, location and so on.

Today's context information is collected through social networking services (e.g. Facebook, Myspace, Twitter and WeChat, etc.) using mobile devices. Some context-aware applications are developed for activity predictions, recommendations and personal assistance. For example, a mobile application may offer location information retrieved from mobile phones to recommend nearby restaurants what a potential customer might like. Another example is Internet-connected refrigeration. The user can check the foods in the refrigerator remotely and decide what to purchase on the way home.

When the user leaves their workplace, the application autonomously does the shopping and guides the user to a particular shopping market so he can collect the goods it has already ordered. In order to perform such tasks, the application must fuse location data, user preferences, activity prediction, user schedules, information retrieved through the refrigerator (i.e. shopping list) and many more. In the light of the above examples, it is evident that the complexity of collecting, processing and fusing information has increased over time. The amount of information collected to aid decision making has also increased significantly.

In the IoT era, there will be a large number of sensors attached to everyday objects. These objects will produce large volumes of sensory data that have to be collected, analyzed, fused and interpreted. Sensory data produced by a single sensor will not provide the necessary information that can be used to fully understand the situation. Therefore, data collected through multiple sensors need to be fused. In order to accomplish sensor data fusion, contexts need to be tagged together with the sensory data to be processed and understood later. Therefore, context annotation plays a significant role in context-aware computing research.

#### 3.5.4.1 IoT Contexts

Context-aware technology provides a methodology to evaluate the performance of an IoT solution. The evaluation is mainly based on three context-aware features in high-level: i) context-aware selection and presentation; ii) context-aware execution; and iii) context-aware-tagging. However, we have also enriched the evaluation framework by identifying sub-features under the above-mentioned three features. In Table 3.7 we give examples to evaluate IoT solutions in the smart city application domain.

The primary context data captured by IoT solutions are listed below: *W* denotes Web-based; *M* denotes Mobile-based; *D* denotes Desktop-based; and *O* denotes Object-based. We identify Touch (*T*), Gesture (*G*) and Voice (*V*) as three common mechanisms. *M* means that interactions are carried out through a PC or a smart phone. *RT* represents that an IoT solution processes data in real time, while *A* means that the IoT solution processes archival data. Other notations in Table 3.7 we have *S* for IoT sensing, *E* for energy, *UD* for user device, *R* for radio technology and *N* for notification. We introduce the name of the IoT project in the left-most column. The web page links are the most reliable references to a given IoT solution. Such links allow readers to probe further and explore the IoT technology more meaningfully.

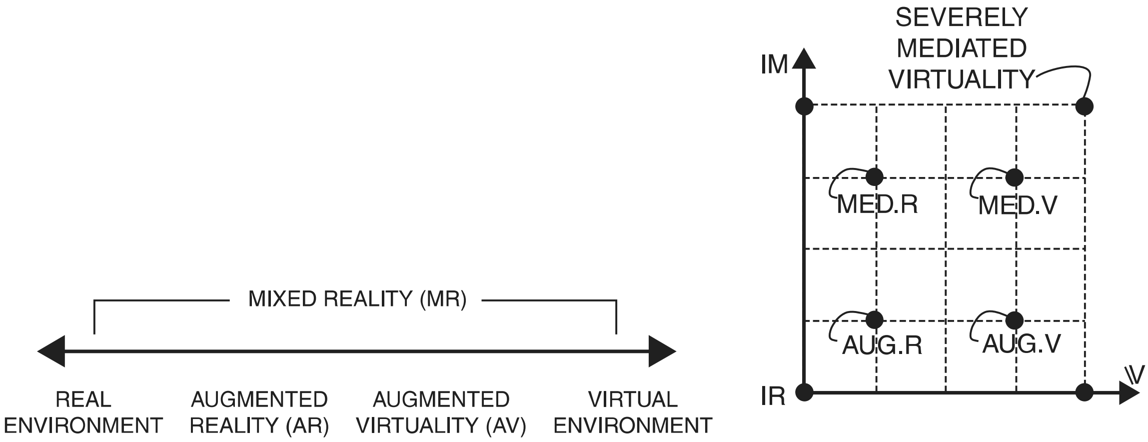
### 3.5.5 Augmented and Virtual Reality Applications

Virtual reality (VR) is a computer technology that replicates an environment, real or imagined, and simulates a user's physical presence and environment to allow user interaction. Virtual realities artificially create sensory experience, which can include sight, touch, hearing and smell. The immersive environment needs to be similar to

**Table 3.7** Representative IoT contexts in smart city applications.

| IoT Project, Builder and (Website) | Primary Context | Secondary Context | Presentation Channel | User Interaction | Real Time Archival | Notification Mechanism | Learning Ability | Notification Execution |
|---|---|---|---|---|---|---|---|---|
| Waste Management Evevo (enevo.com) | Waste fill-level | Efficient routes to pick up waste | *W* | *M* | *RT, A* | *N,R* | *ML,UD* | *E* |
| Indoor Localization, Estimote (estimote .com) | Bluetooth signal strength, Beacon ID | Location, Distance | *M* | *M* | *RT* | *N,R* | *UD* | *T,S,E* |
| Parking Slot Management, ParkSight (streetline.com) | Sound level, Road surface temperature | Route for free parking slot | *M,W* | *M* | *RT, A* | *N,R* | *ML,UD* | *T,S,E* |
| Street Lighting, Tvilight (tvilight.com) | Light, presence, weather, events | Energy usage, patterns, lamp, etc | *W* | *M* | *RT, A* | *N,A* | *ML,UD* | *T,S,E* |
| Movement Analysis, Scene Tap (scenetap.com) | GPS, Video | Crowd profiling by location | *M,W,D* | *M* | *RT* | *N,A* | *ML* | *T,S* |
| Foot Traffic Monitoring (scanalyticsinc.com) | Floor level | Heat maps tracking movements | *W* | *T,M* | *RT, A* | *N* | *ML,UD* | *S,E* |
| Crowed Analysis, Livehoods (livehoods.org) | Foursquare check-ins cloud service | Social dynamics, large cities | *W* | *M* | *RT, A* | - | *ML* | *E* |

**Abbreviations:** *M:* mobile, *W*: web services, *D*; desktop-base, *O*: object-oriented, *T;* touch technology, *S: IoT sensing, G*: gesture, *V*: voice, *RT*: real-time, *N*: notification, *A*: archival, *ML*: machine learning, *DU:* user device



(a) **Mixed reality spectrum** (b) **Four environments**

**Figure 3.20** Spectrum from real environment to AR, AV and VR.

the real world in order to create a lifelike experience – for example, in simulations for pilot or combat training, it can differ significantly from reality, such as in VR games (Figure 3.20).

Augmented reality (AR) is a live view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data. It is related to a more general concept called mediated reality. As a result, the technology functions by enhancing our current perception of reality. By contrast, virtual reality replaces the real world with a simulated one. Augmentation is conventionally in real time and in a semantic context with environmental elements. A mixed reality sits anywhere between the extrema of the virtuality continuum, extending from the complete reality through to a complete virtual environment with augmented reality and augmented virtuality mixed together.

In Figure 3.20, four example points are shown: augmented reality, augmented virtuality, mediated reality and mediated virtuality on the virtuality and mediality axes. This includes, for example, diminished reality (e.g. computerized welding helmets that filter out and diminish certain parts of a scene), accelerometer, gyrometer, proximity sensor, and light sensors are built-in VR headsets, including the HTC Vive, Paystation VR and Samsung Gear VR, etc.

#### 3.5.5.1 Video Games

The use of graphics, sound and input technology in video games can be incorporated into Virtual Reality. Several VR head mounted displays (HMD) have been released for gaming, including the Virtual Boy developed by Nintendo and iGlasses developed by Virtual I-O. Several companies are working on a new generation of VR headsets: Oculus Rift is a head-mounted display for gaming purposes, which was acquired by Facebook in 2014. One of its rivals was named by Sony as PlayStation VR (codenamed Morpheus). Valve Corporation announced their partnership with HTC Vive to make a VR headset capable of tracking the exact position of its user. Other AR/VR products can be found in Table 3.8.

**Table 3.8** Recent AR/VR products developed by high tech companies.

| Company | Product | Introduction |
|---|---|---|
| Microsoft | HoloLens | A pair of mixed reality head-mounted smart glasses by Microsoft. HoloLens gained popularity for being the first computer running the Windows Holographic platform. |
| Google | Google Cardboard | This is a VR platform by Google for use with a head mount for a smart phone. Named for its fold-out cardboard viewer, it is a low-cost system to encourage VR applications. |
| Facebook | Oculus Rift | Oculus Rift is a virtual reality headset developed and manufactured by Oculus VR, released March 28, 2016. |
| Samsung | Gear VR | The Samsung Gear VR is a mobile virtual reality headset developed by Samsung Electronics, in collaboration with Oculus, and manufactured by Samsung. |
| Sony | PlayStation VR | Known by the codename Project Morpheus during development, is a VR gaming head-mounted display developed by Sony Interactive Entertainment and manufactured by Sony. |
| HTC | HTC VIVE | This is a virtual reality headset developed by HTC and Valve Corporation in 2016. This is designed to utilize "room scale" technology to turn a room into 3-D space via sensors |
| Huawei | Huawei VR | Huawei honor VR was released on May 10, 2016 to match the honor V8 smart phone. |
| Alibaba | Buy + Plan | Buy+ program uses VR technology to generate interactive 3-D shopping environment with computer graphics systems and auxiliary sensors. |

#### 3.5.5.2 Education and Training

Strides are being made in the realm of education, although much still needs to be done. The possibilities of VR and education are endless and bring many advantages to pupils of all ages. A few are creating content that may be used for educational purposes, with most advances being made in the entertainment industry, but many understand and realize the future and the importance of education and VR. US Navy personnel use a VR parachute training simulator. The use of VR in a training perspective is to allow professionals to conduct training in a virtual environment where they can improve upon their skills.

## 3.6 Conclusions

The IoT has emerged rapidly to affect our daily life activities. The ultimate goal is to build a worldwide physical web that links everything together. IoT leads to smart classrooms, hospitals, marketplaces, department stores, streets, highways, cities and the Earth. In other words, we want to apply machine intelligence everywhere to aid or promote human safety, comfort, convenience and productivity. Cognitive science and services are also appearing. Both augmented reality (AR) and virtual reality (VR) may appear as commercial devices to expand human experiences in gaming, relaxation and creativity.

This chapter covers the progress of key technologies that make it possible to use smart devices, sensors, tags, phones, tablets, GPS, etc. everywhere and at any time. That is the dream of a paradise on Earth. We have shown how to apply IoT sensing capability in cloud computing with big data. Neural-based CPU chips are appearing at IBM research labs and at the Institute of Computing Technology in the China Academy of Sciences. The use of neuron-modeled computers for future machine learning and deep learning is no longer a dream. It is quite plausible to have innovative applications that involve IoT sensing and machine cognition in the near future.

## Homework Problems

**3.1** Answer the following two questions on IoT development in recent years:

**a)** The early IoT technologies include 4 T's, namely Telemetry, Telemetering, Telenet and Telematics. Perform a literature survey of recent progress in IoT and report your findings.

**b)** How are the following devices or techniques: sensors, smart phones, RFID labels/readers, bar code, 2-D QR code, or smart watch are used for data collection and sensing in IoT applications?

**3.2** Answer the following two updated assessments of GPS technologies and four systems built in the US, Russia, the EU and China. Wikepedia may be a good source of quick information:

**a)** What are the differences in civilian and military applications of various GPS systems?

**b)** Report several interesting civilian applications of GPS services.

**c)** What are the potential military applications of GPS capabilities?

**3.3** Design a healthcare system which consists of body sensors and wearable devices to collect human physiological signals. This system should possess the following functions: real-time monitoring, disease prediction and early detection of chronic diseases. Also, you may need a monitoring and management system that can optimize the distribution of medical resources and facilitate the data sharing for such resources.

**3.4** In recent years, video analytics became a hot topic, especially for security checks through video tracking, which is useful to protect personal and property safety. Traditional security technology emphasizes real-time response and the effectiveness of verification. So video presentation with high-resolution, no loss and low delay has been the main development direction of security industry over the past few years. Nowadays, we can see cameras for city surveillance everywhere.

With increasing use of high-definition cameras, how to effectively transmit the big amount of video data has become a key issue. In addition, tracking criminals to obtain their location information is time-consuming and labor-intensive. Describe how to use artificial intelligence and machine learning technology to analyze massive video samples, automatically track the target and find the moving path.

**3.5** Parkinson's disease (PD) is a chronic disease caused by movement disorder of the central nervous system. Typically, gait is an important indicator to identify and evaluate PD. In order to evaluate gait changes in the elderly with PD continuously without human intervention, pressure of foot step can be measured when PD patients walk, and the mode of center of pressure (CoP) can be obtained. Try to figure out the differences of CoP between normal people and PD patients. Which statement is correct?

**a)** The pressure sensors are deployed under the PD patient's foot.
**b)** The pressure sensors are installed on the ground.
**c)** In order to obtain CoP, the pressure of front, middle or back parts of foot should be collected.
**d)** Measure the pressure data when PD patient is standing or walking.

**3.6** The incidence of leukemia among young people has increased, which needs stem cell transplantation as the compulsory treatment. After the transplant, patients have to stay at home for 12 to 24 months. In order to skip arduous and unpleasant feelings of the patients during rehabilitation, a video system is designed to assist communication between patients and medical teams via smart phone, tablet or personal computer. Meanwhile, the personal data of the patients can be easily accessed through a web-based system. Especially, if a gaming element is added in such remote data retrieving system, the patients' mood can be improved during the daily report. With more practical and frequent healthcare data, the medical team can monitor the patient's health status more accurately and timely and provide more effective treatment.

**1)** About the video system, which statements are correct?
  **a.** We need a highly flexible framework of data, in order to meet the requirements about custom health parameters.
  **b.** External data sources transfer through the e-health data service bus to the database.
  **c.** Data can only be hard to define, not soft definition.
  **d.** Game is given priority with smart phones and tablets use, but can also be conducted on a web browser.

**2)** Video game system workflow includes three steps: data definition, create a configuration file, and plan a game task. When distributing small games to patients, the task can specify a group of patients with physical therapy practice according to their health status evaluated by a medical team. Write your thoughts on the above three steps.

**3.7** Design an intelligent vehicle management system based on IoT, especially RFID technology. The system should enable automatic payment, such that a vehicle can go through an intersection without stopping. When the vehicle exits, the parking fee is deducted automatically.

**3.8** This problem is related to using IoT to promote the green agriculture.

**1)** Based on a research study from the literature, elaborate on each of the following requirements:

   **a.** Real-time collection of a farm's environmental parameters like temperature, humidity, illumination, soil temperature, soil moisture and oxygen levels in a greenhouses or water beds, etc.
   **b.** Real-time intelligent decision on crop growth, and automatic opening or closing the environmental control equipment. The deployment of the system provides a scientific basis and effective means for agricultural monitoring, automatic control and intelligent management.
   **c.** The system will store and analysis the real-time monitoring data on the server to automatically open or close the specified device, such as remote control watering, switching shutter, adding oxygen or $CO_2$, etc.
2) About the solutions of building an intelligent agricultural system, discuss how to implement each solution with up-to-date wireless, sensor and GPS technologies:
   **a.** Wireless sensor network technology is applied in an intelligent agricultural system to achieve data collection and control.
   **b.** A smart agricultural greenhouse equipped with wireless sensors to monitor the environmental parameters like air/soil temperature, humidity, moisture, light and $CO_2$ concentration.

**3.9** Answer the following two updated assessment of GPS technologies and four systems built in the USA, Russia, the EU and China. Wikepedia may be a good source of quick information:
   **a)** What are the differences between civilian and military applications of various GPS systems?
   **b)** Report several interesting civilian applications of GPS services.
   **c)** What are the potential military applications of GPS capabilities?

**3.10** We have studied three IoT applications in Examples 3.1 to 3.8. Make some investigation and search for another meaningful IoT application. Submit an investigated report in similar depth as in the examples. Dig out as much technical information as you can from the literature or other sources. Report on the interesting IoT features, hardware and software advances, interaction models applied, and available performance results both quantitatively and qualitatively. Do not produce a hand-waving report. Everything you report must be substantiated with evidence and analysis.

## References

1 P. Abouzar, K. Shafiee, D. Michelson and V.C.M. Leung, Action-based scheduling technique for 802.15.4/ZigBee wireless body area networks. *IEEE PIMRC,* Toronto, ON, September 2011.
2 Bluetooth Low Energy Specification, Bluetooth Special Interest Group, http://www.bluetooth.com
3 H. Cao, H. Li, L. Stocco and V.C.M. Leung, Wireless three-pad ECG system: Challenges, design and evaluations. *Journal of Communications and Networks*, 13(2), 113–124, 2011.

4 K. Chen and D. Ran, *C-RAN: The Road towards Green RAN.* White Paper, China Mobile Research Institute, Beijing, October 2011.
5 M. Chen, S. Gonzalez, A. Vasilakos, H. Cao and V. Leung, Body area networks: A survey. *ACM/Springer Mobile Networks and Applications* (MONET), 16(2), 171–193, 2010.
6 T. Chen and Z Du, et al., DianNao: A small-footprint high-throughput accelerator for ubiquitous *machine-learning. Proceedings of 19th International Conference on Architectural Support for Programming Languages and Operating Systems* (ASPLOS'14), 2014.
7 J. Dean, Large-scale Deep Learning for Intelligent Computer Systems, Slide Presentation, 2016
8 D, Gardner and G.M. Shepherd, A gateway to the future of neuroinformatics. *Neuroinformatics*, 2(3), 271–274, 2004.
9 E. Farella, A. Pieracci, L. Benini, L. Rochi and A. Acquaviva, Interfacing human and computer with wireless body area sensor networks: The WiMoCA solution. *Multimedia Tools and Applications*, 38(3), 337–363, 2008.
10 J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, Internet of things: A vision, architectural elements and future direction. *Future Generation of Computer Systems*, 29, 1645–1660, 2013.
11 S. González-Valenzuela, M. Chen and V.C.M. Leung, Mobility support for health monitoring at home using wearable sensors. *IEEE Trans. Information Technology in BioMedicine*, 15(4), 539–549, 2011.
12 I. Jantunen et al., Smart sensor architecture for mobile-terminal-centric ambient intelligence. *Sensors and Actuators A: Physical*, 142(1), 352–360, 2004.
13 E.A. Lee, Cyber physical systems: design challenges. *IEEE International Symposium on Object Oriented Real Time Distributed Computing*, 363–369, May 2008.
14 G. Lo, A. Suresh, S. Gonzalez-Valenzuela, L. Stocco and V.C.M. Leung, A wireless sensor system for motion analysis of Parkinson's disease patients. *Proceedings of the IEEE PerCom*, Seattle, WA, March 2011.
15 H. Li and J. Tan, Heartbeat driven medium access control for body sensor networks. *Proceedings of the ACM SIGMOBILE*, San Juan, Puerto Rico, 2007.
16 S. Liu, T. Chen, et al., Cambricon: An instruction set architecture for neural networks. *Proceedings of the 43rd ACM/IEEE International Symposium on Computer Architecture* (ISCA'16), 2016.
17 G.A. Miller, The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences* 7: 141–144, 2004.
18 A. Milenkovic, C. Otto and E. Jovanov, Wireless sensor networks for personal health monitoring: Issues and an implementation. *Computer Communications*, 29(13–14), 2521–2533, 2006.
19 P. Merolla, et al., A million spiking neuron integrated circuit with a scalable network and interface. *Science*, 345(6197), 2014.
20 S. Pentland, Healthwear: Medical technology becomes wearable. *IEEE Computer*, 37(5), 42–49, 2004.
21 C. Perera, A. Zaslavsky, P. Christen and D. Georgakopoulos, Context aware computing for the Internet of Things: A survey, *IEEE Community Surveys Tutorials*, 16(1), 414–454, 2013.
22 TinyOS Website, available at http://tinyos.net/, 2011.
23 D. Terdiman, IBM's TrueNorth processor mimics the human brain, http://www.cnet.com/news/ibms-truenorth-processor-mimics-the-human-brain/, 2014.

**24** N. Torabi and V.C.M. Leung, Robust license-free body area network access for reliable public m-health services. *Proceedings of the IEEE HealthCom*, Columbia, MO, June 2011.

**25** A. Zaslavsky, C. Perera and D. Georgakopoulos, Sensing as a service and big data. *Proceedings of the International Conference Advanced Cloud Computing (ACC),* Bangalore, India, 21–29, July 2012.

**26** ZigBee Specification, ZigBee Alliance, 2005, http://www.zigbee.org