Title

Position Location in AHS by Magnetic Pseudo-Noise Signals

Authors

Bana, Soheila V.
Varaiya, Pravin

Publication Date

1999-07-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Position Location in AHS by Magnetic Pseudo-Noise Signals

Soheila V. Bana, Pravin Varaiya
University of California, Berkeley

California PATH Research Report
UCB-ITS-PRR-99-21

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

Report for MOU 318

July 1999

ISSN 1055-1425

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

Position Location in AHS by Magnetic Pseudo-Noise Signals

University of California at Berkeley
Soheila V. Bana and Pravin Varaiya

February 1999

Contents

Abstract

This paper proposes a novel method for position location by automated vehicles in AHS. The proposed positioning system meets the desired accuracy for AHS and is economically feasible because it takes advantage of the infrastructure and characteristics of the automated roads. This positioning system is very similar to GPS in the sense of using pseudo-noise codes for range measurement. The phase of a pseudo-noise signal can be mapped to the receiver's range from a reference point where the signal correlation properties insure accurate phase estimation. The magnetic markers that are installed on the road for vehicles' lateral control are proposed as the medium to carry the signal. A magnetic pseudo-noise signal that is unique for each lane in the network of automated roads allows a vehicle to resolve its absolute position on the road by estimating the signal phase. The system can detect positioning errors with relatively short delays and correct them.

1 Introduction

Most positioning systems such as GPS use pseudo-noise (PN) signals for range measurement. PN signals are excellent in the accuracy that they can provide in range measurement due to their correlation and autocorrelation properties. However, a radio-navigation system such as GPS may not be the right solution for the automated highway system (AHS) because of the errors such as multi-path that are associated with radio systems. Accuracy in positioning is important for AHS and an automated vehicle at any point must beware of its degree of accuracy in position location to choose its control policy accordingly.

Magnetic markers are provisioned to be used on automated roads for lateral control of vehicles within a lane. Magnetic markers can be binary coded using their dual polarity. Suppose that an automated vehicle knows the code lay out on the road with respect to physical location. If the vehicle has a replica of the signal, it is able to measure the phase of the received signal by autocorrelating the received magnetic signal and its replica and learn its position on the road. The autocorrelation properties of the PN codes allow for error detection capability of the system.

The paper is organized as follows. Section 2 is an overview of the positioning requirements for AHS, and currently available positioning systems. Sections 3 and 4 discuss maximal length PN sequences, and positioning by magnetic maximal PN sequences, respectively. Sections 5 discuses the system's components and design. Section 6 is a detailed evaluation of the system, complemented by section 7 that suggests further improvement. Section 8 contains the cost analysis of the system, and section 9 compares the magnetic positioning system with GPS. Finally, section 10 concludes the paper.

2 Background

A positioning system for the coordination layer must satisfy the following specifications.

Independence The positioning system should be evaluated independent of other components of the AHS system.

High Precision and Accuracy Increased precision of positioning system allows a reduced space for maneuvers. Thus, throughput is improved as the positioning system becomes more accurate.

Reliability Functionality, precision, and accuracy of the system must be independent of the physical environment, weather condition, surrounding walls (tunnels), and road visibility.

Fault Awareness The positioning system should allow an estimate of the positioning error in order to detect faults. Fault detection is important for vehicles and allows them to avoid or abort unsafe maneuvers.

Error Bounds Given an AHS design, the system level requirements of the design will impose constraints on a feasible positioning system and its error bounds that can be tolerated.

Economy Feasibility of the positioning system well depends on its economic feasibility.

We now provide an overview of the position location systems. Most position location systems use radio signals. A radio position location system is often composed of two or more ranging systems each capable of measuring the range between a reference point and the user. The three-dimension position of a user requires its range from three reference points, i.e., three simultaneous ranging systems are required to locate the position of a user [1].

The most classic example of a radio ranging system is radar which transmits a signal and measures the travel time of its echo.

$$R = cT/2$$

where R is the range, c is the speed of the signal, and T is the time taken by the signal to travel to the target and return. Radar is a two-way system that generally cannot detect an object's position, and hence cannot be used for position location. It is only useful in object detection and ranging measurement [2].

In position location systems such as GPS, there are several one-way transmitters (satellites) that send marked signals to users. Each transmitter marks its signal with its own absolute position which can be used by the user as a reference point. A user receives signals from different transmitters and must have a means of measuring the travel time of the signal for range measurement. For this purpose, the user has replicas of the transmitted signals. Furthermore, the transmitters and user are synchronized. A user detects the phase shift of the received signal in comparison with its own replica of the signal, and converts it to the range that the signal has traveled by $R = Tc$. Having its range from three reference points for the user is equivalent to having three equations and three unknowns in terms of the coordinates of the user [3, 4, 5, 1]. GPS has unmodeled errors that we will discuss.

There are other systems where a reverse scheme for position location works and the position of users are resolved by a central office, and not directly by the users [6, 7, 8]. A user sends a marked signal that carries its identity to different receivers. The receivers measure the range of the user with respect to their own positions and relay the information to a central office where the received information are used to locate the user. Central position location systems are used for systems that require central management, such as fleet management and anti-theft systems.

Loran-C is a system where synchronization is only required for transmitters, and not the users. A user measures the phase difference between two received signals, and measures the difference of its range to those transmitters. Such system is useful for position tracking in navigation because initial information about the position is required [9, 7, 10].

Direction finding is another method where the receiver has special antennas that are capable of finding the direction (relative angel) of the received marked signals, and consequently can resolve its position [11].

In general, multi-path is a problem for all radio-navigation systems. A radio-navigation system by itself cannot provide real-time accurate positioning information for a dynamic receiver.

3 Maximal Length PN Sequences

Ranging systems often use digital signals. Code sequences generated by shift registers are most popular because a small number of shift registers can generate relatively long codes. The importance of the *code sequence* to a ranging system is difficult to overemphasize, for the type of code used, its length, and its bit rate set bounds on the capability of the system that can be changed only by changing the code[12]. PN signals are widely used in ranging systems for their special autocorrelation and cross-correlation properties. The bits of PN sequences are called *chips* to underscore that these codes do not carry data, but often modulate data.

3.1 Linear Recursions

Let $\{b_i\}$ be a sequence over a field $\mathcal{F}$. $\{b_i\}$ is said to satisfy a *linear recursion* if there exists a relation

$$b_i = \Sigma_{j=1}^{n} a_j b_{i-j}$$

$\forall i$, with $a_j \in \mathcal{F}$. In other words, each sequence element can be calculated from the n preceding elements. $max\{n|a_n \neq 0\}$ is called the *degree* of the linear recursion. Every periodic sequence $\{b_i\}$ satisfies an infinite number of linear recursions, including the obvious recursions $b_{i+mN} = b_i$ $\forall i, m \in Z^+$ and $N \in Z^+$ where N is the period of the sequence. In binary sequences, the field $\mathcal{F}$ is $(\{0, 1\}, +)$ where the operation $+$ is linear (modulo-2 addition), hence, the term linear codes [13].

Figure 1 shows how a sequence generator can be constructed as an n stage feedback shift register. A shift register sequence generator consists of a shift register working in conjunction with appropriate logic, which feeds back a logical combination of the state of two or more of its stages to its input. The output of a sequence generator, and the contents of its n stages at any clock time is a function of the outputs of the stages fed back at the preceding sample time.

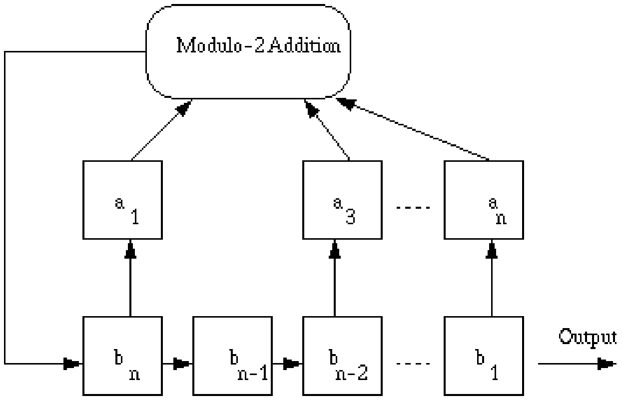

Figure 1: *A linear feedback shift register that is composed of n stages. The feedback to b_n is determined by the characteristic polynomial of the code, $a_n x^n + ... + a_3 x^3 + a_1 x^1$.*

A *characteristic polynomial* can be considered for each code to represent the feedback configuration of the code generator. Consider decreasing indices for the n registers, starting with the register that receives the feed back. By associating each register index i with a power term $n+1-i$ in a polynomial, we obtain an n-terms polynomial that contains the power terms from 1 to n. In Figure 1 , replace b_i with x^{n+1-i}, i.e., the power of x is the same as the index of its feedback. Then the (modulo-2) sum of the feedback terms is the same as the value of the characteristic polynomial. The feedback configuration must include a tap on the last stage in order to generate a maximal length sequence code [12, 13].

3.2 Maximal Length Sequence Codes

The longest codes that can be generated by a given shift register are called *maximal sequence codes.* In binary shift register sequence generators, the maximum length sequence has $N = 2^n - 1$ chips, where n is the number of stages in the shift register. The degree of the characteristic polynomial of the feedback generator, n, is called the *linear span* of the sequence. It is often used as an abstract measure of complexity of the sequence structure.

The number of possible states for each register is 2, and the number of possible states for n registers is 2^n. Excluding the all-zero state, there remains a maximum of $2^n - 1$ states for an n-stage shift register. Every possible state or n-tuple of a given n-stage generator exists at some time during the generation of a complete code cycle. Each state exists for one and only one clock interval.

3.3 Autocorrelation and Index of Discrimination

Autocorrelation function is a measure of the similarity between a signal and a phase-shifted replica of itself. In range measurement, a receiver measures the phase shift needed to maximize the autocorrelation of the received signal and its synchronized replica of the signal. The precision of range measurement depends on phase shift measurement and maximizing the autocorrelation function. Consider a code sequence $\{b_i\}$ and let us define the *periodic autocorrelation function* as

$$R_\tau = \Sigma_{i=1}^{n} b_i b_{i+\tau}$$

where $b_i b_{i+\tau}$ is equal to 1 if $b_i = b_{i+\tau}$, and -1 if $b_i \neq b_{i+\tau}$. When the period $2^n - 1$ is large, the full period correlation loses some of its value as a design parameter. Correlation calculations in this case typically are carried over blocks of K chips, where K may be larger than n, the number of stages of the code generator, and smaller than the code period, $2^n - 1$. A more appropriate statistic for study in this case is the *partial autocorrelation* defined as the following.

$$R(K, i, \tau) = \Sigma_{j=0}^{K-1} b_{i+j} b_{i+j+\tau}$$

This computes the cross-correlation between two blocks of K symbols from $\{b_i\}$, one block located τ symbols from the other.

Index of Discrimination (ID) [12] is a property of a code sequence, pair of sequences, or a sequence and other signal that determines a receiver's ability to recognize the proper point of synchronization. ID denotes the difference in correlation between a fully correlated, i.e., perfectly synchronized, code and the peak of minor autocorrelation or of cross-correlations. A particular code has separate ID values for autocorrelation and cross-correlation with non-coded signals. The higher the ID value, the better the code. High precision position location systems such as GPS use PN codes with high ID values for both autocorrelation and cross-correlation with other signals.

Figure 2 depicts the autocorrelation function for maximal sequence codes. In the region between the zero and plus or minus one chip shifts, correlation increases linearly so that the autocorrelation function for a maximal sequence code is triangular. This characteristic autocorrelation is used to great advantage is ranging systems. A range measurement is ensured of being *accurate within one chip* by using the correlation peak as the marker for measurement. This may be accomplished by setting the correlation detector in such a way that it recognizes the level associated with ±1-chip synchronization and does not recognize a lower level.

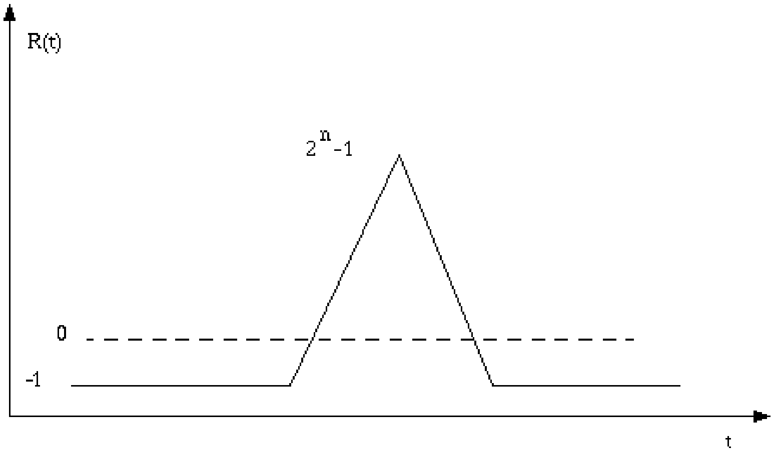

Figure 2: *Maximal sequence autocorrelation function*

Non-maximal sequences have markedly different autocorrelation properties from those of the maximal sequences. Figure 3 illustrates a typical autocorrelation for a non-maximal code. The minor correlation peaks are dependent on the actual code used and are caused by partial correlations of the code with a phase-shifted replica of itself. When such minor correlations occur, a receiving system's ability to synchronize may be impaired because it must discriminate between the major (±1 chip) and minor correlation peaks, and the margin of discrimination is reduced.

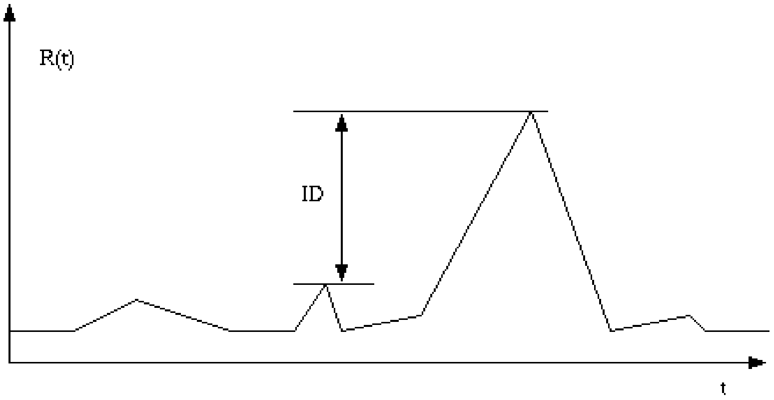

Figure 3: *Typical non-maximal code autocorrelation function*

3.4 Other Properties of Maximal Length Sequences

Maximal sequence codes have the following properties [12, 13].

- The number of ones and zeros are equal in a sequence. For example, a 127-chip code has 64 ones and 63 zeros. The number of ones in any linear maximal code is $2^n/2 = 2^{n-1}$, and the number of zeros is $(2^n/2) - 1 = 2^{n-1} - 1$, where n is the number of stages in the code generator, and the code length is 2^{n-1} chips. This randomness property allows for a low value of either cross-correlation of two PN codes, or autocorrelation of a PN code with its phase-shifted replica. This property is used in phase measurement.
- There is a specific sequence for a shift register, depending on its feedback configuration. A *run* is defined as a finite sequence of repeating ones or zeros. Relative positions of the runs vary from code sequence to code sequence, but the number of each run length does not. Freymodsson [12] has shown that there are exactly $2^{n-(p+2)}$ runs of either ones or zeros of length p in every maximal code sequence. The exception is that there is only one run containing n ones and one containing $n-1$ zeros. There are no runs of zeros of length n or of ones of length $n-1$. Even though the sequences have some randomness properties that can be demonstrated [13], the maximal linear sequences are deterministic as sequences repeat at intervals of $2^n - 1$ chips. Each repetition exhibits the same one-zero distribution. The statistical distribution of ones and zeros is well defined and always the same. As the window of observed values within a period increases, the randomness of one-zero distribution decreases. Finally, when sample size equals $2^n - 1$, the number of ones per sample is $2^n/2$ and the number of zeros is $2^n/2 - 1$, as noted in the previous bullet. We use this property in error analysis of positioning with maximal sequence codes.
- Autocorrelation of a maximal code which is the degree of correspondence between a code and a phase-shift replica of itself is such that for all values of phase shift the correlation value is -1, except for the 0 ± 1 chip phase-shift, in which correlation varies linearly from -1 value to a maximum $2^n - 1$. This is depicted in autocorrelation plot of Figure 2, that shows the number of agreements minus disagreements over the length of the two codes being compared, as the codes assume every shift number in the set of shifts of interest. Such a plot is two-valued, with a peak only at the zero shift point. This is an invaluable property because it allows the receiver to discriminate between signals on a yes-no basis.

  Cross-correlation of two codes is of similar importance. As depicted in Figure 3, it is plotted as a function of phase shift. Unfortunately, cross-correlation is not so well behaved as autocorrelation, and may cause problems for radio waves. In our design, as we will see, cross-correlation does not exist and related difficulties do not arise.

- A modulo-2 addition of a maximal linear code with a phase-shifted replica of itself results in another replica with a phase shift different from either of the originals.

3.5 Predictable Codes

An important property of maximal codes is the predictability of codes. Knowing n, the number of shift registers, and $2n$ consecutive chips of the code, one can predict the entire length of $2^n - 1$-chip code[12, 13]. This is because the feedback configuration can be realized by solving equations of the form $b_i = \Sigma_{j=1}^{n} a_j b_{i-j}$. Since there are n stages and each could be involved in the feedback configuration, n equations are needed.

$$b_{n+1} = a_1 b_n + a_2 b_{n-1} + ... + a_n b_1$$

$$b_{n+2} = a_1 b_{n+1} + a_2 b_n + ... + a_n b_2$$

$$...$$

$$b_{n+n} = a_1 b_{n+(n-1)} + a_2 b_{n+(n-2)} + ... + a_n b_n$$

The equations have overlapping blocks of $n+1$ consecutive chips from the sequence, and n equations are needed, which sets the required number of known chips to $2n$. The feedback structure of an n-stage shift register can uniquely be determined from only $2n$ chips. For example, having 20 chips of a 10-stage shift register, one could predict the entire $2^{10} - 1 = 1023$ chips of the code. This property, a disadvantage for secure spread spectrum communication, can be used as a great advantage in ranging systems. Predictable partial autocorrelation, as will be explained, uses this property in range measurement.

3.5.1 Predictable Partial Autocorrelation

In a PN code ranging system, the receiver has a replica of the transmitted signal. Furthermore, it is synchronized with the transmitter. By shifting the phase of the replica to match it to the received signal, thus achieving maximum autocorrelation, the receiver is able to measure the phase difference between the receiver and transmitter. The phase difference is proportional to the signal travel time from transmitter to the receiver, T. Specifically, the range between transmitter and receiver is measured as $R = Tc$ where c is the velocity of the signal. The length of the code must be long enough to cover the maximum possible range and avoid ambiguity. Autocorrelation of the entire period of the ranging code could take a long time. Therefore, autocorrelation on a partial code period is used to estimate the phase difference between the received signal and the replica. The full period of the code is predictable from $2n$-chip window of the period, and correlating $2n$ chips is sufficient to provide ranging information.

4 Positioning by Magnetic Maximal Sequence Codes

The properties of maximal sequence codes can be used in positioning by magnetic markers. The process of position location by reading magnetic markers is illustrated in Figure 4.

Assume that the magnetic markers in each lane are coded with a unique maximal code with a linear span of n. By reading a $2n$-chip segment of the predictable code, the entire period can be predicted, i.e., the vehicle can figure out the characteristic polynomial of the code, and build the related feedback configuration and generate a replica of the code. Furthermore, it can estimate the exact phase of the magnetic PN signal by *synchronization* of its replica with the received signal.

Measuring the signal phase, or synchronization, is possible because of the shift-and-add property of maximal length sequences: the modulo-2 sum of a code with its phase-shifted replica yields another replica with a different phase shift. Suppose a $2n$-chip block is repeated within one period. Shift a replica of the code, and add it to itself such that the two similar blocks add. This would result in a $2n$-long run of zeros while the maximum run length for zeros is $n - 1$. Therefore, a $2n$-chip block uniquely determines the code phase.

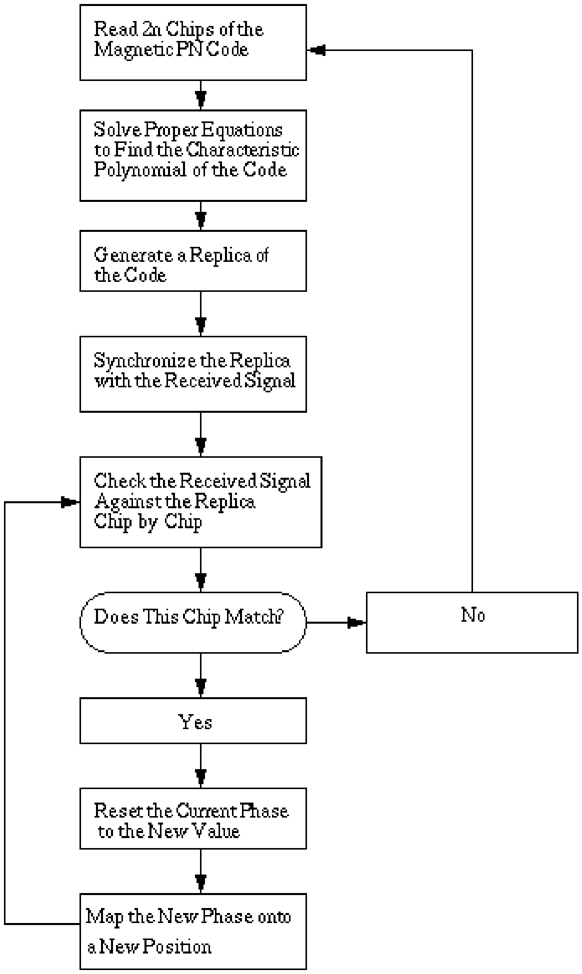

Figure 4: *Position Location by Receiving a PN Coded Magnetic Signal*

By measuring the code phase, a vehicle can have a synchronized replica of the maximal sequence code. After initialization, position updating is done by *tracking* the magnetic code and checking it against the replica. Tracking is basically done by *predictable partial autocorrelation.* The code is predictable to the receiver, and it is partially being autocorrelated with its replica. As long as the autocorrelation function results in the *maximum* value, which is equal to the number of received chips, the reading is assumed to be correct.

Synchronization allows the vehicle to detect the errors in reading the magnetic code. For example, assume the next chip in the replica is "1", but the next reading of the received signal is a "0", or vise versa. In this case an error is detected, and the receiver must read sufficient number of chips correctly before it can assert a re-synchronization. Note that *no false errors* are detected. However, error detection may not happen in real time. For example, assume that the first one in a run of five "1"s is missed by the receiver. The missed "1" is shown by * below.

$$\ldots 0 * 111100 \ldots \quad (magnetic \;\; code)$$

$$\ldots 01111100 \ldots \quad (receiver's \;\; replica)$$

The receiver will not know the missed reading until it reaches the end of the run in the received signal, and starts receiving zeros. The error in positioning is only one magnetic spacing, but it takes a five magnetic spacing trip for the receiver to detect the error.

Figure 4 shows that in case of an error, $2n$ chips must be read again. Actually, at most n chips are needed for re-synchronization, and the optimum number of chips depends on the specific code and its phase, and the probability of error.

Unlike radio signals, magnetic codes do not interfere. Furthermore, multi-path is not an issue for magnetic signal. However, noise is still an issue and will be discussed in error analysis section. It will also be discussed that in case of missing a chip how many more chips must be read before the error is detected. Then, once an error is detected, how many chips must be read correctly for re-synchronization.

5 System Components and Design

The system elements for positioning include both hardware and software elements. Some of the software components can be built as more efficient and faster hardware elements in actual implementation of the design. For example, most of the synchronization and phase tracking processes can be done by hardware.

5.1 Infrastructure

Magnetic markers are already used in the design of AHS for lateral control of automated vehicles and their binary polarity is used to code information about the road curvature, ... [14]. We are proposing a modification to the existing design, i.e., coding magnetic markers by maximal sequences. Magnetic markers that are coded by maximal length sequences provide a very accurate signal for automated vehicles to position their locations on automated roads. Magnetic markers would be similar to chips in a PN code sequence. The uniform distance between magnets sets the accuracy of positioning. The magnet spacing could be set to optimize precision versus cost.

5.2 Software

Consider the set of magnetic sequences $\mathcal{S}$ that are used for positioning by magnetic markers, and the set of road maps $\mathcal{L}$. There must be a one-to-one mapping $\theta : \mathcal{S} \longrightarrow \mathcal{L}$ such that a sequence $S \in \mathcal{S}$ of magnetic markers has a unique physical representation $L \in \mathcal{L}$. Thus, a vehicle that knows the phase of the magnetic signal can obtain its position on the road. Currently, the markers are designed to provide specific information such as road curvature [14]. A mapping that provides information about the absolute position of a vehicle on the road can be linked to other data that provide more specific information about the road such as the exit points, travel information, etc.

5.3 Synchronization and Tracking of Position

Assume that the system can identify the unique code $S \in \mathcal{S}$ of its current lane. The mapping from a magnetic PN sequence S to the physical representation of the road L is meaningful if the mapping is one-to-one, i.e., a magnetic PN sequence has only one physical interpretation. There are, say, α number of automated roads in an area, and each has M automated lanes in both directions. This makes a total of αM magnetic codes and requires the same number of physical interpretations. A vehicle must have some means of knowing the correct range of the mapping for coordination information. Thus, the codes must be different, and a vehicle must be able to specify its code. This can be done by obtaining the characteristic polynomial of a code.

Another issue in phase synchronization is that the code must be long enough to avoid any ambiguity in resolving the phase of the code, i.e., the period of the magnetic sequence must be longer than the road.

After synchronization, the position update during the motion must be tracked by predictable partial autocorrelation.

5.4 Selection of Code Length

Range measurements based on a magnetic signal is similar to radio range measurement in the sense that a set of chips have to be read by the receiver, and correlated with a replica of the signal before a partial autocorrelation is achieved.

As we already mentioned, $2n$ chips are required for predictable partial autocorrelation of a code of length $2^n - 1$. Assume that the distance between two consecutive magnetic markers is one meter. A vehicle must travel $2n$ meters before it can initialize its position and start updating it by tracking the signal. If the linear span of the code, n, is too large, it will take an unacceptedly long ($2n$-chip) travel distance (and computation time) for a vehicle to find its location. If the code period is too short such that the code period is shorter than the road length, it could cause ambiguity about the number of code periods that have repeated on the road. When a vehicle enters the road, or changes lane, it must be able to initialize its position without ambiguity. If the code is short and has been repeated for, say, m periods prior to the current position of the vehicle, the vehicle must be able to find the number m.

Another consideration for code length is the number of different codes that we would like to have for initialization. The maximum number of different codes of length $2^n - 1$ is limited by $\frac{(pf_1-1)(pf_2-1)(pf_3-1)\ldots}{2n}$, where $pf_1, pf_2, \ldots$ are the prime factors of $2^n - 1$ [12].

Initialization is possible by reading $2n$ chips. If we choose different code lengths of linear spans n_1, n_2, …, then a vehicle should read $2n^*$ chips for initialization where $n^* = max_i\{n_i\}$. Alternatively, we can choose a set of codes of equal lengths.

5.5 Examples

We provide two examples of using long and short maximal sequence codes. In both cases, we assume that there are 10 roads in a network, and each road has 4 automated lanes (in both directions), i.e., the total number of automated lanes in the network is $10 \times 4 = 40$. Furthermore, each road is less than 5000 kilometers long, and the reference point for each road is the beginning of the road. The distance between two consecutive magnetic markers is assumed to be one meter.

5.5.1 Long Maximal Code

Let $n = 21$. The prime factors of $2^{21} - 1 = 2,097,151$ are 7, 7, 127, and 337. There are $((7-1) \times (7-1) \times (127-1) \times (337-1))/(2 \times 21) = 36,288$ unique codes of the length $2^{21} - 1 = 2,097,151$ meters, which is well beyond the length of the road. A vehicle has to travel $2n = 42$ meters before identifying the new code. The feedback configuration of the road can be computed from 42 chips, and hence, the entire code can be predicted.

In case of branching roads, such as in Figure 5, each road could have its own code, or the consecutive segments $R1$, $R3$, and $R4$ could have one code, and $R2$ and $R5$ have two other different codes. Assume a vehicle is moving from $R2$ to $R3$ where the code sequence changes at point A. If the receiver is not aware of the code change, it will detect errors in partial autocorrelation and will try to re-synchronize with the code of $R2$, where it really needs to identify code $R3$ and synchronize with it.

The positioning algorithm must be modified such that if error-recovery is not achieved after reading certain number of chips, say n, then code identification and synchronization is in order. In this case, the coding at point A must be such that the starting code of $R3$ has a low partial correlation with the replica of $R2$. Hence, the receiver detects an error as soon as it starts reading $R3$ code, and will give up error recovery within n chips and will identify the new code and perform synchronization. Alternatively, the software could provide road information such that a vehicle that is going from $R2$ to $R3$ is aware of the road change and knows the new code and will only perform a synchronization upon entering $R3$.

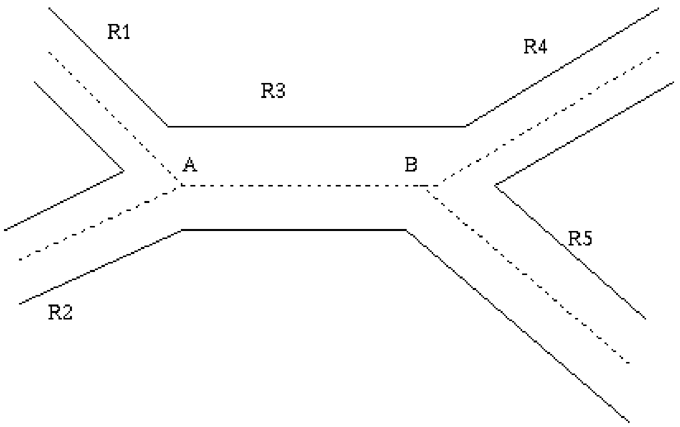

Figure 5: *Branching roads*

5.5.2 Short Maximal Code

Let $n = 11$. There are 88 different codes of the length $2^{11} - 1 = 2047$ meters. A vehicle has to travel $2n = 22$ meters for determining the code phase. Since the length of the code is much shorter than the road length, the code is repeated. The phase of the signal is $k(2\pi) + \phi$ where k is the number of past periods, and ϕ is the phase of the current period. A vehicle needs some initialization that provides information about k, the number of past code periods from the reference point. The number k can be encoded at the entry, and a vehicle keeps track of k by incrementing it every time it travels a code period is passed.

What if a fault causes a vehicle to lose its current k value? Lateral maneuvers are prohibited and a vehicle must keep traveling in one lane until k is resolved. This necessitates an updating system encoded in the magnetic signal. Each road has a total of about $(5,000,000/2047 <)2500$ code periods, which requires 12 binary bits for coding ($2^{12} = 4096 > 2500$). This information can be provided in two different ways:

1) The code can be modulated with such information, i.e., the maximal sequence code can be modulated by another code that carries the number of periods. Modulating one period of the code corresponds to about $(4096/12 >)340$ chips per bit of information for k. Since the vehicle has to travel at least 22 meters to realize its position, it may be ambiguous if the 22 meter is split between a modulated 1 and 0 of data regarding k. Increasing the minimum travel distance to 44 meters before realizing one's position insures that at least 22 meters of a code is read properly.

2)Every other code period is inverted, and then the information about k is marked on the road. Each road has a total of about 2500 code periods. This requires 12 binary bits for coding, i.e., 12 meters of coding for k must be inserted between maximal codes. How would a vehicle who has lost tracking know if it is reading k data, or the maximal code? The answer is that first increase the minimum travel time from 22 meters to 44 meters for initialization. This insures that it will read the maximal code appropriately. Second, every period preceding the k data is the compliment of the maximal code. A vehicle that reads the code will be able to tell if it is reading the maximal code, or its compliment, and hence will know when to expect the k data.

In either case, after at most 4 kilometers of travel, the vehicle will find its absolute position on the road. In case of branching roads as in Figure 5, the short codes could change as a vehicle comes from one road to another. The vehicle needs to identify the new code and synchronize its generated replica with it. The information about the number of passed periods, k, of the new road must be obtained. Mapping the phase of the current code obtains the absolute position. Inverse mapping can be used such that absolute position obtains the phase and k value of the new road segment. The inverse mapping can be used in lane change maneuvers, similarly.

The period ambiguity problem makes long maximal codes favorable.

6 Evaluation

The chip length of a PN magnetic signal is equivalent to the distance between two consecutive magnetic markers. The resolution of this scheme is one chip length. We assume that the distance between two magnetic markers is one meter. A magnetic PN signal with chip length of one meter is 30 times more accurate than the GPS signal that has a chip length of 30 meters. This is because multi-path currently limits the accuracy of GPS to one chip-length.

It is important for vehicles in the coordination layer to choose their control policy based on the information accuracy. In case that accurate information is not available, it is desired that vehicles be aware of the degree of available accuracy and perform accordingly. As an example, if the safety space gap around a vehicle for lane change maneuver is set at 30 meters, and a vehicle's degree of accuracy is ± 1 meter, it must keep a distance of 31 meters to count for a possible error.

6.1 Error Analysis

The error analysis of positioning with magnetic markers is conditioned on the fact that the lateral control of vehicle is working properly. This implies that the error cannot exceed certain amount specified by lateral control regulation. The magnetic markers are the essence of lateral control and the *error probability* p is very small, currently its specification is in the order of 10^{-3}.

The error in positioning by a magnetic signal is due to any error in receiving the signal. Such error in positioning is at most linearly proportional to the number of magnetic markers missed, or misread, or the number of noise pulses read.

6.1.1 Sources of Error

The error in receiving a magnetic signal could be due to

(I) missing a marker,

(II) misreading a marker (reading a "0" as a "1" or vise versa), or

(III) reading a magnetic noise that is not a marker and was not meant to be read.

An error analysis of the signal could be a break down on the probability of having each source of error.

6.1.2 Error Detection and Correction

Assuming one meter distance between markers, and the error probability p in reading markers, the amount of error is one meter with probability of p. Assume uniform distribution of error type (I) over a code period. Errors will not be drifted and accumulated beyond one run. Predictable partial autocorrelation is used in obtaining the phase of the code, and respectively, the position of the vehicle. Perfect partial autocorrelation can be achieved when the received signal and the synchronized replica match, i.e., $R(K,\tau) = K$ where $K \leq 2n$ is the number of received chips that the receiver can hold in the buffer. If a vehicle holds K received chips in its buffer, but $R(K,\tau) < K$, then an error has happened. However, an error may not be detected instantly. As an example, consider the following portion of a code.

...011111000...

The vehicle is reading the run of ones and is checking the received chips against its own replica of the signal, i.e., is calculating the partial autocorrelation function. As long as $R(K,\tau) = K$, where K is the number of received chips in the buffer, it will not detect an error. Suppose the first "1" of the run is missing, and the vehicle starts reading the run with the second "1". The missed marker is shown by $*$ below.

$$...0 * 1111000...$$

It will not detect the missed chip until it reaches the end of the run, i.e., starts reading the run of "0"s when it realizes that the autocorrelation is not maximized. The error is detected at this point, and will be corrected by re-synchronization[1]. At any time, the accuracy holds for the end of the previous run.

An error that happens within one run, will be detected and corrected at the next run. Consequently, probability of having an error of two meters is

$$Prob(two\ errors) \times Prob(two\ errors\ within\ a\ run) \leq p^2 n/(2^n - 1)$$

because the maximum run length is n.

(I) In case of missing markers, we analyze cases of different number of consecutive missed markers. There are exactly $2^{n-(l+2)}$ runs of length l for both "1"s and "0"s in every maximal code sequence except that there is only one run containing n ones and one containing $n-1$ zeros [12]. There are no runs of zeros of length n or ones of length $n-1$.

(I.a) Missing one marker.

In case a marker in a one-chip run is missing, it will be detected at the next marker. The error is happening in run length, and at the end of the run the receiver will notice there is an error. There are a total of $2^n/2^1$ chips next to the end of a run, ..., and $2^n/2^m$ chips are in the position of $m-1$ chips before the last chip in the run. If the last chip of a run is missing, reading the first chip of the next run will reveal that there is a missing chip, i.e., an error has occurred. If the chip next to the last is missing, a vehicle will not notice the miss until it starts reading the next run. Assume uniform probability of missing a chip in a code period. Then, with probability of $1/2$ when a chip is missing, the error will be carried during one meter travel, with probability of $1/2^2$ the error will be carried for two meters, and with probability of $1/2^m$, the error will be carried for m meters. The expected number of meters that a one meter error is carried before the error is detected is 2 meters, $(1 \times 1/2) + (2 \times 1/2^2) + \ldots + (n \times 1/2^n) \leq 2$.

Assume we are using the long maximal codes with $n = 21$. The worst case error is when a vehicle travels 20 meters before it detects an error of one meter it has been carrying in its position.

(I.b) Missing two (or more) markers.

Similar to case (I.a), the expected number of meters that a car travels before an error is detected can be calculated. However, in this case we need to know the exact code to estimate the number of magnets a vehicle reads before it detects an error.

(II) Misreading a marker will immediately result in error detection as the partial autocorrelation function will be reduced from the expected maximum value. A vehicle then can re-synchronize its receiver, and update its position.

The probability of misreading a marker can be significantly reduced by improved software algorithms or implementing hardware.

(III) A magnetic noise is due to presence of a magnetic field, i.e., a magnetic object on the road. Since magnetic field is proportional to the third power of distance, the source of magnetic noise must be physically very close to the markers on the road. There is a great chance that such sources of noise are detected and removed before the road is utilized. However, we must count for the low probability of confronting such noise. When noise is introduced, assume that lateral control remains in place and vehicle continues to read the magnetic markers. Furthermore, assume that a noise has equal probability of being a zero or one. If the polarity of the noise chip matches the polarity of the current run, it will not be detected until the end of the run. This is similar to error detection in case (I.a). If a noise with opposite polarity of the current run happens, it will be immediately detected, similar to the case (II).

6.2 Accuracy

Current specifications for lateral control of vehicles requires an error rate less than 0.1% in receiving the magnetic signal [14], i.e., only one out of a thousand readings of magnetic markers could be erroneous due to the combined effects of all sources of noise.

As explained, the errors do not accumulate, and beginning of each run confirms lack or existence of error in the previous run. In other words, an error could be carried only during one run without

---

[1] The exact method of correction for optimized performance is subject of further research.

being detected. This may be another consideration in choosing n, the linear span of the code, since the maximum run length within a period is equal to n. Furthermore, during each run, a vehicle can have a maximum of $h - j$ chip errors, where h is the length of the run, and j is the number of received chips in the run. Recall that a quarter of chips in one period of the code have a run length of 1, which improves the accuracy of the system.

7 Improvement with Inertial Navigation System (INS)

INS [15, 16, 17] is a set of inertial instruments that are able to measure acceleration and angular velocity (or angular acceleration) of the body to which INS is attached. In order to obtain position information, the measurements of INS must be mathematically integrated. A small deviation in initial conditions will produce largely growing errors with integration over time. The common sources of error in INS are bias drift of accelerometers and gyros, various misalignments, non-linearities, cross-axis sensitivities, etc.

INS provides high update rates and accurate measurements of *relative* motion parameters and relative displacements. However, it needs to be periodically corrected for its growing errors and provided with relatively accurate initial conditions. Small deviations in initial conditions become the cause of error growth during numerical integration. In order to correct these errors INS is integrated with some external positioning system.

INS can be used in conjunction with magnetic PN signal to minimize possible errors. Since INS is relatively accurate in measuring displacement over short distances, it can be used to periodically check the accuracy of positioning by magnetic signal. It is specially useful in checking the accuracy within each run. The error of INS can be zeroed once the vehicle is sure that it is accurately reading the code. Every time a run of "0" or "1" ends, a receiver can assert that its reading at the end of the last run has been accurate.

8 Cost Analysis

The cost of positioning by PN coded magnetic markers can be divided into implementation, maintenance, and application costs.

8.1 Implementation

The implementation cost includes the following cost items.

Infrastructure The magnetic markers are already an element of the AHS design, and their installation cost is already included in the AHS design. However, the exact polarity and location of markers must be recorded for the purpose of mapping the signal phase to absolute position.

Software (1) A code identification algorithm to resolve the feedback configuration of each code from $2n$ chip must be implemented in the receiver. Similarly, algorithms for synchronization, tacking, and error detection and correction must be implemented.

(2) A one-to-one mapping of each magnetic code sequence phase to the absolute position on the road, $\mathcal{S} \longrightarrow \mathcal{L}$, must be provided to each vehicle that uses the AHS network.

Noise reduction In order to reduce the error in the positioning system, it is important to check the automated roads for possible sources of magnetic noise during implementation. This may include costs such as removing a magnetic core from the road asphalt, or displacing some metallic borders around bridges, etc. Alternatively, such permanent sources of error could be included in the magnetic code and the related software. Noise reduction is equally important for lateral control of vehicles, and its cost may be a part of already predicted costs of AHS.

8.2 Maintenance

Periodic maintenance of the roads is necessary to ensure correct polarity of magnetic markers. The roads must also be checked for any new sources of magnetic noise.

8.3 Operation

A receiver's operation includes the following costs.

Code Identification - Every time a vehicle enters an automated road, or changes lane, it must find the characteristic polynomial of the current lane's maximal sequence code. The complexity of an algorithm for finding the characteristic polynomial of a code of linear span n is presented below. A more sophisticated algorithm could decrease the complexity.

Maximal sequence codes always have a feedback on the last stage. Also, the number of feedbacks is even. Recall the set of n equations that reading $2n$ consecutive chips provide. The calculations for finding the feedback configuration from n equations involves the following steps.

(i) For each equation, consider the set of possible feedbacks, i.e., the set of possible polynomials' coefficients.

(ii) Check the set of possible feedback taps of each equation for other $n-1$ equations. If a feedback tap set is not valid for all equations, eliminate it.

(iii) Eliminate the sets with odd number of elements.

(iv) For the remaining sets of possible feedbacks, find the intersection of all feedback tap sets.

This is the conventional approach for breaking security codes by trying the set of potential possibilities. However, the actual number of possibilities for feedback configurations in an automated road network is equal to the number of automated lanes, say η. The set of η possible codes can be initially provided to a vehicle. A vehicle needs to try each configuration for n equations in the worst case of search. This amounts to $\eta \times n \times (n+1)$ arithmetic operations.

Synchronization - Next, the phase of the code must be determined. First, a replica of the code is generated using its characteristic polynomial. By maximizing autocorrelation of the replica with the received $2n$ chips, the phase can be determined. Maximizing $R(K, i, \tau) = \Sigma_{j=0}^{K-1} b_{i+j} b_{i+j+\tau}$ over $\tau \in \{0, 2^n - 1\}$ requires $2^n - 1$ calculations. This complexity can significantly be reduced by providing some information that narrows the window of autocorrelation from the entire code period to a relatively small partial period of the code. For example, a vehicle that already has the absolute position of the current lane and wants to change lane, can use the inverse mapping of the new lane's code phase to absolute position, $\theta^{-1} : L \mapsto S$, and narrow the autocorrelation window in the next lane.

Tracking - Having a replica of the code, the receiver checks the match between the received chip, and the anticipated chip. This is one computation per chip length.

9 Comparison to GPS

The following is a comparison between positioning by GPS and magnetic PN codes.

9.1 GPS Signal

GPS [4, 3, 18, 19, 20] uses Gold codes which are composites of maximal sequence codes and their autocorrelation and cross-correlation properties are very similar to maximal sequence codes. We only discuss high precision GPS mode which is now available to civilian use.

Precision or P code is at fundamental frequency of $f_0 = 10.23$ MHz. Each satellite has a unique P code. Similar to magnetic codes, the unique code sequence reveals the identity of the transmitter and the position of the reference point.

P code is product of two PN codes. The combination of two sequences has a results in a code with approximately 2.3547×10^{14} bits which corresponds to a time span of about 266.4 weeks. By taking advantage of autocorrelation properties of the code, the total code length is partitioned to 37 unique one-week segments and each segment is assigned to a satellite defining its P code which is non-overlapping with every other satellite. Because of high ID of autocorrelation function of the original code, the PN codes of all satellites are orthogonal to each other [21] and the cross-correlation between the signals is very low.

The chip of P code is about $30m$ (the time interval between two chips is 1/10.23MHz which is about 0.1 microsecond, so the chip length is $(3 \times 10^8 \times 0.1\mu sec = 30m)$.

9.2 GPS Receiver

A GPS receiver obtains its absolute position by doing at least three range measurements. Assuming that the receiver and transmitters are synchronized, range measurement is done as the following. The receiver generates a replica of the received signal. It then shifts the phase of the replica to maximize the signal's autocorrelation function. The phase shift of the replica corresponds to the travel time of the received signal, and the range can be measured by $R = Tc$.

Synchrony among transmitters is constantly verified by ground control stations and satellite communication [5]. The assumption about synchrony between the receiver and transmitters is circumvented by considering time as another unknown at the receiver. This would add time as one more unknown parameter to the set of three unknowns position coordinates. Thus, a minimum of four satellites are needed to be visible by a GPS receiver. With latest GPS receivers, the user can receive simultaneous transmissions from up to 12 satellites.

9.3 Accuracy

Higher accuracy than a chip length is achievable by maximizing the autocorrelation function within a chip. GPS uses phase detectors to achieve an accuracy equivalent to a small fraction of a chip, i.e., it can detect a phase difference within one degree (1/360 of a chip) [22].

9.4 Errors

The sources of error for GPS can be categorized in two groups: (i) the atmospheric effects, relativistic effects of Earth's gravity and motion, ephemeris errors, etc. (ii) GPS geometry, and multi-path phenomena.

The first groups of errors can be nulled by using a correction signal from a nearby differential station. For high accuracy, the GPS receiver must be within one kilometer distance of a differential station [23].

The second group of errors cannot be overcome by a correction signal. GPS receiver requires a satellite to be in its line-of-sight for accurate range measurement. The accuracy of GPS is due to the fact that a signal's direct path can be measured very accurately to a fraction of microsecond. If the signal does not reach the user in direct path, then the measurements and calculations based on path measurements are not accurate. If a user is blocked from signal in a tunnel or under an over-pass, it will not receive direct signal and hence will not have accurate measurements of its position.

Even when there is a direct signal received by the user, the existence of indirect signals and multi-path is a major problem for radio positioning systems. There is no general model of the multi-path effect because of the arbitrary different geometric situations. Modern technology can resolve multi-path by multi-antenna tracking only if the indirect path is one wave-length longer than the direct path. If the reflected signal that reaches the user is within one chip-length from the actual signal, the receiver cannot resolve the position accurately [12].

The chip length of the signal that is used for position location in GPS is $30m$. This length could match the difference between a direct and reflected signal off the freeway sound-absorbing walls, or a tall truck. Multi-path is very difficult to characterize and overcome, and hence, it remains a big obstacle in using it for high accuracy in moving systems.

The best estimate for the accuracy of GPS on AHS is $30m$ chip length of the signal. This is mainly because the effect of multi-path can be encountered only down to one chip length, and a resolution beyond that is not possible with the current technology.

9.5 Fault Detection

Erroneous range measurements due to multi-path effect cannot be detected. As a result, GPS fails to detect errors because multi-path error can change continuously.

9.6 Cost

The entire cost of GPS, implementation and operation, is included in the cost of the receiver.

9.7 GPS Integrated with INS

Could GPS be integrated with INS to achieve the high accuracy demanded by the coordination layer of AHS? The problem with such integration for AHS is that the two systems have uncorrelated errors that cannot cancel each other. INS provides an estimate $z(t)$ of the position $x(t)$ such that $z(t) = x(t) + \eta(t)$, where $\eta(t)$ is an error that grows by integration over time. GPS on the other hand provides an estimate $y(t)$ of the position $x(t)$ that has a random error $\epsilon(t)$ due to multi-path, where ϵ is a random function of time, and its probability distribution is not characterized. $\eta(t)$ and $\epsilon(t)$ are uncorrelated

10 Conclusion

Magnetic PN signals can be used for positioning by vehicles in AHS. While the code structure and range measurement technique is very similar to GPS, the accuracy of this technique is remarkable. A novel feature of our proposed scheme is its error detection and correction capabilities. In summary, we compare positioning by PN coded magnetic markers with GPS in the following.

Signal and code Both GPS and magnetic positioning systems use PN codes with high correlation and autocorrelation ID, while one uses digital radio waves, and the other uses magnetic pulses.

Receiver The technology of receiving PN codes, and detecting the phase, and mapping it to a physical position is very similar in both receivers. GPS receivers integrates the entire hardware and software technology, but it depends on differential stations for high accuracy. The components of the magnetic positioning system, magneto-meter and software, may not be physically integrated, but make an independent system.

Accuracy Positioning by magnetic PN codes has an accuracy of 1 meter, which is better than the accuracy of GPS (integrated with INS).

Errors Error probability distribution can be tabulated for magnetic positioning. The main source of error for GPS, multi-path, cannot be characterized.

Fault detection - Unlike GPS, magnetic positioning system has the ability to detect errors and correct them.

Cost In terms of cost, GPS receiver has a higher receiver cost but it imposes no cost on the vehicle's computer.

Combining the magnetic positioning system with INS will result in a very high accuracy positioning system for AHS. Such an integrated system will have negligible errors due to high accuracy of INS in measuring displacement. The errors of positioning by magnetic PN coded signal can be detected and corrected by INS in real time.

References

[1] R. B. Langley, "The mathematics of GPS," *GPS World*, July/August 1991.

[2] M. I. Skolnik, *Introduction to Radar Systems.* McGraw-Hill, Inc., 1980.

[3] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *GPS Theory and Practice.* Springer-Verlag, 1994.

[4] J. J. S. Jr., "GPS signal structure and performance characteristics," *Papers published in Navigation*, vol. 1, pp. 29–54, 1980.

[5] R. B. Langley, "Time, clocks, and GPS," *GPS World*, November/December 1991.

[6] J. Michels and B. Burda, "AVL with GPS and Arcview 2, Park, Ride & Fly SW: A case study," tech. rep., Radio Satellite Integrators, Inc., Torrance, CA, 1994.

[7] E. N. Skomal, *Automatic Vehicle Locating Systems.* Van Nostrand Reinhold Company, 1981.

[8] C. J. Driscoll, "Vehicle tracking- Is this technology useful for local trucking fleets?," tech. rep., C. J. Driscoll and Associates, Rancho Polos, CA, 1994.

[9] G. Lachapelle and B. Townsend, "En-route coverage validation and calibration of LORA N-C with GPS," *GPS World*, March 1990.

[10] R. B. Langley, "The federal radio-navigation plan," *GPS World*, March 1992.

[11] B. Forsell, *Radio-navigation Systems.* Prentice Hall, 1991.

[12] R. C. Dixon, *Spread Spectrum Systems with Commercial Applications, Third Edition.* Wiley Interscience, 1994.

[13] M. K. Simon and J. K. Omura, *Spread Spectrum Communications Handbook, Revised Edition.* McGraw Hill, Inc., 1994.

[14] H.-S. Tan, R. Rajamani, and W.-B. Zhang, "Demonstration of an automated highway platoon system," *American Control Conference (ACC)*, pp. 1823–1827, 1998.

[15] K. R. Britting, *Inertial Navigation Systems Analysis.* New York Wiley-Interscience, 1971.

[16] V. D. Andreev, *Theory of inertial navigation: aided systems.* Springfield, 1969.

[17] A. B. Chatfield, *Fundamentals of high accuracy inertial navigation.* New York Wiley-Interscience, 1997.

[18] E. G. Blackwell, "Overview of differential GPS methods," *Papers published in Navigation*, vol. 3, pp. 89–100, 1980.

[19] R. B. Langley, "Why is GPS signal so complicated?," *GPS World*, May/June 1990.

[20] A. J. Van Dierendock, "Understanding GPS receiver terminology: A tutorial," *GPS World*, January 1995.

[21] R. B. Langley, "The GPS receiver: An introduction," *GPS World*, May/June 1990.

[22] A. Kluesberg, "Precise differential positioning and surveying," *GPS World*, July/August 1992.

[23] Available via WWW URL from http://www.ashtech.com/, April 1998.