CHAPTER 33

AUTOMATED WILDLIFE RADIO TRACKING

Robert B. MacCurdy,[1] Richard M. Gabrielson,[2] and Kathryn A. Cortopassi[2]
[1]Department of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY
[2]Cornell Lab of Ornithology, Cornell University, Ithaca, NY

THE CHAPTER begins with a discussion of currently available wildlife tracking systems and explains why tag mass is the primary design constraint. Current manual direction-finding methods are described, as are several automated implementations. We also discuss the need for generic asset (non-wildlife) tracking tags that are light and inexpensive, and review current asset tracking methods based on cellular and satellite platforms in this context. The shortcomings of existing systems motivate the need for a new approach that offers global positioning system (GPS)-like accuracy, with vastly reduced energy consumption. Terrestrial *time-of-arrival* (TOA) tracking methods are discussed as a lower energy cost solution, with specific sections dedicated to explaining the following concepts and their interplay in an integrated system: *code division multiple access* (CDMA), matched filtering, estimating location with TOA, and efficient signal processing with a digital signal processor (DSP). The section on CDMA and matched filtering introduces basic concepts including spectrum utilization, autocorrelation, cross correlation, signal-to-noise ratio (SNR), link budget, and processing gain. The TOA section describes several different approaches to calculating location from arrival-time measurements, including the estimation methods employed by GPS, the canonical crossed spheres or hyperboloid techniques, and a method that we developed based on stochastic spatial search. We compare the performance of these methods using real and

Handbook of Position Location: Theory, Practice, and Advances, First Edition.
Edited by Seyed A. (Reza) Zekavat and R. Michael Buehrer.

Published 2012 by John Wiley & Sons, Inc.

simulated data. The signal processing section details the computational requirements of real-time matched filtering, including the impact of Doppler shift. We describe several techniques used in order to implement real-time TOA receivers in embedded devices with limited computing resources, including the use of frequency-domain processing via the fast Fourier transform (FFT), and the intelligent reuse of data via time-shifting techniques. The chapter concludes with a summary of the current performance of a TOA wildlife tracking system that we implemented, its design limitations, and likely areas for improvement.

33.1 INTRODUCTION

Radio direction-finding techniques have been widely employed by the wildlife tracking community because they offer powerful, flexible tools for monitoring animal movements and behavior. Reductions in the size and power consumption of GPS chipsets have recently allowed GPS location-finding techniques to also be applied to wildlife monitoring. Despite their successes, these approaches still have significant shortcomings, primarily due to the energy constraints imposed by the allowable mass of the electrochemical battery that can be carried by the animal. This requirement causes tag lifetimes to be shorter than desired. Attaching a tracking collar is a risky procedure for all participants, and maximizing the tag service intervals is extremely important. This is true even for large animals, which can carry significant tag mass without behavior disruptions. Therefore, limiting disturbances to the animal (which are primarily driven by tag energy consumption), followed closely by cost, are the primary design requirements. These requirements motivate a new tracking system based on TOA measurements. This system is similar in many respects to GPS; the primary difference is that in this system, the mobile asset to be tracked emits rather than receives signals. Although transmitting a radio frequency (RF) signal is often the most power-intensive operation for a tag, this choice yields a system with average tag energy requirements that are lower than any current radio tracking method. Though this chapter focuses on the application of this technology to monitoring animal movements, the same set of design criteria apply to generic asset tracking, and we believe that there is a universal need for a local terrestrial tracking system that offers precise positioning with tiny, cheap, long-lived tags.

33.2 A REVIEW OF WILDLIFE TRACKING TECHNIQUES

Radio tracking has been widely used to monitor wildlife movements since the 1960s [2, 38] with countless scientific papers published using some variant of this method. In the majority of these studies, an operator in the field monitors received signal strength while manually changing the orientation of a directional receiving antenna. The direction yielding the maximal signal strength is recorded as a pointing vector to the tagged animal. This simple method is adequate to guide a researcher to the

location of a focal individual, and triangulation using two or more receiving stations can be used to track a few individuals simultaneously. However, this method yields relatively few position fixes per hour and fully absorbs an operator's attention and effort. Automatic or supervised tracking systems have been developed using stationary receiving towers [3–6] in an attempt to increase the number of animals that can be tracked simultaneously. Most efforts involve directional antennas and rely on the beam pattern of the antennas to infer a direction of arrival. These approaches generally show error in the 1°–10° range, depending on the implementation [5, 7–9], and the cross-bearing positional error for each receiving station increases linearly with range. An excellent, comprehensive manual for applying current radio tracking tools to the study of wildlife can be found in Reference 10. Several other technologies, including satellite-based transmitters or receivers, cellular communications, solar geolocation, and radar have been used to study wildlife movements; these methods are discussed later in this section.

33.2.1 Wildlife Tag Design Constraints

Electronic tags offer the possibility of monitoring animals in their native habitat, with minimal disturbance. If this capability is to be realized, these devices must be unobtrusive. Wildlife tracking tags could also enable vastly larger study sample sizes than methods that use unaided observations by field personnel; however, the cost of the equipment required must not prohibit its use. We consider these two parameters: disturbance to the animal and cost, to be the primary factors in the design of wildlife tracking tags, and they motivate all design choices. Animals are disturbed by the use of tags in two primary ways: (1) They are captured in order to apply the tags, and (2) they must then accommodate carrying a foreign object on their body. The relative impacts of these two factors differ depending on the animal being studied. For example, elephants are capable of carrying a large tag on a collar around their neck; however, the process of anesthetizing the animal and applying the tag is dangerous for the animal as well as for the researchers. In contrast, small birds can be easily captured in mist nets, but they are only capable of carrying a small percentage (no more than 2.5–5.0%) of their body mass as additional payload [11, 12]. While the factors differ, both cases require carefully minimizing the energy used by the tag's electronics: The elephant can carry a large battery, but long intervals between servicing the tag are desired, while the bird can carry little mass and therefore requires a small battery.

Bird body mass varies widely, and though no one tag design will be appropriate for all birds, tag designers should strive to make their tags as widely applicable as possible, within functional limits. Figure 33.1 [13] shows a strong peak in bird body mass distribution and, when combined with the loading heuristic mentioned previously, implies that a tag between 1 and 9 g can be safely applied to roughly 50% of all bird species. This metric can be used as an upper limit on the mass of a general-purpose bird-tracking tag. Though larger tags have utility in some specific applications, the majority of migratory bird species cannot carry them.

Limitations on tag mass apply to species other than birds. Even though larger animals are capable of carrying tags with heavier batteries that provide long

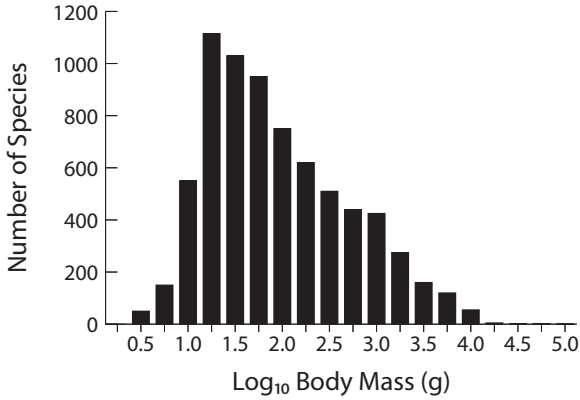

Figure 33.1 Bird body mass distributions. Data from Gaston and Blackburn et al. [13].

runtimes, there may be compelling reasons to limit tag mass. Whale tagging provides an apt example. Whale tags are often applied via ballistic darts because (1) this method is far safer and less intrusive than capturing and anesthetizing an animal (if possible at all), and (2) the thick blubber present in many whales provides a sound anchor for the dart. The darts are delivered by a crossbow, so the overall mass of the dart/tag system must be limited. Many large land mammals could carry a small dart or tag affixed to the ear more easily than the current collar attachment; cattle have long worn plastic ear tags for identification. The primary issue is mass: These tags must be no more than several grams for this approach to succeed.

Tag lifetime, in addition to tag mass, factors strongly into the relative disturbance to the animal and the scientific utility of the method. The movements of migratory birds are strongly tied to seasonal changes, and migration behavior is becoming recognized as an important indicator of climate change [14]; however, recording migrations requires equipment with a useful lifetime of at least 1 year.

Electronic tags offer the potential to dramatically increase the study sample size achievable with a small research team; however, this capability is hindered by cost. Current research tools generally impose severe cost constraints on the study, either in the form of high individual tag costs (as is the case with GPS or Argos) or high labor costs to track and maintain the tags. An automatic wildlife tracking system that used $200 tags would allow its users to track an order of magnitude more animals than a comparably priced system using GPS or Argos tags. Though the cost of installing receivers is obviously greater when using a terrestrial system, this fixed cost becomes negligible if the system can handle large numbers of transmitters.

The design constraints mentioned in this section: cost, mass, energy consumption, and lifetime, are interrelated and must be carefully balanced throughout the design process. Though the examples and motivation that we provide are specific to wildlife tracking, nearly identical constraints apply to mobile asset tracking, since small, low-cost and long-lived devices distinguish successful practical systems from those that work in narrowly defined applications.

Figure 33.2 A 110-mg tag, adapted from Reference 40, built by Julian Kapoor at Cornell University.

33.2.2 Terrestrial Wildlife Transmitters

Early wildlife tracking transmitters [2, 38] began to achieve acceptance in the 1950s and their use accelerated in subsequent decades. The first tag designs used RF tank circuits for frequency control, with a single active element to drive the system into oscillation. These devices broadcast a single frequency carrier, with rudimentary *on-off keying* (OOK) modulation. Subsequent designs employed crystal resonators to achieve tighter frequency specifications and added additional output amplification stages to increase the output power (in order to increase range). Incremental refinements in the intervening 50 years have yielded transmitters that can be extremely light (see Fig. 33.2) and are generally low cost; commercial tags using this technology usually cost no more than $250. Despite these improvements, the underlying technology has remained essentially unchanged: A carrier frequency oscillator circuit is turned on and off by a secondary timing circuit with a period between 2 and 60 seconds. The carrier is typically turned on for a short transmit pulse lasting approximately 20 ms and is off for the duration of the period. The resulting transmit duty cycle determines the tag's average energy consumption. Shorter transmit pulses or longer intervals between transmissions can reduce the average energy consumption but make the tag more difficult to identify and track. Individual tags that will be used in overlapping geographical regions must be assigned unique operating frequencies via crystal selection. These channels are typically spaced in 5- or 10-kHz intervals. Typical carrier frequencies for wildlife tracking range from 140 to 225 MHz. Several bands in this range have historically been reserved for narrowband amateur use, and a few are allocated exclusively for wildlife tracking. In addition to OOK, some transmitters modulate data using *frequency modulation* (FM). This method has been used to telemeter numerous types of analog data, including heart rate and acoustic information.

The tag designs just described have been successful for so long because they use relatively few components and yield tags that are inexpensive and simple to use. Unfortunately, the approach also uses the energy required to send RF transmissions

inefficiently, since the information content of the signal is low and because the signal is transmitted frequently but is rarely actually received. The signal is received infrequently (relative to the number of transmissions) because it would be extremely arduous for human observers to monitor a radio receiver continuously over the lifetime of a tag (tag lifetimes range from weeks to years). Though automated receivers can monitor a single tag frequency continuously, no systems have been built that can monitor multiple tag frequencies simultaneously. This is discussed in the next section.

33.2.3 Terrestrial Wildlife Receivers

The evolution of wildlife tracking receivers has followed a trajectory similar to that of tags: A proven design has been continually refined, with few fundamental changes. These receivers typically use a conventional narrowband heterodyne architecture with analog components, and employ either an FM detector/decoder or a tunable *beat frequency oscillator* (BFO) in the final down conversion stage. The BFO approach mixes the signal at the intermediate frequency with a tunable oscillator that is 1 or 2 kHz away from the intermediate frequency. When a carrier is present, it is mixed into an audio frequency that is easily heard. This approach is simple, reliable, and yields impressive overall sensitivity when paired with a trained operator. However, it requires that the signal being received has no complex modulation and must last long enough for the human operator to hear. In practice, this means about 20 ms, though some tags use slightly shorter transmitter pulse lengths.

Handheld Receivers: Handheld receivers constitute the vast majority of wildlife tracking systems in use. Locating and tracking a tagged animal using a conventional handheld receiver requires an operator to tune the receiver to the channel allocated to the tagged animal. The operator sweeps a directional antenna through all expected bearings while listening (typically through headphones) for a "beep" from the receiver. The audible tone is generated in the BFO circuit, and the system relies on a human listener for detection. This is why the tag's transmissions must be relatively lengthy; the human observer's ear integrates the acoustic signal, so longer tones sound louder. When a signal is detected, the operator makes fine adjustments to the bearing of the antenna while listening to the amplitude of the beep in order to determine the actual line of bearing to the animal. When the direction to the animal has been established, the operator moves to a second location along a baseline roughly perpendicular to the original bearing in order to make a second bearing measurement. The distance between the two locations should be large enough that the second bearing measurement is significantly different from the first. The animal's location is then estimated by the intersection of the two lines of bearing, a process known as triangulation. The operator can also simply follow successive lines of bearing to the animal if the goal is to approach the animal. This approach, though time tested, leaves much to be desired. It is labor-intensive and slow, which limits the number of animals that can be studied and the amount of position data that can be gathered. The cost of achieving round-the-clock observations is prohibitive. It also requires many tag transmissions to achieve a single position estimate.

Automatic Receivers: Automatic receivers have been developed to increase the number of animals that can be simultaneously tracked and to reduce the labor cost of the effort. These devices are functionally very similar to handheld receivers; however, they typically employ a microcontroller to perform the channel scanning and signal detection operations. The simplest automatic receivers do not attempt to estimate transmitter direction at all and are used to determine the presence or absence of a tag within a given detection radius [4]. Directional antennas can be added to these receivers, and the presence/absence information is associated with a particular range of bearings (the main "lobe" of the antenna), which yields a rudimentary location. More sophisticated receivers employ multiple directional antennas whose main sensitivity lobes are uniformly distributed around 360°. A specific variant of this, known as the crossed Adcock antenna, uses two pairs of matched dipole antennas. Each pair of antennas is connected to a phasing element that combines their individual outputs into a single output. The output of each phased pair of antennas reaches a maximum when the phase of the incident signal is equal at both antennas and a minimum when the phase differs by 180°. The two antenna pairs are arranged in a cross, and the signal strengths of the two outputs are compared in order to establish signal direction. Receivers with more antennas have been successfully used; in general, directional receivers of this type use the relative signal strength at each of the directional antennas to establish a line of bearing when a transmitter is detected. Networks of these receivers are established in a study area, and when multiple receivers detect the same transmission event, they can locate the transmitter by intersecting their estimated lines of bearing [3–6, 8]. Digitally steered phased-array approaches have also been employed to determine tag bearing. These methods use multiple antennas, usually in a circular or linear array, and establish signal direction by measuring the phase difference of the signal at each receiving antenna. This method usually requires multiple synchronized receiver signal paths; modern directional receivers usually accomplish the task with high-speed synchronous analog-to-digital converters (ADCs) operating at the IF stage of the receiver. The sampled multichannel signal is then digitally down-converted and the bearing is established via software; the multiple signal classification (MUSIC) [15] algorithm is widely used for this purpose. The complexity of this approach can be problematic for wildlife tracking applications, and receivers of this type are costly. Additionally, phased-array receive antennas must be mounted far from other objects; experiments conducted by our group showed significant variation in signal phase due to nearby vegetation. This constraint requires phased-array antennas to be mounted on tall, sturdy masts rather than opportunistically placed in trees.

A less expensive variation of the phased-array receiver, borrowed from the amateur radio community, uses multiple antennas but only a single receive channel. This receiver, known as a pseudo-Doppler direction finder, arranges the antennas around the circumference of a circle whose diameter is half the wavelength of the received signal. A many-to-one multiplexer sequentially selects each antenna around the circle as the input for a single-channel FM receiver with a phase-locked loop (PLL) detector. As each antenna is switched in, different phases of the signal are presented to the PLL detector, and the phase changes produced each time the receiver is switched from one antenna to the next cause the PLL to produce output pulses

that are proportional to the magnitude of the phase change. This pulse train, when low-pass filtered, has a sinusoidal shape, and the relative phase between this sinusoid and the antenna switching signal provides an estimate of the incident signal's bearing. Though less complex (and far less expensive) than the multichannel receiver described previously, pseudo-Doppler directional receivers generally sacrifice sensitivity for simplicity.

Though receivers that automatically detect transmitter direction have been successful in reducing the labor cost of tracking and can increase the number of location estimates per unit time, they have drawbacks. Their detection sensitivity is lower than that for a system with a human operator, which reduces the radius of detection, and their bearing estimate is several times less accurate than when using a human-rotated antenna.

Most automated receivers operate by continuously scanning a sequence of narrowband channels for tag transmissions. This approach is adapted from the manually tuned receivers that preceded them and restricts the number of channels that can be monitored per unit time. The primary challenge is synchronization with the tag's transmissions, since a typical tag transmits for a few milliseconds every few seconds, while the receiver has to scan tens to hundreds of channels, depending on the number of tags that must be accommodated. The tag must transmit while the receiver is listening for its carrier frequency, an occurrence that becomes increasingly unlikely as the number of channels to be scanned increases. This situation can be partially mitigated, on average, by randomizing the channel scan sequence or by ensuring that a particular channel's scan period is not a multiple of the tag's transmit period. Even so, the scanning approach breaks down when the channel count exceeds 30–50, with many tag transmission events going unnoticed by the receiver. An alternative approach selects each channel for a long enough period to guarantee that a tag will transmit several times and then moves to the next channel. Though this method guarantees that all tags within range will eventually be heard, it also guarantees that most tag transmissions will not be detected. These undetected tag transmissions represent system-level energy inefficiency, and the energy is wasted where it can be tolerated least: in the battery-powered mobile tags.

33.2.4 Satellite Tracking Systems

Two satellite-based systems are widely used for wildlife tracking: GPS and Argos. These two systems provide location information using different techniques. GPS employs a network of orbiting satellites which broadcast signals to terrestrial receivers that use a TOA algorithm to estimate position. Misra and Enge provide a comprehensive review of the GPS system in their book [16]. Tag mass is the primary limiting factor for applying satellite-based systems to wildlife tracking. As mentioned previously, the typical maximum allowable tag-to-body mass ratio is 5%. Commercial GPS tags typically weigh between 22 and 150 g, which limits their application to larger animals (>440 g), and cost between $1500 and $3500. One new, very low-mass (4.5 g) tag is now available [17]; however, the small battery used in order to achieve low mass limits the system to no more than several hundred position fixes. Though continuous refinements have yielded ever smaller and more sensitive receivers, the positioning approach used by GPS makes it unsuitable for ultra-low-

power systems. The principle drawback of GPS is that it does not directly provide a means of reporting position information back to the researcher. The position information is either stored and retrieved later, or downloaded via an auxiliary RF link. The energy cost required to transmit position data from the animal to the researcher is often prohibitive. Additionally, GPS receivers require a relatively low-noise, broadband RF front end, coupled with fast digitizers and signal processing hardware. The power consumption of these elements, integrated over the satellite signal acquisition time, imposes a significant energy demand on the tag's battery because the signal acquisition time can be long. The duration, which depends on several factors, varies from 1 second to 1 minute in modern receivers. A specialized type of GPS logger, which can yield position fixes by postprocessing recorded satellite transmissions, reduces energy consumption by limiting the signal acquisition time to approximately 60 ms [18]. These tags, referred to as fast lock GPS tags, store the raw digital IF data from all satellites in view rather than attempt to acquire each satellite's signal via matched filtering. The trade-off of this approach is that many kilobytes of data must be written to nonvolatile memory. Despite a considerable improvement in energy consumption, relative to conventional GPS, this approach must still transmit the position data if animal recapture is not possible, or if real-time operation is desired. The energy cost of transmitting data from a GPS logger is even higher than for a conventional GPS tag, since the logger must send far more data.

In contrast to GPS's TOA-based operation, the Argos system determines tag (transmitter) position by exploiting the frequency shift in the tag's signal, measured by the satellite's receiver. This frequency shift is caused by the satellite's motion relative to the transmitter and is primarily dependent on the orbital parameters and the earth's rotation. The computation of transmitter location takes place within the Argos satellite control system rather than on the tag, so tag positions are immediately available to the researcher. For a complete description of the Argos system, see Reference 19. The Argos system can achieve reasonably good accuracy; the best service class available advertises 250-m error bounds. However, this level of accuracy is often not available, and the other three accuracy classes range from 250 to 1500 m. The Argos system's link budget requires significant transmit power from the tags; typical tag power consumption varies from 150 to 500 mW during a 300- to 900-ms transmission. These transmissions must repeat with a 90- to 300-second period. These parameters set the minimum energy for operation and necessitate relatively large tags, though some newer models have incorporated solar cells and ultracapacitors to reduce mass. Commercial Argos tags are generally smaller than GPS tags (the lowest mass Argos tag is currently about 10 g) but still only allow animals heavier than 200 g to be tracked. This weight constraint excludes 75% of all bird species [13]. The cost of Argos tags is also prohibitive for large-scale studies. The complexity and low production volumes of these tags lead to typical single-unit prices in the $1500–$4000 range, with little cost reduction at larger volumes.

33.2.5 Solar Geolocation Tracking

Each tracking method mentioned previously has employed a form of RF technology; however, they do so at a price: energy consumption. As shown in Figure 33.1, many birds are so small that their maximum payload is less than 1 g. Remarkably, some

of these tiny animals perform very long migrations and researchers would like to track their movements, yet their size makes the application of continent-scale tracking via GPS or Argos tags impossible. Another approach, using measurements of the time of local dawn and dusk, can be performed at a very low energy cost. The British Antarctic Survey has taken the lead in the application of this technique, known as solar geolocation. Their tiny tags use a microcontroller to keep track of time and to sample the ambient light level every few minutes with a photodiode. The tags create a continuous record of the light level as a function of time (usually, coordinated universal time [UTC] is used as the reference). The time of sunrise and sunset is a function of longitude, while the length of the day is predominantly a function of latitude (each is also affected by the time of year). This method can produce global position estimates with an accuracy better than 100 km, which is sufficient for determining migratory flyways and critical stopover points.

33.2.6 Cellular Tracking

Despite several decades of growth and near-ubiquitous availability (with frustrating exceptions), cellular telephone operators in the United States have historically been reluctant to open their networks to nonvoice traffic. This has begun to change recently, as the network operators recognize the market potential for *machine-to-machine* (M2M) communications. They are beginning to offer nonvoice SMS and à la carte data rate plans that are suitable for M2M use. At least one wildlife tracking company, Cell Track Tech of Rector, Pennsylvania, has begun to use cellular communications technology to enable real-time data downloads and software updates for deployed tags. Their lightest tag weighs about 50 g, making it suitable only for larger birds. Nevertheless, these tags allow researchers to monitor the movements of migratory birds within the global GSM communications network and can archive GPS data for future download when no communication network is available. Cellular data connectivity is now even available to hobbyists via a line of cellular communications modules manufactured by Telit Wireless, an Italian company. Their products can be easily connected to a microcontroller, and abstract the specific details of the cellular network from the developer.

33.2.7 Radar Tracking

Migratory birds have long been visible on weather and transportation radar systems. At times, the considerable backscatter from large flocks of birds or even insects can be a nuisance for the operators. Biologists, however, have begun to use this information to study migratory species [20]. Though individuals cannot be followed, flocks of birds can be clearly tracked over hundreds of square miles, using the existing Doppler radar infrastructure. Smaller, portable radars have been used with extremely small tags, some weighing less than 12 mg, in order to track individual flying insects [21–23]. The tags employ passive nonlinear switching elements in the antenna that reflect a frequency-doubled version of the incident signal. This coherent signal is easily detected by the radar. This approach has been used to determine range and bearing over half-kilometer distances.

33.2.8 Summary and Motivation for Improvements

Although they have been separated into two categories in this description, wildlife receivers and tags should be viewed as integral components of the same system; one is not useful without the other, and neither can be substantially modified without impacting the design of the other. This linkage causes discrete design choices to propagate throughout the system. For example, the use of narrowband transmitters, which has historical as well as practical underpinnings, limits the application of signal processing techniques at the receiver. As a consequence, the tags must transmit long sequences relatively often in order to satisfy the link budget, with an attendant energy cost. The use of narrowband signals also reduces the options for mitigating multipath interference, which can cause errors in the estimated position. Tags are usually designed without a microcontroller in the name of simplicity; however, this limits them to very simple control schemes (e.g., on for 20 ms, off for 2 seconds) that cannot turn off the transmitter during lengthy periods in which tracking will not occur (at night, or when the animal is in a remote segment of its migration). Channel scanning in the receivers causes many transmissions to be missed, which wastes tag battery energy.

33.3 A NEW APPROACH TO WILDLIFE TRACKING

Although traditional narrowband radio tracking systems work well when a single researcher is following a small number of animals, this technology does not lend itself to automation. An alternative approach, using TOA information captured by fixed terrestrial receivers, offers the potential for significant improvements when tracking wildlife within limited regions. The TOA approach, which can be imagined as "inverse GPS," uses small tags that periodically transmit a short carrier that is modulated by a *pseudorandom noise* (PRN) code. A network of nearby receivers continuously listens for tag transmissions and records the arrival time when a transmission is detected. Each of the receivers sends its arrival-time measurement to a centralized server where the transmitter position is computed. This system dramatically reduces tag energy consumption by using very short RF transmissions and by sending the transmissions infrequently. TOA receivers should be capable of detecting a transmission from any tag in the system at any time, so the tags only need to transmit as frequently as a position update is desired. In addition, this requirement avoids any costly synchronization or registration process between the tag and the receiver (the fixed receivers maintain synchronization with each other instead).

Despite (or perhaps because of) the long success of GPS, relatively few TOA-based terrestrial tracking systems have been built. One business, Recon Dynamics of Kirkland, Washington (which acquired the technology from S5 Wireless), is attempting to commercialize a system that uses small transmitters and TOA measurements for asset tracking. A prototype system using TOA to track flying foxes was developed by researchers in Australia [24]. A similar system, using fairly powerful transmitters, was developed to track moose in Sweden [25]. An early vehicle tracking system from Teletrac used terrestrial TOA spread spectrum techniques (though they now use GPS with terrestrial data links). Our group at Cornell has designed,

built, and installed a prototype system based on CDMA and TOA that is capable of automatically locating thousands of tags in real time. In addition, the tags utilize onboard microcontrollers that can implement sophisticated calendar functions; the tags can be put into long periods of "deep sleep" and awakened only when the animal is expected to be within range of the system. This system borrows concepts from the GPS system in general and from *pseudolites* in particular. Pseudolites (or pseudosatellites) are terrestrial devices that transmit GPS signals. An excellent review can be found in Reference 26. TOA-based tracking systems rely on several concepts, including PRN codes, CDMA, matched filters, and digital signal processing. These concepts, along with their applications in TOA tracking systems, are discussed in the following section. For comprehensive resources, see References 16, 27, and 28.

33.3.1 PRNs

Radiotracking equipment has traditionally employed simple modulation schemes, such as OOK or FM, because they are easy to implement with simple analog circuits; however, this simplicity comes at a cost. The signals used are not readily distinguished from noise, and they interfere with each other. Two adjacent tags that use OOK and share a carrier frequency will be difficult to differentiate if they transmit simultaneously. Using PRN signals can reduce these problems. Their autocorrelation properties allow detection and precise synchronization, even when they are contaminated by significantly more powerful noise. Additionally, certain PRN sequences have guaranteed cross-correlation behaviors that allow many different signals to coexist with minimal interference.

We begin the discussion of pseudorandom sequences with a description of random sequences. Consider the random sequence snippet shown in Figure 33.3, which has nine entries whose magnitudes are either 1 or −1. Each of these entries represents a bit (as indicated in the figure), has a duration, a magnitude, and is called a *chip*. The duration of each of the chips is T_C seconds; the entire sequence has N entries; and the sequence's total duration is T_N seconds. The normalized *cross correlation* $C(\tau)$ of two signals, $x(t)$ and $y(t)$, is defined in Equation 33.1, where the superscript * indicates the complex conjugate operator. Cross correlation can be thought of as a measure of how similar two different signals are as they are shifted past each other in time. We define a similar *autocorrelation*, $R(\tau)$ of $x(t)$, by setting $y(t) = x(t)$ in Equation 33.1:

$$C(\tau) = \frac{1}{2T_N} \int_{-T_N}^{T_N} x^*(t)\, y(\tau + t)\, dt. \tag{33.1}$$

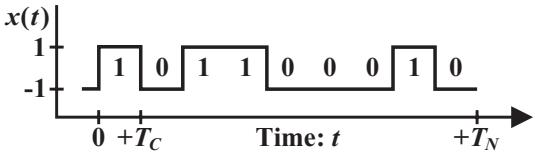

Figure 33.3 A random sequence snippet, $x(t)$, with chip duration T_C and period T_N.

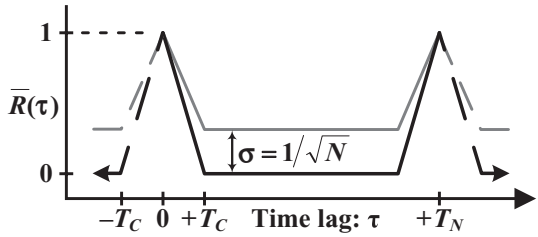

Figure 33.4 The mean value $\bar{R}(\tau)$ of the autocorrelation of the random sequence $z(t)$. The standard deviation of the autocorrelation is represented as an envelope (in gray).

The autocorrelation can be thought of as the degree to which a signal is similar to time-shifted replicas of itself. The magnitude of the time shift is usually referred to as a time *lag*, and has the same units as the domain of the signal that was shifted (in this case, seconds). The mean value of a portion of the autocorrelation function of a signal, $z(t)$, composed of a repeated N-length random sequence snippet, is shown in Figure 33.4 (the N-length sequence $x(t)$ is a finite-length snippet from an infinite-length random signal, and copies are concatenated to form $z(t)$). The mean is shown because the actual autocorrelation values depend on the particular random sequence chosen; however, the mean value of the autocorrelation function of *any* random sequence can be succinctly written in Equation 33.2 [16]:

$$\bar{R}(\tau) = \begin{cases} \dfrac{\tau}{T_C} + 1 & \text{if } -T_C < \tau < 0 \\ \dfrac{-\tau}{T_C} + 1 & \text{if } 0 \le \tau < T_C \\ 0 & \text{else} \end{cases} . \tag{33.2}$$

Notice that the peak in this function occurs at the zero lag. Also note that the function has a prominent peak whose width is equal to twice the chip duration. A sharp peak allows precise time alignment and aids detection, as we will see later. The standard deviation σ of $R(\tau)$ is given in Equation. 33.3 and is depicted as an envelope (gray line) of $\bar{R}(\tau)$ in Figure 33.4:

$$\sigma = \begin{cases} 0 & \text{if } \tau = 0 \\ \dfrac{1}{\sqrt{N}} & \text{else} \end{cases} . \tag{33.3}$$

Although random sequences have appealing autocorrelation properties, the cross correlation of any two distinct equal-length random snippets is not guaranteed to be small. In the worst case, two random codes could differ by a single chip and would have a maximum cross-correlation value close to one, making them very difficult to differentiate in the presence of noise. Additionally, generating a truly random sequence in a simple piece of hardware is not an easy task. For these reasons, pseudorandom noise generators were developed.

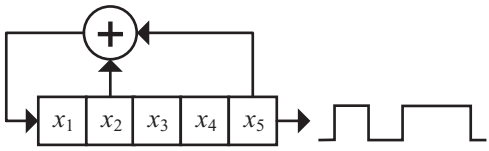

Figure 33.5 A shift register circuit for generating PRN sequences.

These devices generate periodic signals that share many properties with random signals but are easy to implement. A PRN generator can be constructed by performing modulo-2 summation of multiple taps on a shift register and then feeding the result back into the shift register. Figure 33.5 shows a simple example of this approach. The output of this generator will repeat after a certain number of chips, referred to earlier as N. If the position of the taps is properly chosen, N = 2m − 1. This arrangement, known as a maximal-length generator, yields the longest codes possible from a shift register with m cells.

There are many variations on the basic shift register feedback configuration, and they yield sequences with different properties. An important architecture, known as a Gold code generator [29], offers whole families of deterministic codes that provide beneficial autocorrelation properties similar to random codes, but guarantee that the cross correlation of any two member codes will be below a threshold. If the time lag τ is constrained to an integer number of chips ($\tau = iT_C$, $i = \{0, 1, 2, \ldots\}$), the cross correlation of any two different Gold code sequences of length N takes on only three values (Eq. 33.4) [16]:

$$C(\boldsymbol{\tau}) = \left\{ \frac{-1}{\boldsymbol{N}}, \frac{-\boldsymbol{\beta}(\boldsymbol{m})}{\boldsymbol{N}}, \frac{\boldsymbol{\beta}(\boldsymbol{m}) - 2}{\boldsymbol{N}} \right\}, \beta(m) = 1 + 2^{\text{floor}\left(\frac{m+2}{2}\right)}. \tag{33.4}$$

These values are shown for the case when $N = 1023$ in Equation 33.5,

$$C(\boldsymbol{\tau}) = \left\{ \begin{array}{ll} \dfrac{63}{1023} & 10\log_{10}\left|\dfrac{63}{1023}\right|^2 = -24\ \text{dB} \\ \dfrac{-65}{1023} \quad \text{or} & 10\log_{10}\left|\dfrac{-65}{1023}\right|^2 = -24\ \text{dB} \\ \dfrac{-1}{1023} & 10\log_{10}\left|\dfrac{-1}{1023}\right|^2 = -60\ \text{dB} \end{array} \right\}, \tag{33.5}$$

which uses a magnitude-squared metric for differentiating the side lobes from the peak. The expression for the autocorrelation of any Gold code sequence includes the three values in Equation 33.4, and adds a fourth: 1 or 0 dB, which corresponds to $\tau = 0$ or "zero lag." Therefore, if two different Gold codes (from the same family of $N = 2^m - 1$ length codes) are transmitted simultaneously in the same region, with equal signal power, they can be readily distinguished both from random noise and from each other. Finally, there are $2^m + 1$ Gold codes available from a shift register generator of length m, which allows large numbers of codes with low cross correlation to be easily created (see Reference 28 for helpful Gold code generator tables).

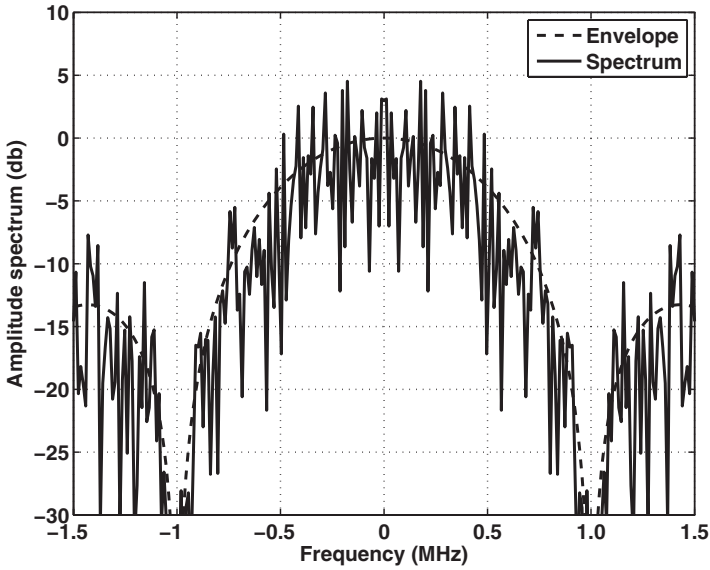

Figure 33.6 Normalized amplitude spectrum of a 1023-chip random noise signal with a 1-MHz chip rate.

This idea forms the basis of CDMA systems, which enable multiple simultaneous transmitters to coexist by assigning each a unique code.

Chip Rate and Bandwidth: The use of CDMA signals has significant implications for the spectrum utilization of a wildlife radio tracking system, since the bandwidth of a CDMA signal is typically several orders of magnitude larger than the signals used by conventional tracking systems. The amplitude spectrum of one period of a random code sequence is given by Equation 33.6,

$$X(f) = \underbrace{T_C\sqrt{N}\,\left|\mathrm{sinc}(\pi T_C f)\right|}_{\text{Envelope}}\left|\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} x_n e^{(-j2\pi fnT_C)}\right|, \tag{33.6}$$

where x_n is the discrete time version of the sequence $x(t)$ sampled at intervals of T_C seconds [16]. This expression has been evaluated for a 1023-chip random noise signal with a 1-MHz chip rate and is plotted in Figure 33.6. The signal amplitude spectrum $X(f)$ is plotted, as is the envelope of the signal. Notice that the 1-MHz chip rate causes the signal to have a 2-MHz-wide main lobe between the two nulls. Though it is not obvious from the figure's log-amplitude scaling, roughly 90% of the signal's power occurs in the 2-MHz-wide main lobe [28]. This fact has practical utility because, although the code's spectral energy extends to $\pm\,\infty$ along the frequency axis, the signal can be band-limited to the 2-MHz main lobe with a filter and will suffer minimal energy loss. Though not identical to Figure 33.6, the spectrum of a PRN sequence is very similar. As Figure 33.6 indicates, modulating the carrier with a random or PRN sequence adds significant bandwidth to the signal, reducing

the signal power at any particular frequency (relative to an unmodulated carrier of equal total power). This feature of direct sequence spread spectrum systems reduces the likelihood that the transmitters will interfere with each other or with conventional narrowband receivers outside the system. Also important to note, though not obvious from the figure, is the fact that modulating a carrier with a PRN sequence actually makes the signal easier to detect than an unmodulated carrier with the same transmitter power and duration. This result, referred to as *processing gain*, will be discussed next. This improvement comes at almost no energy cost to the tag (relative to a narrowband tag) since the primary energy cost during transmission is in the output amplifier stage rather than in the modulation stage.

Detection via Matched Filters: Matched filters exist in many forms, and the term generically refers to the linear filter whose impulse response is the time-reversed replica of the signal to be detected. It can be shown that the matched filter is the optimal linear detector when white Gaussian noise is present. A digital matched filter can be implemented by correlating the incoming signal, which may be heavily contaminated by noise, with an uncorrupted replica (template) of the expected signal. The correlator runs continuously, shifting the incoming signal past the stored template by one sample at each time step. It multiplies aligned samples of the signal and template, and accumulates the result; the output of each time step is a single number that indicates how well the signal and template agree. A detection decision is made when the correlation output exceeds a threshold. Figure 33.7 shows an example of the correlation of a noisy, time-delayed PRN signal with its template. The template in this case is a 31-chip Gold sequence. The incoming signal is delayed

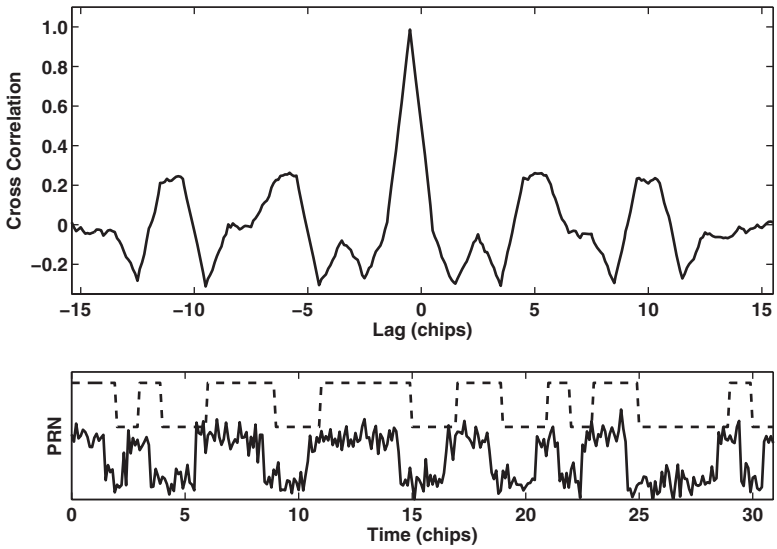

Figure 33.7 Cross-correlation example showing the cross correlation (upper figure) of a received signal with its template. The lower figure shows the time domain signals: matched filter template (dashed line), and noisy and delayed received signal from the transmitter (solid line).

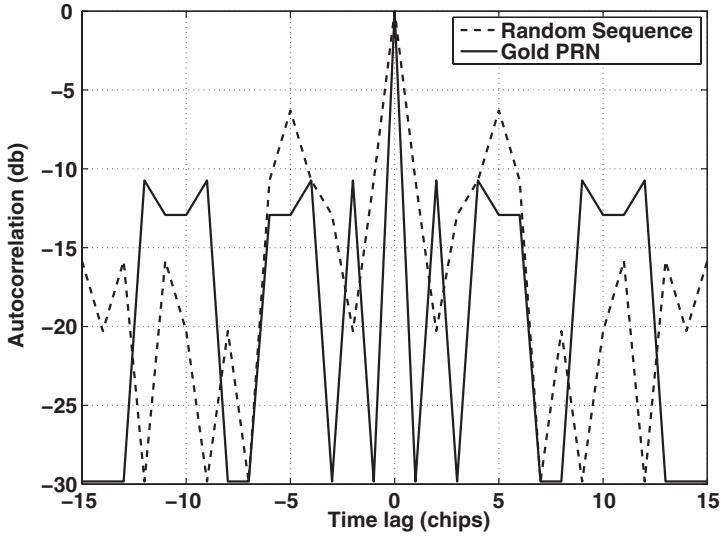

Figure 33.8 Comparison of the autocorrelation functions of two 31-chip sequences: Gold code and random code.

by half a chip, causing the strongest peak to occur at –0.5 lags. The strong peak is easily distinguished from the background noise via thresholding methods. In this case, the four visible peaks to the right and left of the strongest peak are not noise but are the side lobes of the autocorrelation function. The cross-correlation plot in Figure 33.7 reveals an important feature of Gold codes: They provide a guaranteed SNR between the main peak and any side lobes. This relationship is shown exactly in Equation 33.4, and an example of this property is shown in Figure 33.8, which compares the autocorrelation functions of a 31-chip length Gold code and a random code of the same length. The autocorrelation magnitudes are displayed on a log scale, showing the nearly 6-dB advantage in side lobe magnitude that the Gold code has against this particular random code. This difference has a practical significance in signal detection since more prominent autocorrelation peaks yield better SNR.

As stated previously, in addition to providing a sharp correlation peak, which aids precise time synchronization, matched filters offer excellent detection sensitivity. They achieve this sensitivity because the received signal is strongly correlated with the template, while the noise corrupting the signal is not. A system using matched filters for communication effectively replaces each data bit with a chip sequence that is coherently matched at the receiver. Because each bit is represented by a longer sequence of chips, the system is said to yield processing gain. As shown in Figure 33.6, the spectrum of the new sequence is broad, relative to the data that the system is trying to send. The processing gain PG is related to the data rate B_D and the chip rate B_C, or the data bit duration T_D and the chip duration T_C by Equation 33.7:

$$\mathrm{PG} \approx \frac{B_C}{B_D} = \frac{T_D}{T_C}. \tag{33.7}$$

Note that the approximate symbol is used because this is a close approximation, but is not exact for PRN codes like Gold codes. See Reference 16 for specific details. Since each chip sequence has a total duration, $T_D = NT_C$, we can write Equation 33.8:

$$\text{PG} \approx N, \text{ or } \text{PG} \approx 10\log_{10}(N)\text{dB}. \tag{33.8}$$

This convenient result provides an estimate of the improvement in SNR that can be achieved by using a digital matched filter detector of length N chips.

At this point, we have demonstrated the ability of matched filters to provide processing gain; however, we have not said anything about signal detection. Ideally, we desire a method that provides an unambiguous result and detects all true signal transmissions with no false positives. This is a tall order, and in the end, the performance of the system nearly always boils down to SNR—in this case at the output of the matched filter. The wildlife tracking system will need to be able to make detection decisions based solely on the output of the matched filter (no contextual information is available). One approach is to accumulate statistics for the current and several previous cross-correlation buffers, and to use that information to set an adaptive threshold for detections. The GPS system, in contrast, operates under the assumption that a GPS signal is always present to detect, though it may be below the detection threshold. Because of this, GPS receivers can take the signal's history into account during the tracking or acquisition process by tracking with a very narrow loop filter or by applying noncoherent correlation over multiple adjacent data bits.

33.3.2 Signal Processing

The previous section introduced matched filter detectors and described a few core concepts that we have employed in the tracking system's implementation. This section describes a few additional details and addresses some practical design considerations when implementing a matched filter. The first subsection discusses the performance of a matched filter when the received carrier frequency differs from the expected value. The next subsection addresses the computational requirements of the detector algorithm and provides methods to reduce them. The final subsection describes the problems caused by the asynchronous arrival of tag transmissions and invokes the time-shifting property of the Fourier transform to efficiently handle this issue. See Chapter 7 for additional information on the methods used next as well as alternative detection and TOA estimation approaches.

Code Phase Search, Doppler Shift, and Frequency Error: The matched filter detector in a TOA receiver runs continuously, shifting new samples into a buffer as they arrive, and searching for a tag signal by cross correlating the samples with a template. This operation occurs at the *baseband*, after the carrier has been removed completely. At this stage, in theory, the frequency content of the signal is due solely to the PRN sequence, as shown in Figure 33.6.

Unfortunately, the situation is not this simple in reality. Two different sources of carrier frequency uncertainty, clock error and Doppler shift, prevent the carrier from being completely removed. The result is shown in Equation 33.9:

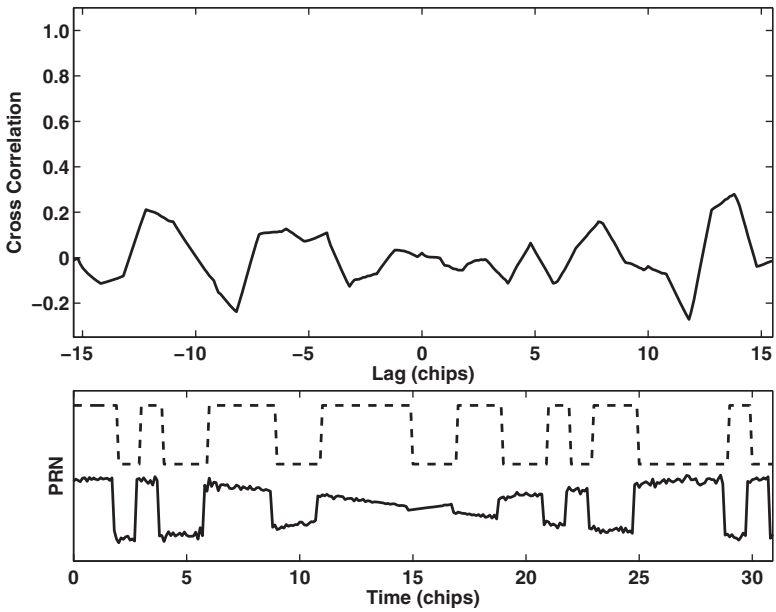

Figure 33.9 Cross correlation between template and received signal with a small carrier and LO frequency mismatch; the cross-correlation peak is eliminated despite high SNR. The lower figure shows the time domain template (dashed line) and the received signal from the transmitter (solid line). Note the low-noise level in the received signal.

$$E\{S\} = Ae^{j(\Delta\theta+\pi\Delta f T_D)}\overline{R}(\Delta\tau)\mathrm{sinc}(\pi\Delta f T_D), \tag{33.9}$$

which shows the expected value of the correlator output signal, S, as a function of $\Delta\theta$, Δf, T_D, and $\Delta\tau$. These parameters are $\Delta\theta$, the difference between the phase of the tag's carrier and the receiver's local oscillator (LO); Δf, the difference between the frequency of the tag's carrier and the receiver's LO; T_D, the duration of the PRN sequence; $\Delta\tau$, the difference in phase between the received PRN sequence and the local template; and A, the carrier amplitude. Assuming that the matched filter finds the correct code phase, which causes $\Delta\tau = 0$ and maximizes $\overline{R}$, $E\{S\}$ might still have a small magnitude if the *sinc* term is small. This situation is illustrated in Figure 33.9, which shows that although the noise level of the received signal is low and the incoming signal is PRN code phase matched to the template, the correlation between these signals is low. This error is caused by a small residual frequency mismatch between the incoming signal and its template. Notice that the sign of the incoming signal is inverted halfway through its length. This causes the correlation disagreements in the second half of the signal to exactly cancel the agreements in the first half. Equation 33.9 provides a simple criterion for ensuring frequency uncertainty does not adversely impact the matched filter detection. The half-power point of the sinc function in Equation 33.9 is reached when $\Delta f T_D \approx 2/5$, so choosing $\Delta f T_D < 2/5$ ensures that a reasonable amount of the signal will always be available to detect. For example, if the PRN duration $T_D = 1.5\,\mathrm{ms}$, we require that the

difference between the transmitter's carrier and the receiver's LO be no more than 266 Hz. This difference could be caused by Doppler shift or by oscillator error. The Doppler shift is shown in Equation 33.10:

$$\Delta f_D = \frac{\Delta v}{c} f_0. \tag{33.10}$$

The velocity difference Δv between the receiver and the animal carrying the tag is assumed to be less than 50 m/s, which yields a maximum Δf_D of about 23 Hz, with a carrier frequency $f_0 = 140$ MHz. Therefore, the Doppler shift is not a significant concern for the animal tracking system. Oscillator error can be bounded by choosing high-precision oscillators, which are readily available. As you can see, choosing relatively short code sequences and accurate clocks allows the sinc term to be ignored in Equation 33.9; however, if longer code sequences are desired, the criterion on $\Delta f T_D$ becomes more difficult to meet, and the detection process becomes a two-dimensional search over code phase and carrier frequency offset.

Computational Requirements and Frequency Domain Operation: The attractive features of a TOA radio tracking system, including low-power tags, automatic detection, and good location accuracy, depend on a network of receivers that can listen continuously and in real time for tag transmissions. The real-time requirement sets a hard limit on performance, which impacts all other design choices. We chose to implement the matched filter detector on a DSP chip that operates at 1 GHz and offers parallel data processing capability, with up to 8 billion *multiply and accumulate* (MAC) operations per second. Despite the powerful DSP chip, the receivers would struggle to maintain real-time operation if they used a straightforward time domain algorithm. To see why, we revisit the cross-correlation function but switch to the discrete time version (Eq. 33.11) (the variables are the discrete time analogs of those in Eq. 33.1),

$$C(k) \equiv \frac{1}{N} \sum_{n=0}^{N} x^*(n) y(k+n), \tag{33.11}$$

which shows that each output sample (lag) requires N multiplications and additions. In our radio tracking system, the incoming RF signal is down-converted to baseband in-phase (I) and quadrature (Q) channels and then sampled at 2.8125 MHz. An11-bit PRN sequence at the tag's 1-MHz chip rate would occupy 5760 samples, and processing the I and Q channels with a straightforward, time domain matched filter implementation would require approximately 32 G MAC per second in order to guarantee real-time operation. This requirement clearly cannot be met by the DSP. The options available are to either reduce the bandwidth of the transmitted PRN signal, which reduces the ranging accuracy, reduce the PRN sequence length, which reduces the processing gain, or to use a frequency domain algorithm to implement the matched filter.

Ordinary time domain correlation is an $O(N^2)$ operation, where N is the number of elements to be cross correlated. Operation in the frequency domain in contrast is approximately an $O(N \log_2 N)$ operation, thanks to the remarkable efficiency of the

FFT. An early application of this technique to GPS was demonstrated in Reference 30. We chose this alternative and used available FFT routines along with the established practice of computing a correlation by conjugate multiplication in the frequency domain to meet the real-time requirements of the system. This approach is shown in Equation 33.12:

$$C(k) = \text{IFFT}\left\{H(f)^{*} G(f)\right\}, \tag{33.12}$$

where * indicates the complex conjugate, IFFT refers to the inverse fast Fourier transform operation, and $H(f)$, $G(f)$ refer to the Fourier transforms of the time series $h(n)$ and $g(n)$. This simple expression masks two important caveats: The signals are finite in duration, since they are stored in RAM, and the signals are not periodic. Real-time operation requires that the FFT lengths be as short as possible, since the processing load scales faster than the length of the buffer to be transformed. As an implementation of the discrete Fourier transform, the FFT assumes that its input data are periodic in N samples, where N is the length of the input buffer. The correlation technique based on the FFT exhibits a circular behavior and will "wrap around" data from the end of one buffer onto its beginning for any lag other than 0 (where the two buffers are exactly aligned). The solution to this problem is to zero-pad the data buffers at their ends. If $\pm k$ lags are desired, then the data buffers must both be padded with k trailing zeros. The *Numerical Recipes* book [31] explains this technique in greater detail.

Time Shifting and Windowing: Operation in the frequency domain offers significant performance improvements but also adds complications. Unlike a time domain correlator, in which incoming samples are continuously shifted and accumulated as data arrive at each time step, frequency domain operation involves processing complete, contiguous blocks of samples, then gathering another whole block and repeating the process. Each block of data represents a "snapshot" or window of the data stream arriving at the receiver. As the previous section demonstrated, we wish to keep the data windows as short as possible in order to reduce the computational load. Of course, the windows must be long enough to at least contain the data from one complete tag transmission, in order to maintain the full autocorrelation of the PRN code. Tag transmissions occur asynchronous to any processing that occurs at the receiver, so the arrival of the first chip from a transmission may fall anywhere in the receiver's window. In the worst case, the incoming signal is misaligned with the buffer boundary by $N/2$ samples, so that the first half of the signal is in one buffer and the second half is in the next buffer. In this case, the matched filter detector will still register a maximum at the $N/2$ lag, but the maximum will be one-half of its autocorrelation value, since only half of the PRN signal is in the buffer. This reduction in signal strength becomes problematic in low signal-to-noise situations. A common solution to this problem is to overlap the buffer by 50%, so that the second half of the last buffer becomes the first half of the next buffer to be cross correlated with the template (in the next iteration of the matched filter). This approach requires twice the processing effort of 0% overlap, and it also computes redundant information, since half of the sample data from the previous cross correlation are present in the next correlation.

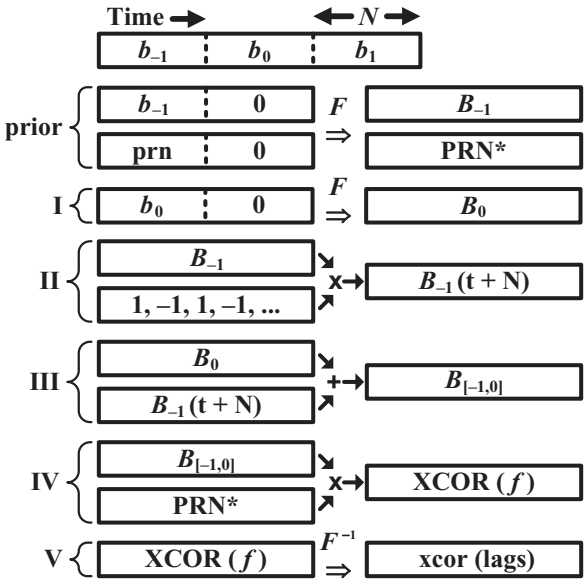

Figure 33.10 Efficient frequency domain cross-correlation algorithm.

An alternative approach exploits the time-shifting property of the Fourier transform, shown in Equation 33.13:

$$x(n-n_0) \overset{F}{\Leftrightarrow} e^{-j\frac{2\pi}{A}mn_0} X(m), \tag{33.13}$$

where x is the sampled time series data, n_0 is the number of samples to shift, A is the length of x, and $X(m)$ is the complex spectra of $x[n]$. This property becomes particularly useful when the time shift is $A/2$, since the complex exponential reduces to the sequence [1, –1, 1, …]. This sequence can be stored in memory rather than computed at runtime. Multiplying $X(m)$ by this simple sequence yields the same spectra as would time shifting $x(n)$ by n_0 seconds and recomputing the Fourier transform. In fact, no actual multiplications need take place at all, since this is merely a sign change on every other data entry.

The complete frequency domain cross-correlation algorithm diagram is shown in Figure 33.10; each conceptual step is identified by a Roman numeral, and time domain data are represented by lowercase letters, while frequency domain data are represented by uppercase. Data arrive from the ADC in a time domain buffer, shown at the top of the figure. This buffer is arranged as a circular buffer and is subdivided into three N-sample segments. The segments are synchronized with the ADC in a way that ensures no samples are changed by the converter while the cross-correlation algorithm is operating on that particular buffer segment. For the sake of clarity, Figure 33.10 was drawn with the assumption that the algorithm always begins at buffer b_0, which has recently been filled with new data. The first step (I) is to compute the discrete Fourier transform of the $2N$-sample buffer that is formed by copying and zero-padding buffer b_0 to length $2N$. The second step (II) time shifts the fre-

quency spectra of the previous buffer, b_{-1}, by multiplication with the simple time-shift sequence [1, −1, 1, …]. Note that the frequency spectra of this buffer were retained from the previous iteration and need not be recomputed. The third step (III) adds the current frequency spectra to the time-shifted spectra from the previous buffer. The fourth step (IV) multiplies the result of (III) by the complex conjugate of the PRN sequence's spectra.

Note that PRN* does not change and can be precomputed. The final step is to convert the cross spectra back into the time domain, yielding the cross correlation. Note that although this buffer contains 2*N* lag entries, only the first *N* are valid. This makes intuitive sense, because we have effectively shifted an *N*-long template inside a 2*N*-long data window. Moving the template beyond *N* lags would cause the template to fall outside the data window, reducing the overlap to below 100%. This block processing algorithm provides an efficient, 100% overlap, frequency domain, digital matched filter.

33.3.3 System Description

The following section delves into the specific implementation of the TOA system that we built. Though certain details are unique to our system, the following ideas can be adapted to any TOA system. We begin with a basic description of the structure of a CDMA-capable transmitter and how we implemented one. The receiver's analog signal chain is discussed next, including the RF and signal conversion components. The final stage in the signal path involves the baseband components, which perform signal detection, timing, and communication. See References 32 and 33 for additional descriptions of the system.

Transmitters: The transmitter is based on an inexpensive, very-low-power microcontroller, along with a precision reference clock, frequency synthesizer, modulator, and amplifier. Our design, shown in Figure 33.11, integrates off-the-shelf components in order to avoid the high cost and long development time of a custom application-specific integrated circuit (ASIC). This choice results in an implementation that is larger than it could otherwise be, but this trade-off allows rapid development. The tag uses a *binary phase-shift keying* (BPSK) modulation scheme to directly modulate and spread the carrier power. Unlike the shift register examples shown earlier, the complete Gold code sequences used in our system are simply stored in the microcontroller's flash memory and are used to toggle a digital output line that is connected to the modulator. The modulation rate (chip rate) is 1 MHz, resulting in a 2-MHz-wide main band. The tag is programmable for center frequency, transmission interval, PRN code, chip rate, RF output power, and operating schedule. This programmability allows tailoring the tag parameters to the application, which maximizes lifetime for a particular tag mass. Typical settings call for operation during the early morning and evening, when birds are most active. During these periods, the tag sends a 1- to 2-ms-long signal once every minute. This signal is actually the concatenation of two different Gold codes. The first code is common to all tags and allows the receiver to achieve phase synchronization. The second code immediately follows the first, is phase synchronous to the first, and is unique to the

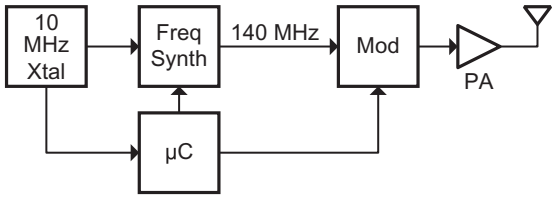

Figure 33.11 Block diagram for the CDMA tracking tags. PA, power amplifier.

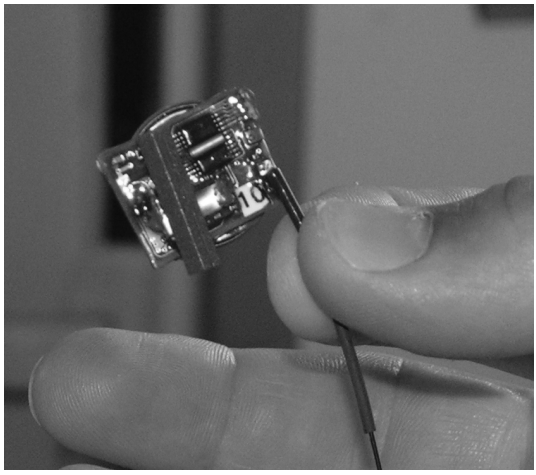

Figure 33.12 BPSK tracking tag, shown with inhibit magnet.

tag that sends it. Although we could, in principle, transmit only the unique tag identifier code, this scheme dramatically reduces the processing load on the receivers since they only need to perform a code phase search on a single synchronization code.

The current tag, shown in Figure 33.12, weighs 1.4 g without the battery and epoxy encapsulation. The tag's 140-MHz center frequency implies a quarter-wave antenna length of approximately 0.5 m. This is too long for most small birds to manage, so the actual antenna used is often between 15 and 25 cm. Despite the efficiency penalty that these electrically short whip antennas impose, they are very common in animal tracking applications because they are relatively unobtrusive and mechanically robust. At the maximum setting, the tag's total output power in the 2-MHz main lobe is 12 dBm (measured into a 50-ohm load). The actual power broadcast into free space is significantly lower than this, since the efficiency of the short antenna is low.

Receiver Architecture: The block diagram for an individual receiver in the automatic tracking system is shown in Figure 33.13. Tag transmissions are received at a 2–7/8 λ phased element monopole antenna that yields approximately 6 dB of gain and is omnidirectional in azimuth. The antenna is mounted atop a 4-m portable mast. The signal is then immediately amplified by a *low-noise amplifier* (LNA), and then passed through an 8-MHz-wide, 6-pole LC passive band-pass filter with a center frequency of 140 MHz. This filter blocks strong, nearby signals before they can

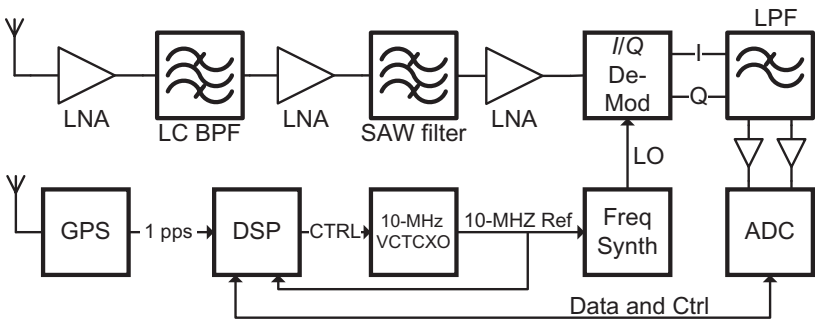

Figure 33.13 Automatic tracking receiver block diagram.

overload subsequent gain stages. The signal then passes through an LNA, a 2-MHz-wide *surface acoustic wave* (SAW) band-pass filter, and another LNA. These additional gain stages are necessary because although SAW filters offer very high selectivity, they are usually fairly lossy. Note that this receiver architecture is not frequency agile; the SAW filter has one set passband and cannot be tuned. This approach is simple and works well if the local RF environment is free from interference at the 140-MHz operating frequency. A more traditional heterodyne architecture would afford the receiver greater flexibility in operating frequency. Next, the signal is down-converted from 140 MHz directly to 0 Hz, or the so-called baseband, by the demodulator. The demodulator uses an internal 90° phase shift circuit to derive sine and cosine signals from the LO, and it multiplies the input with each of these signals. The two outputs, I and Q, are shown in Equations 33.14 and 33.15,

$$\begin{aligned} I(t) &= \cos(2\pi f_{LO}t)\,Ax(t)\cos(2\pi f_C t+\varphi) \\ &= 1/2\,Ax(t)\{\cos(2\pi t(f_C-f_{LO})+\varphi)+\cos(2\pi t(f_C+f_{LO})+\varphi)\} \\ &= 1/2\,Ax(t)\{\cos(\varphi)+\cos(2\pi t(2f_C)+\varphi)\}, \quad \text{if } f_C=f_{LO} \end{aligned} \tag{33.14}$$

and

$$\begin{aligned} Q(t) &= \sin(2\pi f_{LO}t)\,Ax(t)\cos(2\pi f_C t+\varphi) \\ &= \cos(2\pi f_{LO}t-\pi/2)\,Ax(t)\cos(2\pi f_C t+\varphi) \\ &= 1/2\,Ax(t)\{\cos(2\pi t(f_C-f_{LO})+\varphi-\pi/2)+\cos(2\pi t(f_C+f_{LO})+\varphi-\pi/2)\} \\ &= 1/2\,Ax(t)\{\cos(\varphi-\pi/2)+\cos(2\pi t(2f_C)+\varphi-\pi/2)\}, \quad \text{if } f_C=f_{LO}, \end{aligned} \tag{33.15}$$

where f_C is the carrier frequency, f_{LO} is the LO frequency, φ is the phase difference between the LO and the carrier, A is the amplitude of the carrier, and $x(t)$ is the PRN sequence. Since the I and Q signals contain undesirable high-frequency content, the I and Q signals are passed through a low-pass filter. The output of the low-pass filter is the chip sequence that the tag used to modulate the carrier.

After the low-pass filter, we have

$$\begin{aligned} I(t) &= 1/2\,Ax(t)\cos(\varphi) \\ Q(t) &= 1/2\,Ax(t)\cos(\varphi-\pi/2) = 1/2\,Ax(t)\sin(\varphi). \end{aligned} \tag{33.16}$$

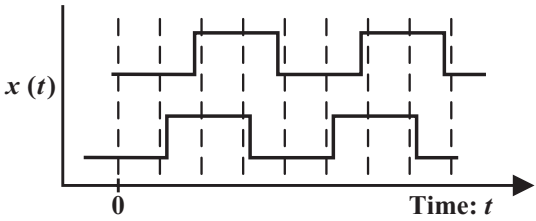

Figure 33.14 A potential problem with chip-synchronous sampling: Two misaligned signals yield identical digital samples.

If we use complex notation, we can write the complex input to the ADC as

$$S(t) = I(t) + jQ(t) = 1/2Ax(t)e^{j\varphi}. \tag{33.17}$$

Note that Equations 33.14–33.17 neglect the case when f_C and f_{LO} differ by a small amount. This situation was discussed in Section 3.2.1 and is addressed in the example at the end of the section.

The complex signal $S(t)$ is buffered and sampled by a two-channel high-speed ADC (one channel for each of the I and Q signals). The sample rate used by the ADC should be greater than the Nyqyist frequency, 2 MHz in this case; our receiver uses a sample clock of 2.8125 MHz for each of the channels. This sample rate, which is a noninteger multiple of the chip rate, ensures that the sampling operation is not synchronous with the chip sequence. This improves the timing resolution by better aligning the samples with the chip edges, on average. Figure 33.14 illustrates the problem with an extreme case. It shows two misaligned PRN signal snippets that are sampled at regular intervals (vertical dashed lines). The intervals are synchronous with the chips, which causes the digital samples to take identical values, even though the two signals are misaligned by nearly half of a chip. Figure 33.14 is an example of aliasing, and in practice, the signals to be sampled are usually band-limited to avoid this problem. There are cases, however, when it is convenient, from a system design standpoint, to violate the sampling theorem. DSP performance limitations may make a lower than desired sampling rate necessary, or may require the samples to be low-resolution (many GPS receivers use only 2 bits per sample). Noninteger rate sampling can sometimes improve the performance of systems operating under these conditions.

The complex samples from the ADC are placed into a circular buffer in the DSP's memory, and the matched filter detector algorithm described in Section 33.3.2 is used to find the cross-correlation peaks within each buffer. Once a peak has been found by the cross-correlation algorithm, its arrival time must be measured. The cross-correlation output includes the sample number (lag) of the peak, so its position within the buffer is known and its time can therefore be calculated, provided that the DSP notes the time when the first sample in the buffer was acquired by the ADC. The timing resolution provided by this method is limited by the ADC sample rate, in this case, 1/2.8125 MHz or about 356 ns, which corresponds to roughly 106 m. Fortunately, we can do better, provided that the SNR is high enough. Recall that the autocorrelation peak from Figure 33.4 is $2T_C$ chips wide, or 5.625 samples wide,

and is triangular. The cross-correlation peak and several samples from either side of it can be used to curve-fit the autocorrelation peak to the cross-correlation data. This method significantly improves the timing resolution, provided that the cross-correlation peak is substantial enough to provide a good fit.

The final step, after the DSP has computed a precise arrival-time estimate, is for each receiver to share that information with the central server that will calculate the position estimate. Each receiver must be connected to the server via a data network (if the positions are required in real time) so that they can submit the arrival information. Our system uses an IP network and submits TCP data packets via the network to the server where they were placed into a database. The server then groups the arrival events by tag ID and computes positions with the stochastic search (SS) method (K. A. Cortopassi, unpublished data) described at the end of Section 33.3.4.

Portability and low cost were significant considerations during the design process. Each receiver, including all antenna components, weighs 30 lb and can be easily transported by a single person. The total power consumption of each receiver, including wireless networking equipment, is 16 W. Power is typically supplied from two 12-V car batteries that are charged by solar panels.

MATLAB Example 33.1 Numerical Simulation of a TOA Receiver

Many of the design issues mentioned in this chapter can be explored with a numerical simulation in MATLAB. Implement a software simulation of the tag/receiver system that allows parameters such as chip rate, carrier offset frequency, SNR, sample rate, and receiver bandwidth to be modified.

Solution

The example code provided ("Chapter_33_Example_1.m") uses MATLAB to generate a carrier signal, modulates that carrier signal with a Gold code, adds Gaussian noise, down-converts the signal into *I*/*Q* baseband signals, cross correlates the signals with the Gold code template, and uses a threshold detector to indicate signal detection. MATLAB codes can be found online at ftp://ftp.wiley.com/public/sci_tech_med/matlab_codes. The script illustrates the importance of *I*/*Q* baseband processing in a communications system that is not phase synchronous and provides a way of investigating the impact of clock frequency offsets in the tag and receiver. Also provided is a Simulink model that illustrates the use of the Gold code sequence generator block.

The example begins by invoking the Gold code generator Simulink block to create the PRN that the tag will send. This block is programmable for a particular sequence; see the Appendix in Reference 28 for appropriate generator polynomial coefficients. The script next generates an appropriate carrier sinusoid and multiplies the PRN with the carrier to create the tag output signal. The following step makes *I* and *Q* LO sinusoids, and multiplies the tag output signal with each, in order to down-convert the signal. These signals are then low-pass filtered using the butter function to generate filter coefficients for filtfilt. The output of the filter is decimated to simulate sampling with an ADC (the resulting sample rate is the ADC's sample

rate). Finally, noise is added using randn; this noise accounts for all noise that the signal would encounter. The cross correlation of the individual I and Q channels with the PRN is calculated via xcorr, and individual plots illustrate the "fading" in each channel as the phase between the carrier and LO changes as the IQ vector rotates. A full, complex cross correlation solves this issue, and the magnitude of this cross correlation is shown in another plot. Finally, a simple threshold detector based on a median magnitude measure of the cross correlation is used to determine if a tag transmission is present.

Time Base: Several components in a TOA receiver require very precise frequency or time information. These include the ADC sampling, the LO generator, and the buffer time stamp. Although very precise quartz frequency references are available, even these devices (which advertise frequency tolerances as low as 0.1 ppm) do not offer sufficient stability to maintain precise synchronization between the receivers over a long period of time. The distance between the receivers, which is typically several kilometers, precludes a cabled or even a point-to-point radio link for synchronization. Fortunately, GPS receivers are capable of providing very precise 1-pulse per second (pps) and 10-MHz signals, which supply the reference signals for the rest of the receiver. Each TOA receiver uses an independent GPS receiver to maintain very tight synchronization with UTC. GPS receivers that are specifically designed for timekeeping purposes are now available for embedded applications. These devices assume a fixed location in order to overdetermine a solution that yields very accurate 1-pps edges. These edges are used to discipline a voltage-controlled, temperature-compensated crystal oscillator (VCTCXO), or in some cases, an oven-compensated crystal oscillator (OCXO). Each of these devices provides excellent short-term stability, and the GPS synchronization maintains their long-term accuracy. See Reference 34 for an overview of modern timekeeping technology.

33.3.4 Arrival-Time Location-Finding Algorithms

Several methods exist for computing location estimates from arrival-time measurements in a TOA system. We briefly present four of them below. Two of the methods (stochastic search [SS] and the Newton–Raphson [NR] method) rely on iterative searching within an assumed solution space; the other two (hyperbolic and spherical intersection) are closed-form solutions based on some simplification of the problem. It should be noted that the scale and requirements of most wildlife tracking systems permit calculations in two dimensions (easting and northing); altitude is ignored. This assumption, which simplifies the system design somewhat, can be made because the primary application of a TOA system is in tracking animals over medium ranges (5–50 km). Any birds being tracked will be near the ground, rather than migrating at altitude, since migrating animals would pass through the relatively small array too quickly for the system to be of use (apart from presence/absence detection). This is an important constraint because tags that are significantly out of plane will incur a significant positioning error if a two-dimensional solution is assumed. Most environments lend themselves to this planar assumption, since the variation in elevation over a typical 5×5 km array cell is small, relative to the 5-km receiver spacing (this

spacing is set by the maximum range of detection, which depends on various factors). In terrain that violates this condition, some receivers must be placed on hilltops so they are significantly out of plane, and a full three-dimensional solution must be found. Note that more detailed treatments of TOA localization computations are available in Chapters 2 and 6.

Hyperbolic Positioning: The hyperbolic positioning method, described by Ho and Chan [7], is a popular approach for determining transmitter position from arrival times. The method works by observing that the locus of points satisfying a signed time difference of arrival (TDOA) between two receivers is one branch of a hyperbola (TDOA pairs can be computed from TOAs). The dashed lines in Figure 33.15 show the hyperbolae induced by three noise-free TDOAs received at the locations represented by the numbered dots. The transmitter is presumed to lie at the intersection of those three (ideal) hyperbolae.

The squared distance between the source at (x, y) and sensor i at (x_i, y_i) is

$$r_i^2 = (x_i - x)^2 + (y_i - y)^2. \tag{33.18}$$

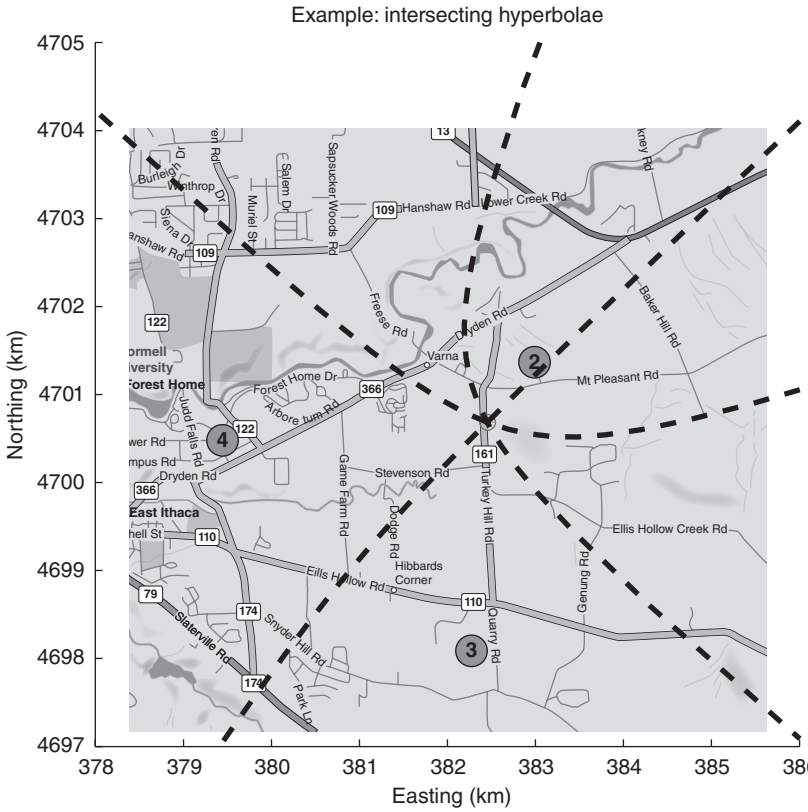

Figure 33.15 Hyperbolic location example.

One of the receivers is chosen as the origin and the computation finds the solution to

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix}^{-1} \times \left\{ \begin{bmatrix} r_2 \\ r_3 \end{bmatrix} r + \frac{1}{2} \begin{bmatrix} r_2^2 - K_2 \\ r_3^2 - K_3 \end{bmatrix} \right\} \tag{33.19}$$

subject to the constraint

$$r^2 = x^2 + y^2, \tag{33.20}$$

where the range difference (TDOA times propagation speed) between receiver 1 and receiver i is

$$r_{i,1} = r_i - r_1 \tag{33.21}$$

and

$$K_i = x_i^2 + y_i^2, \tag{33.22}$$

and the unsubscripted variables are the x and y coordinates of the source and its distance (r) from the receiver at the origin. Where TOAs are available from more than three receivers, replacing the matrix inverse with the pseudoinverse of the receiver coordinates yields a least squares solution.

Spherical Positioning: Another closed-form solution, spherical interpolation (SI), is due to Smith and Abel [1]. The solution is presumed to lie on the surface of a sphere (or a circle in the 2-D case) whose radius is the distance to one of the receiver towers chosen as a reference. The perpendicular distance between the surface of the sphere (the circumference of the circle) and any other receiver tower is the range difference between that tower and the reference tower. The SI method inserts an equation error term (corresponding to measurement noise) into the distance formula and minimizes the error term in a least squares sense to yield the actual solution. One disadvantage of this approach is that it requires one more TDOA than the other methods described here.

Iterative Root Finding (NR Method): The NR method starts with an arbitrary initial guess of the transmitter's location and time of transmission, and proceeds by comparing the measured times of arrival against the TOAs computed from the initial guess. A correction to the guessed position and transmission time is estimated by linearizing the problem at the current transmitter position estimate, and the corrected position is used as the new guess. The process is repeated until a specified convergence criterion is met. Because the method's error term is a nonconvex function of position, this method is sensitive to the quality of the initial guess. In our analysis, because the area of interest was only slightly larger than the bounds of the receiver array, the centroid of the receiver array was used as the initial guess. In more general applications, a suboptimal closed-form solution, such as one of those described earlier, may yield a better initial guess. This method, often used in the GPS system, merits an example.

MATLAB Example 33.2 Estimating Position from TOA Measurements with an NR Method

Use the NR method to estimate the location of a transmitter, given arrival-time measurements at several nearby receivers. Additionally, show the impact of the receiver geometry on the location error.

Solution

We begin by looking at the range $r^{(k)}$ (Eq. 33.23) between a transmitter and the kth receiver, where $t_{Rx(R)}^{(k)}$ is the signal's receive time, measured by the kth receiver's clock; $t_{Tx(R)}^{(k)}$ is the signal's transmit time, measured by the kth receiver's clock; $x^{(k)}$ is the position vector of the kth receiver; x is the position vector of the transmitter; and c is the signal's propagation velocity:

$$r^{(k)} = \left(t_{Rx(R)}^{(k)} - t_{Tx(R)}^{(k)}\right)c = \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\right\|. \tag{33.23}$$

Unfortunately, we do not know $t_{Tx(R)}^{(k)}$, since the transmissions happen at the transmitter, which is asynchronous to the receiver. Instead of ranges, we can express the distance from the transmitter to the kth receiver as a *pseudorange* $\rho^{(k)}$, which is a combination of the true range and some offset; in this case, the pseudorange is measured between the receiver's clock and the tag's clock (these clocks are assumed to have an unknown constant offset). Through substitution of a new variable, we can express $\rho^{(k)}$ as the true range, plus an offset b. Notice that b is the clock offset between the tag and the receivers, expressed in meters. Although $t_{Tx(R)}$ is a column vector with k entries, b is scalar since all receivers have synchronized clocks and therefore will all have the same offset from the tag:

$$\begin{aligned}\rho^{(k)} &= \left(t_{Rx(R)}^{(k)} - t_{Tx(T)}\right)c = \left(t_{Rx(R)}^{(k)} - t_{Tx(R)}^{(k)} + \left(t_{Tx(R)}^{(k)} - t_{Tx(T)}\right)\right)c \\ &= \left(t_{Rx(R)}^{(k)} - t_{Tx(R)}^{(k)}\right)c + b = \left\|\boldsymbol{x}_k - \boldsymbol{x}\right\| + b.\end{aligned} \tag{33.24}$$

Let $\rho_0^{(k)}$ be an approximation of $\rho^{(k)}$ with initial guesses for tag position x_0 and clock offset b_0:

$$\rho_0^{(k)} = \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_0\right\| + b_0. \tag{33.25}$$

The difference between the measured pseudorange $\rho^{(k)}$ and the initial guess $\rho_0^{(k)}$ is $\delta\rho^{(k)}$, as shown in Equation 33.26:

$$\delta\rho^{(k)} = \rho^{(k)} - \rho_0^{(k)},\ \boldsymbol{x} = \boldsymbol{x}_0 + \delta\boldsymbol{x},\ b = b_0 + \delta b. \tag{33.26}$$

This difference starts out large, since our initial guess is poor, and eventually goes to zero as our guess for the tag position and time offset improves. Additionally, we introduce the variables δx and δb, which represent the changes to our initial guesses in order to move closer to the actual position and time. With these relationships, and a Taylor series approximation of the vector norm, we find the relationship in Equation 33.27:

$$\delta\rho^{(k)} = \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_0 - \delta\boldsymbol{x}\right\| - \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_0\right\| + \left(b - b_0\right) \approx -\frac{\left(\boldsymbol{x}^{(k)} - \boldsymbol{x}_0\right)}{\boldsymbol{x}^{(k)} - \boldsymbol{x}_0} \cdot \delta\boldsymbol{x} + \delta b. \tag{33.27}$$

Equation 33.27 can be rewritten in matrix form as Equation 33.28,

$$\delta\boldsymbol{\rho} = \begin{bmatrix} \delta\rho^{(1)} \\ \vdots \\ \delta\rho^{(k)} \end{bmatrix} = \begin{bmatrix} -\dfrac{\left(\boldsymbol{x}^{(1)} - \boldsymbol{x}_0\right)}{\left\|\boldsymbol{x}^{(1)} - \boldsymbol{x}_0\right\|} & 1 \\ \vdots & \vdots \\ -\dfrac{\left(\boldsymbol{x}^{(k)} - \boldsymbol{x}_0\right)}{\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_0\right\|} & 1 \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x} \\ \delta b \end{bmatrix} = \mathbf{G} \begin{bmatrix} \delta\boldsymbol{x} \\ \delta b \end{bmatrix}, \tag{33.28}$$

where $\delta\rho$ is a column vector with as many entries as receivers that participated in this particular arrival-time measurement. Thus far, we have not made use of the actual arrival-time measurements, $t_{Rx(R)}^{(k)}$. We can say that $\rho^{(k)} = t_{Rx(R)}^{(k)} c$ if we choose $t_{Tx(T)} = 0$. This is an arbitrary but allowable choice since we are solving for b, the offset between the receiver and transmitter clocks. The linear system in Equation 33.28 is easily solved numerically for $\begin{bmatrix} \delta x \\ \delta b \end{bmatrix}$; the next estimates for tag position and time, x_1 and b_1, are given by $x_1 = x_0 + \delta x$ and $b_1 = b_0 + \delta b$.

These updated guesses are used to compute new values for $\delta\rho$ and $\mathbf{G}$, and Equation 33.28 is solved again. This process continues with successively better estimates for position and time offset until the error is below a termination threshold. The algorithm converges quickly, usually requiring only a few iterations. The example code provided illustrates this algorithm by creating synthetic TOA measurements from a tag at a known location, and then uses only those measurements to find the location. It plots the array geometry, the true tag location, and shows the positions of the guesses as they converge to the true position, shown in Figure 33.16.

The precision of TOA location estimates in the presence of noise depends on the location of the transmitter, relative to the receivers, with some locations yielding much higher error than others. This phenomenon, known as position dilution of precision (PDOP), is a function of the array geometry and the position of the tag

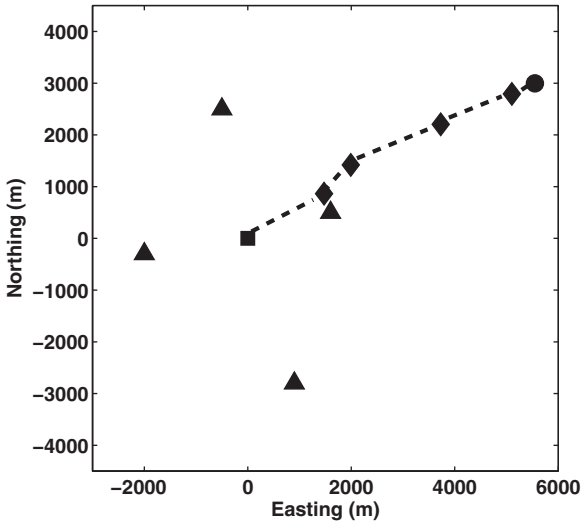

Figure 33.16 Example of NR positioning, showing receivers (triangles), initial guess (square), subsequent guesses (diamonds), and final position (circle).

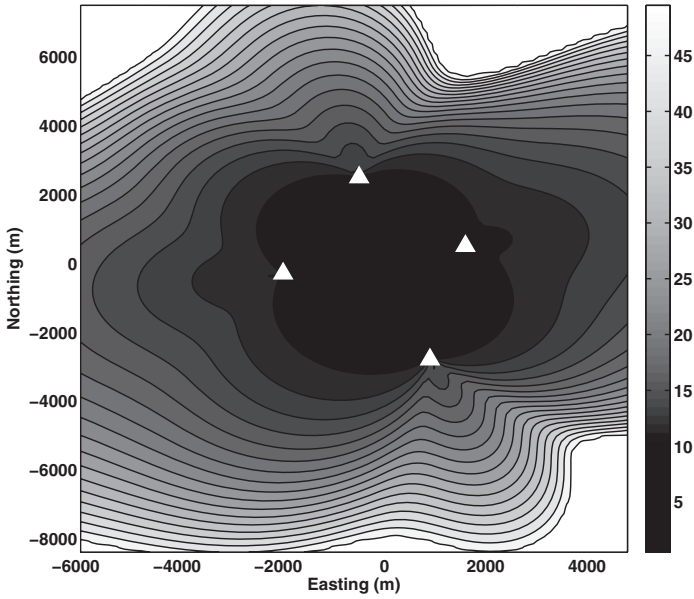

Figure 33.17 A heat map of PDOP for a four-receiver array (receivers are white triangles). Darker areas represent lower PDOP values. See color insert.

within the array. PDOP can be thought of as a scaling factor that makes our measurement errors (from timing resolution, RF noise, etc.) more pronounced in some locations of the array than in others. Equation 3.30 illustrates this point (σ is a lumped error term that represents all sources of timing error and is expressed in meters). The NR method provides a convenient means of estimating PDOP, as shown in Equation 33.29:

$$\mathbf{H} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}, \mathrm{PDOP} = \sqrt{H_{11} + H_{22} + H_{33}} \tag{33.29}$$

$$\text{RMS position error} = \sigma \cdot \mathrm{PDOP}. \tag{33.30}$$

The example code provided ("Chapter_33_Example_2.m") estimates the position from TOA information and plots PDOP (Fig. 33.17) for any desired array geometry. MATLAB codes can be found online at ftp://ftp.wiley.com/public/sci_tech_med/matlab_codes.

Stochastic Search (SS) Method: The algorithm we used for computing actual field location estimates in our system was developed internally, and is a form of stochastic search. A number of initial guesses are spaced within specified search bounds and the corresponding theoretical demeaned TOA vectors are compared with the measured demeaned TOAs. A fraction *n* of the initial guesses with the smallest squared error are retained and duplicated, and small random perturbations are added to the duplicate points. The best *n* of these new guesses are expected to be closer to the actual solution, and so are retained, duplicated, perturbed, and passed to the next iteration. This process is continued until the current guesses converge to within a specified

radius. The random perturbations are drawn from a uniform distribution over an interval that decreases with each iteration (K. A. Cortopassi, unpublished data). Because this process uses multiple guesses throughout, its initial condition is not limited to a single guess. The random perturbations are relatively large at first and decrease in size as the search proceeds, making the search unlikely to converge on a local minimum rather than the best solution.

33.4 PERFORMANCE OF A DEMONSTRATION WILDLIFE TRACKING SYSTEM

Although the ultimate purpose of a TOA tracking network is to determine geographical position, the role of each individual receiver is to make accurate signal arrival-time measurements, since any errors in the arrival-time estimate translate into errors in the position estimate. We measured the timing error of our system by setting up two receivers side by side, and by injecting a test signal into each receiver via a splitter and equal cable lengths. The arrival times of each test transmission were compared and the differences between the two receivers' measurements were computed. The results of this test are shown in Figure 33.18 for a test with high SNR (30 dB). The receivers are able to achieve tight synchronization; however, a slight time offset was evident in this particular test. We later attributed this to cabling differences in each receiver's GPS antenna. The standard deviation of this test was 25 ns, or roughly 7 m.

We employed a similar test to determine the receiver's minimum signal detection. We injected a test signal into the receiver via a variable attenuator and increased the attenuation until the signal was not detected. The minimum signal that can typically be detected (with no added in-band noise) is −124 dBm.

The range of detection is the system parameter that is most often requested and is also the most difficult to estimate because it is so dependent on the applica-

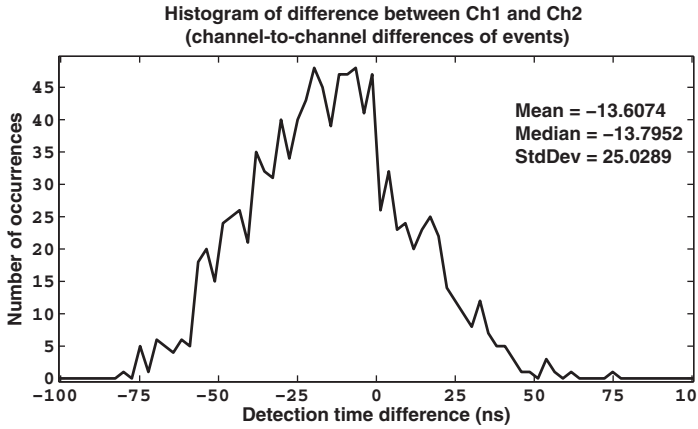

Figure 33.18 Detection time test of two colocated receivers.

tion's location. Wildlife tracking occurs in a wide range of environments, and tag signals can encounter everything from foliage to free space. Many resources exist for estimating the likely attenuation in forests ([35, 36, 37, 39]), though we found a wide variation in practice. We performed numerous field tests in flat and rolling terrain, as well as transmissions through clear areas and transects obstructed by foliage. We also tested the free-space range by placing a receiver on a tall hill or building and by moving the transmitter to another suitably prominent location. As described earlier, the radiation efficiency and radiation pattern of the tags are difficult to measure, and change in response to how and where the tag is mounted to an animal. In general, the range of a tag 1 m above the ground is 3–5 km. A tag in free space can be detected up to 10 km away. These estimates lead to a recommended receiver spacing of 5 km, with the receivers arrayed in a square or hexagonal grid to provide coverage over a large region.

We compared the performance of four location estimation methods—hyperbolic localization (HL), SI, NR iteration, and SS, discussed in Section 33.3.4 using actual field data, acquired on August 23, 2007 in Ithaca, New York, between approximately 12:30 and 14:30. An RF tag transmitter was stationed in each of

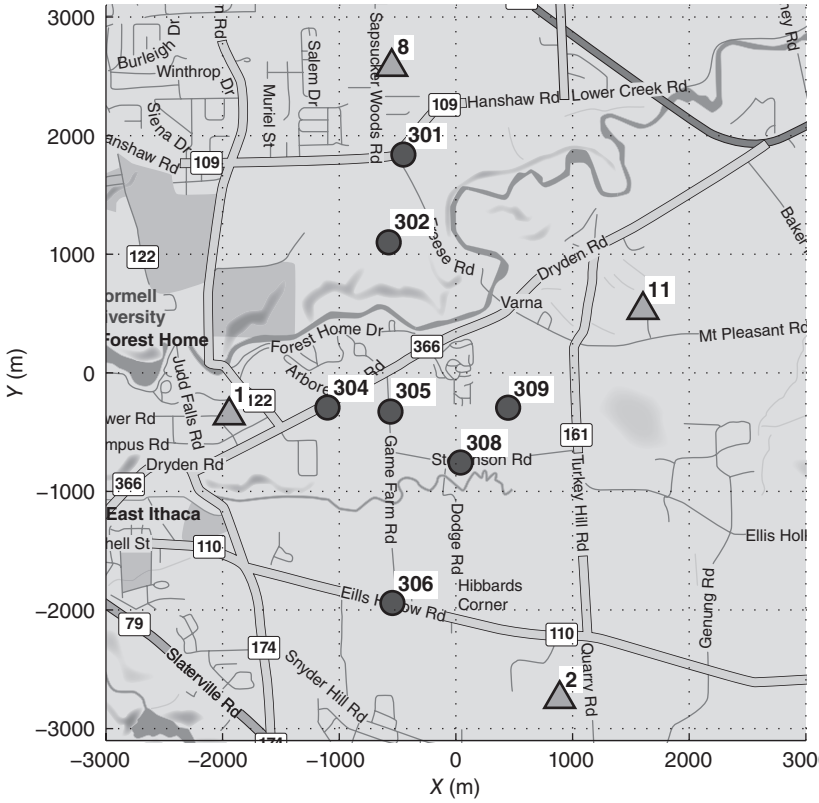

Figure 33.19 Setup for field testing of localization system. Transmitter sites are circles; receiving towers are triangles.

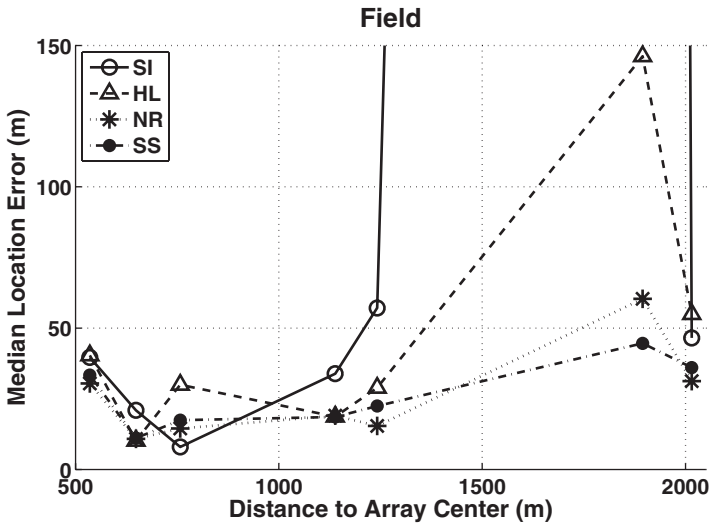

Figure 33.20 Median location error for each of the four estimation methods (SI, open circle; HL open triangle; NR asterisk; SS closed circle) applied to the *field-derived* TOA vectors plotted against the source distance from the array center. The median error for the SI method at site 301 was 3401.00 m.

the seven sites (circles) shown in Figure 33.19; several hundred repeated transmissions were made from each site, and the arrival times were measured by each receiver (triangles). The arrival-time data from our test site were used to evaluate the performance of the four localization techniques mentioned. The field test results are shown in Figure 33.20, which compares the median location error (distance from actual transmission location) for the four different methods with different transmitter locations. Similar experiments were carried out with simulated arrival-time data, which allowed the effect of different noise models to be investigated. In these tests, the two iterative methods (NR and SS) outperformed the closed-form solutions (SI and HL). The iterative methods perform particularly well, relative to the other methods, in the presence of RF noise, timing measurement error, and receiver survey position error.

33.5 CAVEATS AND LIMITATIONS

The use of terrestrial receivers with limited range means that unlike GPS and Argos, fixed TOA systems are not appropriate for tracking over very large spatial scales. For our system, the typical receiver spacing is 5 km, so a grid of 16 receivers could cover an area of 400 km^2 depending on the terrain. Therefore, the system is appropriate for covering a limited study site, but establishing coverage over a large geographic region would require a prohibitive number of receivers. Also, though the system can in principle provide real-time position updates, the tags would likely be configured to provide position updates relatively infrequently, in order to conserve

energy. This feature becomes a liability if field personnel must make an unplanned capture of an animal, since infrequent position updates could make following an animal difficult.

Signals suitable for TOA estimation occupy a fairly wide bandwidth, as illustrated in Figure 33.6. This can be problematic for TOA wildlife tracking systems, which often opportunistically share spectrum with narrowband transmitters, including other wildlife tags, amateur radio operators, and licensed high-power communications systems in the very high-frequency (VHF) band. Although the choice of low transmitter power, the use of short transmission durations, and the low-power spectral density afforded by direct sequence spread spectrum essentially ensure that the TOA system's transmitters will not cause interference for other systems, the receivers are susceptible to narrowband interference. If a strong nearby transmitter happens to fall within the passband of the TOA receiver's front-end filters, the interfering transmitter can overwhelm the receiver's dynamic range or cause the receiver's automatic gain control circuits to adjust, dropping the much weaker desired tag signals below the detection threshold. Though these issues can be partially addressed through careful design of the RF front end, and through adaptive digital notch filters, strong narrowband interference is a persistent design issue for CDMA systems.

Conventional handheld direction-finding equipment is far simpler than a complete TOA system. This leads to two advantages: up-front cost and reliability. Though no TOA-based wildlife tracking system is currently available for sale, the cost of a small, four-receiver TOA system is likely to be substantially higher than four handheld directional receivers. Individual TOA receivers are also components of a complicated, networked system, and individual component failures in any of the receivers could cause substantial portions of the tracking system to fail, since the coordination of multiple receivers is required for proper operation.

33.6 CONCLUSION

Wildlife tracking tools have undergone evolutionary improvements since their introduction over 60 years ago, and the last 20 years have seen a proliferation of complementary technologies brought to bear on the problem. Improvements in technology, including the application of microcircuits to wildlife tracking equipment, have enabled small tags, which in turn allow their use with a much greater diversity of animals. As tag sizes have dropped, their energy consumption has become a critical design parameter. Existing terrestrial wildlife tag technology, though simple and reliable, uses precious transmitter energy poorly; updated communications techniques, including improved modulation and the application of signal processing in the receiver, could enable further tag mass reductions. Additionally, appropriate tag signals will enable future wildlife tracking systems to provide accurate, automated localization via TOA measurements. We have demonstrated the feasibility of this approach with small, inexpensive, portable receivers that can be combined to form a tracking network capable of high-fidelity localizations. This system is appropriate for a wide range of animals and animal tracking studies, and its small, low-cost,

long-lived, spread spectrum transmitters also make it appropriate for generic mobile asset tracking applications.

REFERENCES

[1] J. Abel and J. Smith, "The spherical interpolation method for closed-form passive source localization using range difference measurements," *Proc. IEEE Intl. Conf. on Acoustics Speech and Signal Processing (ICASSP) '87*, vol. 12, pp. 471–474, April 1987.

[2] C. D. LeMunyan, W. White, and E. Nybert, "Design of a miniature radio transmitter for use in animal studies," *J. Wildl. Manage.*, vol. 23, no. 1, pp. 107–110, 1959.

[3] W. W. Cochran, D. W. Warner, J. R. Tester, and V. B. Kuechle, "Automatic radio-tracking system for monitoring animal movements," *BioScience.*, vol. 15, no. 2, pp. 98–100, 1965.

[4] M. Green, T. Piersma, J. Jukema, P. D. Goeij, B. Spaans, and J. Van Gils, "Radio-telemetry observations of the first 650 km of the migration of Bar-tailed Godwits *Limosa lapponica* from the Wadden Sea to the Russian Arctic," *Ardea.*, vol. 90, no. 1, pp. 71–80, 2002.

[5] B. Naef-Daenzer, "A new transmitter for small animals and enhanced methods of home-range analysis," *J. Wildl. Manage.*, vol. 57, no. 4, pp. 680–689, 1993.

[6] M. Wikelski, "Automated Radio Tracking System (ARTS) [Online]." Available: http://www.princeton.edu/%7Ewikelski/research/index.htm.

[7] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal. Process.*, vol. 42, no. 8, pp. 1905–1915, 1994.

[8] W. W. Cochran, G. Swenson Jr., and L. Pater, "Radio direction-finding for wildlife research," [Online] 2002. Available: http://userweb.springnet1.com/sparrow/Direction-finding.html (accessed 2008).

[9] J. E. Lee, "Accessing accuracy of a radiotelemetry system for estimating animal locations," *J. Wildl. Manage.*, vol. 49, pp. 658–663, 1985.

[10] R. Kenward, *A Manual for Wildlife Tagging*. London: Academic Press, 2001.

[11] W. W. Cochran, "Long distance tracking of birds," *Animal orientation and navigation*, NASA SP-262, pp. 39–59, 1972.

[12] B. Naef-Daenzer, F. Widmer, and M. Nuber, "A test for effects of radio-tagging on survival and movements of small birds," *Avian Sc.*, vol. 1, no. 1, pp. 15–23, 2001.

[13] K. J. Gaston and T. M. Blackburn, "The frequency distribution of bird body weights: aquatic and terrestrial species," *Ibis.*, vol. 137, pp. 237–240, 1995.

[14] B. Lyon, A. Chaine, and D. Winkler, "A matter of timing," *Science.*, vol. 321, pp. 1051–1052, 2008.

[15] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propagation.*, vol. 34, no. 3, pp. 276–280, 1986.

[16] P. Misra and P. Enge, *Global Positioning System, Signals, Measurements, and Performance*, 2nd ed. Lincoln, MA: Ganga-Jamuna Press, 2006.

[17] Technosmart, Montecelio, Italy, Web page. Available: http://www.technosmart.eu/.

[18] Wildtrack Telemetry Systems Ltd., Leeds, UK. Fastloc Web page. Available: http://www.wildtracker.com/fastloc.htm.

[19] CLS Inc, "A user manual for the Argos system," [Online] 2007. Available: http://www.argos-system.org/documents/userarea/argos_manual_en.pdf.

[20] S. Gauthreaux, J. Livingston, and C. Belser, "Detection and discrimination of fauna in the aerosphere using Doppler weather surveillance radar," *Integr. Comp. Biol.*, vol. 48, no. 1, pp. 12–23, 2008.

[21] E. T. Cant, A. D. Smith, D. R. Reynolds, and J. L. Osborne, "Tracking butterfly flight paths across the landscape with harmonic radar," *Proc. R. Soc. B.*, vol. 272, pp. 785–790, 2005.

[22] B. Colpitts and G. Boiteau, "Harmonic radar transceiver design: miniature tags for insect tracking," *IEEE Trans. Antennas Propagation.*, vol. 52, no. 11, pp. 2825–2831, 2004.

[23] J. Riley and A. Smith, "Design considerations for an harmonic radar to investigate the flight of insects at low altitude," *Comput. Electron. Agric.*, vol. 35, p. 151, 2002.

[24] F. Savaglio, D. Maskell, and H. Spencer, "Direct sequence spread spectrum burst transmissions in a hyperbolic automatic radio tracking system," *International Conference on Telecommunications*, Melbourne, Australia, pp. 903–908, April 1997.

[25] P. Lemnell, C. Johnsson, H. Helmersson, O. Holmstrand, and L. Norling, "An automatic radiotelemetry system for position determination," *Proc. Int. Conf. Biotelemetry*, vol. 4, pp. 76–93, 1983.
[26] S. Cobb, "GPS pseudolites: theory, design and applications," Ph.D. Thesis, Stanford University, 1997.
[27] A. Bensky, *Wireless Positioning Technologies and Applications*. Norwood, MA: Artech House, 2008.
[28] R. Dixon, *Spread Spectrum Systems with Commercial Applications*, 3rd ed. New York: John Wiley & Sons, 1994.
[29] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. On Inf. Theory.*, vol. 13, no. 4, pp. 619–621, 1967.
[30] D. Van Nee and A. Coenen, "New fast GPS code-acquisition technique using FFT," *IEEE Electron. Lett.*, vol. 27, no. 2, pp. 158–160, 1991.
[31] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press, 1992.
[32] R. MacCurdy, R. Gabrielson, E. Spaulding, A. Purgue, K. Cortopassi, and K. Fristrup, "Automatic Animal Tracking Using Matched Filters and Time Difference of Arrival," *J. Commun.*, vol. 4, no. 7, pp. 487–495, 2009.
[33] R. MacCurdy, R. Gabrielson, E. Spaulding, A. Purgue, K. Cortopassi, and K. Fristrup, "Real-time, automatic animal tracking using direct sequence spread spectrum," *Proc. European Conference on Wireless Technology (EuWiT)*, Amsterdam, pp. 53–56, October 2008.
[34] D. Allan, N. Ashby, and C. Hodge, "The science of timekeeping," Hewlett Packard Corporation, Application note 1289, 1997.
[35] International Telecommunications Union, "Attenuation in vegetation," Rec. ITU-R P.833-4. Available: http://www.itu.int (accessed 2001).
[36] M. L. Palud, T. Dupaquier, and L. Bertel, "Experimental study of VHF propagation in forested environment and modeling techniques," *Proceedings of the IEEE International Radar Conference*, Alexandria, Virginia, pp. 539–544, May 2000.
[37] R. Tewari and S. Swarup, "Radio wave propagation through rain forests of India," *IEEE Trans. Antennas Propagation.*, vol. 38, no. 4, pp. 443–449, 1990.
[38] W. W. Cochran and R. D. Lord, "A radio tracking system for wild animals," *J. Wildl. Mgmt.*, vol. 27, no. 1, pp. 9–24, 1963.
[39] International Telecommunications Union, "Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3000 MHz," Rec. ITU-R P.1546-1. Available: http://www.itu.int (accessed 2003).
[40] B. Naef-Daenzer1, D. Früh, M. Stalder, P. Wetlil, and E. Weise, "Miniaturization (0.2·g) and evaluation of attachment techniques of telemetry transmitters," *J. Exp. Biol.*, vol. 208, pp. 4063–4068, 2005.