IEEE P802.15
Wireless Personal Area Networks

Project	IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs)	
Title	**Application of IEEE Std 802.15.4**	
Date Submitted	[May, 2014	
Source	[IEEE 802.15.4 Maintenance SC] [IEEE 802.15] []	Voice: [] Fax: [] Web: [http://www.ieee802.org/15]
Re:	[]	
Abstract	[This document contains examples of how to satisfy various application requirtements using IEEE Std 802.15.4. If it has been approved by the IEEE 802.15 WG, the introduction will contain the date of the approval. All other versions of the document are unofficial drafts.]	
Purpose	[The information in the document is an informative supplement to the standard and is intended to assist implementers in correctly implementing the standard..]	
Notice	This document has been prepared to assist the IEEE P802.15. It is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein.	
Release	The contributor acknowledges and accepts that this contribution becomes the property of IEEE and may be made publicly available by P802.15.	

1. Introduction

This document describes ways in which IEEE Std 802.15.4 can be used to satisfy application requirements.

This document has not been approved by IEEE 802.15.

2. Japan 950 MHz band considerations

2.1 General

This clause describes the way in which the IEEE 802.15.4 MAC may be configured and used to meet the requirements of the Japanese regulation for 950 MHz, as described in ARIB STD-T96 [B18] and English translation of ARIB STD-T96 [B19].

2.2 Listen before talk (LBT) considerations

The regulation requires that a device uses listen before talk (LBT) prior to transmission if the duty cycle of transmission exceeds 0.1%. There is also a requirement that a device does not continuously transmit. The maximum continuous transmission time and the duty cycle of transmission are dependent on the LBT duration. The detailed operation is described in ARIB STD-T96 [B18].

The regulation applies to a complete product. MAC PIB attributes are provided to permit a higher layer to control both the duty cycle and the listen before talk functionality.

The PIB attribute *macTxTotalDuration* is provided to allow a higher layer to control the transmission duty cycle of operation. The value represents the total number of symbols transmitted since last set to zero. The higher layer may read the value at any time, and it may set the value (typically to zero). This provides a mechanism for the higher layer to calculate the actual transmission time and hence the percentage transmission time over any arbitrary period.

The PIB attributes *macTxControlActiveDuration* and *macTxControlPauseDuration* permit a higher layer to control both the duration for which a device may transmit and the duration of the pause period, i.e., the time during which the MAC will pause to allow other devices access to the channel. These values are dependent on the transmission power and channel.

3. MAC behaviors for industrial and commercial application domains

3.1 Introduction

Industrial applications (and some commercial applications) have critical requirements such as low latency, robustness in the harsh industrial RF environment, and determinism that are not adequately addressed by IEEE Std 802.15.4-2011. Additionally, many of these application domains have unique requirements that conflict with the requirements from other application domains. To allow IEEE 802.15.4 devices to support a wide range of industrial and commercial applications, a wide range of MAC behaviors was found to be necessary.

These MAC behaviors facilitate industrial applications [such as addressed by IEC 62591, ISA100.11a, and wireless network for industrial automation—process automation (WIA-PA)], in addition to those enhancements defined by the Chinese wireless personal area network (CWPAN) standard that also were not included in IEEE Std 802.15.4-2011. The CWPAN standard has identified MAC behaviors to improve network reliability and increase network throughput to support higher duty-cycle data communication applications.

Specifically, there are two categories of MAC enhancements, as follows:

- Behaviors to support specific application domains such as process automation, factory automation
- General functional improvements not specifically tied to application domains

The convention used in this document is to group those MAC behavior modes according to the specific application domain requirements they address under the following headings:

- Timetimeslotslotted channel hopping (TSCH), for application domains such as process automation, described in 3.2.
- Low latency deterministic networks (LLDN), for application domains such as factory automation, described in 3.3.
- Deterministic and synchronous multi-channel extension (DSME), for general industrial and commercial application domains cited by CWPAN includes Channel diversity to increase network robustness, described in 3.4.
- Radio frequency identification blink (RFID), e.g., for application domains such as item and people identification, location, and tracking described in 3.5.
- Asynchronous multi-channel adaptation (AMCA), for large infrastructure application domains described in 3.6.

The additional MAC enhancements not specifically tied to a particular application domain mode are as follows:

- Low-energy (LE) protocol, to allow very low duty cycle devices to send ad hoc data using minimal amounts of energy, described in 3.7.
- Information elements (IE) to provide extensible MAC data transfers, described in 3.8.
- Enhanced Beacons frames and Enhanced Beacon Requests commands, to allow coordinator devices to send beacons with specifically requested data, described in 3.9.
- The MAC multipurpose frame, which provides the scalability and extensibility to allow this standard to address new application needs with minimal MAC changes, as described in 3.10.
- MAC performance metrics to provide upper layers with critical information on the quality of the communication links, described in 3.11.

- FastA to reduce the time required to associate, described in 3.12.

The subsequent subclauses of this document provide tutorial material for a better understanding of the targeted application domain modes and additional enhancements specified in IEEE Std 802.15.4.

3.2 Timestimeslotlotted Channel Hopping (TSCH)

3.2.1 Typical application domains addressed by the TSCH mode

Typical segments of the application domain of TSCH are process automation facilities for the following:

- Oil and gas industry
- Food and beverage products
- Chemical products
- Pharmaceutical products
- Water/waste water treatments
- Green Energy production
- Climate control

3.2.2 TSCH mode overview

The TSCH mode uses time synchronized communication and channel hopping to provide network robustness through spectral and temporal redundancy. Timetimeslotslotted communication links increase potential throughput by minimizing unwanted collisions that can lead to catastrophic failure and provides deterministic yet flexible bandwidth for a wide variety of applications. Channel hopping extends the effective range of communications by mitigating the effects of multipath fading and interference. TSCH is also topology independent; it can be used to form any topology from a star to a full mesh. TSCH MAC primitives can be used with higher layer networking protocols to form the basis of reliable, scalable, and flexible networks.

3.2.3 MAC behaviors unique to TSCH

The TSCH mode introduces no unique commands specific for this mode; however, IEs are used extensively. TSCH uses Enhanced Beacon frames containing the TSCH Synchronization payload IE, TSCH-Slotframe and Link payload IE, TSCH Timeslot payload IE, and Channel Hopping payload IE to advertise the presence of a TSCH PAN and allow new devices to synchronize to it. Devices maintain synchronization by exchanging acknowledged frames within defined Timeslot windows and providing timing corrections in the acknowledgment via the ACK/NACK Time Correction IE.

Communication resources are broken up into timeslots that are organized into repeating slotframes

3.3 Low Latency Deterministic Networks

3.3.1 Typical application domains and requirements addressed by LLDNs

Typical application domains addressed by the LLDN mode are as follows:

- Factory automation (such as for automotive manufacturing)
- Robots
- Overhead cranes

- Portable machine tools
- Milling machines, computer-operated lathes
- Automated dispensers
- Cargo
- Airport logistics
- Automated packaging
- Conveyors

In addition to the already stated industrial requirements, these types of applications have the following additional major requirements:

- Very low latency
- Transmission of sensor data in 10 ms
- Low round-trip time
- Many sensors per LLDN PAN coordinator (there may be more than 100 sensors per LLDN PAN coordinator)

3.3.2 LLDN application domain overview

Factory automation is comprised of a large number of LLDN devices observing and controlling the production. As an example, LLDN devices are located on robots, cranes, and portable tools in the automotive industry; collect data on machine tools, such as milling machines and lathes; and control revolving robots. Further application areas control of conveyor belts in cargo and logistics scenarios and special engineering machines. Depending on the specific needs of different factory automation branches, many more examples could be named.

Common to these applications in the context of factory automation is the requirement of low latency and high cyclic determinism. The performance should allow for reading sensor data from 20 LLDN devices within 10 ms.

Cabling industrial sensors is very time-consuming and expensive. Furthermore, cables are a frequent source of failures due to the harsh environment in a factory and may cause additional costs by production outage. Wireless access to LLDN devices eliminates the cabling issue and also provides advantages in case of mobility and retrofit situations.

Wireless technologies that could be applied to the factory automation scenario include IEEE 802.11™ (WLAN), IEEE 802.15.1™ (Bluetooth®), and IEEE 802.15.4. IEEE 802.15.4 is designed for sensor applications and offers the lowest energy consumption as well as the required communication range and capacity.[1] Moreover, four IEEE 802.15.4 channels can be utilized in good coexistence with three non-overlapping WLAN channels. Bluetooth offers good real-time capabilities, but often interferes with existing WLAN installations.

IEEE Std 802.15.4 is a worldwide and successfully applied standard for wireless and low power transmission of sensor data. Different protocols reside above IEEE 802.15.4 (*Wireless*HART® according to IEC 62591, ISA100.11a, or ZigBee®) in the context of process automation and are already in the process of standardization.[2]

[1] The Bluetooth trademark is owned by Bluetooth SIG, Inc. USA. This information is given for the convenience of users of this standard and does not constitute an endorsement by the IEEE of this product. Equivalent products may be used if they can be shown to lead to the same results.

Those protocols aim at different requirements, but employ the same physical layer hardware as the proposed solution for factory automation, which indicates potential hardware synergies and cost savings. Thus, a solution for factory automation based on IEEE Std 802.15.4 is appropriate.

It can be shown that CSMA-CA operation cannot provide a bounded media access time duration while the IEEE 802.15.4 beacon-enabled PAN's seven guaranteed timeslottimeslots in an interval of 15 ms does not fulfill the factory automation requirements. Therefore, a modification of the IEEE 802.15.4 MAC for the industrial factory automation application domain, i.e., defining a fine granular deterministic TDMA access, is required.

3.3.3 LLDN assumptions

The proposed system should be operated in a controlled configuration to achieve the required performance. Thus, it is assumed that the system is operated in a controlled RF environment with frequency planning. The TDMA channels are allocated in a way that eliminates interference and coexistence issues.

3.3.4 LLDN mode characteristics

Allocating a dedicated timeslot for each LLDN device provides a deterministic system. The IEEE 802.15.4 DSSS coding together with the exclusive channel access for each LLDN device ensures high reliability of the system. Keeping the packets and timeslots short lead to superframes with durations as short as 10 ms. The number of slots in a superframe determines the number of LLDN devices that can access each channel. This solution can be extended by operating the LLDN PAN coordinator with multiple transceivers on different channels to support a high number of LLDN devices.

Due to the stringent requirements of these low-latency applications, the LLDN mode uses the star topology with a minimal superframe structure. Short MAC frames with a 1-octet MAC header (shortened frame control) are deployed to accelerate frame processing and to reduce transmission time.

3.3.4.1 Time Division Multiple Access (TDMA)

The medium in the LLDN mode is accessed by a TDMA scheme that is defined by a superframe of fixed length. The superframe is synchronized with a beacon transmitted periodically from the LLDN PAN coordinator. Access within the superframe is divided into timeslots. The superframe can be configured to provide the full range of operation from a completely deterministic access to a shared access. To provide deterministic access each device is assigned to a particular timeslot of fixed length. Shared group timeslots allow multiple access for a group of nodes within a duration enclosing an arbitrary number (up to the whole superframe) of dedicated timeslots.

3.3.4.2 Addressing

The LLDN mode supports two addressing schemes. The first addressing scheme is based on the timeslot assigned to a device for communication, i.e., the timeslot corresponds exactly to a single device. The second scheme supports the short address format. Both addressing schemes can be used within a single LLDN.

3.3.4.3 Network topology

LLDN devices are connected to a single LLDN PAN coordinator in a star topology shown in Figure 1. As an example, the LLDN devices attached to sensors would send the sensor data uplink to the LLDN PAN

[2] *Wireless*HART is a registered trademark of the HART Communication Foundation. ZigBee is a registered trademark of the ZigBee Alliance. This information is given for the convenience of users of this standard and does not constitute an endorsement by the IEEE of this product. Equivalent products may be used if they can be shown to lead to the same results.

coordinator while LLDN devices attached to actuators would be configured to exchange data uplink and downlink with the LLDN PAN coordinator.

The selection configuration of nodes, channels, and timeslots for communication can be planned in a higher layer. The LLDN devices are configured over the LLDN PAN coordinator possibly based on planning information received from the higher layer.

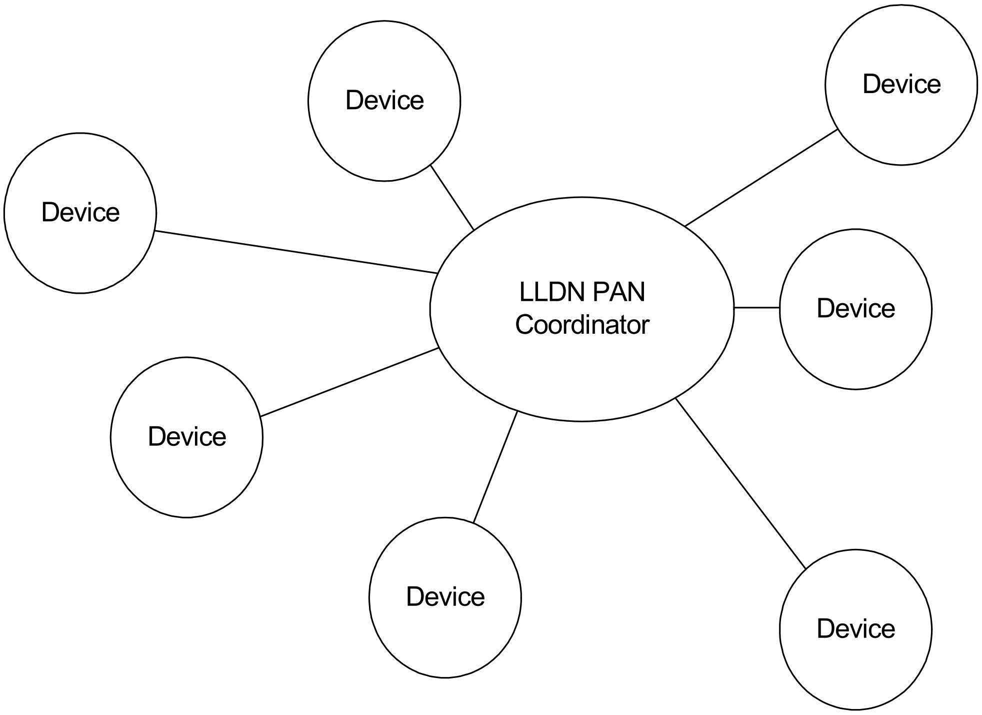

Figure 1—Star topology LL MAC

3.3.5 MAC behaviors unique to the LLDN mode

The following MAC commands are specific for the LLDN mode: LL-Discover response, LL-Configuration request, LL-Configuration status, LL-CTS shared group, LL-Request To Send (RTS), LL-Clear to send (CTS).

3.4 Deterministic and synchronous multi channel extension (DSME)

3.4.1 Typical application domains and requirements

The DSME mode is designed for following types of application domains:

- Industrial applications: process automation, factory automation, smart metering
- Commercial applications: home automation, smart building, entertainment
- Healthcare applications: patient monitoring, telemedicine

These types of application have the following major requirements:

- Deterministic latency

- Flexibility
- High reliability
- Efficiency
- Scalability
- Robustness

3.4.2 Features

To satisfy the previous requirements, the DSME mode provides the following features:

- Multi-channel, multi-superframe, mesh extension to GTS for deterministic latency, flexibility, and scalability
- Group acknowledgment option for high reliability and efficiency
- Distributed beacon scheduling and distributed slot allocation for robustness and scalability
- Two channel diversity modes (channel adaptation and channel hopping) as described 3.4.5 in for robustness and high reliability even in dynamic channel conditions

It is noted that the GTS described in IEEE Std 802.15.4-2011 can support applications requiring deterministic delay. However, this advantage is limited. First, IEEE Std 802.15.4-2011 supports only up to seven guaranteed timeslots (GTSs), which is not enough to support large scale network applications. Second, the coverage is limited because GTSs are supported only within one hop from the PAN coordinator. Finally, GTSs are restricted to use a single channel.

The DSME mode eliminates the noted GTS limitations to provide time bounded and reliable services for a variety of industrial and commercial applications. For capacity enhancement, the DSME mode enhances IEEE Std 802.15.4-2011 in two important directions: extension of number of GTS timeslots and the number of frequency channels used. To accommodate these enhancements, the DSME mode adopts a versatile frame structure to overcome the seven-slot limitation and single channel operation. In order to save energy, the new frame structure provides CAP reduction. Importantly, using DSME, GTS service can be extended to cover multihop mesh networks with deterministic latency. For reliability, the DSME mode supports two channel diversity methods described in 3.4.5 so that designers of mesh networks can choose either channel diversity method to meet the network objectives. The DSME mode is scalable and does not suffer from a single point of failure because beacon scheduling and slot allocation are performed in a distributed manner. In addition, the DSME mode supports a group Acknowledgement option to provide a retransmission opportunity within the same superframe for a data frame that failed in its DSME GTS transmission. The group Acknowledgement also improves efficiency because the Acknowledgement for multiple data frames is aggregated in a single ACK frame.

3.4.3 Recommended DSME parameter settings for various application types

Recommended DSME mode parameter settings for various application types are presented in Table 1. For applications such as factory automation require low delay; setting small *macSuperframeOrder* and *macMultiSuperframeOrder* values can reduce the delay. Applications where reliability is critical, such as patient monitoring, require very low probability of data loss. Group ACK can improve the effectiveness of retransmissions. The loss due to time synchronization error can be reduced using Deferred Beacon and Retrieve Synchronization. Applications that have low duty cycles that require low-energy usage such as infrastructure monitoring can use the DSME setting where a device is allowed to sleep during the CFP except for DSME GTS that are assigned to the device. Note that reducing the CAP increases the duration of CFP and thus increases the duration of sleeping. High throughput applications such as file transfers can be supported by setting relatively large *macSuperframeOrder* value creating a long CAP. For large scale applications such as smart utility networks and process control application domains that have a large number of devices to be supported in a PAN, setting a relatively high value for *macMultiSuperframeOrder* can increase the number of devices supported by DSME GTSs.

Table 1—Recommended DSME parameter setting for various application types, where BO = *macBeaconOrder*, SO = *macSuperframeOrder*, MO = *macMultiSuperframeOrder*

Application type	BO	SO	MO	CAP reduction	Group ACK	Deferred beacon	DSME GTS retrieve synchronization
Delay sensitive	6	0	1	Enabled	Enabled	Enabled	Enabled
Reliability sensitive	8	3		Disabled	Enabled	Enabled	Enabled
Energy critical	14	1	14	Enabled	Disabled	Disabled	Disabled
High throughput	10	5	6	Disabled	Disabled	Disabled	Disabled
Large scale	10	1	8	Enabled	Disabled	Enabled	Enabled

3.4.4 MAC behaviors unique to the DSME mode

The following MAC commands are specific to the DSME mode: DSME Association request, DSME Association reply, DSME GTS request, DSME GTS reply, DSME GTS notify, DSME Information request, DSME Information reply, DSME Beacon allocation notification, DSME Beacon collision notification, DSME Link status report. The following IE is specific to DSME: DSME PAN Descriptor header IE.

3.4.5 Channel diversity

3.4.5.1 Overview

Wireless PANs encounter significant receiver channel quality variations, which degrades the averaged signal reception quality. The main causes of the channel quality variations are multi-path fading and mutual RF interference. As an enhancement to the IEEE 802.15.4 MAC, two types of channel diversity methods, channel adaptation and channel hopping, are provided to overcome these impairments. Channel adaptation is a method whereby the channel is not changed unless the received signal quality drops down lower than a threshold value. In this manner, when channel quality is poor, it switches the channel to another one that is expected to show statistically different reception quality. Channel hopping is a method whereby the channel is periodically switched according to a predefined channel hopping sequence. The channel hopping sequence is defined by a layer above the MAC. The basic idea behind these channel diversity methods is to exploit the large diversity of receiver channel quality over a wide RF channel spectrum. The chances of a channel suffering impairments is statistically much lower than another one suffering deep fading located far apart. Thus, the average reception signal quality is expected to be significantly improved by switching from one channel to other channel frequency significantly different than the first channel.

3.4.5.2 Channel diversity methods

The IEEE 802.15.4 MAC describes two types of PAN operations: beacon-enabled and nonbeacon-enabled. Channel adaptation is implemented over the DSME structure in a beacon enabled PAN, while channel hopping can be implemented in either of the PAN operations. Channel hopping and channel adaptation do not exclude PHY frequency hopping in which the PHY changes the channels independent of the MAC. In this manner, the fundamental difference between the channel hopping done in the MAC versus channel hopping done in the PHY is whether channel switching can occur during the transmission of a PPDU.

Figure 2 illustrates the hopping methods. In the MAC channel hopping scheme, each PPDU is transmitted on different frequency channel, as shown in Figure 2(a). In the PHY frequency hopping a PPDU is divided into three fragments and each fragment is transmitted in a different sub timeslot with a different frequency channel, as shown in Figure 2(b).

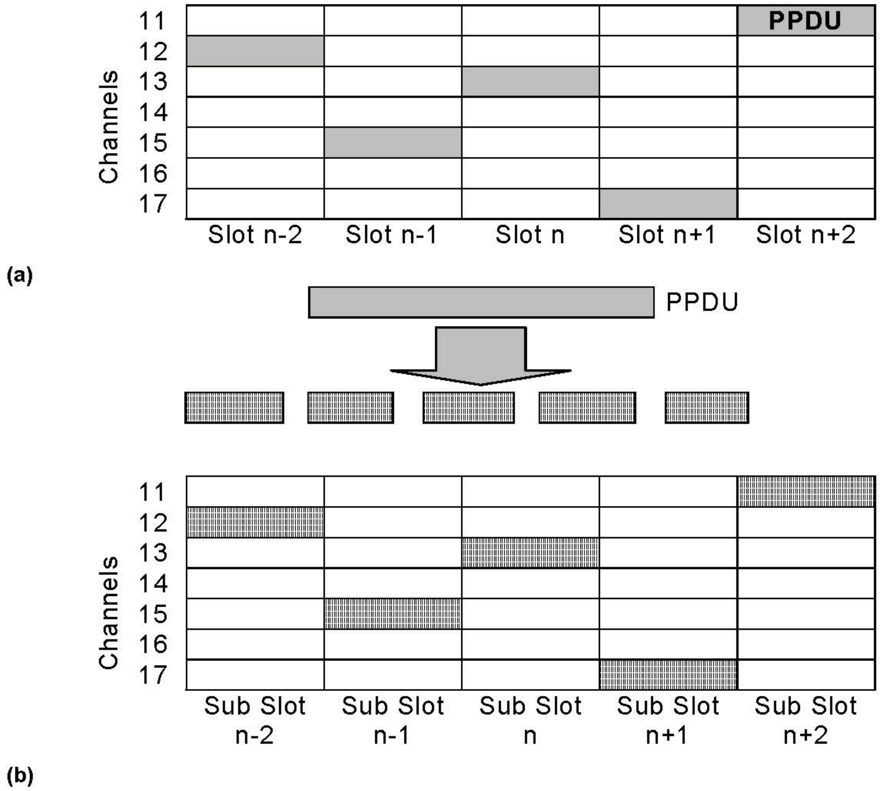

Figure 2—Illustration of channel hopping in (a) MAC (b) PHY Layer

The notion of the channel number gets complex and possibly conflicted when a combined channel diversity scheme (i.e., MAC and PHY) is used. To illustrate this point, consider a Hopping Sequence of {1,2,3,4} to be used by the MAC channel hopping. Before the transmission of the first PPDU, the PHY would configure the physical channel as per information obtained from hopping sequence for MAC channel hopping for while the it would also be configured as per the PHY frequency hopping sequence. Therefore the PHY could have conflicting requirements as to which channel should be used to transmit the frame. To resolve this conflict, the concept of a logical channel number being mapped to a PHY frequency hopping sequence is shown in Table 2. In this example if the PHY frequency hopping employs hopping sequences, {1,3,5,7}, {2,4,6,8}, {9,11,13,15}, and {10,12,14,16}, each sequence would be referred to as a logical channel numbers 1 through 4, respectively. Continuing with this example, if the MAC sets the logical channel number to 1, the PHY would use the hopping sequence {1,3,5,7} for the transmission of a PPDU. Table 3 shows an example of logical channel numbers mapped to PHY hopping sequences.

3.4.5.3 MAC behaviors unique to channel diversity

The following IEs are specific to Channel Diversity: the Channel Hopping IE and the Hopping Timing IE.

Table 2—Logical channel numbering

PHY Hopping Sequence	Logical channel number
{1, 3, 5, 7}	1
{2, 4, 6, 8}	2
{9, 11, 13, 15}	3
{10, 12, 14, 16}	4

Table 3—PHY Hopping Sequences using the notion of logical channels

MAC Hopping Sequence	PHY Hopping Sequences
{1,2,3,4}	{{1,3,5,7},{2,4,6,8},{9,11,13,15}, {10,12,14,16}}
{2,3,4,1}	{{2,4,6,8},{9,11,13,15}, {10,12,14,16},{1,3,5,7}}
{3,4,1,2}	{{9,11,13,15},{10,12,14,16},{1,3,5,7}, {2,4,6,8}}
{4,1,2,3}	{{10,12,14,16},{1,3,5,7}, {2,4,6,8},{9,11,13,15}}

3.5 Blink

3.5.1 Blink mode overview

The Blink mode provides a method for a device to communicate its ID (i.e., the EUI-64 Source Address) and/or an alternate ID (in the payload), and optionally additional payload data to other devices, without prior association and without an Acknowledgement. The frame can be used by “transmit only” devices to co-exist within a network, utilizing an Aloha protocol. Any devices that are not interested in this Blink frame can reject the frame at an early stage during frame processing and not burden the whole MAC or higher communication layers with this data traffic.

The Blink mode is based upon the multipurpose frame to yield a minimal frame consisting only of the header fields necessary for its proper operation.

3.5.2 MAC behaviors unique to the Blink mode

There are no MAC commands specific to the Blink mode.

3.6 Asymmetric multi-channel adaptation (AMCA)

3.6.1 AMCA mode overview

There are deployments where a single common channel approach may not be able to connect all devices in a PAN. The variance of channel quality can be large; additionally link asymmetry can occur between two neighboring devices causing one device to not properly receive the other device’s transmissions. Such cases are likely to happen in large, geographically diverse networks such as smart utility networks, infrastructure monitoring networks, and process control networks. It is for these cases that the AMCA mode is defined. The AMCA mode is used in a nonbeacon-enabled PAN.

3.6.2 MAC behaviors unique to AMCA

The following MAC commands are specific to the AMCA mode: AMCA Beacon Request, and AMCA Channel probe.

3.7 Low energy

3.7.1 LE overview

The low-energy (LE) mechanisms are not specific to any one application domain, rather they are suitable for applications that are willing to trade low latency for low-energy consumption. The LE protocol allows devices to operate down to a fraction of 1% duty cycles while presenting an always-on illusion. The always-on illusion is good for the following:

- Internet protocol (IP) networks, it is what they are used to, i.e., it is a typical assumption
- Manageability, it allows for stateless devices, no prior synchronization of time and state required
- Asynchronous, event-driven, and/or infrequent communications
- Mobility and discovery
- Shifting overhead to transmissions

These behaviors are not exclusive of other low-energy mechanisms, i.e., other low-energy techniques can be layered on top.

Low-energy mechanisms support the conventional superframe structure and they are also applicable in the nonbeacon-enabled PAN as well as in the CAP periods of the beacon mode. It is also possible for the upper layer to temporarily turn off the low-energy mechanisms by operating the radio at 100% duty cycle for emergency messages.

There are two low-energy mechanisms: *coordinated sampled listening* (CSL) and *receiver initiated transmissions* (RIT). CSL is suitable for applications with relatively low latency requirements, e.g., < 1 s. RIT is suitable for applications with a high latency tolerance, e.g., tens of seconds. RIT is also required in cases where the local regulation limits the duration of continuous transmissions to too small a period for CSL to be effective. RIT mode is applicable to low duty cycle, low traffic load type of applications, and especially suitable in the case that consecutive radio emission time is limited by regional or national regulation (e.g., 950 MHz band in Japan).

3.7.2 MAC behaviors unique to LE

The following MAC command is specific to LE: RIT Data Request command The following IEs are specific to LE: LE CSL IE, LE RIT IE, and RZ Time IE.

3.8 Information elements

3.8.1 Overview

The concept of information elements (IEs) has previously been used in IEEE Std 802.11-2007 and IEEE Std 802.15.3-2003. An IE is a well defined, extensible mechanism to exchange data at the MAC sublayer. The IE implementation in this standard leverages the work done in IEEE Std 802.11 and IEEE Std 802.15.3 with alterations to better suit the IEEE 802.15.4 MAC and PHY and the applications served by the standard. It should be noted that the IE implementation allows the MAC to ignore IE types that are not implemented within the MAC.

3.8.2 IE types

There are two IE types: Header IEs and Payload IEs. Header IEs are used by the MAC to process the frame. Header IEs cover security, addressing, etc., and are part of the MHR. Payload IEs are destined for another layer or SAP and are part of the MAC payload.

3.8.3 IE Identification

The ID space is either managed or unmanaged. Managed IDs are defined by IEEE Std 802.15.4-2011. Unmanaged IDs are left for implementers to use with the caveat that these values are implementer specific and could cause misinterpretation in networks consisting of devices from multiple implementers.

3.9 Enhanced Beacons frames and Enhanced Beacon Request command

3.9.1 Overview

The Enhanced Beacon frame is an extension of the IEEE 802.11 beacon frame that provides greater flexibility in content than IEEE 802.15.4 beacons. A device can differentiate the Enhanced Beacon frame from the IEEE 802.15.4 beacon frame by examining the Frame Version field. The Enhanced Beacon frame format is used to construct application specific beacon content, as used in this standard by DSME and TSCH. The different application beacons are constructed by including the relevant IEs. The Enhanced Beacon frame provides a means for application-specific information provided by higher layer protocols to be included in periodic and/or aperiodic beacons. The Enhanced Beacon frame is also used as a query/ response mechanism when used with the Enhanced Beacon Request command.

The Enhanced Beacon Request command is an extension of the beacon request MAC command. The Enhanced Beacon Request command format is differentiated from the IEEE 802.15.4 beacon by the Frame Version field being set to 0b10. The content of the Enhanced Beacon Request command is constructed using IEs. The Enhanced Beacon Request command provides a means for the beacon request to specify specific response filters, providing constraints to qualify the beacon response to the request; in this manner, the beacon responses will be limited to the constraints set by the Enhanced Beacon Request command. In this way a device conducts a scan for a PAN coordinator who is allowing new devices to join the PAN, or scans to discover a subset of like neighbors, for example neighbors or PAN coordinators that support a specific set of MAC and/or PHY capabilities. The Enhanced Beacon Request command may also be used as a general query mechanism, specifying a list of requested IEs that a responding device will, if able, include in the beacons sent in response to the request. This can be used, for example, to convey MAC, PHY, or MAC and PHY capabilities between MAC entities, or exchange the performance metrics collected between MAC entities.

3.9.2 MAC behaviors unique to Enhanced Beacon frame and Enhanced Beacon Request command

Information contained in Enhanced Beacon frame specific to DSME beacons and TSCH beacons is described in the respective functional subclauses.

The enhanced active scan provides for selective scanning using response filters and a generalized query/ response mechanism. The mechanisms for higher layer interaction with the Enhanced Beacon Request command/Enhance Beacon frame exchange are described in the MLME-BEACON-REQUEST service and the MLME-BEACON-NOTIFY service.

3.10 Multipurpose frame

3.10.1 Multipurpose frame overview

The multipurpose frame was introduced to provide an extensible, flexible frame format that can address a variety of MAC operations. Using a single frame type value, the multipurpose frame structure can support multiple MAC operations and provides the extensibility to accommodate new MAC requirements while maintaining backward compatibility.

The multipurpose frame flexibility comes from being based upon IEs. Being based upon IEs allows additional functionality without changing the structure or introducing new frame types.

A higher protocol layer can direct the MAC to configure the multipurpose frame to address unique needs of a specific application as evidenced by the RFID Blink described in 3.5.

3.10.2 MAC behaviors unique to the multipurpose frame

The MCPS-DATA.request primitive includes a parameter unique to the multipurpose frame.

3.11 MAC performance metrics

3.11.1 Requirements of routing (such as in mesh networks)

Many of the deployed networks of IEEE 802.15.4 devices are very large in area of deployment as well as in numbers of devices. These factors stress the networking layer above the MAC, including the mesh networking routing. To allow the networking layer (and other layers) to make optimal decisions, information on the link performances of devices is required.

3.11.2 Requirements of long transmit durations due to large frames or low data rates

The IEEE 802.15.4 MAC was designed for previous PHY layers with upper limits of 127-octet MPDUs. However, to allow direct support of IP packets, PHY amendments have increased the PHY frame length up to a minimum of 1500 octets. Still other PHY amendments have significantly decreased the effective data rates by transmitting fewer symbols per second or by increasing the forward error correction.

The resulting long packets and low rates can yield frame durations of many hundreds of milliseconds while channel impairments and interference bursts may occur in the range of a few milliseconds. Due to this behavior, there is a need to provide a set of metrics to allow higher layers to implement dynamic fragmentation based on channel conditions.

3.11.3 Metrics

Metrics have been added that inform the higher layers on key aspects of data frame transmission and data frame reception. The metrics essentially record the instances of significant link events in counters.

The following attributes relate to data frame transmission:

- *macRetryCount*
- *macMultipleRetryCount*
- *macTXFailCount*
- *macTXSuccessCount*

The following attributes relate to data frame reception:

- — *macFCSErrorCount*
- — *macSecurityFailure*
- — *macDuplicateFrameCount*
- — *macRXSuccessCount*

If *macMetricsEnabled* is TRUE, the MAC collects the requestd metrics listed. The MAC PIB attribute *macCounterOctets* defines the size of the counters.

3.12 Fast association

3.12.1 Overview

Due to low-energy considerations, the IEEE 802.15.4 association response command is sent using indirect transmission, which introduces a significant delay in the association procedure. There are applications that prioritize reduced association time over energy consumption during that operation. The MAC behavior associated with FastA allows a device to associate in a reduced duration time.

3.12.2 Operation

IEEE 802.15.4 devices will continue to be able to use the “association request command” using indirect transmission, but devices enabled with FastA behavior may configure the “association request command” (e.g., when it associates with an IEEE 802.15.4 device) or “direct association request” for Fast A.

3.12.3 MAC behaviors unique to FastA

There are no unique MAC commands for FastA, the existing commands were enhanced to address fast association.

4. Using IEEE 802.15.4 for active RFID applications

4.1 Introduction

Active radio frequency identification (RFID) devices are used to identify and often locate people or objects in industrial or commercial environments. Typical applications include asset management, inventory management, process control and automation, safety and accountability, and many others.

In its simplest form an active RFID system comprises a number of transmit-only tags that periodically transmit a packet containing a unique ID and a small amount of data. The packet is received by one or more readers that may simply register the tag as present, may employ further processing to determine the location of the tag, or forward data to an application server. More complex active RFID systems might employ two-way communications with the tag for control, communication, and coordination.

Active RFID systems are generally characterized by the following attributes:

- Very low cost, low energy consumption tags
- Large populations of tags
- Low duty-cycle transmissions
- A variety of readers from very short range (a few meters) to very long range (hundreds of meters)
- Very short packet lengths, often with no data beyond the device ID and a small number of status bits
- Sensor data may also be transmitted

IEEE Std 802.15.4-2015 contains various MAC mechanisms targeted at enabling active RFID applications and three PHYs particularly suitable for this purpose, as follows:

- LRP UWB PHY
- HRP UWB PHY
- MSK PHY

This document describes how these PHYs and MAC features can be configured for active RFID applications.

4.2 Overview of active RFID PHYs

4.2.1 LRP UWB PHY

4.2.1.1 Description

The low rate PRF ultra wideband (LRP UWB) PHY is a low complexity PHY optimized for active RFID devices. In active RFID systems the hardware components are highly asymmetric, with large populations of very low complexity tags being identified by much smaller populations of potentially complex readers. Typically a tag transmits messages to a reader, although the reverse architecture is also possible (though less common).

The LRP UWB PHY has therefore been predominantly driven by the need for very low complexity transmitters (tags). Low complexity considerations include the following:

- Simple to implement modulation
- No data encoding or whitening in base mode

- Simple to implement PSD mask
- No dithering of pulses for spectral smoothing
- Relaxed timing requirements

Additionally, the low rate PRF is a key feature that reduces location ambiguity and improves the performance of non-coherent receivers in high RF multipath environments.

The low complexity approach drives low energy consumption and low cost in discrete device implementations, which are common in lower volume applications. Where very high volumes of devices are sold, silicon solutions are viable and the HRP UWB PHY becomes a feasible active RFID solution (see 4.2.2).

The following subclauses highlight the key features of each mode of operation.

4.2.1.1.1 Base mode

The base mode is the lowest complexity mode. It is used where there is a requirement for very large tag populations, but no requirement for very long-range operation. Typically long range is not an issue in environments with a large number of line-of-sight obstructions.

The key base mode attributes are as follows:

- Very simple modulation (OOK)
- No further encoding or whitening
- Shortest packet length
- Packet length designed to achieve maximum pulse amplitude under most global UWB regulations (192 pulses at 1 MHz PRF)

This last point makes the base mode particularly useful for non-coherent (RF energy detect) receivers that benefit from instantaneous pulse amplitude.

A receiver encountering an incoming stream of UWB pulses will go through a process similar to the following in ascertaining whether it is receiving a base mode packet:

- A tone check on the incoming preamble to identify a 1 MHz pulse train (i.e., not a long-range mode packet)
- An SFD search, with a match confirming one of either base mode or enhanced mode
- A check of the first three bits of the PHR (the three bits immediately following the SFD) to confirm base mode
- Three zeros = base mode
- Three ones = extended mode
- The receiver may choose to vote on the first three bits of the PHY header in a “best of three” manner in order to introduce simple error-proofing on the Encoding Type bits.

It should be noted that the preamble length is variable between 16 pulses and 128 pulses at the discretion of the implementer. Sixteen pulses have been shown to be sufficient for a wide variety of use cases, but the implementer may desire to improve acquisition performance by increasing the preamble length. However, doing so has three important negative effects on overall system performance, so the choice should be carefully considered. A longer preamble will:

- Risk increasing the packet length beyond 192 pulses, which may require the pulse amplitude to be reduced in order to comply with local UWB average emission limits.
- Increase power consumption in the tag (more pulses transmitted and more processor on-time).
- Increase tag packet collisions (due to longer packets).

4.2.1.1.2 Extended mode

The extended mode adds rate 1/4 convolutional code to provide simple forward error correction to the base mode for improved performance in certain circumstances. Additionally, the extended mode allows for a longer preamble (up to 256 pulses) if the implementer needs to increase acquisition robustness. The encoding scheme is simple to implement in the transmitter, and allows for two different decoding schemes in the receiver.

Since the extended mode packet length is longer than a base mode packet, the signal is likely to become constrained by the regulatory average emission limits for UWB. This requires individual pulse amplitudes to be reduced, which causes a loss in terms of link budget when using a non-coherent receiver. This loss is to be weighed against an expected coding gain of up to 4 dB to determine whether extended mode has net benefit in any given circumstance.

In general the use of extended mode is a trade-off, balancing coding gain with packet length (i.e., pulse energy loss). The packet length will be determined by a number of factors including preamble length, addressing mode used, and payload size.

The receiver process for identifying an extended mode packet is the same as for the base mode.

4.2.1.1.3 Long-range mode

The long-range mode is targeted at coherent receivers that can leverage coherent pulse integration in order to increase symbol energy. Since multiple pulses are integrated together to form a single symbol, packet length can be long, meaning the pulse amplitude is certainly defined by regulatory average emission limits for UWB. However, coherent pulse integration may compensate for the loss in pulse energy for a net gain in link budget.

The performance of a coherent receiver depends primarily on two factors: the accuracy of the template pulse used in the receiver, and the accuracy of the timing synchronization at the pulse level between the receiver and the transmitter. The LRP UWB PHY does not make stringent demands on either of these parameters; this is intentional in order to allow for low cost implementations (without sophisticated timing or pulse shaping). For this reason there is a limit to the extent that coherent gain can be achieved by simply adding more pulses per symbol. It should be noted that increased pulses per symbol decreases the maximum potential population of tags within a physical region for any given message rate.There is also a pulse amplitude penalty when adding more pulses due to regulatory limits. The parameters selected for the long-range mode therefore represent the peak performance for a coherent receiver with relatively relaxed timing and pulse shaping; longer symbols would add little to coherent gain and only serve to reduce pulse energy.

The long-range mode uses Manchester encoded binary PPM as a modulation scheme, rather than simple OOK as in the base and extended modes, and operates at a 2 MHz PRF. It also uses a more complex preamble necessary to support this encoding scheme. The preamble consists of the following:

- A sequence of between 1024 and 8192 pulses, then
- The SFD encoded as per the base and extended modes, then
- A sequence of between 16 and 64 binary 1 symbols encoded as per the long-range mode

The preamble is then followed by the SFD encoded as per the long-range mode. The purpose of this more complex preamble is to allow a coherent receiver to detect a long-range mode packet and achieve symbol synchronization before the SFD, and also to allow a non-coherent receiver to achieve synchronization (see 4.2.1.2).

A receiver encountering an incoming stream of UWB pulses will go through a process similar to the following in ascertaining whether it is receiving a long-range mode packet:

- A tone check on the incoming preamble to identify a 2 MHz pulse train (i.e., not a base or extended mode packet).
- A wait for the start of pulse transitions in the form of 32 present pulses followed by 32 absent pulses, repeating. At this point the receiver will use these transitions to achieve symbol synchronization.
- An SFD search, with a match confirming long-range mode.
- A check of the first three bits of the PHR (the three bits immediately following the SFD) with “000” reconfirming long-range mode (other values are reserved for future use).

4.2.1.2 Mixed mode networks

4.2.1.2.1 Performance considerations

The long-range mode is primarily targeted at coherent receivers, whereas the base and extended modes are primarily targeted at non-coherent receivers. Since active RFID systems are generally part of a fixed infrastructure, it is unlikely that a mix of coherent and non-coherent receivers will be deployed in any given location. Instead, the characteristics of the location will demand one or the other type of receiver for best overall system performance.

For example, an indoor environment with many obstructions, such as a warehouse or manufacturing facility, may not afford a long line of sight distances and so extended range is not useful. Instead, a higher density of non-coherent receivers is likely to be deployed with tags operating either in base mode or extended mode.

By contrast, large open space such as outdoors will afford very much longer line of sight distances, so a lower density of longer range receivers might be more cost-effective. In this case a coherent receiver infrastructure may be deployed with tags operating in long-range mode.

There are many cases, however, where tags may roam from coherent to non-coherent infrastructure, and so this standard requires that all receivers, coherent or non-coherent, are able to receive and demodulate all modes. The “cross” cases do suffer from performance limitations as follows:

A non-coherent receiver will have a short range of operation with a long-range mode tag when compared to either a base or extended mode tag. A long-range mode tag has relatively long packets that cause the emissions to be regulated by UWB average emission limits. This means that long-range mode pulses are smaller in amplitude than base or extended mode pulses. Since a non-coherent receiver relies on single pulse amplitude for reception, it cannot receive the smaller long-range mode pulses over a comparable range. Because of the shorter range of operation, a long-range tag may be only intermittently detected within a network of non-coherent receivers, and location of the tag may not be possible.

A coherent receiver will have a short range of operation with a base or extended mode tag when compared to a long-range mode tag. The base and extended mode tags do not provide sequences of pulses sufficient for coherent pulse integration, which negates the coherent receiver advantage. However, a coherent receiver may operate with similar range to a non-coherent receiver when operating with base and extended mode tags. Since coherent receivers will generally be deployed more sparsely than non-coherent receivers, a base or extended mode tag may be only intermittently detected within a network of coherent receivers and location of the tag may not be possible

The cross-case characteristics are summarized in Table 4.

Table 4—Cross-case characteristics

Receiver type	Tag mode	Reception range	Notes
Non-coherent	Base	Good	Typical configuration for large populations of very simple tags
	Extended	Potentially Better	Extended range operation in some circumstances
	Long range	Shortened	Short-range reception means tag detection coverage will be intermittent in typical non-coherent networks
Coherent	Base	Good	Wider spacing of coherent receivers means tag detection coverage will be intermittent in typical networks
	Extended	Better	
	Long range	Best	Typical configuration for smaller tag populations in open areas

4.2.1.2.2 Receiver processing considerations

Any of the three modes can be easily decoded by any receiver type once the RF signal has been converted to baseband. The key to operation in the cross cases is to be able to synchronize and demodulate the incoming RF.

When a coherent receiver detects a pulse train at 1 MHz PRF, indicating a base or extended mode packet, it should start to search for the preamble encoded as per the base and extended modes. Once the preamble is detected, and synchronization achieved, the remainder of the packet can be demodulated as per a non-coherent receiver.

When a non-coherent receiver detects a 2 MHz pulse train, then a long-range mode packet is indicated. The receiver may run its normal acquisition/SFD sync process, but at a 2 MHz rate. The insertion of the base/extended mode SFD in the long-range mode preamble facilitates this functionality.

After the inserted SFD, the non-coherent receiver will start to demodulate bits with allowance for the pulse repetition and Manchester encoding. This might be accomplished by running the normal OOK engine at 2 MHz and appending a compression step that translates the 32 pulse repetition Manchester PPM into actual bits. This is depicted in Figure 3.

Alternatively the non-coherent receiver may take an analog approach, averaging pulses over a 32 pulse window, and then applying OOK demodulation and Manchester decoding as depicted in Figure 4.

After detecting the SFD, a non-coherent receiver should ignore the sequence of “1” symbols (transmitted to aid coherent receiver synchronization) and search for the “000” normal SFD, encoded as long-range mode symbols, as a trigger to start demodulation of the rest of the packet. It is likely that some implementers will choose to utilize the inserted “1” symbols for positioning purposes if desired.

4.2.1.3 Timing and synchronization

A receiver requires regular pulses in order to maintain bit synchronization. The Manchester encoding in the long-range mode ensures that there are plenty of synchronization pulses, since both data “1” and data “0”

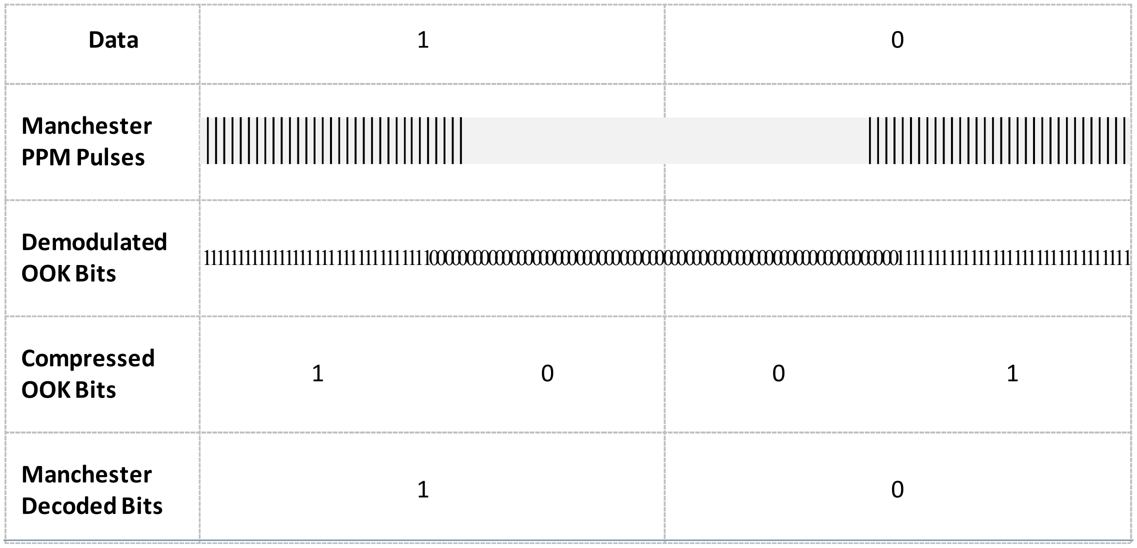

Figure 3—Converting Manchester PPM to OOK using digital bit compression

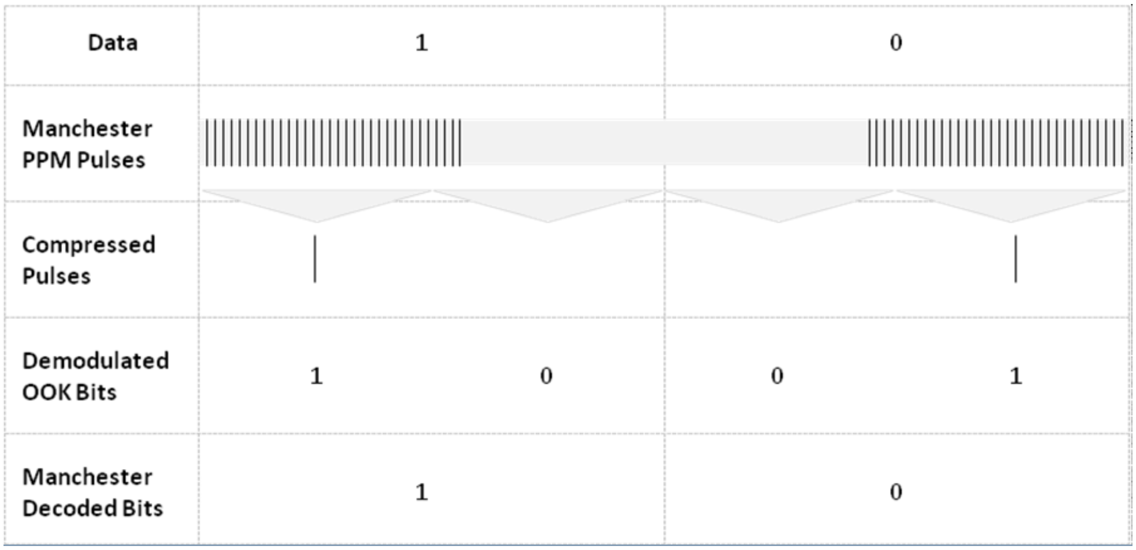

Figure 4—Converting Manchester PPM to OOK using analog pulse compression

contain pulses. However, the base and extended modes use OOK modulation, where a data “0” is denoted by the absence of a pulse. In this case a long string of zeros will cause a long period with no pulses, which in turn causes a synchronization issue.

One method of dealing with long periods with no pulses is for devices to employ very high quality timing systems that enable synchronization to be maintained over long periods with no inputs. However, in keeping with the desire to allow the use of low cost components in active RFID devices, a better approach is to ensure that there are never any long periods without pulses.

The LRP UWB PHY therefore requires that the transmitter insert a sequence of four pulses after every 128th chip (pulse). These pulses ensure a regular synchronization signal and are to be ignored by the receiver PHY.

The standard states that the transmission time of any individual pulse shall not drift more than 11 ns from its nominal transmission time during a 128 symbol period over the specified operating temperature range of the device. The inserted sync pulses ensure that this can be achieved using standard AT-cut crystals over normal temperature ranges. Devices specified for wider temperature ranges may need to use temperature compensated crystals to maintain synchronization. The 11 ns figure ensures that bit boundaries are maintained sufficient to minimize bit errors due to timing offsets.

4.2.2 HRP UWB PHY

The HRP UWB PHY was introduced into the standard as a high mobility, long-range PHY for industrial environments, with the capability to locate devices with better than 1 m accuracy. Although the modulation and encoding schemes to achieve this goal are complex, device cost can be minimized through silicon-based solutions.

Generally speaking, the LRP UWB PHY is advantageous when the market dynamics favor discrete device implementation; whereas the HRP UWB PHY becomes attractive when device volumes are sufficiently high to justify silicon integration. The high bit rate mode of the UWB PHY allows its packet duration to be very short to support a large population of active RFID tags as a trade off for longer range.

4.2.2.1 Interoperability with the LRP UWB PHY

The LRP UWB PHYs characteristics have been deliberately chosen to be complementary to the HRP UWB PHY. They use the same center frequencies, both are impulse radio UWB PHYs and both can be implemented with coherent or non-coherent receivers. These characteristics allow a receiver architecture to be designed that can demodulate either type of PHY with maximal reuse of receiver blocks. This provides for interoperability shown in Figure 5.

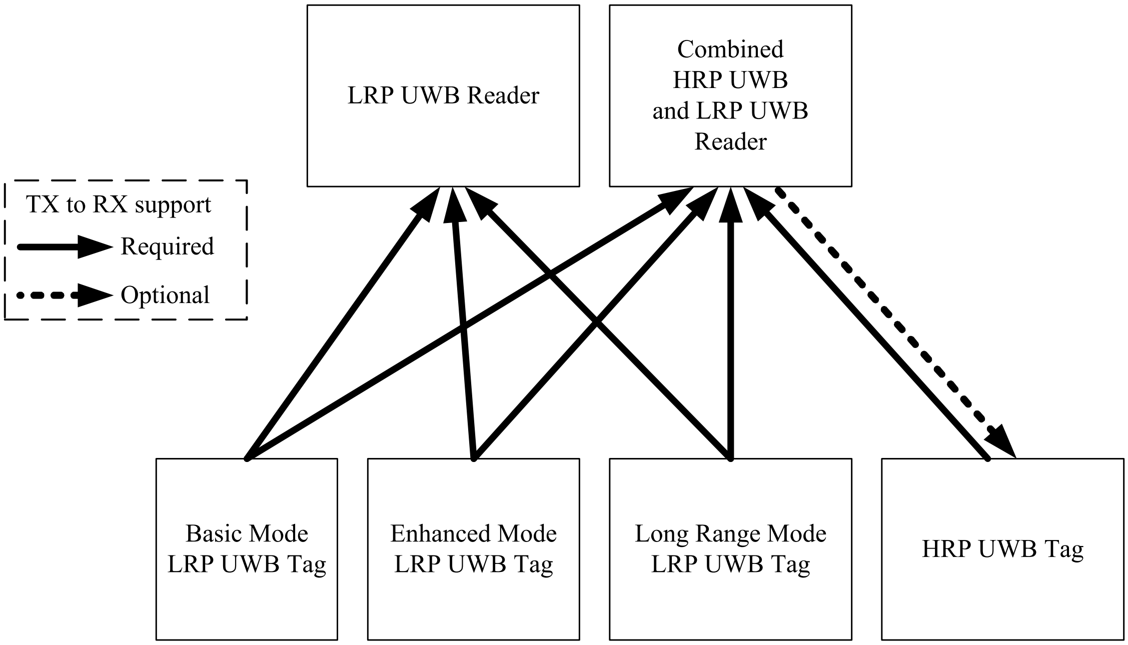
LRP UWB Reader
Combined HRP UWB and LRP UWB Reader
TX to RX support
Required
Optional
Basic Mode LRP UWB Tag
Enhanced Mode LRP UWB Tag
Long Range Mode LRP UWB Tag
HRP UWB Tag

Figure 5—LRP UWB PHY and HRP UWB PHY interoperability

Both UWB PHYs are highly suitable for active RFID/RTLS systems based on time difference of arrival (TDOA). Indeed the choices available for its preamble length and data rate allow it to operate usefully in multiple scenarios. For instance, where long range is necessary a long preamble and the 110 kb/s data rate can be employed to maximize the range, or, where a higher density of tags and/or more frequent updates are needed, over a shorter range, the UWB PHY's high data rates and short preambles can be employed.

Readers that support demodulation of frames from tags employing the HRP UWB PHY are facilitated to receive frames from LRP UWB tags by the alignment of the LRP UWB PHY band plan with channels 5, 6, 7, 9, and 11 of the HRP UWB PHY. So for instance, in an integrated circuit supporting reception of the HRP UWB PHY channel 5, which is 500 MHz centered at 6489.6 MHz, the analog portion can easily be capable of receiving the LRP UWB PHY channel 0, guaranteed to have at least 400 MHz bandwidth about 6489.6 MHz, its nominal frequency. In such an integrated circuit the digital decoding of the LRP modulation is a relatively small addition to make a device capable of supporting both the HRP UWB and LRP UWB modulations.

In mixed mode networks, it is required that the reader nodes be able to identify the modulation and appropriately demodulate it. For the LRP PHY the ability to do this is an integral part of this specification intended to be a base requirement for all its receivers as described in 4.2.1.2. With the capability to support both the HRP UWB and LPF UWB PHYs, such universal UWB active RFID readers provide for a single infrastructure that simultaneously supports all these standard UWB active RFID/RTLS tags, see Figure 5.

4.2.3 MSK PHY 2450 MHz band

4.2.3.1 Description

The MSK PHY 2450 MHz band is a narrowband PHY, operating in the band 2400 MHz to 2483.5 MHz and targeted at active RFID devices. It builds on the protocols used in traditional IEEE 802.15.4 systems, but the PHY uses non-spread-spectrum techniques that allow non-interfering operation in (typically industrial) sites where other 2.4 GHz ISM-band systems are already in use.

It can be used stand-alone for non-precision active RFID and data transfer applications or in conjunction with other IEEE 802.15.4 PHYs (e.g., LRP UWB PHY) for control and regulatory compliance. An example of a device that might use the LRP UWB PHY and the MSK PHY 2450 MHz band together is a small location tracking tag, which is operating in a regulatory domain that prohibits LRP UWB transmissions unless the tag is in the vicinity of a fixed reader infrastructure. The reader infrastructure might periodically transmit an "OK to transmit" signal via the MSK PHY 2450 MHz band, and the tag would inhibit its LRP UWB transmissions unless it had recently received such a signal.

4.2.3.2 Channel selection and interoperability with other 2.4 GHz services

The MSK PHY 2450 MHz band is compatible with global regulations relating to the 2.4 GHz ISM band and offers bidirectional capability with low circuit complexity, lower sidelobes than O-QPSK (leading to improved compatibility with neighboring users of the radio spectrum), and is implemented by commonly available transceiver ICs.

A typical target environment for the MSK PHY 2450 MHz band is an industrial site. These sites tend to have many existing systems (e.g., Wi-Fi®, ZigBee®) that use the 2.4 GHz ISM band. In these production environments, where frequency usage is carefully-planned, spectrum managers will insist that any new systems operate only on currently unused frequencies.[3]

[3] ZigBee is a registered trademark of the ZigBee Alliance. Wi-Fi is a registered trademark of the Wi-Fi Alliance. This information is given for the convenience of users of this standard and does not constitute an endorsement by the IEEE of this product. Equivalent products may be used if they can be shown to lead to the same results.

Fortunately, in a typical environment of this kind, there tend to be small gaps in the RF spectrum, at the edges of the regions used by existing systems, where narrowband channels could coexist gracefully with those existing systems. The MSK PHY 2450 MHz band is designed with two key characteristics to support the use of these gaps, as follows:

- First, the channel bandwidth of the MSK PHY 2450 MHz band is small, because of its moderate data rate (250 kb/s) and its non-spread-spectrum nature.
- Second, the MSK PHY 2450 MHz band defines a large number of possible channels (42), so that it is likely that a suitable channel will fall in the region of the gap.

Furthermore, active RFID systems (for which the MSK PHY 2450 MHz band is targeted) tend to involve mobile active RFID tags and a fixed reader/control infrastructure. It is anticipated that frequency managers at these sites will select a set of suitable channels on which the MSK PHY 2450 MHz band can operate without causing interference to (or receiving interference from) existing systems.

One channel (channel page 7, channel number 47: 2463.75 MHz) is designated as a default channel.

4.2.3.3 Discovery and orphan recovery

Since devices (e.g., active RFID tags) may roam between sites and the configuration of a site may change, and since there are a large number of possible channels, it is necessary to consider how devices may determine the set of MSK PHY 2450 MHz band channels that are in use at a particular site, and how they can recover in the event that they lose contact with other devices in the system.

4.2.3.3.1 Tag initiated

It is expected that tags that are not associated with other devices in the system will search and/or beacon periodically on a subset of the MSK PHY 2450 MHz channels that are considered to be likely candidates for non-interference with common 2.4 GHz ISM services (e.g., Wi-Fi, ZigBee), in order to receive channel guidance from infrastructure and/or establish a connection to infrastructure. Once aware of any infrastructure operating on one of these channels (which may be a subset of the infrastructure at the site), the tag can determine the complete set of MSK PHY 2450 MHz channels that are in use at that site, either using information broadcast from the infrastructure or via a query/response mechanism, and direct its future channel selection activity appropriately.

NOTE—The precise subset of MSK PHY 2450 MHz channels that would be used for discovery is application-dependent and is out of the scope of this document.

In the case where a tag is operating on a particular MSK PHY 2450 MHz channel, it is expected that it will attempt to receive periodic messages on that channel from the surrounding infrastructure. If the tag fails to receive a number of these messages, it may conclude that either (a) it has been removed from the site, or (b) the configuration of the infrastructure has changed. In either case, it will likely revert to the discovery behavior detailed above.

4.2.3.3.2 Network initiated

It is expected that infrastructure at a particular site will beacon and/or listen periodically on a subset of the MSK PHY 2450 MHz channels that are considered to be likely candidates for non-interference with common 2.4 GHz ISM services (e.g., Wi-Fi, ZigBee), in order to provide channel selection guidance and/or establish a connection with tags at the site. Tags can then determine the complete set of MSK PHY 2450 MHz channels that are in use at that site, either using information in beacon messages from the infrastructure or via a query/response mechanism, and can then direct their future channel selection activity appropriately.

NOTE—The precise subset of MSK PHY 2450 MHz channels that would be used for discovery is application-dependent and is out of the scope of this document.

Infrastructure operating on a particular MSK PHY 2450 MHz channel may also send out channel selection or hand-off information to tags in the event that the infrastructure believes that a tag would be best served by the use of a different channel. This may occur, for example, when the link quality of a connection with a tag operating on a particular MSK PHY 2450 MHz channel tag is poor, when the tag is known to be physically at the edge of coverage of infrastructure operating on a particular MSK PHY 2450 MHz channel, or when the infrastructure knows that its own channel configuration is about to change (because of, say, system administration activity).

4.2.4 MSK PHY 433 MHz band

4.2.4.1 Description

The MSK PHY 433 MHz band is a narrowband PHY, operating in the band 433.05 MHz to 434.79 MHz, that targets active RFID/RTLS devices and systems where determination of presence and approximate position satisfies location requirements.

One of the primary characteristics of the MSK PHY 433 MHz band is a long communication range due to lower signal path loss when compared to other higher frequency PHYs within the IEEE 802.15™ family of protocols.

Its operation in frequency bands that are outside traditional IEEE 802® PHYs allows full independence and non-interfering operation in an environment where other IEEE 802 wireless protocols are used.

It can be used stand-alone for active RFID, and low resolution RTLS, sensor and data transfer applications, or in conjunction with other IEEE 802.15.4 PHYs (e.g., LRP UWB PHY) for control and regulatory compliance.

The MSK PHY 433 MHz band is compatible with global regulations relating to the 433 MHz band, which is legal band in most of the countries. It offers bidirectional capability with low circuit complexity, mature and widely available silicon technology, which is low cost to implement because it takes advantage of existing investment in integrated circuits operating in a sub-GHz band.

The MSK PHY 433 MHz band uses MSK modulation, which has low side lobes, leading to more efficient usage of radio spectrum in this narrow frequency band and which can be implemented by commonly available transceivers.

4.2.4.2 Channel selection and data rates

The MSK PHY 433 MHz band uses 15 optional channels that use one of the following three data rates:

- 31.25 kb/s data rate provides additional 10 dB in receiver sensitivity at expense of longer RF packets.
- 100 kb/s provides the optimum data rate in cases in which a balanced combination of long range, channel occupancy, and power consumption is desired.
- 250 kb/s data rate can be used for applications where short RF packets minimize channel occupancy and power consumption at the expense of reduced communication range.

There are three 250 kb/s non-overlapping channels that occupy the entire 433 MHz band.

In addition to providing longer communication range, the 31.25 kb/s channel is optimized to comply with Japanese and Korean regulatory requirements.

4.2.4.3 433.92 MHz frequency

The 433.92 MHz frequency is a central frequency in 433 MHz band and in some regions of the world that is the only central frequency that can be used in this band. In order to allow compliance with these regulations, 433 MHz PHY can use all three data rates at this frequency.

4.2.4.4 Discovery and orphan recovery

Since devices (e.g., active RFID/RTLS tags) may roam between sites and the configuration of a site may change, and since there are a large number of possible channels, it is necessary to consider how devices may determine the set of channels or a single channel of MSK PHY 433 MHz band that are in use at a particular site, and how they can recover in the event that they lose contact with other devices in the system.

4.2.4.4.1 Tag initiated

It is expected that tags that are not associated with other devices in the system will search and/or beacon periodically on a default MSK PHY 433MHz band (channel 7), in order to receive channel guidance from infrastructure and/or establish a connection to infrastructure. Once aware of any infrastructure operating on one of these channels (which may be a subset of the infrastructure at the site), the tag can determine the complete set of MSK PHY 433 MHz channels that are in use at that site, either using information in beacon messages from the infrastructure or via a query/response mechanism, and direct its future channel selection activity appropriately.

In the case where a tag is operating on a particular MSK PHY 433 MHz channel, it is expected that it will attempt to receive periodic beacon messages on that channel from the surrounding infrastructure. If the tag fails to receive a number of these messages, it may conclude that either (a) it has been removed from the site, or (b) the configuration of the infrastructure has changed. In either case, it will likely revert to the discovery behavior detailed above.

It is expected that some of the devices compliant to MSK PHY 433 MHz band will not have RF receive capability (relying exclusively on ALOHA channel access). In such a case, other, alternative methods are used to set device channel at the time of device deployment (e.g., switch setting or other communication methods, unrelated to IEEE 802.15.4 PHYs), which would guarantee the most optimal operation of ALOHA channel access.

4.2.4.4.2 Network initiated

It is expected that infrastructure at a particular site will beacon and/or listen periodically on the default channel 7 of the MSK PHY 433 MHz band in order to discover nodes that are not associated in order to establish a connection. Nodes can then determine the complete set of MSK PHY 433 MHz channels that are in use at that site, by either using information in beacon messages from the infrastructure or via a query/response mechanism, and can then direct their future channel selection activity appropriately.

Infrastructure operating on a particular MSK PHY 433 MHz channel may also send out channel selection or hand-off information to nodes in the event that the infrastructure believes that a node would be best served by the use of a different channel. This may occur, for example, when the link quality of a connection with a node operating on a particular MSK PHY 433 MHz channel is poor, when the node is known to be physically at the edge of coverage of infrastructure operating on a particular MSK PHY 433 MHz channel, or when the infrastructure knows that its own channel configuration is about to change.

4.3 Overview of active RFID MAC features

The MAC requirements for active RFID are small. The main requirement for an active RFID tag is to transmit its ID number. A secondary requirement is to also provide a sequence number to aid in location determination for techniques that utilize time measurements. For IEEE 802.15.4 active RFID tags, the ID may be the device's unique 64-bit address.

The application driving the active RFID tag periodically sends a message that includes an ID number uniquely identifying the sending tag. This periodic message is termed a *blink*. For maximum battery life the application layer will sleep in its lowest possible power state with its radio off, awaken briefly to send its blink message, and then return to its low power sleep state.

For lowest power consumption the blink frame should be as short as possible. The multipurpose blink frame is ideal for this purpose. Channel access using ALOHA is also a necessary part of the low power strategy in applications in which the active RFID tags are typically transmit-only and do not have a receiver capability to do carrier sensing. For ALOHA channel access to work well, the air-utilization of the network needs to be kept below 18%. This gives a good probability that tags' transmissions get through to the reader infrastructure.

The MAC level primitive MCPS-DATA.request is used in LRP UWB and MSK PHY active RFID tag applications to initiate transmission of the multipurpose blink frame. This primitive allows selection of the preamble length, the data rate of the LRP UWB PHY that defines its modulation mode, and specification of the inclusion of a LEIP sequence. The selection of operating mode, preamble length, and inclusion of the LEIP depend on the on the capability of the PHY and the requirements of the application in terms of the desired operating range and the desired precision of any RTLS functionality.

The active RFID infrastructure consists of fixed location active RFID reader devices that receive the multipurpose blink frames sent by the active RFID tags. If a blink is received by multiple readers and then analyzed as a group at some central point, then localization may be possible.

The application driving the IEEE 802.15.4 active RFID reader then will typically initialize and permanently turn on the receiver to continually report the arrival of blink messages received from the active RFID tags.

The MAC level primitive MCPS-DATA.indication is used by the MAC layer in an active RFID reader to deliver details of any received multipurpose blink frames to the active RFID reader upper layers. This primitive contains the source address of the blink, which identifies the sending tag, and depending on the capabilities of the receiving PHY, additional parameters that can be used to perform an RTLS function when data from multiple readers is gathered at some central localization function. These additional parameters include angle-of-arrival, receive signal strength indication, and the time of arrival as provided by the *RangingCounterStop* parameter.

4.4 Location determination

4.4.1 Locating an object through received signal strength

A transmitting device (usually active RFID/RTLS tag) is transmitting a periodic signal (multipurpose blink frame) that is received by multiple readers. Each reader measures received signal strength during the PHY Header (Preamble and SFD) and produces an RSSI value. Due to wide variations in antenna transmit patterns for small and inefficient antennas found in active RFID tags that can vary in size and form factor, the absolute value of RSSI cannot be relied upon for location determination. In order to compensate for this variation; RSSI values from multiple readers may be collected by an application on a centralized server and used to calculate approximate location of the transmitting device based on the relative RSSI values.

One of the methods to calculate location based on RSSI is “weighted center of mass” method where location is calculated based on known location of readers and RSSI values received by corresponding readers as given by:

$$X_{\mathrm{RSSI}} = \frac{\sum_{i=1}^{n} x_i r_i}{\sum_{i=1}^{n} r_i}$$

$$Y_{\mathrm{RSSI}} = \frac{\sum_{i=1}^{n} y_i r_i}{\sum_{i=1}^{n} r_i}$$

Where x_i,y_i define location of the reader and r_i is corresponding RSSI value for respective reader. Calculated values X_{RSSI} and Y_{RSSI} represent location of the transmitting device.

4.4.2 Locating an object through trilateration using time difference of arrival

4.4.2.1 Overview of trilateration

A transmitter (tag) sends a signal that is received by readers in at least three different locations. Each reader notes the signal’s time-of-arrival (TOA). The difference in arrival times at any pair of readers implies that the transmitter was located somewhere on a known hyperbola. Using two reader pairs (one reader may be common between these pairs) implies that the tag resided at the intersection of two different hyperbolas. Figure 6 illustrates this situation.

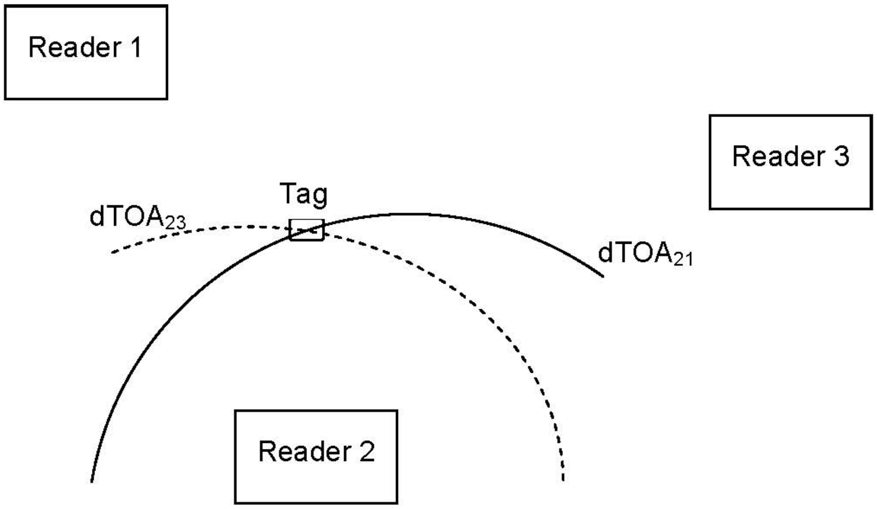

Figure 6—Locating a tag through trilateration

Figure 6 and the equations to follow all pertain to a two-dimensional (i.e., x, y) problem. If the tag is known to reside on some surface (e.g., in a pane or on a geoid) the equations can be readily modified to incorporate tag and reader elevation (z), without increasing the dimensionality of the solution. However, if tag elevation is unknown and to be evaluated, the equations must be altered to include z as a free variable and four readers will be required to produce a solution.

4.4.2.2 Mathematical location solution

Let t_x denote the time when the tag transmitted from location (x, y). The k^{th} reader, at location (x_k, y_k), will detect the signal at time $toa_k = t_x + p_k$ where p_k denotes the time required for the signal to propagate from the tag to the reader. This propagation time is equal to the tag-to-reader separation divided by the speed of light c as indicated is given by:

$$p_k = \frac{\sqrt{(x-x_k)^2+(y-y_k)^2}}{c}$$

The unknown transmission time becomes unimportant when considering the difference in arrival times at two readers ($dTOA_{jk}$) defined by:

$$dTOA_{jk} \equiv toa_j - toa_k = \frac{\sqrt{(x-x_j)^2+(y-y_j)^2}}{c} - \frac{\sqrt{(x-x_k)^2+(y-y_k)^2}}{c} \qquad (1)$$

The above equation was used to generate the hyperbolic segments shown in Figure 6 for both $dTOA_{21}$ and $dTOA_{23}$.

The following subclause details a method of solving for the intersection of these two hyperbolas. Equation pair will produce two possible locations for this intersection, one real and the other (almost always) extraneous. An extraneous solution will not satisfy the original pair of *dTOA* equations and may be recognized by this failure.

4.4.2.3 Detailed equation derivation

For convenience, define $\delta_{jk} \equiv c \cdot dTOA_{jk}$. By squaring Equation (1) and collecting terms, one obtains the relationship shown in Equation (2):

$$\frac{1}{2}\{\delta_{jk}^2 + (x_j^2 - y_k^2) - (x_k^2 - y_j^2)\} + (x_k - x_j)\cdot x + (y_k - y_j)\cdot y = \delta_{jk}\cdot\sqrt{(x-x_j)^2+(y-y_j)^2} \qquad (2)$$

Again, to simplify notation define $\lambda_{jk} \equiv \frac{1}{2}\{\delta_{jk}^2 + (x_j^2 - y_k^2) - (x_k^2 - y_j^2)\}$, then readers 1, 2, and 3 give the hyperbolic relationships shown in Equation (3) and Equation (4):

$$\lambda_{21} + (x_1 - x_2)\cdot x + (y_1 - y_2)\cdot y = \delta_{21}\cdot\sqrt{(x-x_2)^2+(y-y_2)^2} \qquad (3)$$

$$\lambda_{23} + (x_3 - x_2)\cdot x + (y_3 - y_2)\cdot y = \delta_{23}\cdot\sqrt{(x-x_2)^2+(y-y_2)^2} \qquad (4)$$

Subtracting these two shows that the location solution must lie on a line given by Equation (5):

$$\{\delta_{23}\cdot(y_1 - y_2) - \delta_{21}\cdot(y_3 - y_2)\}\cdot y + \{\delta_{23}\cdot(x_1 - x_2) - \delta_{21}\cdot(x_3 - x_2)\}\cdot y + (\delta_{23}\cdot\lambda_{21} - \delta_{21}\cdot\lambda_{23}) = 0(5)$$

Or equivalently $y = m \cdot x + b$ where, as shown in Equation (6) and Equation (7):

$$m = -\frac{\delta_{23} \cdot (x_1 - x_2) - \delta_{21} \cdot (x_3 - x_2)}{\delta_{23} \cdot (y_1 - y_2) - \delta_{21} \cdot (y_3 - y_2)} \tag{6}$$

$$b = -\frac{\delta_{23} \cdot \lambda_{21} - \delta_{21} \cdot \lambda_{23}}{\delta_{23} \cdot (y_1 - y_2) - \delta_{21} \cdot (y_3 - y_2)} \tag{7}$$

All that remains is to find the intersection of this line with either of the previously identified hyperbolas. Substituting $y = m \cdot x + b$ into hyperbolic relationship Equation (3) and Equation (4), squaring and collecting terms gives the quadratic equation $A \cdot x^2 + B \cdot x + C = 0$ where, as shown in Equation (8), Equation (9), and Equation (10):

$$A = (x_1 - x_2)^2 + 2 \cdot m \cdot (x_1 - x_2) \cdot (y_1 - y_2) + m^2 \cdot (y_1 - y_2)^2 - \delta_{21}{}^2 \cdot (1 + m^2) \tag{8}$$

$$\begin{aligned} B = {} & 2 \cdot \{\lambda_{21} \cdot (x_1 - x_2) + \lambda_{21} \cdot m \cdot (y_1 - y_2) + b \cdot (x_1 - x_2) \cdot (y_1 - y_2) + m \cdot b \cdot (y_1 - y_2)^2 + \\ & \delta_{21}{}^2 \cdot (x_2 - m \cdot b + m \cdot y_2)\} \end{aligned} \tag{9}$$

$$C = \lambda_{21}{}^2 + 2 \cdot \lambda_{21} \cdot b \cdot (y_1 - y_2) + b^2 \cdot (y_1 - y_2)^2 - \delta_{21}{}^2 \cdot (x_2^2 + (b - y_2)^2) \tag{10}$$

Two possible solutions for the tag location are given by the quadratic formula along with the linear equation shown in Equation (11) and Equation (12):

$$x = \frac{-B \pm \sqrt{B^2 - 4 \cdot A \cdot C}}{2 \cdot A} \tag{11}$$

$$y = m \cdot x + b \tag{12}$$

The second solution to the pair of equations [Equation (11) and Equation (12)] is (almost always) extraneous, introduced by the squaring operation. It can be recognized by the fact that it will not match the observed time differences.

4.4.3 Locating an object through angle of arrival

4.4.3.1 Overview of triangulation

A transmitter (tag) sends a signal that is received by readers in (at least) two different locations. Each reader notes the signal's angle-of-arrival (AOA), in azimuth, elevation or both. The tag resides at the intersection of the angle vectors measured at two readers, as shown in Figure 7:

If both azimuth and elevation are measured at each reader (Figure 7 shows only one measurement being made at each reader) then it is possible to find the 3D location of the tag using information from only two readers.

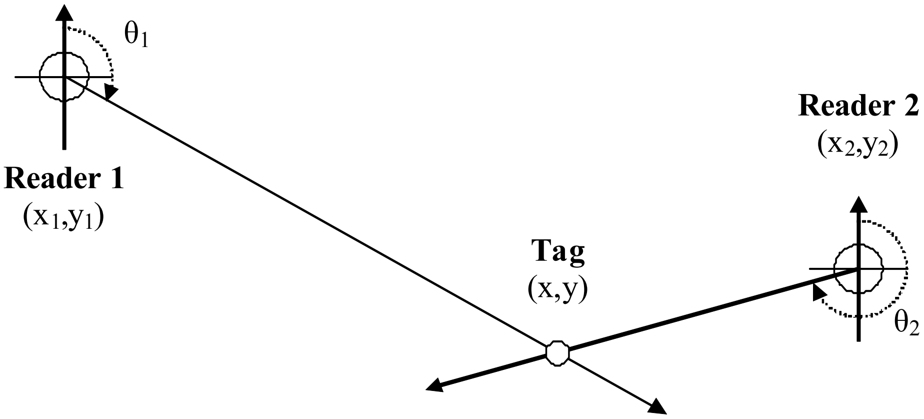

Figure 7—Locating a tag through triangulation

4.4.3.2 Mathematical location solution

In Figure 7, the tag position (x,y) can be calculated using Equation (13) and Equation (14):

$$d_1 \text{ (distance from Reader 1 to Tag)} = \text{sqrt}((x - x_1)^2 + (y - y_1)^2) \tag{13}$$

$$d_2 \text{ (distance from Reader 2 to Tag)} = \text{sqrt}((x - x_2)^2 + (y - y_2)^2) \tag{14}$$

Then:

$$x = x_1 + d_1 \times \sin(\theta_1)$$
$$y = y_1 + d_1 \times \cos(\theta_1)$$

And also:

$$x = x_2 + d_2 \times \sin(\theta_2)$$
$$y = y_2 + d_2 \times \cos(\theta_2)$$

These equations can be solved to determine the two unknowns (x,y) using the two measurements θ_1 and θ_2.

5. Features to assist MBAN devices

5.1 Coordinator switching

There are several MBAN usage scenarios where the devices that are associated with one coordinator may need to be moved to another coordinator. For example, a patient with an array of sensors is in transportation from an operating room to a recovery room and monitored by a portable monitor to provide continuous care services. The patient is then moved back to his/her patient room in which a bedside monitor is available. Switching the sensor devices from the portable monitor to the bedside monitor would provide better data processing and display and save the battery life of the portable device. It is important that an orderly switching of the devices from the coordinator in the portable monitor to the desired coordinator in the bedside monitor takes place. Since the switching is instigated and controlled by a clinician and the procedure to enable this resides in the higher layer of the coordinators, a detailed description of how this is done will not be described here, though some of the steps that could be used are noted here.

First, the coordinator sends a message to the associated devices informing the devices of the identity of the coordinator with which they should associate after they have been disassociated from their existing coordinator. This identity information is supplied to the coordinator by the clinician. On receipt of this information the devices should then cease further data transmission and await notification from the coordinator that they have been disassociated. Upon disassociation, the devices then commence a scanning procedure to find the new coordinator whose identity is being broadcast within the beacon to determine the channel number that the coordinator is using, its PAN ID and MAC extended address. Once this information is obtained the devices can begin association and then communication between the devices and the new coordinator can commence. This description does not necessarily include all of the steps that need to be taken and it only gives an indication of what may be required.

5.2 Channel bitmap

To assist with coexistence to the primary users of the 2360 MHz – 2390 MHz band, MBAN devices may be excluded from using some of the channels in this band (Channels 0 – 5 and 7 – 12). The channels that are excluded may change over time. Regulations require that the operation of MBAN hub devices have access to a mechanism that provides information on the portions of the band that cannot be used. The description of this mechanism is beyond the scope of this standard. To assist other MBAN devices, such as those connected to sensors, that do not have direct access to this mechanism, a bitmap may be used by a hub device to indicate the channels that are allowed to be used by the MBAN devices.

The bitmap can be conveyed as an Encapsulated Service Data Unit (ESDU) Information Element (IE) that can be used to encapsulate a higher layer payload. The definition of the format of this payload is beyond of the scope of this standard.

6. Low energy, critical infrastructure monitoring systems

6.1 Introduction

Globally there are many definitions of critical infrastructure. For example, as per Public Law 107–56 [B1], the term “critical infrastructure” means systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters. Most commonly associated with the term are facilities for:

- Electricity generation, transmission, and distribution
- Gas production, transport, and distribution
- Oil and oil products production, transport, and distribution
- Telecommunication
- Water supply (e.g., drinking water, waste water/sewage, stemming of surface water [e.g., dikes and sluices])
- Agriculture, food production, and distribution
- Heating (e.g., natural gas, fuel oil, district heating)
- Public health (e.g., hospitals, ambulances)
- Transportation systems (e.g., fuel supply, railway network, airports, harbors, inland shipping)
- Financial services (e.g., banking, clearing)
- Security services (e.g., police, military)

6.1.1 LECIM characteristics

The LECIM portions of this standard form the MAC and PHY behaviors that implement a minimal network infrastructure, enables the collection of scheduled and event data from a large number of non-mains powered end points that are widely dispersed, or are in challenging propagation environments. To facilitate low energy operation necessary for multi-year battery life, MAC protocols minimize network maintenance traffic and device wake durations. In addition, LECIM addresses the changing propagation and interference environments encountered over many years.

The following is a list of LECIM characteristics and the underlying behaviors that form them:

a) Minimal infrastructure
  - Star topology, i.e., PAN coordinator communicates directly with devices.
  - Mains energy supply only for the PAN coordinator.

a) Commissioned network (not ad hoc)
  - PAN coordinators and devices are configured specifically for the deployed network.
  - PAN coordinators and devices are stateful, i.e., they are preconfigured with parameters that eliminate the need for wireless messages sending configuration information.

b) Long range
  - High receiver sensitivity, resulting from methods such as narrow bandwidth or high processing gain.
  - Interference robustness.
  - Challenging environments and widely dispersed devices.

c) Very limited energy supplied devices, such as:

  - Ten to twenty year battery life with no maintenance, i.e., original battery supplies all energy for the life of the device.
  - Energy harvesting with limited power supplies, i.e., short and infrequent transmission and reception durations.

d) Significant difference between PAN coordinator and devices

  - Some PAN coordinators have significantly higher performance and a larger energy supply than the devices.
  - Does not preclude distributed systems.

e) Asymmetrical data flows

  - Sensor end point: up-link dominates data flow with limited down-link data needs.
  - Actuator end point: down-link dominates data flow with limited up-link data needs.

The following MAC enhancements are included to support the LECIM PHYs defined in IEEE Std 802.15.4:

- Enhanced timing and synchronization capabilities to support synchronous and asynchronous channel access in both beacon-enabled and nonbeacon-enabled operation
- Enhanced low energy mechanisms
- MPDU fragmentation to support extremely low data rates and limited PSDU sizes
- PCA
- MLME-SAP and PIB extensions for PHY control and configuration

6.1.2 Use case examples

The following use cases exemplify LECIM applications.

6.1.2.1 Oil and gas pipeline monitoring

The key drivers of pipeline monitoring are as follows:

- Environmental protection
- Reliability (critical resources)
- Cost savings (increasing cost)
- Compliance (regulators)

6.1.2.2 Water leak detection

The key drivers of water leak detection are as follows:

- Permanent installation of large number of sensors underground
- Long range and ability to penetrate underground vaults
- Battery operated and long lifetime
  - Small data messages once per day and in case of alarm event (e.g., leak detected)
- Low installation cost (easy deployment) and low cost of maintenance

6.1.2.3 Soil monitoring

The key drivers of soil monitoring are as follows:

- Power consumption

  - Low-cost batteries that last over many years
- Networking
  - Long range links to cover large fields
  - Ability to use mesh or tree networking for complicated environment
  - Ability to connecting WPAN with mobile networks
- Reliability and cost
  - Very low maintenance requirements

6.1.2.4 Inventory control - event driven with query

The application is for a warehouse floor with thousands of parts bins. Each bin has a battery operated RF link for communicating current quantity and changes in quantity to the central inventory control (CIC) system. Battery life is important.

Each bin contains only one part number. The RF link has an LCD display showing the quantity in the bin. It also has an "Increase Button" and a "Decrease Button." When an operator adds units to the bin, he presses the Increase Button, and when parts are removed, he presses the Decrease Button. Each time a button is pressed, it generates an event to the RF module, which then transmits the change to the CIC. This would most likely use a contention access method for transmission, since events occur in an unscheduled manner.

The CIC receives events from all of the bins, as changes are made to the quantity contained in each bin. Both the local RF module and the CIC maintain the quantity in the bin.

For inventory auditing, it is necessary for the CIC to query each bin to check the quantity. This requires the CIC to initiate a transaction with each bin, either individually or as a broadcast/multicast message. The desire is to have all bins report within a reasonable time (minutes).

Also, since changes in quantity are event driven, the CIC needs a means to query each bin to make sure that it is still operational and that no "change in quantity" events were missed.

To minimize battery drain, the LECIM device is only activated when necessary:

- A change in quantity as indicated by a button event
- Some type of synchronous sniff/query operation for receiving queries from the CIC
- A response to query messages

6.1.2.5 Building monitoring - time and event driven data with query

A building (or any structure) is being monitored by sensors that report measurement or state information over long periods, e.g., several minutes to several hours. There are also be sensors that report events or changes in state that are event driven and not time driven. Battery life is important.

Each measurement sensor is set to report its information at a certain interval, using either a GTS or the CAP. This gives very low duty cycle for normal operation, which is 99% of the usage. There are also be sensors that are event driven and report change in state, such as door open/closed, door locked/unlocked, switch on/off, etc. This is also low duty cycle.

Occasionally there is an event, such as an emergency, where the central monitoring system requires readings from all sensors as soon as possible. The central controller sends a request to all sensors to report their current measurement or state. This requires a low latency response mechanism capable of maintaining long battery life.

6.1.3 LECIM behaviors

The following assumptions and precepts are essential to address the needs of LECIM applications:

- Commissioning
- Low energy
- Coverage extension

6.1.3.1 Commissioning

Commissioning by a professional installer allows the network to reduce the amount of data to be sent by creating statefulness. The commissioning parameters are not expected to change over the duration of the network.

The following is one, simple example of commissioning to enable communication among three LECIM DSSS devices (devices *A*, *B*, and *C*), where device *A* is a powered coordinator and devices *B* and *C* are end devices.

PHY PIB attributes (commissioned):

Modulation-related PIB attributes

- *phyLECIMDSSSPPDUModulation* = O-QPSK
- *phyLECIMDSSSPPDUModulationRate* = 1000

SHR-related PIB attributes

- *phyLECIMDSSSPreambleSize* = 16
- *phyLECIMDSSSSHRGoldCodeResetPerSymbol* = TRUE
- *phyLECIMDSSSSHRGoldCodeSeed*
- *phyLECIMDSSSSHRSpreadingFactor*

PSDU-related PIB attributes

- *phyLECIMDSSSPSDUSize* = 32
- *phyLECIMDSSSPSDUGoldCodeResetPerSymbol* = FALSE
- *phyLECIMDSSSPSDUGoldCodeSeed*
- phyLECIMDSSSPSDUSpreadingFactor

The PIB attribute *phyCurrentChannel*, the Gold code seed value, and the spreading factor value need to be shared among communicating parties but do not necessarily need to be the same for each direction in a link. For example, if three devices (A,B,C) make up a star network where device A acts as the PAN coordinator, device A uses one set of code/spreading for its beacon transmission and use two additional, unique codes/ spreading for reception from devices B and C. This ability to use different codes, or logical channels, provides a mechanism to address hidden node problems and interference from other co-located networks. For example, two nodes hidden from each other can have overlapping transmissions that can successfully be decoded by a receiver, provided the receiver has the ability to simultaneously demodulate two or more different sets of PN sequences at the same time (e.g., additional processing resources, CDMA). This ability eliminates the need for the end devices to re-transmit hidden node collisions, thus saving energy and improving system capacity.

For the purposes of this example, the frequency band is set to 902 MHz. When the transmitter is device A, and the receivers are devices B and C, the values of *phyCurrentChannel*, the Gold code seed, and the spreading factor are (5, 0x0123, 7). When the transmitter is device B and the receiver is device A, the set of values is (5, 0x0789, 7). When the transmitter is device C and the receiver is device A, the set of values is (5, 0x0def, 7).

6.1.3.2 Low energy

LECIM applications require significantly low energy operation, in order to be able to either last 20 years on the original battery supply or on energy harvesting mechanisms. Achieving low energy operation is made very difficult given the low data rates necessary for long range operation. Accordingly, LECIM networks should be capable of eliding any overhead octets not absolutely necessary in order to minimize transmit and receive durations, schedule link times to minimize device "on" durations, and maximize link reliability to minimize retransmissions.

The maximum PPDU size that a LECIM PHY is able to receive is 2047 octets. To facilitate the multi-year battery life operation expected for envisaged LECIM applications, and given the low over the air (OTA) chip rate provisioned by the PHY in terms of allowable spreading factor (32,768 chips per symbol for the DSSS PHY and 16 chips per symbol in the case of the FSK PHY), it is recommended that small PPDU sizes be used.

As an example, the on-air time for an unfragmented 2047 octet frame at a symbol rate of 12.5 kb/s will require a minimum of 1.3 seconds total air time to be transmitted or received. If that same frame has a spreading factor of eight applied to each symbol, the OTA broadcast time will exceed 10 seconds. This increase in transmission and reception time will have a significant impact upon the battery longevity of a LECIM device, even allowing for low duty cycle operation. It should also be noted that the increased transmission time can lead to regulatory duty cycle limitations, especially in the case of the PAN coordinator.

6.1.3.3 Coverage extension

To keep infrastructure costs to a minimum, LECIM devices have large link margins to achieve long ranges without requiring mesh devices or repeater devices. Requiring a mesh topology would increase the number of devices needed to sustain the network and, in most cases, require mains power for these devices. To extend the coverage for supporting sparse dispersed devices beyond the link margin or to maintain connections in dramatically changing environments, optional frame relaying repeaters located between the concentrator and devices are included in this standard to sustain the connections without reconfiguring the whole LECIM network.

6.1.3.4 Device sensitivity and interference robustness

To support long range operations, the LECIM device is intended to have high receiver sensitivity and interference robustness. The high receiver sensitivity (capable of supporting 120 dB path losses) is achieved via low data rates, FEC, high processing gains, and other such mechanisms. The interference rejection specification for the LECIM device needs to support the criticality aspect of infrastructure monitoring such as co-channel rejection, improved receiver interference rejection, and blocking immunity. Attention is drawn to the class two receiver requirements of European standard ETSI EN 300 220-1 [B2] as an example of typical receiver immunity requirements for a LECIM device.

6.2 Functionality added

The following functions have been added to this standard in order to implement LECIM applications: DSSS, FSK, fragmentation, frame priority, TRLE, PIB attributes, and IEs.

6.2.1 DSSS

The DSSS devices used by LECIM networks differ from the other DSSS devices defined in this standard in that they have significant processing gain to allow devices to receive messages with very low or negative carrier-to-noise ratios. High processing gain also allows for CDMA operation to reduce the possibility of collisions.

With high spreading factors and CDMA, transmitted signals from other devices within radio range of the receiver that are operating on different codes are likely to be undetectable, because they fall below the effective noise floor. This should be taken into account when configuring the CCA mode to be used; however, in many applications, sensing of the medium is not practical or useful, so selecting the CCA mode for use with the CSMA algorithm that equates to CCA Mode 4 (ALOHA) is recommended.

There are a great many options available in LECIM DSSS PHY that provide the ability to best address the applications throughout a diverse and changing set of regulatory environments. Some options are not be valid in some regulatory regimes, and it is up to the OEM and/or higher layers to specify options which comply with local regulations.

For example, under the current FCC regulations it would be perfectly legal to use a *phyLECIMDSSSPPDUModulationRate* of 400 (kHz), with certain restrictions. Specifically, the device would be required to use frequency hopping and would need to limit transmission to ≤ 400 ms. A *phyLECIMDSSSPSDUSize* of 32 octets, after FEC, yields a minimum of 512 modulation symbols per fragment. Using BPSK modulation, at a spreading factor of 2^8 (256), the fragment duration would be 328 ms.

Higher spreading factors would not be allowed under FCC rules. Under other regulatory domains at this modulation rate, frequency hopping is not required and the maximum duration (and spreading factor) is not be limited to 256 (≤ 400 ms).

6.2.2 FSK

The LECIM FSK PHY uses a transmit signal characterized by a constant envelope, which allows for low cost implementation and good transmit power efficiency.

LECIM FSK devices are typically narrow bandwidth (hence low data rate) to permit higher sensitivity and an increased number of channels in each band, which can reduce the probability of packet collision.

Features, such as forward error correction, with a relatively high constraint length and robust interleaving, as well as spreading capability, are included to allow for further sensitivity gains.

6.2.3 Fragmentation

With the addition of very low data rate PHY operating modes, the resulting increase in the over-the-air duration of a MAC frame can lead to increased interference potential, susceptibility to channel conditions changing during the duration of a MAC frame transmission, and other effects that can reduce reliable transfer in some environments typical of LECIM applications. The long packet duration also brings a large cost for retransmission, both in terms of energy consumed and interference footprint.

MPDU fragmentation can improve the probability of successful transmission and reduce the cost of retransmission. With fragmentation, each fragment is packaged into a PPDU for transmission, and this smaller PPDU has a reduced interference footprint. Also, retransmissions can be performed on a per fragment basis without needing to retransmit the entire original packet.

The Simplified Superframe Specification IE is intended to reduce the amount of communications between the PAN coordinator and end devices. This allows an enhanced beacon with the Simplified Superframe Specification IE to be transmitted in a single PPDU when the LECIM DSSS PHY is in use.

MPDU fragmentation operates on the MHR and MSDU portions of the MPDU transparent to other MAC MPDU processing. MPDU fragmentation is specified so that it can be used with any of the PHYs defined in this standard. Optimum use of fragmentation depends on many variables, including channel performance, the interference environment, and characteristics of the PHY selected (e.g., data rate and maximum PSDU size supported). An overview of the fragmentation process is shown in Figure 8.

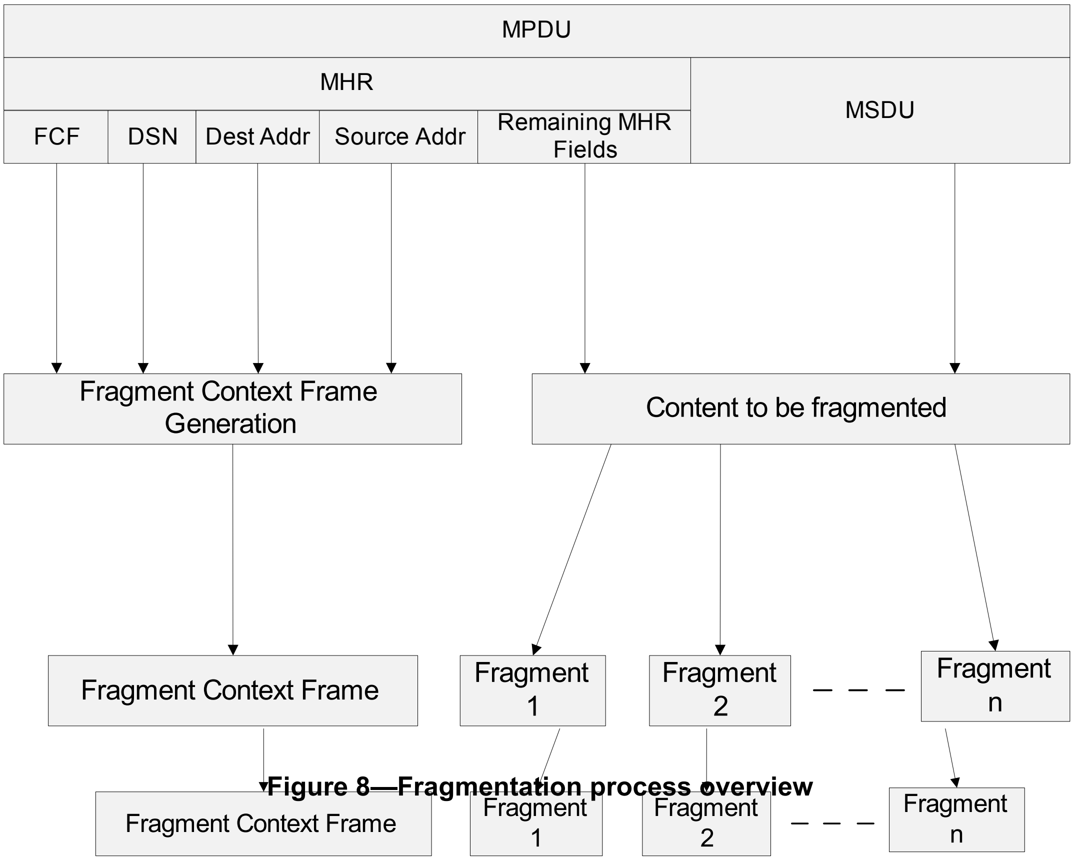

Figure 8—Fragmentation process overview

When the LECIM DSSS PHY is being used, the PSDU size varies from 16 to 32 octets. A typical MPDU can be substantially larger than the available PSDU size. Fragmentation enables all the MAC frame formats defined by this standard to be carried in the constrained PSDU size. Fragmentation operates on the complete MPDU and adapts it to the characteristics of the specific PHY and PHY operating mode. The fragment size is determined by the value of *phyLECIMDSSSPSDUSize*. The following example illustrates how MPDU fragmentation is applied to the LECIM DSSS PHY.

The LECIM DSSS PHY allows for a unique code to be assigned to a link between the end-point and the PAN coordinator. For this example, consider a unique code that is assigned for the down-link between the PAN coordinator and the device. Down-link addressing information is implied, and the probability of a TID collision is nil. A context setup frame in this example, shown in Figure 9, would be a multipurpose frame with no DSN, no addressing, no auxiliary security header, and exactly one MPDU Fragment Sequence Context Description IE.

Octets: 2	6	4
FCF	MPDU Fragment Sequence Context Description IE	FCS
MHR	MAC Payload	MFR

Figure 9—Context setup frame when using LECIM DSSS PHY

The fragment size is

$$phyLECIMDSSSPSDUSize - 2 - V$$

where *V* is 2 when *macFragmentCVSType* is set to 16, and *V* is 4 when *macFragmentCVSType* is set to 32.

6.2.4 Frame priority

Frame priority allows LECIM networks to exhibit low latencies for truly critical event messages versus those latencies for link maintenance or other lower priority messages.

Frame priority is established by two means: PCA allocations and the PCA backoff algorithm for CSMA-CA and for CCA Mode 4 (ALOHA). Both algorithms are used during contention access, but the PCA allocations can only be used when operating in beacon-enabled mode. The PCA is only usable for critical event messages, but the critical event messages do have to compete with each other for access to the channel.

The PCA backoff algorithm is used whenever contention access is applied. It operates slightly differently based on whether CCA Mode 4 (ALOHA) is used or not. When CCA Mode 4 (ALOHA) is not used, the transmitting device remains in persistent mode.

6.2.5 TRLE

For extending the range of a link in a star network composed of beacon-enabled devices or DSME-enabled devices, the TRLE PAN relays residing between the PAN coordinator and devices support transparent link connectivity without additional networking overhead to an end device.

The TRLE PAN relay operates with the frame filtering in relaying mode and relays MAC frames either in the direction of the PAN coordinator or in the direction of a device. The TRLE PAN relay provides a one-hop relaying link extension for the beacon-enabled PAN. The TRLE-enabled PAN coordinator and the TRLE PAN relays provide multi-hop relaying link extension for the DSME-enabled PAN.

6.2.6 PIB attributes

LECIM mechanisms and protocols in this standard require the following additional PIB attributes.

MAC PIB attributes

- *macLECIMAlohaUnitBackoffPeriod*
- macLECIMAlohaBE
- macMPDUFragPadValue
- macFragmentFVSType

- macFVSoffset
- macFVSRIV
- macRelayingMode
- macTRLEenabled

PCA-specific PIB attributes

- macPriorityChannelAccess
- macPCAAllocationSuperRate
- macPCAAllocationRate
- macCritMsgDelayTol

DSME specific MAC PIB attributes

- *macAllocationOrder*
- macBIIndex
- macExtendedDSMEcapable
- macExtendedDSMEenabled

LE-specific MAC PIB attributes

- *macIRITOffsetInterval*
- *macIRITListenDuration*
- macIRITEnabled

6.2.7 IEs

LECIM mechanisms and protocols in this standard require the following IEs:

- DSME PAN Descriptor IE
- PHY Parameter Change IE
- MPDU Fragment Sequence Context IE
- LECIM Capabilities IE
- DSSS Operating Mode IE
- FSK Operating Mode IE
- TRLE Descriptor IE

6.2.8 PHY parameter changes for fragment sequence exchange

Given the potentially long duration of the MPDU transaction in time, there is a possibility that channel conditions will change significantly. If the higher layer determines that the channel is becoming unusable and it can change to another channel for subsequent transaction.

The PHY Parameter Change IE in included in a directed frame, which facilitates changing PHY operating parameters between the specific sender and receiver.

If a device determines that a switch in band, channel, or other PHY operating parameter is necessary during a fragmentation sequence transaction, the device should terminate the transaction context. Following the termination, a MAC frame with the PHY Parameter Change IE is sent by the device initiating the change. The originator switches to the new PHY parameters upon reception of an acknowledgment.

The higher layer network management entity controls which channel and/or PHY configurations are used to communicate with which neighboring devices; the process by which this is done is outside the scope of this standard.

7. Ranging and location topics

7.1 Ranging and location for UWB

7.1.1 Overview

Ranging capability is achieved through support of a number of specific PHY capabilities as well as defined MAC behaviors and protocols.

7.1.1.1 Two-way ranging

UWB devices that have implemented ranging support are called ranging-capable devices (RDEVs). UWB PHYs have a bit in the PHY header (PHR) called the ranging bit that serves to signal to the receiver that this particular frame is intended for ranging. A UWB frame with the ranging bit set in the PHR is called a ranging frame (RFRAME). There is nothing else (beyond the ranging bit set in the PHR) that makes an RFRAME unique. RFRAMEs can carry data, RFRAMEs can be acknowledgments, and RFRAMEs do not even (for the case of one-way ranging) necessarily require an acknowledgment. As far as ranging is concerned, the critical instant in a frame is the first pulse of the PHR. The first pulse of the PHR is the ranging marker (RMARKER). This standard is primarily structured to support the two-way time-of-flight computation of distance between two RDEVs. Figure 10 illustrates the complete sequence for two-way ranging.

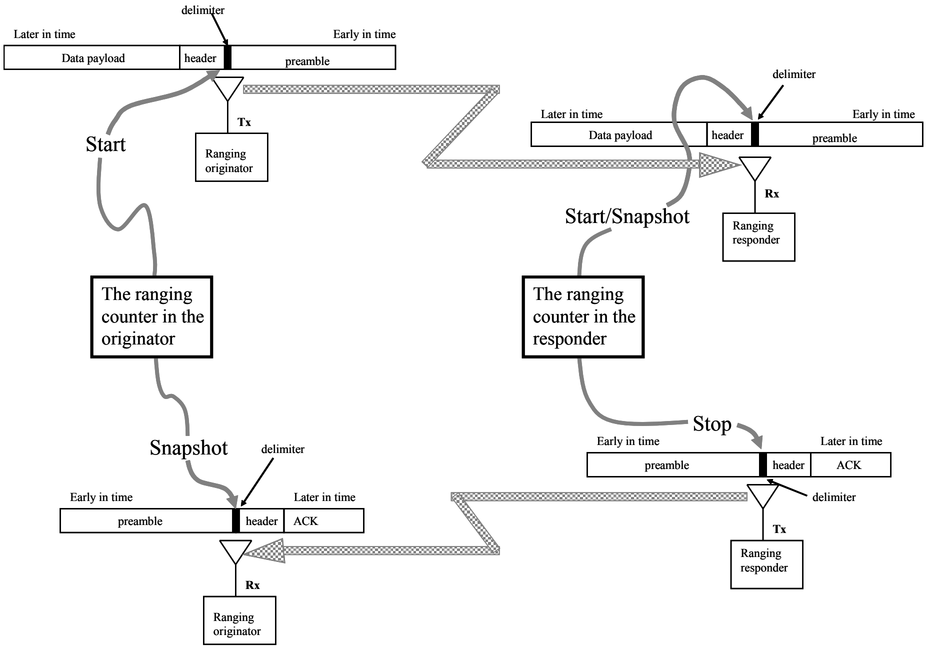

Figure 10—The complete two-way ranging technique

The first frame of the two-way exchange is an RFRAME that is sent from the originating device to the responding device, as shown in the top half of Figure 10. A ranging counter start value is captured in the originator device upon the RMARKER departure from the originator, and a ranging counter start value is captured in the responding device upon RMARKER arrival at the responder. The RFRAME has the acknowledge request bit set in the MHR. In the most general case, the counter in the responder PHY will have already started running when a previous RFRAME arrived, but the previous RFRAME was not intended for this device and thus did not get an acknowledgment from this device. In Figure 10, the counter activity is labeled “start/snapshot” from the PHY perspective. For the PHY, the counter function is “start” for the first arriving frame and “snapshot” for subsequent frames. Snapshot means that the value of the counter is captured and stored at the instant of the snapshot, but the counter continues to count as if the snapshot had not happened. The responder PHY initiates the counter snapshot values for all arriving RFRAMEs. The responder MAC discards the snapshot values that are for RFRAMEs not intended for the responder device. At the end of the first frame transmission in Figure 10, the counters are running in both devices.

The RFRAME shown in the bottom half of Figure 10 is an acknowledgment sent from the responding device to the originating device. The ranging counter stop value is a snapshot in the responding device upon RMARKER departure from the responder. The responder PHY is now in transmit mode, and the counter is still running. Because the PHY is in transmit mode, it will not be receiving any frames or taking any counter snapshots. Leaving the counter running in the responder at this point in the algorithmic flow only serves to deplete the battery of a mobile device. In Figure 10, the counter action is labeled stop, not because it really is stopped (it is not), but because the algorithmic flow is done with it and because it will appear to the application as if it has stopped because it will generate no more snapshots. The counter stop value in the originator device is snapshot and saved upon RMARKER arrival at the originator. The originator MAC verifies that the frame was from the responder, and ultimately the application will then stop the counter using MLME primitives.

When the application in the responding device learned that the acknowledgment had been sent, it stopped the ranging counter in the PHY. When the application at the originator device discovered that the acknowledgment frame was for the originator device, it stopped the counter in the PHY. Thus, in Figure 10, all the counters have stopped, and the values are located in the respective devices.

The discussion in this subclause risks confusion because it includes the general case of arriving frames not intended for the devices. While that behavior is important for algorithmic robustness, for understanding basic ranging, it is a distraction. In Figure 10, the ranging pair holds two different sets of counter values, with a start value and a stop value in each set. Even if some counter snapshots have been discarded by the application due to frames destined for other devices, what remains at the end of the exchange are pairs of counter values that when subtracted represent the elapsed times between the arrival and the departure of the intended frames.

At the system state represented by Figure 10, the necessary information required to compute the range between the devices is known. However, the information is still distributed in the system; and before the ranging computation can be accomplished, the data are brought to a common compute node. The difference of the counter start and stop values in the originator device represents the total elapsed time from the departure of the first message to the arrival of the acknowledgment. The difference of the counter start and stop values in the responder represents the total elapsed time from arrival of the data message to the departure of the acknowledgment. After these values are all brought together at a common compute node, they are subtracted, the difference is divided by two, and the time of flight (and thus the inferred range) is known.

While management of the ultimate disposition of the counter values is outside the scope of this standard, the most immediately obvious resolution is for the application at the responder device to send the counter value to the originator device in a data frame. The originator device started the exchange; therefore, it might be assumed that it is the point at which the application desires to have the range information. The responder device was just in radio contact with the originator device; therefore, a communication channel is very likely to be available.

The obvious solution suggested in the previous paragraph is often the wrong solution. This standard supports applications that bring a large number of ranging measurements together at a single computation device, and there the application solves not just for the ranges between individual devices but also the two- or three-dimensional relative location of the devices. A discussion of typical activities at a central compute node is included in 7.

7.1.1.2 Position awareness through one-way transmissions

The primary intent of the ranging support in this standard is to support ranging through two-way time-of-flight measurements. To establish that capability, this standard defines a whole suite of capabilities and behaviors by which two-way ranging is enabled. The capabilities required to accomplish one-way ranging are sufficiently similar so that this standard allows operation in that mode as well. One-way ranging requires an infrastructure of RDEVs and some means to establish a common notion of time across those devices. The protocol to establish the common notion of time is outside the scope of this standard. What this standard does provide is a bit in the PHR that serves to signal an RDEV that location awareness is desired. To operate in a one-way environment, any UWB device (not necessarily an RDEV) simply sends an RFRAME having an appropriate preamble length. As described in 7.1.1.1, after the PHY counter is running in an infrastructure RDEV, the MAC initiates an MCPS-DATA.indication primitive to the next higher layer for all arriving RFRAMEs. By this mechanism, the application acquires a list of counter values representing the arrival times of all one-way ranging messages received at all infrastructure devices. As with the case of two-way ranging, nothing useful can happen with the lists of counter values until they are brought together at a common compute node; and as with two-way ranging, that bringing together activity is beyond the scope of this standard. After the counter values have been brought together, the computation of the relative locations proceeds as shown in 7.1.5.2.

The PHY will initiate an MCPS-DATA.indication for all arriving RFRAMEs. That can be thought of as a great goal; and if the RFRAMEs arrive infrequently, it is a very achievable goal. However, the PHY does not get to choose how quickly a sequence of RFRAMEs might arrive; and in real-world applications, it is possible that the RFRAMEs arrive more quickly than the PHY can deal with them. The processing time for an RFRAME is very dependent on the implementation of the PHY. While the PHY has no control over the inter-arrival time of RFRAMEs, the application very well might. The application can discover the RFRAME processing time by reading the PHY PAN information base (PIB) attribute *phyRFRAMEProcessingTime*. The intended use of this attribute is that if an application discovers the processing capabilities of the devices in the network, it can structure the traffic so that devices are not overrun. Because the only purpose is to help prevent overruns, there is no need for high precision in expressing this value.

7.1.1.3 The ranging counter

At the most fundamental level, ranging capability as described in this standard is enabled by the ranging counter shown in the center boxes of Figure 10. The ranging counter has the capability of assigning values to the precise instant that RMARKERS are transmitted and received from the device. Once that counter and the ability for it to precisely snapshot a timestamp are in place, then conceptually, the computation of the time of flight is simple.

An actual physical counter that exhibits the behavior attributed to the "ranging counter" does not need to exist in any PHY implementation. The ranging counter described is simply an abstraction that is used to specify the required PHY behavior.

The LSB of the ranging counter represents a time interval so small that an actual physical counter would have to run at a nominal 64 GHz to produce values with the required resolution. It is unlikely that a device intended for low-cost, battery-powered operation would implement a counter running at 64 GHz.

The lack of an actual physical realization does not in any way preclude the use of the ranging counter as an abstraction in this standard to visualize and specify the behavior of an RDEV. From an algorithmic and computational viewpoint, the RDEV will appear to an application as if it possesses a 64 GHz counter with the ability to start and stop based on the state of bits before they arrive.

The implementation of the ranging counter is beyond the scope of this standard; however, the following ideas are suggested:

- Take snapshots of the counter at every event when counter value might be required and then discard unneeded snapshots after the future becomes known.
- Do not build a 64 GHz counter, but rather generate the less significant bits of the ranging counter values using computational techniques like those described in 7.1.2.

7.1.1.4 Accounting for signal arrival time

The start counter values represent the time of arrival of the first pulse of the first symbol of the header of a PPDU. That necessary task is not trivial when the signal arrives in a channel with significant multipath. The first pulse of the header arrives at the antenna once for every multipath reflection in the environment. The result is that in an indoor environment, the “first pulse” can seem to arrive at a multitude of different times. Accurate ranging requires discriminating the leading edge of the cluster of signal arrivals that accounts for the first pulse of the header. The technique for achieving this discrimination is outside of the scope of this standard, but it is helpful to discuss a typical approach. In a typical approach, the counter value is a snapshot relative to a position on the arriving waveform where the PHY-tracking loop has achieved and is maintaining a “signal lock spot.” Then the offset from the lock spot to the leading edge of the pulse energy is determined. After a time offset from the UWB signal lock spot to the leading edge of the energy cluster is found, the rest is very straightforward: that offset is just subtracted from the counter value exactly as the other correction factors are. The time offset to the leading edge is discovered by sampling the energy ahead of the acquisition lock spot over multiple different offsets to discover the earliest point with discernible energy. To achieve the required precision, the sample values in the vicinity of the leading edge could be further refined using techniques like those shown in 7.1.2. For those computations (and other techniques, generally known as upsampling), it is critical that the noise in the samples be well suppressed. This standard supports this leading edge activity by allowing a very long acquisition preamble keeping the signal steady and data free for a protracted time.

The necessary characterization of the channel multipath response is generically called a channel sounding. The techniques that accomplish that task can be numerically intense. A system designer could accomplish that task in the PHY. However, a sounding mechanism involving the MLME primitives has been defined. This allows the PHY to present raw data to a higher layer should the PHY lack significant computational capability and the system designer wishes to employ numerically intensive channel sounding algorithms. The raw data can then be moved to whatever device has sufficient computational resources to support the desired algorithms. The primitives supplied are very simple and assume that both the MAC and the application are well behaved and give priority attention to sounding activities. For example, the primitives do not include tags to associate a particular sounding with a particular packet. The application is responsible for making sure that sounding activities are conducted in a timely way so that the sounding information is associated with the last packet received. The actual handling of the raw data after the sounding operation facilitates uploading to a next higher layer is beyond the scope of this standard.

7.1.1.4.1 Leading edge search during the acquisition preamble

Upon acquisition of the signal, the PHY is not aware of how much time remains in the preamble before the delimiter. A reasonable goal is to do the best possible job of bracketing the leading edge with whatever time is available and then reporting how well the leading edge was bracketed by way of the ranging FoM. In a typical implementation, if the delimiter arrives very quickly after the acquisition threshold was satisfied, then the leading edge equipment will still be using coarse steps to characterize the energy. The PHY will make the best judgment it can about the leading edge based on the coarse steps and then report a FoM value appropriate for coarse steps. If the leading edge search engine has ample time before the delimiter arrives, then not only can it have progressed to using a very fine search step, but it can also have integrated many samples to drive down the noise in the computation. In this case, the correction representing the leading edge offset is applied to the counter value (and it might have been the very same correction value as was applied in the previous case when the search time was short), but this time the FoM value is reported for a very small characterization bin and very high confidence that the leading edge truly was in that bin. Again note that the counter value returned with a good FoM can have the same value as the counter value returned with a bad FoM; i.e., the counter value is independent of the FoM.

7.1.1.4.2 FoM for bad times

If the PHY performs a short leading edge search (as will happen after recovering from an acquisition false alarm, for example), it still makes its best guess for a leading edge correction and goes on with the ranging algorithm. Even when the final counter value represents a known error-prone measurement, the PHY should not return a FoM of zero. Zero means “no FoM,” which is neither correct nor useful. An appropriate FoM to report for the most error-prone cases is 0x79. That value decodes to tell the application that even if the other RDEV calculated its half of the measurement perfectly, given the expected error just due to this RDEV’s measurement alone, the PHY is 80% confident that the computed range will be wrong by more than 2 m.

7.1.1.4.3 Other opportunities for leading edge search refinement

The previous discussion was framed as if all channel characterization had to stop upon the arrival of the delimiter. In fact, PHYs can do additional things after the delimiter to further refine the estimate of the leading edge offset. The UWB CCA pulses described in 802.15.4, if used, offer additional opportunities to look at the signal in a known state after the delimiter. A very sophisticated PHY can perform additional characterization during the time that data are on the air if the algorithm designer is willing to “back out” the effects of the data after demodulation (and thus after the data are “known”). From the application’s point of view, all the application sees for the difference between a very capable PHY and a sloppy, mediocre PHY is a difference in the FoM values being reported.

7.1.1.4.4 Managing the preamble length for leading edge search

One of the most distinguishing traits of ranging UWB radio transmissions is the long preambles. This standard allows the application to specify preambles that are either 1024 or 4096 symbol repetitions long. The selection is a function of the channel multipath, the signal-to-noise ratio (SNR) in the link, and the capability of the receiving PHY. It is theoretically possible that a very capable PHY that does leading edge refinement using the data could do ranging accurately with a preamble that is 16 symbols long. It is also possible (likely, in fact) that a PHY with a poorly designed search engine will not do a good job in a heavy multipath even with a 4096 symbol preamble. The upper layers are responsible for picking the preamble length. It is suggested that the application start ranging operations using the 1024 symbol preamble and keep a history of how the FoMs are reported. The FoMs are the critical feedback information that tells the application how the various PHYs are doing, and the application can make future adjustments to the preamble length based on that history.

7.1.1.4.5 PHY deferral of the computations for leading edge search

This standard provides a mechanism to allow the PHY to pass the computational burden of leading edge processing to a higher layer. If the computations are not done in the PHY, then the value in the timestamp report for RangingCounterStart is not corrected for the leading edge. The RangingFOM is used to signal this condition to the higher layer, which will have to compute a correction based on data acquired using the sounding primitives, described 802.15.4. The higher layer issues a MLME-SOUNDING.request primitive. The associated MLME-SOUNDING.confirm response returns a list SoundingPoints where each SoundingPoint is a pair of integers representing data taken by the PHY at time offsets from the point on the waveform represented by the uncorrected value in RangingCounterStart. A time of zero in the list designates an amplitude value taken at the point indicated by RangingCounterStart. Positive time values indicate amplitudes that occurred earlier in time than the zero point. The amplitude values do not represent any particular voltage. They are only meaningful in a relative sense and in the context of each other. The values are linear (not logarithmic). The amplitude values are all consistent with each other. For example, it is acceptable for an automatic gain control (AGC) circuit to change the gain during the measurement of the amplitudes, if the numbers are corrected so that the effect of the gain change is removed and the numbers returned in a SOUNDING.confirm primitive are the values that would have been measured had the gain been perfectly stable and unchanged for all measurements. The list of measurements returned by the SOUNDING.confirm primitive begins with the size of the list. The maximum size that can be represented is 65 K value pairs. That large value is only because a single octet would not be adequate to represent lists larger than 255 pairs. In practical systems, lists larger than 255 pairs can occur. Two octets would be the next choice to represent the list size, but that does not mean that lists approaching 65 K pairs would be appropriate. There is no particular acceptable or unacceptable list size. Generally, a larger list is superior, as described in 7.1.1.4.1. In the case where the PHY is deferring the leading edge computation to an upper layer, the PHY does not assign a FoM to the timestamp report. That does not mean that a FoM is unneeded by algorithms at the higher layers; it just means that the PHY will not be the source. In cases where the PHY defers computation, the upper layer will typically compute a FoM for itself based on the size and quality of the list returned with the SOUNDING.confirm primitive.

7.1.1.4.6 PHY deferral of the computations for self-calibration

The sounding primitives provide a mechanism to offload from the PHY the computational burden of analyzing a channel sounding for the leading edge of an arriving signal. A very similar problem arises in the self-calibration of a ranging UWB PHY. One excellent technique for self-calibration is the “sounding” of a loopback path in the radio. In this technique, the PHY actually transmits to itself through reflections associated with the transmission path (like a transmit/receive switch or the antenna itself). When using this technique, the issue comes up again that it can be computationally intense to discover the moment of arrival of the (often small) amplitude disturbances associated with elements in the path to the antenna. Implementers can choose to achieve this computational effort in the PHY. In an alternate implementation, the sounding mechanism can be used to offload this computational burden to a higher layer. When performing a sounding for leading edge computation, the instant associated with time zero is the point associated with signal tracking. Sounding for calibration is slightly different in that the time associated with zero is the launching of the pulse event from logic at the level of the ranging counter.

7.1.1.5 Management of crystal offsets

The numbers that will be subtracted in the range computation will typically represent times on the order of 5 ms. The time of flight for a 10 m link is about 30 ns. The expected difference of the counter values will be twice the time of flight, or something like 60 ns, in this example.

When a subtraction of values representing 5 ms is supposed to yield a meaningful answer on the order of 60 ns, even small percentage errors in the relative measurement of the 5 ms numbers yield large errors in the difference. The root cause of these errors is the fact that each of the individual 5 ms measurements was made with different crystals in different devices.

The crystals’ oscillators in different devices can generate frequency errors of 20 ppm. A 20 ppm error in a 5 ms number can account for 100 ns. This 100 ns error is disconcerting considering the 60 ns time-of-flight result.

Management of the errors due to crystal differences is essential to ranging. Correcting for a crystal difference algebraically at the time of the subtraction computation is straightforward if the difference is precisely known. The crux of the problem is to determine the crystal difference.

The mechanisms to characterize the crystal difference will be present and functioning in typical UWB PHY implementations. This crystal characterization equipment is the signal tracking loop in the receiver. A UWB pulse occupying 500 MHz has a nominal envelope width of 2 ns. The receiver tracking loop in a UWB PHY will stay “locked on” to this envelope for the duration of a packet. In the case of ranging packets, this will typically amount to milliseconds. In actual practice, the tracking loop will hold the sampling point on the envelope much tighter than 2 ns; therefore, by the end of the transmission, the tracking loop has the information to hold its sample point steady (with respect to its local crystal) on the received signal (which is sourced by the other devices crystal). In other words, by the end of the packet, the tracking loop has exactly measured the crystal difference. This crystal difference is the very thing necessary to correct the values in the ranging computation.

7.1.1.5.1 Characterizing crystal offsets with digital tracking loops

For a digital tracking loop, the most convenient way to express the crystal difference is with two numbers. A tracking interval number is the total number of units during which the signal was tracked, and a tracking offset number is a count of the number of times the tracking loop had to add or drop a unit to hold the sample point steady on the incoming signal. If the other oscillator frequency was lower than the tracking loop's local oscillator, then the tracking loop would be adding units to hold the sample point steady. The tracking offset is simply a count and a sign bit. All numbers are expressed from the local device's point of view; therefore, positive counts are characterizing crystals in the other device that are running slower (so the receiver was adding time units to match it) and negative numbers characterize faster crystals in the other device (so the receiver was dropping local units to keep up). The offset is thus a signed magnitude integer (not the twos-complement number that might be expected). The actual units (generally called "parts") that are called out in the count are whatever units happen to be convenient for a given PHY implementation. Since the numbers are used only as a ratio, the type of unit need not be specified as long as the numbers express the same unit.

7.1.1.5.2 Characterizing crystal offsets with analog tracking loops

PHYs that use analog phase-locked loops (PLLs) to track the received signal do not lend themselves as directly to the expression of the tracking offsets as counts. However, the PLL steady-state error signal is still a direct measure of the crystal offset. The analog PLL-based PHY can convert the PLL error signal to a number [for example, with an analog-to-digital converter (ADC)], put that result in the offset count field (taking care to get the sign bit correct), and put a convenient scaling number (like a million) into the total tracking interval field so that the ratio again expresses the difference of the crystals from the local oscillator's point of view.

7.1.1.5.3 Characterizing crystal offsets with different tracking loops

The use of the receiver's tracking loop to characterize the crystal offset is convenient for some PHY implementations, but it is not required for compliance. In fact, RDEVs are not required to support crystal characterization at all. If two RDEVs are involved in a ranging exchange and only one of them is supporting crystal characterization, all the information needed for a good ranging computation is available. If both RDEVs support crystal characterization, they will get the same ratio with opposite sign; therefore, there is little new information.

If neither RDEV supports crystal characterization, the application puts more traffic on the air to support ranging. For this situation, the application does the measurement twice. The first time is simply the normal exchange. On the second measurement, the roles are reversed. The device that was the originator on the first measurement is the responder for the second measurement, and likewise the responder on the first measurement becomes the originator for the second measurement. Then the application does the range computation twice. Because neither measurement provided for correcting for crystal offsets, the answers for both measurements are likely to be totally wrong. But, the computations did involve the same crystals so the error in the measurements is the same. Because the application reversed the measurement sequence between the two measurements, the answers have errors with opposite signs. The bottom line is that while the two individual answers are both hopelessly inaccurate by themselves, the average of the two individual answers will be exactly correct. A further refinement of this technique is called symmetric double-sided two-way ranging (SDS-TWR) and is discussed in 7.1.4.2. The refinement shown in 7 seeks greater efficiency by combining the two independent measurements into a single stream with the originator sending an acknowledgment for the responder's acknowledgment. While the "acknowledge for an acknowledge" approach is absolutely sound mathematically for ranging and the additional efficiency is tempting, the "acknowledge for an acknowledge" message sequence construct is beyond the scope of this standard.

7.1.1.5.4 Size of units

The units of measurement in the crystal characterization ratio are not rigidly defined in this standard to allow vendors the freedom to choose a value that works well with their PHY implementation. Design freedom is good, but to ensure at least a minimum level of ranging accuracy, this standard insists on a value that allows the ratio to express individual parts per million of oscillator difference. When the ranging computation is done for typical packet sizes and turnaround times, the desired answer will typically only be tens of parts per million compared to the numbers being subtracted. It is in this context that this standard calls for using the nominal 500 MHz chip time (nominally 2 ns) as the largest acceptable unit for the crystal characterization numbers. While an implementation using this value will be compliant, it would typically yield ranging errors on the order of a meter due to poor crystal characterization. To achieve reduced ranging errors, it is recommended that smaller units for crystal characterization be used since this translates directly to reduced errors in the ranging computation. When both RDEVs of a two-way ranging exchange are supporting crystal characterization, the application has a choice of which set of numbers will be used to make the correction. The ranging application would be wise to manage the reality that different devices might present different quality results for the crystal characterization.

The LSB of the ranging counter caps the highest ranging accuracy that can be achieved by compliant devices (which use the straightforward two-way ranging techniques described here). The LSB represents a nominal 16 ps, which corresponds to about half a centimeter of flight for energy in free space.

7.1.1.6 Accounting for internal propagation paths

The ranging counter is supposed to be capable of measuring events at the device antenna very precisely. It is understood that an actual implementation will not be trivial. Typically the device's PHY will have a counter somewhere in the digital section, multiple correction values stored in registers, and some arithmetic hardware to apply the correction values. The end result of all this is that it appears (from the numbers reported) that the PHY has a counter that is somehow magically positioned right at the antenna of the device and is taking snapshots of the counter values for events as they happen right at the antenna. That is important because the computation is supposed to be for the time of flight through the air, not through some impedance matching network feeding an antenna. Subtracting the correcting values for internal propagation times is not hard. What is hard is actually knowing the values of the internal propagation times. This standard provides a CALIBRATE mechanism (involving MLME primitives) that allows an application to cause a PHY (at a time that the application chooses) to invoke whatever capability that the PHY might have to characterize the internal propagation times of the PHY. Inclusion of a device capability for antenna loopback with an associated self-calibration algorithm is encouraged, but beyond the scope of this standard. A defensively written ranging application can maintain tables of correction factors at the computation nodes where the table entries are individually associated with the unique devices it is using for ranging. In this way, the application can compensate (after the fact) for devices in the ranging environment that have done a poor job of self-calibration.

7.1.1.6.1 PIB attributes for internal propagation paths

This standard provides a defined place to go to for the correction factors characterizing the delays of the internal propagation paths. There are two separate PHY PIB attributes that separately cover the transmit and receive paths. The intended use of these PIB attributes is for them to be written by the application at a time of the application's choosing and for the values to stay with the device until rewritten. One possible way for the application to learn what values to write to the PIB attributes is to invoke the CALIBRATE primitives. This standard does not mandate that the CALIBRATE primitives are used. The standard simply makes them available for use, if desired.

7.1.1.6.2 Support for self-calibration and one-way ranging

This standard does not preclude position awareness through one-way ranging. Successful one-way ranging requires that the internal propagation paths to the transmit antenna and from the receive antenna be accounted for separately. The PHY PIB attribute *phyTXRMARKEROffset* represents the time from the internal ranging counter reference to the transmit antenna. Likewise, the PHY PIB attribute *phyRXRMARKEROffset* represents the time from the receive antenna to the internal ranging counter reference.

7.1.1.6.3 Use of the calibrate primitives

The CALIBRATE.confirm primitives return either the values that are correct for the RMARKER offsets, if the PHY takes care of all computations itself, or the status COMPUTATION_NEEDED. The actual implementation of the self-calibration could be as simple as returning hardwired values for the RMARKER offsets. In this situation, the hardwired values would be selected by the vendor at the time of device manufacture to represent a best guess of what the offsets might ultimately be. Alternatively, the self-calibrate implementation might involve a full channel sounding of a loopback path and a sophisticated pattern-matching algorithm to determine from the sounding waveform when an internally generated pulse reflected back from an antenna. In either of the two scenarios, the status COMPUTATION_NEEDED would not be used because either (in the first case) the calibrate implementation was so crude that there was nothing to do, or (in the second case) the calibrate implementation was so sophisticated that the PHY took care of all of the computations without assistance.

It is a property of the algorithms that a node which only does ranging transmissions within a one-way infrastructure-based ranging application need not support calibration in any form.

7.1.1.6.4 Use of the COMPUTATION_NEEDED status

Two extreme implementations of PHY self-calibration were described in 7.1.1.6.2. This standard supports a reasonable middle ground implementation where the PHY does a channel sounding of a loopback path using the same hardware resources as normally used for leading edge scanning, but then the PHY defers the actual computations associated with the channel sounding to a higher layer. As discussed in 7.1.1.4.5, the computations associated with processing a channel sounding can be numerically intense and can be beyond the resources of a particular PHY implementation.

When the higher layer receives a status of COMPUTATION_NEEDED in response to a CALIBRATE.request primitive, the higher layer will then use the sounding primitives, as described in 7.1.1.4.6, to get a list of SoundingPoints from the PHY. The higher layer will then process the SoundingPoints to determine the values of the RMARKER offsets.

7.1.1.7 Timestamp reports

The two ranging counter values (start and stop), the ranging FoM value, and the two values that (as a ratio) characterize the crystal offsets all characterize a single ranging measurement. These five individual numbers that characterize a measurement are referred to in a group as a timestamp report. It then takes (at least) two timestamp reports to do a time-of-flight computation. The numbers in a single timestamp report have meaning in the context of each other. As such, they are generated by the PHY as a set and not split apart during subsequent data handling.

7.1.1.7.1 Presentation of timestamp reports

It should be noted that timestamp reports will occur at seemingly nonintuitive times in the actual primitives and the message sequence charts. For example, a timestamp report is included in the MCPS-DATA.confirm primitive. When the MCPS-DATA.confirm primitive is used following an initial transmission, the elements of the timestamp report are not all known. However, when the MCPS-DATA.confirm primitive is used following an acknowledgment in a ranging message sequence, all of the elements are known. Likewise, the MCPS-DATA.indication primitive includes a timestamp report, but when the MCPS-DATA.indication primitive is used in response to the initial reception of the first message of a ranging message sequence, not all of the elements of the timestamp report are known. However, when the MCPS-DATA.indication primitive is used following reception of the acknowledgment message, all of the elements are known.

7.1.1.7.2 Start and stop times in the timestamp report

The timestamp report as both a start time and a stop time. This can appear to be counterintuitive since either start or stop number by itself is useless and that the only real utility for the numbers is in their difference. A different strategy would be to have the PHY do arithmetic on the pair of numbers and present only the difference in the timestamp report. In this standard, the numbers are handled separately by the PHY to allow ranging by PHY implementations having few arithmetic or logic resources. Another reason is to allow an infrastructure node in a one-way ranging environment to issue a new timestamp report for each arriving RFRAME without being concerned about when the "start time" was.

7.1.1.8 Private ranging

It is important to note that for some applications of this standard, the range information will be the critical deliverable information for the entire system. As such, it is reasonable to protect this information as well as safeguard the integrity of the ranging traffic itself.

7.1.1.8.1 Simple encryption of the timestamp reports

At the end of the two-way exchange, half of the information necessary for the range computation is in each of two devices. Either half, by itself, is useless to the desired ranging application as well as to any undesired hostile device. When the two halves of the information are brought together, the range is computed by simple arithmetic. The single most critical and effective thing that an application can do to keep hostile devices from learning range information is encrypting the time reports whenever they are being transmitted. There is no problem doing this, as the reports are moved after the time-critical ranging exchange is complete and there is nothing time critical about movement of the timestamp reports.

7.1.1.8.2 Dynamic preamble selection (DPS)

It is anticipated that typical ranging traffic will take place using the normal channel codes and preambles in regular network use. Therefore, even if the time reports are encrypted and a hostile device is denied knowledge of the ranges, a hostile device can monitor traffic and listen for long preambles. It can then turn on its transmitter, spoof the acknowledgment transmission, and generally disrupt the ranging traffic. This hostile behavior creates a race between the hostile device and the legitimate responder, but the hostile device can expect to win the race because the legitimate responder will be parsing the MAC header to discover whether an acknowledgment is appropriate before it starts transmitting. To defeat this spoof attack, this standard offers the DPS option, where RDEVs are allowed to move their long preamble RFRAMEs to codes that are altogether different from the codes in normal use. Furthermore, the different preambles are coordinated using encrypted messages so that the hostile device is denied knowledge of the preambles that will be used. And finally, there are no retries allowed with these preambles so that a "jam and spoof the retry" attack will also be defeated.

When the DPS option is invoked by the devices that support it, the hazard is created where if the two-way ranging packet is not received as expected, the devices waiting and listening for special unique length 127 preambles will have become lost. To render this hazard safe, this standard provides for an additional timeout, DPSIndexDuration, which is used whenever DPS is used to ensure that devices at risk of becoming lost are always returned to an interoperable state, as described in 802.15.4™-2015.

DPS provides no additional ranging capability beyond resistance to attacks by hostile nodes. PHYs that do not implement DPS do not be give up any one-way or two-way ranging capabilities.

7.1.2 Time-of-arrival estimation from channel sounding

The range between a pair of transmitter and receiver devices can be estimated from the measured multipath profile characterizing the wireless channel between them. The peaks of the profile correspond to the arrivals, the first denoted as the time of arrival. Given τ and knowing that the signal travels at the speed of light *c*, the range between the two devices can be estimated as $c \times \tau$. The multipath profile appears in the form of a cross-correlation function of the received signal and the transmitted pseudo-random template sequence. In the proposed circuit in Figure 11 to estimate the delay in an AWGN channel, the receiver first samples the received signal through an ADC and then digitally correlates it with the template to generate a cross-correlation function.

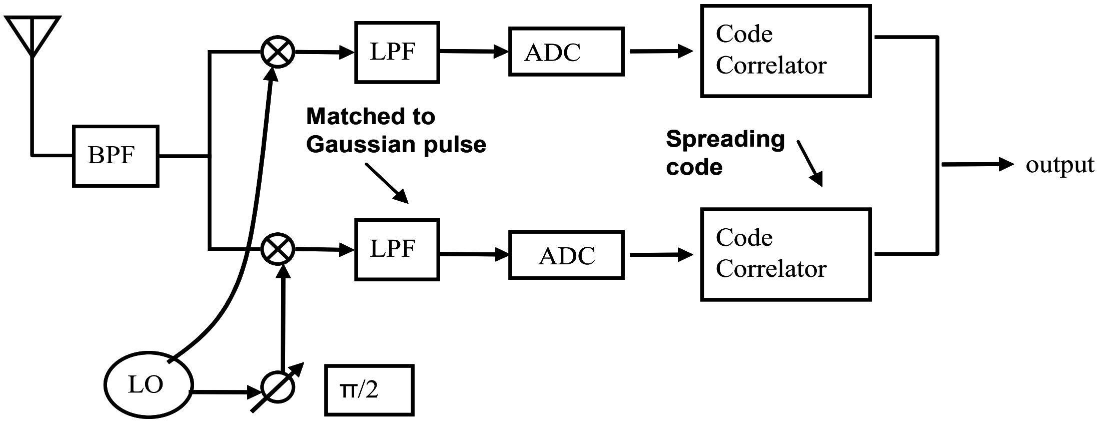

Figure 11—Circuit to compute multipath profile at receiver

Narrowband communication systems can afford a sampling rate equal to and up to five times the Nyquist rate[4] in the conventional approach to furnish good estimation accuracy. However, such a high sampling rate is difficult to implement with UWB devices demanding low cost and low power consumption. Motivated by low sampling rate, Qi and Kohno [B13] propose an approach tailored to such UWB devices that resorts to linear interpolation of the cross-correlation function through a second-order approximation of the maximum likelihood estimate. This estimate exploits both the given autocorrelation function of the template sequence and the given statistical characteristics of the noise, as shown in Figure 12.

Let the three largest adjacent correlation samples be denoted as $h(t_3) = [h(t_1)\ h(t_2)\ h(t_3)]^T$, the corresponding time instants as $t_3 = [t_1\ t_2\ t_3]^T$, and the inverse of the correlation matrix as:

$$W_3 = \begin{bmatrix} g(0) & g(T_s) & g(2T_s) \\ g(T_s) & g(0) & g(T_s) \\ g(2T_s) & g(T_s) & g(0)) \end{bmatrix}$$

[4] This rate is twice the bandwidth of the transmitted signal.

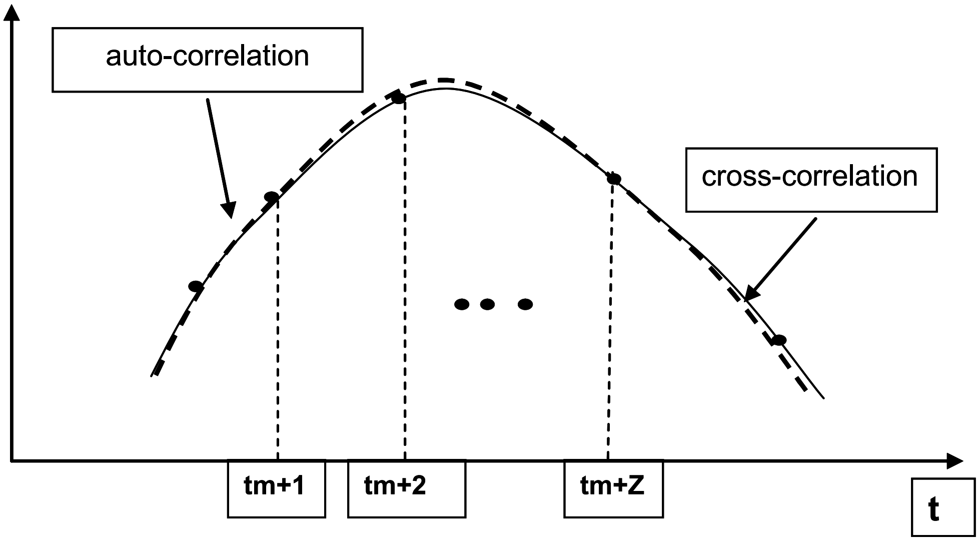

Figure 12—Autocorrelation and cross-correlation functions for simplified maximum likelihood estimator

where T_s is the sampling period. Let:

$$g(a) = \int s(t) \times s(t-w)dt$$

with $s(t)$ being a ternary pseudo-random sequence of length 31. The delay estimate can be expressed as a simple algebraic solution:

$$\hat{\tau} = \frac{t_3^T \times W_3 \times h(t_3)}{1_3^T \times W_3 \times h(t_3)}$$

where $1_3 = [1\ 1\ 1]^T$. This solution can be shown to be optimal in the sense that the estimate is approaching the theoretical lower limit as the sampling rate grows sufficiently large.

Figure 13 shows simulation results for a Gaussian UWB pulse of width 500 MHz, a PRF of 30.875 MHz, and an ADC sampling rate of 494 MHz (or 16×PRF). The conventional approach denotes choosing the sample time with largest peak, the interpolation approach denotes simple interpolation without the autocorrelation function, and the simplified maximum likelihood denotes the proposed approach. Figure 13 shows a decreasing RMS estimation error with increasing SNR, and Figure 14 shows the same decreasing error with increasing ADC sampling rate. In both plots, the simplified maximum likelihood outperforms the other two approaches.

7.1.3 Time-of-arrival estimation in non-line-of-sight (NLOS) conditions

In line-of-sight (LOS) conditions, the dominant peak in the cross-correlation function generated at the receiver corresponds to the first arrival. Since its strength is generally much greater than the subsequent peaks in the profile, it proves easy to isolate. However, in NLOS conditions, the first peak corresponding to the direct path is seldom the strongest, attenuated by transmission through walls and other objects; the strongest peak often corresponds to a reflected path whose travel time is greater than the direct path.

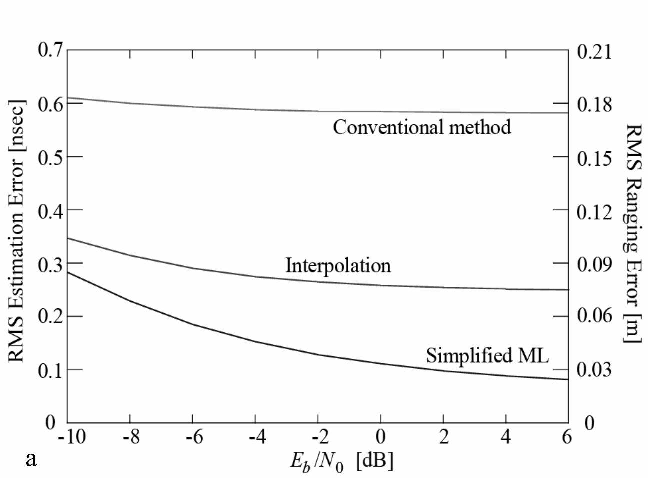

Figure 13—Performance comparison between the simplified maximum likelihood estimator and other approaches as a function of E_b/N_0

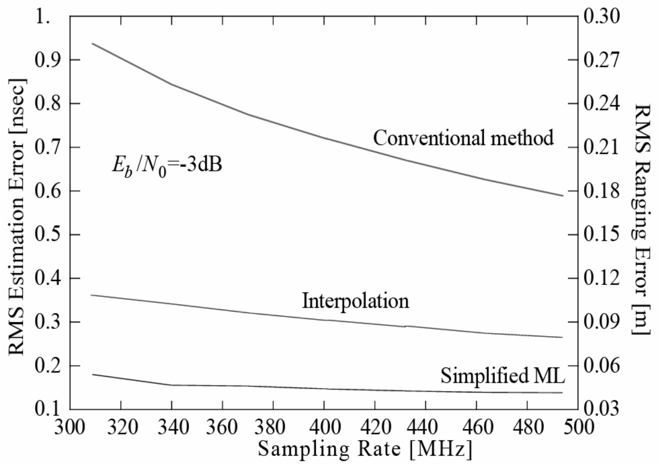

Figure 14—Performance comparison between the simplified maximum likelihood estimator and other approaches as a function of ADC sampling rate

In the latter case, a sequential linear cancellation scheme (Qi et al. [B12]) is devised for leading edge detection based on the aforementioned simplified maximum likelihood scheme. It can cope with the accuracy degradation when the first arriving signal component is weak compared to a dominant multipath component. This scheme reduces to an iterative algorithm. In each step, the amplitude Â of the present strongest component in the cross-correlation function is estimated based on a sliding delay $\hat{\tau}$:

$$\hat{A} = \frac{g(\hat{\tau}) \times W_3 \times h(t_3)}{g(\hat{\tau}) \times W_3 \times g(\hat{\tau})}$$

The autocorrelation samples, scaled by the amplitude Â and the time delay of the strongest component, are subtracted from the cross-correlation samples, effectively eliminating this component as in Figure 15. Since only the delay of the first arrival is of interest, components with delays greater than τ are subsequently removed in the following iterations, as shown in Figure 16, until no such components greater than a certain threshold exist. Guvenc and Sahinoglu [B7], [B8] and Lee and Scholtz [B10] cover threshold estimation techniques for UWB systems based on correlation.

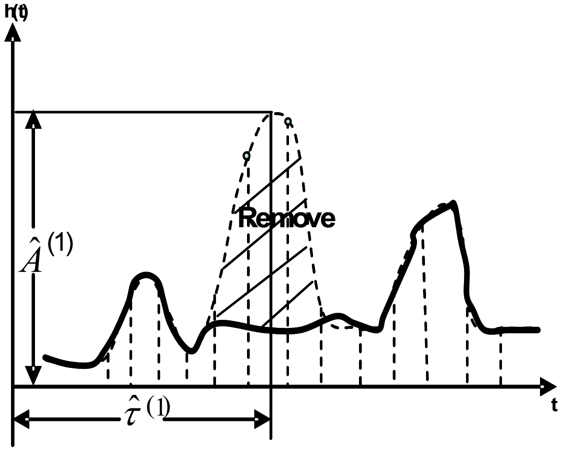

Figure 15—Leading-edge detection in NLOS conditions

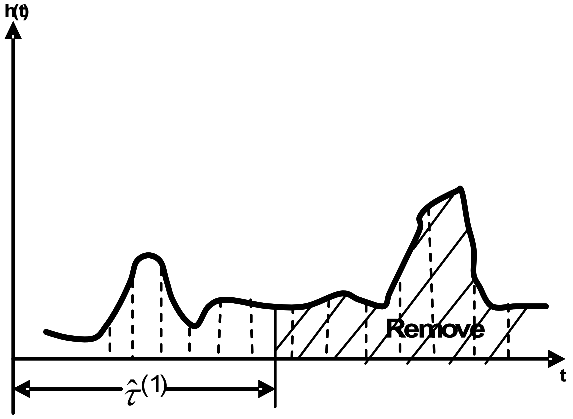

Figure 16—Leading-edge detection in NLOS conditions after removing delays

7.1.4 Asynchronous ranging

As described in the previous subclause, the time of arrival is extracted from the cross-correlation function generated at the receiver. It is used for the computation of the time of flight, t_p, defined as the time for propagation of the signal between the transmitter and receiver. The latter is found through an exchange of messages between the two devices in order to estimate range. The number of messages depends on the backbone structure of the network, which is described in detail in 7.1.5. With clock synchronization between the devices in the network, a single message suffices in one-way ranging to estimate t_p; in the absence of such synchronization, more messages are required. The finite crystal tolerance of the clocks is susceptible to drift and, therefore, has an effect on the number of messages required. This subclause considers two schemes for asynchronous ranging.

7.1.4.1 Two-way ranging (TWR)

In the absence of clock synchronization between two ranging devices, request device A uses its own clock as a time reference, as depicted in Figure 17.

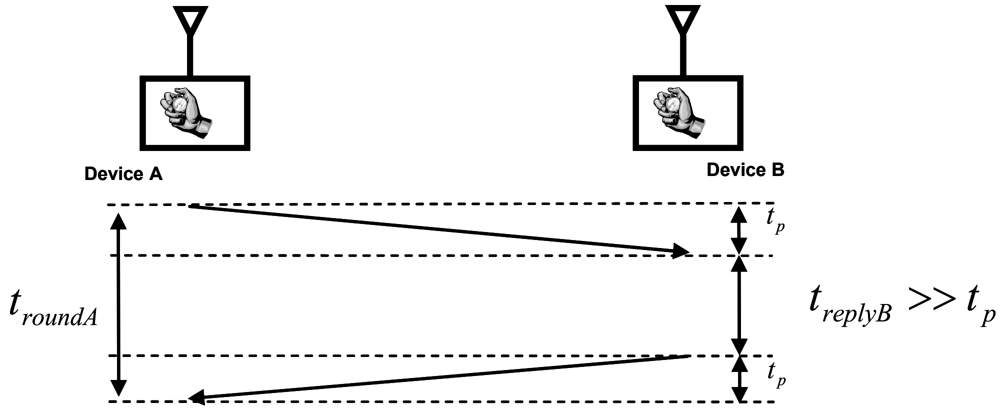

Figure 17—Exchange of message in two-way ranging

Device A begins the session by sending a range request message to device B. While device B can measure the absolute time of arrival of the message, lacking synchronization with device A, it does not know the time of departure of the message and, therefore, cannot extract t_p. Rather device B waits a time t_{replyB}, known to both devices, to send a request back to device A. Now device A can measure the round-trip time $t_{roundA} = 2t_p + t_{replyB}$ and extract t_p with respect to its own reference time.

Concern should be given to the finite tolerance of the device crystal reference frequency since frequency error introduces error in the measurement of t_p. In reference to Figure 17, the true value of t_p is computed in terms of the transmitted and received times (denoted by subscripts T and R, respectively) at devices A and B:

$$2t_p = (\underbrace{\tau_{BR} - \tau_{AT}}_{t_p}) + (\underbrace{\tau_{AR} - \tau_{BT}}_{t_p}) = (\tau_{AR} - \tau_{AT}) + (\tau_{BR} - \tau_{BT})$$

Therefore, the estimated value $\hat{t}_p$ follows as:

$$2\hat{t}_p = (\underbrace{\tau_{AR} - \tau_{AT}}_{t_p}) \times (1 + e_A) + (\underbrace{\tau_{BR} - \tau_{BT}}_{t_p}) \times (1 + e_B)$$

where e_A and e_B represent the crystal tolerances of the respective devices expressed in parts per million. Substituting for $\tau_{AR} - \tau_{AT} = 2t_p + t_{replyB}$ and $\tau_{BR} - \tau_{BT} = -t_{replyB}$ and simplifying gives:

$$\hat{t}_p - t_p = \frac{1}{2}(t_{replyB} \times e_A - t_{replyB} \times e_B + 2t_p \times e_A)$$

Note the t_{replyB} is not the turnaround time between the received message from device A and the sent message from device B, but rather includes both the packet duration and this turnaround time. Since the packet duration is on the order of several milliseconds, this duration implies $t_{replyB} >> t_p$ and, therefore:

$$\hat{t}_p - t_p \approx \frac{1}{2} \times t_{replyB} \times (e_A - e_B)$$

Table 5 presents some typical values for $\hat{t}_p - t_p$ according to the other system parameters.

Table 5—Typical errors in time-of-flight estimation using TWR

$t_{replyB}/(e_A – e_B)$	2 ppm	20 ppm	40 ppm	80 ppm
100 µs	0.1 ns	1 ns	2 ns	4 ns
5 ms	5 ns	50 ns	100 ns	200 ns

The scope specifies a ranging precision of 1 m; therefore, the estimated t_p needs to lie within 3 ns of the true time of flight given the speed of light. Obviously for the normative packet duration, even with high-quality crystals with tolerance of 2 ppm, the measurement error is greater than the required resolution of the ranging system.

7.1.4.2 Symmetric double-sided two-way ranging (SDS-TWR)

In order to compensate for the shortcomings of simple two-way ranging, Hach [B9] proposes an additional message exchange in the ranging session to reduce the effect of the finite crystal tolerances of the devices. Figure 18 shows the message exchange.

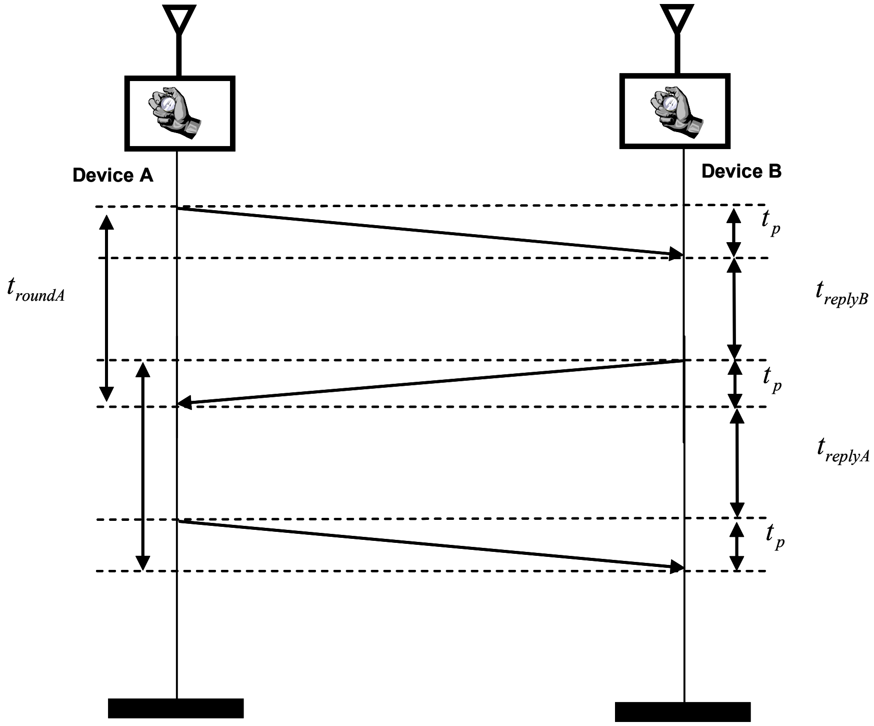

Figure 18—Exchange of message in SDS-TWR

The diagram shows that the round-trip times t_{roundA} and t_{roundB} can be expressed in terms of t_p nd the respective t_{replyA} and t_{replyB} as follows:

$$t_{roundA} = 2t_p + t_{replyB}$$

$$t_{roundB} = 2t_p + t_{replyA}$$

Combining the two equations allows for isolating the true value of t_p as:

$$4t_p = t_{roundA} - t_{replyA} + t_{roundB} - t_{replyB}$$

and the estimated $\hat{t}_p$ follows by introducing the finite crystal tolerance e_A and e_B:

$$4\hat{t}_p = (t_{roundA} - t_{replyA}) \times (1 + e_A) + (t_{roundB} - t_{replyB}) \times (1 + e_B)$$

Without loss of generality, replacing t_{replyA} and t_{replyB} with:

$$t_{replyA} = t_{reply}$$

$$t_{replyB} = t_{reply} + \Delta_{reply}$$

reduces it to:

$$\hat{t}_p - t_p = \frac{1}{2} \times t_p \times (e_A + e_B) + \frac{1}{4} \times \Delta_{reply} \times (e_A - e_B)$$

Assuming that $t_p \ll \Delta_{reply}$, the equation simplifies further to:

$$\hat{t}_p - t_p \approx \frac{1}{4} \times \Delta_{reply} \times (e_A - e_B)$$

Table 6 shows the typical errors in the SDS-TWR time-of-flight estimation versus frequency tolerance.

Table 6—Typical errors in time-of-flight estimation using SDS-TWR

$\Delta_{reply}/(e_A - e_B)$ (μs)	2 ppm (ns)	20 ppm (ns)	40 ppm (ns)	80 ppm (ns)
1	0.0005	0.005	0.01	0.02
10	0.005	0.05	0.1	0.2
100	0.05	0.5	1	2

The extra message in the SDS-TWR accommodates a much smaller error margin even with low-quality crystals of 80 ppm.

7.1.5 Location estimation from range data

The ranging capabilities of the devices can be used to estimate their locations through network collaboration. A device wishing to determine its location gathers at least three or four ranges to neighboring devices with known location in a two- or three-dimensional network. The technique used to triangulate the estimated ranges to an estimated location of the device depends largely on the network topology and communication protocols. The subclause considers the two main approaches to simple triangulation.

7.1.5.1 Time of arrival

Consider device A initiating a ranging session by sending a request message to device B. Device B can estimate the time of arrival τ_A from the message through one-way ranging. If the two clocks are synchronized to the same time reference, device B can also extract the time of departure τ_D included in the message by device A and hence compute the time of flight $t_p = \tau_D - \tau_A$. In practice, it proves difficult or inefficient to establish and/or maintain synchronization between two mobile devices; therefore, a network lacking a wired backbone will need to resort to asynchronous ranging described in 7.1.4.

The time-of-arrival technique for triangulation of ranges applies to the general network lacking synchronization between devices and/or a priori organization. The technique assumes that three (or more) ranges $c \times t_1$, $c \times t_2$, and $c \times t_3$ are gathered from anchor devices $i = 1, 2, 3$, respectively, with known locations (x_i, y_i), as in Figure 19.

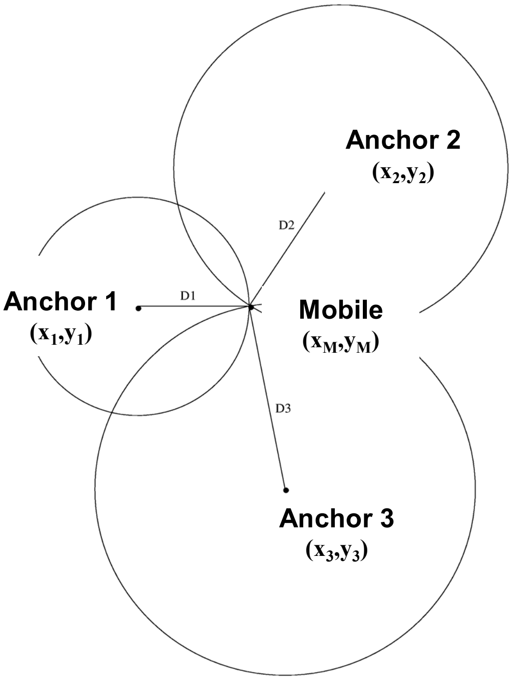

Figure 19—Time-of-arrival triangulation of ranges to determine location

$$c \times t_1 = \sqrt{(x_1 - x_M)^2 + (y_1 - y_M)^2}$$
$$c \times t_2 = \sqrt{(x_2 - x_M)^2 + (y_2 - y_M)^2}$$
$$c \times t_3 = \sqrt{(x_3 - x_M)^2 + (y_3 - y_M)^2}$$

Solving the set of equations translates to finding the intersection of the three circles (or spheres in three dimensions), yielding the unknown coordinates of the mobile device (x_M, y_M).

Time of arrival is appealing due to its application to the general network architecture; however, the associate asynchronous ranging requires two or more messages per ranging session. This requirement may potentially increase the network traffic considerably.

7.1.5.2 Time difference of arrival

A pre-installed network can be configured so that the mobile devices within a deployment area can maintain connectivity with at least three anchor devices positioned at known locations and connected through wire to maintain clock synchronization. Clock synchronization enables one-way ranging in conjunction with the time-difference-of-arrival technique, as opposed to time of arrival. In order to carry out a range request, the devices in the network classified as stationary anchor devices and mobile devices operate in one of the two following modes: Mode 1 or Mode 2.

7.1.5.2.1 Mode 1

The synchronized anchor nodes jointly send range requests to the mobile node at the same time instant; therefore, the number of messages equals the number of anchors. Although the mobile node lacks synchronization with the three anchors indexed through $i = 1, 2, 3$, it can still measure the time difference of arrivals $t_{32} = \tau_3 - \tau_2$ and $t_{31} = \tau_3 - \tau_1$ within its own time reference, where τ_i denotes the arrival time of the message from anchor i at the mobile M. Figure 20 depicts the locations of anchors i and respective locations (x_i, y_i). Solving the following equations translates to finding the intersection of two hyperbolae and yields the unknown location (x_M, y_M) of the mobile device. The burden of the calculation lies on the mobile device, which may have limited resources with respect to the anchor devices due to its smaller dimensions to preserve battery life.

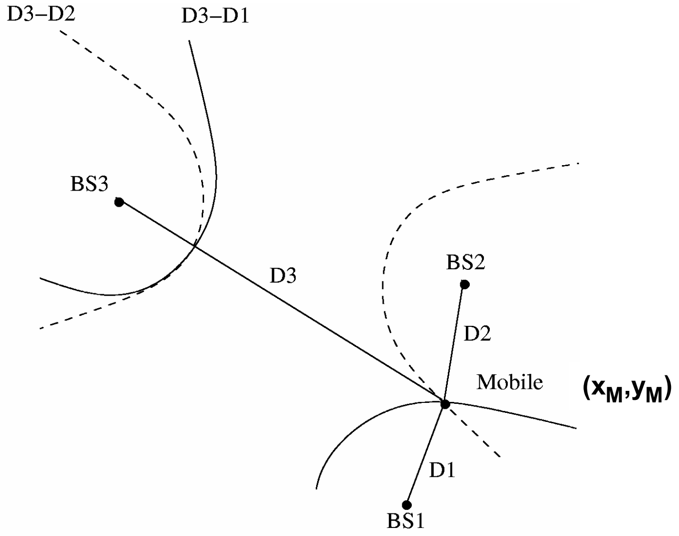

Figure 20—Time-difference-of-arrival triangulation of ranges to determine location

7.1.5.2.2 Mode 2

This mode operates in a similar fashion to Mode 1; however, the mobile device initiates the range request by sending a single message received by all the neighboring anchor nodes. The request message arrives at the anchor nodes at different times of arrival measured individually. These values are either distributed throughout the wired backbone or sent to a central controller to compute the time difference of arrivals t_{32} and t_{31} and in turn the location of the mobile device. This mode offers the advantage of reducing the wireless network traffic to just one message per request and also does not place the burden of computation on the mobile node with potentially limited resources.

7.1.6 Network location algorithms

In the basic approach to location estimation described in 7.1.5, a mobile device gathers range measurements from a number of anchor devices with known locations and triangulates these ranges to a single point (or area) through simple geometrical relationships. This approach assumes that the mobile device receives at least three or four estimates in a two- or three-dimensional coordinate system.

Interest in dense sensor networks due to falling price and reduced size has motivated research in network location algorithms in recent years, where the number of mobile devices vastly outnumbers the number of fixed stations. Consider a rapidly deployable network whose nodes are scattered about an area of interest and self-organize in an ad hoc fashion to determine their locations through simple messaging and ranging. Most of the sensors in such networks lack connectivity to fixed stations. Rather, the high sensor-to-sensor connectivity allows them to infer their locations from the locations of other sensors that do have connectivity to the fixed stations. This class of algorithms to process large amounts of range data is commonly known as *network location algorithms*. These algorithms can render good location accuracy despite significant errors in range estimates between sensors.

This subclause provides a survey of the benchmark network location algorithms developed in recent years and divides them into three main classes: ad hoc algorithms, centralized algorithms, and convex optimization algorithms.

7.1.6.1 Ad hoc algorithms

Network location algorithms arose from the need to locate nodes in ad hoc networks characterized by high mobility and dynamic architecture, with nodes joining and leaving the network at random times. Here greater importance is placed on continuously updating location in a distributed manner rather than furnishing precision. Niculescu and Nath [B11] contributed to this pioneering effort. Their work describes several algorithms to this end based on the range measurements available to the nodes. The strength of these algorithms lies in their simplicity and in turn their applications to ad hoc networks. In general they foster reduced complexity for larger networks.

The first algorithm known as *DV-Hop* generates location based on mere connectivity quantified as the minimum number of hops between two nodes, or the minimum-hop distance. Hence the nodes in the network lack any ranging capabilities. The network is categorized into two types of nodes: *anchor nodes* whose locations are known and *sensor nodes* whose locations are unknown. Rather than estimate range between each other, the nodes simply determine the number of hops between each other. Given the ground-truth ranges between the anchors through simple messaging of their known locations through the ad hoc network, each anchor node i computes a factor f_i, which is simply the sum of the ground-truth ranges between all the anchors in the network divided by the sum of the respective minimum hops.

Consider the network in Figure 21. Anchor L1 computes factor f_1 by summing the ground-truth ranges from L2 (40 m) and L3 (100 m), divided by the sum of the minimum hops 6 and 2, or $f_1 = (40 + 100)/(2 + 6)$ m. Anchor L2 and L3 compute their factors f_2 and f_3 in the same manner as $f_2 = (40 + 75)/(2 + 5)$ m and $f_2 = (75 + 100)/(5 + 6)$ m, respectively, and then distribute them to all the sensors in the network. Ultimately the sensor node A, pictured in Figure 22, finds its location from its three closest neighboring anchor nodes; however, rather than through the triangulation of three measured ranges R_1, R_2, and R_3 to them, respectively, it finds its location through DV-ranges $R_1 = 3 \times f_1$, $R_2 = 2 \times f_2$, and $R_3 = 3 \times f_3$ given through the minimum hops and the factors.

In the same paper, the authors present an alternative algorithm known as *DV-Distance* as an adaptation of the ad hoc positioning system to accommodate nodes with ranging technology for enhanced precision without compromising simplicity. Rather than scaling the ground-truth distance by the minimum-hop distance between two anchors in computing the factors of each anchor node, it is scaled by the sum of the measured distances on the multihop path between two anchors. It directly follows in the triangulation step

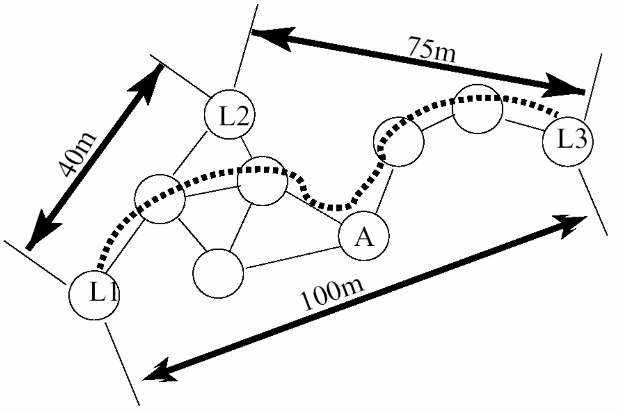

Figure 21—Network for comparison between the simplified maximum likelihood estimator and other approaches

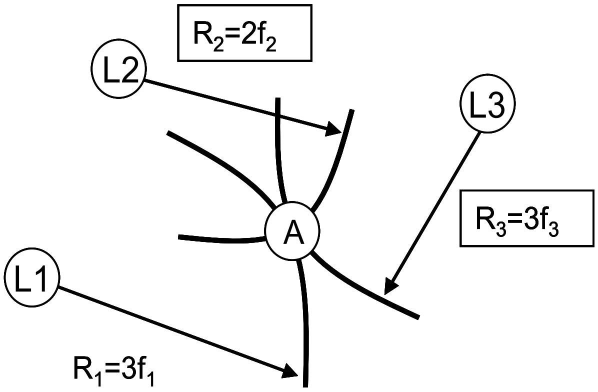

Figure 22—Performance comparison between the simplified maximum likelihood estimator and other approaches

that the distances between the unknown sensor and the three closest anchor nodes are scaled by the measured distances on the multihop path between the two.

Saverese et al. [B14] present another algorithm designed for ad hoc networks, which also incorporates range estimates. Knowing its coordinates (x_1, y_1), anchor L1 orients a local coordinate system pictured in Figure 23 in the direction of an arbitrary neighbor, e.g., L2. Given the measured range r_{12}, the coordinates $(x_2, y_2) = (x_1 + r_{12}, y_1)$ of sensor 2 are easily computed in the same coordinate system. Given further the range estimates r_{13} between anchor L1 and sensor L3 and r_{23} between sensors L2 and L3, the coordinates of $(x_3, y_3) = (x_2 + dx$, y2 + $dy)$ are computed through:

$$dx = \frac{r_{12}^2 + r_{13}^2 + r_{23}^2}{2 \times r_{12}}$$

$$dy = \sqrt{r_{12}^2 - dx^2}$$

In this manner, anchor L1 propagates the coordinates of its successive neighbors throughout the network as an initialization step. The other anchor nodes independently do likewise.

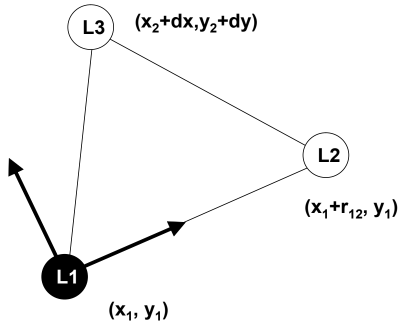

Figure 23—Ad hoc network location algorithm proposed by Savarese et al.

In propagation through the individual neighbors of the n_A anchor nodes in the network, node 2 receives n_A estimates of its coordinates. It can then refine its estimate $(\hat{x}_2,\hat{y}_2)$ as the equally weighted average of all the accumulated estimates. In sequence, sensor 3 refines its estimate through the propagating of the coordinates from anchor L1 and sensor 2 as in the initialization step; however, it replaces (x_2, y_2) with $(\hat{x}_2,\hat{y}_2)$. Propagation then continues and eventually converges to an acceptable solution after a number of iterations, although the authors provide no proof. In the network simulations, the algorithm furnishes a location error of about 5% even with range errors up to 50%.

7.1.6.2 Centralized algorithms

Centralized algorithms gather all the data available from the network to process it collectively. As expected, these algorithms in general render better results than the ad hoc algorithms, which consider only local data and process it independently of the data available to other parts of the network. The drawback of the former involves designating a central controller, a condition which may be unsuitable for some applications. Even so, the dynamic links of nodes in motion may require rapid updating; alternatively, relaying information across a large network sanctions the centralized processing of obsolete data at the controller, and this condition limits scalability. It is worth noting that most of these algorithms have distributed versions, which, however, compromise the quality of the results. This subclause considers the architecture of two centralized algorithms commonly referred to in literature.

Many centralized algorithms estimate the locations of the unknown nodes by minimizing an objective function such as in Savvides et al. [B15]. A popular function, as used here, is the least-squared sum of the residuals; the residual is defined as the difference between a measured range and the range estimated through the algorithm. The minimization is performed through Kalman filtering, which ultimately finds only a local minimum. In fact, an initialization step is required to estimate the positions of the nodes from the measured ranges and the anchor nodes. Consider sensor node C in Figure 24 as an example. The simplistic initialization step finds a bounding rectangle for it as the intersection of the individual bounding squares of each anchor node (here A and B) to the sensor: the length of half the side of the square is the sum of the measured distances on the multihop between the anchor and the sensor:

Shang et al. [B16] have designed another centralized algorithm to minimize the least-squared sum of the residuals, but use a more powerful technique called *multidimensional scaling* to find the global maximum. As in Savvides, the initial step consists of computing the minimum-hop distances between all nodes in the network, but here they are stored in a range matrix *R* having the following structure:

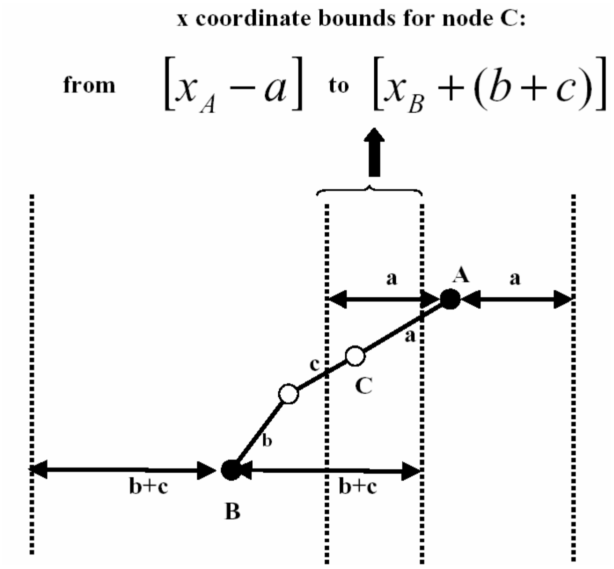
x coordinate bounds for node C:
from $[x_A - a]$ to $[x_B + (b+c)]$
a
A
a
a
c
C
b
b+c
B
b+c

Figure 24—Autocorrelation and cross-correlation functions for the maximum likelihood estimator

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots \\ r_{21} & r_{22} & r_{23} & \cdots \\ r_{31} & r_{32} & r_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

The minimum-hop distance between two nodes is just the sum of the estimated ranges between them.[5] Performing the singular value decomposition of R^2 yields a matrix of eigenvectors in V and respective squared eigenvalues on the diagonal matrix A. The relative coordinates of the nodes are subsequently reconstructed as $X = V\sqrt{A}$. These coordinates are then scaled, translated, and rotated to fit the anchor nodes.

7.1.6.3 Convex optimization algorithms

The most powerful class of network location algorithms places the geometrical constraints of the physical world on the nodes while minimizing an objective function similar to the one in 7.1.6.2. The advantage of the technique over the ones in 7.1.6.2 lies in its ability to achieve the global maximum independent of any initial estimates. Doherty et al. [B4] first proposed the technique in defining the following optimization problem with quadratic constraints to minimize a linear objective function:

$$\min \sum |\alpha_{ij}|$$

such that:

[5] The estimated ranges are actually normalized to fit the framework of the multidimensional scaling.

$$\begin{aligned}\|\hat{\bar{x}}_i - \hat{\bar{x}}_j\|_2 &= d_{ij},\ \forall(i,j) \in N \\ \|\hat{\bar{x}}_i - \hat{\bar{x}}_j\|_2 &\geq R,\quad \forall(i,j) \notin \bar{N}\end{aligned}$$

where

$d_{ij} = \hat{d}_{ij} + \alpha_{ij}$ is the estimated range resultant of the algorithm

$\hat{d}_{ij}$ is the measured range

α_{ij} is the residual between the two

N is the set containing the index of neighboring nodes

$\bar{N}$ is the set of non-neighboring nodes

R is the radio range, i.e., the range at which a node loses connectivity with another node in the network

Since many of the constraints are nonconvex, the paper relaxes the problem simply by removing these constraints. The corresponding solution not only offers less precision, but also forces the sensors to lie within the convex hull of the anchors. An alternative approach proposed by Biswas and Ye [B3] relaxes the problem to a semi-definite program that yields an average and standard variation for the positions of the unknown nodes.

Rather than relaxing the problem from the original, Gentile [B6] directly applies linear constraints given through the triangle inequality:

$$\min \sum |\alpha_{ij}|$$

such that:

$$\left.\begin{aligned} d_{ij} + d_{jk} &\geq d_{ik} \\ d_{ij} + d_{ik} &\geq d_{jk} \\ d_{jk} + d_{ik} &\geq d_{ij} \end{aligned}\right\},\left\{\begin{aligned} &\forall(i,j) \in N \\ &\forall(j,k) \in N \\ &\forall(i,k) \in N \end{aligned}\right.$$

The original convex constraints necessitate no relaxation and hence render a much tighter solution than the other two approaches mentioned. A distributed version of the algorithm yields the same results as the centralized version with no compromise in achieving an optimal result (Gentile [B5]).

In order to substantiate the effectiveness of the network location algorithms, Table 7 presents the results from Gentile [B6]. The simulation platform consists of a network with 50 sensor nodes uniformly distributed in a unit area. The number of anchor nodes varies as a parameter {3, 5, 7} in the table and R as {0.20, 0.25, 0.30}. The noise parameter controls the percentage of Gaussian-distributed noise perturbing the measured radio range from the ground-truth range and varies as {0.0, 0.1, 0.2, 0.3}. The table shows that despite the measured range errors up to 30%, the location errors yielded by the algorithm can lie on the order of only 5%.

Table 7—Location results according to varying network parameters

noise	*R* = 0.20			*R* = 0.25			*R* = 0.30		
	3	5	7	3	5	7	3	5	7
0.0	0.043	0.042	0.041	0.007	0.006	0.006	< 1e–6	< 1e–6	< 1e–6
0.1	0.075	0.064	0.064	0.053	0.044	0.027	0.045	0.036	0.025
0.2	0.085	0.080	0.068	0.077	0.065	0.049	0.057	0.057	0.046
0.3	0.107	0.088	0.087	0.095	0.083	0.077	0.077	0.074	0.046

7.1.6.4 Location estimation using multipath delays

For location estimation in a multipath environment, the conventional approach is based on leading-edge detection, where the time-of-arrival estimate of the first arriving signal is taken as the distance between the transmitter and the receiver of interest up to a constant and the delays of other multipath signals are completely ignored. This approach works well when the first arrival signal is sufficiently strong and via a LOS propagation path. However, in a typical UWB channel, the first arrival signal is usually weak, e.g., 6 dB lower than a dominant multipath component, and can be subject to NLOS propagation. Hence the conventional approach can cause severe degradation of the positioning accuracy. To address this problem, one method is to utilize time-of-arrival estimates of multipath components in addition to the first arriving signals for location estimation. Although subject to NLOS propagation, the second and later arriving signals should also carry information regarding the position of interest. Hence the method incorporating the multipath delays can improve the positioning accuracy under certain conditions.

For simplicity, consider that a mobile node is synchronized with B anchor nodes, whose locations $\{(x_b, y_b), b = 1, 2, \ldots, B\}$ are known. Each anchor node receives radio signals transmitted from the mobile node via multipath propagation. A received signal at the b^{th} anchor node is expressed as:

$$r_b(t) = \sum_{i=1}^{N_b} A_{bi} \times s(t - \tau_{bi}) + n_b(t)$$

where τ_{bi} is the delay of the i-th multipath component, given by:

$$\tau_{bi} = \frac{1}{c}\sqrt{(x - x_b)^2 + (y - y_b)^2} + l_{bi}$$

which consists of the LOS delay corresponding to the distance between the mobile node and the anchor node, and the NLOS induced path length error l_{bi}. The quantity $l_b = (l_{b1}, l_{b2}, \ldots, l_{bN})^{\mathrm{T}}$ is usually modeled as a multivariate random variable, which can be determined by field experiments or theoretical models. Noise $n_b(t)$'s are independent white Gaussian processes, and A_{bi} is the signal amplitude. Estimation of the multipath delays yields:

$$\hat{\tau}_{bi} = \tau_{bi} + \xi_{bi}$$

where $\xi_b = (\xi_{b1}, \xi_{b2}, \ldots, \xi_{bN})^{\mathrm{T}}$ is a multivariate Gaussian random variable with zero mean and an explicit covariance matrix F_ξ. Based on the equation for $r_b(t)$ and the probability density functions of l_b and ξ_b, location estimation of (x, y) can be formulated as the maximum a priori estimation. It is shown that the positioning accuracy enhancement depends on two principal factors, the strength of multipath components and the variance of the NLOS induced errors. In certain situations, significant accuracy improvement, e.g., greater than 50%, can be obtained. The limitation of this approach is that its computation complexity is higher than the conventional approach. The exact formula of accuracy improvement and detailed discussion on this approach can be found in Qi et al. [B12].

7.2 Ranging considerations for operation in TVWS

6.2 describes a ranging mechanism for a TVWS WPAN.

7.2.1 Introduction

The geolocation requirements for TVWS specify that the accuracy of a geolocation capability to determine its geographical coordinates is ± 50 m for Mode II fixed and personal/portable devices. Mode I devices may also require location capability. It may be possible to provide a geolocation capability by incorporating a GPS receiver on a device. However, the GPS service may not always be available in some situations, such as when the receiver is inside a building or urban canyon, or is under attack through jamming or spoofing. Moreover, battery-powered Mode I devices may not be equipped with GPS receiver. Therefore, it is advisable to provide optional RF localization for TVWS WPANs.

7.2.2 General

The ranging mechanism for TVWS WPAN PHYs is basically the same as that of the UWB PHY, shown in 7.1. Similar to the UWB PHY, a TVWS WPAN frame with the ranging bit set in the PHR is called a ranging frame (RFRAME). The critical instant in this RFRAME is the start of the PHR for both TVWS-FSK, TVWS-OFDM, and TVWS-NB-OFDM PHYs, known as the ranging marker (RMARKER). In the two-way ranging technique, ranging counter values in the ranging originator are captured upon RMARKER departure and arrival, while ranging counter values in the ranging responder are captured upon RMARKER arrival and departure. In this ranging counter operation, the exact timing of RMARKER for any RFRAME transmission can be easily determined. However, the timing of the RMARKER arrival at the receiver that determines the ranging performance is susceptible to noise, signal bandwidth, and operation clock frequency tolerance. As a result, a major issue in TVWS WPAN based ranging is how to obtain the accurate arrival time of TVWS-FSK, TVWS-OFDM, and TVWS-NB-OFDM signals.

The technique for achieving appropriate accuracy of this signal arrival time is outside the scope of this standard, but it is helpful to discuss a typical approach for TVWS WPAN PHYs, e.g., TVWS-FSK, TVWS-OFDM, and TVWS-NB-OFDM PHYs. In the following, the symbol transition timing (STT) estimation for the TVWS-FSK PHY and the time of arrival (ToA) estimation for the TVWS-OFDM and TVWS-NB-OFDM PHYs are briefly described.

7.2.3 Estimation for TVWS-FSK PHY

Generally, the FSK system has not been used for accurate ranging due to its narrowband characteristics. However, the accuracy of ± 50 m in TVWS enables FSK-based ranging to assist a geolocation capability of TVWS devices. Unlike the UWB and OFDM PHYs that exploit a correlation property of the preamble sequence, the timing of the RMARKER arrival in an FSK system can be obtained from STT estimation during the preamble, whose sequence is multiple repetitions of "01010101."

One approach for STT estimation is to use the phase difference vector of the received FSK signal. The phase of the FSK signal is reversed in every symbol during the preamble. Therefore, the phase difference vector between the received signal and its delayed signal shows a phase transition, from which the symbol transition time can be estimated. The TVWS-FSK PHY WPAN allows applications to specify the length of the preamble (between 4 to 1000 bytes); this feature can be used to enhance the ranging performance by increasing the number of preamble symbols involved in STT estimation.

7.2.4 ToA estimation for TVWS-OFDM PHYs

The conventional autocorrelation-based schemes can be used for ToA estimation in the TVWS-OFDM and TVWS-NB-OFDM PHYs since the STF and LTF sequences in the SHR show a good autocorrelation property.

8. Rail communications and control (RCC) systems

8.1 General description

For the purposes of the 802.15.4 standard, an RCC system refers to a wireless communications system used for rail vehicle data communications between the rail vehicle and fixed infrastructure. This standard applies to information exchange and sensor or control communications.

Portions of RCC systems are deployed as:

- A wireless link between trains, locomotives, or other mobile rolling stock to fixed trackside or network infrastructure
- A link between connected fixed, remote trackside infrastructure and fixed network infrastructure
- A link between vehicles in the same train or between two or more trains

This standard is capable of supporting fixed-to-fixed, fixed-to-mobile, and mobile-to-mobile communications.

RCC PHYs are designed in such a manner that ranges of over 50 km are practical, subject to propagation path loss, transmitter output power, carrier frequency, data rate, and antenna placement/height above average terrain. RCC PHYs are intended to support mobile rail vehicle communications at speeds up to 600 km/h and data rates from 9.6 kbps to nearly 1 Mbps. The PHYs are designed to take advantage of relatively small amounts of spectrum where spectrum is costly or scarce, with the ability to operate in channel widths from 12.5 kHz (licensed spectrum) to nearly 2 MHz (license-exempt spectrum).

8.2 Introduction to communications-based train control (CBTC)

As defined by IEEE Std 1474.1™-2004 [B17], a CBTC system is a “continuous, automatic train control system utilizing high-resolution train location determination, independent of track circuits; continuous, high-capacity, bidirectional train-to-wayside data communications; and trainborne and wayside processors capable of implementing Automatic Train Protection (ATP) functions, as well as optional Automatic Train Operation (ATO) and Automatic Train Supervision (ATS) functions.” This standard provides the bidirectional train-to-wayside data communications function. Many of the largest metropolitan rail transit systems in the world use CBTC.

This standard provides a bi-directional wireless communications link that can be used for CBTC systems, and has the flexibility to employ either licensed or license-exempt frequency bands to provide flexibility and robustness.

8.3 Example: positive train control (PTC)

In 2008, the United States Congress enacted a law called the Rail Safety Improvement Act of 2008, in order to improve rail safety. The law uses the phrase “positive train control system” to describe a safety system designed to prevent train-to-train collisions, over-speed derailments, incursions into established work zone limits, and the movement of a train through a switch left in the wrong position. The law does not specify implementation of such a system.

As interpreted by the US Federal Railroad Administration, a PTC system includes four components:

- Equipment deployed on the locomotive/train
- Equipment deployed trackside

- Network access points deployed at or near trackside that are connected to systems operating at a remotely located control center
- A bi-directional wireless data link that connects all these elements

PTC systems are integrated command, control, communications, and information systems for controlling train movements with safety, security, precision, and efficiency. PTC systems will improve railroad safety by significantly reducing the probability of collisions between trains, casualties to roadway workers and damage to their equipment, and overspeed accidents.

PTC systems are composed of digital data link communications networks, continuous and accurate positioning systems such as National Differential GPS, on-board computers with digitized maps on locomotives and maintenance-of-way equipment, in-cab displays, throttle-brake interfaces on locomotives, wayside interface units at switches and wayside detectors, and control center computers and displays. PTC systems also interface with tactical and strategic traffic planners, work order reporting systems, and locomotive health reporting systems. PTC systems issue movement authorities to train and maintenance-of way crews, track the location of the trains and maintenance-of-way vehicles, have the ability to automatically enforce movement authorities, and continually update operating data systems with information on the location of trains, locomotives, cars, and crews. A remote intervention capability of a PTC system permits the control center to stop a train should the locomotive crew be incapacitated.

A number of radio frequency bands currently used or planned for rail and rail transit communications are included in this standard. Also included are modulation modes and error-correction techniques that enhance functionality for low-data rate rail and rail transit communications.

8.4 RCC network

A typical RCC network (RCCN) concurrently supports both star and peer-to-peer topologies in order to allow communication between the control center and various endpoints, as depicted in Figure 25.

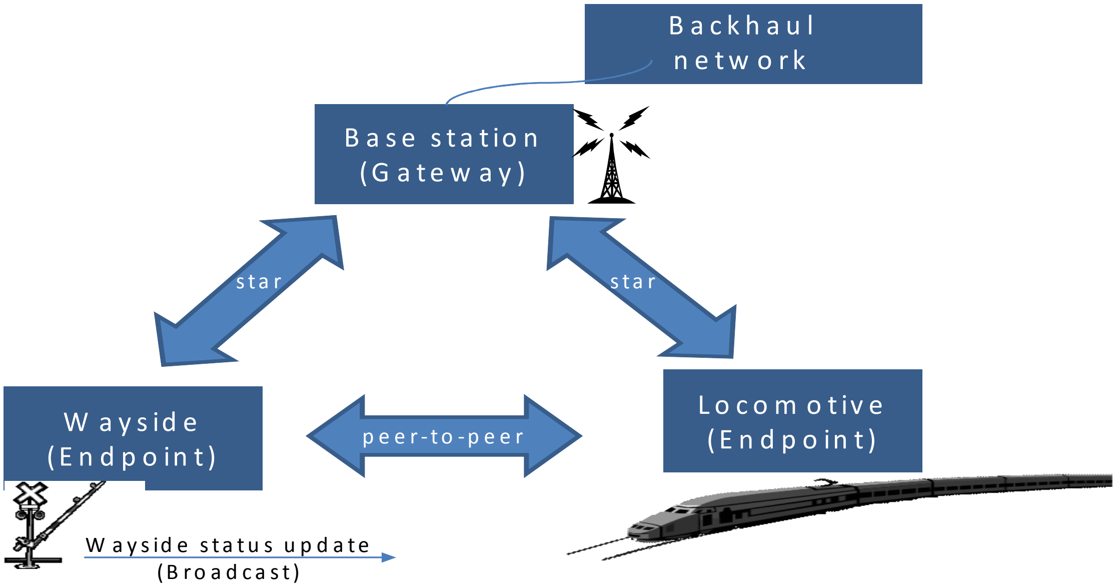

Figure 25—A typical RCCN

AAn RCCN base station is a fixed device that provides gateway access to the backhaul network. RCCN endpoints are mobile (e.g., a locomotive) or fixed (e.g., a wayside device). Communication between the fixed endpoints and mobile endpoints are either through a base station or directly peer-to-peer. A base station is not always in range of the endpoints.

When a base station is available, it acts as an RCCN PAN coordinator, transmitting a periodic beacon and defining an RCCN superframe, as described in 802.15.4.

When an RCCN endpoint is not receiving beacons from an RCCN PAN coordinator, the RCCN endpoints communicate directly using contention access.

Annex A

Bibliography

Bibliographical references are resources that provide additional or helpful material but do not need to be understood or used to implement this standard. Reference to these resources is made for informational use only.

[B1] Public Law 107–56 (42 U.S.C. 5195c(e)), Section 1016(e), Critical Infrastructure Protection Act of 2001, October 2001.

[B2] ETSI EN 300 220-1, Electromagnetic compatibility and Radio spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment to be used in the 25 MHz to 1 000 MHz frequency range with power levels ranging up to 500 mW; Part 1: Technical characteristics and test methods.

[B3] Biswas, P,. and Ye, Y., “Semidefinite programming for ad hoc wireless sensor network localization,” *IEEE Conference on Information Processing in Sensor Networks*, pp. 46–54, Apr. 2004.

[B4] Doherty, L., Pister, K. S. J., and. El Ghaoui, L., “Convex position estimation in wireless sensor networks,” *IEEE Conference on Information Theory and Communications*, pp. 1655–1663, Apr. 2001.

[B5] Gentile, C., “Distributed sensor location through linear programming with triangle inequality constraints,” *IEEE Conference on Communications*, June 2006.

[B6] Gentile, C., “Sensor location through linear programming with triangle inequality constraints,” *IEEE Conference on Communications*, pp. 3192–3196, May 2005.

[B7] Guvenc, I., and Sahinoglu, Z., “Threshold-based TOA estimation for impulse radio UWB systems,” *IEEE International Conference on Ultra Wideband Systems and Technologies*, pp. 420–425, Sept. 2005.

[B8] Guvenc, I., and Sahinoglu, Z., “Threshold selection for UWB TOA estimation based on Kurtosis analysis,” *IEEE Communication Letters,* vol. 9, no. 12, pp. 1025–1027, Dec. 2005.

[B9] Hach, R., “Symmetric double sided two-way ranging,” IEEE P802.15 Working Group for Wireless Personal Area Networks (WPAN), Doc. IEEE P.802.15-05-0334-00-004a, June 2005.

[B10] Lee, J.-Y., and Scholtz, R. A., “Ranging in sense multipath environment using an UWB radio link,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, Dec. 2002.

[B11] Niculescu, D., and Nath, B., “Ad hoc positioning system (APS),” *IEEE Conference on Global Communications*, pp. 2926–2931, Nov. 2001.

[B12] Qi, Y., Kobayashi, H., and Suda, H., “On time-of-arrival positioning in a multipath environment,” *IEEE Transactions on Vehicular Technology*, 2006.

[B13] Qi, Y., and Kohno, R., “Mitigation of sampling-induced errors in delay estimation,” *Proceedings of the IEEE International Conference on UWB 2005 (ICU2005)*, Zurich, Switzerland, Sept. 2005.

[B14] Saverese, C., Rabaey, J. M., and Beutel, J., “Location in distributed ad-hoc networks,” *IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 2037–2040, May 2001.

[B15] Savvides, A., Park, H., and Srivastava, M. B., "The bits and flops of the N-hop multilateration primitive for node localization problems," *ACM Conference on Wireless Sensor Networks and Applications*, pp. 112–121, Sept. 2002.

[B16] Shang, Y., Rumi, W,. Zhang, Y., and Fromherz, M. P. J., "Localization from mere connectivity," *ACM Conference on Mobile Ad Hoc Networking and Computing*, pp. 201–212, June 2003.

[B17] IEEE Std 1474.1™-2004, IEEE Standard for Communications-Based Train Control (CBTC) Performance and Functional Requirements.

[B18] ARIB STD-T96, 950 MHz-Band Telemeter, Telecontrol and Data Transmission Radio Equipment for Specified Low Power Radio Station, 2010.7.15 (H22.7.15) Version 1.1.

[B19] English translation of ARIB STD-T96, 950 MHz-Band Telemeter, Telecontrol and Data Transmission Radio Equipment for Specified Low Power Radio Station, 2008.6.6 (H20.6.6) Version 1.0.