Traffic Detector Handbook:

Third Edition—Volume I

Publication No. FHWA-HRT-06-108 October 2006

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

The objective of the third edition of the *Traffic Detector Handbook* is to provide a comprehensive reference document to aid the practicing traffic engineer, planner, or technician in selecting, designing, installing, and maintaining traffic sensors for signalized intersections and freeways. Judicious application of the concepts and procedures set forth in the Handbook should result in improved installations and operations of traffic sensors and a long-term savings of public funds.

Sensor types include both in-roadway and over-roadway sensors. Topics covered include sensor technology, sensor applications, in-roadway sensor design, sensor installation techniques and sensor maintenance. The sensor technology chapter discusses the operation and uses of inductive loop detectors, magnetic sensors and detectors, video image processors, microwave radar sensors, laser radars, passive infrared and passive acoustic array sensors, and ultrasonic sensors, plus combinations of sensor technologies. Sensor application topics include safety, operation, multimodal issues, and physical and economic factors that affect installation and performance. The appendixes include a variety of research, background papers, and implementation guidance. The information contained in this Handbook is based on the latest research on available treatments and best practices in use by jurisdictions across the United States and elsewhere. References are provided for the student, practitioner, researcher, or decisionmaker who wishes to learn more about a particular subject.

The third edition is published in two volumes, of which this is the first, Volume I (FHWA-HRT-06-108), containing Chapters 1 through 4. Volume II (FHWA-HRT-06-139) contains Chapters 5 and 6 and all Appendixes.

Antoinette Wilbur, Director
Office of Operations
Research and Development

NOTICE

QUALITY ASSURANCE STATEMENT

Technical Report Documentation Page

1. Report No. FHWA-HRT-06-108	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Traffic Detector Handbook: Third Edition—Volume I		5. Report Date October 2006
		6. Performing Organization Code
7. Author(s) Principal Author: Lawrence A. Klein Coauthors: Milton K. Mills, David R.P. Gibson		8. Performing Organization Report No.
9. Performing Organization Name and Address Lawrence A. Klein 3 Via San Remo Rancho Palos Verdes, CA 90275		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. Contract No. DTFH61-03-P00317
12. Sponsoring Agency Name and Address Federal Highway Administration Turner-Fairbank Highway Research Center 6300 Georgetown Pike, HRDO-04, Room No. T204 McLean, VA 22101-2296		13. Type of Report and Period Covered Implementation Package
		14. Sponsoring Agency Code

15. Supplementary Notes

David R.P. Gibson (David.Gibson@fhwa.dot.gov) at the Turner-Fairbank Highway Research Center (http://www.tfhrc.gov) was the Technical Representative for the Federal Highway Administration (FHWA). Milton K. Mills, Advanced Research Team FHWA, contributed many technical papers. Traffic sensor researchers and practitioners contributed significantly to document organization, content, and exhibits. The peer review panel, consisting of Darcy Bullock of Purdue University, Dan Middleton of Texas Transportation Institute, and Tom Urbanik of the University of Tennessee, provided peer review and advice based on their use and testing of sensors. Tom Potter of Reno A&E gave technical advice on the electrical functioning of loop detectors. Sensor vendors provided technical information and exhibits on their technologies. In addition, many FHWA staff members participated as sensor group members and/or provided comments throughout the project, including Pamela Crenshaw, Ralph Gillman, Peter Huang, David Jones, and Raj Ghaman.

16. Abstract

The objective of this Handbook is to provide a comprehensive resource for selecting, designing, installing, and maintaining traffic sensors for signalized intersections and freeways. It is intended for use by traffic engineers and technicians having responsibility for traffic sensors, whether in-roadway or over-roadway sensors. These two families of sensors have different characteristics and thus corresponding advantages and disadvantages that are discussed throughout the Handbook. Topics covered include sensor technology, applications, in-roadway sensor design, installation techniques, and maintenance. The sensor technology chapter discusses the operation and uses of inductive loop detectors, magnetic sensors and detectors, video image processors, microwave radar sensors, laser radars, passive infrared and passive acoustic array sensors, and ultrasonic sensors, plus combinations of sensor technologies. The sensor application topics addresses safety, operational performance, multimodal issues, and physical and economic factors that the practitioner should consider. Appendixes include research, background papers, and implementation guidance. The information contained in this Handbook is based on the latest research available on treatments and best practices in use by the surveyed jurisdictions. References are provided for the student, practitioner, researcher, or decisionmaker who wishes to learn more about a particular subject.

The third edition is published in two volumes, of which this is the first, Volume I (FHWA-HRT-06-108), containing Chapters 1 through 4. Volume II (FHWA-HRT-06-139) contains Chapters 5 and 6 and all Appendixes.

17. Key Words		18. Distribution Statement
Traffic detectors, sensors, detector installation, detector maintenance, signalized intersections, intersection safety, intersection treatments, infrared sensor, inductive loop, magnetometer, video image processor, microwave radar sensor, laser radar sensor, acoustic sensor, ultrasonic sensor, magnetic sensor.		No restrictions. This document is available to the public through the National Information Technical Service, Springfield, VA, 22161 and Research and Technology Product Distribution Center, 9701 Philadelphia Court, Unit Q, Lanham, MD 20706; telephone: 301–577–0818; fax: 301–577–1421.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	288	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

CONTENTS OF VOLUME I

CONTENTS OF VOLUME II

FIGURES IN VOLUME I

FIGURES IN VOLUME II

TABLES IN VOLUME I

TABLES IN VOLUME II

There are no Tables in Appendix F.

TABLE OF CONTENTS: CHAPTER 1.

LIST OF FIGURES: CHAPTER 1.

LIST OF TABLES: CHAPTER 1.

CHAPTER 1. INTRODUCTION

This edition of the *Traffic Detector Handbook* (the *Handbook*) is an updated version of the previous *Handbook* originally published as Implementation Package FHWA-IP-85-1, and supersedes the previous two editions. While the basic philosophy of the original document has been retained, the *Handbook* has been restructured and revised to update discussions of concepts and equipment that reflect the state of the practice, particularly as they relate to incorporation of new sensor and controller technologies in traffic management applications.

SCOPE AND OBJECTIVES OF THE HANDBOOK

The overall objective of the *Handbook* is to provide a reference to assist the practicing engineer and technician in planning, designing, installing, and maintaining vehicle sensors that support traffic management on surface streets, arterials, and freeways. In accordance with this objective, the *Handbook*:

The objective of the Traffic Detector Handbook is to provide a single resource and basic reference for the design, installation, and maintenance of traffic flow sensors.

- Provides a compendium of existing sensor technology.
- Facilitates the understanding of basic sensor technology.
- Aids in the application of new sensor technology.
- Identifies the best current practices.
- Serves as a training aid for traffic engineers, technicians, and field personnel.

The National Electrical Manufacturers Association (NEMA) Standards define a vehicle detection system as ".... a system for indicating the presence or passage of vehicles." These systems provide traffic flow data for traffic-actuated signal control, traffic-responsive signal control, freeway surveillance and traffic management, and data collection systems.

A traffic flow sensor is a device that indicates the presence or passage of vehicles and provides data or information that supports traffic management applications such as signal control, freeway mainline and ramp control, incident detection, and gathering of vehicle volume and classification data to meet State and Federal reporting requirements.

The *Traffic Detector Handbook* describes the theory of operation, installation, and applications of in-roadway and over-roadway sensors. An in-roadway sensor is one that is placed as one of the following ways:

- Embedded in the pavement of the roadway.
- Embedded in the subgrade of the roadway.
- Taped or otherwise attached to the surface of the roadway.

In-roadway and over-roadway sensor technologies and applications are described in this Handbook.

By contrast, an over-roadway sensor is one that is mounted above the surface of the roadway in one of the following two ways:

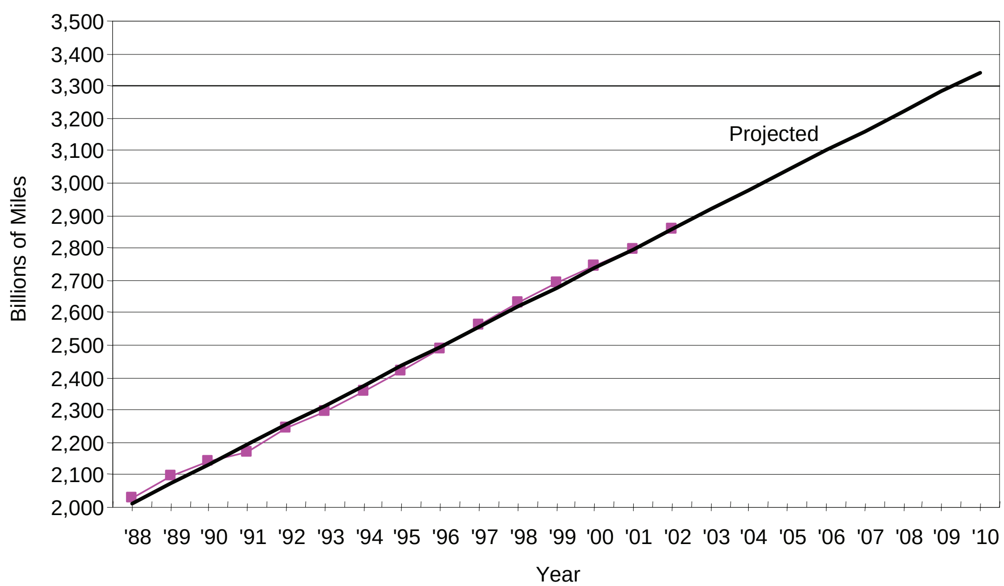
- Above the roadway itself.
- Alongside the roadway, offset from the nearest traffic lane by some distance.

Examples of in-roadway sensors include inductive-loop detectors, which are sawcut into the pavement; magnetometers, which may be placed underneath a paved roadway or bridge structure; and tape switches, which are mounted on the roadway surface. Examples of over-roadway sensors are video image processors that utilize cameras mounted on tall poles adjacent to the roadway or traffic signal mast arms over the roadway; microwave radar, ultrasonic, and passive infrared sensors mounted in a similar manner; and laser radar sensors mounted on structures that span the lanes to be monitored. Some emerging applications for wide area surveillance envision over-roadway sensors mounted on tall buildings and radio towers near the roadway and on aerial platforms.

THE NEED FOR SENSORS IN MODERN TRAFFIC MANAGEMENT SYSTEMS

Maximizing the efficiency and capacity of existing transportation networks is vital because of the continued increase in traffic volume and the limited construction of new highway facilities in urban, intercity, and rural areas. In the United States for example, highway miles traveled increased by 33 percent while public road mileage increased by less than 2 percent from 1987 to 1997.[1,2] Figure 1-1 shows the projected growth in U.S. highway demand to the year 2010 in terms of highway miles traveled per year.[2,3] The increase in demand, relative to the limited construction of new roads, has caused recurring congestion in the U.S. and throughout the industrialized world, as well as in developing nations.

Figure 1-1 shows the growth in U.S. highway demand from 2 trillion miles (mi) (3.2 trillion kilometers (km) in 1988, to 2.600 trillion mi (4.1 trillion km) in 1998, to a projection of over 3.3 trillion miles (5.3 trillion km) by 2010 in terms of highway miles traveled per year.[2,3] The increase in demand relative to capacity has caused recurring congestion. The annual growth is approximately linear.



1 mile (mi) = 1.6 kilometers (km)

Figure 1-1. Growth in highway miles traveled in the United States. The 2003–2010 trend is based on linear extrapolation of previous years' data.

Even when additional facilities are built to ease congestion and promote the use of multiple occupancy vehicles, the cost is often quite high. For example, the freeway-to-freeway high-occupancy vehicle (HOV) bypass structure, illustrated in Figure 1-2, costs approximately $150 million (U.S.) to construct. Construction and striping of HOV freeway lanes in Los Angeles County vary between $400,000 and $750,000 (U.S.) per lane mile ($640,000 to $1,200,000 per lane km), depending on the freeway configuration. Striping only costs $100,000 (U.S.) per lane mile ($160,000 per lane km). The latter estimates include costs associated with providing safety for the construction workers and building temporary roadbeds to maintain traffic flow during construction.

An alternative to expensive new highway construction is the implementation of strategies that promote more efficient utilization of current road, rail, air, and water transportation facilities. These strategies are found in Intelligent Transportation Systems (ITS) roadway and transit programs that have among their goals reducing travel time, easing delay and congestion, improving safety, and reducing pollutant emissions. ITS that contain electronic surveillance, communications, and traffic analysis and control technologies bring benefits to transportation system users and managers. Users gain from the information and guidance provided by ITS. Transportation managers and agencies profit from improved ability to monitor, route, and control traffic flows and disseminate information.

Figure 1-2 shows a freeway-to-freeway high occupancy vehicle (or HOV) bypass lane ramp structure under construction at intersection of CA–57 and CA–91 freeways in Anaheim, CA. HOV bypass lanes are a way of encouraging use of carpooling to reduce congestion.

Figure 1-2. Freeway-to-freeway HOV bypass lane structure under construction at intersection of CA–57 and CA–91 freeways in Anaheim, CA (Photograph courtesy of Lawrence A. Klein).

Millions of research and operations dollars are budgeted for managing traffic and alleviating congestion and delay on the Nation's existing streets and freeways. ITS applications for Advanced Traffic Management Systems, Advanced Traveler Information Systems, Commercial Vehicle Operations, Advanced Vehicle Control Systems, Advanced Public Transit Systems, and Archived Data User Services rely on traffic flow sensors to provide vehicle detection, incident detection, automatic traveler surveillance, real-time traffic adaptive signal control, archival data, and data for traveler, commercial, and emergency information services. The success of these intelligent transportation systems depends to a large extent on the proper design, installation, and maintenance of the sensor component of the overall system. Consequently, it is incumbent on the jurisdictions or agencies implementing or operating ITS to assure that appropriate attention is directed toward this relatively straightforward, but critical system element.

EVOLUTION OF TRAFFIC FLOW SENSOR TECHNOLOGY

Charles Adler, Jr., of Baltimore, MD, developed the first traffic sensor based on detecting acoustic energy from the vehicle's horn.

In the 1920s, when manually operated traffic signals were being replaced by automatic, pretimed traffic signal control devices, engineers soon realized they needed a method to collect the traffic data previously obtained visually by the police officer on duty. Among those concerned was Charles Adler, Jr., of Baltimore, MD, a railway signal engineer. He developed a sensor that was activated when a driver sounded his car horn at an instrumented location. This device consisted of a microphone mounted in a small box on a nearby utility pole. First installed in 1928 at a Baltimore intersection, Adler's device enabled the first semiactuated signal installation to assign right-of-way by means of a vehicle sensor.

The first vehicle passage sensor used a pressure-sensitive device.

At nearly the same time, Henry A. Haugh, an electrical engineer, developed an in-roadway pressure-sensitive sensor, utilizing two metal plates that acted as electrical contacts. The wheel pressure of passing vehicles brought the plates together. This pressure-sensitive, treadle type sensor proved more popular than the horn-activated sensor. In fact, this sensor enjoyed widespread use for over 30 years as the primary means of detecting vehicles at actuated signals.

Adler continued his work with sound detectors and in l931 introduced another sound detector, which employed hollow steel boxes embedded in the intersection approach. These boxes picked up the sound of passing wheels, which was transmitted to microphones.

Mechanical problems with the contact-plate sensor led to the introduction of the electro-pneumatic sensor. Although this device found some application, it was costly to install, capable of only passage (motion) detection, and its (axle) counting accuracy was limited by the generation of air pressure waves and capsule contact bounce.

In retrospect, it seems unfortunate that the treadle detector, which utilized the most obvious and most easily detected property of vehicles—their weight—could not be economically produced. Snow plows could lift the plate from the roadway, resulting in costly repairs. There was also the expense of reinstalling the detector after roadway resurfacing. These problems led to the search for traffic flow sensors based on more subtle properties such as:

- Sound (acoustic sensors).

- Opacity (optical and infrared sensors and video image processors).
- Geomagnetism (magnetic sensors, magnetometers).
- Reflection of transmitted energy (infrared laser radar, ultrasonic sensors, and microwave radar sensors).
- Electromagnetic induction (inductive-loop detectors).
- Vibration (triboelectric, seismic, and inertia-switch sensors).

Not all of these concepts have been commercially exploited. Today, the inductive-loop detector is, by far, the most widely used sensor in modern traffic control systems. Magnetometers, magnetic sensors, video image processors, microwave and laser radar sensors, ultrasonic, acoustic, and passive infrared sensors are also produced commercially and used for various traffic management applications. The optical sensor has found use for detecting priority and overheight vehicles.

NEED FOR SENSOR ALTERNATIVES

While single inductive-loop detectors give direct information concerning vehicle passage and presence, other traffic flow parameters such as density and speed must be inferred from algorithms that interpret or analyze the measured data. When these parameters are calculated from inductive loop data, the values may not have sufficient accuracy for some applications (such as rapid freeway incident detection) or the available information may be inadequate to support the application (such as calculation of link travel time). Furthermore, the operation of inductive-loop detectors is degraded by pavement deterioration, improper installation, and weather-related effects. Street and utility repair may also impair loop integrity. Thus, a good loop installation, acceptance testing, repair, and maintenance program is required to maintain the operational status of an inductive-loop-based vehicle detection system.

Evaluations of modern over-roadway sensors show that they provide an alternative to inductive-loop detectors. The traffic flow parameters measured with over-roadway sensors satisfy the accuracy requirements of many current freeway and surface street applications, provided suitable mounting is available. The mounting location must provide an unobstructed view of vehicles for optimum performance. In general, when sensors are installed over the lane of traffic they are intended to monitor, their view and hence their data collection ability is not occluded by other vehicles that are present within the viewing area of the sensor. Over-roadway sensors that are mounted on the side of a roadway and view multiple lanes of traffic at angles perpendicular to or at an oblique angle to the flow direction may experience two types of data anomalies. The first occurs when tall vehicles block the sensor's view of distant lanes. The occlusion may potentially cause an undercount or false average speed measurement. The second anomaly occurs when tall vehicles project their image into adjacent lanes. When a sensor is sensitive to this effect, it will overcount and again may report a misleading average speed. Thus, sensor type, mounting height and location, vehicle mix, road configuration, and sensor viewing angles must be analyzed with respect to the intended application. Some over-roadway sensors may be more susceptible to these anomalies than others.

Installation and maintenance of over-roadway sensors mounted over the traffic lanes they monitor may require lane closure for bucket trucks to be parked on the mainline. This can disrupt traffic and pose a safety risk to the installers. Traffic should be managed in accordance with "Part 6—Temporary Traffic Control," as found in the *Manual on Uniform Traffic Control Devices* (MUTCD).[4]

The issue of how to handle sensor data more intelligently rather than as simple contact closures is now beginning to be addressed by researchers. See especially the work with NCHRP 3-66 "Traffic Signal State Transition Logic Using Enhanced Sensor Information" and Small Business Innovative Research Project DTRS57-03-R-SBIR "Real Time Linux Operating System Software for Advanced Traffic Controller to Host Traffic Control Software."

Most over-roadway sensors have relay or solid state outputs that are compatible with systems that accept inductive loop data. Some also have serial outputs that directly provide multilane traffic volume, occupancy, speed, vehicle length, and classification that are not ordinarily available from inductive-loop detectors. As there are presently no generally accepted standards for serial data formatting, software code called drivers must be written before the serial data stream can be decoded by the field controller or central computer at a traffic management center. A standard National Transportation Communications for ITS Protocol (NTCIP) protocol is needed to ameliorate this situation.

An emerging potential source of traffic flow data is from cellular telephone companies who monitor the transmitting status of telephones that are engaged in conversations in support of the wireless enhanced all automatic location identification (ALI) directive of the Federal Communications Commission (FCC). This directive mandates providing the caller's location to within:

- 328 feet (ft) (100 meters (m)) for 67 percent of the calls and 984 ft (300 m) for 95 percent of the calls when network-based solutions are implemented.
- 164 ft (50 m) for 67 percent of the calls and 492 ft (150 m) for 95 percent of the calls when handset-based solutions are implemented.

The location of these telephones can potentially be made available to traffic management agencies to anonymously track vehicles on a noncooperative basis. This information can assist in estimating congestion and travel time over wide areas, while protecting the identity of the telephone subscriber. This method of gathering traffic flow data is outside the scope of the handbook and, therefore, will not be discussed further.

Another unconventional source of traffic monitoring data is from non-stationary and airborne platforms. Information gathered from satellite, aircraft, and unmanned aerial vehicles can be used to estimate arterial and freeway traffic characteristics over long time scales and large geographic areas, including those where data were previously unavailable. The spatial coverage provided from air- and satellite-based sensors can potentially support the development of new metrics that better represent highway utilization and congestion.

Although the installation and maintenance of in-roadway sensors such as inductive-loop and magnetic field sensors can disrupt traffic and pose a safety risk to the installers, a requirement for in-roadway sensors continues for several reasons. These include aesthetic considerations that dictate their use for traffic management when over-roadway sensors are excluded from consideration, axle counting and weigh-in-motion applications requiring sensors (such as pneumatic tubes, fiber-optic, bending plates, piezoelectric, pressure sensitive resistance, load cells, and capacitance mats) under or on the road surface, cost and safety issues associated with mounting over-roadway sensors where existing structures are not available, and policy that prohibits over-roadway sensors in certain locations. Newly installed inductive-loop detectors may also provide more accurate data than over-roadway sensors when they are coupled with advanced electronics units available from several manufacturers. Sensors used in weigh-in-motion applications are not described in this *Handbook*, but information concerning their operation and installation may be found elsewhere.[5,6]

SENSOR TECHNOLOGY CHARACTERISTICS

Table 1-1 compares the strengths and weaknesses of current sensor technologies with respect to installation, parameters measured, and performance in inclement weather, variable lighting, and changeable traffic flow. Most over-roadway sensors are compact and mounted above or to the side of the roadway, making installation and maintenance relatively easy. Some sensor installation and maintenance applications may require the closing of the roadway to normal traffic to ensure the safety of the installer and motorist. All the sensors listed operate under day and night conditions.[5]

Table 1-2 lists the types of data typically available from each sensor technology, coverage area, communication bandwidth requirements, and purchase costs.

Several technologies are capable of supporting multiple lane, multiple detection zone applications with one or a limited number of units. These devices may be cost effective when larger numbers of detection zones are needed to implement the traffic management strategy.

A low to moderate communication bandwidth is indicated if only data and control commands are transmitted between the sensor, controller, and traffic management center. Larger bandwidth is required if real-time video imagery is transmitted at 30 frames/second (s). The requirement for large bandwidth communications media such as T1 telephone lines, which support transmission rates of 1.544×10^6 bits/s (baud) at a bandwidth of 125 Megahertz (MHz), and fiber can be reduced if compressed imagery (e.g., transmission rates of 256,000 bits/s at a bandwidth of 20.5 MHz) is suited for the application. The required transmission rate increases when large numbers of sensors, roadside information devices such as changeable message signs and highway advisory radio, signal timing plans, and traveler information databases are used to implement traffic management strategies.

The range of purchase costs for a particular sensor technology reflects cost differences among specific sensor models and capabilities. If multiple lanes are to be monitored and a sensor is capable of only single lane operation, then the sensor cost must be multiplied by the number of monitored lanes.

Table 1-1 summarizes the strengths and weaknesses of inductive loop, magnetometer, microwave radar, active infrared, passive infrared, ultrasonic, acoustic, and video image processor sensors.

The good performance of in-roadway sensors such as inductive loops, magnetic, and magnetometer sensors is based, in part, on their close location to the vehicle. Thus, they are insensitive to inclement weather due to a high signal-to-noise ratio. Their main disadvantage is their in-roadway installation, necessitating physical changes in the roadway as part of the installation process. Over-roadway sensors often provide data not available from in-roadway sensors and some can monitor multiple lanes with one unit.

Table 1-1. Strengths and weaknesses of commercially available sensor technologies.

Technology	Strengths	Weaknesses
Inductive loop	• Flexible design to satisfy large variety of applications. • Mature, well understood technology. • Large experience base. • Provides basic traffic parameters (e.g., volume, presence, occupancy, speed, headway, and gap). • Insensitive to inclement weather such as rain, fog, and snow. • Provides best accuracy for count data as compared with other commonly used techniques. • Common standard for obtaining accurate occupancy measurements. • High frequency excitation models provide classification data.	• Installation requires pavement cut. • Improper installation decreases pavement life. • Installation and maintenance require lane closure. • Wire loops subject to stresses of traffic and temperature. • Multiple loops usually required to monitor a location. • Detection accuracy may decrease when design requires detection of a large variety of vehicle classes.
Magnetometer (two-axis fluxgate magnetometer)	• Less susceptible than loops to stresses of traffic. • Insensitive to inclement weather such as snow, rain, and fog. • Some models transmit data over wireless radio frequency (RF) link.	• Installation requires pavement cut. • Improper installation decreases pavement life. • Installation and maintenance require lane closure. • Models with small detection zones require multiple units for full lane detection.
Magnetic (induction or search coil magnetometer)	• Can be used where loops are not feasible (e.g., bridge decks). • Some models are installed under roadway without need for pavement cuts. However, boring under roadway is required. • Insensitive to inclement weather such as snow, rain, and fog. • Less susceptible than loops to stresses of traffic.	• Installation requires pavement cut or boring under roadway. • Cannot detect stopped vehicles unless special sensor layouts and signal processing software are used.
Microwave radar	• Typically insensitive to inclement weather at the relatively short ranges encountered in traffic management applications. • Direct measurement of speed. • Multiple lane operation available.	• Continuous wave (CW) Doppler sensors cannot detect stopped vehicles.

Table 1-1. Strengths and weaknesses of commercially available sensor technologies—*Continued*

Technology	Strengths	Weaknesses
Active infrared (laser radar)	• Transmits multiple beams for accurate measurement of vehicle position, speed, and class. • Multiple lane operation available.	• Operation may be affected by fog when visibility is less than ≈20 feet (ft) (6 m) or blowing snow is present. • Installation and maintenance, including periodic lens cleaning, require lane closure.
Passive infrared	• Multizone passive sensors measure speed.	• Passive sensor may have reduced vehicle sensitivity in heavy rain, snow and dense fog. • Some models not recommended for presence detection.
Ultrasonic	• Multiple lane operation available. • Capable of overheight vehicle detection. • Large Japanese experience base.	• Environmental conditions such as temperature change and extreme air turbulence can affect performance. Temperature compensation is built into some models. • Large pulse repetition periods may degrade occupancy measurement on freeways with vehicles traveling at moderate to high speeds.
Acoustic	• Passive detection. • Insensitive to precipitation. • Multiple lane operation available in some models.	• Cold temperatures may affect vehicle count accuracy. • Specific models are not recommended with slow-moving vehicles in stop-and-go traffic.
Video image processor	• Monitors multiple lanes and multiple detection zones/lane. • Easy to add and modify detection zones. • Rich array of data available. • Provides wide-area detection when information gathered at one camera location can be linked to another.	• Installation and maintenance, including periodic lens cleaning, require lane closure when camera is mounted over roadway (lane closure may not be required when camera is mounted at side of roadway) • Performance affected by inclement weather such as fog, rain, and snow; vehicle shadows; vehicle projection into adjacent lanes; occlusion; day-to-night transition; vehicle/road contrast; and water, salt grime, icicles, and cobwebs on camera lens. • Reliable nighttime signal actuation requires street lighting • Requires 30- to 50-ft (9- to 15-m) camera mounting height (in a side-mounting configuration) for optimum presence detection and speed measurement. • Some models susceptible to camera motion caused by strong winds or vibration of camera mounting structure. • Generally cost effective when many detection zones within the camera field of view or specialized data are required.

Table 1-2 lists traffic flow sensor technologies and their capabilities. Most measure count, presence, and occupancy. Some single detection zone sensors, such as the range-measuring ultrasonic sensor and some infrared sensors, do not measure speed. CW Doppler radar sensors do not detect stopped or slow moving vehicles. Magnetometer, magnetic, passive infrared and sound-based, acoustic sensor models do not classify vehicles.

Table 1-2. Traffic output data (typical), communications bandwidth, and cost of commercially available sensors.

Sensor technology	Count	Presence	Speed	Output data	Classification	Multiple lane, multiple detection zone data	Communication bandwidth	Sensor purchase cost[a] (each in 1999 U.S. $)
Inductive loop	✓	✓	✓[b]	✓	✓[c]		Low to moderate	Low[i] ($500–$800)
Magnetometer (two axis fluxgate)	✓	✓	✓[b]	✓			Low	Moderate[i] ($900–$6,300)
Magnetic induction coil	✓	✓[d]	✓[b]	✓			Low	Low to moderate[i] ($385–$2,000)
Microwave radar	✓	✓[e]	✓	✓[e]	✓[e]	✓[e]	Moderate	Low to moderate ($700–$2,000)
Active infrared	✓	✓	✓[f]	✓	✓	✓	Low to moderate	Moderate to high ($6,500–$3,300)
Passive infrared	✓	✓	✓[f]	✓			Low to moderate	Low to moderate ($700–$1,200)
Ultrasonic	✓	✓		✓			Low	Low to moderate (Pulse model: $600–$1,900)
Acoustic array	✓	✓	✓	✓		✓[g]	Low to moderate	Moderate ($3,100–$8,100)
Video image processor	✓	✓	✓	✓	✓	✓	Low to high[h]	Moderate to high ($5,000–$26,000)

[a] Installation, maintenance, and repair costs must also be included to arrive at the true cost of a sensor solution as discussed in the text.
[b] Speed can be measured by using two sensors a known distance apart or estimated from one sensor, the effective detection zone and vehicle lengths.
[c] With specialized electronics unit containing embedded firmware that classifies vehicles.
[d] With special sensor layouts and signal processing software.
[e] With microwave radar sensors that transmit the proper waveform and have appropriate signal processing.
[f] With multidetection zone passive or active mode infrared sensors.
[g] With models that contain appropriate beamforming and signal processing.
[h] Depends on whether higher-bandwidth raw data, lower-bandwidth processed data, or video imagery is transmitted to the TMC.
[i] Includes underground sensor and local detector or receiver electronics. Electronics options are available to receive multiple sensor, multiple lane data.

Direct hardware and software purchase costs are not the only costs associated with a sensor. Installation, maintenance, and repair should also be factored into the sensor selection decision. Installation costs include fully burdened costs for technicians to prepare the road surface or subsurface (for inductive loops or other surface or subsurface sensors), install the sensor and mounting structure (if one is required for over-roadway sensors), purchase and install conduit, close traffic lanes, divert traffic, provide safety measures where required, and verify proper functioning of the device after installation is complete. Environmental concerns may warrant providing for the removal of cutting water and debris from the site. Maintenance and repair issues are discussed in Chapter 6.

The technologies listed in Tables 1-1 and 1-2 are mature with respect to traffic management applications, although some may not provide the data required for a specific application. Some technologies, such as video image processing, microwave and laser radars, and inductive-loop detectors, continue to evolve by adding capabilities that measure additional traffic parameters, track vehicles, improve spatial resolution, or link data from one sensor to those from another.

MODERN VEHICLE SENSORS

The following discussion provides a broad overview of the operation of in-roadway and over-roadway traffic flow sensors most used today. These sensors include inductive-loop detectors, magnetometers, video image processors, microwave radar sensors, laser radar sensors, passive infrared sensors, ultrasonic sensors, and passive acoustic sensors. Typical applications include traffic signal control, freeway ramp metering, freeway mainline control, incident detection, and gathering of vehicle volume and classification data to meet State and Federal reporting requirements. These devices are either installed in, below, or above the roadway. Subsequent chapters of the *Handbook* describe the installation and operation of these sensors in more detail.

INDUCTIVE-LOOP DETECTORS

An inductive-loop detector senses the presence of a conductive metal object by inducing currents in the object, which reduce the loop inductance. Inductive-loop detectors are installed in the roadway surface. They consist of four parts: a wire loop of one or more turns of wire embedded in the roadway pavement, a lead-in wire running from the wire loop to a pull box, a lead-in cable connecting the lead-in wire at the pull box to the controller, and an electronics unit housed in the controller cabinet as shown in Figure 1-3. The electronics unit contains an oscillator and amplifiers that excite the embedded wire loop. The electronics unit also supports other functions such as selection of loop sensitivity and pulse or presence mode operation to detect vehicles that pass over the detection zone of the loop.

Inductive-loop detectors sense the presence of a conductive metal object by inducing electrical currents in the object. The induced current decreases the loop inductance, which is sensed by the inductive-loop electronics unit. The electronics unit interprets the decreased inductance as a vehicle detection and sends an appropriate call to the controller.

Figure 1-3 illustrates an inductive-loop detector system consisting of one or more turns of wire laid in a loop in the roadway, which is connected by lead-in wire to a pull box located at the side of the road. The pull box, in turn, is connected by lead-in cable to an electronics unit in a cabinet. The electronics unit energizes the wire loop, analyzes the signal, and transmits vehicle detection information to the controller.

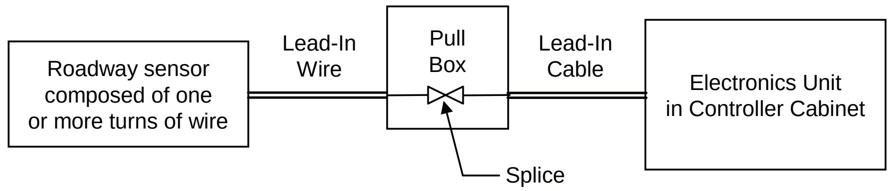


Figure 1-3. Inductive-loop detector system.

When a vehicle passes over the wire loop or is stopped within the area enclosed by the loop, it reduces the loop inductance, which unbalances the tuned circuit of which the loop is a part. The resulting increase in oscillator frequency is detected by the electronics unit and interpreted as a vehicle detection by the controller.

Conventional inductive loops are constructed by cutting a slot in the pavement and placing one or more turns of wire in the slot as indicated in Figure 1-4. The wire is then covered with sealant.

The size, shape, and configuration of the loop vary depending upon the specific application, ranging from the common 6- x 6-ft (1.8- x 1.8-m) loops, to long rectangular loops 6- x 40- to 70-ft (1.8- x 12- to 21-m) for actuated signal control. Because of the flexibility of its design, the inductive-loop detector is capable of detecting a broad range of vehicles.

An alternate, more durable construction is to place the turns of wire in a plastic conduit just below the pavement surface. Another option is to encase the wire in a plastic sleeve before installing the wire loop in the sawcut slot in the pavement. A wide variety of loop sizes and shapes are available to meet specific needs as described in Chapter 4.

Figure 1-4 depicts an inductive-loop detector installation, which contains a 3-turn 6- x 6-ft (1.8- x 1.8-m) square loop and connecting wire and cable. The loop is centered in a 12-ft (3.7 m) roadway, 3 ft (0.9 m) from either side.

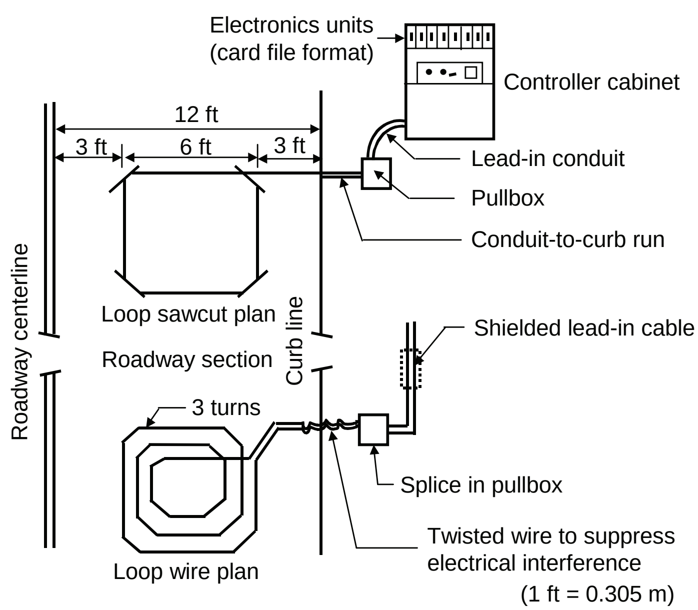


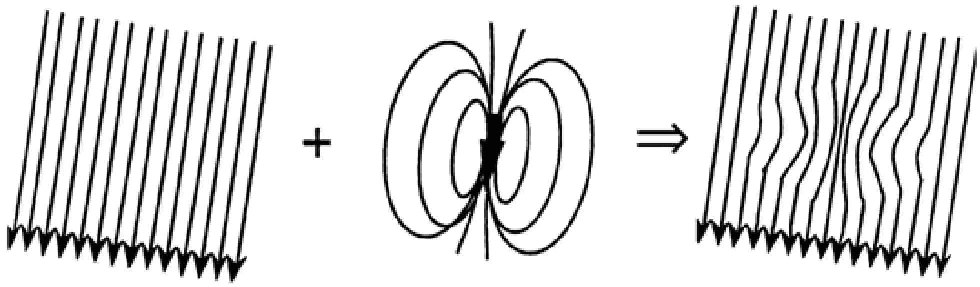
Figure 1-4. Inductive-loop installation example.

MAGNETIC SENSORS

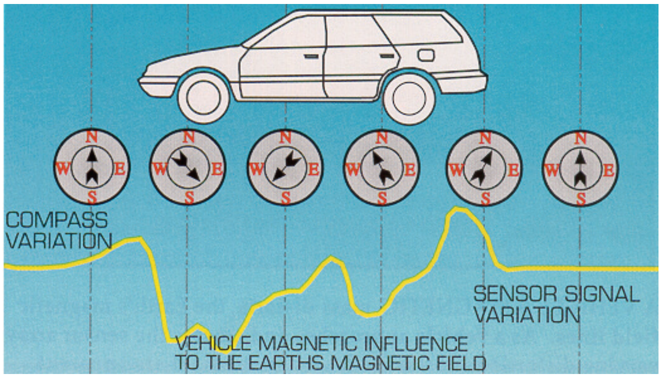
Magnetic sensors are passive devices that detect the presence of a ferrous metal object through the perturbation (known as a magnetic anomaly) they cause in the Earth's magnetic field. Figure 1-5 shows the magnetic anomaly created by the magnetic dipoles, i.e., energy fields, on a steel vehicle when it enters a magnetometer's detection zone. The upper part of Figure 1-5 indicates how the vector addition of the dipole magnetic field to the quiescent Earth's magnetic field produces the magnetic anomaly. The lower portion of the figure depicts several dipoles on a vehicle and their effect on compass readings and sensor output.

Magnetic sensors are passive devices that detect the presence of a ferrous metal object through the perturbation they cause in the Earth's magnetic field

Figure 1-5 shows that the magnetic dipole produced by the ferrous materials in vehicles causes an anomaly in the Earth's magnetic field, which can be measured by sensors as the vehicle passes over them.



(a) Magnetic anomaly induced in the Earth's magnetic field by a magnetic dipole.



(b) Perturbation of Earth's magnetic field by a ferrous metal vehicle (Source: Nu-Metrics, Vanderbilt, PA).

Figure 1-5. Magnetic anomaly in the Earth's magnetic field induced by magnetic dipoles in a ferrous metal vehicle.

Two types of magnetic field sensors are used for traffic flow parameter measurement. The first type, the two-axis fluxgate magnetometer, detects changes in the vertical and horizontal components of the Earth's magnetic field produced by a ferrous metal vehicle. The two-axis fluxgate magnetometer contains two primary windings and two secondary "sense" windings on a bobbin surrounding a high permeability soft magnetic material core. In response to the magnetic field anomaly, i.e., the magnetic signature of a vehicle, the magnetometer's electronics circuitry measures the output voltage generated by the secondary windings. The vehicle detection criterion is for the voltage to exceed a predetermined threshold. In the presence or stopped vehicle mode of operation, the detection output is maintained until the vehicle leaves the detection zone.[5]

The second type of magnetic field sensor is the magnetic detector, more properly referred to as an induction or search coil magnetometer. It detects the vehicle signature by measuring the distortion in the magnetic flux lines induced by the change in the Earth's magnetic field produced by a moving ferrous metal vehicle. These devices contain a single coil winding on a permeable magnetic material rod core. Similar to the fluxgate magnetometer, magnetic detectors generate a voltage when a moving ferromagnetic object perturbs the Earth's magnetic field. Induction magnetometers do not detect stopped vehicles since they require a vehicle to be moving or otherwise changing its signature characteristics with respect to time. However, multiple units of some magnetic detectors can be installed and utilized with specialized signal processing software to generate vehicle presence data.

Magnetic detectors are inserted horizontally below the roadway. Since they provide only passage data and not occupancy or presence data, their use is limited to special applications.

Another device similar to the magnetic detector is the microloop probe. As a vehicle passes over the microloop, the change in inductance is sensed by a conventional inductive-loop detector electronics unit. Some models are inserted into holes bored into the roadway surface. Other models are inserted into sleeves below the road surface using horizontal drilling from the side of the road. Often two or more microloop probes are connected in series or with conventional wire loops to detect a range of vehicle sizes and obtain required lane coverage. One microloop probe model can be connected in rows of three to generate signals that detect stopped vehicles. Application-specific software from its manufacturer is also needed to enable stopped vehicle detection.

VIDEO IMAGE PROCESSORS

Video image processors (VIPs) detect vehicles and provide traffic flow data across several lanes and in multiple areas in one lane. They require a line-of-sight view of the area they monitor and are subject to occlusion and the effects of inclement weather.

Video cameras were introduced to traffic management for roadway surveillance based on their ability to transmit closed-circuit television imagery to a human operator for interpretation. Present-day traffic mangers utilize video image processing to automatically analyze the scene of interest and extract information for traffic surveillance and management. A video image processor (VIP) system typically consists of one or more cameras, a microprocessor-based computer for digitizing and analyzing the imagery, and software for interpreting the images and converting them into traffic flow data. A VIP can replace several in-ground inductive loops, provide detection of vehicles across several lanes, and perhaps lower maintenance costs. Some VIP systems process data from more than one camera and further expand the area over which data are collected.

VIPs can classify vehicles by their length (usually three length classification ranges are available) and report vehicle presence, volume, lane occupancy, and speed for each class and lane. VIPs that track vehicles may also have the capability to register turning movements and lane changes. Vehicle density, link travel time, and origin-destination pairs are potential traffic parameters that can be obtained by analyzing data from a series of image processors installed along a section of roadway.[5] The types of information provided by VIPs makes suitable for arterial and freeway applications. An example of a camera mounted to transmit imagery to a VIP for an arterial traffic signal control application is shown in Figure 1-6.

VIP systems detect vehicles through the analysis of black and white or color imagery gathered by cameras at a section of roadway. Black and white image analysis is performed by algorithms that examine the variation of gray levels in groups of pixels (picture elements) contained in the video frames. Research has been conducted on algorithms that are sensitive to color features, for example those that assist in eliminating shadow artifacts or enhance vehicle discrimination in inclement weather. Along with vehicle size and class data, color fingerprints or signatures have been proposed to determine traffic volume, lane changes, turning movements, and link travel time by re-identifying a vehicle or group at a downstream site.[7,8]

Figure 1-6 shows a camera in a weatherproof housing mounted on a vertical pole that extends upright from a signal mast arm.

Figure 1-6. Overhead camera mounting on a mast arm as typically used to provide imagery to a VIP for arterial signal control (Photograph courtesy of Lawrence A. Klein).

Algorithms utilized in video image processing are designed to ignore gray level or color variations in the stationary image background. The algorithms are intended to also ignore variations caused by weather conditions, shadows, and daytime or nighttime artifacts, but retain objects identified as automobiles, trucks or buses, motorcycles, and bicycles. Traffic flow parameters are calculated by analyzing successive video frames.

MICROWAVE RADAR SENSORS

Microwave radar was developed for detecting objects in the period before and during World War II. Radar is defined as "a device for transmitting electromagnetic signals and receiving echoes from objects of interest (i.e., targets) within its volume of coverage."[9] Radar was originally an acronym for **RA**dio **D**etection **A**nd **R**anging.

The term microwave refers to the wavelength of the transmitted energy, usually between 0.4 inch and 11.8 inches (1 and 30 centimeters (cm)). This corresponds to a frequency range of 1 gigahertz (GHz (10^9 Hertz (Hz)) to 30 GHz 10^9 Hz. Microwave sensors designed for traffic data collection in U.S. roadside applications are limited by FCC regulations to operating frequency intervals near 10.5, 24.0, and 34.0 GHz. The sensor manufacturers satisfy these requirements, as well as others that restrict the transmitted power and bandwidth. Thus, the end users are not required to possess special licenses or test equipment to verify the output frequency or power of the devices. Radars at frequencies above 30 GHz operate in the millimeter-wave spectrum since the wavelength of the transmitted energy is expressed in terms of millimeters (mm). Most commercially available microwave radar sensors utilized in roadside applications transmit electromagnetic energy at the X-band frequency of 10.525 GHz. Higher frequencies illuminate smaller ground areas with a given size antenna and thus are capable of greater spatial resolution. FCC-approved frequencies for vehicle-mounted radars utilized in collision avoidance, obstacle detection, and automatic cruise control are 47.5 to 47.8 GHz and 76 to 77 GHz.

Figure 1-7 shows the transmission of energy by an overhead-mounted microwave radar toward an area of roadway. The beamwidth or area in which the radar energy is concentrated is controlled by the size and the distribution of energy across the aperture of the antenna. The sensor manufacturer usually establishes these design constraints. When a vehicle passes through the antenna beam, a portion of the transmitted energy is reflected back towards the antenna. The energy then enters a receiver where the detection is made and traffic flow data, such as volume, speed, and vehicle length, are calculated.

Figure 1-7 shows that a microwave radar mounted on overhead bridges, poles or mast arms transmits signals that are reflected off vehicles back to the radar sensor. The reflected energy is analyzed to produce traffic flow data.

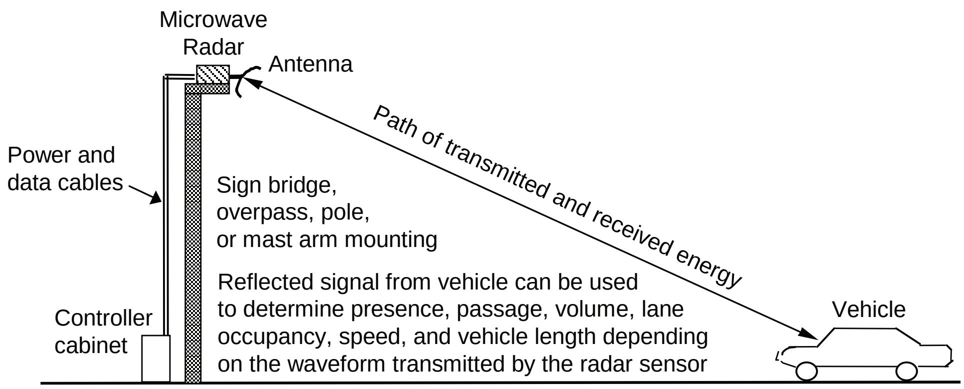


Figure 1-7. Microwave radar operation.

The radar sensor may be mounted over the middle of a lane to measure approaching or departing traffic flow parameters in a single lane, or at the side of a roadway to measure traffic parameters across several lanes as shown in Figure 1-8. Forward-looking wide beamwidth radars gather data representative of traffic flow in one direction over multiple lanes. Forward-looking narrow beamwidth radars monitor a single lane of traffic flowing in one direction. Side-mounted, multiple detection zone radars project their detection area (i.e., footprint) perpendicular to the traffic flow direction. These sensors provide data corresponding to several lanes of traffic, but generally not as accurately as can the same radar mounted in the forward-looking direction. Side-mounted, single detection zone radars are typically used to detect vehicle presence in one or more lanes at signalized intersections.[5]

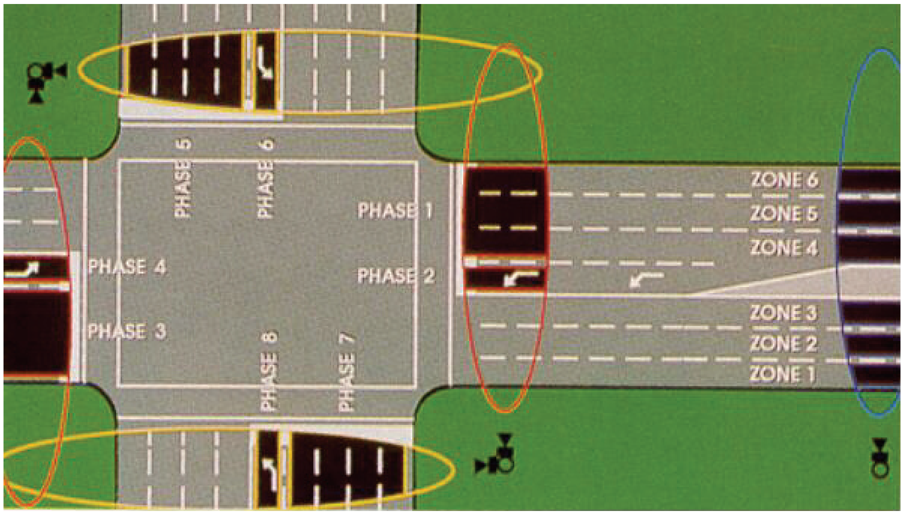


Figure 1-8. Mounting of presence-detecting microwave radar sensors for multilane vehicle detection and signal actuation at an intersection (Picture courtesy of EIS, Toronto, Canada).

Figure 1-8 shows that presence-detecting microwave radar sensors can detect traffic flow in multiple lanes when mounted on poles adjacent to pedestrian crosswalks or roadway shoulders.

Microwave sensors that transmit a continuous wave (CW) Doppler waveform detect vehicle passage and provide measurements of vehicle count and speed. They cannot detect stopped vehicles. Microwave sensors that transmit a frequency modulated continuous wave (FMCW) detect vehicle presence as well as vehicle passage. They can detect stopped vehicles and provide measurements of lane occupancy, vehicle count, speed, and vehicle length grouped into several length bins.

Microwave radar sensors that transmit a frequency modulated waveform detect stopped and moving vehicles. Microwave radar sensors that transmit a continuous wave signal constant in frequency cannot detect stopped vehicles.

INFRARED SENSORS

Active and *passive* infrared sensors are manufactured for traffic flow monitoring applications. Active infrared sensors illuminate detection zones with low power infrared energy transmitted by laser diodes operating in the near infrared region of the electromagnetic spectrum at 0.85 mm. A portion of the transmitted energy is reflected or scattered by vehicles back towards the sensor. Although light-emitting diodes may also be utilized as the energy source in an active IR sensor, there are currently no commercial models marketed in the U.S. that exploit this design. A prototype sensor system using modulated light emitting diodes was designed to measure the speed and height of high and long trucks entering a curved freeway-to-freeway interchange. The diodes operated in the near infrared spectrum at 880 nanometers (nm). The signal modulation prevented interference from other sources of infrared energy, including sunlight. Two transmitter-receiver systems measured the vehicle speed and one measured the vehicle height. When trucks susceptible to rollover or jackknifing were encountered, flashers were activated to warn drivers to reduce speed.[5,10]

Passive infrared sensors transmit no energy of their own. Rather they detect energy emitted from roadways and vehicles or energy that is reflected from them.

Passive sensors transmit no energy of their own. Rather they detect energy from two sources:

- Energy emitted from vehicles, road surfaces, and other objects in their field-of-view.
- Energy emitted by the atmosphere and reflected by vehicles, road surfaces, or other objects into the sensor aperture.

The energy captured by active and passive infrared sensors is focused by an optical system onto an infrared-sensitive material mounted at the focal plane of the optics. This material converts the reflected and emitted energy into electrical signals. Real-time signal processing is used to analyze the signals for the presence of a vehicle. The sensors are mounted overhead to view approaching or departing traffic. They can also be mounted in a side-looking configuration. Infrared sensors are utilized for signal control; volume, speed, and class measurement; detection of pedestrians in crosswalks; and transmission of traffic information to motorists.

A laser radar is an active infrared sensor, since it transmits energy in the infrared spectrum.

LASER RADAR SENSORS

Laser radars are active sensors in that they transmit energy in the near infrared spectrum. Models are available that scan infrared beams over one or two lanes or use multiple laser diode sources to emit a number of fixed beams that cover the desired lane width. An example of a laser radar beam-scanning configuration is shown in Figure 1-9. Laser radars provide vehicle presence at traffic signals, volume, speed, length assessment, queue measurement, and classification. Multiple units can be installed at the same intersection without interference from transmitted or received signals. Modern laser sensors produce two- and three-dimensional imagery of vehicles suitable for vehicle classification as illustrated in Figure 1-10. Their ability to classify 11 types of vehicles has found application on toll roads.

Figure 1-9 depicts a scanning laser radar mounted over a lane. This mounting configuration optimizes the sensor's view of vehicular traffic flow.

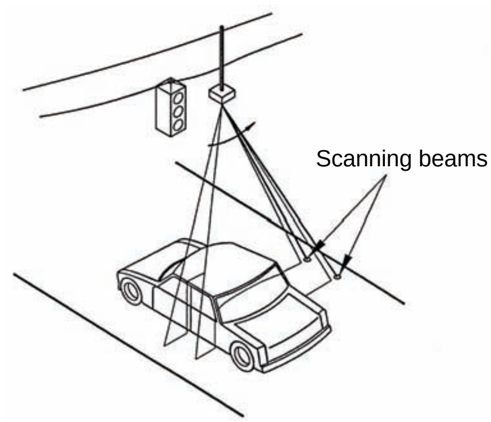


Figure 1-9. Scanning infrared laser radar two-beam pattern across a traffic lane (Drawing courtesy of Schwartz Electro-Optics, now OSI Laserscan, Orlando, FL).

Figure 1-10 illustrates a three-dimensional image of a van pulling a boat as produced by a scanning laser radar.

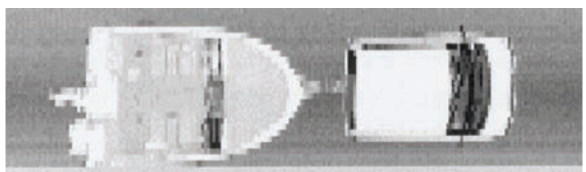


Figure 1-10. 3-D laser radar range image of a van pulling a boat (Photograph courtesy of Schwartz Electro-Optics, now OSI Laserscan, Orlando, FL).

OTHER SENSOR TECHNOLOGIES

In addition to the technologies discussed above, others find application to traffic management. These include ultrasonic sensors, passive acoustic sensors, and devices that use a combination of sensor technologies. These devices are described further in Chapter 2. Chapter 2 also discusses the operation of passive infrared sensors in more detail.

Although not discussed in this chapter, other sensor technologies find application in traffic management. These include ultrasonic sensors, passive acoustic sensors, and devices that use a combination of sensor technologies.

DEFINITION OF TERMS

Traffic sensor literature often uses different terms to describe the same traffic flow sensors or their characteristics. To reduce potential confusion, the terms utilized in this *Handbook* are defined below. Words or phrases having the same meaning are listed along with the primary word being defined. Additional terms are defined in the glossary that appears in Appendix P.

Crosstalk: The adverse interaction of any channel of a sensor or sensor electronics unit with any other channel in that or another device. Crosstalk can occur via mutual coupling of magnetic fields in nearby inductive loops. The mutual coupling causes an interaction between two or more electronics units in the same cabinet when the units operate at the same or nearby frequencies. Crosstalk results in a sensor output actuation in the absence of a vehicle.

Detector Electronics Unit (Electronics Unit, Sensor Electronics Unit, Amplifier): An electronic device that energizes an inductive-loop detector, monitors loop inductance, and responds to a predetermined decrease in inductance with outputs that indicate the passage or presence of vehicles in the detection zones (NEMA).

An inductive-loop detector electronics unit is sometimes called an amplifier or detector, although it performs other functions as well, e.g., sensitivity adjustment, failure indication, and delayed actuation of controlled signals. Electronics units are also used with magnetic detectors and magnetometers. The electronics unit is typically located in a controller cabinet.

Detection Zone (Area of Detection, Detection Area, Zone of Detection, Effective Loop Area, Field of Influence, Field of View, Sensing Zone, Footprint): The area of the roadway within which a vehicle is detected by a sensor system.

Inductive-loop Detector (Loop Detector System): A sensor capable of detecting vehicle passage and presence. It consists of four parts, namely one or more turns of wire embedded in the pavement, a lead-in wire running from the wire loop in the pavement to the pull box, a lead-in cable spliced to the lead-in wire at the pull box, which connects to the controller, and an electronics unit housed in the controller.

Fluxgate Magnetometer: Two-axis fluxgate magnetometers are sensors that detect changes in the vertical and horizontal components of the Earth's magnetic field produced by a ferrous metal vehicle. They detect moving and stopped vehicles and thus provide passage and presence information.

Large Area Sensor (Area Sensor): For inductive-loop detectors utilized for traffic signal actuation, it is an inductive loop or combination of inductive loops connected in series, parallel, or series and parallel covering an area in the approach to an intersection. Detection area varies from 6 x 40 ft (1.8 x 12 m) to 6 x 100 ft (1.8 x 30 m) or larger. One of the more common configurations is four 6- x 6-ft (1.8- x 1.8-m) loops spaced 9 or 10 ft (2.75 or 3 m) apart for a length of 51 or 54 ft (15.5 or 16.5 m).

Lead-In Cable (Feeder Cable, Home-Run Cable, Transmission Line): The electrical cable that is spliced to the lead-in wire in the pull box and connects to the input of the inductive-loop detector electronics unit.

Lead-In Wire: That portion of an inductive-loop wire between the physical edge of the loop and the pull box. For a magnetic detector and magnetometer, it is the wire that runs from the sensor (probe) to the pull box.

Magnetic Detector (Induction or Search Coil Magnetometer): A passive device that detects changes in the Earth's magnetic field caused by the movement of a ferrous-metal vehicle in or near its detection area. It is placed under or in the roadway to detect the passage of a vehicle over the sensor. These sensors generally detect only moving vehicles. Their output is connected to an electronics unit.

Magnetic Sensor: Passive devices that detect the presence of a ferrous metal object through the perturbation (known as a magnetic anomaly) it causes in the Earth's magnetic field. Its output is connected to an electronics unit. The two types of magnetic sensors are fluxgate magnetometers and induction magnetometers, also referred to as magnetic detectors in this *Handbook*.

Passage Sensor (Motion Detector, Motion Sensor, Dynamic Detector, Movement Detector): A traffic flow sensor that detects the passage of a vehicle moving through the detection zone and ignores the presence of a vehicle stopped within the detection zone.

Presence Sensor: A traffic flow sensor that detects the presence of a vehicle within its detection zone and holds the call for a specified minimum time.

Pull Box: (Hand Hole, Junction Box, Junction Well, Splice Box): A container usually at least 1 cubic foot (e.g., approximately 1 ft^3 (0.028 m^3)) in size that is placed underground with a removable cover flush with the ground surface. Splices between lead-in cable and loop lead-in wire are located here.

Sensor: A device for indicating the presence or passage of vehicles or pedestrians. This general term is usually supplemented with a modifier indicating type (e.g., inductive-loop detector, magnetic detector, video image processor, microwave sensor, and infrared sensor); operation (e.g., point sensor, large area sensor, and presence sensor); or function (e.g., calling sensor, extension sensor, and classification sensor).

Sensor Amplifier: A device capable of intensifying the electrical energy or signal produced by a sensor. An example is a magnetic detector amplifier. An inductive-loop detector electronics unit is sometimes called an amplifier, although it performs other functions in addition to signal amplification.

Small Area Sensor (Point Sensor): A sensor that detects vehicles at a spot location, i.e., a small area usually not exceeding 6 x 6 ft (1.8 x 1.8 m).

Splashover: An unwanted actuation caused by a vehicle in a lane adjacent to that in which the sensor is located.

ORGANIZATION OF HANDBOOK

The *Traffic Detector Handbook* is structured to parallel the progression of decisions, activities, and functions related to the design, installation, and maintenance of sensor systems. Chapter 1 addresses the need for sensors as an integral part of modern traffic control and management systems and discussed the operation of several of the sensor technologies currently exploited for traffic management.

Chapter 2 describes the theory of operation of the inductive-loop detector, magnetometer, video image processor, microwave radar sensor, laser radar sensor, passive infrared sensor, ultrasonic sensor, passive acoustic sensor, and other sensors that utilize a combination of technologies. It addresses the needs of traffic and electrical engineers who have responsibility for selecting or specifying sensors that meet specific operational requirements. It also describes the NEMA Standards and the Type 170 and 2070 Controller Specifications.

Chapter 3 provides an overview of traffic control and management applications that rely on vehicle detection and monitoring of environmental conditions. It identifies how sensors support these applications.

The design and operating characteristics of in-roadway sensors such as inductive loops, magnetometers, and magnetic detectors for surface street and freeway traffic management are discussed in Chapter 4. This topic should be of particular interest to traffic engineers who develop plans and specifications for local intersections, traffic signal systems, and freeway surveillance and control systems.

Chapter 5 illustrates the installation procedures and best current practices for in-roadway and above-roadway sensors. The information is directed toward project engineers, contractors, inspectors, field crew supervisors, and traffic technicians.

Chapter 6 describes the broad spectrum of maintenance activities associated with in-roadway and over-roadway sensors. It provides management and supervising engineers with information needed to identify and resolve maintenance-related issues. The chapter contains detailed guidelines for maintenance supervisors and technicians to assist in identifying sensor failure mechanisms and corresponding corrective actions.

The appendices provide additional detailed information concerning several of the topics discussed in the main body of the *Handbook*.

REFERENCES

1. *Transportation Statistics Annual Report 1999*, BTS99-03. U.S. Department of Transportation, Bureau of Transportation Statistics, Washington, DC. 1999.
2. Davis, S.C., and S.W. Diegel. *Transportation Energy Data Book*: *Edition 24-2004*, Table 8.1, ORNL-6973. Oak Ridge National Laboratory, Oak Ridge, TN. December 2004.

3. Sweedler, B.M. "Toward a Safer Future," *TR News*. Transportation Research Board, National Research Council, no. 201, 3-6. Washington, DC. March–April 1999.
4. *Manual on Uniform Traffic Control Devices*. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 2003. http://mutcd.fhwa.dot.gov/pdfs/2003/pdf-index.htm (Accessed Nov. 26, 2004.)
5. Klein, L.A. *Sensor Technologies and Data Requirements for ITS*. Artech House, Norwood, MA. 2001.
6. McCall, W., and W. Vodrazka. *States' Successful Practices Weigh-In-Motion Handbook*. Iowa State University, Center for Transportation Research and Education, Ames, IA. December 1997. http://www.ctre.iastate.edu/research/wim_pdf/index.htm (Accessed Nov. 26, 2004.)
7. Zeng, N., and J.D. Crisman. "Evaluation of Color Categorization for Representing Vehicle Colors," *Proc. SPIE*, vol. 2902. Bellingham, WA. 1996. pp. 148–155.
8. Chachich, A., A. Pau, A. Barber, K. Kennedy, E. Olejniczak, J. Hackney, Q. Sun, and E. Mireles. "Traffic Sensor Using a Color Vision Method," *Proc. SPIE*, vol. 2902. Bellingham, WA. 1996. pp. 156–165.
9. Kurpis, G.P., and C.J. Booth, eds. *The New IEEE Standard Dictionary of Electrical and Electronics Terms*, 5th ed. Institute of Electrical and Electronics Engineers, New York, NY. 1993.
10. Lee, C.E., D.W. Borchardt, and Q. Fei. *Truck-Monitoring and Warning Systems for Freeway to Freeway Connections*, report number 2915-1. University of Texas, Texas Transportation Institute and Center for Transportation Research, Austin, TX. October 1999.

TABLE OF CONTENTS: CHAPTER 2.

LIST OF FIGURES: CHAPTER 2.

LIST OF TABLES: CHAPTER 2.

CHAPTER 2. SENSOR TECHNOLOGY

The operation of in-roadway and over-roadway sensors is described in this chapter. The technologies represented include inductive-loop detectors, magnetometers, video image processors, microwave radar sensors (presence detecting and Doppler), laser radar sensors, passive infrared sensors, ultrasonic sensors, passive acoustic sensors, and devices that utilize a combination of technologies. The information is intended to provide the practicing traffic engineer and electrical engineer with the knowledge needed to select the proper sensor technology for specific applications.

INDUCTIVE-LOOP DETECTORS

Since its introduction in the early 1960s, the inductive-loop detector has become the most utilized sensor in a traffic management system. The principal components of an inductive-loop detector system include:

- One or more turns of insulated loop wire wound in a shallow slot sawed in the pavement.
- Lead-in cable from the curbside pull box to the intersection controller cabinet.
- Electronics unit housed in a nearby controller cabinet.

Figure 2-1 displays a notional diagram of an inductive-loop detector system and the vehicle and steel reinforcement elements in the roadway with which it reacts.

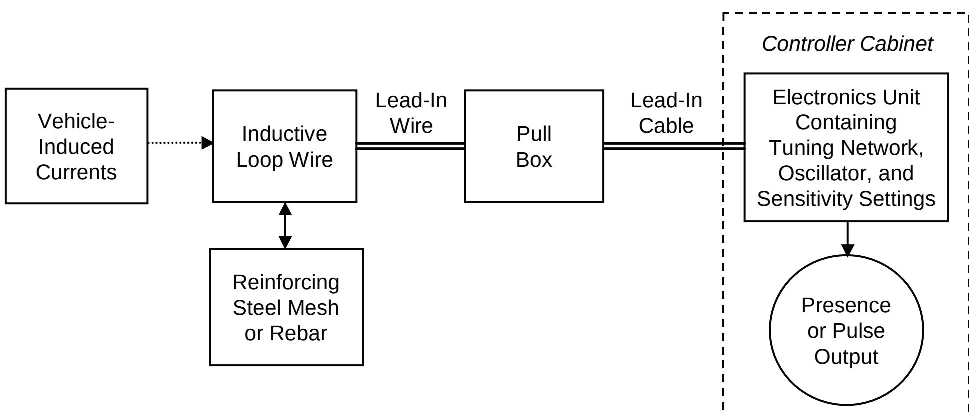


Figure 2-1. Inductive-loop detector system (notional).

The electronics unit transmits energy into the wire loops at frequencies between 10 kHz to 200 kHz, depending on the model. The inductive-loop system behaves as a tuned electrical circuit in which the loop wire and lead-in cable are the inductive elements. When a vehicle passes over the loop or is stopped within the loop, the vehicle induces eddy currents in the wire loops, which decrease their inductance. The decreased inductance actuates the electronics unit output relay or solid-state optically isolated output, which sends a pulse to the controller signifying the passage or presence of a vehicle.

Vehicles passing over or stopped within the detection area of an inductive-loop detector decrease the inductance of the loop. The electronics unit senses this event as a decrease in frequency and sends a pulse to the controller signifying the passage or presence of a vehicle.

The following sections describe inductive-loop system theory, loop characteristics, and the electronics unit.

THEORY OF OPERATION

The principles of operation of the inductive-loop detector system discussed below are common to all of the inductive-loop system designs described in Chapter 4. The loop wire and lead-in cable contain a combination of resistance, inductance, and capacitance (both wire-to-wire and wire-to-Earth capacitance).

LOOP WIRE AND CABLE RESISTANCE

Inductive-loop wire, lead-in wires, and lead-in cables typically use #12, #14, or #16 American Wire Gauge (AWG) wire with the low frequency or direct current resistance measured in units of ohms (Ω). The wire resistance is inversely proportional to the square of the wire diameter and increases as the wire diameter decreases. A volt-ohmmeter (VOM) measures direct current resistance. The wire resistance to alternating current flow increases as the frequency increases because the conducting area of the wire decreases due to the nonuniform flux inside the wire. The high frequency resistance cannot be measured with a VOM, but can be obtained from a measurement of quality factor as defined later in this chapter.

The loop in the roadway also contains an induced resistance (called the ground resistance) caused by transformer coupling between the loop and induced currents flowing in the roadway and subgrade materials. Appendix A provides a detailed derivation of ground resistance. Table 2-1 contains DC or low frequency resistance values for commercially available loop wire and lead-in cables.

Table 2-1. Resistance of cables commonly found in inductive-loop detector systems.

Manufacturer's wire or cable type	Function	Wire gauge (AWG)	DC resistance (Ω/ft)
9438	Loop wire	14	0.0025
8718	Lead-in cable	12	0.0019
8720	Lead-in cable	14	0.0029
8719	Lead-in cable	16	0.0045

LOOP INDUCTANCE

All wire conductors carrying an electrical current produce magnetic flux lines, which encircle the current flow that forms them. The magnetic flux induces the electrical property called inductance, measured in henrys (H). The inductance of the wire is called self-inductance. If the flux from current flowing in one wire couples to other wires, the resulting inductance is called mutual inductance.

Figure 2-2 displays the flux around a single turn wire loop. The plane containing the flux is normal to the current flow in the wire, where the flux direction is determined by the right hand rule. This rule is applied as follows: Place the right hand under the wire with fingers curled in the direction of the flux lines. The thumb points in the direction of the current flow. All flux lines are in the same direction inside the loop.

Figure 2-3 illustrates the magnetic flux lines for a solenoid or coil whose length is greater than the diameter. The magnetic flux is uniform inside the coil except near the ends. The magnetic field for this coil geometry is given by

$$H = \frac{NI}{\ell} \tag{2-1}$$

where

H = Magnetic field, ampere turns per meter, not to be confused with the units of inductance in henrys

N = Number of turns

I = Coil current, amperes

ℓ = Length of coil, meters.

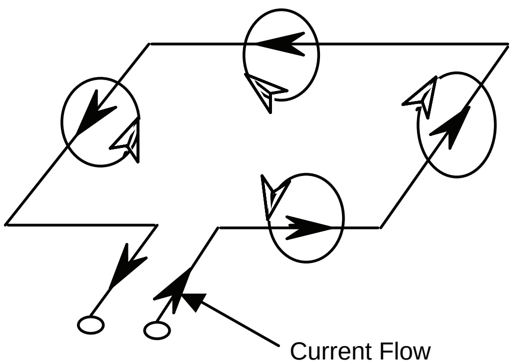


Figure 2-2. Magnetic flux around loop. The black arrows represent the current flow in the wire and the white arrows the induced flux as determined by the right hand rule.

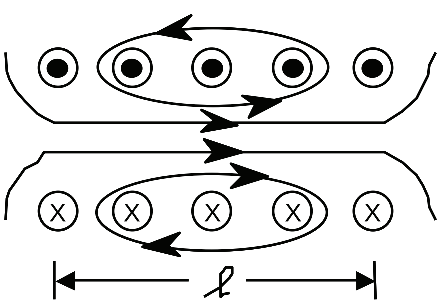


Figure 2-3. Magnetic flux for solenoid (coil). The black arrows represent the current flow, while the circles with the black and "X" centers represent the induced flux flow out of and into the plane of the figure, respectively.

Since the magnetic flux is uniform inside the coil, the flux is given by

$$\phi = BA \tag{2-2}$$

where

ϕ = Magnetic flux, webers

B = Magnetic flux density, webers per m^2

A = Cross sectional area of coil, m^2.

The magnetic flux density is expressed as

$$\phi = BA \tag{2-3}$$

where

μ_r = Relative permeability of material (1 for air)

μ_0 = $4\pi \times 10^{-7}$ henrys per meter.

The inductance of a coil is defined as

$$L = \frac{N\phi}{I} = \frac{NBA}{I} \tag{2-4}$$

where

L = Inductance, henrys

N = Number of turns

I = Coil current, amperes.

The inductance of a coil, with a length much greater than the coil area to ensure uniform magnetic flux inside the coil, is given by

$$L = \frac{N\mu_r\mu_0 HA}{I} = \frac{\mu_r\mu_0 N^2 A}{\ell}. \tag{2-5}$$

A roadway inductive loop has a nonuniform flux field that produces an inductance value given by Equation 2-6.

This equation shows that coil inductance is directly proportional to the turns squared and the coil area, and inversely proportional to coil length. Although the inductance formula as written is not directly applicable to a roadway inductive loop, the formula can be modified by a factor F' to account for the nonuniform flux in the roadway inductive loop. Thus

$$L = \frac{\mu_r\mu_0 N^2 AF'}{\ell}. \tag{2-6}$$

Equation 2-6 is applied to a loop inductance calculation in Appendix B. In this case, ℓ is referred to as the "length of the current sheet." Equation 2-6 shows that iron, with a relative permeability greater than one, will increase the loop inductance. Although the greatest increase in inductance occurs when an iron core passes directly through the loop, the iron mass of a vehicle engine, transmission, or differential will slightly increase the loop inductance. This condition is called the "ferromagnetic effect."

FERROMAGNETIC EFFECT AND VEHICLE DETECTION

The ferromagnetic effect increases loop inductance. However, vehicle-induced eddy currents decrease loop inductance even more. Therefore, net effect is decreased loop inductance when a vehicle passes through the detection zone of an inductive loop.

However, the ferromagnetic effect produced by the iron mass of the engine, transmission, or differential does not create a presence or passage indication by the controller. When the heavy ferrous engine enters the inductive loop's detection area, it increases the inductance of the wire loop. This effect occurs because the insertion of any iron core into the field of any inductor reduces the reluctance (i.e., a term that corresponds to the resistance of a magnetic circuit) of the flux path and, therefore, increases the net inductance. However, the peripheral metal of the vehicle has an opposite effect on the inductance due to eddy currents that are produced. The decrease in inductance from the eddy currents more than offsets the increase from the ferrous mass of the engine, and the net effect is an overall reduction in the inductance of the wire loop.

LOOP CHARACTERISTICS

The inductive-loop detector provides a wide range of geometries to the traffic engineer for satisfying diverse traffic signal control applications, as discussed in Chapter 4. The size and the number of turns of a loop or combination of loops, together with the length of the lead-in cable, must produce an inductance value that is compatible with the tuning range of the electronics unit and with other requirements established by the traffic engineer. NEMA standards for inductive-loop detectors (see Appendix J) specify that an electronics unit must be capable of operating satisfactorily over an inductance range of 50 to 700 microhenrys (μH). Some units tolerate much larger inductance values, for example, from several loops wired in series. While larger inductance values are technically feasible, NEMA has specified a conservative upper limit to promote practices compatible with all existing electronics units.

LOOP CAPACITANCE

Figure 2-4 depicts the dominant capacitive coupling phenomena that exist between (1) loop wires themselves and (2) loop wires and the sidewalls of the sawcut slot. The capacitance related to the sawcut slot is directly proportional to the dielectric constant of the slot sealing material. Figure 2-5 is an equivalent electrical circuit representing the inductive-loop wire resistance R_S, inductance L_S, and capacitance C_P that are created when a loop is installed in roadway pavement.

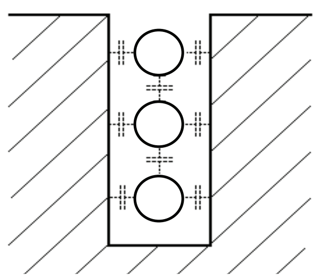


Figure 2-4. Capacitive coupling between the loop wires themselves and the sawcut slot sidewalls.

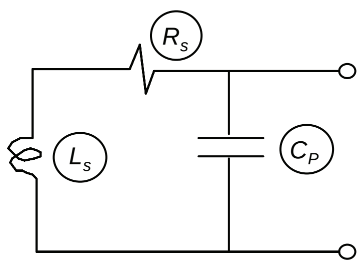


Figure 2-5. Equivalent electrical circuit for an inductive loop with capacitive coupling to the sidewalls of a sawcut slot.

The measurement data in Figure 2-6 show the effect capacitance C_P has on increasing the inductance at the loop terminals as operating frequency increases.(1) If the slot sealing material is hygroscopic (i.e., readily absorbs and retains water) or incomplete (i.e., does not fill the slot or encapsulate the wires, allowing water to enter the slot and penetrate between the loop wire turns), the variation in capacitance, and hence inductance, will be large because of the large dielectric constant of water.

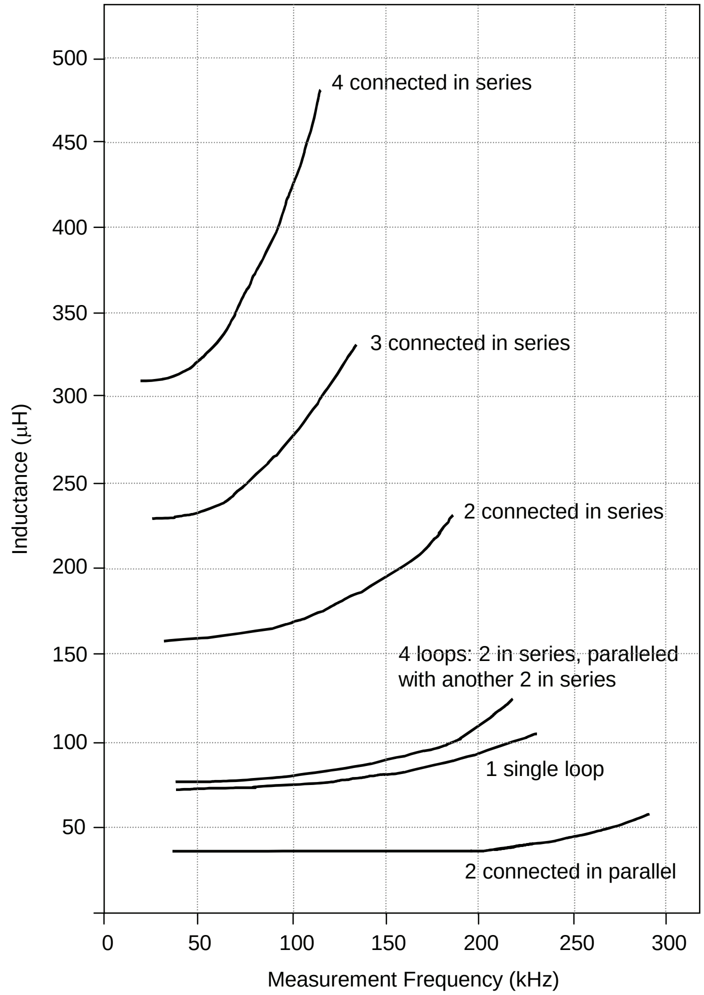


Figure 2-6. Average values of loop inductance vs. measuring frequency for series, parallel, and series-parallel connections of 6- x 6-ft (1.8- x 1.8-m) inductive loops. The plots represent curve fits to the measured data.

Loop system capacitance must be minimized for proper operation at frequencies of 10 kilohertz (kHz) and greater.

The capacitance change due to water can, therefore, result in unstable inductive-loop detector operation. At frequencies of 1 kilohertz (kHz), the capacitance effect is insignificant. At frequencies of 10 kHz or greater, the capacitance effect is important. When loop inductance is measured at 20 kHz or greater, the measurement frequency must be specified since the measured inductance is frequency dependent. A large number of turns on large area loops further increases the loop capacitance and lowers the self-resonant frequency of the loop (i.e., no loop inductance is measured at the loop terminals when the loop is self resonant).

Figure 2-6 also illustrates how different series, parallel, and series-parallel configurations of wire loops affect the resultant loop inductance and its rate of change with frequency. The effect of the connection method on system inductance is discussed further under "Loop System Inductance Calculations," later in this chapter.

LOOP QUALITY FACTOR *Q*

Loop quality factor Q is a measure of the losses in an inductive-loop detector system.

The resonant efficiency of a circuit is expressed through the dimensionless quality factor Q. If the losses of the inductor are large, Q is low. A perfect inductor has no losses; therefore, there is no dissipation of energy within the inductor and Q is infinite.

Total energy loss in a lossy inductor is calculated by modeling the inductor as an equivalent lossless inductor in series with a resistor. The quality factor is equal to the ratio of the inductive reactance to the resistive loss of the inductor. Since inductive reactance is a frequency-dependent quantity, the frequency must be specified when measuring quality factor. The formula for Q is written as

$$Q = \frac{2\pi f L_S}{R_S} = \frac{\omega L_S}{R_S} \tag{2-7}$$

where

- Q = Quality factor
- π = 3.14159 (a constant)
- f = Inductive-loop system excitation frequency, Hz
- L_S = Loop series inductance, henrys
- R_S = Loop series resistance, ohms
- ω = Radian frequency = $2\pi f$.

The resonant frequency ω_0 of an inductive-loop-equivalent electrical circuit represented by Figure 2-5 is given by

$$\omega_0 = \frac{1}{\sqrt{L_S C_P\left(1+\frac{1}{(Q_0)^2}\right)}} . \tag{2-8}$$

From Equation 2-7,

$$\omega_0 = \frac{Q_0 R_S}{L_S} . \tag{2-9}$$

Therefore, the equation for the loop quality factor Q_0 of the resonant circuit becomes

$$Q_0 = \sqrt{\frac{1}{(R_S)^2}\frac{L_S}{C_P} - 1} . \tag{2-10}$$

The electronics unit adds a load resistance R_L in parallel with the capacitor C_P shown in the inductive-loop-equivalent electrical circuit of Figure 2-5. The effect of R_L is to reduce the quality factor. The resulting quality factor is

$$Q_P = \omega_0 C_P R_L \tag{2-11}$$

or

$$Q_P = \omega_0 C_P R'_P \tag{2-12}$$

where R'_P is the transformed series resistance in parallel with R_L.

The loaded quality factor Q_L of the circuit in Figure 2-5 with a load resistance R_L in parallel with the capacitor C_P is

$$Q_L = \frac{Q_P Q_0}{Q_P + Q_0}. \tag{2-13}$$

Quality factors of 5 and above are recommended when installing inductive-loop detectors as oscillators in most electronics units will not operate with low Q. Moisture in the pavement and subgrade can increase the loop ground resistance such that the Q of the inductive-loop system decreases below 5, thereby reducing the sensitivity of most inductive-loop electronics units. Loop capacitance will also reduce Q.

Losses caused by the high frequency loop excitation and ground currents in the pavement near the wire further reduce the quality factor. As a result, the Q of an identical wire configuration will vary from location to location.

The loaded quality factor Q_L given by Equation 2-13 applies to low loss applications, where the quality factor is large and f, L_S, and R_S can be readily measured. Inductive-loop detectors used in roadways, on the other hand, are not as adaptable to the above analysis because the inductance is distributed over the loop and lead-in cable and is difficult to measure. Calculation of the quality factor for roadway loops is further complicated by the larger actual resistances of the loop wire and lead-in cable as compared to the series value measured with an Ohm-meter. The extra losses are due to the high frequency excitation and ground currents in the pavement associated with the loop configuration and the roadway environment near the wire. As a result, the Q of an identical wire configuration will vary from location to location.

Figure 2-7 illustrates an inductive-loop system quality factor calculation using Q_0 and Q_P. Tables 2-2 through 2-4 list calculated quality factors for rectangular, quadrupole, and circular inductive loops, respectively, of 1, 2, 3, 4, and 5 turns. Loops are excited at 20 kHz in these tables, with conductor and/or quadrupole lateral spacing of 200 mils. All inductance and quality factors are apparent values (i.e., loop capacitance and resistance are included).

LOOP SYSTEM QUALITY FACTOR *Q* CALCULATION

Assumptions:

Loop Type: 3-turn, 6 x 6 ft (1.8 x 1.8 m) of #14 AWG wire

Loop Inductance: 74 μH at 20 kHz from Appendix C

Loop Resistance (in air): 0.0025 Ω/ft (0.0083 Ω/m) from Appendix D

Lead-in Cable Type: 100 ft (30 m) of Belden 8718

Lead-in Cable Inductance: 0.20 μH/ft (0.67 μH/m) from Appendix D

Lead-in Cable Resistance: 0.0031 Ω/ft (0.0103 Ω/m) from Appendix D

Operating Frequency: 20 kHz

Total Loop System Series Inductance: 74 μH + 20 μH = 94 μH

Total Loop System Series Resistance: 0.25 Ω + 0.62 Ω = 0.87 Ω

Note: Wire length for resistance calculation is per wire (i.e., twice the cable length)

Total Inductive-Loop System Capacitance:

$$C_P = \frac{1}{\omega_0^2 L_S} = \frac{1}{(2\pi \times 20 \times 10^3)^2 (94 \times 10^{-6})}$$

$$= 6.74 \times 10^{-7} \text{F}$$

Quality Factor of Inductive-Loop System:

$$Q_0 = \sqrt{\frac{1}{R_S^2}\frac{L_S}{C_P} - 1} = \sqrt{\frac{94 \times 10^{-6}}{(0.87)^2 \times (6.74 \times 10^{-7})} - 1}$$

$$= 13.54$$

This value is the unloaded inductive-loop system quality factor with 100 ft (30 m) of Belden 8718 #12 AWG lead-in cable.

Assume that the inductive-loop electronics unit adds a shunt parallel resistance of 1,000 ohms. Then

$$Q_P = \omega_0 C_P R_P = (2\pi \times 20 \times 10^3) \times (6.74 \times 10^{-7}) \times 1000$$

$$= 84.70$$

Therefore, the total loaded loop system quality factor is

$$Q_L = \frac{Q_P\, Q_0}{Q_P + Q_0} = \frac{84.70 \times 13.54}{84.70 + 13.54}$$

$$= 11.67$$

Figure 2-7. Loop system quality factor sample calculation.

Table 2-2. Rectangular loop inductance and quality factor parameters at f = 20 kHz.*

Wire gauge (AWG)	1 Turn inductance (μH)	1 Turn quality factor *Q*	2 Turn inductance (μH)	2 Turn quality factor *Q*	3 Turn inductance (μH)	3 Turn quality factor *Q*	4 turn inductance (μH)	4 Turn quality factor *Q*	5 Turn inductance (μH)	5 Turn quality factor *Q*
12	10	20	35	30	73	37	123	43	184	47
14	11	16	36	24	74	30	125	35	186	40
14**	63	12	89	14	128	18	180	21	243	25
16	11	12	37	18	75	23	126	28	188	31
18	11	8	37	13	77	17	127	20	189	23

*6 x 6 ft (1.8 x 1.8 m) loop. ** With lead-in cable.

Table 2-3. Quadrupole loop inductance and quality factor parameters at f = 20 kHz.

Wire gauge (AWG)	1 Turn inductance (μH)	1 Turn quality factor *Q*	2 Turn inductance (μH)	2 Turn quality factor *Q*	3 Turn inductance (μH)	3 Turn quality factor *Q*	4 Turn inductance (μH)	4 Turn quality factor *Q*	5 Turn inductance (μH)	5 Turn quality factor *Q*
12	17	22	60	33	125	40	211	46	315	50
14	18	17	61	27	127	33	213	38	317	43
16	18	13	62	20	129	26	215	30	320	34
18	19	9	63	14	130	19	217	22	323	25

Table 2-4. Circular loop inductance and quality factor parameters at f = 20 kHz.*

Wire gauge (AWG)	1 Turn inductance (μH)	1 Turn quality factor *Q*	2 Turn inductance (μH)	2 Turn quality factor *Q*	3 Turn inductance (μH)	3 Turn quality factor *Q*	4 Turn inductance (μH)	4 Turn quality factor *Q*	5 Turn inductance (μH)	5 Turn quality factor Q
12	10	20	34	31	71	38	120	44	179	49
14	10	16	35	25	72	32	121	37	181	41
16	10	12	35	19	73	24	122	29	182	32
18	11	8	36	13	74	17	123	21	184	24

*7-ft (2.1-m) diameter loop.

LOOP LEAD-IN WIRE

Table 2-5 contains loop-to-pull box lead-in wire inductance, capacitance, and resistance values for two common types of wire. The two lead-in wires from the start and end of the loop turns should be twisted together to form a symmetrically twisted pair from the loop to the pull box. The twisting reduces crosstalk and noise pickup in the lead-in wire. Most manufacturers recommend at least five turns per foot (16.5 turns per meter). The wire twists form small loops along the wire, which alternate in winding direction. An external magnetic field from noise or crosstalk induces voltages in the small loops, which almost cancel, thus reducing interference. The importance of twists in the lead-in wire is discussed further in Chapter 5.

Table 2-5. Twisted loop lead-in wire specifications.

Wire manufacturer and type	Wire insulation type	AWG number	Jacket diameter (mils)	Number of twists per foot	Induc-tance (μH/ft)	Capaci-tance (pF/ft)	Resis-tance (Ω/ft)
XHHW	Cross-linked polymer	14 stranded	130	3 to 4	0.24	10	0.006
Belden 9438	High density polyethylene	14 stranded	139	5.5	0.22	10	0.00252

LEAD-IN CABLE

Shielded, twisted wire pairs are used for the lead-in cable (home run cable), which runs from the pull box to the electronics unit terminals in the controller cabinet. The conducting shield reduces interference from external electric fields. Lead-in cable inductance, capacitance, and resistance values for several types of cable are given in Table 2-6.

Table 2-6. Commercial lead-in cable specifications.

Cable manufacturer and type	Wire insulation type	AWG number	Insulation diameter (mils)	Cable insulation type	Induc-tance (μH/ft)	Capaci-tance (pF/ft)	Resis-tance (Ω/ft)
Belden							
8718	Polyethylene	12	37	Vinyl	0.2	25	0.0019
8720	Polyethylene	14	32	Vinyl	0.2	24	0.0029
8719	Polyethylene	16	32	Vinyl	0.2	23	0.0045
Clifford							
IMSA Specification 50-2-1984	Polyethylene	12	30	Polyethylene	0.2	25	0.0016
	Polyethylene	14	30	Polyethylene	0.2	24	0.0025
	Polyethylene	16	30	Polyethylene	0.2	23	0.0040

The measurements of loop system quality factor (with 100 ft (30 m) of shielded lead-in cable connected to a loop) in Appendix D show that little benefit is gained from using larger conductor diameters in the shielded lead-in cable. For example, the quality factor associated with #14 AWG shielded lead-in cable is not substantially reduced by substituting #12 cable. The principal loss results from the type of shielding rather than the conductor diameter. Table 2-7 illustrates how lead-in cable type and length affect the quality factor.

CALCULATING INDUCTANCE

Several simplified formulas are available for calculating the approximate inductance of an inductive-loop detector. More accurate inductance values are obtained from the mutual coupling method discussed in Appendix A.

The simplified formulas provide acceptable accuracy for the self inductance of multiturn, rectangular, quadrupole, and circular loops, which have a large area relative to the conductor spacing. The approximations compare favorably with a range of measured inductive-loop inductance values.

Appendix C contains calculated loop inductance values for various size loops and shapes (rectangular, quadrupole, and circular). Inductance and quality factor for several numbers of turns of wire were calculated using the mutual coupling formula discussed later in this chapter.

Table 2-7. Influence of lead-in cable type and length on *Q*.

Turns of #14 AWG wire in loop	Lead-in cable type, Belden	Lead-in cable length (ft)	Cable wire gauge (AWG)	Total parallel capac. (μF)	Series loop induct. (μH)	Lead-in cable induct† (μH)	Total series induct. (μH)	Loop resist.* (Ω)	Lead-in cable resist.†,** (Ω)	Total series resist. (Ω)	Loop system *Q*	Loop system loaded *Q* (kΩ)
3	8718	100	12	0.674	74	20	94	0.25	0.62	0.87	14	12
3	8720	100	14	0.670	74	21	95	0.25	0.80	1.05	11	10
3	8719	100	16	0.670	74	21	95	0.25	1.00	1.25	10	9
4	8718	100	12	0.437	125	20	145	0.33	0.62	0.95	19	14
4	8720	100	14	0.434	125	21	146	0.33	0.80	1.13	16	13
4	8719	100	16	0.434	125	21	146	0.33	1.00	1.33	14	11
5	8718	100	12	0.312	186	20	206	042	0.62	1.04	25	15
5	8720	100	14	0.306	186	21	207	0.42	0.80	1.22	21	14
5	8719	100	16	0.306	186	21	207	042	1.00	1.42	18	12
5	8718	1,000	12	0.172	186	200	386	0.42	6.20	6.62	7	5
5	8720	1,000	14	0.160	186	210	396	0.42	8.00	8.42	6	5
5	8719	1,000	16	0.160	186	210	396	0.42	10.00	10.42	5	4

Loop size is 6 x 6 ft (1.8 x 1.8 m). Excitation frequency is 20 kHz.
* Measured series resistance of loop 3 ft (0.9 m) above the laboratory floor.
** 8719 resistance value estimated.
† Lead-in cable length is 100 ft.

LOOP SYSTEM INDUCTANCE CALCULATIONS

Inductance attributed to the lead-in cable is added to wire loop inductance at the rate of 21 μH per 100 ft (30 m) of #14 AWG lead-in cable. For example, a 6- x 6-ft (1.8- x 1.8-m) rectangular loop should have three turns, according to Appendix C, and an inductance of 74 μH. If the lead-in cable is 200 feet (61 m) in length, the total inductance is

$$L = 74 + \left(\frac{200}{100}\right)(21) = 74 + 42 = 116 \ \mu\text{H} \ . \qquad \text{(2-14)}$$

The inductance L of two or more loops connected in series is additive such that $L = L_1 + L_2 \pm 2M$, where L_1 and L_2 represent the inductance of each of the individual series-connected loops, M is the mutual inductance between the two loops, and the sign of M is positive if flux is increased by current flowing in the same direction in the closest spaced loop wires.

The mutual inductance is negligible when the loops are separated by a large distance. In this case, $L = L_1 + L_2$, i.e., the loops are connected in series producing maximum loop inductance.

If the loops are connected in parallel, then the combined inductance is calculated as $1/L = 1/L_1 + 1/L_2$. For example, the combined inductance of two 6- x 6-ft (1.8- x 1.8-m) loops of three turns each connected in parallel is given by

$$\frac{1}{L} = \frac{1}{74} + \frac{1}{74} = \frac{2}{74} \ . \qquad \text{(2-15)}$$

Thus, $2L = 74$ μH and $L = 37$ μH.

Thus, parallel connection of loops reduces the inductance. Good design practice requires that the combined loop inductance be greater than the lower limit of 50 μH. Therefore, the parallel connection described above is not suitable as a vehicle sensor.

In some cases, both series and parallel connections of inductive loops are desirable. Consider, for example, four 6- x 6-ft (1.8- x 1.8-m) three-turn loops installed 9 ft (2.7 m) apart to provide detection in a left-turn lane. Three possible types of connections are shown in Figure 2-8. Connection in series produces an inductance of 4 x 74 = 296 μH. Parallel connection produces only 18.5 μH ($4L$ = 74 μH, L = 18.5 μH). A series-parallel configuration, where the upper two loops are connected in series and the bottom two loops are connected in series, produces two loop pairs, which are then connected in parallel to give a combined inductance of 74 μH.

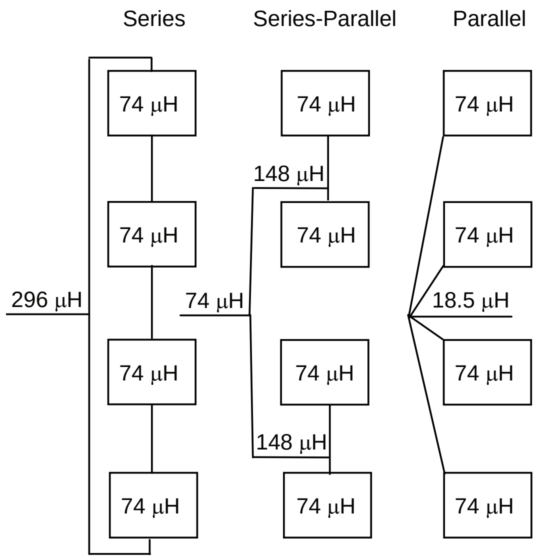


Figure 2-8. Four 6- x 6-ft (1.8- x 1.8-m) three-turn loops connected in series, parallel, and series-parallel.

NUMBER OF TURNS REQUIRED

Wire loops should have a sufficient number of turns to provide a nominal minimum inductance of 100 μH per loop to ensure stable operation of the inductive-loop system. A rule of thumb for the number of turns needed to produce an inductance value within the required range is:

- If the loop perimeter is less than 30 ft (9 m), use three turns of wire.
- If the loop perimeter is over 30 ft (9 m), use two turns of wire.

LOOP SENSITIVITY TO AN ELECTRICALLY CONDUCTIVE OBJECT

The current flowing through the loop wire creates a magnetic field around the wire as given by Equations 2-1, 2.1 and 2-3. If a vehicle (or any other electrically conductive object) enters this magnetic field and the magnetic field or a component of the magnetic field is normal to the area of the object, eddy currents are induced in the conducting object. The eddy currents generate another magnetic field that opposes the magnetic field of the loop, causing a decrease in the total magnetic field around the loop. Since the loop inductance is proportional to magnetic flux, the loop inductance decreases.

Eddy currents are induced in an electrically conductive object, such as a metal vehicle, by the magnetic field created by the current flowing through the wire loop. The eddy currents then generate a magnetic field that opposes the original magnetic field produced by the inductive loop. The result is a decrease in loop inductance.

Loop sensitivity to a conductive object can be tested with a 12 inches (30 cm) long wire formed into a circle whose diameter is approximately 4 inches (10 cm). The circular loop forms an open electrical circuit when the wire ends are held such that they do not touch one another. No actuation should occur when the open circular loop is rapidly moved horizontally over the in-roadway inductive loop. When the ends of the circular loop are made to touch, forming a closed circuit before being thrust over the in-roadway loop, an actuation will occur because of the flow of eddy currents. This demonstrates that it is the shorted turn, and not the wire or vehicle mass, which is important in producing the actuation.

BICYCLE AND MOTORIZED VEHICLE DETECTION MODELS

Figure 2-9 illustrates the detection of a bicycle or motorcycle by an inductive loop. These conveyances can be modeled as a vertical conducting object relative to the plane of the loop. When the cycle travels along the loop wire, eddy currents are induced in the conducting wheel rims and frame. When the cycle is directly over the loop wire, coupling between the inductive loop and the cycle is maximized.

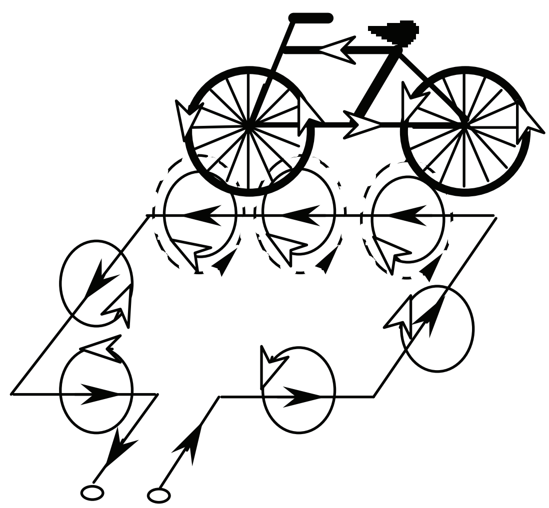


Figure 2-9. Bicycle detection showing induced eddy currents. The black arrows represent the current flow in the loop wire and the white arrows the induced flux.

A vehicle undercarriage, on the other hand, is a horizontal target. As shown in Figure 2-10, the undercarriage is modeled as a conducting rectangular plate, whose width is equal to the width of the vehicle and whose length is equal to the length of the vehicle at some average undercarriage height.

A conducting mesh can be used to approximate the electrical characteristics of the continuous plate. When the mesh is symmetrically located over the inductive loop to produce maximum sensitivity, all induced internal mesh currents cancel. This results in a single induced current flowing around the perimeter of the mesh, which is equivalent to a single turn rectangular wire loop or shorted turn. The air core transformer on the right of Figure 2-10 models the coupling between the vehicle undercarriage, as represented by a shorted turn of wire, and the inductive-loop wire.

Maximum vehicle detection sensitivity is produced by a shorted turn at minimum distance from the loop wires. Consequently, the ideal inductive-loop detector has a shape that approximates the vehicle's periphery. That is, a 6- x 6-ft (1.8- x 1.8-m) square loop would be preferable to one the size of a vehicle's engine.

Because of undercarriage height, high-bed trucks are difficult to detect. Detection of these vehicles is maximized when the width of the loop is equal to the width of the truck, lane width permitting. The length of the loop should not be less than its width to avoid a loss in sensitivity.

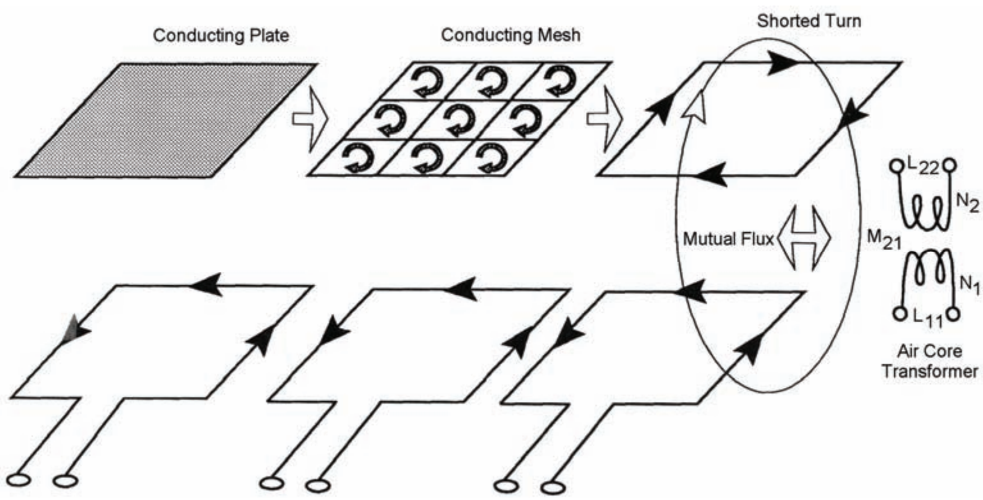


Figure 2-10. Vehicle undercarriage model. The upper part of the figure depicts the vehicle undercarriage electrical models and the lower part the inductive-loop wire.

MUTUAL INDUCTANCE

Inductive-loop self-inductance is defined using the loop's magnetic flux. When the magnetic flux of a loop couples to a vehicle, the coupled flux is used to define mutual inductance.

Figure 2-10 showed the magnetic coupling between a loop and shorted turn, which behaves as an air core transformer. The mutual inductance between the primary circuit (i.e., the inductive loop) and secondary circuit (i.e., the shorted turn) is given by

$$M_{21} = \frac{N_2 \times \phi_{21}}{I_1} \qquad \text{(2-16)}$$

where

M_{21} = Mutual inductance between circuit one (loop) and circuit two (shorted turn), henrys

N_2 = Number of turns (equals 1 for a shorted turn)

ϕ_{21} = Magnetic flux normal to shorted turn area, webers

I_1 = Current flowing in the loop, amperes.

LOOP SENSITIVITY

Loop sensitivity is equal to the change in loop system inductance induced by a conductive metal object divided by the original inductance of the loop system.

The loop sensitivity S_L of an inductive-loop detector is defined as

$$S_L = 100 \times \frac{L_{NV} - L_V}{L_{NV}} = 100 \times \frac{\Delta L}{L} \qquad \text{(2-17)}$$

where

L_{NV} = Inductance in absence of vehicle, henrys

L_V = Inductance with vehicle present, henrys.

The sensitivity S_L for the air core transformer shown in Figure 2-10, assuming a quality factor Q greater than 10, is given by

$$S_L = 100 \times K = \frac{(M_{21})^2}{L_{11} L_{22}} \text{ percent} \tag{2-18}$$

where

K = Coefficient of coupling

M_{21} = Mutual coupling between loop and shorted turn, henrys

L_{11} = Self inductance of loop, henrys

L_{22} = Self inductance of shorted turn, henrys.

Simplified expressions for the self inductance and mutual coupling can be derived by assuming the effect of vehicle iron is negligible. Then $\mu_r = 1$ and the self inductance of the roadway loop of length ℓ_1 is found from Equation 2-6 as

$$L_{11} = \frac{\mu_0 (N_1)^2 A F_1}{\ell_1} \ . \tag{2-19}$$

The inductance of the shorted turn loop of length ℓ_2 is given by

$$L_{22} = \frac{\mu_0 (N_2)^2 A_V F_2}{\ell_2} \ . \tag{2-20}$$

The mutual inductance between the shorted turn loop and the roadway loop is given by

$$M_{21} = \frac{\mu_0 N_1 A_V F_1}{d_{21}} \tag{2-21}$$

where

A_V = Area of vehicle undercarriage, (meters)^2

d_{21} = Distance between loop and shorted turn, meters.

The sensitivity is then expressed as

$$S_L = \frac{A_V \ell_1 \ell_2 F_1}{A (d_{21})^2 F_2} \tag{2-22}$$

where $A_V \leq A$.

Equation 2-22 shows that the sensitivity decreases for loop areas larger than the vehicle undercarriage area. The sensitivity decreases as the square of the vehicle undercarriage distance from the loop. The sensitivity is independent of the number of loop turns; however, pulling the turns apart slightly increases sensitivity by increasing ℓ_1 at the expense of a deeper sawcut slot in the roadway.

Appendix E contains more complex formulas for calculating S_L for two-turn and other multi-turn inductive loops. Comparisons of measured and calculated sensitivities are also available in this appendix.

Figure 2-11 illustrates the variation of loop sensitivity with vehicle undercarriage height for 6- x 2-ft (1.8- x 0.6-m), 6- x 4-ft (1.8- x 1.2-m), and 6- x 6-ft (1.8- x 1.8-m) three-turn inductive loops. The sensitivity of the 6- x 2-ft (1.8- x 0.6-m) loop is small because of its short length l_1.

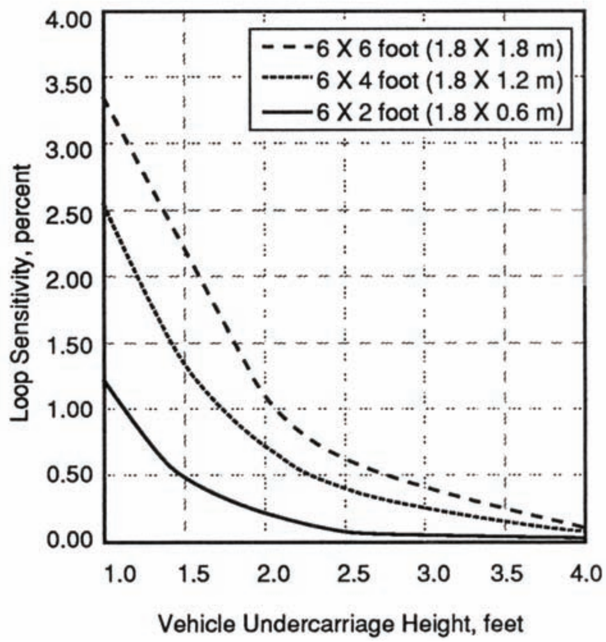


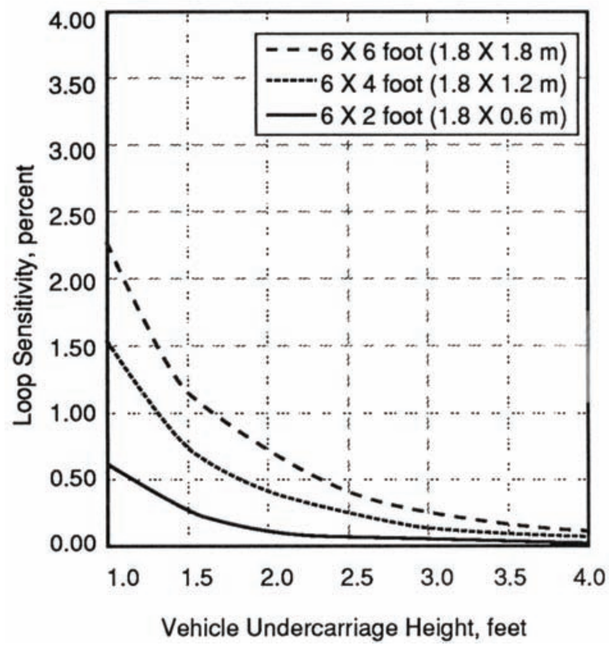
Figure 2-11. Calculated sensitivity of three-turn inductive loops as a function of vehicle undercarriage height.

Figure 2-12 depicts the decrease of loop sensitivity that occurs when a 200-ft (60-m) lead-in cable is added to the loops specified in Figure 2-11. The 6- x 2-ft (1.8- x 0.6-m) loop will probably double count a high-bed truck under this condition.

Figure 2-13 shows the decrease in loop sensitivity for a vehicle centered in two-turn long inductive loops as compared to the sensitivity of three-turn loops. Loop sensitivity decreases further when a lead-in cable is added.

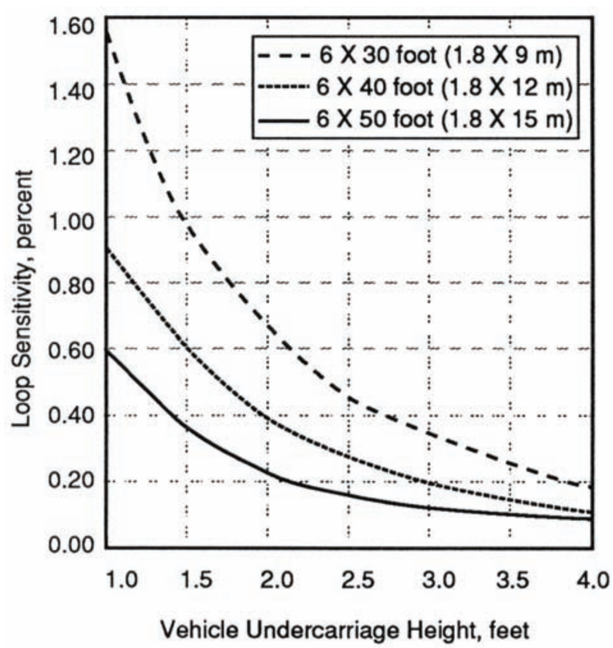
EFFECT OF REINFORCING STEEL

Figure 2-14 illustrates the reduction in loop sensitivity that occurs when an inductive-loop detector is installed over reinforcing steel mesh. The effect of the reinforcing steel is modeled as a shorted turn at twice the mesh spacing from the loop. The reinforcing steel reduces the magnetic field around the loop wire conductors, which causes a decrease in loop inductance and in loop sensitivity. Table 2-8 shows the effect on loop inductance when reinforcing steel is added to the pavement subsurface. The values are conservative since the mesh is assumed to be a perfect conductor. Modern inductive-loop detector electronic units are capable of detecting vehicles even though the loop wire is laid on the rebar before concrete is poured.



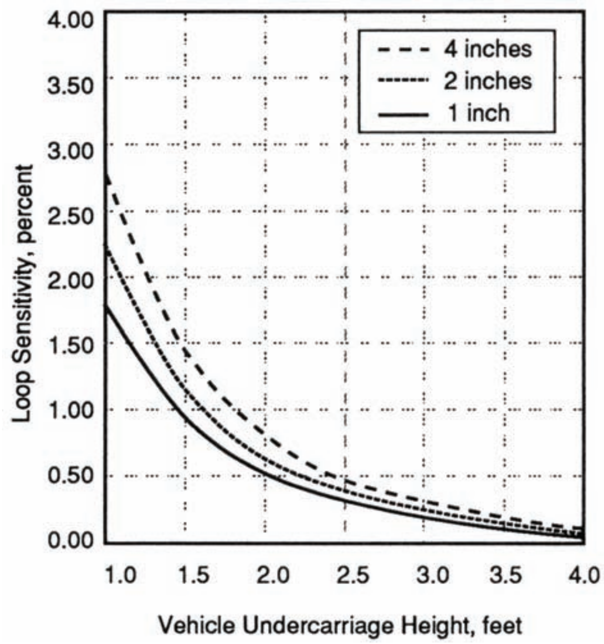
1 ft = 0.3 m

Figure 2-12. Calculated sensitivity of three-turn inductive loops with 200 ft (60 m) of lead-in cable as a function of vehicle undercarriage height.



1 ft = 0.3 m

Figure 2-13. Calculated sensitivity of two-turn long inductive loops as a function of vehicle undercarriage height.



1 ft = 0.3 m

Figure 2-14. Calculated sensitivity of 6- x 6-ft (1.8- x 1.8-m) inductive loop over reinforcing steel.

Inductive loops do not function as vehicle sensors when installed above steel rebar, whose pieces are connected such that current flows through the rebar. This induced current fully or partially cancels the vehicle-induced current in the inductive loop. If the rebar spacing is sufficiently large, then the current flows may not cancel. Conversely, if the rebar is not shorted together when it is installed, it will not support the flow of counter currents that inhibit the performance of the inductive loop.

Table 2-8. Influence of reinforcing steel on loop inductance (μH).

Number of turns	No reinforcing steel	Steel of 1-inch diameter	Steel of 2-inch diameter	Steel of 4-inch diameter
1	11	9	10	10
2	35	28	31	33
3	73	56	63	68
4	121	89	103	112
5	179	127	151	166
6	248	167	206	228
7	325	206	266	298

1 inch = 2.5 cm

Epoxy coatings normally placed on rebar are insulating in nature. However, the nature of the coating process usually places voids in the coating, which allow currents to flow. The number of allowable voids may be specified in the construction documents. The counter-current flow may increase in winter months in cold climates where salts are placed on the roadway or bridge deck.

LOOP SYSTEM SENSITIVITY

Loop system sensitivity is defined as the smallest change of inductance at the electronics unit terminals that will cause the controller to actuate. This sensitivity must be equal to or greater than the threshold for the electronics unit. Many states specify that the electronics unit must respond to a 0.02 percent change in inductance. NEMA Standards (see sSection 15.3.2 of Appendix J), recognizing the differences in electronics unit design ($\Delta L/L$ or ΔL), specify the sensitivity threshold for three classifications of test vehicles when they are centered in a single 6- x 6-ft (1.8- x 1.8-m) three-turn loop with 100 ft (30.5 m) of lead-in cable. The vehicle classes are:

- **Class 1:** 0.13 percent ($\Delta L/L$) or 0.12 µH (ΔL) inductance change (small motorcycle).
- **Class 2:** 0.32 percent ($\Delta L/L$) or 0.3 µH (ΔL) inductance change (large motorcycle).
- **Class 3:** 3.2 percent ($\Delta L/L$) or 3.0 µH (ΔL) inductance change (automobile).

An inductance in series or parallel with an inductive-loop detector will reduce the loop system sensitivity at the input terminals of the electronics unit.

A study performed for the SCANDI project in Detroit found that the duration of a call is impacted by the height of the flux field, which, in turn, depends on the presence and depth of reinforcing steel and other location-specific factors.[2] The study indicated that adjustable, diamond-shaped loops compensate for such factors at each location, resulting in a uniform duration from loop to loop for a given vehicle at a given speed.

SENSITIVITY OF TWO SERIES INDUCTORS

Figure 2-15 illustrates the total inductance calculation for the combination of two separate inductive loops connected in series as one equivalent loop. (Figure 2-19 illustrates the connection of two loops in this manner.) The equivalent total series inductance L_{TS} is

$$L_{TS} = L_A + L_B \tag{2-23}$$

where L_A and L_B are the individual inductance values of the loops.

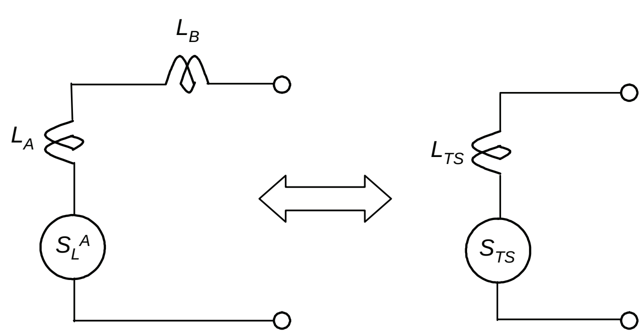


Figure 2-15. Equivalent total inductance from two inductive loops in series.

The equivalent total series sensitivity S_{TS} is

$$S_{TS} = S_L^A \frac{1}{1+\frac{L_B}{L_A}} \tag{2-24}$$

where

S_L^A = Loop sensitivity as vehicle enters Loop A.

SENSITIVITY OF TWO PARALLEL INDUCTORS

Figure 2-16 illustrates the sensitivity calculation for two separate inductive loops connected in parallel as an equivalent single loop. (Figure 2-21 illustrates the connection of two loops in this manner.) The equivalent total parallel inductance L_{TP} is

$$L_{TP} = \frac{L_A \times L_B}{L_A + L_B}. \tag{2-25}$$

The equivalent total parallel sensitivity S_{TP} is

$$S_{TP} = S_L^A \frac{1}{1+\frac{L_A}{L_B}}. \tag{2-26}$$

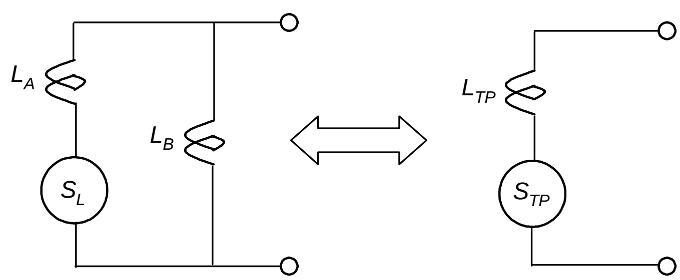


Figure 2-16. Equivalent total inductance from two inductive loops in parallel.

SINGLE LOOP EXAMPLE

1. What is the loop sensitivity at the pull box assuming a high-bed vehicle (4-ft (1.2-m) undercarriage) passes over the loop? Figure 2-17 illustrates this case and gives the lead-in wire lengths. The equivalent electrical circuit is shown in Figure 2-18.

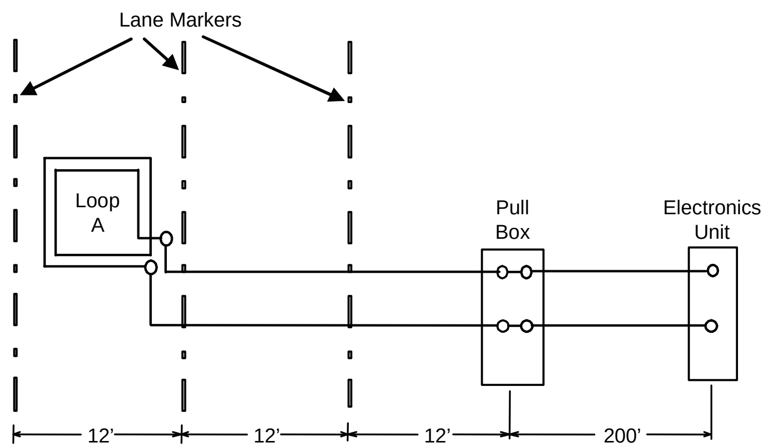


Figure 2-17. Single inductive loop connected to a pull box and electronics unit.

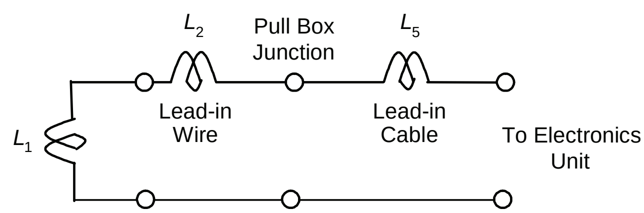


Figure 2-18. Equivalent single loop electrical circuit.

The sensitivity S_L for a 4-ft (1.2-m) high undercarriage and a three-turn, 6- x 6-ft (1.8- x 1.8-m) loop of #14 AWG wire is 0.1 percent from Figure 2-11. The twisted loop wires form an approximately 24-ft (7.3-m) lead-in wire to the pull box. The inductance per foot for #14 AWG loop wire with 5 twists per foot is 0.22 μH/ft (0.7 μH/m). The lead-in inductance L_S is

$$L_S = (0.22\,\mu\text{H/ft}) \times (24\text{ ft}) = 5.3\,\mu\text{H}\,. \tag{2-27}$$

The self inductance L_L of a three-turn, 1.8-1.8-m (6- x 6-ft) loop of #14 AWG wire at 20 kHz from Appendix C is 74 μH. Therefore, the sensitivity S_P (in percent) at the pull box is

$$S_P = \frac{S_L}{1 + \frac{L_S}{L_L}} = \frac{0.1\%}{1 + \frac{5.3\,\mu\text{H}}{74\,\mu\text{H}}} = 0.093\%\,. \tag{2-28}$$

2. What is the inductive-loop system sensitivity at the input terminals of the electronics unit with a 200-ft (61-m) length of Type 8720 shielded lead-in cable between the pull box and the electronics unit?

From Table 2-6, the inductance of type 8720 cable is 0.22 μH/ft. The total series inductance between the loop and the input terminals of the electronics unit is

$$L_S = [(0.22) \times (24)] + [(0.22) \times (200)] \tag{2-29a}$$

$$L_S = 5.3\,\mu\text{H} + 44\,\mu\text{H} = 49.3\,\mu\text{H}\,. \tag{2-29b}$$

Then the sensitivity S_D at the input terminals of the electronics unit is

$$S_D = \frac{S_L}{1+\frac{L_S}{L_L}} = \frac{0.1\%}{1+\frac{49.3\,\mu\text{H}}{74\,\mu\text{H}}} = 0.060\% \,. \quad (2\text{-}30)$$

3. What is the inductive-loop system sensitivity at the input terminals of the electronics unit with a 200-ft (61-m) length of Type 8720 shielded lead-in cable between the pull box and the electronics unit if a four-turn, 6- x 6-ft (1.8- x 1.8-m) loop #14 AWG wire is used?

The sensitivity S_L for a 4-ft (1.2-m) high undercarriage and four-turn, 6- x 6-ft (1.8- x 1.8-m) loop is 0.1 percent. From Appendix C, the loop self inductance is 125 µH at 20 kHz. The series inductance is the same as in the previous example.

Therefore

$$S_D = \frac{S_L}{1+\frac{L_S}{L_L}} = \frac{0.1\%}{1+\frac{49.3\,\mu\text{H}}{125\,\mu\text{H}}} = 0.072\% \,. \quad (2\text{-}31)$$

TWO LOOPS IN SERIES EXAMPLE

1. What is the inductive-loop system sensitivity at the input terminals of the electronics unit when a second identical loop is placed in series with the loop sensing the vehicle? Figure 2-19 illustrates the loop configuration and shows lead-in wire lengths. The series connection is made in the pull box.

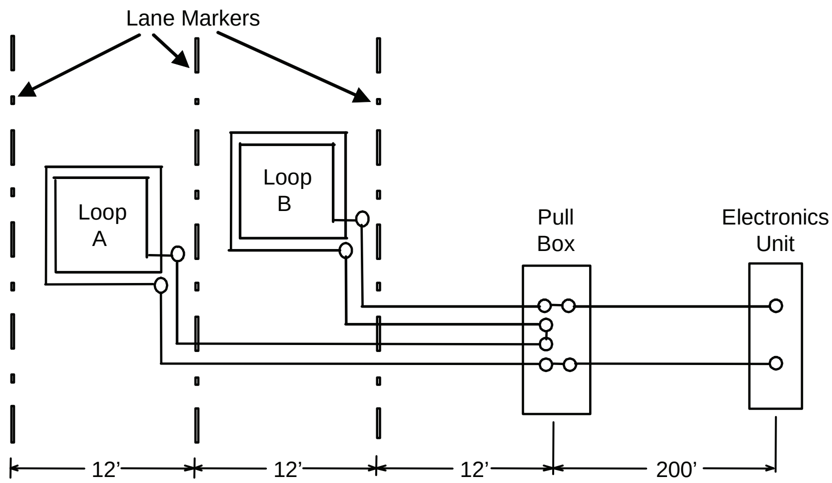


Figure 2-19. Two inductive loops connected in series to a pull box and electronics unit.

Figure 2-20 shows the equivalent electrical circuit. The sensing loop is a 6- x 6-ft (1.8- x 1.8-m), three-turn loop of #14 AWG wire. The self inductance of the second loop (i.e., series Loop B) is 74 µH. The lead-in wire inductance for Loop B is

$$L_S = [(0.22\,\mu\text{H/ft}) \times (12\text{ ft})] = 2.6\,\mu\text{H} \,. \quad (2\text{-}32)$$

The total series inductance of Loop B and lead-in wire to the pull box is

$$L_T^B = 2.6\,\mu\text{H} + 74\,\mu\text{H} = 76.6\,\mu\text{H} \quad (2\text{-}33)$$

and the total series inductance between the two loops and the input terminals of the electronics unit is

$$L_T^S = L_T^B + L_T^A + L_S \tag{2-34a}$$

$$= 76.8\,\mu\text{H} + 79.3\,\mu\text{H} + 47.6\,\mu\text{H} = 203.7\,\mu\text{H}\,. \tag{2-34b}$$

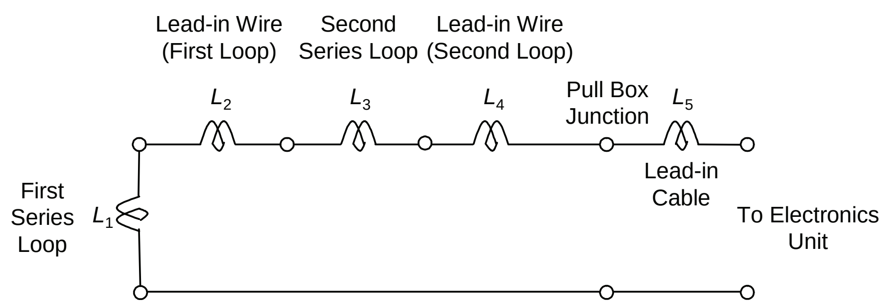


Figure 2-20. Equivalent electrical circuit for two loops connected in series to a pull box and electronics unit.

Then

$$S_D = \frac{S_L}{1+\dfrac{L_T}{L_L}} = \frac{0.1\%}{1+\dfrac{203.7\,\mu\text{H}}{74\,\mu\text{H}}} = 0.066\%\,. \tag{2-35}$$

TWO LOOPS IN PARALLEL EXAMPLE

1. What is the loop system sensitivity at the electronics unit terminals with two identical loops connected in parallel? Figure 2-21 illustrates the loop configuration and shows lead-in wire lengths. The equivalent electrical circuit is shown in Figure 2-22. All parameters are the same as in the previous series loop example. The total inductance and sensitivity at the input to the electronics unit are calculated as

$$L_{TS} = L_1 + L_2 \tag{2-36}$$

and

$$S_{TS} = \frac{S_L}{1+\dfrac{L_2}{L_1}}\,. \tag{2-37}$$

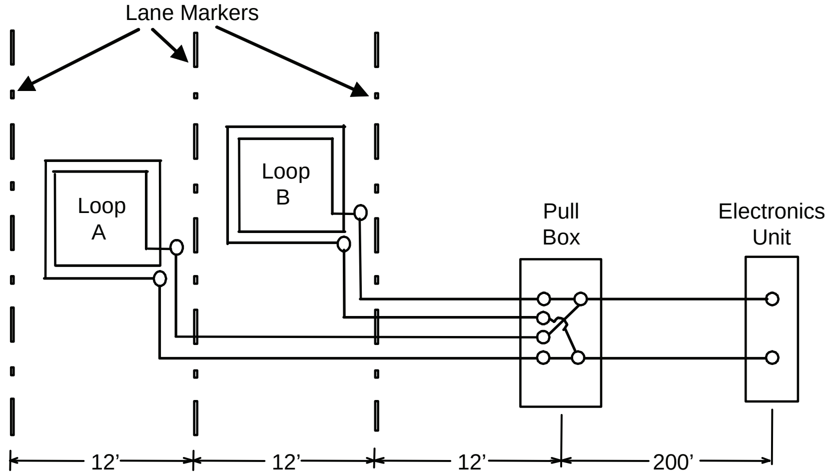


Figure 2-21. Two inductive loops connected in parallel to a pull box and electronics unit.

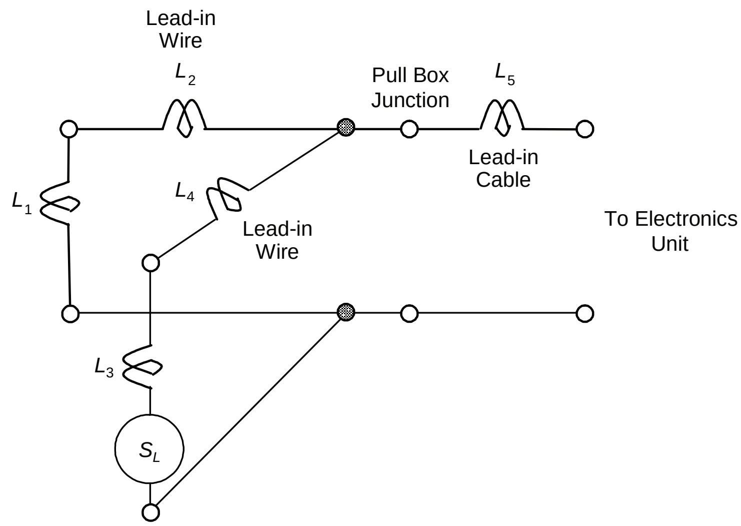


Figure 2-22. Equivalent electrical circuit for two loops connected in parallel to a pull box and electronics unit.

Let

$$L_A = L_1 + L_2 \tag{2-38}$$

and

$$L_B = L_3 + L_4 \,. \tag{2-39}$$

Then

$$L_{TP} = \frac{L_A \times L_B}{L_A + L_B}$$
$$= \frac{(L_1 + L_2) \times (L_3 + L_4)}{L_1 + L_2 + L_3 + L_4} \tag{2-40}$$

$$S_{TP} = S_{TS}\frac{1}{1+\dfrac{L_A}{L_B}}$$

$$= S_{TS}\frac{1}{1+\dfrac{L_1+L_2}{L_3+L_4}} \tag{2-41}$$

and

$$L_D = L_{TP} + L_S = \frac{L_A \times L_B}{L_A + L_B} + L_S \tag{2-42}$$

$$= \left[\frac{(L_1+L_2)\times(L_3+L_4)}{L_1+L_2+L_3+L_4}\right] + L_S\,. \tag{2-43}$$

Therefore

$$L_D = S_D = S_L\frac{1}{1+\dfrac{L_2}{L_1}} \times \frac{1}{1+\dfrac{L_1+L_2}{L_3+L_4}} \times \frac{1}{1+\dfrac{1}{\dfrac{(L_1+L_2)\times(L_3+L_4)}{L_1+L_2+L_3+L_4}}}\,. \tag{2-44}$$

RESONANT CIRCUIT

Many self-tuning inductive-loop electronics units use a frequency shift or change in period of an oscillator to indicate the passage or presence of a vehicle. The frequency of the oscillator is controlled by a parallel resonant circuit, sometimes called a tank circuit, composed of the equivalent loop system inductance and the tuning capacitance found in the electronics unit. The equivalent loop system capacitance also includes capacitive effects due to the placement of the loop wires in the sawcut. The associated equivalent quality factor accounts for the effect of system resistance losses. If the equivalent loop system inductance is too small, the oscillator will not oscillate. The manufacturer of the electronics unit specifies the range of loop system inductance and minimum loop system quality factor that are acceptable.

The oscillator frequency is calculated as

$$f_D = \frac{1}{2\pi\sqrt{L_D C_D\left(1+\dfrac{1}{(Q_D)^2}\right)}} \tag{2-45}$$

where L_D, C_D, Q_D are the inductance, capacitance, and quality factor, respectively, of the tank circuit.

Equation 2-45 shows that a decrease in the inductance increases the resonant frequency. Furthermore, a quality factor greater than five will have negligible effect on the performance of the resonant circuit.

TEMPORARY LOOPS

Several manufacturers and State agencies have sought to develop a durable and cost effective temporary loop that satisfies the needs of speed monitoring, vehicle counting, vehicle classification, and portable weigh-in-motion (WIM) programs. Two types of temporary and portable loop systems are described below.

MAT-TYPE LOOPS

The mat-type temporary loop consists of a durable rubber mat into which multiple turns of wire are embedded. The mats are usually smaller in width than the typical 6-ft (1.8-m) inductive loop. Standard sizes vary, ranging from 4 x 6 ft (1.2 x 1.8 m) to 3 x 6 ft (0.9 x 1.8 m). The mats are positioned in the center of the traffic lane with the longer dimension parallel to the flow of traffic so that most vehicles straddle the mat, thereby extending the life of the mat. A typical installation is shown in Figure 2-23. Nails and washers are commonly used to secure the mat to the road surface. A wide 3-inch (7.6-cm) heavy-duty adhesive tape is applied to prevent the mat edges from lifting. The lead-in wires from the mat to the data collection equipment at the roadside are encased between two layers of tape.

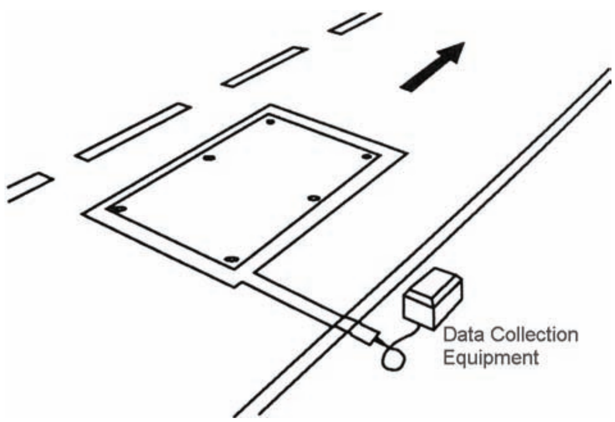


Figure 2-23. Typical installation of mat-type temporary inductive-loop detector.

Some agencies have manufactured this type of sensor in their own shop. However, hand-producing these mats was too labor intensive to be cost-effective. The mats were reliable, but with heavy truck traffic, some of the mats did not last more than a few hours.

OPEN LOOP CONFIGURATION

One manufacturer produces a preformed temporary portable loop that is 4 x 6 ft (1.2 x 1.8 m). The loop is composed of a sandwich of five layers as illustrated in Figure 2-24. The bottom layer is a 4-inch (101.6-mm) wide paper release sheet, which protects the 2-inch (50.8-mm) wide strip of adhesive bituminous rubber compound. Its upper surface is finished with a high-density polyethylene film. This padding strip is the bed for three turns of #22 AWG loop wire. An identical 2-inch (50.8-mm) padding strip covers the loop wires. The top layer is a 4-inch (101.6-mm) wide strip of adhesive bituminous compound reinforced with an overlay of woven polypropylene mesh.

The preformed, open loop configuration can be transported to the selected location and installed by one man in a few minutes. Installation consists of removing the bottom backing, positioning the loop on the roadway, and applying sufficient pressure to ensure adhesion. Five feet of protected lead-in wire is standard. Other loop dimensions and protected lead-in wire lengths are available.

Another approach to the open loop configuration was developed by the Special Study Section of the Nevada Department of Transportation (DOT).[3] The Nevada DOT previously used a 6- x 6-ft (1.8- x 1.8-m) portable loop constructed of three turns of #14 AWG stranded copper wire wrapped with black duct tape. Time consuming durability and maintenance issues increased as the use of the portable loops increased. This led to the testing of a variety of tapes, rubber tubing, and a rubber mat material as replacements for the duct tape covering of the original loops.

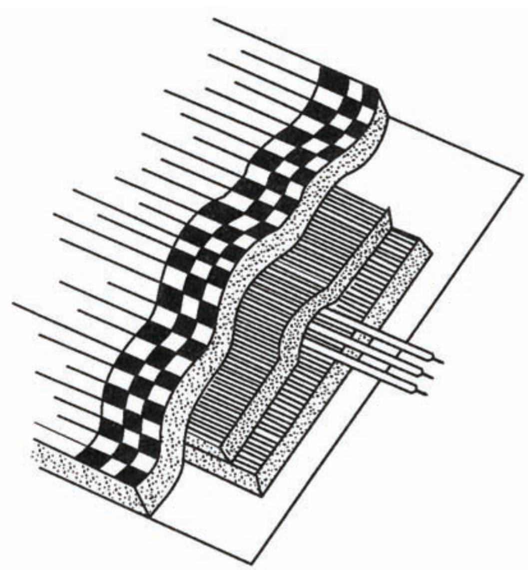
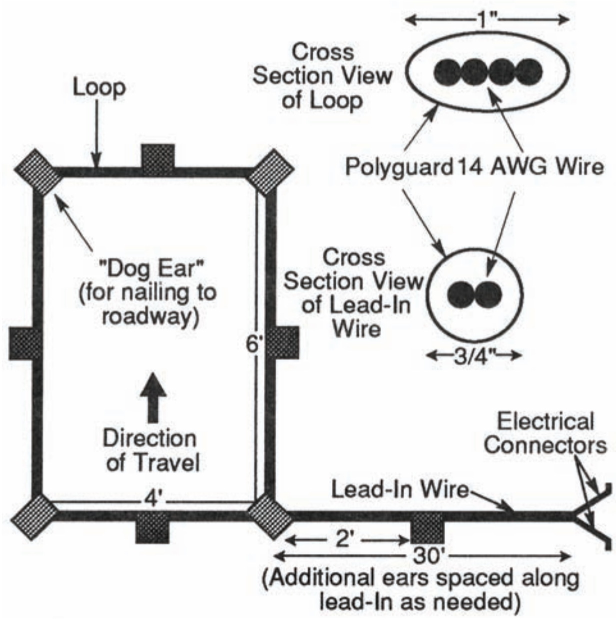


Figure 2-24. Five-layer temporary open loop detector configuration.

A bitumen tape manufactured by Polyguard Products was eventually selected to enclose the wire loops. It is a fabric-reinforced rubber-like material with one adhesive side. The final configuration consists of four turns of #14 AWG copper wire, wound in the shop, and taped together for easy handling. The loops are encased in two wraps of the Polyguard material and installed as shown in Figure 2-25.



1 ft = 0.3 m
1 inch = 2.5 cm

Figure 2-25. Nevada portable open loop installation.

A number of tests were conducted to measure the durability and accuracy of the loops as compared with conventional loops installed in sawcut slots. Other tests compared the 4- x 6-ft (1.2- x 1.8-m) configuration to the 6- x 6-ft (1.8- x 1.8-m) configuration. The test loops were installed on a rural two-lane FAP roadway with a high percent of multiple unit trucks. Both series of tests used the same counter/classifier recorder.

After almost 5,000 actuations, there was less than 1 percent difference between the number of vehicles counted by this type of portable loop and a saw-cut installed loop. It was also found that the 4- x 6-ft (1.2- x 1.8-m) loop size performed virtually the same as the 6- x 6-ft (1.8- x 1.8-m) size whether the loop was in a saw-cut or in the portable form.

The portable loops were still functioning after a more than yearlong product durability test consisting of over a million activations. This evaluation, on U.S. 395 between Reno and Carson City, NV, showed the loops to be extremely durable and capable of withstanding a wide range of weather conditions. The roadway contained an asphalt surface and, after several months, the loops became embedded in the pavement, which may have contributed to their longevity. On a concrete surface, these loops are expected to last well past one-half million activations. The loops have also been used with overlays and were able to withstand the heat involved in this process.

Testing in a semipermanent location enhanced loop longevity, as the loops were not subjected to repeated removal and reinstallation. However, other loops of the same type have been repeatedly installed without signs of undue wear. As a result of these tests and the experience with these loops, the Nevada DOT is now using the Polyguard loop in all of their portable loop installations.

ELECTRONICS UNITS

The electronics unit, which generates the inductive-loop excitation frequency and monitors the operation of the inductive-loop system, has changed significantly since the 1970s. Early versions of inductive-loop electronic units operated at a fixed resonant frequency, using a crystal to stabilize the frequency. There were many problems with the crystal electronics units, particularly when used with long lead-in cables.

One of these was resonant-frequency drift due to environmental changes in temperature and moisture. These units were removed from service in the 1970s and were initially replaced with designs that incorporated analog phase shifters, which were capable of compensating for (or tracking) drift caused by environmental changes. Modern electronics units stabilize the oscillation frequency and detect vehicles with configurations that incorporate digital frequency shift, digital ratioed frequency shift, digital period shift, and digital ratioed period shift designs. The theory of operation of these devices is described below. Analog phase shift electronics units are still in limited use for vehicle classification.

ANALOG PHASE-SHIFT ELECTRONICS UNIT

This device was developed to meet the demands of the European market, where bicycles must be detected. Like the crystal model, it operates as a phase shift sensor, but uses two variable-frequency oscillators rather than one crystal-controlled oscillator. The loop oscillator operates at a frequency between 25 and 170 kHz, as determined by the loop and lead-in wire. The loop oscillator is coupled to a second internal oscillator such that the initial manual tuning procedure brings the two oscillators into synchronization in frequency and phase.

The tuning knob moves a ferrite core back and forth inside an inductor, causing the oscillator connected to it to change its frequency (and phase) to match that of the loop oscillator. Arrival of a vehicle into the loop decreases loop inductance and the loop oscillator attempts to pull out of synchronization with its companion oscillator. It is not able to change frequency because of a cross-coupling resistor, but a phase shift is developed that is the basis for vehicle detection.

With this design concept, the electronics unit is able to compensate for (or track) environmental drift. As the temperature within the controller cabinet changes, the two oscillators drift identically. The output of the two oscillators are fed to a phase comparison circuit, which develops a DC voltage proportional to the amount of shift; thus, the term analog because it uses varying voltages rather than digital counts to indicate vehicle passage or presence.

When there are no vehicles within the detection zone, the DC voltage is stored and remembered by a memory capacitor. When a vehicle causes a change in the phase comparator output, the difference between it and the memory capacitor causes a relay to change state. Very slow changes in the DC voltage are followed by the memory capacitor, which allows the circuit to compensate for drift due to environmental changes. The memory circuit will ultimately forget a vehicle parked over the loop and drop that call. Detail on compensating for environmental drift is included later under the “Digital Frequency Shift Electronics Unit.”

OVERVIEW OF DIGITAL ELECTRONICS UNITS

The stability and added features afforded by electronic digital processing have lead most inductive-loop detector electronics unit manufacturers to produce digital devices. Digital techniques allow more reliable, accurate, and precise measurements than the analog techniques.

In using digital electronics units, one must be aware of the relation between increased sensitivity and the resulting increased response time. A large response time can result in a significant error in vehicle velocity measurements when two loops are utilized in a speed trap configuration (i.e., separated by a known and measured distance). Response times vary with electronics unit manufacturers.

Digital electronics units sense either a change in the frequency or period of a waveform. The oscillator frequency or period shift is caused by the decrease in loop inductance created when a vehicle is within the loop's detection zone. The oscillator frequency for a quality factor Q of 5 or greater is given by

$$f_D = \frac{1}{2\pi\sqrt{L_D C_D}} \qquad \text{(2-46)}$$

where

f_D = Oscillator frequency, Hz

L_D = Total inductance (i.e., loop plus lead-in cable) across the input terminals of the electronics unit, henrys

C_D = Total capacitance across the input terminals of the electronics unit, henrys.

The normalized oscillator frequency change due to a normalized change in inductance at the input terminals to the electronics unit when the Q is 5 or greater is given by

$$\frac{\Delta f_D}{f_D} = -\frac{1}{2}\frac{\Delta L_D}{L_D} = -\frac{1}{2}S_D \qquad \text{(2-47)}$$

where

Δf_D = Change in oscillator frequency of the electronics unit, Hz

ΔL_D = Change in inductance at the input terminals of the electronics unit, henrys

S_D = Sensitivity of the electronics unit to the inductance change.

Vehicle detection by an inductive-loop detector system is primarily induced by vehicle proximity to a buried inductive wire loop, which causes a change in the loop inductance in the inductance-capacitance (LC) oscillator circuit formed by the loop, lead-in cable, and the input capacitor located in the electronics unit. Some manufacturers detect vehicles through the percent change of loop inductance $\Delta L_L/L_L$, while others simply use the change in loop inductance ΔL_L. Neither of these quantities can be measured directly at the input terminals to the electronics unit. However, to indicate sensitivity, several manufacturers provide frequency meters to measure the resonant frequency and the amount of frequency change.

Experience shows that the percentage change of inductance ($\Delta L_L/L_L$) from an unoccupied loop to an occupied loop is extremely repeatable for a given loop size and geometry, a given vehicle size and geometry, and a given location of the vehicle with respect to the loop. Since parameters such as actual loop inductance and loop operating frequency do not affect $\Delta L_L/L_L$ but do affect ΔL_L, the following discussions and computations address the $\Delta L_L/L_L$ concept. The term "electronics unit sensitivity," in the context of this discussion, is defined as the value of $\Delta L_L/L_L$ that actuates the electronics unit with smaller values, which are interpreted as denoting greater sensitivity.

For short lead-in cables with negligible series cable inductance

$$\frac{\Delta L_D}{L_D} = \frac{\Delta L_L}{L_L} = S_L \tag{2-48}$$

where

$\Delta L_L =$ Change in loop inductance when sensing a vehicle, henrys

L_L = Loop inductance, henrys

S_L = Sensitivity of loop to vehicle in detection area.

The period of the oscillator T_D is defined as the inverse of the frequency f_D. For a Q of 5 or greater, T_D is given by

$$T_D = \frac{1}{f_D} = 2\pi\sqrt{L_D C_D}\ . \tag{2-49}$$

The normalized change in oscillator period caused by a normalized change in inductance at the input terminal to the electronics unit when Q is 5 or greater is approximately equal to

$$\frac{\Delta T_D}{T_D} = -\frac{1}{2}\frac{\Delta L_D}{L_D} = -\frac{1}{2}S_D. \tag{2-50}$$

The negative sign indicates that the change in period is opposite in direction to the change in inductance.

With the advent of sophisticated digital microprocessors and the availability of loop network resonant frequency information at the electronics unit input terminals, precise measurements of the following parameters can be obtained with relative ease:

- Frequency shift (Δf_D).
- Ratioed frequency shift ($\Delta f_D/f_D$).
- Period shift (ΔT_D).
- Ratioed period shift ($\Delta T_D/T_D$).

The four types of digital electronics units, each utilizing one of these measurement techniques, are introduced below. Detailed analyses and block diagrams of each device are provided in Appendices F through I.

DIGITAL FREQUENCY SHIFT ELECTRONICS UNIT

This type of unit is not manufactured. However, the theory and operating characteristics associated with this concept are included so that the operation of the digital ratioed frequency shift electronics unit may be better understood.

The digital processor in a digital shift electronics unit would compare counts proportional to the oscillator frequency when a vehicle is present to a reference count taken periodically when no vehicles are present. The reference count is stored in a memory. During vehicle detection, when the count exceeds the reference count by a preset sensitivity threshold count, a vehicle call is initiated.

The sensitivity S_D^f of a frequency shift electronics unit is calculated from Equation 2-47 as

$$S_D^f = -2\frac{\Delta f_D}{f_D}. \tag{2-51}$$

Appendix F shows that

$$S_D^f = K_f\sqrt{L_D C_D} = \frac{2N_{ft}}{N_{fc}} \tag{2-52}$$

where

N_{ft} = Fixed frequency threshold count selected by sensitivity switch

N_{fc} = Number of oscillator cycles counted by the variable frequency counter

K_f = Frequency sensitivity constant.

In the digital frequency shift method, S_D is proportional to the square root of the $L_D C_D$ product. Since larger values of S_D represent decreased sensitivity, it follows that sensitivity decreases in proportion to the square root of the $L_D C_D$ product with a Δf_D measurement. Hence, every time the frequency switch is changed to a new position (e.g., to avoid crosstalk), the sensitivity would change and, if critical, would necessitate a new setting of the sensitivity switch.

Increased lead-in cable length increases the inductance of the lead-in cable and, hence, causes some loss of sensitivity. The increased $L_D C_D$ product would cause even more loss of sensitivity. Therefore, this type of measurement does not appear to be practical.

DIGITAL RATIOED FREQUENCY SHIFT ELECTRONICS UNIT

The digital processor in the digital ratioed frequency shift electronics unit compares counts proportional to the oscillator frequency when a vehicle is present to a reference count taken periodically when no vehicle is present. The reference count is stored in memory. When the count during vehicle detection exceeds the reference count by a preset sensitivity threshold count, a vehicle call is initiated.

The ratioed frequency shift electronics unit differs from the frequency shift unit in that the frequency counter is held approximately constant (as explained further in Appendix G).

The sensitivity S_D^f is independent of the inductance L_D and the capacitance C_D across the terminals of the electronics unit. The sensitivity is calculated as

$$S_D^f = 2\frac{N_{ft}}{N_{fc}} \tag{2-53}$$

where

N_{ft} = Fixed frequency threshold count

N_{fc} = Count produced by the fixed frequency counter.

From Appendix G, the measurement response time t^f is

$$t^f = 2\frac{N_{ft}}{m f_D S_D^f} \tag{2-54}$$

where m is a frequency multiplier.

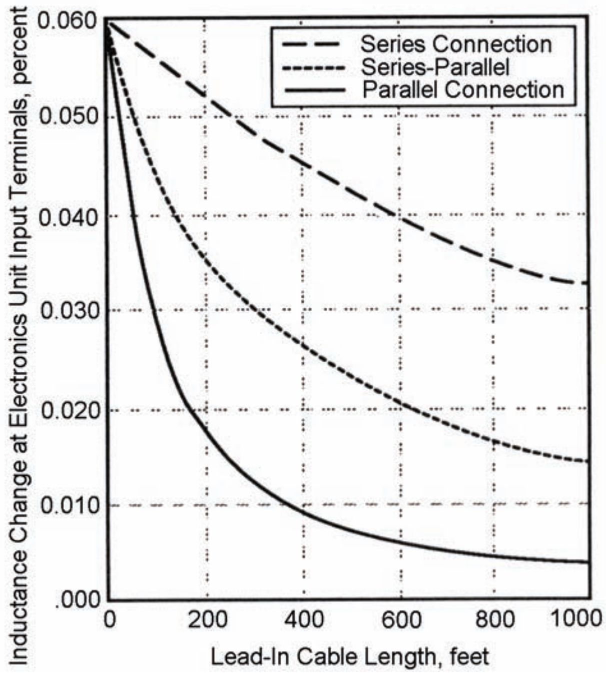
The advantage of having the sensitivity independent of the inductance and capacitance across the input terminals of the electronics unit is illustrated by the following example. This example also applies to the digital ratioed period shift electronics unit discussed later.

Assume four equal size loops, say 6 x 6 ft (1.8 x 1.8 m) with an equal number of turns, say three. Connect the loops as shown in Figure 2-8, namely

- All series (296 μH).
- Series-parallel (74 μH).
- All parallel (18.5 μH).

For simplicity, lead-in cable length is not considered. The sensitivity of the ratioed frequency or ratioed period shift electronics unit is identical for the above three loop connection configurations. Accordingly, a sensitivity threshold sufficient to detect a small motorcycle over one of the four loops when wired in series need not be changed when rewired in series-parallel or all parallel.

Although lead-in cable length was not considered above, an extra long lead-in cable will produce varying amounts of change in inductance due to inductance sharing. The amount of change depends on the length of the lead-in cable and the wiring scheme used for the multiple loops. Figure 2-26 provides an estimate of the inductance change at the electronics unit input terminals generated by a small motorcycle traveling over one of the four loops.[1]



1 ft = 0.3 m

Figure 2-26. Inductance change produced by a small motorcycle as a function of lead-in cable length for series, parallel, and series-parallel connections of four 6- x 6-ft loops.

DIGITAL PERIOD SHIFT ELECTRONICS UNIT

The digital period shift concept utilizes the period of the loop oscillator frequency, where period is defined as the time required for one full cycle of the oscillator frequency. The period is calculated by dividing one by the frequency in Hz or equivalently dividing one by the frequency in cycles per second.

Digital period shift electronics units use a reference clock running at megahertz (MHz) frequencies, i.e., between 20 and 100 times faster than the inductive-loop oscillation frequency, to measure the period of the loop oscillation as shown in Figure 2-27. Measurement precision is enhanced without sacrificing a great deal of time between measurements by determining the time for 32 oscillation cycles for sensitivity one, 64 cycles for sensitivity two, and so on. The loop oscillation period is calculated in terms of the number n of reference clock cycles contained within the period. Since the oscillation frequency increases when a vehicle passes over the loop, the period of the oscillation decreases, as it is equal to the inverse of the frequency. The reduction in the oscillation period results in a smaller number of cycles of the reference clock within the oscillation period. When the number of reference cycles is reduced more than a preselected threshold, a call is initiated to indicate the presence of a vehicle.

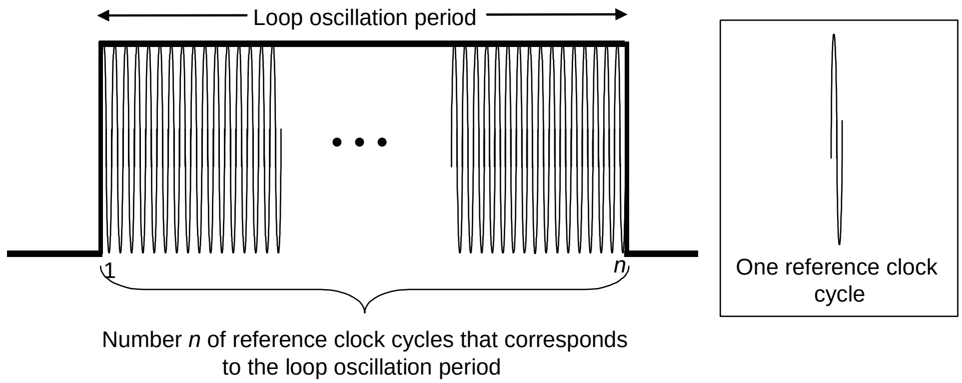


Figure 2-27. Measurement of inductive-loop oscillation period by a reference clock (Source: L.A. Klein, *Sensor Technologies and Data Requirements for ITS* (Artech House, Norwood, MA, 2001)).

A judicious choice of reference clock frequency and threshold value (4 counts ± 2 counts) makes the digital period shift design practical at any frequency encountered in practice. The time to detect is sufficiently short to enable the electronics to sequentially scan or operate four small loops, one at a time, several times a second. (Multichannel operation is discussed later.)

The period shift electronics unit is entirely self-tuning on installation and, similar to most other designs, is able to track environmental drift. Like the digital frequency shift unit, most models stop tracking for a time after a vehicle enters the loop, to guarantee that the call placed by a small vehicle is held long enough to bring the green to that approach.

The sensitivity S_D^p of a period shift electronics unit is found from Equation 2-50 as

$$S_D^p = 2\frac{\Delta T_D}{T_D}. \tag{2-55}$$

Appendix H shows that

$$S_D^p = \frac{K_p}{\sqrt{L_D C_D}} = 2\frac{N_{pt}}{N_{pc}} \tag{2-56}$$

where

N_{pt} = Fixed frequency threshold count selected by the sensitivity switch

N_{pc} = Count produced by the variable frequency counter

K_p = Frequency sensitivity constant.

Sensitivity S_D^p is inversely proportional to the square root of the LC product with a ΔT measurement. When S_D^p takes on small values, the sensitivity is increased. Hence, with increased lead-in cable length, part of the loss of sensitivity due to the added lead-in cable inductance is automatically compensated for by the increase in the LC product. Unfortunately, the compensation is not perfect because of the square root relationship.

The response time t^p of the electronics unit, as derived in Appendix H, is

$$t^p = 2\frac{N_{pt}}{f_c S_D^p}. \tag{2-57}$$

Most electronics units use a transformer to connect the external inductive-loop terminals to the internal oscillator. A loosely coupled transformer produces a series leakage or swamping inductance. This inductance reduces the influence of the lead-in cable on sensitivity at the expense of overall sensitivity.

If a swamping inductance L_T is used in the electronics unit, then

$$t^p = 2\frac{N_{pt}}{f_c S_D^p \dfrac{1}{1+\dfrac{L_T}{L_D}}} = \frac{2N_{pt}\left[1+1+\dfrac{L_T}{L_D}\right]}{f_c S_D^p}. \tag{2-58}$$

For example, let

N_{pt} = 4

L_T = 150 μH

L_D = 75 μH

f_D = 2.22 MHz

S_F^p = 0.005%.

Then

$$t^p = \frac{2\times 4\left[1+\dfrac{160\ \mu\mathrm{H}}{75\ \mu\mathrm{H}}\right]}{(2.22\times 10^6\ \mathrm{Hz})\times(5\times 10^{-5})} = 216\ \mathrm{ms}. \tag{2-59}$$

The percentage error in vehicle speed obtained from a speed trap using two inductive loops a known distance apart is given by

$$\frac{\Delta S}{S} = 100\times\frac{\Delta T\times S}{X} \tag{2-60}$$

where

$\Delta S/S$ = Error in vehicle speed, percent

ΔT = Error in measured time, seconds

X = Distance between leading edges of the loop, distance units

S = Vehicle speed, distance units/second.

The maximum measured time error in vehicle speed or occupancy measurements is attributed to the response time of the electronics unit. The speed measurement error caused by finite response times is illustrated by the following example.

Let

$\Delta S/S$ = Unknown error in vehicle speed, percent

ΔT = 2 x 216 milliseconds (ms) = 432 ms (0.432-second error in measured time)

X = 100-ft (30.5-m) spacing between leading edges of the loop

S = 60 miles per hour (mi/h) = 88 ft/s (96.6 kilometers per hour (km/hr) = 2.68 m/s) vehicle speed.

Then

$$\frac{\Delta S}{S} = 100 \times \frac{0.432 \times 2.68}{30.5} = 0.38 = 38 \text{ percent} . \qquad (2\text{-}61)$$

This example indicates that the loop system should be designed so that the system sensitivity is as large as possible. By setting the electronics unit to a less sensitive range, the response time is decreased, producing a more accurate vehicle speed measurement.

An increase in the electronics unit's clock frequency from 2.22 MHz to 22.2 MHz reduces the percentage velocity error from 38 percent to 3.8 percent. Many of the newer electronics units use clock frequencies between 20 and 25 MHz and, thus, are capable of reducing the percent velocity error.

DIGITAL RATIOED PERIOD SHIFT ELECTRONICS UNIT

The digital processor in this design compares counts proportional to the oscillator period when a vehicle is present to a reference count taken periodically when no vehicle is present. The reference count is stored in memory. When the count during detection is less than the reference count by a preset sensitivity threshold count, a vehicle call is initiated. The ratioed period shift electronics unit differs from the period shift electronics unit in that the threshold count N_{pt} is not fixed.

The threshold count (see Appendix I) is given by

$$S_D^p = \frac{2S_T N_{pc}}{N_{pc}} = 2S_T . \qquad (2\text{-}62)$$

Since the sensitivity of the electronics unit is independent of the count measured by the period counter, the sensitivity is also independent of frequency. The response time is identical to that of the digital period shift electronics unit.

COMPARISON OF DIGITAL ELECTRONICS UNITS

Table 2-9 compares the various inductive-loop digital electronics unit concepts in terms of sensitivity and response time.

Table 2-9. Comparison of sensitivities and response times of digital electronics units.

Digital electronics unit type	Sensitivity (S_D)	Response time	Interpretation
Frequency shift	$S_D^f = K_f \sqrt{L_D C_D}$	t^f = frame time	Sensitivity decreases when the inductance increases from the use of a larger loop or longer lead-in cable. Response time is fixed by the frame time of the electronics unit.
Ratioed frequency shift	$S_D^f = \frac{2N_{ft}}{N_{fc}}$	$t^f = 2\frac{N_{ft}}{mf_D S_D^f}$	Sensitivity is independent of the inductance value at the input terminals to the electronics unit. Response time increases directly with an increase in inductance at the input terminals to the electronics unit and inversely with sensitivity of the unit.
Period shift	$S_D^p = \frac{K_p}{\sqrt{L_D C_D}}$	$t^p = \frac{2N_{pt}\left[1+\frac{L_T}{L_D}\right]}{f_c S_D^p}$	Sensitivity is increased by increase in size of loop and longer lead-in cable. However, the effect of a change in inductance is mitigated by a swamping inductance from a loosely coupled transformer. Response time is dependent on loop inductance.
Ratioed period shift	$S_D^p = 2S_T$	$t^p = \frac{2N_{pt}}{S_D^p}$	Sensitivity is independent of inductance value at the input terminals to the electronics unit. Response time is dependent on loop inductance.

MULTICHANNEL DIGITAL MODELS

Controller cabinet space can be conserved if the electronics unit can operate more than one loop. Most digital electronics unit manufacturers offer products that can operate four or more loops. Some models address the crosstalk problem by providing a frequency-separation switch, while others separate the loops by a time-division scanning process.

One manufacturer's scanning electronics unit energizes and analyzes each of four or more channels sequentially up to 100 times per second. The digital period shift electronics unit is inherently fast enough to permit scanning. The time to analyze a channel depends on the sensitivity desired, as high precision in thresholding demands more time for counting reference pulses.

For example, if three 150 μH loops were connected and used with sensitivities of 1, 4, and 6, and the fourth channel was switched to off, then the four channels would require 2.3, 9, 63, and 0.9 ms respectively, for a total of 75 ms. Each channel would be energized and analyzed 1,000 ÷ 75 = 13 times per second. The use of loops with larger inductance values reduces the scan rate, as does the selection of the highest sensitivity settings on the unit. If more than four similar and nearby loops are involved, the frequency switch or the size and/or number of turns in the loops can be varied to provide crosstalk protection. Equation 2-40 and Tables 2-2, 2-3, and 2-4 can be used to design frequency separations of 7 percent or more.

Manufacturers utilize higher clock speeds to provide faster scanning rates. For example, on the lower sensitivity settings, the sample time is 0.5 ms per channel. Therefore, the total time to scan all four channels is 2 ms. When a channel is switched off, the scan time for that channel is zero.

ADDED FEATURES FOR DIGITAL ELECTRONICS UNITS

During the l980s, several refinements were incorporated into inductive-loop detector digital electronics units. Recognizing the heavy demand on maintenance dollars, some manufacturers added circuitry that reduces the frequency of trouble calls to reset units attached to faulty loops. These features, intended to reduce maintenance costs and maximize traffic performance, include open loop test, automatic reset, and remote reset, as discussed below.

Open Loop Test

This feature allows the electronics unit to continue to operate an intermittently open loop system. A momentary open caused by a broken wire, poor splice, or loose connection will be stored in memory. If the connection remakes, the unit will promptly retune and continue to operate properly. If the open continues, it will result in a constant call.

When making a service call to the intersection, the technician may observe an indicator light that will flash a distinct pattern if an open has occurred. With other brands of electronics units, the technician presses the "Open Loop Test" button to determine whether an opening has occurred since the last service call. The open loop memory can be queried repeatedly as it can be reset only by power interruptions (such as removing the module from its card rack and reinserting) or by pressing the common reset button on the electronics unit. This constitutes a system reset, which will clear the open loop memory.

Automatic Reset

Some electronics units can be programmed to generate an internal reset if a call (i.e., the output of the electronics unit) exceeds the programmed time. The reset is controlled by the termination of the associated phase green. One agency claims that this feature reduced their electronics unit maintenance costs by 42 percent.

Remote Reset

Remote reset allows automatic investigation of suspicious calls generated by computer or software-program control systems. A remote master monitoring the actuations of each system sensor may suspect that an electronics unit is malfunctioning. By asserting the Reset command, the unit can frequently be returned to normal operation. The reset causes presence calls to be cleared, but it does not clear open loop memory nor does it prevent an open loop from calling.

If the reset fails to restore normal operation, the fault can be recognized and printed out for maintenance attention. An open loop that is constantly calling can be taken off line so that it does not falsely influence system operating parameters.

INDEPENDENT LOOP FAIL OUTPUT

In addition to the normal output of the electronics unit, a second output for loop status is provided on some models. Whenever the loop inductance undergoes a step change of ± 25 percent or more, the loop fail output is turned on. If the inductance returns to a value less than ± 25 percent of the reference, the loop fail output turns off. This enables remote interrogation of the loop status.

Other fault detection algorithms are embedded in the microprocessors found in modern controllers. These algorithms output digital codes that identify the fault type to a controller, which transmits the information to a central location.

VEHICLE CLASSIFICATION

Newer inductive-loop detector electronics units and loop configurations are capable of vehicle classification. The electronics module shown in Figure 2-28 uses artificial neural network software to classify the traffic stream into the 23 categories depicted in Figure 2-29. The first 13 are the standard FHWA classes, while the remaining ones represent vehicles with unique characteristics.(4)

Figure 2-28. Model S-1500 inductive-loop vehicle classifier and speed sensor (Photograph courtesy of Reno A&E, Reno, NV).

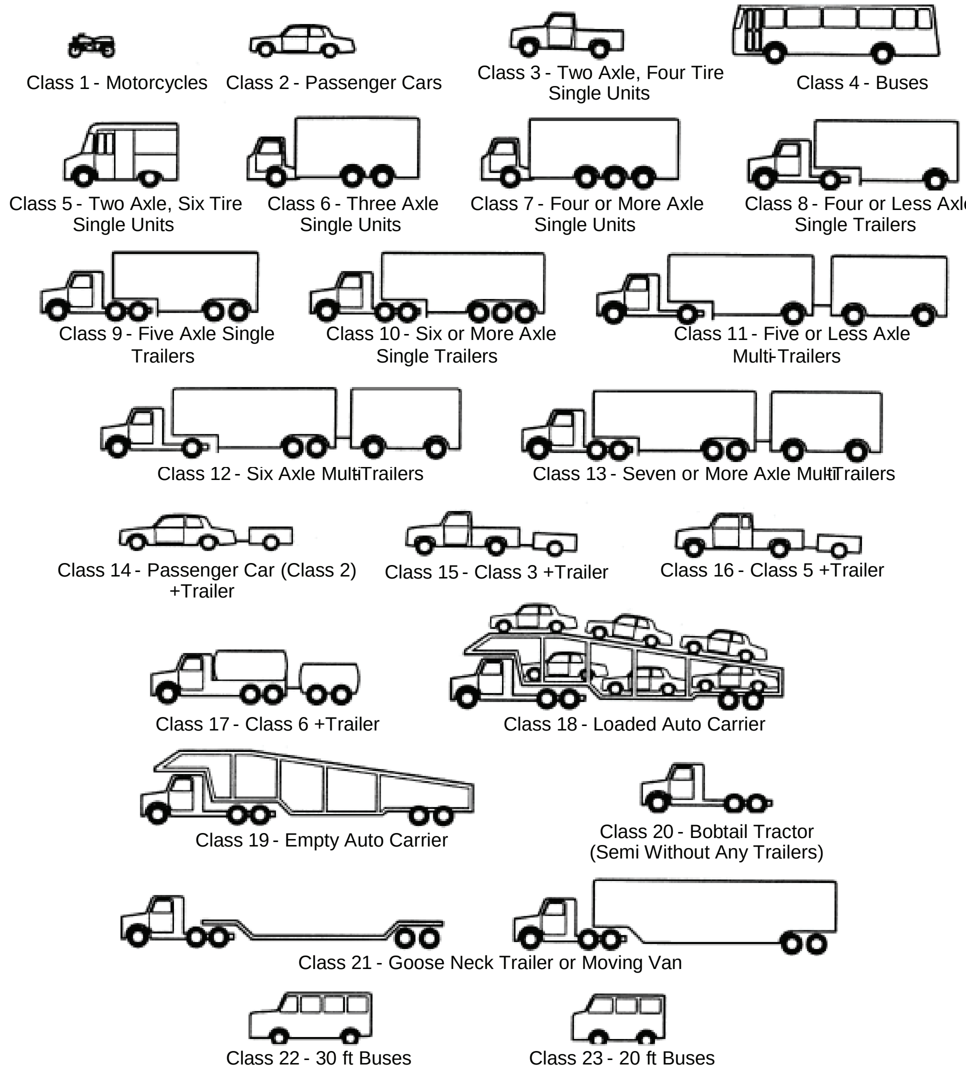


Figure 2-29. Classes available from inductive-loop classifying sensor (Source: *Model IVS-2000 Installation and Operation Manual*, Rev. 1.53 (Intersection Development Corp, Downey, CA, Sept. 1997)).

Special configurations of inductive loops have been developed to detect axles and their relative position in a vehicle. Such systems are used at toll plazas to elicit the correct payment for the vehicle class. In the application shown in Figure 2-30, an axle loop array is situated between two main loops. Axle presence is detected by the axle loop array. The relative position of the axles in the vehicle is determined from the signatures provided by the main loops. The data obtained are vehicle length, speed, acceleration, vehicle type, number of axles, and axle separation. Profile information can also be obtained to refine and validate classification in ambiguous cases. This electronics unit, as well as the one shown in Figure 2-28, can be used to identify transit buses and provide priority treatment at traffic signals.

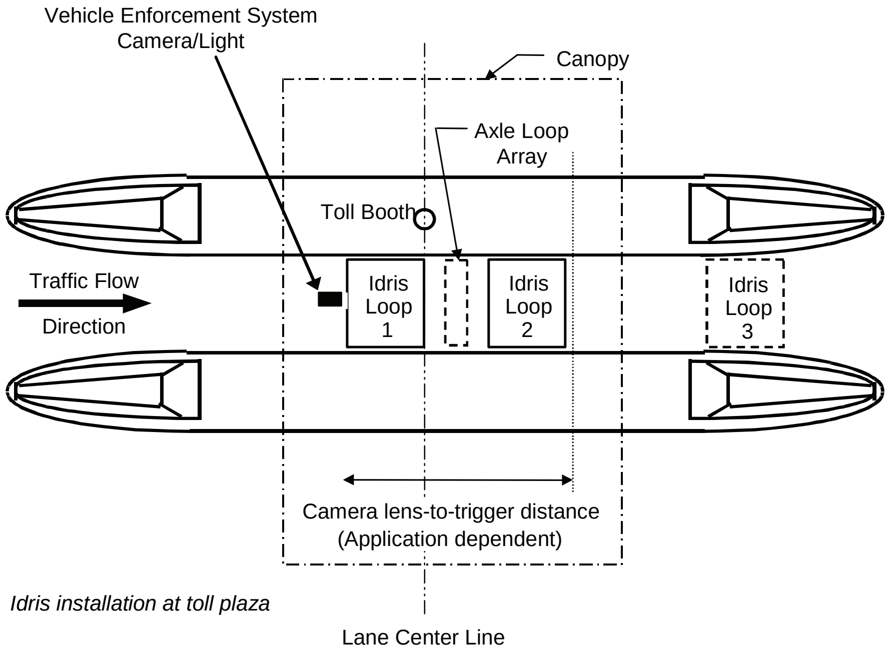


Figure 2-30. Axle location and vehicle classification using an array of inductive loops (Drawing courtesy of Peek Traffic, Inc.–Sarasota, Sarasota, FL).

NEMA STANDARDS

As with traffic signal controllers, loop detector electronics units were developed and marketed by numerous manufacturers, each using a different type of harness connector and detection technique. To overcome subsequent interchangeability problems, NEMA developed a set of standards known as “Section 7. Inductive-Loop Detectors.” These were released early in 1981. This section of the NEMA Standards defined functional standards, physical standards, environmental requirements, and interface requirements for several inductive-loop electronics unit configurations.

Section 7 described only the basic functions associated with inductive-loop detector electronics units. Users identified the need for additional functions for specific locations, particularly delay and extension timing. To cover this gap, NEMA developed and in 1983 released “Section 11. Inductive-Loop Detectors with Delay and Extension Timing.” This section was basically identical to Section 7 with the addition of requirements for the timing of delayed call and extended call features. A further revision resulted in a new Section 15, which was released February 5, 1987 (a reproduction is provided in Appendix J). This new standard combines, updates, and supersedes Sections 7 and 11.

NEMA ELECTRONICS UNIT CONFIGURATIONS

The NEMA Standards define two basic types of electronics unit configurations: shelf mounted and card-rack mounted. Shelf mounted units are commonly used in NEMA controllers and are available in both single-channel and multichannel (two- or four-channel) configurations. Figure 2-31 shows an example of a shelf-mounted unit, which is powered by the 120-volt AC supply in the cabinet. Outputs are generated by electromechanical relays or by electrically isolated solid-state circuits. Physical dimensions and connector requirements are included in the NEMA Standards in Appendix J.

Card-rack mounted electronics units, illustrated in Figures 2-32 and 2-33, fit into a multiple card rack and operate with external 24-volt DC power generated in the rack assembly or elsewhere in the controller cabinet. These devices are an effective way to reduce cabinet space requirements where large numbers of inductive-loop detector electronics units are needed.

OUTPUT TYPES

Electronics units have one of two output types: relay or solid-state optically isolated. Relay outputs use electromechanical relays to generate a circuit closure and generate a detection call to the controller. Solid-state outputs have no moving parts and are, therefore, generally more reliable and more accurate in tracking the presence of vehicles. This factor can be important in some traffic signal timing operations.

The relay outputs are designed to fail "on" (contacts closed) when power to the electronics unit is interrupted. The solid-state output fails "off" (nonconducting) in the same circumstances. Therefore, a relay output may be more desirable for use with intersection actuation because a constant-call would be safer than a no-call situation. A solid-state output is more desirable where accurate presence detection is desired.

NEMA electronics units are free-standing shelf modules. They are connected to a NEMA-style controller using a cable that attaches to circular connectors on the electronics unit and the controller.

Figure 2-31. Shelf-mounted NEMA electronics unit.

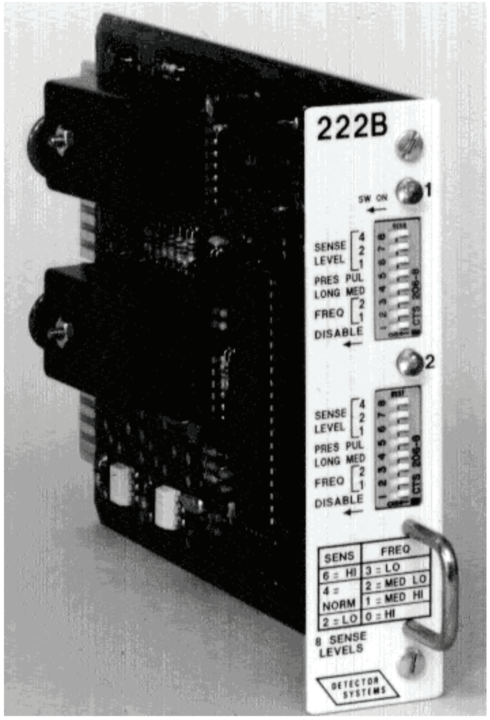


Figure 2-32. Two-channel card rack mounted electronics unit.

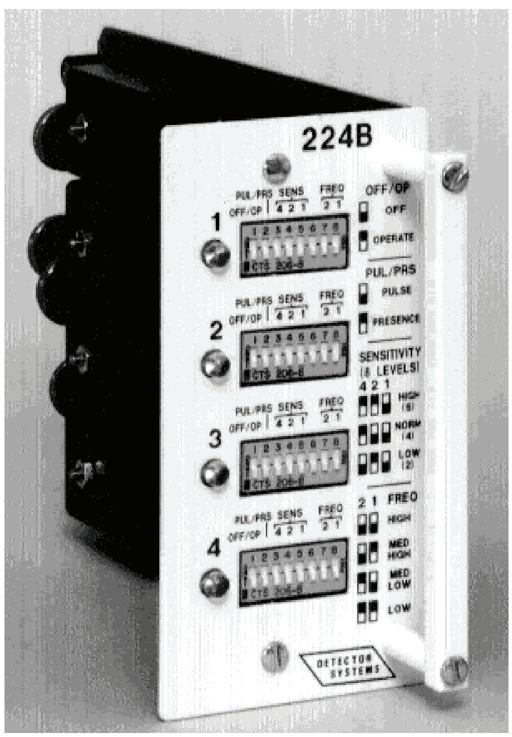


Figure 2-33. Four-channel card rack mounted electronics unit.

Card-rack mounted electronics units contain an edge connector. The card slides into the rack cage where the male edge connector fits snuggly into the female connector in the card cage.

PRESENCE AND PULSE MODES OF OPERATION

Two modes of operation are selectable for each electronics unit channel: presence and pulse. The presence mode of operation provides a constant output while a vehicle is over the loop detection area.

NEMA Standards require the electronics unit to sustain a presence output for a minimum of 3 minutes before tuning out the vehicle. Most units will maintain the call for periods up to 10 minutes. This mode is typically used with long loop installations on intersection approaches with the controller in the nonlocking detection memory mode.

Nonlocking detection memory is a controller function whereby the controller retains a call (vehicle detection) only as long as the presence inductive loop is occupied (vehicles are passing over or stopped on the loop). The controller drops the call if the calling vehicle leaves the detection area.

The pulse operating mode generates a short pulse (between 100 and 150 ms) each time a vehicle enters the loop detection area. Pulse operation is typically used when inductive-loop detectors are located well upstream of the intersection with the controller in the locking detection mode. That is, the controller does not drop the vehicle call when the calling vehicle leaves the detection area. Chapter 4 contains additional information about applications of locking and nonlocking controller operation.

CROSSTALK

When two loops constructed of the same wire diameter have the same loop dimensions, number of turns, and lead-in length, they have the same resonant frequency. When these two loops are near each other or when the lead-ins from these loops are in close proximity (perhaps running in the same conduit), a phenomenon known as “crosstalk” can occur. This effect is caused by an electrical coupling between the two loop channels and will often manifest itself as brief, false, or erratic actuations when no vehicles are present.

NEMA Standards require inherent, automatic, or manual techniques to be utilized to prevent crosstalk. The most common feature is a frequency selection switch that varies the operating frequency of the adjacent loop channels.

TIMING FEATURES

As contained in Section 15 of the NEMA standards (Appendix J), the timing features include delay and extension timing. Delay timing can be set from 0 to 30 seconds, indicating the time that the electronics unit waits, from the start of the continuous presence of a vehicle until an output begins, as shown in Figure 2-34. The output terminates when the vehicle leaves the detection area. If the vehicle leaves the detection area before the delay time has expired, no output is generated. Extension time, shown in Figure 2-35, defines the amount of time the output is extended after the vehicle leaves the detection area and can be set from 0 to 15 seconds.

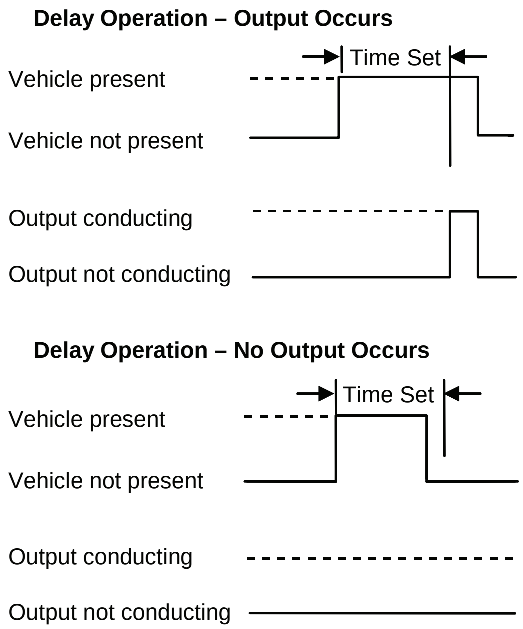


Figure 2-34. Delay operation.

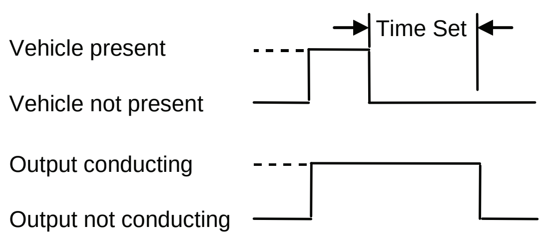


Figure 2-35. Extension operation.

Timing features can be controlled by external inputs to the electronics unit. For units with relay outputs, a delay/extension inhibit input is provided and requires 110 volts to activate. Electronics units with solid-state outputs have a delay/extension enable input, which requires a low-state DC voltage (0 to 8 volts).

A typical delayed call installation might be a semiactuated intersection with heavy right-turns-on-red from the side street, which has long loop presence detection. Using a relay output electronics unit, the side street green field output (110 volts AC) is connected to the delay/extension inhibit input. Thus, a delay is timed, allowing right-turns-on-red to be made without unnecessarily calling the controller to the side street. (Heavy right-turn movements will bring up the green anyway, as the loop will be occupied by following vehicles.) However, when the side street has the green, the delay is inhibited, permitting normal extensions of the green.

Conversely, extended call detectors could be used on high-speed approaches to an intersection operated by a basic (non-volume-density) actuated controller. Using this technique, the apparent zone of detection is extended, and different "gap" and "passage" times are created without the volume-density controls (this does not, however, replace volume-density functions). The delay/extension enable input on a solid-state output unit could be tied to the controller's "Phase On" output.[5]

TYPE 170 CONTROLLER SPECIFICATION

Simultaneously with the evolution of the NEMA Standards, the States of California and New York developed the Type 170 controller specification.[6,7] As with the NEMA Standards, the development of this new controller was a direct response to the problems of noninterchangeability. The Type 170 system was to be interchangeable between all manufacturers supplying equipment for either state.

Unlike the NEMA Standards, which standardize functions, the Type 170 Specification standardizes hardware. The Type 170 system stipulates cabinet, controller, and all required accessories including sensors. A 170 controller is shown in Figure 2-36. The California 170 system, shown in Figure 2-37, specifies a large, base-mounted cabinet with full component layout for 28 two-channel electronics units. The New York system features a smaller, pole-mounted cabinet with 14 two-channel electronics units. New York subsequently revised its specification to require a new microprocessor. This system, denoted as a Type 179, includes the same cabinet and electronics units as the original Type 170.

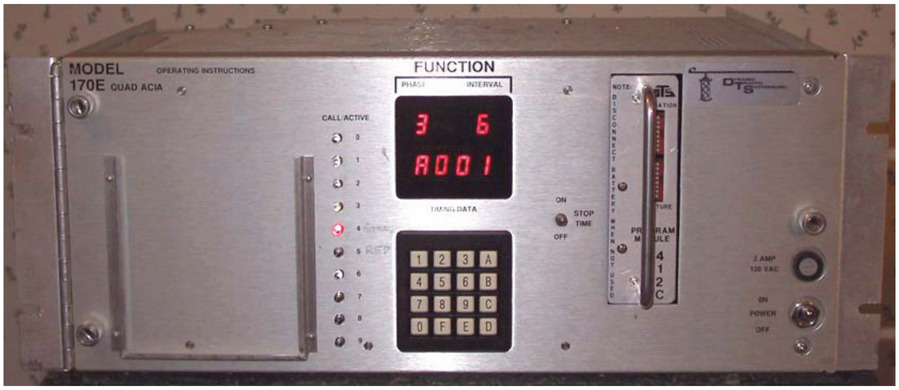


Figure 2-36. Model 170 controller (Photograph courtesy of Lawrence A. Klein).

All 170-system modules are shelf or card-rack mounted as the controllers are housed in standardized cabinets. The electronics units are mounted on an edge-connected, printed circuit board. One card slot in the cabinet will support two detector input channels. Cards supporting four detector input channels require two card slots. The electronics unit's front panel has a hand pull for insertion and removal from the input file.

2070 ADVANCED TRANSPORTATION CONTROLLER SPECIFICATION

An Institute of Transportation Engineers (ITE) committee composed of members from industry, Federal Highway Administration, various States and cities, and NEMA developed the concept for the advanced transportation controller (ATC). The development of a functional standard was subsequently assigned to a committee composed of affiliates from the American Association of State and Highway Transportation Officials (AASHTO), ITE, and NEMA. This group was charged with completing a consensus-based standard.[8]

The advanced transportation controller is designed to control or process data from one or more roadside devices. It operates as a general-purpose computer with a real-time operating system. One or more applications are stored in a FLASH drive. The application software is loaded and launched from the FLASH drive into dynamic random access memory (DRAM). The open architecture found in the ATC allows hardware modules and software to be purchased from a variety of vendors. A future central processor unit (CPU) option will support a daughter-board CPU that incorporates an applications program interface (API), allowing use of different microprocessors and real-time operating systems.

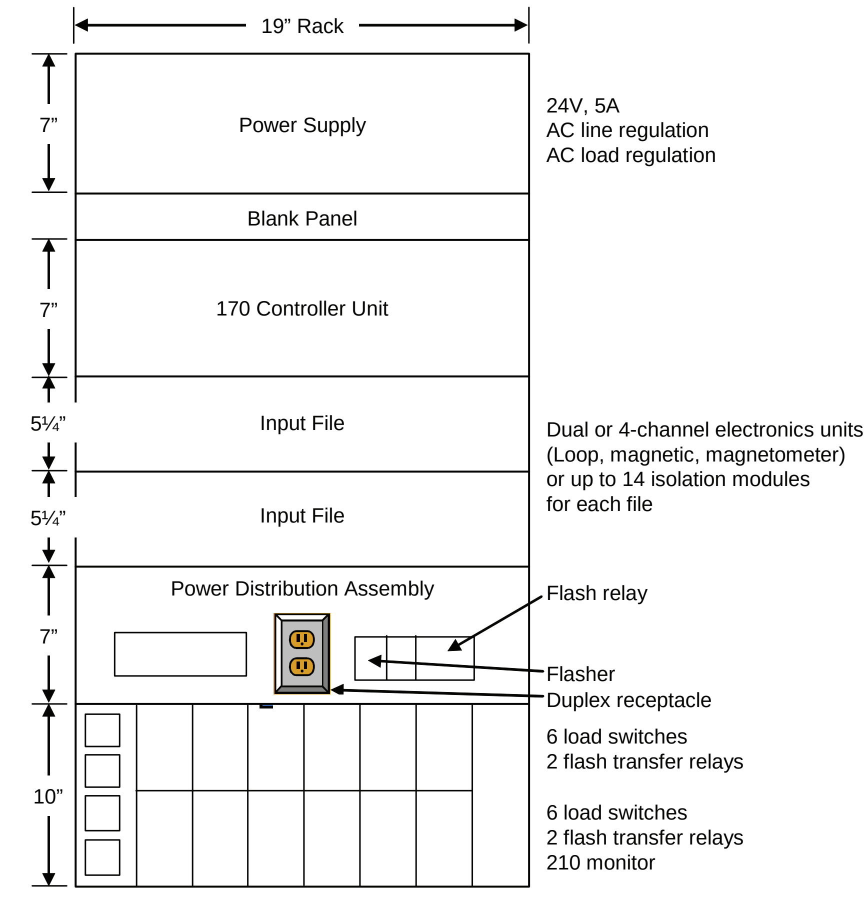


Figure 2-37. Type 170 cabinet layout (California).

The California Department of Transportation (Caltrans) has continued development of the ATC and issued standards for the development of a Model 2070 ATC.[9]

The 2070 controller contains eight module types as follows:

- Central processor unit (CPU).
- Field input/output (I/O) module.
- Front panel assembly (FPA).
- Power supply.
- Chassis.
- Modem.
- NEMA interface.
- Back cover.

Figure 2-38 shows a 2070 ATC controller. An optional NEMA module may be mounted below the main chassis. Table 2-10 describes the options available for each of the modules.[9]

Figure 2-38. Model 2070 controller. The 2070 ATC controller is manufactured by Econolite, Eagle, Naztec, Safetran, GDI Communications, and McCain Traffic Supply at present.

Table 2-10. 2070 ATC module options.

Module	Option	Description
CPU	1A	Two-board CPU using Motorola 68360 microprocessor and MicroWare OS-9 operating system or PTS real-time Linux operating system.
	1B	One-board CPU using Motorola 68360 microprocessor and MicroWare OS-9 operating system or PTS real-time Linux operating system.
	1C	Daughter-board CPU using applications program interface (API) to allow use of different processors and real-time operating systems.
Field I/O module		Isolates the EIA-485 internal signal line voltages between the CPU module and external equipment, and the drive and receive interfaces on external equipment.
	2A	170 cabinet module: Field controller unit (FCU); 64-bit parallel I/O ports; other circuit functions (60-Hz square wave line synchronization signal, jumper to activate monitoring of the watchdog timer input signal (used with Model 210 cabinet monitor unit only), watchdog timer, 1-kHz reference signal, 32-bit millisecond counter, loss of communications message, logic switch to disconnect C12S connector upon command); EIA-485 serial communications; C1S, C11S, and C12S connectors; +12Vdc to +5Vdc power supply; any required software. Directly accommodates 64 inputs and 64 outputs.
	2B	2070 cabinet module: EIA-485 serial communications, DC power supply, C12S connector. Current serial bus #1 address scheme accommodates a maximum of 120 inputs and 42 outputs.

Table 2-10. 2070 ATC module options—*Continued*

Module	Option	Description
Front panel assembly	3A	FPA controller that interfaces with the CPU, 16-key pad for hexadecimal alphanumeric entry, 12-key pad for cursor control and action symbol entry, liquid crystal backlighted display of 4 lines with 40 characters per line, minimum character dimensions of 0.26-inch- (5.00-mm-)wide by 0.4- inch- (10.44-mm)-high, auxiliary switch, alarm bell.
	3B	FPA controller that interfaces with the CPU, 16-key pad for hexadecimal alphanumeric entry, 12-key pad for cursor control and action symbol entry, liquid crystal backlighted display of 8 lines with 40 characters per line, minimum character dimensions of 2.65 mm wide by 4.24 mm high, auxiliary switch, alarm bell.
	3C	No display, system serial port.
Power supply	4A	+5VDC (±5%) at 10 A, ±12VDC (±8%) at 0.5 A, +12VDC (±8%) at 1.0 A.
	4B	+5VDC (±5%) at 3.5 A, ±12VDC (±8%) at 0.5 A, +12VDC (±8%) at 1.0 A.
Chassis	5A	Contains VME cage assembly compatible with 3U boards, metal housing, serial motherboard, backplane, power supply module supports, slot card guides, wiring harness, cover plates.
	5B	Contains two-board CPU (option 1A) mounting assembly, metal housing, serial motherboard, backplane, power supply module supports, slot card guides, wiring harness, cover plates.
Modem	6A	Two channels of 300 to 1,200 baud asynchronous EIA-232 serial communications.
	6B	Two channels of 0 to 9,600 baud asynchronous EIA-232 serial communications.
Modem	7A	Two channels of 300 to 1,200 baud asynchronous EIA-485 serial communications.
	7B	Two channels of 0 to 9,600 baud asynchronous EIA-485 serial communications.
NEMA interface	8	Contains four NEMA connectors and EIA-485 serial communications to interface with the 2B field I/O module. Accommodates 118 inputs and 102 outputs.
Back cover	9	Protects the interface harness connecting the NEMA interface module with the 2070 controller chassis.

Five Intelligent Transportation System (ITS) cabinet styles are manufactured to house the 2070 ATC controller and associated electronics, which includes the cabinet monitoring system, power distribution system, control and data communications systems, inductive-loop detector electronics assemblies, batteries, and power supplies. The ITS cabinets incorporate three serial buses. The first bus supports command and response communications needed for real-time control of the modules in the cabinet. A serial interface unit (SIU) can be connected to serial bus #1 to accommodate 54 inputs or outputs, including detector calls or status reports. Alternatively, each SIU can support up to 14 load switches. Serial bus #1 also sends polling commands to the emergency cabinet monitor unit (CMU) located in the power distribution assembly. The function of the CMU is to query various cabinet conditions, such as power supply voltages and line synchronization signal, and if warranted, transfer control from the ATC to a safe control mode. The second bus supports command and response communications between the 2070 controller and the controlled field devices. Serial bus #3 supports the cabinet emergency system. This system senses and monitors load outputs, various operational functions, and the bus controls and communications. When an abnormal operational condition is encountered, the cabinet emergency system controls cabinet emergency actions, through the CMU, and reports the condition to the ATC.

An advantage of serial communications between the controller and the field devices is the simplified wiring and connector utilized in the ITS cabinets as contrasted with the 104-pin C1S parallel connector found on the 170 controller or the three NEMA connectors on TS-1 controllers. The C1S connection may still be made in the 2070 using the C1S connector found on the front panel of the 2070–2 field input/output (I/O) module. Any required NEMA connections may be made through the optional NEMA module, which is mounted below the 2070.

The cabinets, constructed of aluminum sheet, contain front and rear doors. Designed to be rainproof, they are ventilated using a thermostatically controlled fan. Table 2-11 describes the cabinet styles recommended for traffic signal control and traffic management applications.

Table 2-11. 2070 ATC ITS cabinet options.

Application	Model	Description
Traffic signal control[1]	340	4-door cabinet with P-base ground mount
	342	2-door cabinet with 170-base ground mount
	346	2-door cabinet with 170-base adapter mount
Traffic management[1] (e.g., ramp metering)	354	2-door cabinet with 170-base ground mount
	356	2-door cabinet with 170-base adapter mount

1. The cabinets utilized for traffic signal control and traffic management differ in the types and numbers of cages, power supplies, power distribution and output assemblies, and serial bus harnesses.

Five models of the 2070 ATC have been designated to incorporate the VME cage and cabinet options as shown in Table 2-12.

Table 2-12. 2070 ATC models and cabinet options.

Model number	Description
2070 V	2070 containing a VME cage and mounted in a 170 or TS2 cabinet
2070 N	2070 containing a VME cage and mounted in a TS2 cabinet
2070 L	2070 without a VME cage and mounted in a 170 or TS2 cabinet
2070 LC	2070 without a VME cage and mounted in an ITS serial communications cabinet
2070 LCN	2070 without a VME cage and mounted in a TS2 cabinet

Table 2-13 relates the 2070 ATC module options of Table 2-10 to the model and cabinet options of Table 2-12.

Table 2-13. 2070 ATC modules corresponding to model and cabinet options.

Module designator	Module description	Model 2070 V	Model 2070 N	Model 2070 L	Model 2070 LC	Model 2070 LCN
—	Unit chassis	1	1	1	1	1
2070-1A	Two-board CPU	1	1	–	–	–
2070-1B	One-board CPU	–	–	1	1 or	1 or
2070-1C	Daughter-board CPU	–	–	–	1	1
2070-2A	Field I/O for 170 cabinet	1	–	1	–	–
2070-2B	Field I/O for ITS and NEMA cabinets	–	1	–	1 or none	1 or none
2070-3A	Front panel 4-line display	1	1	–	–	–
2070-3B	Front panel 8-line display	–	–	1	–	–
2070-3C	No front panel display	–	–	–	1	1
2070-4A	+5 VDC (±5%) at 10 A power supply	1	1	1 or	1 or	1 or
2070-4B	+5 VDC (±5%) at 3.5 A power supply	–	–	1	1	1
2070-5A	VME cage assembly	1	1	–	–	–
2070-5B	Two-board CPU mounting assembly	1	1	–	–	–
2070-8	NEMA interface	–	1	–	–	1
2070-9	2070 N back cover	–	1	–	–	1

MODEL 222/224 INDUCTIVE-LOOP DETECTOR ELECTRONICS UNITS

Chapter 5 of the California Department of Transportation (Caltrans) *Transportation Electrical Equipment Specifications* (TEES), (reproduced in Appendix K) provides a general description, functional requirements, and electrical requirements for the Model 222 Two-Channel Loop Detector Electronics Unit and Model 224 Four-Channel Loop Detector Electronics Unit. These were illustrated in Figures 2-32 and 2-33.

Each channel in an electronics unit has panel-selectable sensitivity settings for presence and pulse operation. As with the NEMA standards, the TEES requires some mechanism to prevent crosstalk with other modules. It also requires that the selected channel not detect moving or stopped vehicles at distances of 3 ft (1 m) or more from any loop perimeter. The timing features are incorporated into the various Type 170 software programs such as the Caltrans Local Intersection Program (LIP).

SELECTION FACTORS

Several operational characteristics influence the selection of an electronics unit model. These characteristics depend on the application and its requirements.

Tuning Range

NEMA requires that an inductive-loop detector electronics unit be capable of tuning and operating as specified over a range of inductance from 50 μH to 700 μH. For most applications, this range is adequate. Some units, however, are capable of tuning and operating over a range of 1 μH to 2,000 μH. This larger operating range permits extra long lead-in cables and/or several loops to be connected in series to one unit.

Response Time

The time required for an electronics unit to respond to the arrival and departure of a vehicle is critical when the output is used to calculate speed and occupancy. Systems for the control of surface street traffic signals and monitoring of freeway conditions usually perform these calculations.

If the time for vehicle call pick-up is close to the time for drop-out, little or no bias in the vehicle occupancy time is introduced. If there is a significant difference, but the difference is about the same from unit to unit, a correction for the bias is easy to apply.

NEMA specifies that an electronics unit respond to the arrival or departure of a small motorcycle into and out of a 6- x 6-ft (1.8- x 1.8-m) loop within 125 ms. An automobile call must be initiated or terminated within 50 ms. NEMA also states that shorter response times might be required for specific surveillance applications that involve vehicle speeds in excess of 45 mi/h (72 km/h).

Modern traffic management systems often demand small response times. Whereas speed and occupancy measurements averaged over many vehicles to a marginal accuracy were adequate in the past, current and future applications require faster response times with greater traffic flow parameter measurement accuracies. Many of the manufacturers are responding to this need by supplying new models of electronics units with enhanced capabilities.

Recovery from Sustained Occupancy

Recovery time can become critical when the electronics unit is operated in the presence mode with a long loop, say 6- x 50-ft (1.8- x 15-m), at the stop line, or four 6- x 6-ft (1.8- x 1.8-m) loops in a left-turn lane wired in a combination of series-parallel. NEMA requires that after a sustained occupancy of 5 minutes by any of the three test vehicles, the electronics unit shall recover to normal operation with at least 90 percent of the minimum specified sensitivity within 1 second after the zone of detection is vacated. If an electronics unit does not recover quickly enough, the next vehicle may not be detected at all and will be trapped until a new vehicle arrives on the loop.

Loss of Detection During Saturated Flow

During peak periods, a long loop or combination of loops may be held in detection without a break for an hour or more. In these situations, the electronics unit must continue outputting for at least an hour without dropping the detection because of an environmental tracking feature or other design defect. NEMA does not address this condition.

Sensitivity with Pavement Overlay

Depth of loop wires has traditionally been considered to be a critical factor when the pavement is overlaid. However, tests conducted in Texas suggest that with high sensitivity, proper installation, and calibration, the depth at which a loop is buried should have little effect on automobile detection. In these tests, a 6- x 6-ft (1.8- x 1.8-m), five-turn loop was buried at a depth of 18.5 inch (50 cm) and encased in a 1.5-inch (1.2-cm) PVC conduit with no filler. There was no appreciable difference in the detection of large cars between the near-surface mounted loops and the deeply buried loops.

Bicycle detection was only slightly less efficient with deeply buried loops. The difference occurred at the medium sensitivity setting. The surface loop detected the bicycle 1 ft (0.3 m) outside the loop at the medium setting, while the deeply buried loop did not. The rectangular 6- x 6-ft (1.8- x 1.8-m) loop did not detect bicycles in the center portion of either the surface mounted or deeply buried loop regardless of the electronics unit sensitivity setting.

Pulse-Mode Reset

The pulse mode of operation provides an output (100 to 150 ms) that is useful in counting vehicles. An electronics unit should be capable of resetting or rephasing correctly to avoid either overcounts or undercounts when specific conditions occur. NEMA requires only that an electronics unit produce one, and only one, output pulse for a test vehicle moving at 10 mi/h (16 km/h) in the detection zone of a 6- x 6-ft (1.8- x 1.8-m) loop.

When vehicles are counted simultaneously in two or more lanes, a separate electronics unit and loop are recommended for each lane to guarantee accuracy. This approach can be implemented for a small additional cost over the use of a single electronics unit and a wide loop.

OPERATION WITH GROUNDED OR OPEN LOOPS

Although NEMA does not address this point, some electronics units provide failsafe operation with faulty loops, i.e., grounded or open loops. The use of a loop isolation transformer permits these models to operate if the loop insulation is leaky or even shorted completely to ground at a single location.

Isolation allows “balanced to ground” operation of the loop circuit. This reduces the effect of the loop and lead distributed capacitance and, hence, thermal- and moisture-induced changes in the capacitance. The balanced operation also minimizes loop circuit coupling from lead-in cables in common conduits. If the loop breaks, this design fails in a “safe” way by holding a constant call, thus keeping traffic moving and avoiding trapping any vehicles. Such operation is, however, very inefficient.

Lightning Damage and Electrical Interference

An extensive study into the effects of lightning related electrical surge problems was conducted by the Ontario Ministry of Transportation.[10] The major conclusions of the study were:

- Most problems occurred on common grounding systems.
- Isolated grounding techniques reduced maintenance problems by almost 80 percent.
- If a system using isolated grounding is also upgraded to include both primary and secondary surge protection, maintenance problems are reduced by more than 90 percent.

NEMA requires that electronics units withstand the same power-line transients specified for controller units. The input terminals to the electronics unit must be able to withstand 3,000 volts. The primary to secondary insulation of the input (loop-side) transformer protects against “common mode” lightning voltages.

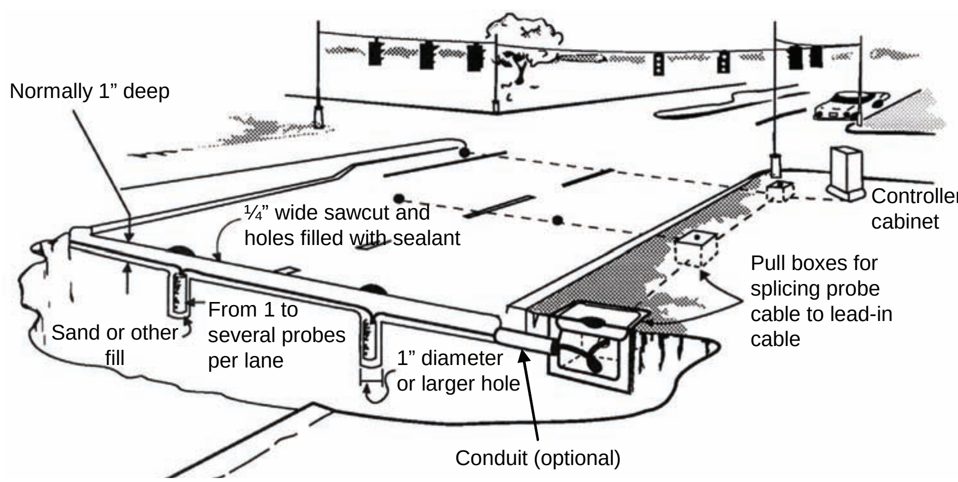
With one electronics unit model, differential lighting protection consists of a four-element protective circuit composed of the input resistors, a neon tube, transformer leakage inductance, and the diode path across the secondary. Differential lightning-induced currents are also potentially damaging. These currents are shunted through neon bulbs, limiting the voltage across the transformer.

Specifications for Model 170 controller cabinets require that lightning protection be installed within the inductive-loop detector electronics unit. The protection allows the electronics unit to withstand the discharge of a 10-microfarad capacitor charged to ±1,000 volts directly across the electronics unit input pins with no loop load present. The protection must also withstand the discharge of a 10-microfarad capacitor charged to ±2,000 volts directly across either the electronics unit input inductance pins or from either pin to earth ground. The electronics unit chassis is grounded and a dummy resistive load of 5 ohms is attached to the pins.

The Model 170 specifications also include provisions for preventing interference between channels in a given electronics unit as well as between units. The prevention techniques may be either manual or automatic.

MAGNETIC SENSORS

Two types of magnetic field sensors are used for traffic flow parameter measurement, the two-axis fluxgate magnetometer (that detects stopped and moving vehicles) and the magnetic detector, more properly referred to as an induction or search coil magnetometer (that typically detects only moving vehicles). Magnetic sensors were introduced in the l960s as an alternative to the inductive-loop detector for specific applications. A magnetic sensor is designed to detect the presence or passage of a vehicle by measuring the perturbation in the Earth’s quiescent magnetic field caused by a ferrous metal object (e.g., a vehicle) when it enters the detection zone of the sensor. An example of a magnetic sensor installation is shown in Figure 2-39.



1 inch = 2.5 cm

Figure 2-39. Magnetic sensor installation.

Early magnetic sensors were utilized to determine if a vehicle had arrived at a “point” or small-area location. They actuated controller phases that operated in the locking detection memory mode. They were also effective in counting vehicles. Modern two-axis fluxgate magnetometers are used for vehicle presence detection and counting. Unlike the inductive-loop detector, the magnetometer will usually operate on bridge decks where uncoated steel is present and cutting the deck pavement for loop installation is not permitted. The magnetometer probe and its lead-in wire tend to survive in crumbly pavements longer than ordinary loops. Another benefit is that they require fewer linear feet of sawcut.

THEORY OF OPERATION

Magnetic sensor operation is based on an Earth magnetic field model that depicts the Earth as a large bar magnet with lines of flux running from pole to pole as illustrated in Figure 2-40. A vertical axis magnetometer requires the vertical component of the Earth’s magnetic field to exceed 0.2 Oersteds. Therefore, magnetometers that contain only vertical axis sensors cannot be used near the Equator, where the magnetic field lines are horizontal. This situation is illustrated in Figure 2-41 by the cross-hatched area near the equator, which defines the region that is not suitable for vertical axis magnetometers. However, modern fluxgate magnetometers are built with both horizontal and vertical axis sensors. Therefore, they can operate anywhere on the face of the Earth.

Magnetic sensor operation is based on detecting the concentration of the Earth’s magnetic flux above and below a ferrous metal vehicle.

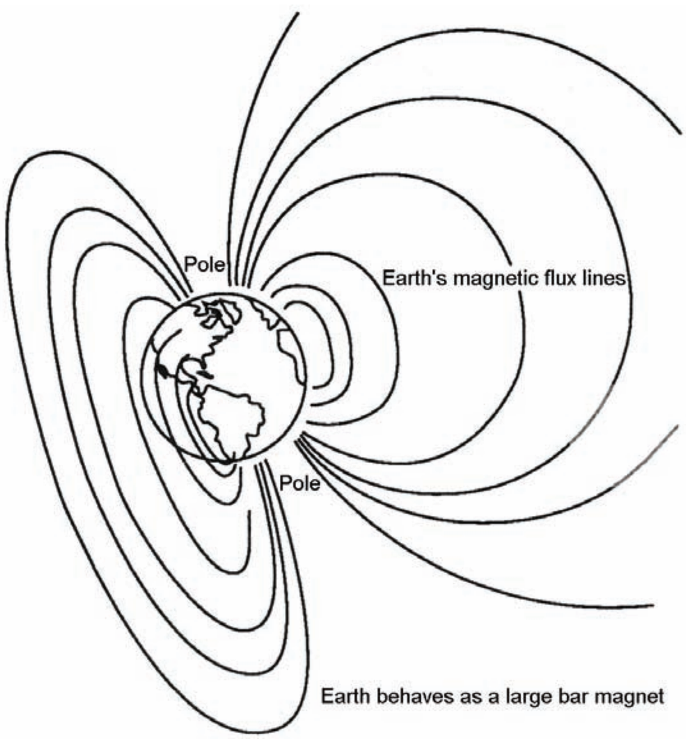


Figure 2-40. Earth’s magnetic flux lines.

Modern fluxgate magnetometers built with both horizontal and vertical axis sensors can operate in the equatorial belt region shown in Figure 2-41.

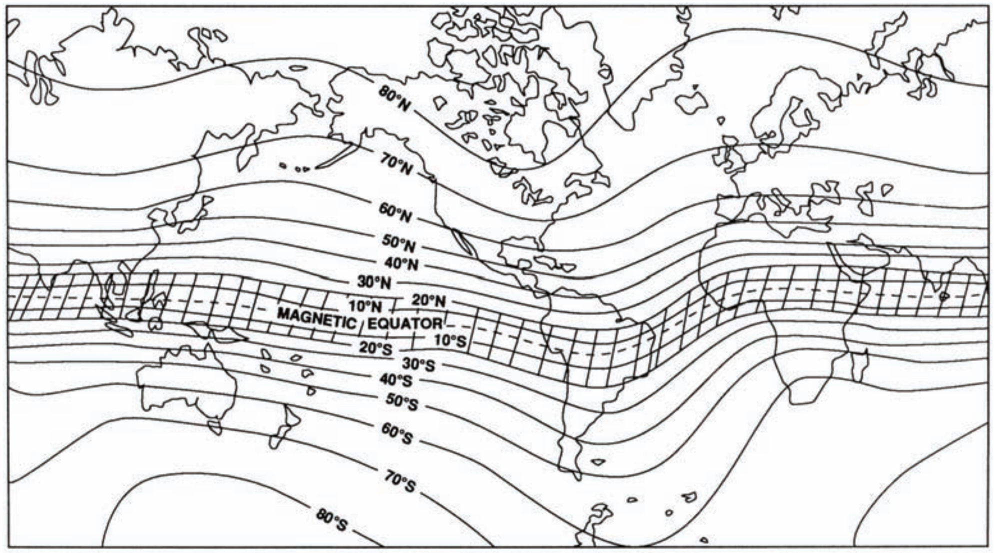


Figure 2-41. Equatorial belt showing where Earth's magnetic field is too small for deployment of simple vertical axis magnetometers between 20 degrees north and 20 degrees south of the Equator.

An iron or steel vehicle distorts the magnetic flux lines because ferrous materials are more permeable to magnetic flux than air. That is, the flux lines prefer to pass through the ferrous vehicle. As the vehicle moves along, it is always accompanied by a concentration of flux lines known as its "magnetic shadow" as illustrated in Figure 2-42. There is reduced flux to the sides of the vehicle and increased flux above and below it. A magnetometer installed within the pavement detects the increased flux below the vehicle.

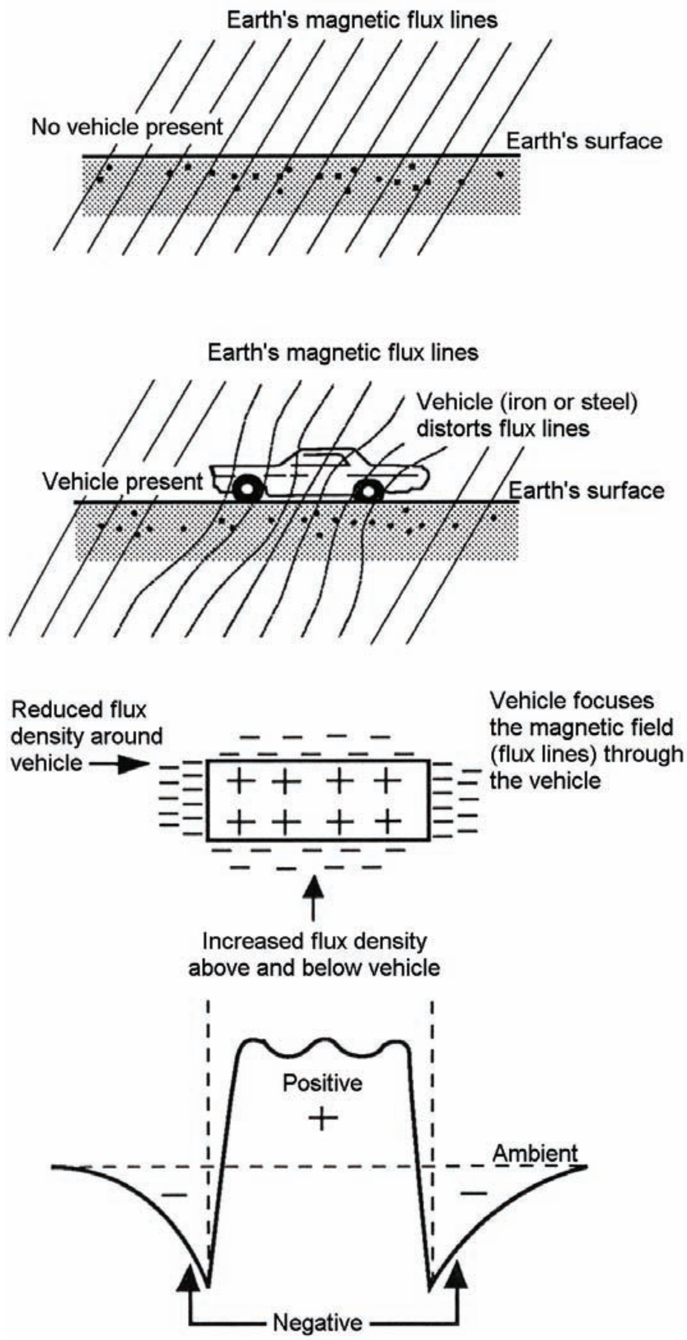


Figure 2-42. Distortion of Earth's quiescent magnetic field by a ferrous metal vehicle.

TWO-AXIS FLUXGATE MAGNETOMETERS

Two-axis fluxgate magnetometers contain sensors that detect both the vertical and horizontal components of the Earth's magnetic field and any disturbances to them. One of the secondary windings in a two-axis fluxgate magnetometer senses the vertical component of the vehicle signature, while the other, offset by 90 degrees, senses the horizontal component of the signature. The horizontal axis of the magnetometer is usually aligned with the traffic flow direction to provide in-lane presence detection and adjacent lane vehicle rejection. Fluxgate magnetometers measure the passage of a vehicle when operated in the pulse output mode. In the presence mode, they give a continuous output as long as either the horizontal or vertical signature exceeds a detection threshold.

The data provided by fluxgate magnetometers are the same as from inductive loops. Typical applications are vehicle presence detection on bridge decks and viaducts where inductive loops are disrupted by the steel support structure or weaken the existing structure, temporary installations in freeway and surface street construction zones, and signal control.

The infusion of modern digital processing and radio frequency (RF) communications technology in the area of magnetic anomaly detection have generated new designs, such as the self-powered vehicle detector (SPVD) and Groundhog magnetometers, justifying a reassessment of their supplementary role in vehicle detection. In addition, an array of magnetometers sharing a common signal processor has the potential to locate, track, and classify vehicles in a multilane scenario using a row of above-ground sensors.[11]

MAGNETOMETER SENSOR AND COMPANION ELECTRONICS UNIT DESIGN

Figure 2-43 shows the electrical configuration of the core and primary and secondary windings typically found in a magnetometer sensor. The core and windings create a small, stable transformer-like element. The two primary and two secondary windings are placed over a single strip of Permalloy core material, which is treated to produce special satur4able magnetic properties.

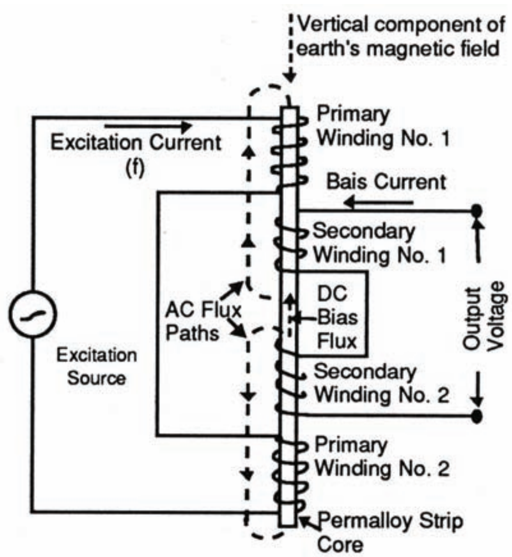


Figure 2-43. Magnetometer sensor electrical circuit (notional).

Sensor operation is based on generating a second harmonic signal from the open saturable core, which is oriented to provide maximum sensitivity to disturbances in the Earth's magnetic field. A triangular wave excitation current of suitable frequency (typically 5 kHz) is applied to the two primary windings connected in series opposition. The secondary windings, which also carry the DC bias to neutralize the quiescent Earth field, are connected in series, enhancing the supply of second-harmonic signal to the sensor's electronics unit.

A polyurethane casing is often used for abrasion resistance. It also makes the probe impervious to moisture and chemically resistant to all normal motor vehicle petroleum products. If the pavement is soft, some probes may tilt, causing them to change their orientation and lose sensitivity. A length of PVC pipe can be used to hold these probes vertical.

Figure 2-44 contains an electrical circuit diagram of a magnetometer sensor and companion electronics unit that performs signal amplification. Provided the ambient magnetic field is stable and exceeds about 16 ampere per meter (20,000 gamma), the magnetic shadow cast by a vehicle causes a local field increase of the order of 20 percent. The switch-like action of the magnetic material in the sensor induces a signal change several times this amount. This detection principle provides high sensitivity and signal-to-noise ratio.

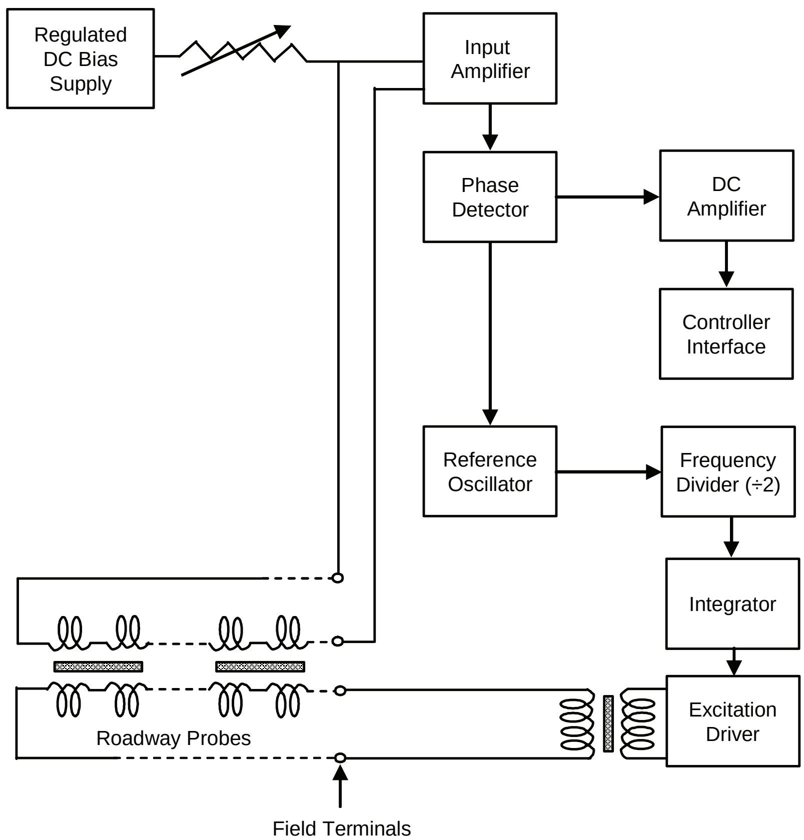


Figure 2-44. Magnetometer sensor probe and electronics unit equivalent electrical circuit.

MAGNETOMETER DETECTION SENSITIVITY

Since magnetometers are passive devices, they do not transmit an energy field. Therefore, a portion of the vehicle must pass over the sensor for it to be detected. Consequently, a magnetometer can detect two vehicles separated by a distance of 1 foot (0.3 m). This potentially makes the magnetometer as accurate as or better than the inductive-loop detector at counting vehicles.

Conversely, the magnetometer is not a good locator of the perimeter of the vehicle, with. There is an uncertainty of about ±1.5 ft (45 cm). A single magnetometer is, therefore, seldom used for determining occupancy and speed in a traffic management application. Two closely spaced sensors are preferred for that function.

Magnetometers are sensitive enough to detect bicycles passing across a 4-ft (1.2-m) span when the electronics unit is connected to two sensor probes buried 6 inches (16 cm) deep and spaced 3 ft (0.9 m) apart. Magnetometers can hold the presence of a vehicle for a considerable length of time and do not exhibit crosstalk interference. Vehicle motion is not required for detection by two-axis fluxgate magnetometers.

One channel of an electronics unit can support as many as 12 series-connected sensors. However, the sensitivity is divided among the probes so that there is a loss in sensitivity per probe when more than one is used per channel. The electronics unit will detect with a vehicle over one out of five probes in the series (i.e., 20 percent sensitivity on each probe).

ELECTRONICS UNIT

Specific magnetometer electronics units are matched with various sensor probes. Some units provide only signal amplification and relay or solid state-state outputs that indicate the passage or presence of a vehicle, while others add additional features such as background compensation and sensitivity selection. Some include 2 or more independent detection channels, which can accept 1 to 12 series-connected sensors on each channel.

Calibration procedures are utilized to compensate the electronics unit for the magnetic environment around the roadway sensors. A check is made prior to tuning to assure that no vehicles or movable ferrous equipment are within 20 ft (6 m) of a sensor. Some magnetometer electronics units include, for each channel, a calibration knob that is turned until a pilot light flashes, indicating that the ambient magnetic field has been neutralized. Some magnetometer sensors do not offer a delayed-call timing feature.

Four operating modes are typically available. These are selected on the front panel of the electronics unit and include:

- Presence—Output is maintained throughout the time the vehicle is over the sensor probe. Hold time is unlimited.
- Extended Presence—Output is held for a preselected time interval of up to 5 seconds after each vehicle departs.
- Pulse—A single output occurs during the first 40 ms a vehicle presence is sensed. A subsequent output cannot occur while that presence is maintained.
- Inhibited Pulse—Subsequent output pulses are inhibited for a preselected time up to 5 seconds after each vehicle departs. This eliminates multiple pulses from trailers and long vehicles.

MODEL 227/228 MAGNETOMETER SYSTEM

Chapter 6 of the Caltrans Type 170 Specification contains specifications for the Model 227 magnetometer sensing element and the Model 228 two-channel electronics unit. The electronics unit is defined only for a two-channel version. Each channel operates one to six Model 227 sensors. The electronics unit produces an output signal whenever a vehicle passes over one or more of the sensors.

Since a magnetometer measures the passage or presence of a vehicle, two modes of operation are provided by the electronics unit: pulse, which gives an output closure of 125 ± 25 ms for each vehicle entering the zone of detection, and presence, which gives a continuous output as long as a vehicle occupies the detection zone.

SPVD TWO-AXIS FLUXGATE MAGNETOMETER

Figure 2-45 shows the self-powered vehicle detector. Applications include permanent and temporary installations on freeways and surface streets or where mounting under bridges or viaducts is desired. The SPVD-2 approximates the shape of a 5-inch (13-cm) cube and fits into a cylindrical hole of 6-inch (15-cm) diameter and about 7-inch (18-cm) depth. The upper 2 inches (5 cm) of the hole is filled with cold patch or other sealant that can be removed when the battery needs replacing.

This two-axis fluxgate magnetometer has a self-contained battery and transmitter that broadcasts passage or presence information at 47 MHz over a 400- to 600-ft (122- to 183-m) range to a receiver that can be located remotely in a controller cabinet. The transmitting frequency can be selected to operate in a quiet frequency band, i.e., one where other licensed and unlicensed devices are not operating in the local region. A direct connection (lead-in cable) is not required—a transmitting antenna is built into the housing that encloses the magnetometer electronics and battery. When operated in the presence mode, presence is maintained with a voltage latching circuit for as long as the vehicle is in the detection zone. The sensitivity can be adjusted to restrict vehicle detection to the lane in which the device is installed.

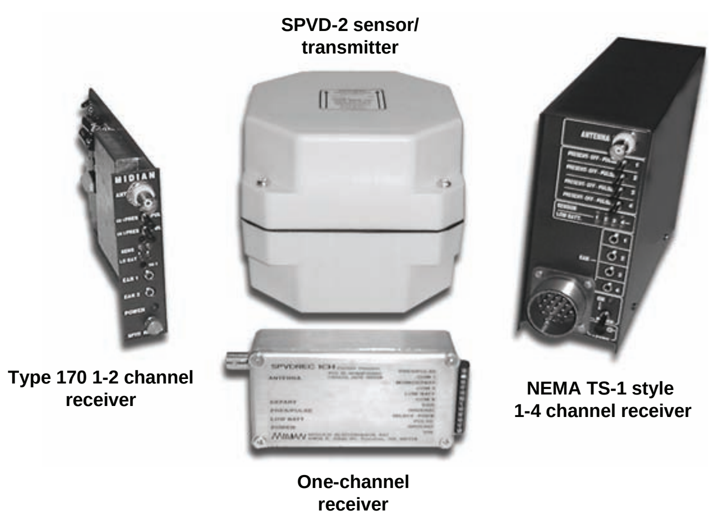


Figure 2-45. SPVD-2 magnetometer system (Photograph courtesy of Midian Electronics, Tucson, AZ).

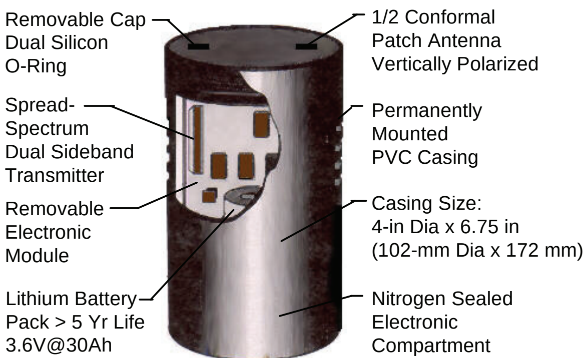
The SPVD is microprocessor controlled and self-calibrating, i.e., it can be installed and adapt to a location in either the Northern or Southern Hemisphere. The battery life is a function of traffic volume and battery type. With an average of 10,000 arrivals and 10,000 departures per day, the alkaline battery is quoted by the SPVD manufacturer as lasting approximately 4 years. Turning off the departure pulse adds another year or two of battery life. A low battery warning is transmitted to the receiver to indicate that 5 to 6 months of battery life remain.

GROUNDHOG MAGNETOMETER

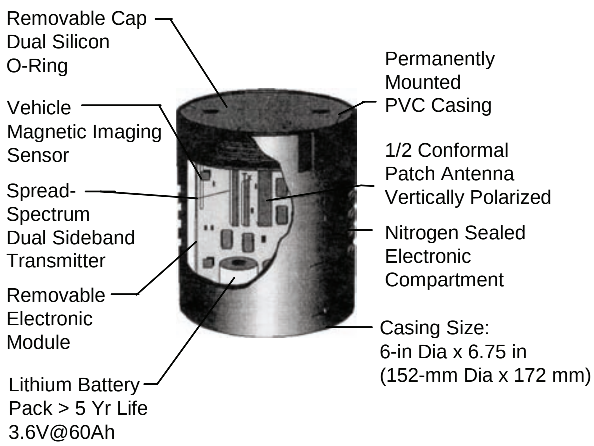
Figure 2-46 illustrates the Groundhog magnetometer sensors. The G-1 fits into a hole of 4.5-inch diameter and 7.5-inch depth (115-mm diameter x 190-mm depth). The G-2 models require a 6.75-inch diameter hole with 7.5-inch depth (172-mm diameter x 190-mm depth). A shorter G-3 model series was designed for installation under a bridge deck. These models had the same functionality as the G-1 and G-2 models. The G-4 series sensors are designed to fit into the G-3 housing, which requires a 6-inch diameter hole of 3.25-inch depth (152.4-mm diameter x 82.6-mm depth).

Buried in the roadway, the G-1 and G-2 sensors transmit their data over the 908 to 922 MHz spread spectrum band to a local base unit located within 200 m (656 ft) of the sensor. The G-4 series sensors transmit data using the 2.45 GHz spread spectrum band. The base unit can be powered from batteries recharged by solar energy.

Groundhog magnetometers incorporate a bridge circuit that is balanced to output zero voltage in the absence of a vehicle and a nonzero voltage when a vehicle enters the detection zone. The G-1 model provides volume, lane occupancy, and road surface temperature. The G-2 adds vehicle speed reported in up to 15 bins, vehicle class in terms of vehicle length reported in up to 6 bins, and a wet/dry pavement indicator. In addition to the previously mentioned traffic parameters, the G-2wx provides chemical analysis for measuring the quantity of anti-icing chemicals on the road surface. The G-4C provides vehicle count only, while the G-4CS provides vehicle count, speed, and classification. The G-4WX adds environmental monitoring of road surface temperature from –67 °Fahrenheit (F) to 185 °F (–55 ° Celsius (°C) to 85 °C) and road surface wet or dry condition to the G-4CS data set.



G-1



G-2 and G-2wx

Figure 2-46. Groundhog magnetometer sensors (Photographs courtesy of Nu-Metrics, Uniontown, PA).

The G-1 sensor is powered from 2 lithium batteries with a quoted battery life of 5 years based on 18,000 cars/day and report intervals of 2 minutes. The G-2 requires 4 lithium batteries with a quoted battery life of 5 years based on 18,000 cars/day and report intervals of 1 minute. The G-4 also operates from 4 lithium batteries for up to 5 years, depending on annual average daily traffic (AADT) and polling interval.

MAGNETIC DETECTORS

The magnetic detector is a simple, inexpensive and rugged device that is capable of only a pulse output. It can be used for traffic-actuated signal control or to simply count vehicles. Some magnetic sensors are installed inside a nonferrous conduit by boring under the roadway. Others are mounted under bridges or in holes cored into the road surface.

THEORY OF OPERATION

Magnetic detectors, i.e., those based on induction or search coil magnetometers, also respond to perturbations in the Earth's magnetic field produced by vehicles passing through the detection zone. The axis of the coil in all magnetic detectors is installed perpendicular to the traffic flow and has an associated spherical detection zone. The distortion and change of the magnetic flux lines with respect to time induces a small voltage that is amplified in an electronics unit located in the controller cabinet. This signal is interpreted by the controller as the passage of a vehicle and produces a call if appropriate.

Magnetic detectors contain a highly permeable magnetic core on which several coils are wound, each with a large number of turns of fine wire, and connected in series. Disturbed lines of magnetic flux cut the turns of the coil and create an output for as long as the vehicle is in motion through the detection zone. Minimum speeds of 3 to 10 mi/h (5 to 16 km/h) are required to produce an actuation. A single magnetic detector does not detect stopped vehicles; therefore, it cannot be used as a presence detector. However, multiple units of some devices can be installed and used with specialized signal processing to generate vehicle presence.

Magnetic detectors respond to flux changes in the lane under which they are buried and to flux changes in adjacent lanes. However, the signal processing is designed to ignore the lower magnitude signals generated in adjacent lanes and analyze only the larger signals produced by vehicles in the lane containing the sensor.

These detectors provide volume, occupancy, and speed data based on the detection zone size and an assumed vehicle length. The historical criteria for their selection are traffic volume accuracy, sensitivity, output data rate, no requirement to detect stopped vehicles, and cost. Magnetic detectors are well suited for snow-belt States where deteriorated pavement and frost break wire loops and where subsurface sensors are desired but the pavement cannot be cut.[2] They also perform well in hot climates where asphalt pavements can become soft from the sun heat load.

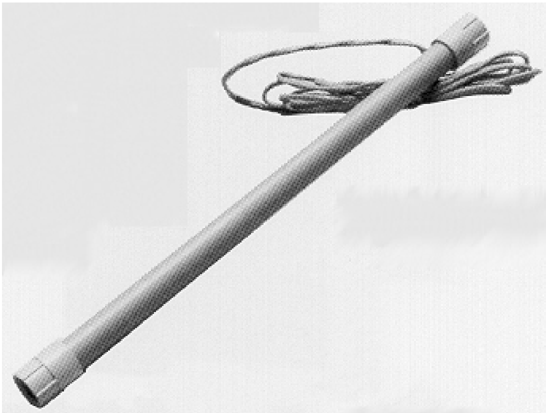
Magnetic detector models differ in their installation and size. Summary specifications of several models are given below.

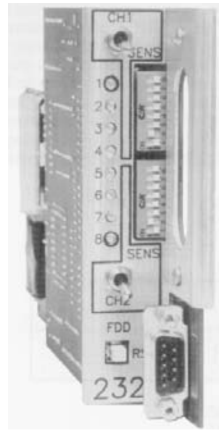
MODEL 231/232 MAGNETIC DETECTOR SYSTEM

Specifications for the Model 231 sensing element and Model 232 two-channel electronics unit are found in Chapter 5 of the Caltrans TEES document (see Appendix K).

Each individual sensor channel and its associated magnetic sensing element operate independently and produce an output signal when a vehicle passes over the embedded sensing element. The solid-state electronics unit, located in the controller cabinet, activates the controller by amplifying the voltage induced in the sensing element by a passing vehicle.

The Model 231 sensor, shown in Figure 2-47, is installed by tunneling under the roadway 14 to 24 inches (356 to 610 mm) and inserting the sensor into a nonferrous plastic or aluminum conduit. In bridge or overpass installation, the sensor is fastened to the understructure directly below the center of the traffic lane. The sensor has a diameter of 2.25 inches (57 mm) and a length of 21 inches (533 mm). The speed data, available from the electronics modules, are based on a detection zone length of 6 ft (1.8 m) and an assumed vehicle length of 18 ft (5.5 m).





231E Detector Probe   232E Detector Electronics

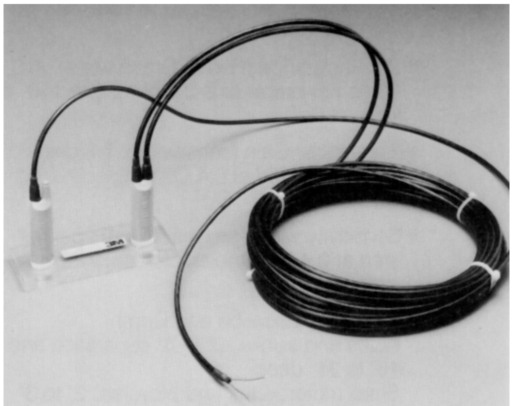
Figure 2-47. Magnetic probe detector (Photographs courtesy of Safetran Traffic Systems, Colorado Springs, CO).

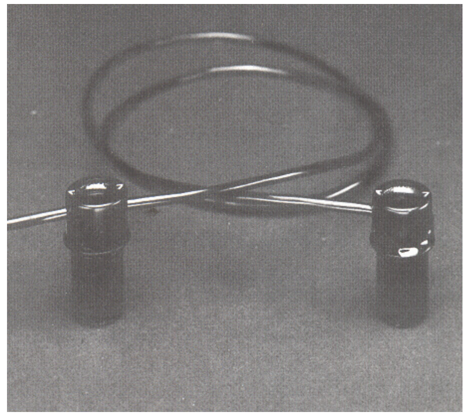
FLUSH-MOUNTED MAGNETIC DETECTOR

Another type of passive magnetic detector is flush-mounted with the surface of the roadway. It is approximately 3 x 5 x 20 inches long (76 x 127 x 508 mm) encased in a cast aluminum housing.

MICROLOOP PROBES

Figure 2-48 depicts other forms of magnetic detectors called microloop probes. The Model 701 probe is inserted into 1-inch (25-mm) diameter holes bored to a depth of 16 to 24 inches (406 to 610 mm). The Model 702 probe is inserted into 3-inch (76-mm) Schedule 80 PVC placed 18 to 24 inches (457 to 610 mm) below the road surface using horizontal drilling from the side of the road. Often two or more microloop probes are connected in series or with conventional wire loops to detect a range of vehicle sizes and obtain required lane coverage. The Model 702 microloop probe can be connected in rows of three to generate signals that detect stopped vehicles. Application-specific software from 3M is also needed to enable stopped vehicle detection.





Model 701 Model 702

Figure 2-48. Microloop probes (Photographs courtesy of 3M Company, St. Paul, MN).

VIDEO IMAGE PROCESSORS

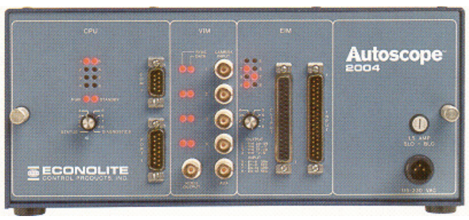
The use of video and image processing technology as a substitute for inductive-loop detectors was first investigated during the mid-1970s. This research was funded by the U.S. Federal Highway Administration (FHWA) and performed by the Jet Propulsion Laboratory. The project combined television camera and video processing technology to identify and track vehicles traveling within the camera's field of view.[12] During the 1970s and 1980s, parallel efforts were undertaken in Japan, the United Kingdom, Germany, Sweden, and France. These investigations addressed the problems and limitations of existing roadway sensors in attempting to fulfill requirements for state-of-the-art control of traffic and detection of incidents.

As a follow-on to the initial FHWA project, a video image processor was developed by University of Minnesota research personnel.[13] Dubbed the *Vi*deo *D*etection *S*ystem (VIDS), it was jointly funded by FHWA, Minnesota Department of Transportation, and the University of Minnesota. The sensor provided volume and occupancy data equivalent to those from multiple inductive-loop detectors. Full bandwidth imagery and the traffic data were transmitted to a central location for interpretation and management of traffic.

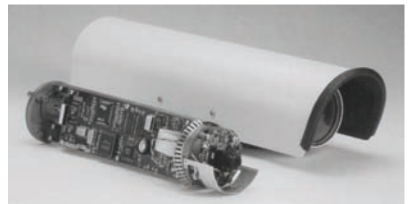
VIDEO IMAGE PROCESSING VEHICLE DETECTION CONCEPTS

Three classes of VIP systems are now produced or in development: tripline, closed-loop tracking, and data association tracking. Tripline systems, as used by the sensors in Figures 2-49a, 2-49b, 2-49c, and 2-49d, operate by allowing the user to define a limited number of linear detection zones on the roadway in the field-of-view of the video camera. When a vehicle crosses one of these zones, it is identified by noting changes in the properties of the affected pixels relative to their state in the absence of a vehicle. Tripline systems that estimate vehicle speed measure the time it takes an identified vehicle to traverse a detection zone of known length. The speed is found as the length divided by the travel time.

Surface-based and grid-based analyses are utilized to detect vehicles in tripline VIPs. The surface-based approach identifies edge features, while the grid based classifies squares on a fixed grid as containing moving vehicles, stopped vehicles, or no vehicles.



(a) Autoscope 2004



(b) Autoscope Solo

(Photographs courtesy of Econolite Control Products, Anaheim, CA)

(c) Traficon VIP 3 (Photograph courtesy of Traficon, Heule, Belgium)

(d) Iteris Vantage processors (Photograph courtesy of Iteris, Anaheim, CA)

Figure 2-49. Tripline video image processors.

The advent of VIP tracking sensors has been facilitated by low-cost, high-throughput microprocessors. Closed-loop tracking systems, as used in some VIPs, as illustrated in Figure 2-50, detect and continuously track vehicles along larger roadway sections. The tracking distance is limited by the field of view, mounting height, and resolution of the camera. Multiple detections of the vehicle along the track are used to validate the detection and improve speed estimates. Once validated, the vehicle is counted and its speed is updated by the tracking algorithm.(14) These tracking systems may provide additional traffic flow data such as lane-to-lane vehicle movements. Thus, they have the potential to transmit information to roadside displays and radios to alert drivers to erratic behavior that can lead to an incident. Their ability to monitor turning movements on arterials may allow more frequent updating of signal timing plans.

Figure 2-50. Closed-loop tracking video image processor.

Data association tracking systems identify and track a particular vehicle or groups of vehicles as they pass through the field of view of the camera. The computer identifies vehicles by searching for unique connected areas of pixels. These areas are then tracked from frame-to-frame to produce tracking data for the selected vehicle or vehicle groups.[15–17] Objects are identified by markers that are derived from gradients and morphology. Gradient markers utilize edges, while morphological markers utilize combinations of features and sizes that are recognized as belonging to known vehicles or groups of vehicles.[18] Data association tracking can potentially provide link travel time and origin-destination pair information by identifying and tracking vehicles as they pass from one camera's field of view to the next camera's field of view.

Modern video image processors conserve transmission bandwidth by performing signal processing in microprocessors located either in the camera (as for example in the Autoscope Solo) or in modules mounted in a controller cabinet at the roadside as shown on the upper shelf in Figure 2-51. The data are used locally by the controller for signal timing and to characterize the level of service on freeways, for example. The data can also be transmitted to the traffic operations center over low-bandwidth communications media for incident detection and management, database update, and traveler information services.

By multiplexing video images from several cameras on one transmission line and sending the video only when requested, operating costs associated with leased transmission media are further reduced. By extending the identification of vehicles or groups of vehicles from the field of view of one camera into that of another, vehicles can be tracked over greater distances to compute link travel times and generate origin-destination pair data. The output products from the traffic management center can be forwarded to other transportation management centers, emergency service providers, and information service providers.

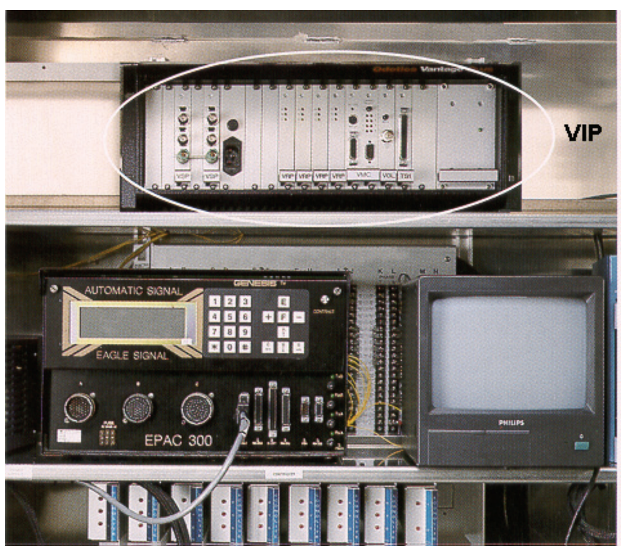


Figure 2-51. Video image processor installed in a roadside cabinet
(Photograph courtesy of Iteris, Anaheim, CA).

SIGNAL PROCESSING

Image formatting and data extraction are performed with firmware that allows the algorithms to run in real time. The hardware that digitizes the video imagery is commonly implemented on a single formatter card in a personal computer architecture or in microprocessors located in the camera housing. Once the data are digitized and stored by the formatter, spatial and temporal features are extracted using a series of image processing algorithms to differentiate vehicles from roadway or background pixels.

In the concept illustrated in Figure 2-52, a camera is used to acquire imagery of the traffic flow. The images are digitized and stored in memory. A detection process establishes one or more thresholds that limit and segregate the digitized data passed on to the rest of the image processing algorithms. It is undesirable to severely limit the number of potential vehicles during detection, for once data are removed they cannot be recovered. Therefore, false vehicle detections are permitted at this stage since the declaration of actual vehicles is *not* made at the conclusion of the detection process. Rather algorithms that are part of the classification, identification, and tracking processes still to come are relied on to eliminate false vehicles and retain the real ones.[19] Image segmentation is used to divide the image area into smaller regions (often composed of individual vehicles) where features can be better recognized. The feature extraction process examines the pixels in the regions for preidentified characteristics that are indicative of vehicles. When a sufficient number of these characteristics are present and recognized by the processing, a vehicle is declared present and its flow parameters are calculated.

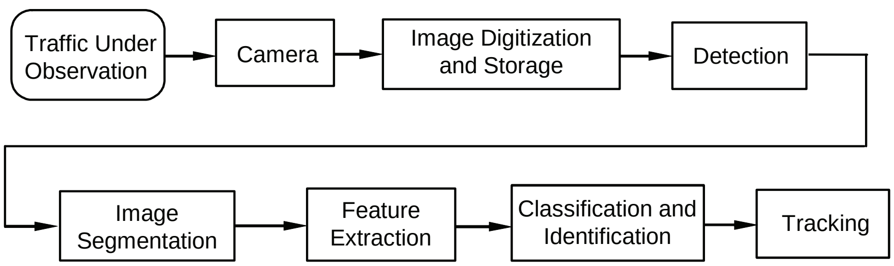


Figure 2-52. Conceptual image processing for vehicle detection, classification, and tracking. (Source: L.A. Klein, *Sensor Technologies and Data Requirements for ITS* (Artech House, Norwood, MA, 2001)).

Artificial neural networks are another form of processing used to classify and identify vehicles, measure their traffic flow parameters, and detect incidents.[20] Features are not explicitly identified and sought when this processing approach is used. Rather an electrical network that emulates the processing that occurs in the human brain is trained to recognize vehicles. The digital imagery is presented to the trained network for vehicle classification and identification.

VIPs with tracking capability use Kalman filtering techniques to update vehicle position and velocity estimates.[21] The time trace of the position estimates yields a vehicle trajectory. By processing the trajectory data, local traffic parameters (e.g., flow and lane change frequency) can be computed. These parameters, together with vehicle signature information (e.g., time stamp, vehicle type, color, shape, position, and speed), can be communicated to the traffic management center.[22] Tracking vehicle subfeatures. such as edges, corners, and two-dimensional patterns, rather than entire vehicles, has been proposed to make the VIP robust to partial occlusion of vehicles in congested traffic. Preliminary results with this technique show insensitivity to shadows since shadow subfeatures tend to be unstable over time, especially in congestion. Camera mounting position and continued full occlusion of smaller vehicles by trucks still deteriorate performance.[23,24]

A signal processing technique implemented by Computer Recognition Systems (Wokingham, Berkshire, England; Knoxville, TN) incorporates wireframe models composed of line segments to represent vehicles in the image. This approach claims to provide more unique and discriminating features than other computationally viable techniques.

The artificial neural network approach is incorporated by Nestor Traffic Systems, Inc. (Providence, RI) in their VIP products. An advantage of the Nestor implementation is that the camera can be repositioned for data acquisition and surveillance.[25] VIPs that utilize tracking offer the ability to warn of impending incidents due to abrupt lane changes or weaving, calculate link travel times, and determine origin-destination pairs. The tracking concept is found in the VideoTrak 905 and 910 by Peek Traffic-Transyt, the Traffic Analysis System by Computer Recognition Systems, MEDIA4 developed by Citilog (Paris, France), and the IDET-2000 by Sumitomo (Japan).

PERFORMANCE

VIP signal processing is continually improving its ability to recognize artifacts produced by shadows, illumination changes, reflections, inclement weather, and camera motion from wind or vehicle-induced vibration. However, artifacts persist and the user should evaluate VIP performance under the above conditions and other local conditions that may exist. In their 1998 report to the TRB Freeway Operations Committee, the New York State Department of Transportation (NYSDOT) stated that one VIP model had difficulty detecting vehicles on a roadway lightly covered with snow in good visibility. Another model did not experience this problem. A 2004 evaluation of VIP performance by Purdue University described significant false and missed vehicle detections as compared with loops, even when the cameras were installed at vendor-recommended locations at a well-lighted intersection.[24]

Figure 2-53 illustrates the effect day-to-night illumination change has on VIP performance. The VIP cameras viewed downstream traffic from a mast arm position approximately 25 ft (7.6 m) above the curb and middle lanes of a three-lane roadway. Shortly after 1,900 hours, there are changes in the slopes of the VIP vehicle count data due to either degradation in performance of the daytime algorithm or the different performance of the nighttime algorithm.[11]

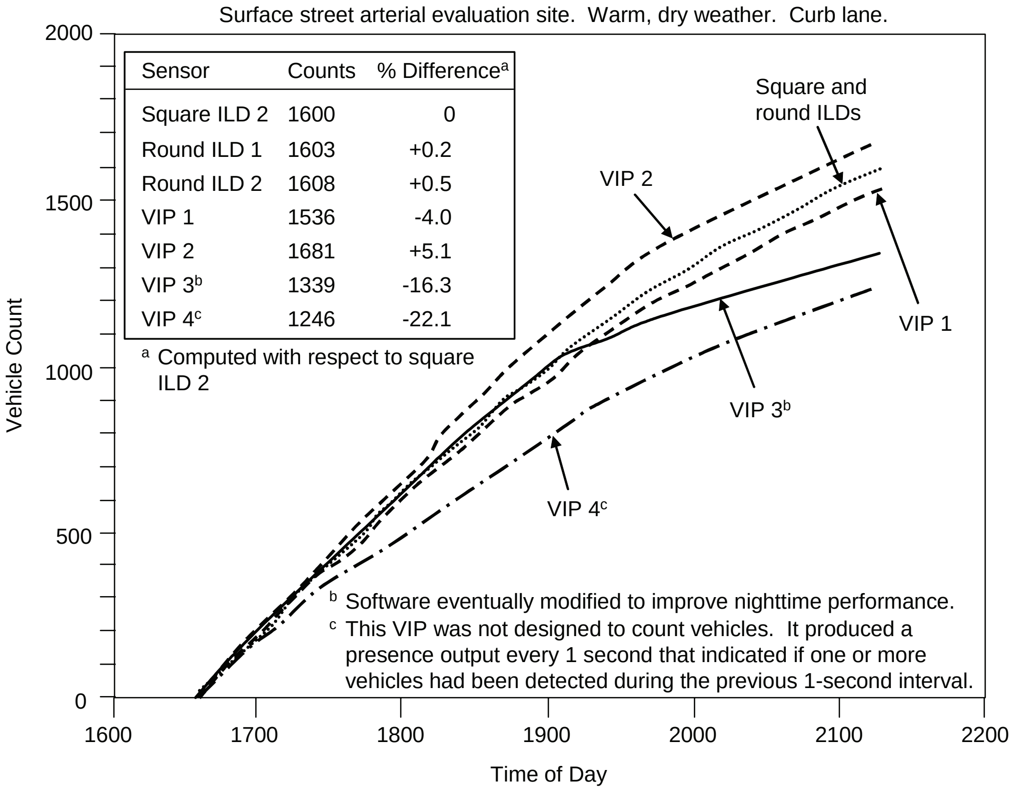


Figure 2-53. Vehicle count comparison from four VIPs and inductive-loop detectors (source: L.A. Klein and M.R. Kelley, *Detection Technology for IVHS*: *Final Report*, Report No. FHWA-RD-95-100 (USDOT, FHWA, Washington, D.C.,1995)).

Figure 2-54 shows the effects of day-night illumination changes from another VIP test.[26] The camera faced upstream traffic, mounted approximately 25 ft (7.6 m) above the road surface. The camera was placed on a mast arm extension on the left side of a roadway, which contained three through lanes and a dedicated right-turn lane. The VIP tracked vehicles and provided flow rate, lane speed, queue length, density, number of left and right turning vehicles, and approach stops. VIP performance was not optimized because of a number of factors: low camera sensitivity that was more problematical at low light levels, focal length of lens not matched to the viewing distance, low camera resolution, low and offcenter camera mounting height, inadequate video signal output from camera to drive both the VIP and video cassette recorder, no sun shade, and camera vibration with winds greater than 10 mi/h (16 km/h).

Because of these nonideal conditions, a quantitative comparison of the VIP-generated traffic flow parameters with ground truth is not given here. Instead, the focus is on the effect that illumination changes have on VIP performance, albeit exaggerated by many of the conditions cited above. The effects of low camera sensitivity and inadequate video signal are most apparent at dawn, night, and dusk where the false alarm and missed vehicle errors (0.15, 0.44, and 0.91, respectively) are largest. These results demonstrate the importance of proper camera and lens selection, camera mounting, and illumination in maximizing VIP performance.

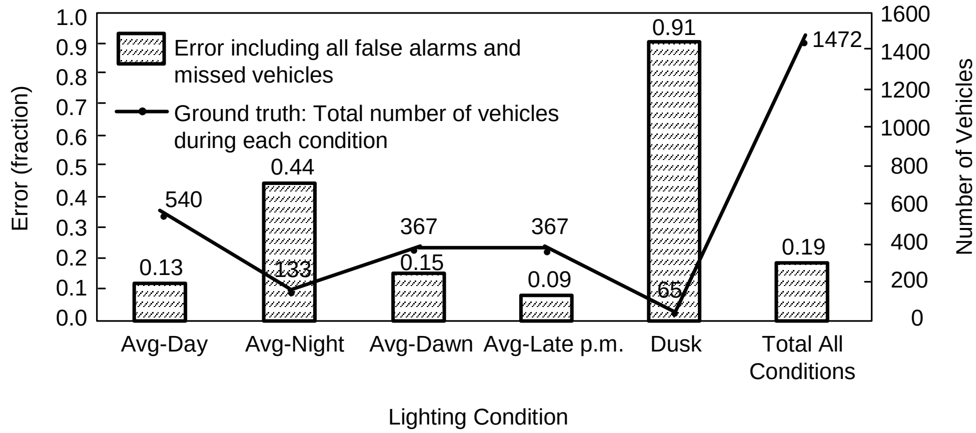


Figure 2-54. VIP vehicle count errors under varying illumination conditions. Low camera sensitivity, improper focal length of lens, low camera resolution, nonideal camera mounting height, inadequate video signal, and lack of sun shade contributed to degraded performance (source: *Traffic Surveillance and Detection Technology Development*: *Phase II and III Test Results*, *Troy, MI*, JPL Pub. D-15779 (California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, May 1998)).

Heavy congestion that degraded early VIPs does not appear to present as great a problem to more modern systems. Combined results for clear and inclement weather show vehicle volume, speed, and occupancy measurement accuracies in excess of 95 percent using a single detection zone and a camera mounted at a sufficiently high height. VIPs with single or multiple detection zones per lane can be used to monitor traffic on a freeway. For signalized intersection control, where vehicle detection accuracies of 100 percent are desired, the number of detection zones per lane is increased to between two and four, depending on the camera mounting and road geometry. However, even with multiple detection zones, vehicle detection accuracy can degrade to 85 percent or less with side-viewing cameras that are not mounted high enough (on the order of 30 ft (9 m) rather than 50 ft (15 m)) or are not directly adjacent to the roadway.[27–29] The study that produced these results also reported that vehicle detection was sometimes sensitive to vehicle-to-road color contrast.

Additional merits and issues associated with application of video image detection at signalized intersections are reported in a 2004 survey conducted by the *Urban Transportation Monitor*.[30] The survey contained responses from 120 jurisdictions. The survey results were summarized by the *Monitor* as follows:

> …the use of video detection at signalized intersections is clearly controversial with many respondents having very strong opinions (both positive and negative) about its application. For example, 35% of the respondents experienced more complaints with the application of video detection, 19% indicated they experienced fewer complaints, and the balance indicated that the number of complaints have remained the same. On the other hand, 66% of respondents indicated that they will increase their application of video detection while 12% indicated that they will actually decrease their applications.
>
> These differences in experience and attitude are also reflected in the answers to the questions related to the main advantages and disadvantages compared to inductive loops. About the same number

of positive replies (advantages) and negative replies (disadvantages) were received. Most frequent positive replies had to do with the ability of video detection to cope with changes in the detection zones due to restriping and the most frequent negative replies had to do with the inability of video detection to provide adequate results (or any results at all) during inclement weather (fog, heavy snow) and when the sun shines directly at the video camera.

What seems to be clear is that agencies who are contemplating the use of video detection should approach it carefully as there are many pitfalls, as indicated in the survey results. It seems clear that it is also important to make sure a vendor is selected that can provide the latest improvements in video detection technology.

MOUNTING AND TRAFFIC VIEWING CONSIDERATIONS

Video image processor cameras can be deployed to view upstream or downstream traffic. The primary advantage of upstream viewing is that incidents are not blocked by the resultant traffic queues as described in Table 2-14. However, tall vehicles such as trucks may block the line of sight, and headlights may cause blooming of the imagery at night. With upstream viewing, headlight beams can be detected as vehicles in adjacent lanes on curved road sections. Downstream viewing conceals cameras mounted on overpasses so that driver behavior is not altered. Downstream viewing also makes vehicle identification easier at night through the information available in the taillights and enhances track initiation because vehicles are first detected when close to the camera.[31]

Table 2-14. Performance comparison of a visible spectrum video image processor using upstream and downstream viewing.

Upstream viewing	Downstream viewing
• Headlight blooming, glare on wet pavement, headlight beams detected in adjacent lanes on curved road sections. • More blockage from tall trucks. • Traffic incidents not blocked by resulting traffic queues.	• Camera on overpass concealed from drivers. • More information from taillights available for braking indication, vehicle classification, turning movement identification, and tracking. • Easier to acquire vehicles that are closer to the camera for tracking algorithm implementation.

Although some manufacturers quote a maximum surveillance range for a VIP of ten times the camera mounting height, conservative design procedures limit the range to smaller distances because of factors such as road configuration (e.g., elevation changes, curvature, and overhead or underpass structures), congestion level, vehicle mix, and inclement weather. The impact of reduced headway on the effective surveillance range is calculated from the distance d (along the roadway from the base of the camera mounting structure to the vehicles in question) at which the VIP can distinguish between two closely spaced vehicles. The distance d depends explicitly on camera mounting height, vehicle separation or gap, and vehicle height, as illustrated in Figure 2-55.

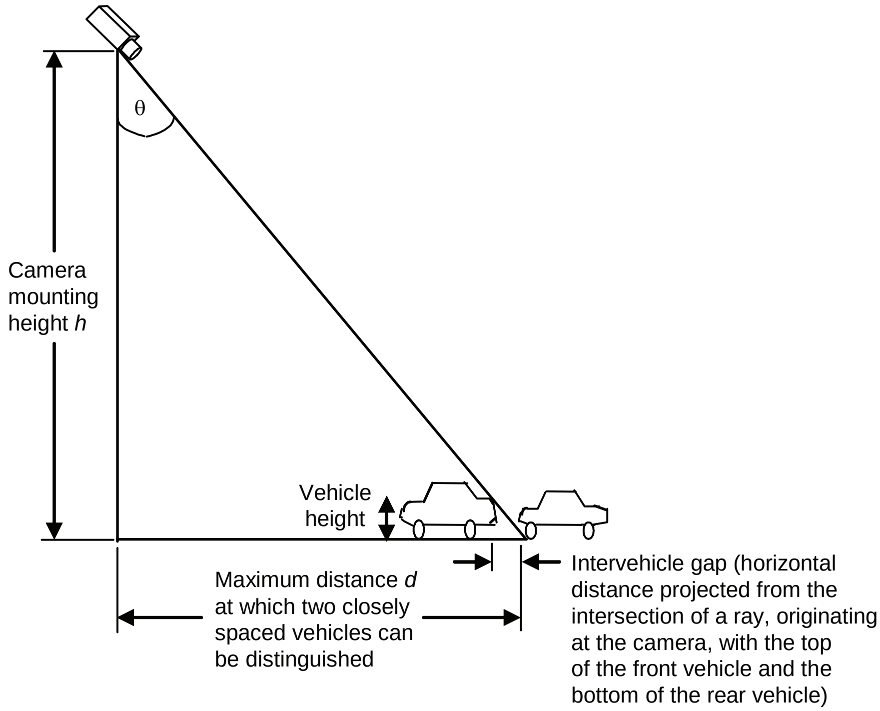


Figure 2-55. Distinguishing between two closely spaced vehicles (source: L.A. Klein, *Sensor Technologies and Data Requirements for ITS* (Artech House, Norwood, MA, 2001)).

If the roadway does not have a grade, d is approximated by

$$d = h \frac{Veh_{gap}}{Veh_{height}} \tag{2-63}$$

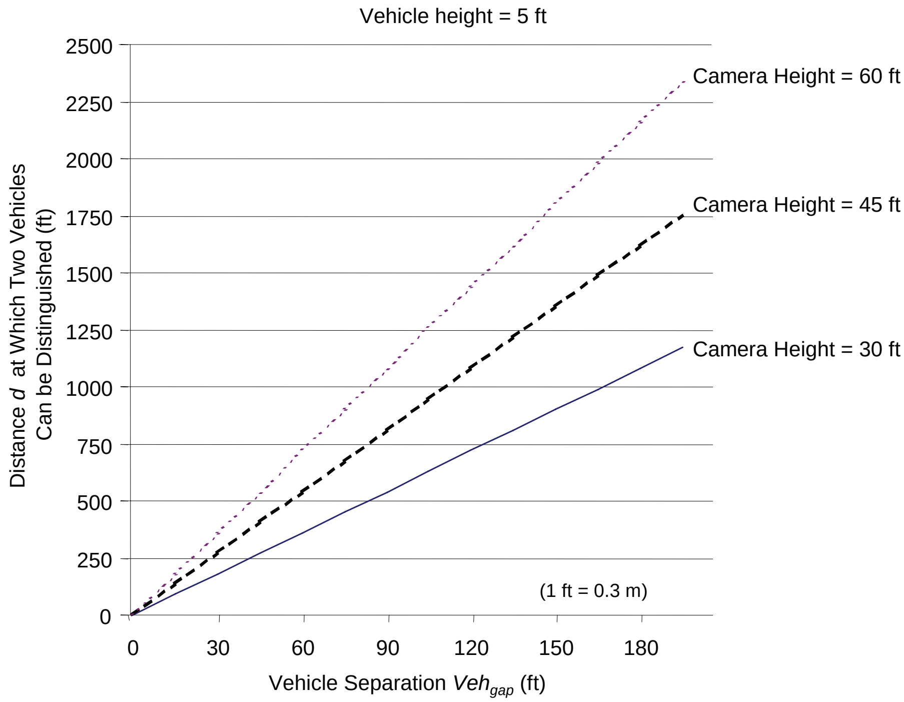
where:

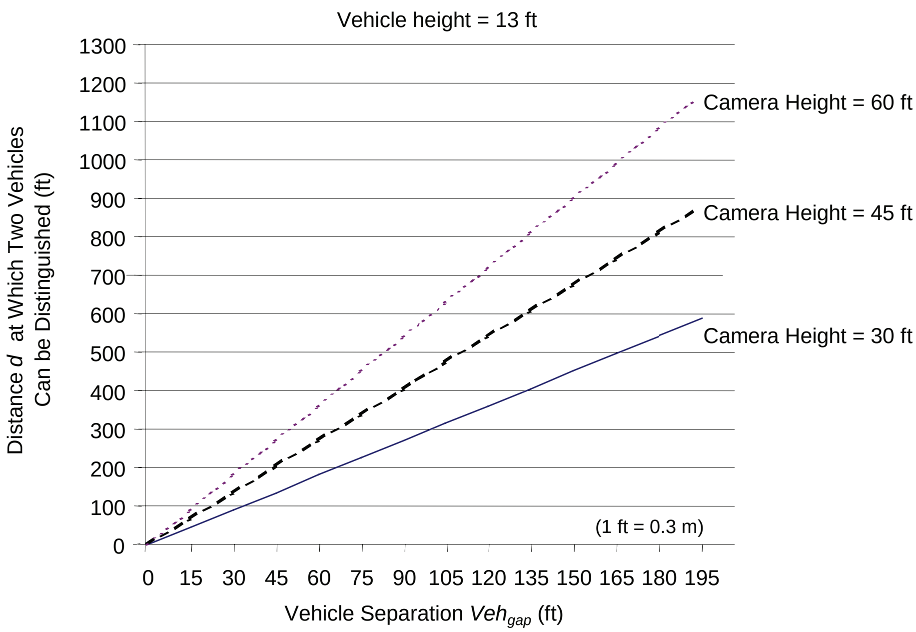
h = camera mounting height

Veh_{gap} = intervehicle gap

Veh_{height} = vehicle height.

Distance d is also implicitly dependent on the pixel size or instantaneous field-of-view of the camera, as larger values of d may not be realized without a correspondingly small pixel size. Figure 2-56 contains plots of distance d versus vehicle separation based on Equation 2-63 for vehicle heights of 5 ft (1.5 m), representative of a passenger automobile, and 13 ft (4 m), representative of a larger commercial vehicle.





1 ft = 0.3 m

Figure 2-56. Distance *d* along the roadway at which a VIP can distinguish vehicles: (top) vehicle height = 5 ft (1.5 m), (bottom) vehicle height = 13 ft (3.9 m). These values may be further limited by road configuration, congestion level, vehicle mix, inclement weather, and pixel size.

Other factors that affect camera installation include vertical and lateral viewing angles, number of lanes observed, stability with respect to wind and vibration, and image quality. VIP cameras can be mounted on the side of a roadway if the mounting height is high, i.e., 50 ft (15.2 m) or higher. For lower mounting heights of 25 to 30 ft (7.6 to 9.1 m), a centralized location over the middle of the roadway area of interest is required. However, the lower the camera mounting, the greater is the error in vehicle speed measurement—the measurement error is proportional to the vehicle height divided by the camera mounting height. False detection and missed detections also increase at lower camera mounting heights. (See references 14, 24, and 32–34.)

The number of lanes of imagery analyzed by the VIP is of importance to ensure that the required observation and analysis region is supported by the VIP. For example, if the VIP provides data from detection areas in three lanes, but five must be observed, that particular VIP may not be appropriate for the application.

VIPs that are sensitive to large camera motion may be adversely affected by high winds since the processor may assume that the wind-produced changes in background pixels correspond to vehicle motion. Software that automatically learns the new operating environment and road configuration on which a VIP is operating has entered the market place. One such tool developed by Citilog advertises that no configuration, calibration, or parameter entry are necessary for their software to operate with existing pan, tilt, and zoom cameras for detection of stopped vehicles, congestion, and accidents.

Image quality and interpretation can be affected by cameras that have automatic iris and automatic gain control. An automatic iris adjusts the light level entering the camera not only when roadway background lighting changes, but also when headlights, reflections from windshields and bumpers, or white or bright objects are in the field of view. Automatic gain control momentarily reduces the sensitivity of the camera to the same phenomena. These controls impair the ability of cameras with a low signal-to-noise ratio to detect following vehicles when they restrict entering light levels.(26) When the camera responds by quickly darkening the entire picture and then recovering, some systems interpret this as an extra vehicle.

In tests conducted by California Polytechnic University at San Luis Obispo (Cal Poly SLO), automatic iris and gain controls were disabled in an effort not to handicap the detection ability of the VIPs they evaluated.(35) In followup tests several years later, Cal Poly SLO found VIPs better able to compensate for light level changes when the automatic iris responded slowly to variations in light entering the camera.(36) This finding was confirmed by the Texas Transportation Institute (TTI), which specifies its VIP cameras to have damped iris and automatic gain controls.(32)

When a traffic management agency wishes to use a single camera to provide imagery to a VIP and to obtain video surveillance with pan, tilt, and zoom controls, it is necessary to reposition the camera to its calibrated position for each VIP application. If the camera is not in the calibrated position, the performance of the VIP is degraded. If remote control of cameras and their return to calibrated fields of view is not feasible, then separate cameras may be required to perform automated traffic data collection and video surveillance. Some VIPs can automatically recalibrate the field of view for a new camera position using specialized algorithms.

VIDEO SURVEILLANCE ON FREEWAYS AND ARTERIALS

Even without its association with VIPs, closed-circuit television (CCTV) has become a valuable asset for traffic management. In addition to its primary task of incident verification on freeways, CCTV is used for other applications in freeway and corridor traffic management. These include:

- Monitoring traffic movements on the mainline, HOV lanes, and the entrance and exit ramps (e.g., driver response to messages on changeable message signs, HOV lane use, compliance with ramp metering, and verification of incidents).
- Overlooking surface streets that run parallel to a freeway to verify that the local arterial system has unused vehicle capacity before implementing freeway diversion and then monitoring system performance as diversion occurs.
- Monitoring the operation of critical signalized intersections and evaluating the signal timing and related functions.
- Verifying message displays on changeable message signs.

Current CCTV technology allows viewing of 0.25 to 0.5 mi (0.4 to 0.8 km) in each direction if the camera mounting, topography, road configuration, and weather are ideal. The location for CCTV cameras is dependent on the terrain, number of horizontal and vertical curves, desire to monitor weaving areas, identification of high-incident locations, and the need to view ramps and arterial streets. Each prospective site must be investigated to establish the camera range and field of view that will be obtained as a function of mounting height and lens selection.

MICROWAVE RADAR SENSORS

Two types of microwave radar sensors are used in roadside applications, those that transmit continuous wave (CW) Doppler waveforms and those that transmit frequency modulated continuous waves (FMCW). The traffic data they receive are dependent on the respective shape of the transmitted waveform.

CW DOPPLER RADAR

Figure 2-57 depicts the constant frequency waveform transmitted by a CW Doppler radar. This sensor is also referred to in some literature as a microwave or microwave Doppler sensor. The constant frequency signal (with respect to time) allows vehicle speed to be measured using the Doppler principle. Accordingly, the frequency of the received signal is decreased by a vehicle moving away from the radar and increased by a vehicle moving toward the radar. Vehicle passage or count is denoted by the presence of the frequency shift. Vehicle presence cannot be measured with the constant frequency waveform since only moving vehicles are detected. Two microwave radars that use the Doppler principle to measure speed are shown in Figure 2-58.

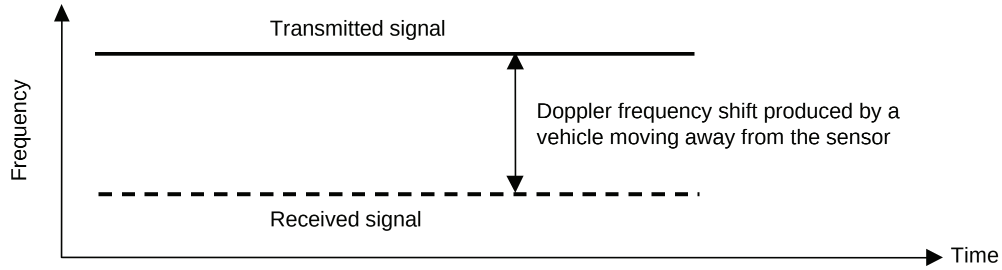
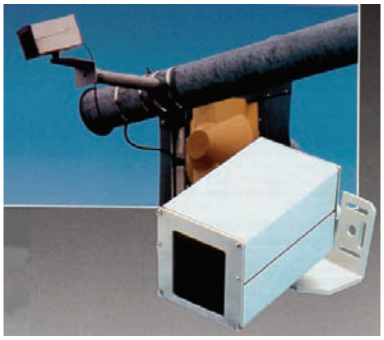


Figure 2-57. Constant frequency waveform.



TC-20 Doppler microwave radar. (Photograph courtesy of Microwave Sensors, Ann Arbor, MI)

TDN-30 Doppler microwave radar. (Photograph courtesy of Whelen Engineering Company, Chester, CT)

Figure 2-58. CW Doppler microwave radars.

Vehicle speed is proportional to the frequency change between transmitted and received signals in Doppler radars. The relation between the transmitted frequency f, Doppler frequency f_D, and vehicle speed S is expressed as

$$f_D = 2Sf/c \cos\theta \tag{2-64}$$

where θ is the angle between the direction of propagation of radar energy (i.e., the angle that represents the center of the antenna beam) and the direction of travel of the vehicle, c is the speed of light (3×10^8 m/s), frequency is in units of Hz, and speed in units of m/s. The received frequency is given by $f \pm f_D$.[31]

FMCW RADAR

The second type of microwave sensor used in traffic management and control applications transmits an FMCW waveform in which the transmitted frequency is constantly changing with respect to time, as depicted in Figure 2-59a. The FMCW radar operates as a vehicle presence sensor and can detect motionless vehicles.

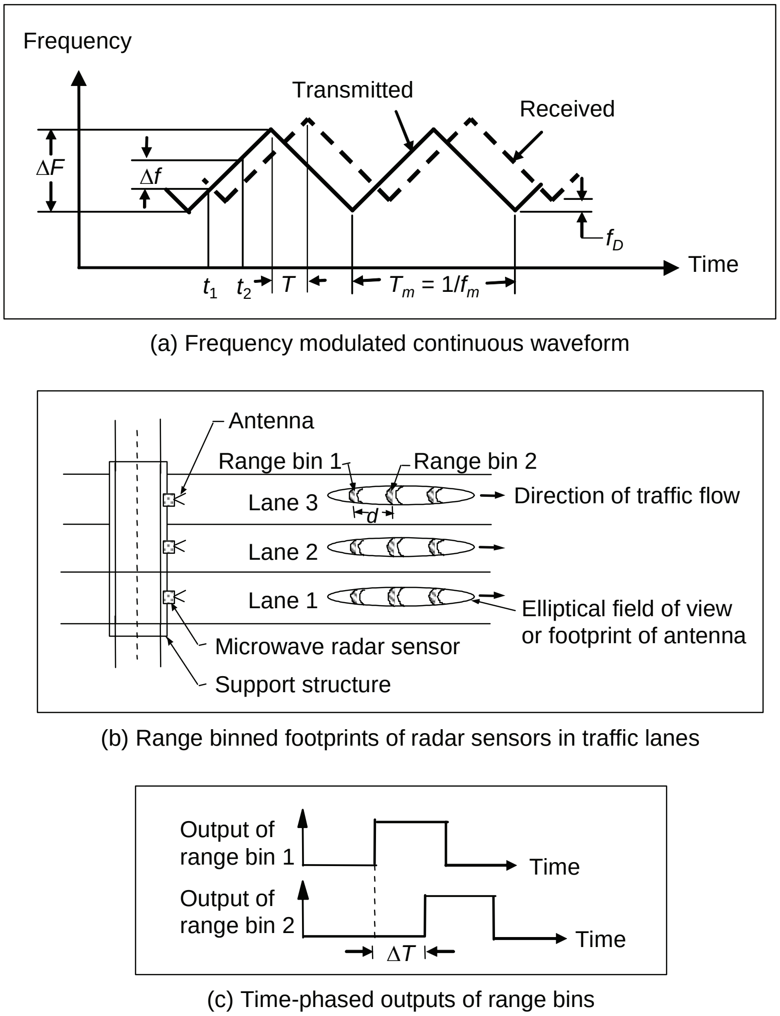


Figure 2-59. FMCW signal and radar processing as utilized to measure vehicle presence and speed (Source: L.A. Klein, *Sensor Technologies and Data Requirements for ITS* (Artech House, Norwood, MA, 2001)).

The range R to the vehicle is proportional to the difference in the frequency Δf of the transmitter at the time t_1 the signal is transmitted and the time t_2 at which it is received, as given by

$$R = \frac{c\ \Delta f}{4\ \Delta F\ f_m} \tag{2-65}$$

where

$\Delta f =$ instantaneous difference in frequency, in Hz, of the transmitter at the times the signal is transmitted and received

ΔF = radio frequency (RF) modulation bandwidth in Hz

f_m = RF modulation frequency in Hz.

Alternatively, the range may be calculated from $t_2 - t_1$ or from the time difference T between consecutive peaks in the received and transmitted signals as indicated in Figure 2-59a. The range R, in terms of $t_2 - t_1$ or T, is

$$R = c(t_2 - t_1)/2 = cT/2 \tag{2-66}$$

when the transmitter and receiver are collocated. The quantity c is the speed of light (3×10^8 m/s).

The range resolution ΔR or minimum distance, in meters, resolved by an FMCW radar is

$$\Delta R = \frac{c}{2\Delta F}. \tag{2-67}$$

Therefore, if the radar operates in the 10.500 to 10.550 GHz band and the bandwidth is limited to 45 MHz (rather than the full 50 MHz) to ensure that the field strength is reduced by at least 50 dB outside the band, the range resolution is at best 10.8 ft (3.3 m).

Speed or Doppler resolution Δf_D is given by

$$\Delta f_D = 2/T_m \text{ Hz} \tag{2-68}$$

where $T_m = 1/f_m$ is the time for an up and down frequency sweep, as depicted in Figure 2-59a.

The FMCW radar measures vehicle speed by dividing the field of view in the direction of vehicle travel into range bins as shown in Figure 2-59b. A range bin allows the reflected signal to be partitioned and identified from smaller regions on the roadway. Vehicle speed S is calculated from the time difference ΔT corresponding to the vehicle arriving at the leading edges of two range bins a known distance d apart as illustrated in Figure 2-59c. The vehicle speed is given by

$$S = \frac{d}{\Delta T} \tag{2-69}$$

where

$d =$ distance between leading edges of the two range bins

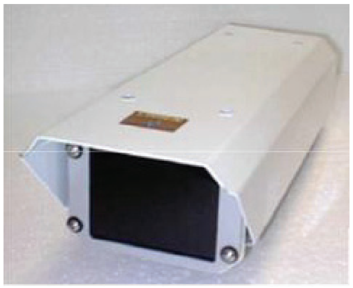
$\Delta T =$ time difference corresponding to the vehicle's arrival at the leading edge of each range bin.

When mounted in a side-looking configuration, multilane FMCW radar sensors can monitor traffic flow in as many as eight lanes. With the sensor aligned perpendicular to the traffic flow direction, the range bins are automatically or semiautomatically (depending on the sensor model) adjusted to overlay a lane on the roadway to enable the gathering of multilane traffic flow data. FMCW radars can also use Doppler to calculate the speed of moving vehicles. A discussion of this technique is beyond the scope of this *Handbook*, but can be found in References 19 and 37.

Presence-detecting radars, such as the models illustrated in Figure 2-60, control left turn signals, provide real-time data for traffic adaptive signal systems, monitor traffic queues, classify vehicles in terms of vehicle length, and collect occupancy and speed (multi detection zone models only) data in support of freeway incident detection algorithms. CW Doppler radars are used to measure vehicular speed on city arterials and freeways, but cannot detect stopped vehicles. Multi detection zone microwave presence-detecting radars are gaining acceptance in electronic toll collection and automated truck weighing applications that require vehicle identification based on vehicle length.

RTMS multizone presence-detecting microwave radar. (Photograph courtesy of Lawrence A. Klein).

SmartSensor multizone presence-detecting microwave radar. (Photograph courtesy of Wavetronix, Provo, UT).



Loren multizone presence-detecting microwave radar. (Photograph courtesy of Electronic Control Measurement Inc, Manor, TX).

150LX single zone presence-detecting microwave radar. (Photograph courtesy of Naztec, Inc., Sugar Land, TX).

Figure 2-60. FMCW microwave radars. These are examples of microwave radars that detect vehicle presence.

LASER RADARS

Laser radar sensors have two sets of optics. The transmitting optics split the pulsed laser diode output into two beams separated by several degrees, as displayed in Figure 2-61. The receiving optics has a wider field of view so that it can better collect the energy scattered from the vehicles. The multiple beams allow laser radars to measure vehicle speed by recording the times at which the vehicle enters the detection area of each beam. Since the beams are a known distance apart, the speed is easily calculated from the familiar speed equals distance over time.

The laser radar illustrated in Figure 2-62 mounts 20 to 25 ft (6.1 to 7.6 m) above the road surface with an incidence angle (i.e., forward tilt) of 5 deg. A rotating polygon is utilized to line scan a laser diode rangefinder across the traffic lane as shown in Figure 2-61.[38] A sister model scans the laser across two lanes.[39] Their ability to classify 11 types of vehicles has found application on toll roads.

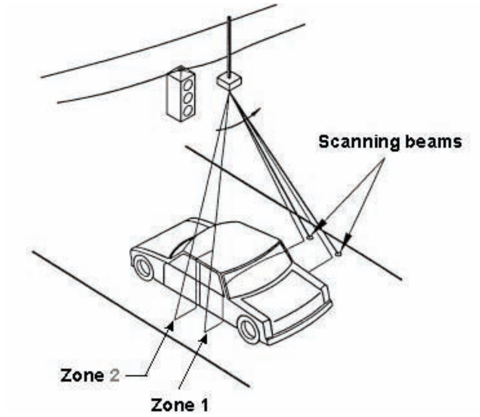


Figure 2-61. Laser radar beam geometry (drawing courtesy of Schwartz Electro-Optics, now OSI Laserscan, Orlando, FL).

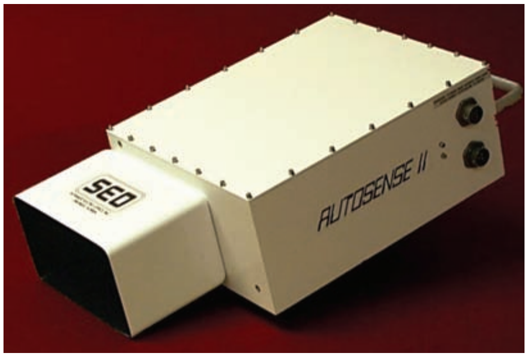


Figure 2-62. Autosense II laser radar sensor (Photograph courtesy of Schwartz Electro-Optics, now OSI Laserscan, Orlando, FL).

Another commercial laser radar for collecting traffic flow data is shown in Figure 2-63. This model does not continuously scan the laser beam across a traffic lane. Instead it uses six individual pairs of transmitters and receivers to produce a beam that traverses the lane. The number of active transmitter-receiver pairs can be specified by the user to control the width of the laser scan across a single lane. The effective range of the device is up to 23 ft (7 m).

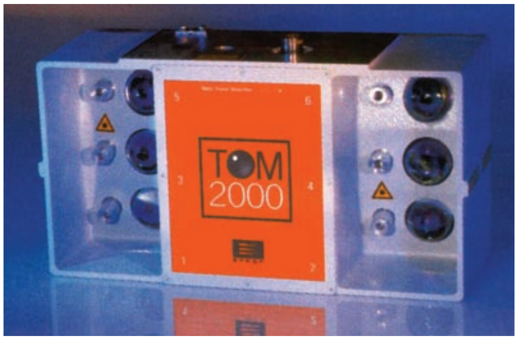


Figure 2-63. EFKON Traffic Observation Module (TOM) laser radar sensor (Photograph from EFKON product literature).

Active infrared sensors also communicate traffic information to motorists by modulating and coding the infrared beam with the appropriate data. Figure 2-64 depicts a sensor system used for this application.

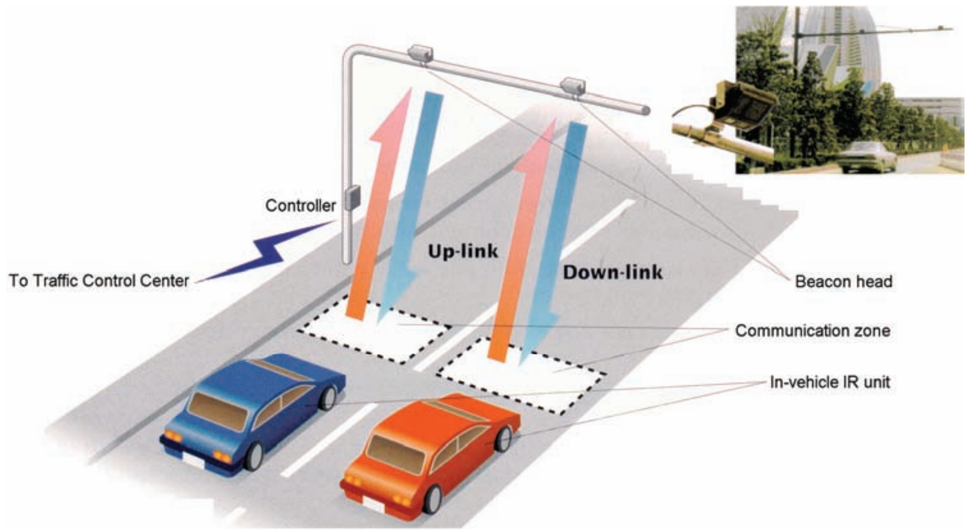


Figure 2-64. Active infrared sensor installed for transmitting traffic conditions to motorists (Photograph courtesy of Sumitomo Electric U.S.A. Inc., Santa Clara, CA).

PASSIVE INFRARED SENSORS

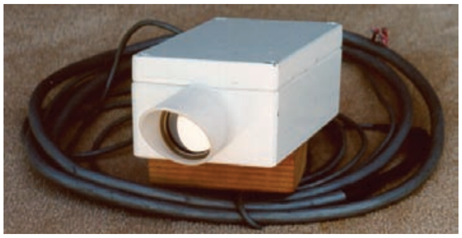
Passive sensors transmit no energy of their own. Rather they detect energy from two sources: (1) energy emitted from vehicles, road surfaces, and other objects in their field of view and (2) energy emitted by the atmosphere and reflected by vehicles, road surfaces, or other objects into the sensor aperture. The energy captured by passive infrared sensors is focused by an optical system onto an infrared-sensitive material mounted at the focal plane of the optics. With infrared sensors, the word detector takes on another meaning, namely the infrared sensitive element that converts the reflected and emitted energy into electrical signals. Real-time signal processing is used to analyze the signals for the presence of a vehicle. The sensors are mounted overhead to view approaching or departing traffic. They can also be mounted in a side-looking configuration. Infrared sensors are used for signal control; volume, speed, and class measurement; detection of pedestrians in crosswalks; and transmission of traffic information to motorists.

Nonimaging passive infrared sensors used in traffic management applications contain one or several (typically not more than five) energy-sensitive detector elements on the focal plane that gather energy from the entire scene. The detector in a nonimaging sensor generally has a large instantaneous field of view. Instantaneous field of view is equal to the angle, e.g., in the x-y plane, subtended by a pixel. Objects within the scene cannot be further divided into subobjects or pixels with this device.

Passive infrared sensors with a single detection zone measure volume and lane occupancy by responding to vehicle passage and presence. Those with multiple detection zones can also measure vehicle speed and length (to the extent that vehicles are detected in one or more detection zones a known distance apart). Imaging sensors, such as modern charge-coupled device (CCD) cameras, contain two-dimensional arrays of detectors, each detector having a small instantaneous field of view. The two-dimensional array gathers energy from the scene over an area corresponding to the field of view of the entire array. Imaging sensors display the pixel-resolution details characteristic of the imaged area, as illustrated in Figure 2-65.[40] An alternative method of gathering two-dimensional images is by scanning one-dimensional arrays over the scene of interest. However, this approach is not currently applied to passive infrared traffic flow sensors.

The emitted energy detected by passive sensors is produced by the nonzero surface temperature of emissive objects in their field of view. Emission occurs at all frequencies by objects not at absolute zero (-459.67 °F or -273.15 °C). If the emissivity of the object is perfect, i.e., emissivity = 1, the object is called a blackbody. Most objects have emissivities less than 1 and, hence, are termed graybodies. Passive sensors can be designed to receive energy at any frequency. Cost considerations make the infrared band a good choice for vehicle sensors with a limited number of pixels. Passive infrared sensors, such as those shown in Figure 2-66, operate in the long-wavelength infrared band from 8 to 14 μm and thus minimize the effects of sun glint and changing light intensity from cloud movement. Passive vehicle sensors operating at microwave frequencies have been evaluated, but their costs are greater.[40,41]

Figure 2-65. Visible spectrum CCD camera imagery of approaching traffic typical of dawn and dusk lighting (Source: C.A. MacCarley, B.M. Hemme, and L.A. Klein, *Advanced Image Sensing Methods for Traffic Surveillance and Detection, Final Report*, Document No. TEL5859F98 (Rev. 1) (California Polytechnic University, San Luis Obispo, CA, Jun. 30, 1998).



Eltec 842 passive infrared vehicle presence sensor. [*Source*: L.A. Klein, *Sensor Technologies and Data Requirements for ITS* (Norwood, MA: Artech House, 2001)].

ASIM IR 250 series passive infrared sensor. This multizone sensor performs vehicle counting, speed measurement, classification by length, and presence detection. (Photograph courtesy of ASIM Technologies, Uznach, Switzerland).

Siemens Eagle PIR-1 sensor. Performs vehicle counting, stop line presence detection, occupancy detection, and queue detection. (Photograph courtesy of Siemens ITS, Austin, TX).

Figure 2-66. Passive infrared sensors

THEORY OF OPERATION

When a vehicle enters a passive sensor's field of view, the detected energy changes due to the presence of the vehicle. The difference in detected energy created by the vehicle is described by radiative transfer theory.(19,31,42) The emissivities of the vehicle and road surface in the wavelength region of interest are denoted by ε_V and ε_R and their surface temperatures in degrees Kelvin by T_V and T_R, respectively, as depicted in Figure 2-67.

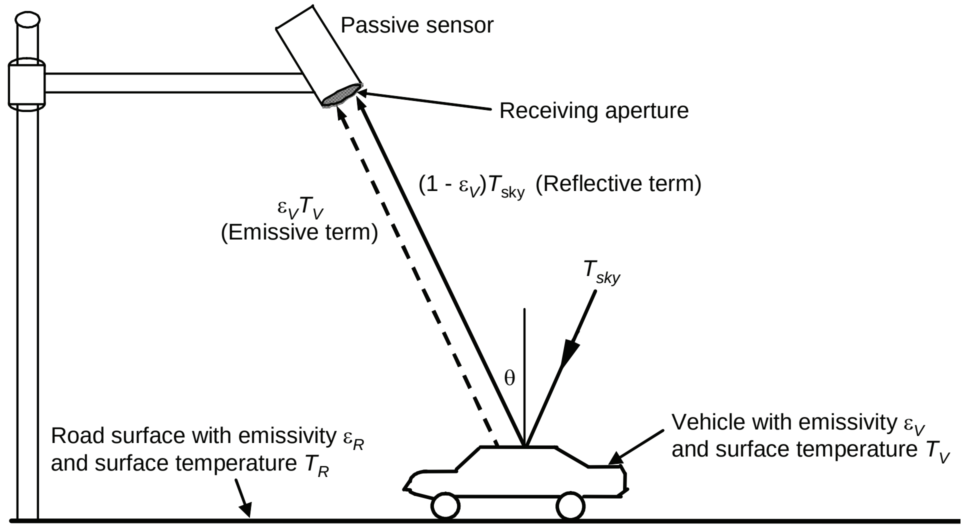


Figure 2-67. Emission and reflection of energy by vehicle and road surface.

The apparent temperature T_{BV} of the vehicle is

$$T_{BV}(\theta, \phi) = \varepsilon_V T_V + (1 - \varepsilon_V)\, T_{sky} \tag{2-70}$$

assuming infrared energy emission by the sensor is negligible. The sky temperature T_{sky} is a function of atmospheric, galactic, and cosmic emission. The atmospheric emission in the infrared spectrum is dependent on the water, ozone, carbon dioxide, nitrous oxide, and methane concentrations. The angles θ and ϕ are the incident angle with respect to nadir (i.e., directly downward) and the angle in the plane of the road surface (the x-y plane), respectively. The term $\varepsilon_V T_V$ represents the energy emitted from the vehicle, referred to as the brightness temperature. The term $(1 - \varepsilon_V)\, T_{sky}$ is the portion of the sky temperature that is reflected from the metal vehicle into the passive sensor. Thus, surfaces that are highly emissive (large ε) have a low reflective component (small $1 - \varepsilon$).

One can also write an expression similar to Equation 2-70 for the apparent temperature of the road surface as

$$T_{BR}(\theta, \phi) = \varepsilon_R T_R + (1 - \varepsilon_R)\, T_{sky}\ . \tag{2-71}$$

The terms in Equation 2-71 have interpretations analogous to those in Equation 2-70. By subtracting the apparent temperature of the vehicle from that of the road, one gets an expression for the temperature difference $\Delta T_B(\theta, \phi)$ sensed by the passive infrared sensor when a vehicle passes through its field of view. Thus

$$\Delta T_B(\theta, \phi) = (\varepsilon_R\, T_R - \varepsilon_V\, T_V) + (\varepsilon_V - \varepsilon_R)\, T_{sky}\ . \tag{2-72}$$

When $T_V = T_R$, then

$$\Delta T_B(\theta, \phi) = (\varepsilon_R - \varepsilon_V)\,(T_R - T_{sky})\ . \tag{2-73}$$

Hence, a vehicle entering the sensor's field of view generates a signal that is proportional to the product of an emissivity difference term and a temperature difference term when the surface temperatures of the vehicle and road are equal. The emissivity term is equal to the difference between the road and the vehicle emissivities. The temperature term is equal to the difference between the absolute temperature of the road surface and the temperature contributed by atmospheric, cosmic, and galactic emission. On overcast, high humidity, and rainy days, the sky temperature is larger than on clear days and the signal produced by a passing vehicle decreases. This, in itself, should not pose a problem to a properly designed passive infrared sensor operating at the longer wavelengths of the infrared spectrum, especially at the relatively short operating ranges typical of traffic management applications.

Fields of view can be tailored through the optical design to accommodate different requirements such as stopline presence detection and presence detection in the approach to an intersection [e.g., a detection zone 68 to 100 ft (20.7 to 30.5 m) in advance of the stopline]. A long focal length lens eliminates adjacent lane detection when sensing vehicles over 100 ft (30.5 m) from the sensor, e.g., as they approach an intersection.

Multichannel (i.e., incorporating more than one type of sensor technology as in Figure 2-74) and multizone (i.e., including more than one detection region) passive infrared sensors measure speed and vehicle length as well as the more conventional volume and lane occupancy. These models are designed with dynamic and static thermal energy detection zones that provide the functionality of two inductive loops. Their footprint configuration is shown in Figure 2-68. The time delays between the signals from the three dynamic zones are utilized to measure speed. The vehicle presence time from the fourth zone is used to calculate the lane occupancy of stationary and moving vehicles.

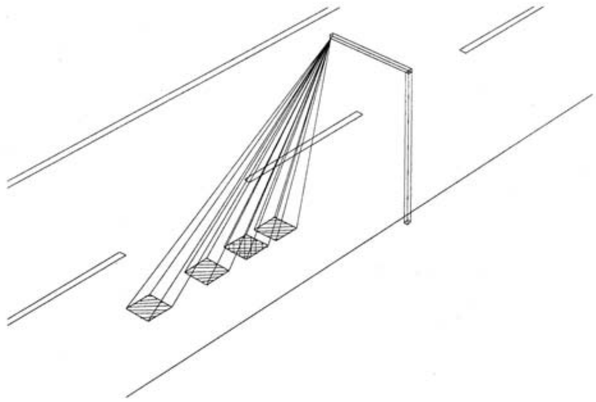


Figure 2-68. Multiple detection zone configuration in a passive infrared sensor.

Several disadvantages of infrared sensors are sometimes cited. Glint from sunlight may cause unwanted and confusing signals in passive sensors operating at near- or midinfrared wavelengths. Atmospheric particulates and inclement weather can scatter or absorb energy that would otherwise reach the focal plane. The scattering and absorption effects are sensitive to water concentrations in fog, haze, rain, and snow as well as to other obscurants such as smoke and dust. At the relatively short operating ranges encountered by infrared sensors in traffic management applications, these concerns may not be significant. However, some performance degradation (i.e., undercounting) in heavy rain and snow has been reported.[40] A rule of thumb for gauging when an infrared sensor may experience difficulty detecting a vehicle is to note if a human observer can see the vehicle under the same circumstances. If the observer can see the vehicle, there is a high probability the infrared sensor will detect the vehicle.

ULTRASONIC SENSORS

Ultrasonic sensors transmit pressure waves of sound energy at a frequency between 25 and 50 kHz, which are above the human audible range. Most ultrasonic sensors, such as the model shown in Figure 2-69, operate with pulse waveforms and provide vehicle count, presence, and occupancy information. Pulse-shape waveforms measure distances to the road surface and vehicle surface by detecting the portion of the transmitted energy that is reflected towards the sensor from an area defined by the transmitter's beamwidth. When a distance other than that to the background road surface is measured, the sensor interprets that measurement as the presence of a vehicle. The received ultrasonic energy is converted into electrical energy. This energy is then analyzed by signal processing electronics that is either collocated with the transducer or placed in a roadside controller.

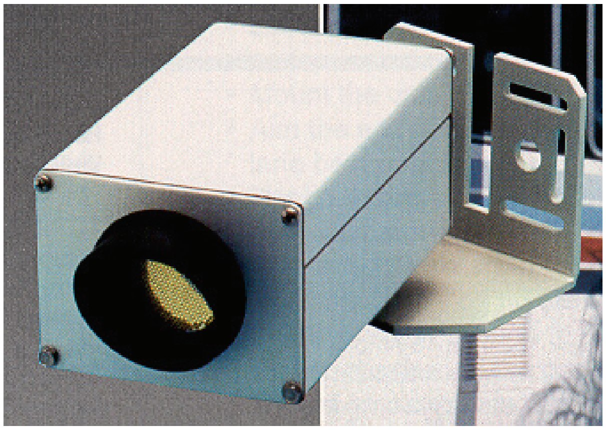


Figure 2-69. TC-30C ultrasonic range-measuring sensor (Photograph courtesy of Microwave Sensors, Ann Arbor, MI).

Pulsed energy transmitted at two known and closely spaced incident angles allows vehicular speed to be calculated by recording the time at which the vehicle crosses each beam. Since the beams are a known distance apart, the speed can be calculated as beam separation distance divided by the time to traverse the beams. The preferred mounting configurations for range-measuring, pulsed ultrasonic sensors are at nadir, looking from an overhead position, and side viewing, as shown in Figure 2-70.

Constant frequency ultrasonic sensors that measure speed using the Doppler principle are also manufactured. However, these are more expensive than pulsed models. The speed-measuring Doppler ultrasonic sensor is designed to interface with the highway infrastructure in Japan. It is mounted overhead facing approaching traffic at a 45-deg incidence angle. It has two transducers, one for transmitting and one for receiving, as illustrated in Figure 2-71. The Doppler ultrasonic sensor detects the passage of a vehicle by a shift in the frequency of the received signal. Vehicle speed can be calculated from the pulse width of an internal signal generated by the sensor's electronics that is proportional to the speed of the detected vehicle.

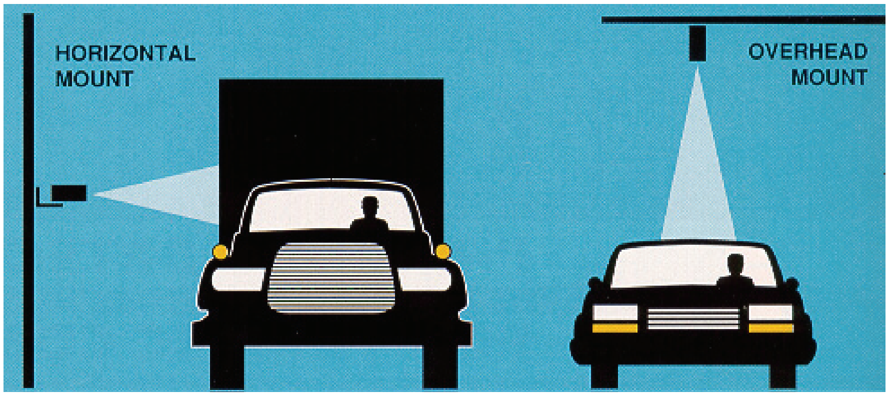


Figure 2-70. Mounting of ultrasonic range-measuring sensors (Illustration courtesy of Microwave Sensors, Ann Arbor, MI).

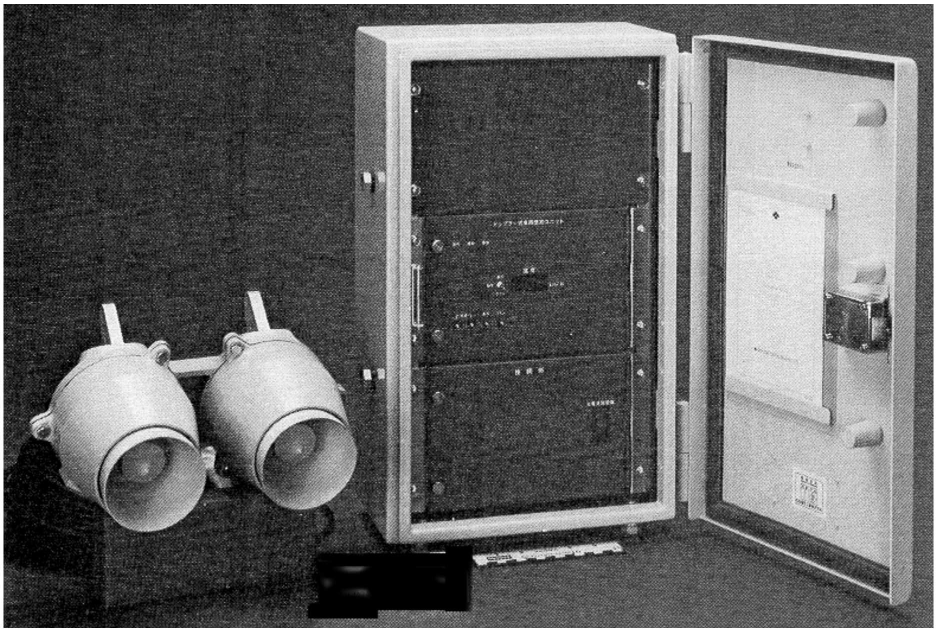


Figure 2-71. Speed-measuring RDU-101 Doppler ultrasonic sensor (manufactured by Sumitomo Electric, Japan) with separate transmitting and receiving transducers, as shown in the left portion of the figure. The cabinet on the right contains the electronic systems needed to supply power and control signals to the transducers, receive the data, and interface with the highway infrastructure.

The range-measuring ultrasonic sensor transmits a series of pulses of width T_p (typical values are between 0.02 and 2.5 ms) and repetition period T_0 (time between bursts of pulses), typically 33 to 170 ms, as illustrated in Figure 2-72. The sensor measures the time for the pulse to arrive at the vehicle and return to the transmitter. The receiver is gated on and off with a user-adjustable interval that differentiates between pulses reflected from the road surface and those reflected from vehicles. The detection gates of various models are adjusted to detect objects at distances greater than approximately 0.5 to 0.9 m above the road surface. This is achieved by closing the detection gate several milliseconds before the reflected signal from the road surface arrives at the sensor.

Automatic pulse repetition frequency control reduces effects of multiple reflections and improves the detection of high-speed vehicles. This control makes the pulse repetition period as short as possible by transmitting the next pulse immediately after the reflected signal from the road is received.[43] A hold time T_h (composite values from manufacturers range from 115 ms to 10 s) is built into the sensors to enhance presence detection.

Ultrasonic sensors are widely used in Japan in keeping with the government policy that discourages pavement cutting on existing highways. In Tokyo, ultrasonic sensors are a major component of the traffic control system. A central computer monitors traffic signals and vehicle motion, resets timing patterns, activates motorist information displays, and relays real-time information to motorists and police. The presence or range-measuring type of sensor is used more extensively than the Doppler type. The Japanese applications appear to be the most extensive use of ultrasonic sensors.

Temperature change and extreme air turbulence may affect the performance of ultrasonic sensors. Temperature compensation is built into some models.

Large pulse repetition periods may degrade occupancy measurement on freeways with vehicles traveling at moderate to high speeds as an insufficient number of pulses is transmitted and reflected from the vehicle while in the sensor's detection zone.

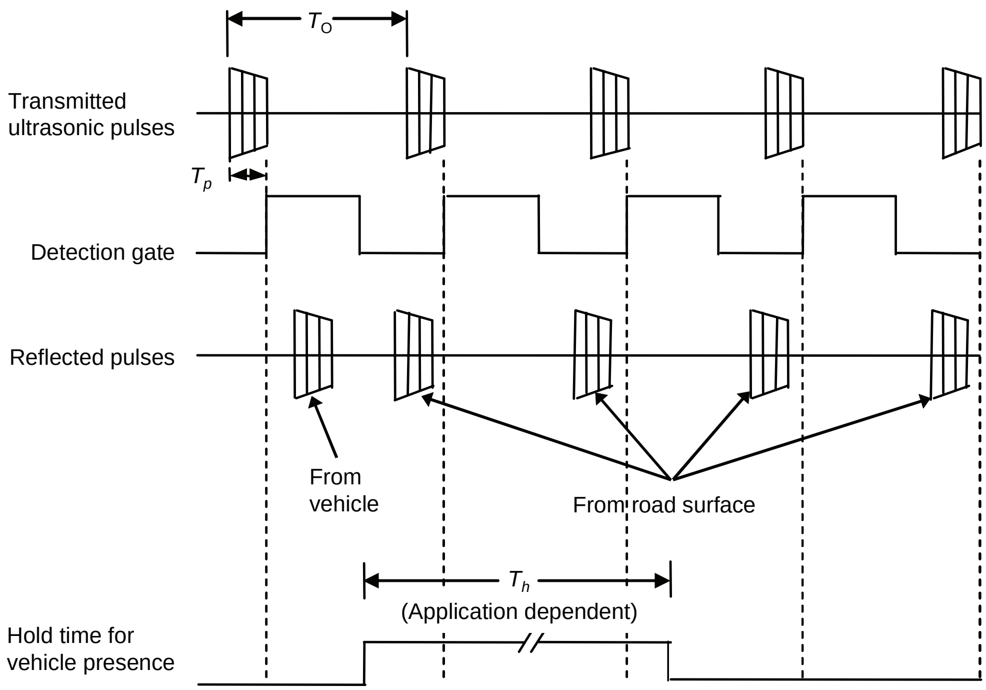


Figure 2-72. Operation of range-measuring ultrasonic sensor.

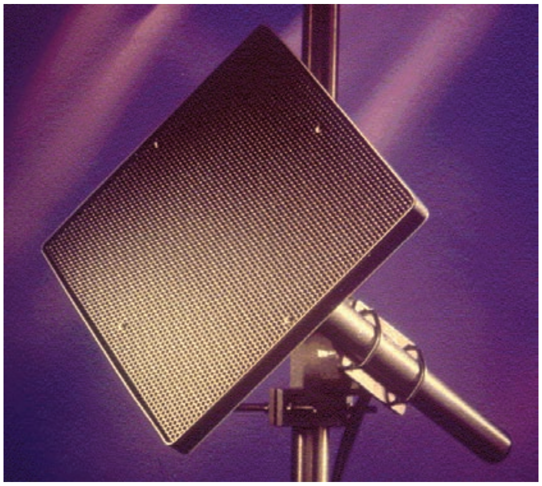
PASSIVE ACOUSTIC ARRAY SENSORS

Acoustic sensors measure vehicle passage, presence, and speed by detecting acoustic energy or audible sounds produced by vehicular traffic from a variety of sources within each vehicle and from the interaction of vehicle's tires with the road. When a vehicle passes through the detection zone, an increase in sound energy is recognized by the signal processing algorithm and a vehicle presence signal is generated. When the vehicle leaves the detection zone, the sound energy level drops below the detection threshold, and the vehicle presence signal is terminated. Sounds from locations outside the detection zone are attenuated.

Single lane and multiple lane models of acoustic sensors are marketed. Both use a two-dimensional array of microphones to detect the sounds produced by approaching vehicles.

SINGLE LANE MODEL OPERATION

The SmartSonic acoustic sensor, shown in the upper part of Figure 2-73, detects vehicles by measuring the time delay between the arrival of sound at the upper and lower microphones, which are arranged in a vertical and horizontal line through the center of the aperture. The time delay changes as the vehicle approaches the array. When the vehicle is inside the detection zone, the sound arrives almost instantaneously at the upper and lower microphones. When the vehicle is outside the detection zone, sound reception at the upper microphone is delayed by the intermicrophone distance.



SmartSonic acoustic sensor (Photograph courtesy of IRD, Saskatoon, SK, Canada).

SAS-1 acoustic sensor (Photograph courtesy of SmarTek Systems, Woodbridge, VA).
Figure 2-73. Acoustic array sensors.

The size and shape of the detection zone are determined by the aperture size, processing frequency band, and installation geometry of the acoustic array. The SmartSonic sensor is tuned to a center frequency of 9 kHz with a 2 kHz bandwidth. Preferred mounting is at 10 to 30 deg from nadir with a detection range of 20 to 35 ft (6 to 11 m).

The speed of a detected vehicle is determined with an algorithm that assumes an average vehicle length. Vehicle presence detection is through an optically isolated semiconductor. A serial interface on the controller card installed with the sensor provides volume, lane occupancy, speed, vehicle classification (cars, light trucks, heavy trucks, and buses), and sensor status messages. When the optional acoustic sensor controller board is installed in a NEMA or 170 cardfile, two detection zones can be used in a speed trap mode to measure vehicle speed. The speed trap activates relay outputs that simulate two inductive loops connected to a NEMA or 170 controller. The SmartSonic is recommended for data collection applications on bridges and other roads where nonintrusive sensors are required, providing either slow moving vehicles (speeds < 20 mi/h (32 km/h)) in stop-and-go traffic flow or free-flow traffic is present. The sensor is not recommended where a mix of stop-and-go and free-flow traffic occurs, such as on a freeway with the potential for congestion, because the vehicle detection algorithm cannot switch between these two flow conditions fast enough to detect the onset of the change in flow.

MULTIPLE LANE MODEL OPERATION

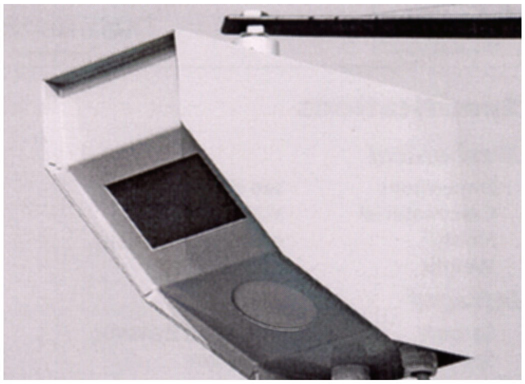
The SAS-1 acoustic sensor in the lower part of Figure 2-73 utilizes a fully populated microphone array and adaptive spatial processing to form multiple detection zones. The SAS-1 is designed to monitor up to 5 lanes when mounted in a side-looking configuration. During setup, the detection zones are steered to positions that correspond to the monitored traffic lanes. The detection zones are self-normalized and polled for vehicles every 8 ms. Detection zones are adjustable to 6 ft (1.8 m) or 12 ft (3.6 m) in the direction of traffic flow and have user-specified values in the cross-lane direction. Acoustic frequencies between 8 and 15 kHz are processed by this sensor, which accommodates mounting heights of 20 to 40 ft (6 to 12 m). The output data are volume, lane occupancy, and average speed for each monitored lane over a user-specified period (e.g., 20 s, 30 s, 1 minute). Vehicle presence is provided by an optional relay interface.

SENSOR COMBINATIONS

Figure 2-74 illustrates sensors that combine passive infrared presence detection with ultrasound or CW Doppler microwave radar.[44] The passive infrared-ultrasonic combination, shown in the upper portion of the figure, provides enhanced accuracy for presence and queue detection, vehicle counting, and height and distance discrimination.

The passive infrared-CW Doppler radar sensor, in the lower portion of the figure, is designed for presence and queue detection, vehicle counting, speed measurement, and length classification. It relies on the radar to measure high to medium vehicle speeds and the passive infrared to measure vehicle count and presence. At medium speeds, the multiple detection zone passive infrared automatically calibrates its speed measurements against the radar's. This calibration permits the infrared to measure slow vehicle speeds and detect stopped vehicles.

ASIM DT 272 Passive infrared-ultrasonic sensor.



ASIM DT 281 Passive infrared-CW Doppler microwave radar sensor.

Figure 2-74. Passive infrared combination sensors (Photographs courtesy of ASIM Technologies, Uznach, Switzerland).

REFERENCES

1. Ingram, J.W. *The Inductive Loop Vehicle Detector: Installation Acceptance Criteria and Maintenance Techniques*, Report No. TL 631387, prepared by California Department of Transportation, Sacramento, CA, for Federal Highway Administration, Washington, DC. U.S. Department of Commerce, National Technical Information Service, PB-263 948, Washington, DC, March 1976.
2. Gordon, R.L., R.A. Reiss, H. Haenel, E.R. Case, R.L. French, A. Mohaddes, and R. Wolcott. *Traffic Control Systems Handbook*, FHWA-SA-95-032. U.S. Department of Transportation, Federal Highway Administration, Washington, DC, February 1996.
3. *Test Results of "Polyguard" as a Portable Loop Covering and Portable Loops in a 4- x 6-ft Configuration*, Research Report. Special Studies Section, Nevada Department of Transportation, October 1988.

4. *Model IVS-2000 Installation and Operation Manual*, Rev. 1.53. Intersection Development Corp., Downey, CA, September 1997.

5. *NEMA Standards Publication No. TS-1: Traffic Control Systems.* National Electrical Manufacturers Association, 1989.

6. *Traffic Signal Control Equipment Specifications.* State of California Department of Transportation, Sacramento, CA, January 1989.

7. *Traffic Control Hardware Specifications*, Release 1.4. New York State Department of Transportation, Albany, NY, June 1, 1990.

8. *Advanced Transportation Controller (ATC) Standard for the Type 2070 Controller*, ATC 2070 V02.03. Joint Committee on the ATC Ballot Copy for Joint Adoption by AASHTO, ITE, and NEMA, March 12, 2004.

9. Workmon, F. *Transportation Electrical Equipment Specifications* (TEES). State of California Department of Transportation, Sacramento, CA, November 19, 1999.

10. Thomas, P. *Lightning Problems on 170/332 Traffic Control Systems-Investigative Report.* Ontario Ministry of Transportation and Communications, Downsview, Ontario, Canada, June 1985.

11. Klein, L.A., and M.R. Kelley. *Detection Technology for IVHS*, *Vol. I*: *Final Report*, FHWA-RD-95-100. U.S. Department of Transportation, Federal Highway Administration, Washington, DC, December 1996.

12. Hilbert, E.E., P.A. Rennie, and W.A. Kneidl, “A Sensor for Control of Arterials and Networks,” *IEEE Trans. Vehicular Technology*, VT-29(2), May 1980, pp. 208–215.

13. Michalopoulos, P.G. *Machine Vision System for Traffic Surveillance and Control—Summary and Status.* University of Minnesota, Department of Civil and Mineral Engineering, Minneapolis, MN, 1988.

14. MacCarley, C.A., S. Hockaday, D. Need, and S. Taff. “Evaluation of Video Image Processing Systems for Traffic Detection,” *Transportation Research Record No.* 1360. Transportation Research Board, National Research Council, Washington, DC, 1992.

15. Jet Propulsion Laboratory (JPL). *Traffic Surveillance and Detection Technology Development*, *Final Report*, JPL publication 97-10. California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, March 1997.

16. Zeng, N., and J.D. Crisman. “Evaluation of Color Categorization for Representing Vehicle Colors,” *Proc. SPIE*, vol. 2902. Bellingham, WA, 1996, pp. 148–155.

17. Chachich, A., A. Pau, A. Barber, K. Kennedy, E. Olejniczak, J. Hackney, Q. Sun, and E. Mireles. “Traffic Sensor Using a Color Vision Method,” *Proc. SPIE*, vol. 2902. Bellingham, WA, 1996, pp. 156–165.

18. Wentworth, J., C. Dougan, D. Green, W. Kaufman, E. Kent, T. O’Keefe, and H. Wang. *International Scanning Report on Advanced Transportation Technology*, FHWA-PL-95-027. Department of Transportation, Federal Highway Administration, Washington, DC, December 1994.

19. Klein, L.A. *Millimeter-Wave and Infrared Multisensor Design and Signal Processing*. Artech House, Norwood, MA, 1997.
20. Chang, E. C-P, and H. Kunhuang. "Incident Detection Using Advanced Technologies," Paper 930943, 72nd Annual Meeting. Transportation Research Board, National Research Council, Washington, DC, 1993.
21. Gil, S., R. Milanese, and T. Pun. "Feature Selection for Object Tracking in Traffic Scenes," *Proc. SPIE*, vol. 2344. Bellingham, WA, November 1994, pp. 253–266.
22. Coifman, B., D. Beymer, P. McLauchlan, and J. Malik. "A Real-Time Computer Vision System for Vehicle Tracking and Wide-Area Traffic Surveillance," 78th Annual Meeting. Transportation Research Board, National Research Council, Washington, DC, 1999.
23. Beymer, D., and J. Malik. "Tracking Vehicles in Congested Traffic," *Proc SPIE*, vol. 2902. Bellingham, WA, November 1996, pp. 8–18.
24. Rhodes, A., D.M. Bullock, J. Sturdevant, Z. Clark, and D.G. Candey, Jr. "Evaluation of Stop Bar Video Detection Accuracy at Signalized Intersections," 84th Annual Meeting. Transportation Research Board, National Research Council, Washington, DC, 2005.
25. Video Image Processor Data Sheets. Nestor Corp., Providence, RI, 1999.
26. *Traffic Surveillance and Detection Technology Development*: *Phase II and III Test Results*, *Troy, MI*, JPL Pub. D-15779. California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, May 1998.
27. Klein, L.A. *Final Report*: *Mobile Surveillance and Wireless Communication Systems Field Operational Test—Vol. 1*: *Executive Summary*, California PATH Res. Rpt. UCB-ITS-PRR-99-6. University of California, Berkeley, Richmond, CA, March 1999.
28. Klein, L.A. *Final Report: Mobile Surveillance and Wireless Communication Systems Field Operational Test—Vol. 2: FOT Objectives, Organization, System Design, Results, Conclusions, and Recommendations*, California PATH Res. Rpt. UCB-ITS-PRR-99-7. University of California, Berkeley, Richmond, CA, March 1999.
29. Mastako, K.A., and L.A. Klein. "Evaluation of Mobile Surveillance and Wireless Communication Systems FOT," *Transportation Research Record* 1683. Transportation Research Board, National Research Council, Washington, DC, 1999, pp. 14–21.
30. "The Use of Video Detection at Signalized Intersections," *The Urban Transportation Monitor*. Lawley Publications, Burke, VA, December 2004.
31. Klein, L.A. *Sensor Technologies and Data Requirements for ITS*. Artech House, Norwood, MA, 2001.
32. Hartmann, D., D. Middleton, and D. Morris. *Assessing Vehicle Detection Utilizing Video Image Processing Technology*, FHWA/TX-97/1467-4. Texas Transportation Institute, College Station, TX, 1996.

33. Middleton, D., and R. Parker. *Initial Evaluation of Selected Detectors to Replace Inductive Loops on Freeways*, FHWA/TX-00/1439-7. Texas Transportation Institute, College Station, TX, April 2000.
34. *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies*, vol. 5, Task Three Report: Extended Field Test Results. Minnesota Guidestar and SRF Consulting Group, St. Paul, MN, 1996.
35. Hockaday, S. *Evaluation of Image Processing Technology for Applications in Highway Operations Final Report*, TR 91-2. California Polytechnical State University, Transportation Research Group, San Luis Obispo, CA, June 29, 1991.
36. Ponce, L.A., A. Chatziioanou, S. Hockaday, E. Sullivan, and S. Kaighn. "Field Tests for Freeway Traffic Video Image Processing Systems Installations—A Preliminary Examination of VIPs Standardization," in *Proc. of the National Traffic Data Acquisition Conference* (NATDAC 1994), vol. II, FHWA Report FHWA-CT-RD 2304-F2-94-1. Connecticut Department of Transportation, Rocky Hill, CT, Sept. 18-22, 1994, pp. 551–562.
37. Edde, B. *Radar: Principles, Technology, Applications.* PTR Prentice Hall, Englewood Cliffs, NJ, 1993.
38. Schwartz, W.C. "Laser Vehicle Detector/Classifier," *Proc. SPIE*, vol. 2344. Bellingham, WA, November 1994, pp. 81–87.
39. Schwartz, W.C., and R.A. Olson. "Wide-Area Traffic-Surveillance (WATS) System," *Proc. SPIE*, vol. 2902. Bellingham, WA, November 1996, pp. 124–133.
40. MacCarley, C.A., B.M. Hemme, and L.A. Klein. *Advanced Image Sensing Methods for Traffic Surveillance and Detection, Final Report*, document no. TEL5859F98 (revision 1). California Polytechnic University, San Luis Obispo, CA, June 30, 1998.
41. Kranig, J., E. Minge, and C. Jones. *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies*, FHWA-PL-97-018. U.S. Department of Transportation, Federal Highway Administration, Washington, DC, May 1997.
42. Ulaby, F., R.K. Moore, and A.K. Fung. *Microwave Remote Sensing: Active and Passive, Vol. I, Microwave Remote Sensing Fundamentals and Radiometry.* Artech House, Norwood, MA, 1981.
43. Kumagai, Y., T. Yamamoto, M. Deguchi, and S. Yamaoka, *Ultrasonic Detector and New Type Sensors for Urban Traffic Control Systems.* Sumitomo Electric Industries, 1992.
44. *DT 281 data sheet.* ASIM Technologies, Ltd., Uznach, Switzerland.

TABLE OF CONTENTS: CHAPTER 3.

LIST OF FIGURES: CHAPTER 3.

LIST OF TABLES: CHAPTER 3.

CHAPTER 3. SENSOR APPLICATIONS

Sensor applications to traffic control and management continue to evolve and escalate. Originally utilized for signalized intersection control, sensors are now used to supply real-time data for traffic adaptive signal control and mitigating recurring and nonrecurring congestion on freeways. Many advances in traffic control system technology during the past decade have been supported by the evolution of microprocessors and other electronics components. The relative ease with which research and widespread user knowledge can be retrieved has also assisted agencies in selecting appropriate sensor technologies and deployment configurations to satisfy their operational needs. The Internet with its convenient access to public and private libraries that contain evaluation reports of sensor performance and traffic management strategies enables the rapid sharing of test and operational experiences.

Traffic control concepts that rely on traffic flow data from sensors are common to central business district, arterial, and freeway traffic management strategies. The data needs of the applications can generally be satisfied by one or more sensor technologies. This chapter presents an overview of several of these traffic management strategies so that the traffic engineer or other practitioner can better understand their detection requirements. Additional information on how to use traffic sensors in traffic signal control systems can be found in two recommended references: *Manual of Traffic Signal Design*[(1)] and *Traffic Control System Handbook.*[(2)] The *Freeway Management and Operations Handbook*[(3)] may be consulted for information concerning freeway operation and appropriate freeway traffic management strategies.

PRESENCE AND PASSAGE SENSORS

Presence sensors detect slow moving and stopped vehicles. Passage sensors detect vehicles moving faster than 3 to 5 mi/h (5 to 8 km/h).

Most vehicle sensors in use today monitor the movement of vehicles past a given point on the road. The data acquired are transmitted to a traffic signal controller, traffic counter, or other device. The controller or counter processes some data locally, while others are transmitted to a central computer or display monitor, in the case of camera imagery, at a traffic management center.

A sensor that detects the passage of vehicles in a specified direction may be used to issue a warning to alert the driver of a vehicle traveling in the opposite or forbidden direction. A presence sensor is generally used at locations where vehicle speeds are less than approximately 5 mi/h (8 km/h) or where stopped vehicle detection is required. A passage (motion-detecting) sensor will record the passage of a vehicle in the detection zone as long as the vehicle is moving more than 3 to 5 mi/h (5 to 8 km/h).

When vehicles are forced to stop or move very slowly when approaching a traffic-actuated signalized intersection, it is desirable to use presence detection to ensure that stopped vehicles waiting in the detection zone are detected. Early inductive-loop electronics units operating in the presence mode did not have a directional feature and could not differentiate between vehicles entering or leaving the detection zone. Some newer units contain this feature. Presence sensors that use over-roadway technologies such as video image processing incorporate directional features. At intersections, there may be less need for a directional feature, as the call is not normally retained in the controller once the vehicle leaves the detection zone. Passage sensors will not retain the call either once the vehicle leaves the detection zone. The difference is in controller operation (i.e., whether the controller is in locking or nonlocking mode).

Passage sensors may be used to count vehicles in individual lanes, multiple lanes where traffic flows in a particular direction, or all lanes where traffic flows in both directions. Passage sensors also measure vehicle speed and volume when speeds are greater than 3 to 5 mi/h (5 to 8 km/h).

Sensors based on video image processing, microwave radar, laser radar, passive infrared, ultrasound, passive acoustic, magnetometer, and inductive-loop sensor technologies discussed in Chapter 2 detect either vehicle presence or passage, depending on their particular design or the selected operating mode.

Modern traffic sensors can record traffic flow parameters by direction and by lane. Accurate arterial vehicle count and speed data are obtained by locating the sensors far enough in advance of the signalized intersection so that traffic does not back up to the detection zone. If traffic backs up to the sensor, only volume is measured.

Magnetometers can be used to count vehicles, provided the detection zone is placed in the desired portion of the lane of interest. Since magnetometer detection zones are generally less than 3 ft (1 m) in diameter, multiple units may be needed to detect vehicles over the entire lane width.

A properly functioning inductive-loop detector is an excellent sensor for detecting vehicle presence, providing it is properly installed and maintained. Loop size can be varied to accommodate different applications. For small area detection, the conventional loop and two-axis fluxgate magnetometer may be interchanged.

Some techniques for high-speed intersection signal control utilize conventional loops or magnetometer sensors with electronics units that are programmed for normal outputs. Other designs use electronic units that extend or hold the call of the vehicle after it leaves the detection zone (extended call detectors or extended call sensors). Another plan incorporates electronic units that delay an output until the detection zone has been occupied for a preset time (delayed call detectors or delayed call sensors). These are discussed in Chapter 4.

Over-roadway sensors are becoming more popular as sources of real-time data for signal control and freeway traffic management. This is because of their ability to provide multiple lane data from a single sensor, reduced maintenance, increased safety to installation personnel, data types not available from loops or magnetometers, and competitive purchase and installation costs.

SPEED MONITORING WITH INDUCTIVE-LOOP AND OTHER POINT SENSORS

A 1980 study suggested that the optimum characteristics of a sensor used for vehicle speed measurement were:[4]

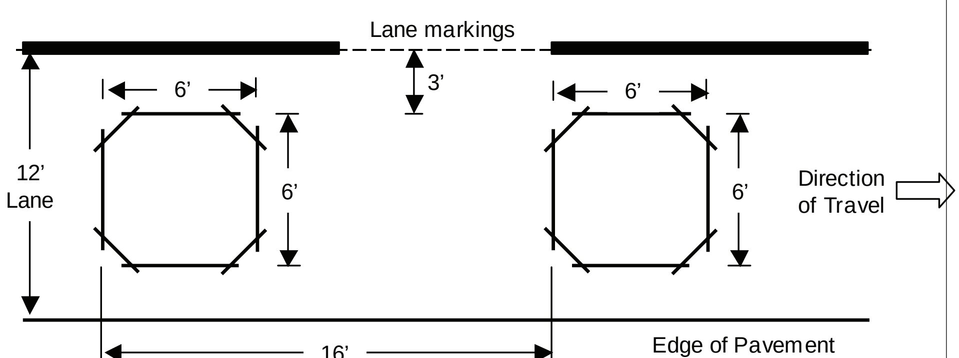
- Self-tuning electronics to reduce drift.

- Short response time from intercept to output.
- High sensitivity without appreciable time delay.
- Consistency of vehicle location at beginning and end of detection, independent of vehicle speed or length of lead-in cable.

When using two loops to measure speed, the loops should be large enough to sense high-body vehicles and provide a sharply defined wave front output as the vehicle passes over the loop. Any time differences in the detection of different vehicle types passing over the loops should be minimized. The loops should be spaced sufficiently far apart so that any difference in the time of intercept of the two inductive-loop detectors is small as compared to the transit time from the first loop to the second loop.

A rule of thumb for loop design states that the height of the magnetic field intercepted by the vehicle is two-thirds the distance of the shorter loop dimension. Therefore, a 6- x 6-ft (1.8- x 1.8-m) loop has intercepts of approximately 4 ft (1.2 m), as does a 6- x 100-ft (1.8- x 30.5-m) loop. Both 5-ft (1.5-m) wide and 6-ft (1.8 m) wide loops have proven effective at consistently detecting high-body vehicles. The choice depends on lane width. A spacing of at least 2.5 ft (0.8 m) should be allowed from the centerline to the edge of the loop to avoid actuation by traffic in adjacent lanes. In a 12-ft (3.6-m) lane, the 6-ft (1.8-m) loop should be used to ensure no counts are missed.

The spacing between loops for speed measurements is often specified as 16 ft (4.9 m) between the leading edges of two 6-ft (1.8-m) loops as shown in Figure 3-1. This sensor layout also applies to any pair of point sensors that might be used in a dual sensor speed-trap configuration. The sensitivity of the electronics unit connected to each loop must be the same. If not, the vehicle position with respect to the loop's leading edge, which induces the critical change in loop inductance needed to activate the electronics unit and which is proportional to the sensitivity, will vary from loop to loop, thereby introducing a measurement error. In the late 1980s, very fast response times at sensitivity levels appropriate for roadway vehicles were made possible by new electronic component technology.



1 ft = 0.3 m

Figure 3-1. Vehicle speed measurement using two inductive-loop detectors placed a known distance apart.

TRAFFIC MANAGEMENT CONCEPTS FOR FREEWAYS

Freeways were originally conceived as limited access, free-flowing facilities with little need of traffic control. Rapid growth in freeway traffic volume and resulting congestion have led to development of freeway surveillance and control systems. These systems employ techniques such as ramp control, mainline control, corridor control, and traveler information systems to mitigate freeway congestion. Strategies used today to assist in the management of freeways include:

- Restricted entry (ramp closure, ramp metering).
- Priority treatment HOV and bus priority lanes).
- Surveillance.
- Incident management (detection, identification, and response).
- Display and broadcast of advisory information (speed, travel time, route guidance, diversion).

Each city street arterial and freeway traffic management concept has its own set of operating parameters and components, including sensors, as discussed below in the sections on signal control concepts for city streets, freeway surveillance and control, and coordinated operation of freeways and surface arterials.

SIGNAL CONTROL CONCEPTS FOR CITY STREETS

Traffic signal control concepts for city street intersections may be grouped into strategies for individual intersections and strategies for groups of intersections as follows:

- Strategies for individual intersections:
  - Pretimed.
  - Traffic actuated.
- Strategies for groups of intersections:
  - **Uncoordinated control**: Traffic flow is controlled without considering the operation of adjacent traffic signals.
  - **Time-based coordinated control (TBC)**: TBC systems have no system sensors. These provide signal progression that allows platoons of vehicles to proceed along arterial routes without stopping. It can also provide area-wide control to minimize total delay and number of stops over an entire network. Control information on the time-of-day/day-of-week (TOD/DOW) plan is provided by means of a real time clock. Most modern controllers have this capability built in.
  - **Interconnected control**: Basic coordination is provided by wireline or wireless interconnect. This information is used to determine the local signal timing of offset and cycle/split or actuated timing based on TOD/DOW. The operator can select timing and download timing plans and changes as well as monitor and record system status. System response to traffic changes is measured in weeks or months.

- Traffic adjusted control: Systems operate with few sensors (or at sensor web density levels of 1.0 and 1.5) and use fixed timing plans, the selection of which is adjusted by a minimum number of system sensors at periodic intervals of 15 minutes or more. The sensors in these systems measure system wide variations in traffic demand using either volume plus weighted occupancy or predominant direction of flow. They then use pattern matching to choose from among a set of precalculated plans. This definition builds on the original definition by Kell and Fullerton in the *Manual of Traffic Signal Design*,[1] which in turn built on that of Automatic Signal. System response to traffic changes is relatively slow.

Sensor web density levels and their relation to traffic control systems are described in Appendix L.

- **Traffic responsive control**: Systems operate with at least one sensor per link up to one sensor per lane per link. They use flexible timing plans in which the offsets, splits, and phase durations can be promptly changed in reaction to changes in traffic on a nearly cycle-by-cycle basis. Cycle, split, and offset are each optimized in the selection or calculation of the flexible timing plan to be implemented. These systems typically use macroscopic measures of traffic flow on individual links such as platoon and other characteristics. System response to traffic changes is prompt.
- **Traffic adaptive control**: Systems operate with two sensors per lane per link for optimum efficiency. Adaptive systems do not have cycle, split, and offsets in the classic sense. They forecast traffic into the near future and proactively reoptimize selection, sequence and duration of phases every several seconds. These systems measure the flow of individual vehicles to predict the future flow of the individual vehicles. System response to traffic changes is proactive.

- Strategies for groups of intersections can be implemented through the selection of timing plans from among a library of prestored fixed timing plans that best match current traffic flow conditions, from fixed or flexible timing plans generated online in real time based on current traffic flow. Flexible timing plans are updated incrementally once each cycle.
  - Open network arterial control: Primary consideration is the progressive flow of traffic along the arterial by operating as a system.
  - Closed network control: Applied to a group of adjacent signalized intersections whose signal operations are coordinated, e.g., the control of signals in a central business district (CBD).
  - Areawide system control: Treats all of the traffic signals within an area, a city, or a metropolitan area as a total system. Individual signals within an area may be controlled by isolated, open network, or closed network concepts.

Signal control concepts for special functions include:

- Preemption and priority vehicle signal control: Preempts or alters a normal signal sequence for the movement of emergency and transit vehicles, respectively.

- Driver alerts and warnings: Lighted signs that alert drivers to a red signal ahead and warn of excess vehicle speed when approaching sharp curves or other roadway conditions that require reduced speed or other caution.
- Pedestrian signal actuation: Devices that request the pedestrian crossing phase and alert drivers to pedestrians in a crosswalk.

STRATEGIES FOR LOCAL INTERSECTION SIGNAL CONTROL

Types of local intersection signal control include pretimed control and actuated control.

Local intersection control is the strategy residing in the local controller that manages traffic flow independently of other traffic signals. Two types of local control exist, pretimed and actuated.[2,5] The type of control selected is frequently subject to local policy and practice. If the local control is run independently of strategies for other intersections, the system control is called "uncoordinated control." Since each signal in uncoordinated control or isolated intersection control operates independently, offset is not a controlled parameter when isolated intersection control is implemented.

Pretimed Control

Sensors are not required for pretimed control when right-of-way is assigned based on a predetermined fixed time duration, as determined from historic data, for all signal display intervals. Therefore, pretimed control is generally inefficient for controlling intersections that undergo changes in demand. Pretimed control may be used in conjunction with traffic adjusted or traffic responsive timing plan selection where closely spaced signals dictate fixed offsets, such as with diamond interchanges or central business districts.

Actuated Control

Actuated control utilizes sensors to provide data to a local traffic signal controller as illustrated in Figure 3-2. Sensors are typically located at stoplines (A), upstream of the stopline (B), left turn lanes (C), and at positions to detect emergency (D) and transit vehicles (E). Inductive-loop detectors are the most common sensor used for this application, although multiple-lane, multiple-detection zone sensors such as the video image processor or true presence microwave radar may have merit for this type of signal control. The suitability of over-roadway sensors for a specific application should be evaluated through field testing by the responsible agency to ensure that the required calls are provided reliably. The information gathered by the sensors can be processed as indicated in Figure 3-3 or in another manner, depending on the particular traffic management requirements and strategies.

Actuated control can be semiactuated or fully actuated. In semiactuated control, the major street operates in a nonactuated mode such that green is always present unless a minor street actuation is received. Therefore, sensors are required only for the minor cross-street phases. In the absence of cross-street demand, semiactuated signals are recalled to the major street phase. Semiactuated operation is appropriate when vehicles on the minor streets approach the intersections in a random manner, that is, where platoons (groups of closely spaced vehicles traveling at the same speed) cannot be sustained. Such a condition is likely where there are long distances between signalized intersections, unpredictable or relatively low minor-street volumes (e.g., less than 20 percent of volumes on the major street), and a large proportion of turning movements.

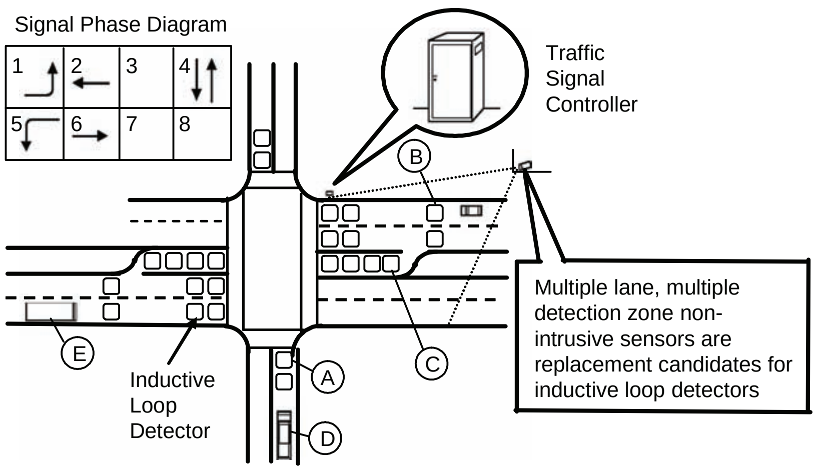


Figure 3-2. Isolated intersection control. The letters represent data sources that influence signal timing as explained in the text.

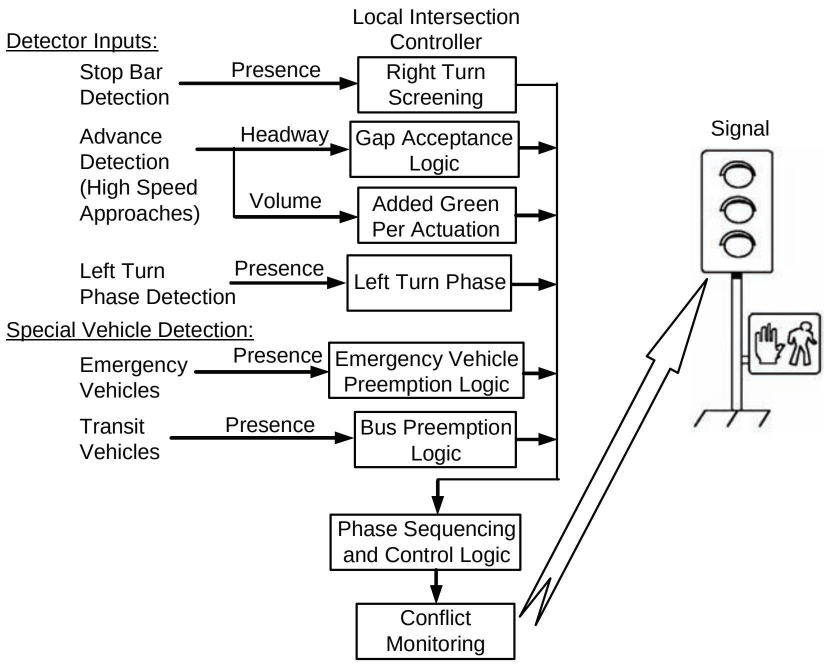


Figure 3-3. Data processing at an intersection with isolated intersection control.

In Figure 3-2, letter A indicates placement of stop bar detection sensors that are used to detect vehicles that turn right and thus avoid the need to call the green. Letter B indicates advance detection sensors used for measuring headways for gap acceptance logic, red light running minimization, and measuring volumes for added green per phase. Letter C indicates left turn phase detection sensors that measure presence and sometimes queue length of vehicles needing to call left turn phase logic. Letter D indicates emergency vehicle sensors that detect the approach of emergency vehicles and invoke emergency preemption logic. Letter E represents transit vehicle sensors, which invoke transit vehicle priority logic. All of the sensors provide traffic flow data used to actuate appropriate phases of isolated intersection signal control. The data enter a traffic signal controller located at one of the corners of the intersection. The controller uses logic processing, illustrated in Figure 3-3, to display the correct indications on the traffic signals and pedestrian signals.

Fully actuated control operates with traffic detection on all approaches to the intersection for all signal phases. It is the most widely applied control strategy for isolated intersections. Because the cycle length varies from cycle to cycle, it can be utilized at street intersections with sporadic and varying traffic distribution.

Volume-density control, a variant of actuated control, provides a complex set of criteria for allocating green time (“added initial” and “time waiting-gap reduction”). This mode can be utilized with both semiactuated and fully actuated signals. It normally operates on a continuously variable cycle length and requires accurate traffic flow data to accommodate changing conditions in a timely manner. Although the time waiting gap reduction can be utilized with presence sensors, passage sensors are normally installed far in advance of the intersection, e.g., from 200 to 600 ft (60 to 180 m) before the intersection, depending on approach speed.

Figure 3-4 shows a typical detection zone placement for fully actuated intersection control with advance sensors only on the primary phases. The sensors near the stopline detect vehicles that would otherwise be trapped in the crosswalks or in front of the stopline. Passage and advance sensors are normally located in accordance with estimated approach speeds to place a call to the controller until the phase is serviced.

The controller provides memory latching when passage sensors are used. Passage sensors generate a pulsed output (generally between 100 and 150 ms) whenever a vehicle enters the sensor's detection zone.[2]

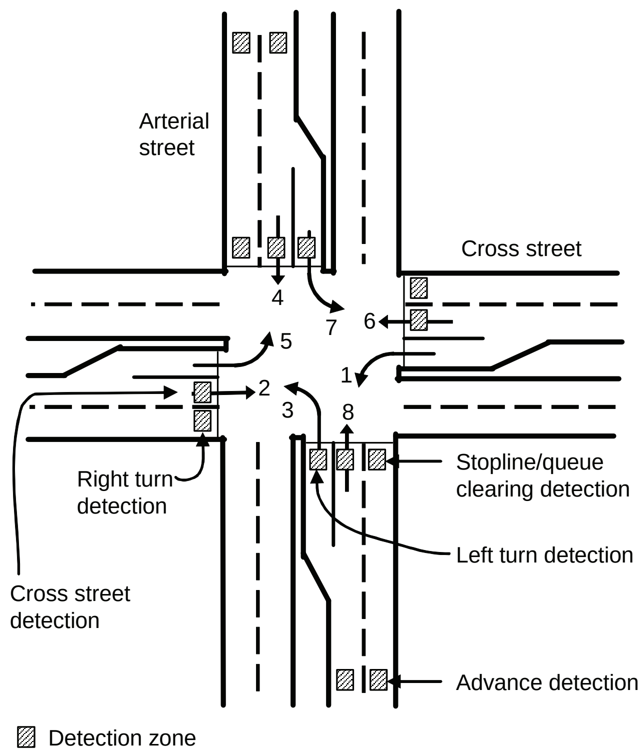


Figure 3-4. Placement of sensors for fully actuated intersection control. The numbers represent a four-phase signal sequence consisting of traffic movements 1 and 5, 2 and 6, 4 and 8, and 3 and 7.

STRATEGIES FOR CONTROL OF GROUPS OF INTERCONNECTED INTERSECTIONS

Interconnected intersection control provides signal progression that allows platoons of vehicles to proceed along arterial routes without stopping. It also imparts area-wide control to minimize total delay and number of stops over an entire network. This form of control is effective when traffic moves in platoons and their arrival time can be predicted at downstream intersections. Interconnected intersection control can function in a variety of ways. The first category selects from among a library of prestored signal timing plans based on TOD/DOW(classic pretimed). The prestored plans are generated offline from average or historical data. The second uses timing plans that best match current traffic flow conditions which may be generated online (see

1.5 GENERATION and *SECOND GENERATION* sections). The third uses flexible signal timing plans that are generated online in real time based on current traffic flow conditions based on centralized control with fixed offsets. These traffic responsive may update their flexible timing plans incrementally at each signal cycle. The fourth uses continuous adjustments to intersection timing with continuous communications between adjacent intersections and decentralized control. (See section titled Traffic Adaptive Control and Distributed Systems). These categories are detailed in Table 3-1.

Sensor Data

Sensors are utilized in interconnected intersection signal control to gather traffic flow data for signal timing plan selection or real-time calculation and to support critical intersection control. The signal timing selection process is similar for arterial and network systems. The operation of signal timing plans is determined by the roadway configuration and the goals of the corresponding plans. Figure 3-5 shows examples of surface street traffic signal configurations found in arterial open network systems and closed network systems typified by central business districts. An open network typically has coordination timing constraints on only two of the approaches to the signal. This is typical of arterial and multiple arterial systems. A closed network typically has constraints on all approaches to each intersection. This is typical of central business district systems and multiple parallel arterial systems.

Arterial Systems

Arterial signal control systems are often implemented using an open network. The goals of arterial system timing plans are to provide arterial progression in the direction that carries heavier traffic volumes, to maximize arterial capacity, and to minimize arterial delay. Cycle length, split, and offset timing plan parameters are varied to reflect the current traffic conditions. The maximum phase times for minor cross streets in arterial systems are often controlled by pedestrian crossing time requirements. Detection of cross-street demand was discussed earlier as a part of isolated intersection control.

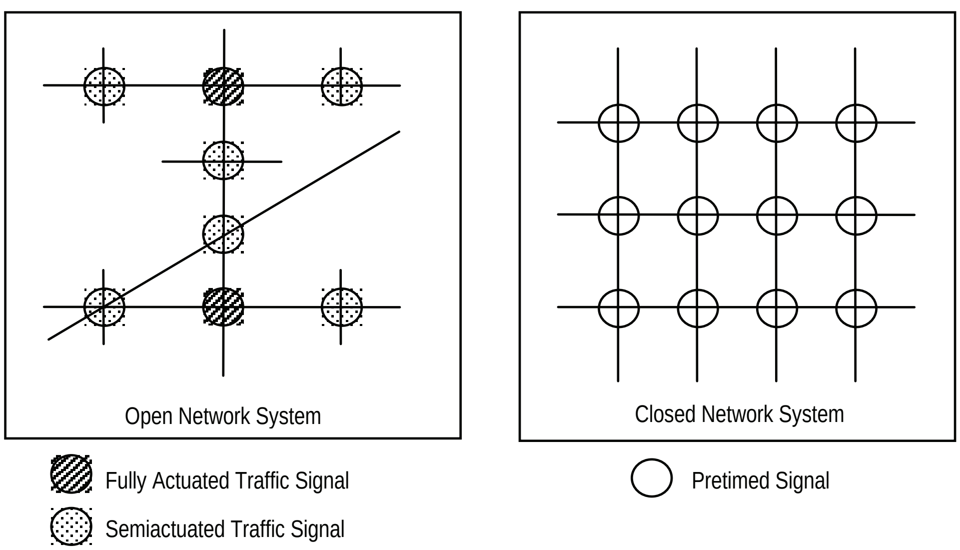


Figure 3-5. Arterial open network and surface street closed network traffic signal configurations typical of those found in interconnected intersection control.

Interconnected Control and Classic Pretimed Network Systems

These network systems typify the era before urban traffic control system (UTCS) software was available. Two types of network systems were used: a closed-grid roadway configuration typical of urban central business districts and an area-wide system that controlled all or a major portion of the signals in a city or metropolitan area. The goals of network system timing plans were essentially the same as those for arterial systems.

Many urban jurisdictions installed pretimed and semiactuated traffic signals for a closed network. Vehicle detection sensors were used at the semiactuated intersection. As with the arterial system, cycle length, split, and offset varied from plan to plan. The plans could be computed offline and implemented on a TOD/DOW basis as in UTCS 1st-generation software; computed offline and implemented on a traffic-adjusted basis as in UTCS 1.5-generation software; or computed online in real time as with 2nd-generation software. In all cases, the plans were pretimed with sensor inputs used to generate the data needed to calculate the pretimed plans.

Timing Plan Implementation

Timing plans were normally implemented by dividing a system into a number of sections, each of which had homogeneous traffic conditions. System sensors, assigned to each section, provided the information required for plan selection from a database library of applicable timing plans.

Control of Groups of Intersections Using Online Generated Timing Plans

The intersection control strategies described in previous sections are typical of those employed in pre-UTCS systems using uncoordinated control, time base coordinated control or interconnected control systems (see definitions in this chapter under “Signal Control Concepts for City Streets”). The UTCS project began the large-scale use of sensors and computers for controlling traffic signal systems in the United States. This project drove the initial U.S. development of traffic adjusted control and traffic responsive control, which led to traffic adaptive control.

Traffic Adjusted Control and UTCS 1.5-Generation Systems

First-generation interconnected traffic signal control systems are characterized by the TOD/DOW selection of a timing plan from a set of timing plans, which are computed offline. Traffic adjusted control such as UTCS 1.5-generation added the capability for timing plans to be selected based on a combination of volume V and weighted occupancy O sensor data, referred to as VPLUSKO (i.e., volume plus weighted occupancy), where the K represents the weighting factor. Figure 3-6 illustrates the 1.5-generation UTCS timing plan selection process.

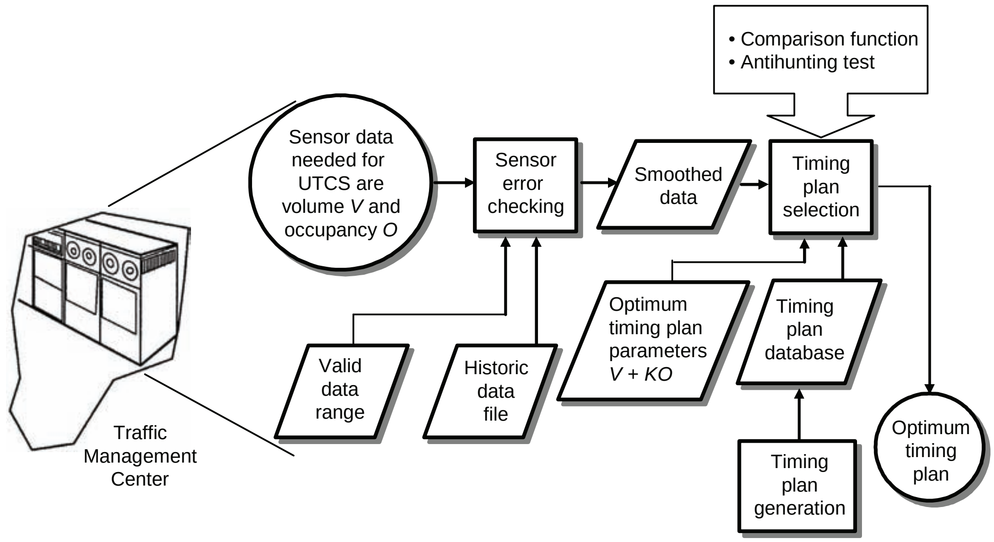


Figure 3-6. UTCS timing plan selection procedure.

Each timing plan has a VPLUSKO value corresponding to each of the sensors in the system. The sensor outputs are compared to the VPLUSKO values from each plan, and the plan that most closely matches the sensor outputs is selected. If a potential timing plan is found to be more favorable than the current plan, then the new plan is subjected to an antihunting test. The purpose of the antihunting test is to verify that the new plan is sufficiently better (by a predefined amount) to warrant implementation. This prevents needless transitions between timing plans that have similar benefits. The comparison operation is repeated at user-selected intervals, e.g., 4 to 15 minutes.

However, 1.5-generation systems do not address all the drawbacks of pre-UTCS signal control such as unexpected traffic flow scenarios (e.g., from unanticipated incidents). Traffic adjusted control (such as UTCS 1.5) responds successfully to precalculated traffic demands such as those produced by large vehicle flows from parking lot exits at sports venues.

Second-Generation Traffic Signal Control Systems

Second-generation UTCS was a first attempt at real-time, online computing of optimized splits and offsets, while keeping cycle length fixed within variable groups of intersections. Trials with this technique demonstrated some reduction in vehicle-minutes of travel time (with respect to the base system) on the arterial, but increased travel time in the network as a whole.[(7)] More advanced strategies are found in third-generation traffic signal control systems. These traffic systems support the online generation and implementation of signal timing parameters derived from real-time sensor data and modeling, prediction, and optimization techniques.

Third-Generation Traffic Signal Control Systems

Traffic responsive and traffic adaptive systems can overcome several limitations of signal control systems that rely solely on prestored timing plans. For example, prestored timing plans developed offline are best suited for traffic flow on a normal day or for events that produce predictable traffic patterns. Their major disadvantage is that they are developed from specific traffic flow scenarios and, therefore, cannot respond to situations that are significantly different from those used to generate them. Their major advantage is that they can be implemented at the beginning of a preplanned traffic event or incident such as the exiting of a large number of vehicles from a sports venue parking facility. Furthermore, data collection and manpower costs limit the ability of many traffic management organizations to maintain timing plans that are representative of current traffic volumes and patterns. Traffic responsive and traffic daptive systems attempt to overcome these limitations by providing signal timing that more quickly responds to real-time traffic flow sensor data than traditional traffic adjusted or TOD/DOW systems.

Traffic responsive and traffic adaptive systems generally require a greater number of sensors than conventional first-generation systems and usually require extensive initial calibration and validation.[8] Therefore, total system life-cycle costs, including software licensing, purchase of local controllers and central computers, and ongoing operating and field maintenance costs, are often compared to expected benefits when evaluating traffic signal operating strategies. Nevertheless, continued advancements in sensor and computer system technology plus improving traffic control algorithms are making traffic responsive and traffic adaptive systems increasingly attractive as compared to conventional systems when traffic volumes and roadway network design warrant their use. The ability to adapt to changes in traffic flow patterns over long-term intervals (i.e., to respond to aging of prestored timing plans) frequently makes traffic responsive systems cost effective.[9-26]

Centralized traffic responsive and distributed traffic adaptive system concepts have been developed for signal control. Typical of the traffic responsive concept are SCOOT and SCATS.

Traffic Responsive Control and Centralized Systems

SCOOT and SCATS have been called both traffic responsive and traffic adaptive. This edition of the Traffic Detector Handbook *uses the definitions in Appendix L, "Classification of Sensor Systems by Sensor Density Level," and Appendix P, "Glossary," and classifies them as traffic responsive systems.*

SCOOT: The Split, Cycle and Offset Optimization Technique (SCOOT) is a real-time system developed by the Transport and Road Research Laboratory (TRRL) in Great Britain.

SCATS: The Sydney Coordinated Adaptive Traffic System (SCATS). A real-time system developed by the Roads and Traffic Authority (RTA) of New South Wales, Australia.

SCOOT continuously measures traffic demand on most approaches to intersections in the network and optimizes signal cycle lengths, splits, and offsets to minimize delay and stops. SCOOT sensors are located upstream from the signal stopline, approximately 15 meters downstream of the adjacent upstream intersection. Timing changes per cycle are small to avoid major disruptions to traffic flow, but frequent enough to allow rapid response to changing traffic conditions. The prototype SCOOT systems reduced peak period average delay at traffic signals by approximately 11 percent as compared to fixed signal control plans generated by TRANSYT. Off-peak delay was reduced by an average of 16 percent.[10,16]

Similar to SCOOT, SCATS adjusts cycle time, splits, and offsets in response to real-time traffic demand and system capacity. The principal goal of SCATS is to minimize overall stops and delay when traffic demand is less than system capacity. When demand approaches system capacity, SCATS maximizes throughput and controls queue formation. SCATS sensors are installed in each lane immediately in advance of the stopline to collect volume and occupancy data during the green of the approach. Studies have found that the reduction of stopped and approach delay for the main approaches to intersections is greater during low-volume time periods than during peak volume periods.[27-29] Sensor configurations for SCOOT and SCATS are discussed in Appendix L.

Traffic Adaptive Control and Distributed Systems

Traffic adaptive systems are typified by the adaptive control system (ACS), which uses OPAC algorithms in one rendering and RHODES algorithms in another. These distributed systems are executed at the local intersection using an advanced traffic controller such as the 2070. They employ a rolling horizon, which recalculates predictions and optimizations every second for several seconds. Sensors placed upstream of the stopline give at least 10 to 15 seconds of travel time until the upstream vehicle platoon reaches the downstream intersections.[(2)] Sensor configurations for ACS are discussed in Appendix L. An implementation of ACS with the OPAC algorithms on Reston Parkway in Virginia resulted in a 5 to 8 percent reduction in stops and delay over an optimized actuated signal control system.[(30)] In simulations conducted as part of the RT-TRACS program (now called ACS), RHODES was found to increase throughput and reduce delay on the test network and, hence, performed better than the fixed-timing plan generated from TRANSYT-7F.[(31-33)] The advent of sensors that can be used to measure turning movements and travel times on links will expand the application of algorithms used in traffic adaptive systems. See Table 3-1 for more detail on differences between traffic responsive and traffic adaptive control.

The key differences between interconnected control, traffic adjusted control, traffic responsive control and traffic adaptive control are shown in Table 3-1.

Wireless communications may be substituted for wireline for most communications speeds.

Review of strategies for control of groups of intersections

Table 3-1 summarizes the characteristics of different categories of control strategies for groups of intersections. The key difference is how rapidly each control strategy can make changes in response to variations in traffic demand. From this criterion flows differences in the data requirements and data flow rate needed to implement the strategies. This in turn enables different timing features to be changed at different frequencies in each category.

All of the categories and characteristics are for systems and not for the individual intersection.

This row describes the system response and not any local traffic actuated response.

In some interconnected systems, operators can select plans in real time. In all traffic adjusted and traffic responsive systems, plans may be selected in real time.

In most traffic adjusted and all traffic responsive systems, operator may select and download changes in signal settings and plans in real time.

Fixed timing plans—cycle, split, offset, and actuated settings are not adjusted between switching of timing plans. These are used in uncoordinated signal control, time based coordinated control, interconnected control and traffic adjusted control systems.

Wireless communications may be substituted for wireline for most communications speeds.

Table 3-1. Categories of traffic signal systems and their characteristics or requirements.

Category title to right character-istic below	Uncoord-inated control	Time-based coordinated control (TBC) and interconnected control categories	Traffic adjusted control	Traffic responsive control	Traffic adaptive control
Response to variations in traffic demand	Very slow reactive response based on historic traffic flows, prompt local actuated response possible, no current system response	Very slow reactive response based on historic traffic flows, prompt local actuated response possible, no current system response	Slow reactive response based on overall network trend—volume plus weighted occupancy or general direction of flow	Prompt reactive response based on changes in platoons	Very rapid predictive proactive response based on predicted movements of individual vehicles or small packets of vehicles
Frequency of change in control plan	No changes related to adjacent intersections	Plans for peaks, midday and evening off peaks, and weekends	Minimum of 15 minutes, usually several times a rush period	Minimum of 15 minutes with adjustments possible once per cycle	No overall timing plan as continuous adjustments are made to all parameters.
What timing features are changed	No changes related to adjacent Intersections Timing for individual intersections may change at preset TOD/DOW	Total timing plan may be changed out at preset TOD/DOW—Fixed plan with splits, offsets cycle length, etc., unchanged during period	Total timing plan changed out—Fixed plan with splits, offsets, cycle length, etc., unchanged during period; variable TOD when plan is changed	Total timing plan changed out—Flexible plan with cycle, splits, offsets, and actuated settings changeable once per cycle; variable TOD plan changes	Phase selection, sequence, and durations adjusted every time step; cycle and offset not required; splits are dynamic as above
Data communi-cations require-ments	None	Not required for TBC; low speed for interconnected real-time comm not required (frequently twisted pair)	Low speed, real time communica-tion may be needed; twisted pair or better preferred	Low speed but faster is better; real-time communication is needed	Fast and real time
System sensor require-ments	No *system* sensors required	No *system* sensors required	Minimal, one system sensor per intersection *on average*	Moderate, minimum of 1 per approach, preferred 1 per lane per approach	High, minimum of 1 per lane per approach, RHODES 2 per lane per approach

Table 3-1. Categories of traffic signal systems and their characteristics or requirements—*Continued*

Category title to right character-istic relow	Uncoord-inated control	TBC and interconnected control categories	Traffic adjusted control	Traffic responsive control	Traffic adaptive control
Local inter-section controller and coresident local computer require-ments	For new installations NEMA/170 and no coresident computer	For new installations NEMA/170 and no coresident computer	NEMA/170 and no coresident computer	Moderate—NEMA/170/ATC and no coresident computer for local control algorithms	High, ATC preferred, coprocessor card or separate computer box for running local algorithms
Central computer or system master require-ments	None	TBC None—interconnect minimal, vendor system master or PC level	Moderate, vendor system master or PC level	High end system master or PC required—much central processing	Medium most processing is local
Mainte-nance of database, communi-cations, sensors and system master	Low cost —local settings, local sensors, traffic studies	Low cost —add setting coordination settings, some communi-cations traffic studies for them	Moderate cost —add to interconnected more database, communi-cations and minimal system sensors	Medium cost — add more system sensors, central database, communications but fewer traffic studies	High cost —add more system sensors, communi-cations, less central database
Installation	Low cost	Low cost	Moderate cost	Medium cost	High cost
System design and operation	Requires lowest skill level and staffing	Requires moderate skill level and staffing	Complex and requires good skill level and staffing	More complex and requires high skill level and staffing	More complex and requires high skill level and staffing

SYSTEMS ENGINEERING PROCESS IN DESIGN OF TRAFFIC SIGNAL SYSTEMS

A discussion of the systems engineering process involved in designing traffic signal systems is found in NCHRP Synthesis 307.[8] The major steps involved in the systems engineering process as applied to traffic signal systems begin with the identification of requirements for:

- Function.
- System design.
- Project management.
- Operations.
- Logistics.
- Evaluation.

Once requirements are established, plans and methods to implement them and evaluate system performance are developed.

The *Traffic Control Systems Handbook* and the *Traffic Detector Handbook* describe the types of coordinated traffic signal systems through a category of characteristics as listed in Table 3-1. The number of sensors required for each level are described in more detail in Appendix L.

Freeway surveillance and control are used to mitigate recurring and nonrecurring congestion. Common techniques applied for this application include entrance ramp metering, mainline metering, freeway-to-freeway metering, corridor management, high occupancy vehicle lanes, congestion-based pricing for toll facilities, advanced traveler information systems, and automatic incident detection.

FREEWAY SURVEILLANCE AND CONTROL

Sensors are used in freeway surveillance and control to detect recurring and nonrecurring congestion and to assist in mitigating their effects. *Recurring* congestion occurs when both the location and time of congestion are predictable such as during weekday peak travel periods. *Nonrecurring* congestion is caused by random, temporary incidents such as stalled vehicles, accidents, spilled loads, or other unpredictable events.

Recurring congestion develops when traffic demand exceeds freeway capacity. Recurring congestion may be reduced by decreasing peak period demand through techniques such as entrance ramp metering, mainline metering, freeway-to-freeway connection control, corridor management, control of the number of occupants needed to access high-occupancy vehicle lanes, congestion-based pricing for use of toll facilities, and advanced traveler information systems that inform motorists of congestion ahead and perhaps of alternate routes. Sensors play a major role in alleviating recurring congestion, particularly in entrance ramp metering.

Nonrecurring congestion is more difficult to manage because of its unpredictability. Detecting the incident and removing its cause as quickly as possible minimize the effects of the nonrecurring events.

A variety of information gathering techniques are used for incident detection, including periodic sensor placement along freeways, closed-circuit television, aerial surveillance, emergency call boxes, freeway patrols, and cellular telephone calls from motorists. Nonimaging mainline sensors are not as effective for incident identification, but are often used to detect the beginning of congestion at off peak hours, which usually indicates some type of incident has occurred. They can also be used to determine the extent of the incident's impact. Automatic incident detection algorithms face a more difficult task during peak traffic hours because the characteristics of nonrecurring congestion caused by an incident are often difficult to differentiate from those of recurring congestion.

The following sections further explore techniques that utilize sensors for managing traffic on freeways.

FREEWAY INCIDENT DETECTION

Fifty to sixty percent of the delay on urban freeways is associated with incidents, rather than with recurring congestion due to capacity shortfall.[34,35] Often recurring and nonrecurring congestion occur simultaneously during peak periods. In fact, since most algorithms detect incidents by measuring perturbations in traffic flow, automatic incident detection techniques must distinguish between shock waves formed by recurring peak-period congestion and those formed by incidents during similar periods. An incident can be defined as any anomaly that disrupts the smooth flow of traffic. Debris on the road, flooding, cargo spills, vehicle collisions or accidents, special events such as ball games and festivals, and highway work are examples of incidents.

Traffic Flow Characteristics During an Incident

Traffic flow characteristics during a freeway incident can be characterized in terms of the four flow regions illustrated in Figure 3-7.[36] Flow region A is far enough upstream of the incident so that traffic moves at normal speeds with normal density. Flow region B is the area located directly behind the incident where vehicles are queuing if traffic demand exceeds the restricted capacity caused by the incident. In this region, characterized by the upstream propagation of a shock wave, speeds are generally lower, and a greater vehicle density may exist. Flow region C, also shown in Figure 3-8, is the region directly downstream from the incident where traffic is flowing at a metered rate, or incident flow rate, due to the restricted capacity caused by the incident. Depending on the extent of the capacity reduction, traffic density in region C can be lower than normal, while the corresponding traffic speed is generally higher than normal. Flow region D is far enough downstream from the incident such that traffic in D flows at normal density and speed, as in region A.

Incident clearing and the return of traffic flow to normal occur in several phases. These are detection, verification and identification, response, removal, and recovery. Detection determines that something extraordinary has occurred. Then the location, types of vehicles, presence of hazardous materials, and possible injuries related to the incident have to be verified and identified. This step facilitates timely dispatching of appropriate personnel and equipment to remove affected vehicles and people from the site.

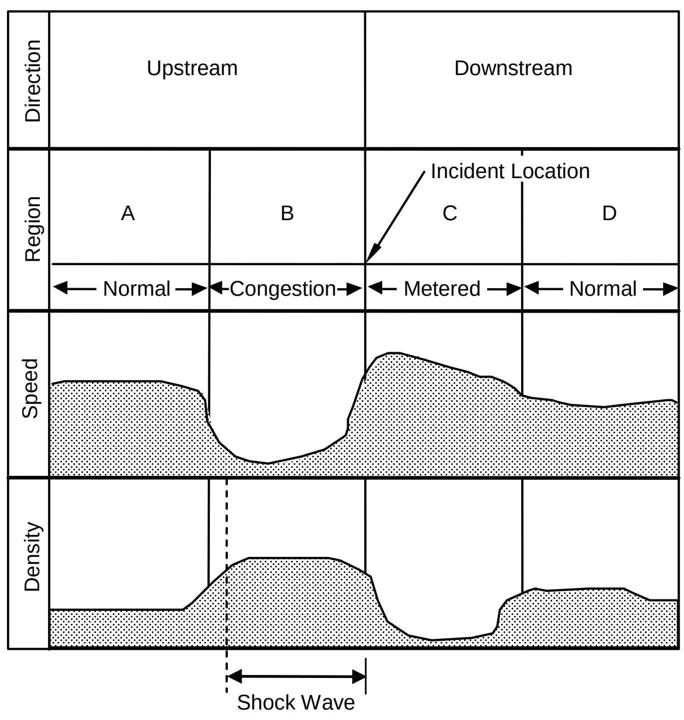


Figure 3-7. Traffic flow characteristics during an incident (source: *Intelligent Vehicle Highway Systems: The State of the Art*, prepared for Massachusetts Department of Highways (JHK and Associates, New York, NY, March 1993)).

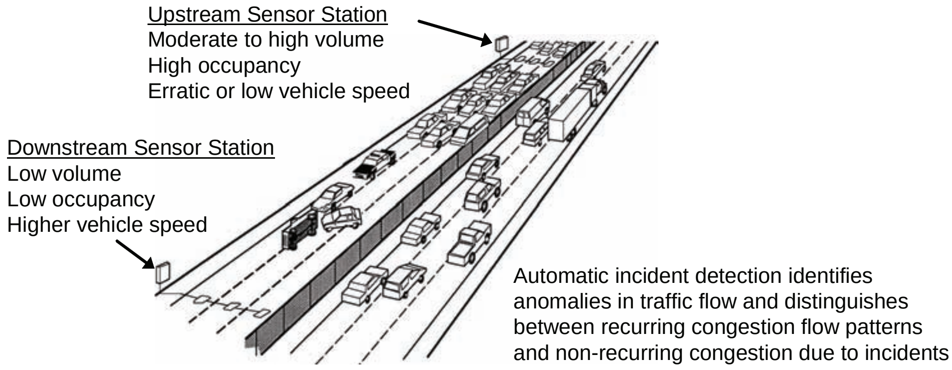


Figure 3-8. Freeway incident detection scenario.

Detection and verification are the areas in which improved sensor and communications technology can be of most benefit to incident clearing. Closed circuit television, roving highway and freeway service patrols, aerial surveillance, roadside emergency call boxes, reports from fleet operators, and some repeated number of cellular telephone calls to 911 or the traffic management center are methods used to verify the occurrence of an incident.

Timeliness is the key to incident clearing and the minimization of congested flow conditions. Rapid detection and verification allow faster and perhaps less complex response options and more effective aid to victims. Quick response and the proper equipment permit rapid removal of affected vehicles and debris, and reduced delay and exposure to secondary accidents at the scene and at the end of any queues that form. When an incident blocks a traffic lane, the flow is choked and slowed, resulting in a traffic queue upstream of the incident. The queue and vehicle-hours of delay continue to build until the incident is cleared and normal traffic flow is restored. If the normal flow of traffic into the incident site is reduced by diversion onto alternate routes, then the vehicle-hours of delay are minimized. If normal traffic flow is not diverted, then additional vehicle-hours of delay are realized.(35)

ITS technology, with its emphasis on real-time operation and rapid communication, can assist in reducing the delay and crash costs associated with incidents. However, without the capability for rapid response and the cooperation and coordination among personnel in the responsible agencies (e.g., highway patrol, tow truck operators, fire, hazardous material team, emergency medical services, local police, traffic and transit management), the potential benefits of these high technology systems may not be realized.

FREEWAY METERING

Several forms of freeway metering exist. The most common is *onramp metering*, which restricts freeway demand at limited access onramps in an effort to prevent breakdown of flow on the mainline. *Freeway-to-freeway connector metering*, also used to control mainline demand, is designed so that queues at the metering signals do not compromise safety on the high-speed roadways. *Mainline metering* is applied at selected locations, such as bridges and tunnels, to maximize traffic flow through these facilities. It is usually implemented where there is sufficient storage capacity (such as at a toll plaza) and in conjunction with HOV lanes that bypass the queues upstream of the signals.

Ramp closure, a technique that does not require sensors on the ramp or mainline, can be used to restrict the number of vehicles entering a freeway. The more modern approach is to use sensors to monitor mainline congestion and transmit that information to ramp signals, which moderate the number of vehicles entering the freeway mainline.

Ramp Closure

Closing an entrance ramp during peak period is a positive technique for limiting the number of vehicles entering a congested freeway. It is, however, the most restrictive and least popular with the public. If applied in an inappropriate situation, it could result in underutilization of the freeway and the overloading of alternative routes. Ramp closures are effective where the entrance ramp introduces serious weaving or merging problems under congested conditions.

Manually placed barriers (including law enforcement vehicles), automated barriers, and signing are used for ramp closures. Sensors are not required except, perhaps, during the changeover operation of automated barriers.

Ramp Metering

The most common technique for addressing recurring congestion on freeways is ramp metering. It limits the rate at which vehicles enter the freeway's mainline so that the downstream mainline capacity is not exceeded. Ramp metering redistributes the freeway demand over space and time. Excess demand is either stored on the ramp or diverted. The diverted vehicles may choose less traveled alternate routes, or their occupants may select another mode of transportation. Metering rates range from a minimum of 180 to 240 vehicles per hour (v/h) to a practical maximum of 750 to 900 v/h.

Ramp meters assist in dispersing platoons of vehicles that are released from nearby signalized intersections. By releasing a limited number of vehicles into the mainline traffic stream, turbulence is reduced in the merge zone. This leads to a reduction in sideswipe and rear-end accidents, which are associated with stop-and-go traffic flow. Maximum mainline flow rates can be achieved by controlling ramp flow rates such that freeway traffic moves at or near optimum speed throughout the network.

Management of nonrecurring congestion created by freeway incidents is a secondary benefit associated with ramp metering. Once an incident is detected, ramp metering can potentially reduce the number of vehicles impacted by the incident. For example, meters upstream of a detected incident can be adjusted to allow fewer vehicles to enter the affected facility, potentially diverting some trips to onramps downstream of the incident location. Conversely, the downstream ramps can operate with relaxed metering rates in order to accommodate the increased demand. This incident management strategy works well during peak periods, especially when integrated with a smart corridor operation. Incidents occurring during off-peak hours may be mitigated with a strategy that closes upstream onramps rather than with metering because of the reduced onramp volumes.

A seminal study that evaluated the benefits of ramp metering was conducted by the Minnesota Department of Transportation (MnDOT) in accordance with a bill passed in the year 2000 session by the Minnesota Legislature.[37] The bill required MnDOT to examine the effectiveness of ramp meters in the Minneapolis-St. Paul, MN, region by conducting a shutdown study before the next legislative session. The goal was to evaluate and report any relevant fact comparisons or statistics concerning traffic flow and safety impacts associated with deactivating system ramp meters for a predetermined amount of time. The study, completed at a cost of $651,600, occurred in the fall of 2000, with the results presented to the Legislature and the public in early 2001.

A summary of the conclusions reached by MnDOT concerning the annual benefits of ramp metering is as follows:

- Traffic volumes and throughput: After the meters were turned off, traffic volume decreased by an average of 9 percent on freeways, with no significant traffic volume change on parallel arterials included in the study. Also, during peak traffic conditions, freeway mainline throughput declined by an average of 14 percent in the "without meters" condition.
- Travel time*:* Without meters, the decline in travel speeds on freeway facilities more than offset the elimination of ramp delays. This gives an annual system-wide savings of 25,121 hours of travel time with ramp meters.
- Travel time reliability: Without ramp metering, freeway travel time was almost twice as unpredictable as with ramp metering. The ramp metering system reduced unexpected delay by 2.6 million hours.
- Safety: In the absence of metering and after accounting for seasonal variations, peak period crashes on previously metered freeways and ramps increased by 26 percent. Ramp metering reduced annual crashes by 1,041 or approximately 4 crashes per day.
- Emissions: Ramp metering reduced net annual emissions by 1,160 tons.
- Fuel consumption: Ramp metering increased annual fuel consumption by 5.5 million gallons, based on a simple straight-line estimation technique that does not address the tempering of flow, typically because of ramp metering, by smoothing the travel speed variability (less acceleration and deceleration). This was the only criteria category that was degraded by ramp metering. However, four other regions of the country examined for fuel consumption impacts of ramp metering as part of the MnDOT study showed fuel savings ranging from about 6 percent to 13 percent.
- Benefit/Cost Analysis: Ramp metering produces an annual savings of approximately $40 million to the Twin Cities traveling public. The benefits of ramp metering outweigh the costs by a significant margin and result in a net benefit of $32 to $37 million per year. The benefit/cost ratio indicates that benefits are approximately 5 times the cost of the entire congestion management system and more than 15 times the cost of the ramp metering system alone.

Figure 3-9 depicts a conceptual freeway ramp meter installation. The sensors on the mainline serve a dual purpose: adjustment of the ramp-metering rate in response to real-time demand and collection of historical volume and occupancy data. A demand sensor on the ramp indicates the arrival of a vehicle at the stopline and the commensurate start of the metering cycle. Demand at the stopline is typically required before the ramp signal is allowed to turn green. A passage sensor detects when the vehicle passes the stopline and returns the ramp signal to red for the next vehicle. The passage sensor can also be used to monitor meter violations (i.e., drivers who ignore the red stop signal) and provide historical data about the violation rate at each ramp.

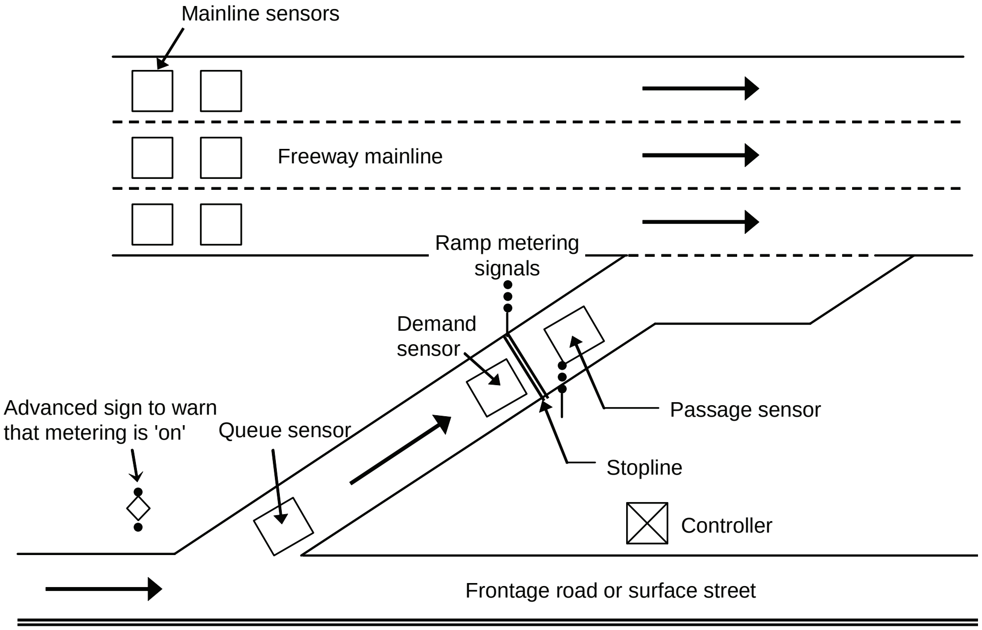


Figure 3-9. Conceptual ramp metering sensor and roadway configurations.

Ramps that contain two metered lanes or one metered and one unmetered HOV lane add a count sensor after the passage sensor to obtain the total count of vehicles entering the mainline. The queue sensor is used at locations where ramp backup impacts surface street operation. A high occupancy rate over a queue sensor can signal the ramp metering logic to increase vehicle passage or, in special situations, to suspend metering to prevent vehicles from backing onto feeder roads.[2] The advanced warning sign alerts drivers to the operational status of metering activity. Where ramp geometrics are poor, a merge sensor may be used as a feedback mechanism to prevent additional vehicles from proceeding down the ramp if a vehicle is stopped in the freeway merge area.

Current ramp metering strategies include:[2]

- Pretimed metering.
- Local traffic responsive control.
- Coordinated or global traffic responsive control.

Pretimed Ramp Metering

The traffic signal in a pretimed metering system operates with a constant cycle using a metering rate calculated from historical data. Metering and signage can accommodate single vehicle entry or platoon entry onto the mainline. The major operational advantage of pretimed metering is the regularity of the rate that is easily accommodated by drivers. The principal drawback of pretimed metering is its inherent insensitivity to changes in traffic conditions. Pretimed metering is often implemented as an initial operating strategy until traffic responsive control can be initiated. The sensors needed for pretimed metering are a queue sensor, demand sensor, and passage sensor.

The method for calculating pretimed metering rates depends on the primary purpose of the metering, namely to minimize congestion or to improve safety.

Minimizing congestion through pretimed metering—Freeway congestion reduction is implemented by choosing a metering rate approximately equal to the difference between upstream freeway demand and the downstream freeway capacity. Metering rates that address congestion reduction are based on:

- Ramp storage capacity.
- Availability of alternative routes for diverted traffic.
- Conditions at other ramps upstream and downstream of the controlled ramp.

Improving safety through pretimed metering—Pretimed metering that improves merging safety allows each vehicle enough time to enter the mainline before the following vehicle enters the merging area. This prevents rear-end and lane-changing collisions by breaking up platoons of vehicles that compete for gaps in the freeway traffic stream. Metering rates that primarily treat safety depend on the distance from the stopline to the merging point, ramp geometry, and vehicle type.

Local Traffic Responsive Metering

Local responsive metering uses real-time mainline traffic flow information near the ramp or just downstream of the ramp to determine metering rates. The traffic flow parameters assist in evaluating freeway operation with respect to upstream demand and downstream capacity and in determining the maximum number of ramp vehicles permitted to enter the freeway without causing congestion. Traffic flow parameter measurements such as occupancy and volume are often smoothed to filter short-term random fluctuations by calculating running averages over a 5-minute period. Some facilities include sensors to determine traffic composition and weather to account for the effects of these factors. The sensors needed for pretimed metering are a queue sensor, demand sensor, passage sensor, merge sensor, and mainline sensors.

Advantages and disadvantages—The advantage of local traffic responsive metering over pretimed metering is the ability of the metering rate to respond to short-term variations in traffic demand or to reduced capacity caused by incidents downstream of the ramp. Results with traffic responsive metering are 5 to 10 percent better than pretimed metering in terms of reduced overall delay. A best case example of traffic responsive ramp metering benefits is from a ramp control experiment in Los Angeles, CA, that produced a 100 percent increase in average speed from 25 to 52 mi/h (40 to 84 km/h), a 20 percent decrease in ramp wait time, and a 3 percent increase in freeway volumes.[38] The main disadvantage of local traffic responsive control strategies is that mitigation must wait until the congestion reaches the local section controlled by the ramp. Table 3-2 contains a recommended range of metering rates based on mainline occupancy as the controlling parameter.[39,40]

Table 3-2. Local responsive ramp metering rates based on mainline occupancy.

Occupancy (percent)	Metering rate vehicles per minute (v/min)[a]
<10	12
11–16	10
17–22	8
23–28	6
29–34	4
>34	3

Occupancy (percent)	Metering rate (v/min)[b]
<20	12+
20–22	10
22–25	8
25–27	6
27 or higher	4

[a] Blumentritt, C.W., et al. *Guidelines for Selection of Ramp Control Systems*, NCHRP Report 232. National Research Council, Transportation Research Board, Washington, DC, May 1981.
[b] McDermott, J.M., S.J. Kolenko, and R.J. Wojcik. *Chicago Area Expressway Surveillance and Control: Final Report*, Report Number FHWA-IL-ES-27. Illinois Department of Transportation, Oak Park, IL, 1979.

Improving safety—The impact of local traffic responsive ramp metering on the crash rate was studied in an early operational test in Chicago, IL. The metering rate was controlled by stream measurements of lane occupancy in one of the center lanes near the metering device, plus measurements of occupancy at the nearest downstream bottleneck. The prevailing metering rate was the most restrictive calculated from local or freeway segment conditions. An analysis of merges that occurred during peak traffic flow periods showed an 11.6 percent reduction in all traffic conflicts when the metering rate was 12 or more v/min. Metering at a more restrictive rate was predicted to reduce conflicts further.(40) Although strategies were developed for controlling metering rates based on mainline flow rate, ramp flow rate, and occupancy measurements, the advantages of these techniques were outweighed by operational and maintenance disadvantages created by the need for more sensors and more precise and accurate data.

Implementation—Local traffic responsive metering algorithms may be implemented with techniques other than lookup tables. Some algorithms attempt to maintain mainline occupancy at a preselected value through a recursive process. Asservissement LINéaire d'Entrée Autoroutière (ALINEA), is such an algorithm that uses feedback to control mainline occupancy by adjusting the metering rate R (v/min) at time t+1 based on its value at time t according to(41-43)

$$R|_{t+1} = R|_{t} + K_R\,[\hat{O} - O|_{t+1}] \qquad (3\text{-}1)$$

where

$R|_t$ = metering rate at time t in v/min

K_R = rate adjustment parameter in v/min

$\hat{O}$ = nominal or target occupancy

$O|_{t+1}$ = measured mainline occupancy at time t+1.

Papageorgiou uses a default value of 1.17 for K_R, while Mahmassani reports a value of 0.32. Typical values for $\hat{O}$ are 0.17 to 0.30. If the calculated value for R violates physical or operational constraints, it is suitably adjusted to a feasible value. The metering rate at any time may be constrained by maximum and minimum values to suit the traffic management objectives.

The default values of K_R are for illustration only and are not meant to imply values suitable for any or all applications.

In the Netherlands, fuzzy logic control (FLC) is used on the A12 freeway between The Hague and Utrecht.[44] The algorithm restricts the metering rate when the downstream speed is less than the upstream speed. FLC produced 35 percent faster travel times and a 5 to 6 percent greater bottleneck capacity than two other controllers over a 6.8-mi (11-km) freeway section.

The Washington State Department of Transportation (WSDOT) operated two ramp metering algorithms in the greater Seattle area, the local metering algorithm and the bottleneck algorithm.[45] The local metering algorithm uses linear interpolation between measured mainline occupancy and appropriate metering rates for that occupancy. The bottleneck algorithm reduces the number of vehicles entering the freeway by the number of vehicles stored in a downstream bottleneck section. Both of these algorithms use queue overrides to flush the ramp queue when it becomes excessive. A fuzzy logic algorithm (FLA) is replacing the older algorithms to meter traffic on more than 100 ramps on Interstates 5, 405, and 90 and on State Route 520. Congestion at the I–90 study site was 8.2 percent lower with the FLA than with the local metering algorithm, while throughput was 4.9 percent greater. Some ramp queues decreased, while other increased slightly. However, the ramps had sufficient storage space, and given the mainline benefits, slightly longer ramp queues were acceptable. Results at the more congested I–405 test site were mixed. Mainline congestion increased by 1.2 percent with the FLA, while throughput increased slightly by 0.8 percent (as compared to the bottleneck algorithm). However, the fuzzy logic algorithm significantly reduced the time each ramp was congested by an average of 26.5 minutes.[46] A fuzzy logic control algorithm is under investigation to simultaneously control multiple ramps along a freeway section.[47,48]

Local traffic responsive metering, as used by Caltrans in Orange County, CA, establishes a floor below which the rate cannot fall. The local responsive ramp-metering algorithm that resides in the 170 controllers at the onramps operates as follows. First, a TOD rate is programmed into the 170 controller. This rate may be: alternatively(1) manually entered by a traffic engineer into the 170 controller in the field, (2) sent by a traffic management center operator to the 170 in the field, or (3) sent by area engineers from their office to the 170 in the field. Second, critical values for the average 3-min mainline volume/lane (currently 75 vehicles) and the average 1-min mainline occupancy/lane (currently 20 percent) are entered into the 170 controller. Third, a local responsive rate R_{LR}, based on the actual mainline volume, for the number of vehicles released by the ramp is calculated as:

$$R_{LR} = (V_C - V_M)(N_{RL})/3 \text{ veh/min} \tag{3-2}$$

where

V_C = critical value of mainline traffic volume per lane for a 3-min period

V_M = real-time mainline traffic volume per lane over 3 min

N_{RL} = number of ramp lanes

and the factor of 3 converts the 3-min values into 1-min values.

Fourth, if the locally measured mainline volume is less than the critical value, then the local responsive meter rate is used, provided R_{LR} is greater than the value in the TOD table. If the TOD rate is larger, then it is used. If the locally measured mainline volume is greater than the critical value, then the TOD rate is also used. Other versions of a local responsive ramp-metering algorithm can compare the actual mainline occupancy with the critical occupancy value and adjust the meter rate accordingly.[49]

Coordinated Traffic Responsive Control

In this strategy, traffic flow data are analyzed at a central traffic management center that simultaneously adjusts the metering rates at several ramps. The metering rates are found from the analysis of the demand and capacity of an entire freeway section rather than traffic conditions in the immediate vicinity of individual ramps. A conceptual representation of a coordinated traffic responsive ramp control system is shown in Figure 3-10.

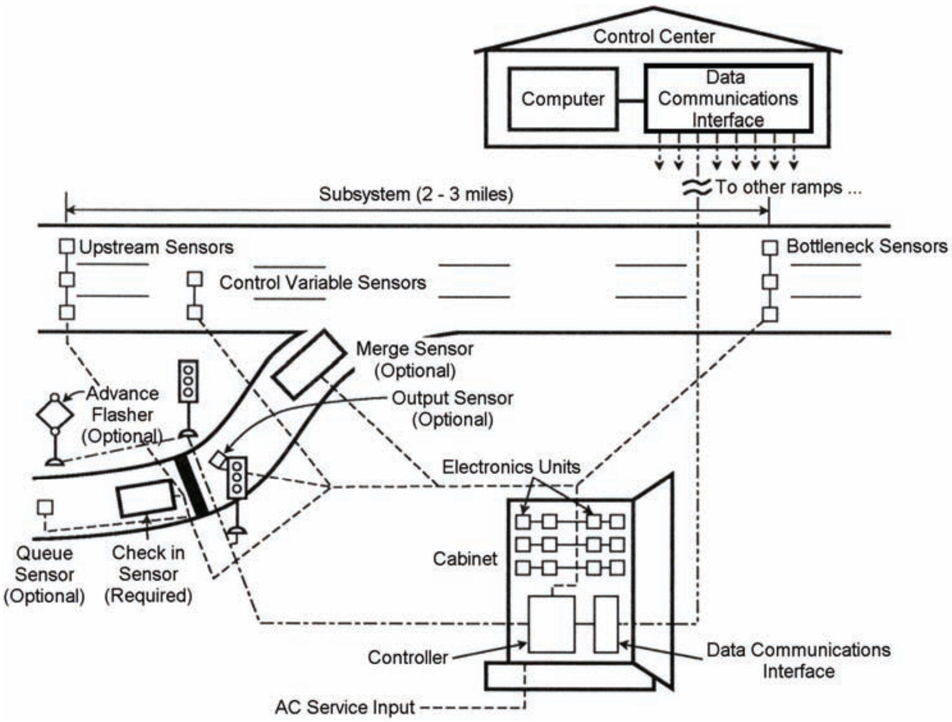


Figure 3-10. Conceptual coordinated traffic responsive ramp control system.

A linear programming model is often used to calculate sets of integrated metering rates for each ramp based on the expected range of capacity and demand. The appropriate metering rate at each ramp is then selected from these precomputed sets, based on real-time measurement of freeway conditions.[38] In another approach, the temporal and spatial characteristics of the traffic flow pattern on the freeway and affected surface streets are used as inputs to a metering rate optimization algorithm.[50]

In the United States, applications of coordinated ramp metering are found in Long Island, NY, Seattle, WA, Denver, CO, and Los Angeles, CA. The Los Angeles system wide adaptive ramp metering (SWARM) model implemented by Caltrans can activate one of three algorithms in an attempt to avoid breakdown in the flow-density regime.[51] The first algorithm, SWARM 1, adjusts rates at ramps upstream of a bottleneck in an attempt to maintain an acceptable level of service in the region of the bottleneck. The second and third algorithms, SWARM 2 and 3, are local responsive in nature. SWARM 2 is headway based, and SWARM 3 is density based. SWARM 1 estimates freeway conditions 10 to 15 min into the future using a Kalman filter.

As depicted in Figure 3-11, the SWARM 1 algorithm clamps down on the metering rates at time t_0 to prevent a bottleneck from occurring later at time t_1. The algorithm establishes the extent of the flow rate reduction propagated upstream through user-selected factors that control the propagation rate of the flow restriction. Data from either several links or the entire freeway can be used to establish the metering rate. When a queue override condition is in force, the metering goes to the maximum rate, rather than continuous green, to maintain platoon dispersion.

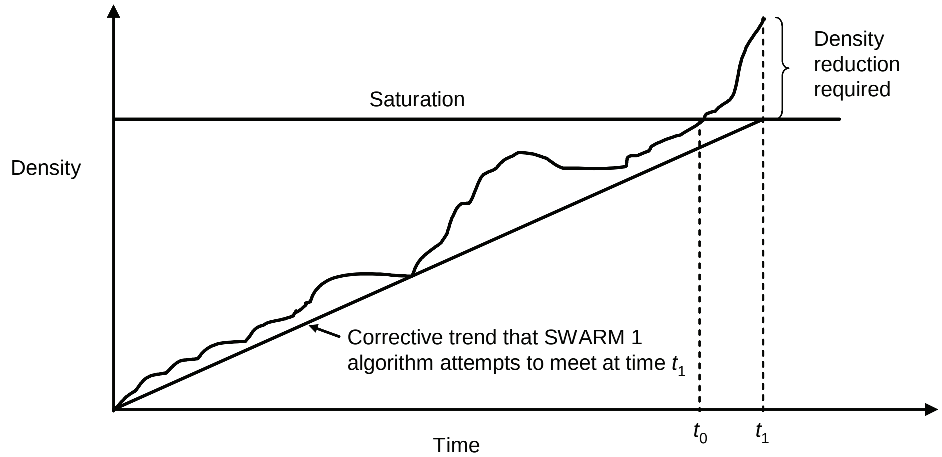


Figure 3-11. Principles of SWARM 1 ramp metering algorithm.

In Europe, two coordinated ramp metering approaches have been developed as part of DRIVE projects: the Metaline strategy and the Sirtaki strategy. The Metaline strategy is based on a linear quadratic integral control law that minimizes the deviation of selected bottleneck densities from their desired values by modifying the onramp flows. The Sirtaki strategy is built around the SIMAUT simulation model. SIMAUT incorporates coordinated ramp metering strategies based on demand-responsive control.(52) In case of an incident, control strategies can be simulated and analysis aided by the model's ability to reconstruct the current traffic stream in real time.

A summary of ramp metering rates provided by different metering methods is found in Table 3-3 in order of metering method preference.(3)

Table 3-3. Ramp metering rates by metering strategy.

Ramp metering method	Number of metered lanes	Approximate metering rate, v/h	Comments
Single vehicle entry per green interval	1	240–900 [54]	• Full stop at meter usually not achieved at 900 v/h metering rate.
Tandem metering with single vehicle entry per green interval per lane	2	400–1700	• Applies when required metering rate exceeds 900 v/h. • Requires two lanes for vehicle storage. • Vehicles may be released from each lane simultaneously or sequentially. • Requires sufficient distance beyond metering signal for vehicles to achieve tandem configuration before merging with freeway traffic.
Platoon metering with single lane multiple vehicle entry per green interval	1	240–1100 [54]	• Platoon lengths permit passage of 1 to 3 vehicles per green interval. • Principally used to increase metered volumes when geometrics do not permit more than one metered lane. • May require changeable sign indicating permitted number of vehicles in green interval. • *Manual on Uniform Traffic Control Devices* requires yellow interval after green.

Mainline Metering

Mainline metering is one form of freeway mainline traffic management that also makes use of driver information systems, variable-speed control, lane closure, and reversible lane control. Individually or in combination, these control techniques are gaining acceptance at many highly congested freeway locations throughout the Nation. The role of sensors in mainline control is dependent on the selected operational strategy and its data requirements.

When mainline metering is applied, signals on the freeway mainline control flow rates in a manner similar to ramp metering. Mainline metering manages traffic demand at a mainline control point to maintain a desired level of service on the freeway downstream of the control location.

The desired level of service for mainline metering is selected to achieve one or several of the following objectives:

- Flow maximization through a downstream bottleneck.
- High level of service downstream of the control point.
- Distribution of total delay on the freeway system more equitably.
- Diversion of traffic to other routes or modes.
- Increase in the overall safety of the facility.

Mainline metering is found on the westbound lanes crossing the San Francisco-

Oakland Bay Bridge, as illustrated in Figure 3-12. The signal bridge used for metering up to16 lanes of traffic is located about 800 ft (244 m) downstream from the toll plaza, which contains 20 toll booths and 2 bypass lanes (used for carpools during peak traffic periods and buses all the time). During peak periods, two of the toll booth lanes convert into additional HOV lanes that bypass the metering. The lanes from the remaining toll booths are reduced to 12 metered lanes at the signal bridge prior to merging onto the 5-lane Bay Bridge. During offpeak periods and weekends, 14 lanes are metered. The metering produces a more flexible system that accommodates a mix of bus, carpool, and single passenger vehicles, allowing the bridge to operate at peak efficiency.

Figure 3-12. Mainline metering configuration on I–80 westbound crossing the San Francisco-Oakland Bay Bridge (Photograph courtesy of Caltrans District 4, Oakland, CA).

Traffic accidents were reduced by 15 percent after the metering system was installed, with as many as 500 additional vehicles per hour crossing the bridge during peak periods. Fuzzy logic was utilized to control the queue length at the metering stations.[53] A new and enhanced fuzzy control system was developed to function with fewer sensor stations, lane occupancy data rather than station occupancy data, and new data validation algorithms. As reported by Caltrans (Sean Coughlin, Caltrans District 4, August 24, 1999), fuzzy control will be activated again once the original infrared sensors in each lane (set approximately 600 ft (183 m) apart) are replaced by dual magnetic sensors (approximately 1200 ft (366 m) apart).

Another application of mainline metering has operated since 1978 on Route 94 upstream of its junction with Route 125 in San Diego, CA. Ramps on Route 125 are metered, but the ramps on Route 94 are not. Mainline lanes of Route 94 are metered to offset the unbalanced upstream control.

A gantry, which contains a separate signal head for each of the three lanes (two conventional lanes and one HOV lane), is mounted over the roadway. During normal system operation from 6 to 9 a.m., each signal displays a green ball to allow one vehicle at a time to proceed in that lane. A passage sensor is located immediately downstream of the stop line. As the vehicle is detected, the signal turns red (no yellow interval is used during the metering operation, only during startup). The signal then remains red for the remainder of the metering cycle, which is set to achieve the desired flow rate. Extensive signing is placed in advance of the metering station to warn freeway traffic of the unusual event that they are about to encounter.

The success of the Route 94/125 mainline metering has encouraged the use of three other mainline metering stations on San Diego area freeways. One of these, located in El Cajon, is shown in Figure 3-13. Three lanes of southbound State Route 67 are metered as they join westbound Interstate 8. An internally illuminated sign displays the message “PREPARE TO STOP” during the metering operation. The metering signals are 12-inch (300-mm) standard 3-section heads centered on each lane and mounted on a 50-ft (15-m) mast arm.

Figure 3-13. Freeway mainline metering at El Cajon, CA.

In Tokyo and Osaka, Japan, mainline metering has been implemented by regulating the number of toll booths that are open at any given time on the mainline of the expressways. Traffic entering the expressway system via the mainline is controlled based on expressway demand and downstream capacity. Such control is rather coarse, however, and might not be appropriate for all bottleneck conditions.

Freeway-to-Freeway Metering

A third technique for managing recurring congestion is to meter freeway-to-freeway connector ramps. Experiences in Minneapolis, MN, and San Jose, San Diego, Los Angeles, and Orange County, CA, indicate that significant benefits can be achieved with connector metering under conditions similar to those associated with ramp metering. Freeway connectors often have per-lane flow rates greater than 900 v/h (the maximum possible with single entry metering). Metering rates exceeding this figure are achieved by two-lane metering or possibly platoon metering. Such configurations work best when there is an added lane downstream from the onramp.

Figure 3-14 illustrates connector metering from the eastbound I–105 Freeway onto the southbound I–605 Freeway in Norwalk, CA. Meters on the two I–105 lanes allow three vehicles per lane during each green cycle to enter the four lanes of the I–605 mainline. One of the two I–105 ramp lanes continues as a fifth mainline lane.(49)

Storage capacity is another issue that affects the metering of freeway-to-freeway connectors. As with onramps, queues are not allowed to extend upstream where they might interfere with other freeway movements. In addition, end-of-queue protection must be provided with automatic warning signs to prevent rear-end collisions.

Figure 3-14. Freeway-to-freeway metering at the junction of the I–105 and I–605 freeways in Norwalk, CA (Photograph courtesy of Lawrence A. Klein).

SPEED MONITORING AND DRIVER NOTIFICATION AT CRITICAL FREEWAY SEGMENTS

Safety problems occur when the design speed for certain curves is below that of other portions of the freeway. To lessen the crash potential at such locations, a speed measurement system that incorporates a flashing display to alert the driver to an unsafe speed can be used. Thus, if a vehicle is traveling faster than the desired or safe speed, a flashing sign or signal is activated to advise the driver to reduce the speed of the vehicle, as shown in Figure 3-15.

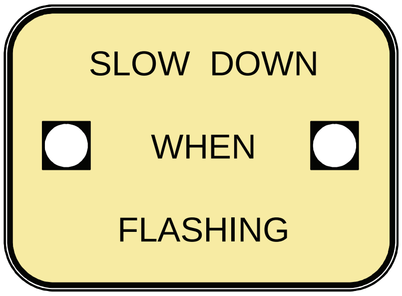


Figure 3-15. Warning sign alerting drivers to unsafe speed.

One system that was evaluated used loops spaced 16 ft (4.8 m) apart to measure speed.[54] A display was attached to a bridge structure downstream of the loop to alert the driver. If the vehicle was detected traveling 62 mi/h (100 km/h) or less, only the speed was displayed. If the vehicle was traveling faster, an additional SLOW DOWN message was displayed along with the speed. The study concluded that a speed detection system was effective in inducing drivers to reduce vehicle speed.

Unfortunately, at one such installation, drivers were observed deliberately accelerating to see how high a reading they could achieve on the speed sign. A potential solution to this problem would be to display speed values only up to the existing speed limit. Any speed above the limit would receive the message SLOW DOWN or YOU ARE EXCEEDING SPEED LIMIT.

Another speed measuring sensor was designed using modulated light emitting diodes. It was deployed to measure the speed and height of high and long trucks entering a curved freeway-to-freeway interchange. The diodes operated in the near-infrared spectrum at 880 nanometers. The signal modulation prevented interference from other sources of infrared energy, including sunlight. Two transmitter-receiver systems measured the vehicle speed and one measured the vehicle height. When trucks susceptible to rollover or jackknifing were encountered, flashers were activated to warn drivers to reduce speed.[55]

COORDINATED OPERATION OF FREEWAYS AND SURFACE ARTERIALS

Corridor control applications that coordinate traffic flow on freeways and major surface arterials require improved monitoring of traffic to support faster incident detection, quicker prediction and notification of congested locations, ramp queue detection, and motorist information services. The strategies below assist in optimizing the design of corridor control systems and maximizing their utilization:

- Coordination of actions between local agencies by sharing information and providing a unified decision support mechanism.
- Balancing traffic flow between freeways and arterials to minimize the effects of capacity-reducing incidents and maximize the use of existing roadway capacity.
- Providing motorists with access to current traffic information that assists them in planning their routes and avoiding congested areas.

The I–10 Smart Corridor Program in Los Angeles, CA, applies this congestion management approach. The service region consists of the freeways and arterial surface streets in a geographic area along a 14-mile segment of the I–10 (Santa Monica) Freeway. Building on the existing infrastructure, it adds new or modifies existing capabilities, resources, and policies to provide interagency coordination that maximizes agency effectiveness. An expert system helps automate responses to congestion along the corridor. Fourteen changeable message signs (also referred to as variable message signs), 24 trailblazer signs, 350 real-time traffic controlled signals, 45 ramp meter stations, and about 3,000 inductive-loop detectors are used in the corridor.

The Information for Motorists (INFORM) corridor in the New York City metropolitan area is another example of the corridor approach to traffic management. INFORM incorporates 136 mi (219 km) of roadway, consisting of two freeways (the Long Island Expressway-I–495 and the Northern State Parkway-Grand Central Parkway combinations) and a number of parallel and crossing arterial streets. The corridor extends east from the Borough of Queens in New York City, through Nassau County and into Suffolk County. This corridor uses 80 changeable message signs, 112 real-time controlled signals, 70 ramp meter stations, and about 2,400 inductive-loop detectors.

TRAFFIC DATA COLLECTION

Data collection is an essential part of any traffic engineering, planning, or operational activity. Advanced traffic management and traveler information services are supported by the collection of real-time traffic flow information on highway segments and surface street networks, especially during peak traffic flow periods.

Advanced traffic management and traveler information systems require real-time, online traffic data to effectively:

- Operate traffic adjusted, traffic responsive and traffic adaptive signal systems for surface streets and highways.
- Determine location and extent of highway congestion.
- Regulate congestion pricing.
- Operate automatic hazard warning devices.
- Detect and verify traffic incidents.
- Develop a historical traffic flow database to support planning and evaluation.
- Provide traveler information and traffic advisories.

Traffic management data requirements are dependent on the application, whether it is to support real-time operational strategies, offline planning and administration, computation of measures of effectiveness, compilation of related statistics, verification of proper sensor operation, or research. The following sections discuss the traffic management functions supported by real-time and offline data acquisition and analysis and measures of effectiveness that require data collection to evaluate the performance of a traffic management strategy.

DATA FOR REAL-TIME TRAFFIC OPERATIONS

Table 3-4 lists typical demand and capacity management services and strategies that support real-time traffic operations. Demand management is typically advisory, while capacity management involves enforceable controls. Each strategy has its individual data requirements, including data type and format, measurement and update rates, precision, and data transmission bandwidth. For example, automatic incident detection associated with real-time freeway traffic management requires, as a minimum, traffic volume, occupancy, and speed data updated in 20- to 30-second intervals. Volume and occupancy provide measures of congestion and alert operations personnel to incidents. Speed data provide estimates of travel delay and level of service and are inputs to incident detection algorithms.

Environmental sensors are needed to supply data about current weather conditions and information that leads to warnings concerning blowing sand, roadway icing, or the presence of other hazards. This information can augment and assist in the interpretation of volume, occupancy, and speed data. Travelers may report some incidents over cellular or roadside telephones, while others are deduced from the volume, occupancy, and speed data provided by sensors. Once an incident is verified, complementary data from patrol vehicles or surveillance cameras are needed to determine incident severity and the resources required to clear the incident. Cameras, in turn, require the transmission of control signals for pan, tilt, zoom, and focus and the return of video imagery to the operations center.[49]

Traveler information services provided by kiosks, internet web sites, and roadside devices require the transmission of data that reflect the real time operational state of the highways, including location of incidents, predicted travel times, suggested alternate routes and transportation modes, and maintenance, construction, and special event locations and impacts. Often interagency and public-private partnership agreements are required to implement these strategies.

Table 3-4. Strategies supporting real-time traffic management.[49]

Service	Strategy
Demand management	
Traffic demand management	High occupancy vehicle lanes, ramps, and connectors
	Toll road congestion pricing (public and private)
	Electronic road pricing (ERP) for entry into central business district during congested periods
	Diversion to surface streets
	Alternative transportation mode identification
	Fees for single occupant use of high occupancy vehicle lanes
	Parking advisories
Motorist and traveler information	Changeable message sign activation and updating
	Highway advisory radio operation
	Kiosk maintenance
	Web site maintenance
	Personal information access device data transmission
	Information service provider data exchange
	Interagency coordination

Table 3-4. Strategies supporting real-time traffic management(49) —*Continued*.

Service	Strategy
Capacity management	
Traffic signal control	Adaptive signal control algorithms that calculate and adjust cycle length, split, offset, and green band interval (bandwidth) in real time
	Left turn lanes activated by vehicle detection
	Automated red light enforcement
Metering	Ramp metering
	Mainline metering
	Freeway-to-freeway metering
Incident management	Automatic incident detection
	Incident verification
	Incident removal
	End of queue warning
	Alternative route diversion
	Onramp closure and metering
	Incident response teams for large or severe incidents
	Interagency coordination
Hazard warning	Rain and fog detection and motorist alert
	Ice and snow detection and motorist alert
	Dangerous curve, merge area, high wind, and rough pavement warnings
	Construction and maintenance lane closures and alerts
	End of queue detection
	Overheight and overweight detection and alert
Commercial vehicle operations	Data exchange in support of electronic clearance, credentialing, and weigh-in-motion
Pollutant emissions enforcement	Monitoring of carbon monoxide, hydrocarbons, and nitrogen oxides emitted at the tailpipe and evaporated
	Mandating use of alternate transportation modes

Strategies for real-time arterial signal control generally u travel times from an upstream sensor to the stopline at the intersection, startup or lag time for vehicles discharging from the stopline, and discharge rates for vehicles at the stopline and in queues between an upstream sensor and the stopline. The types of data are gathered and the time interval over which the data is gathered depends on the category of traffic control system strategy being utilized.

Metering is performed through fixed time-of-day plans or by using sensors that monitor mainline traffic flow and adjust the ramp or feeder highway metering rates accordingly. In the latter application, the metering rate may be responsive to local mainline flow (e.g., flow upstream of the ramp) or to flow over a wider area or section of the freeway where metering among several ramps is coordinated.

Automated hazard warning requires sensors to detect inclement weather, road condition, or oversize or overweight vehicle. The information is then conveyed to the motorist through mobile or fixed changeable message signs, highway advisory radio, normal commercial radio, or special radios that allow sideband reception of traffic information. Commercial vehicles can receive warnings concerning overweight status through transducers mounted in the cab when weigh-in-motion sensors are installed. Large, lighted, and bright roadside signs are needed to warn drivers of fog. These signs must be repeated at frequent intervals to alert the motorist to the continued danger and reduced speed limits.[56,57]

Sensors used for electronic clearance and credentialing and weigh-in-motion enhance commercial vehicle operations. Analysis of transportation vehicle emissions also relies on sensors to gather the needed pollutant data.

Weigh-in-motion is the focus of two separate FHWA documents: McCall, Bill; Vodrazka, Walter C. Jr., States' Successful Practices Weigh-In-Motion Handbook, *Dec. 15, 1997, available from the National Transportation Library, http://ntl.bts.gov/card_view.cfm?docid=6243, http://ntl.bts.gov/lib/6000/6200/6243/wim.pdf*

Data Collection Guide for SPS WIM Sites*, August 31, 2001, www.tfhrc.gov/pavement/ltpp/spstraffic/tmg.pdf*

DATA FOR OFFLINE TRAFFIC PLANNING AND ADMINISTRATION

Much of the information collected to execute the real-time functions of a traffic management system serves a secondary role as a valuable data resource for operations analyses and reviews, new construction and safety planning, highway research, and other administrative and planning services. The offline applications support a broad spectrum of users and uses, each requiring specific data, accuracy, precision, and spatial and temporal sampling. In recognition of the importance of archived data, an Archived Data User Service (ADUS) has been added to the U.S. National ITS Architecture and a five-year plan developed to support ADUS implementation.[58] Table 3-5 describes the types of archived data generally of interest in offline applications.[49]

Table 3-5. Data supporting offline traffic planning and administration.[49]

Function/use	Data
Traffic census volumes	Vehicle volumes by variable time intervals
	Vehicle classification counts
	Average speed by time periods
	High occupancy vehicle priority entry and facility usage
	Ramp usage by variable time intervals
	Vehicles by person occupancy
	Person and vehicle miles of travel
Traffic congestion	Average speed or travel time by time-of-day periods
	Congestion delay measurement
Crash and incident events	Crashes, types, trends, and rates
	Incidents, types, trends, and rates
	Commercial vehicle crash and incident locations and causes
Event planning, construction, and maintenance lane closures	Traffic volumes, time distribution, congestion levels
	Volume/capacity relationships
	Lane closure statistics (spatial and temporal)
	Volumes from parking facilities and their impact on local arterials and freeways
	Equivalent single axle loads

Table 3-5. Data supporting offline traffic planning and administration[49]—*Continued*

Function/use	Data
Resource allocation and policy analyses	Spatial and temporal traffic activity levels
	Spatial and temporal resource deployment
	Measures of effectiveness to support cost effectiveness analyses
	Origin-destination (O-D) data to support trend and policy option studies
	Electronic fare, transit automatic vehicle identification, and O-D data for intermodal transportation and land use analyses
	Incident statistics by type, location, time, season, weather
	Flow rate, density, truck use, speed, O-D, etc. data for research
	Pollutant emission concentrations by vehicle type, time-of-day, flow rate, speed, and vehicle miles traveled for air quality analyses
	High occupancy vehicle and transit usage to support effectiveness analyses

DATA FOR EVALUATING MEASURES OF EFFECTIVENESS

The strategies used by a traffic management system should be evaluated periodically to assess the degree to which the system objectives are satisfied. Evaluations are also important to ensure that the needs of the stakeholders (e.g., the system operators and agencies that respond to real-time incidents and congestion, those involved with planning for future improvements, users of historical data, travelers, and allied agencies and private partners) are supported. The evaluation criteria are usually expressed in terms of measures of effectiveness. Many measures of effectiveness (MOEs) are based on sensor surveillance data collected by the traffic management system.

System-Wide MOEs

A number of measures have traditionally been used to quantify the performance of freeway systems. These include:[59-62]

- Total travel time in vehicle hours traveled (VHT): The product of the total number of vehicles using the roadway or roadway segment during a given time period multiplied by the average travel time of the vehicles.
- Total vehicle miles traveled (VMT): The product of the total number of vehicles using the roadway or roadway segment during a given time period multiplied by the average trip length of the vehicles.
- VMT/VHT: The average travel speed throughout the network.
- Vehicle delay: Excess travel time as compared to the free-flow travel time.

- Total minute-miles of congestion: Extent of freeway congestion in both time and space. This MOE was developed by the Chicago Area Expressway Surveillance Project to quantitatively measure freeway congestion.[(40)] Congestion is defined as a 5-min lane occupancy of 30 percent or more measured at a mainline sensor. When this condition occurs, the number of minute-miles of congestion at the mainline sensor is equal to the product of the minutes of congestion multiplied by half the distance between adjacent sensors upstream and downstream of the mainline sensor at which congestion is measured. Minute-miles of congestion is a summary measure that accounts for all the variables contributing to the traffic flow condition, e.g., wet pavement, accidents, disabled vehicles, and construction. Analysis of this congestion measure in the Chicago area has shown correlation with total travel time. A completely accurate sensor network is not required to implement this MOE.
- Emissions: Usually the concentrations of three pollutants targeted by the Clean Air Act Amendments of 1990 are calculated in the United States. These are carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NO_x) emitted at the tailpipe and through evaporative emissions in units of grams per mile and grams per hour.
- Fuel consumption: Gallons (or liters) of fuel per VMT by link and aggregated link data to obtain total gallons (or liters) of fuel consumed per day.
- Crash statistics: Crash rates, types of crashes (fatalities, injuries, rear end, sideswipe, property damage only, etc.), correlation with congestion level, and trends of these statistics can be used to document improvements in motorist safety.

DETECTION OF PRIORITY VEHICLES

Detection of priority vehicles is associated with signal preemption and signal priority. Preemption interrupts or alters the normal signal sequence at an intersection in deference to a special situation, such as the passage of a train, a bridge opening, or the granting of the right of way to an emergency vehicle such as a fire engine, ambulance, or police car. Signal priority holds the green longer than normal or allows the signal to change to green as soon as possible in order to provide priority treatment for transit vehicles.

Priority vehicles may either be emergency vehicles such as fire engines, ambulance, and police cars, or mass transit vehicles such as buses and trolleys. Emergency vehicles *preempt* the right-of-way by interrupting the normal phase sequence of the signals along the route to allow for their safe and rapid passage. After a preset time period, the signal returns to normal operation. With railroad preemption, train predictors (sensors provided by the railroad) detect the approach of a train and trigger a reaction, which first clears the track area and then allows nonconflicting green phases to operate during the passage of the train. Transit vehicles receive *priority* treatment by extending the green signal phase or changing the signal to green as soon as possible to expedite movement of the vehicle along its route. Transit priority is used less frequently than preemption for emergency vehicles, primarily because of concerns over disruption to surrounding traffic.[(63)] The response to different types of priority vehicles can be coordinated among several agencies and jurisdictions to provide a seamless travel corridor. In urban and rural traffic, travel time savings of 10 to 30 seconds and savings greater than 30 seconds per intersection, respectively, were reported when an emergency vehicle preempted a red signal during an ITS field operational test (FOT).[(64)]

Some emergency systems preempt the right-of-way without using sensors to detect emergency vehicles along the route. In these systems, a green indication is displayed at all signalized intersections along the selected route traveled by the emergency vehicle. The right-of-way is assigned by activating a switch from a central location, such as a fire station. A number of centrally controlled computerized signal systems preprogram the emergency progression, which leads to more efficient traffic flow than the manual switch-activation technique.

Detection systems used to detect priority vehicles either use transmitters mounted on the vehicles or roadside-mounted sensors that identify the priority vehicle by its signature (i.e., the information they receive by regularly scanning the roadway for the presence of these vehicles).

There are two broad classes of priority vehicle detection systems. The first class uses transmitters or beacons mounted on the vehicle to identify it to receivers located near the controlled signal. Some transit vehicles transmit a signal to a central station, which in turn sends a priority command to the signal. The first class of priority vehicle detection systems includes high-intensity light beacons, RF beacons, infrared beacons, and inductive loops operating in conjunction with vehicle-mounted identification devices. The second class of priority vehicle detection systems does not require a beacon or transmitter on the vehicle. These sensors identify the priority vehicles through information they receive from the vehicle signature produced by the sensor. Included in this class of sensors are inductive loops that classify vehicles using algorithms that are incorporated into the detector electronics, laser radar sensors, and sound detection systems.

HIGH-INTENSITY LIGHT BEACON

One of the first priority systems developed is the high-intensity light emitter/receiver system, which flashed light on and off at a high frequency coded to identify different vehicle types.(65) The receiver is able to distinguish between emergency vehicles coded for command preemption and transit vehicles that receive only priority treatment. Each intersection to be preempted is equipped with one or more optical receivers, depending on the number of approaches that provide preemption. As the emergency vehicle with the light emitter approaches the intersection, the appropriate directional receiver senses the coded light, triggering the preemption circuitry in the intersection controller. When the light from an emergency vehicle is detected, a signal is conveyed to a phase selector (connected to the controller) or directly to the controller. The status of the controller is examined and the controller either holds the green interval for the emergency vehicle or terminates the green on the opposing street(s), thus transferring the green to the direction of travel of the emergency vehicle. Priority vehicle operation is similar, except that the control function is not preempted. If the signal is green, the controller attempts to hold the green long enough for the priority vehicle to enter. If the light is not green, the controller attempts to activate the green phase as quickly as possible. Newer transit priority systems, which are centralized, first transmit the request for priority from the vehicle to a centralized transit management center and then from the center to the affected intersection. The components of the high-intensity light beacon system are similar to those shown below for the infrared beacon.

RADIO BEACON

Another early signal priority system was the radio beacon. Each emergency or transit vehicle is equipped with a RF transmitter and each controlled intersection with a receiver. Since the transmitter in this system broadcasts the RF frequencies in all directions, it must also incorporate a direction code or provide limited preemption such as having all signals flash red. Furthermore, the detection range must be limited to avoid preempting nearby signals unnecessarily.

Another form of RF priority system is found on the buses in Aalborg, Denmark.[66] These vehicles are equipped with a driver's console, computer, global positioning system (GPS), and data radio. The driver inputs pertinent details about the current trip and an identifying journey number. The console informs the driver whether the vehicle is ahead or behind schedule, at which point the driver inputs the journey time in plus or minus form on the console display. Priority request information is transmitted by the vehicle to a centrally located base station when the vehicle crosses a report line. The base station can analyze all priority requests received and, if necessary, decide among conflicting requests before relaying the priority command to the affected signal. The base station also downloads GPS correction data every 10 to 20 seconds and new schedules and report lines as needed.

The GPS approach has become increasing popular in the United States as well. Communication techniques include transmission of priority requests from the bus to the control center by digital channels on the bus radio system and transmission of the signal priority command from the control center to the intersection using spread spectrum radio or optical links.

INFRARED BEACON

The infrared priority system relies on beacons or transmitters mounted on vehicles. As the vehicle approaches the intersection, it saturates the intersection with an encoded, infrared signal that is received, decoded, and validated to give the requesting vehicle priority. The type of signal priority is programmed into the electronics unit located in the controller cabinet. Figure 3-16 shows an example of beacon, receiver, and electronics units.

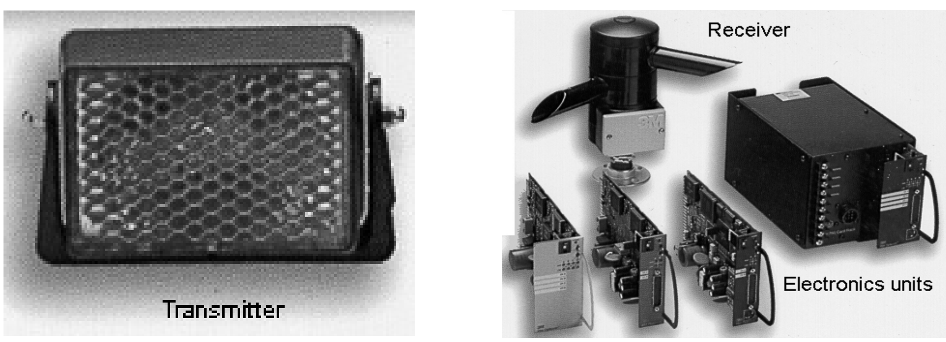


Figure 3-16. Infrared beacon priority system (Courtesy of 3M Company, St. Paul, MN).

INDUCTIVE LOOP WITH VEHICLE-MOUNTED TRANSMITTER

Inductive-loop detectors, in conjunction with electronics units designed to receive dozens of different codes from transmitters mounted on the underside of the vehicles, allow discrimination among emergency and transit vehicles. The transmitter in Figure 3-17 is disk shaped and is attached to the vehicle with a threaded shaft that extends from the disk. Its continuously transmitted programmed code is detected by any standard loop in the roadway. A special digital electronics unit interprets the code and identifies the vehicle or vehicle type. The electronics unit also provides conventional vehicle detection data. In addition to preemptive and priority control, the vehicle-mounted transmitter can also be used to override gate control, recognize vehicles at control positions or gasoline pumps, and correlate transit bus passage with schedules.

INDUCTIVE LOOP WITH VEHICLE CLASSIFICATION ALGORITHM

Another type of digital inductive-loop detector electronics unit eliminates the need for mounting a special transmitter on the vehicle. This system consists of a conventional loop buried in the roadway and a special digital electronics unit that transmits a high frequency excitation signal to the loop and receives a unique waveform (also termed the "signature" or "footprint") corresponding to each vehicle class it detects. When a waveform similar to that expected for a priority vehicle is identified, priority signal treatment is provided. The system can also be used to classify vehicles as described in Chapter 2.[67]

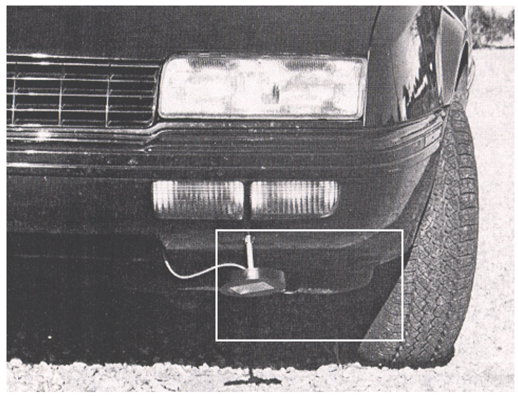


Figure 3-17. Transmitter mounted under a vehicle identifies it to a subsurface inductive-loop detector.

Its application to bus detection is illustrated in Figure 3-18. The digitized signature is input to a microprocessor that seeks and compares preidentified features to those of known bus profiles stored in memory. If a bus is identified, an output is generated to request priority treatment. Typical signatures for various classes of vehicles are shown in Figure 3-19.

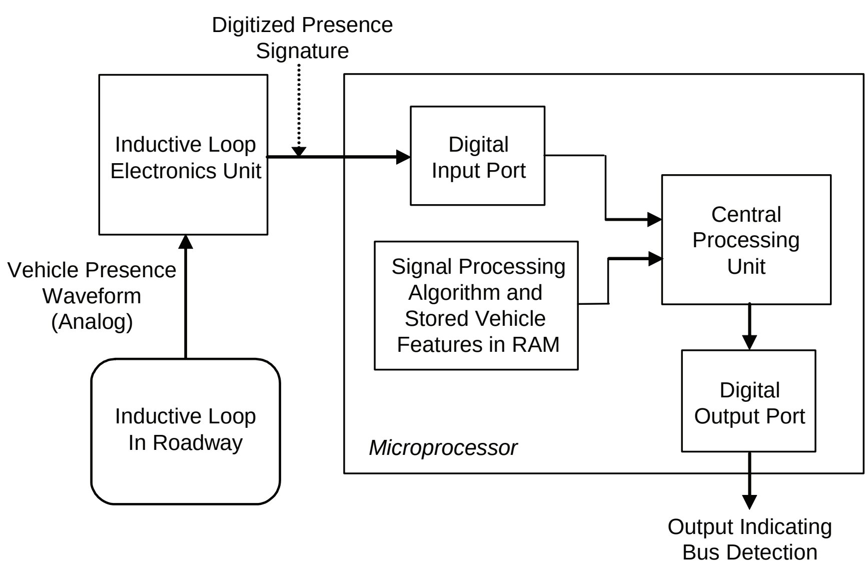


Figure 3-18. Inductive-loop detection system for transit vehicles.

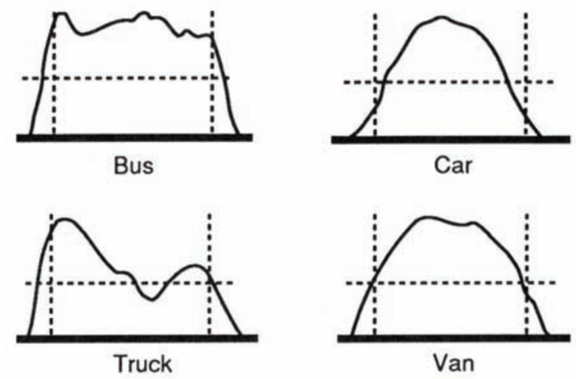


Figure 3-19. Representative vehicle signatures obtained from loops excited by high frequency signals from specialized electronics units.

LASER RADAR

Laser radar sensors that scan one or more traffic lanes are capable of supplying unique vehicle profiles. The vehicle's profile is classified by an algorithm incorporated into the sensor or stored in an auxiliary computer connected to the sensor through a serial interface. The traffic signal priority is adjusted according to the vehicle class. Although these sensors have the potential to differentiate between emergency and transit vehicles, this application of the sensor has not been demonstrated.

SOUND DETECTION SYSTEMS

Sound detection systems, such as that illustrated in Figure 3-20, detect the sound energy emitted by sirens on emergency vehicles. Optional ultrasonic emitters are available to identify transit vehicles. A set of phased-array microphones, mounted on the mast arm, provides directionality to the source of the sound. A light can also be mounted to give visual verification that the signal priority request has been received.

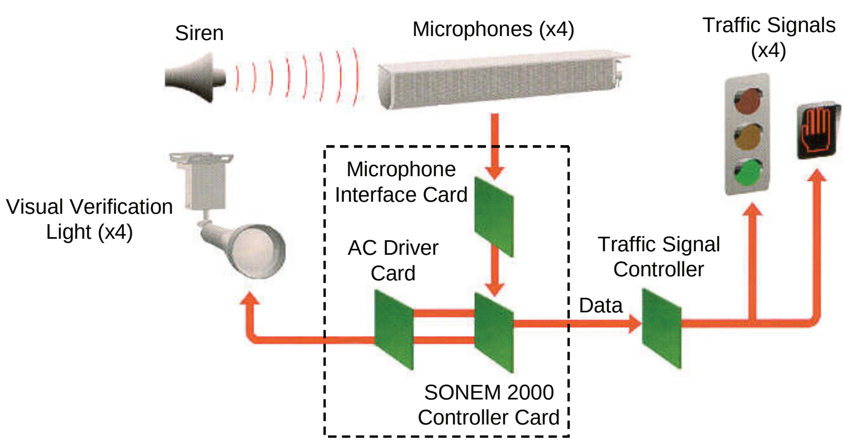


Figure 3-20. Sound detection priority system for 4-way intersection control (Courtesy of Sonic Systems, Vancouver, British Columbia, Canada).

PEDESTRIAN DETECTION AND SIGNAL ACTUATION

Pedestrian detection is often a major consideration when developing traffic signal timing plans. Manually operated and automatic pedestrian sensors may be installed for this application.

The applications described above involved the detection of vehicles. However, properly timed actuated signal control also requires the detection of pedestrians. Unlike vehicles, pedestrians do not induce changes in the Earth's magnetic field, cause inductance variations, or otherwise produce signals that are easily detected by other sensors (an exception might be video detection, although this application has not been exploited to date). Moreover, pedestrians cannot be depended on to follow a specific path toward their intended destination, nor can they be expected to take a specific action to make their presence known to the signal controller.

The push-button sensor, shown in Figure 3-21, is the most common form of pedestrian detection. The microswitch contact closure created when a pedestrian pushes the button causes a low-voltage current to flow to the controller and registers a "demand" for pedestrian service.

The weak link in the manually operated push-button system is the pedestrian, as it requires the pedestrian to be proactive in registering a demand. Unfortunately, many pedestrians do not make the necessary effort and, in those cases, are likely to cross the intersection illegally or unsafely. These pedestrians do not realize that pushing the button will extend their green time as well as service their needs faster.

Where two buttons for crossing in different directions are located on the same support, each button's relationship to the corresponding crosswalk must be clearly and unequivocally indicated. Although the *Manual on Uniform Traffic Control Devices* (MUTCD[68] specifies a series of standard signs for use with pedestrian push-button devices, many agencies augment these standard signs with additional instructional signs.

Figure 3-21. Manually operated pedestrian push button.

Active response pedestrian push buttons are common in the United States and Europe. When activated, they turn on a small light (usually green) or a small sign that flashes the message "WAIT PLEASE" or "WAIT FOR WALK." An example from England is shown in Figure 3-22. Such a response is a confirmation of the pedestrian's call for service similar to that of a lighted elevator button. This response to a call for a walk message appears to alleviate pedestrian anxiety from the lack of feedback from conventional manually operated push buttons and promotes understanding of the pedestrian phasing of the traffic signal.

Figure 3-22. Interactive pedestrian push button.

Another responsive crosswalk device is shown in Figure 3-23. Although not presently in the MUTCD, the device is increasing in popularity as it displays the time remaining for the pedestrian to cross the street.

Figure 3-23. Countdown timer showing time remaining for pedestrian crossing.

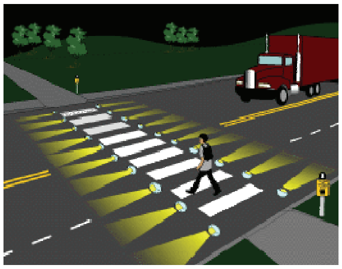
When visually impaired pedestrians use push buttons, some type of guiding device can be used to enable the pedestrian to locate the button. Texturing the concrete on the sidewalk approaching the push button, audible locators, or handrails can be used for this purpose.

Audible pedestrian signals are devices that emit buzzing, whistling, beeping, or chirping sounds that are correlated with the visual WALK/DON'T WALK signs used by sighted individuals. Intermittent pulses in the frequency range of 300–1,000 Hz (with 750 Hz being optimal) are the most effective sounds for the human ear to localize and do not require a high volume level to be effective. The devices may be either pedestrian activated or automatic. Pedestrian-activated signals are connected to the pedestrian call (push) button, which is connected to the signal controller. Automated signals are activated by the cycle change at pretimed intersections. Many of the automatic devices, such as that in Figure 3-24, emit different sounds to indicate which direction to cross and how much time is available for crossing. The two most popular audible pedestrian signals used in the United States emit either a buzzer or a birdcall sound. The audible signal most frequently used in the western United States emits a "peep peep" tone for the east-west direction and a “cuckoo” tone for the north-south crossings. Audible signals indicate only that the WALK indication is displayed, not that the intersection is clear. The devices are meant to complement rather than be a substitute for a visually impaired person's orientation and mobility skills.

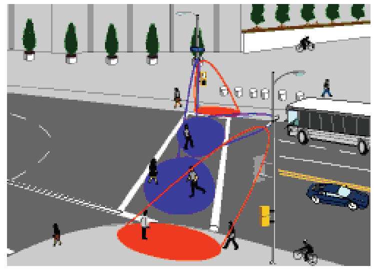
Figure 3-24. Audible pedestrian signal device.

Several innovative pedestrian signal alternatives have been tested to detect the presence of pedestrians and provide safer crossing intervals. Instead of traditional pedestrian pushbutton devices, which require actuation, the new signal installations use pressure-sensitive mats or infrared sensors (mounted on poles) to detect the presence of pedestrians waiting to cross the street. After being detected, a pedestrian is given a white light, indicating that a WALK signal will follow. If pedestrians wait until the WALK signal is displayed, a second infrared sensor monitors their progress within the crosswalk and holds the pedestrian clearance interval (and red signal for motor vehicles) until the pedestrian reaches the other side of the street. Thus, slower pedestrians (e.g., elderly, wheelchair users) are given extended walking intervals. If the pedestrian steps off the curb too soon or leaves the crossing area, the request for a WALK signal is no longer valid, and NO WALK interval is given.

Other systems that alert drivers to the presence of pedestrians are illustrated in Figure 3-25. The pedestrian crosswalk lights in Figure 3-25a only flash when activated by the pedestrian. Therefore, the motorist receives real-time information indicating that a pedestrian is in the vicinity of the crosswalk. Figure 3-25b shows automatic detection of pedestrians at curbside and within an intersection using microwave radar sensors that activate a pedestrian call feature. At the same time, slower pedestrians detected within the on-street detection zones receive more time to cross the street. The detection of motion in the target area (for example, curbside) has the same effect as a pedestrian pressing a call button.



(a) Crosswalk lighting



(b) Microwave radar sensing

Figure 3-25. Pedestrian crossing safety devices.

DRIVER WARNING FOR RED SIGNAL AHEAD

Curves, hills, inclement weather, and natural and man-made visual obstructions often require advance warning signs to be installed to alert drivers to a red signal that lies ahead. Such a warning provides the extra time needed to bring the vehicle to safe stop.

Advance driver alerts to a red signal ahead are often required when the geometrics of intersections do not permit the signal to be seen in time for drivers to react. In addition, there may be sight-distance restrictions due to overhead obstructions such as bridges and large trees, which cannot be removed. Intersections located on a downgrade may increase the actual required stopping distance.

Other potentially hazardous situations that involve traffic signals occur where signals are difficult to see at vertical curves that hide the queue at a signal, while not obstructing the signal itself. Another poor signal visibility situation arises in areas subject to dense ground fogs that reduce signal visibility below normal minimums.

To alleviate these dangerous situations and modify driver behavior, additional information may be provided to the driver and not be limited to a simple fixed sign that warns of a potential RED signal ahead. This type of tailored response is provided through changeable message advance warning signs. These signs have the capability to alter the displayed information based on whether the driver should slow or proceed. The warning sign shown in Figure 3-26 is used to alert approaching drivers that a hidden signal around a curve is red and that they must stop.

The criteria for a changeable message sign used in a red signal warning system include:

- A sign mounting adjacent to the roadway or overhead (at least 17 ft (5.2 m) above the pavement); the lettering should be at least 12 inches (30 cm) high with 12-inch (300-mm) alternating beacons.
- A legend that reads “PREPARE TO STOP WHEN FLASHING.”

Figure 3-26. PREPARE TO STOP lighted sign.

- A warning sign message and beacons that begin to flash 50 to 60 times per minute just prior to the signal display of the yellow change indication, so that the driver will observe the signal display just after it has turned yellow. The last car passage feature of some of the early density controllers can be used for this purpose.
- Continued flashing of the sign except when the approach signal is green.
- Additional control logic to ensure safety; that is, the sign must go to flash if the signal fails, is in conflict, or is placed in a manual or red flashing mode.

When warranted, this type of flashing sign provides an effective solution to the problem of rear-end accidents. When actuated in such a manner that it provides the approaching drivers with accurate information as to whether they will have to stop for the signal, it is a beneficial addition to the signal installation.

TRAFFIC COUNTING AND VEHICLE CLASSIFICATION

Roadway mounted sensors are often used for vehicle counting and classification, over-height detection, and road surface and weather reporting. Both in-roadway and over-roadway sensors are used in these applications.

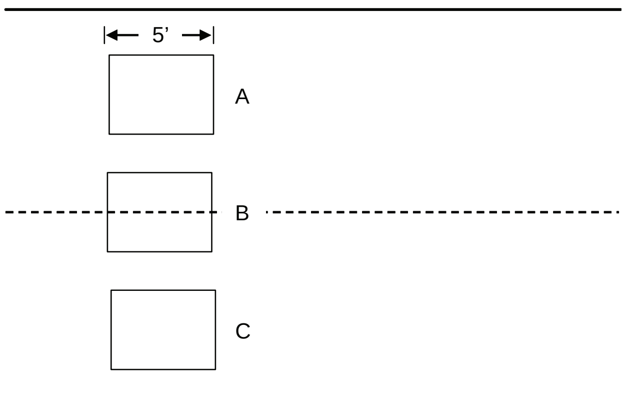
In many locations, inductive-loop detectors and over-roadway sensors have replaced pneumatic tubes for counting vehicular traffic. Since the tubes are laid over the roadway, they are especially vulnerable to the wear and tear of passing traffic. The buried loops or over-roadway sensors provide a better means of counting. Moreover, should the loop or over-roadway sensor fail, all counting would stop. On the other hand, the tube tends to distort counts prior to failure. Another favorable aspect is that the loop and over-roadway sensors give only a single output for most vehicles, regardless of the number of axles, thus providing a more accurate count.

In addition to conventional traffic counts, a growing number of computerized signal systems are using loop and over-roadway sensors to provide volume data. The data collected by the sensors are output directly to the system computers, which control signal timing throughout the system.

COUNTS BY LOOP INDUCTIVE DETECTORS

It is not appropriate to generate multiple-lane vehicle counts by extending a single loop across several lanes. If a second vehicle in a different lane moves over part of a loop before the first vehicle has left its part of the loop, only one continuous output will be registered. This could lead to significant undercounting.

When lane discipline is good (i.e., traffic stays in its own lane), a separate loop should be placed in each lane. However, in some cases lane discipline is poor (i.e., traffic is continually changing lanes) and lanes are wide. If count accuracy is important for these instances, then an additional loop could be placed between the lanes, as shown in Figure 3-27. This depicts the three-loop layout for a two-lane roadway.



1 ft = 0.3 m

Figure 3-27. Three-loop layout for counts.

Ideally, the loops should be placed according to the following constraints:

- The widest vehicle will not straddle more than two loops.
- The narrowest vehicle will not pass between any two loops.
- Two side-by-side vehicles must cross three loops.

The logic to be inserted into the vehicle count model is described as follows:(69)

- The operation of loops A or B or C produces one count immediately.
- Operation of loop A and C together or A, B, and C together produces one count and a second count after a short delay.
- The operation of single loops or adjacent pairs (A and B or B and C) produces only one count, but the operation of A and C together produces the first count followed by a second count a short time later.

COUNTS USING LONG LOOPS

Electronics units provide normal inductive-loop detector operation with timing functions (if desired) and also a secondary output of one pulse per vehicle in a single long loop or in a series of sequential short loops. Count accuracy for a single loop of any size is greater than 95 percent for properly functioning loop systems. Accuracy for the four loops in series configuration (discussed in Chapter 4) is lower because of the complexities associated with the analysis of its count output values.

DIRECTIONAL DETECTION USING INDUCTIVE LOOPS

When it becomes necessary to distinguish between the directions of travel (such as where two directions of traffic must use the same roadway facility as on a reversible lane), two loops, two electronics unit channels, and directional logic can be used as illustrated in Figure 3-28. With this system, separate counts are accumulated for each direction of travel. An alternative approach is to activate the appropriate loops and reversible lane control signals according to time-of-day programming.

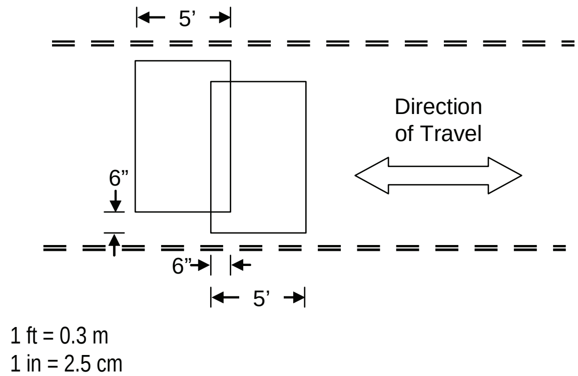


Figure 3-28. Directional detection using point sensors.

VEHICLE CLASSIFICATION SENSORS

As part of the traffic counting process, many agencies wish to obtain vehicle counts by class of vehicle. There are several vehicle classification counters commercially available. Many of these devices use loops and axle sensors to obtain the information required to classify vehicles as was discussed in Chapter 2. Other approaches use laser radar sensors to obtain vehicle class data. Many video image processors and presence-detecting microwave radar sensors can classify vehicles into three to five user-selected length bins.

OVERHEIGHT SENSORS

Overheight sensors are deployed at approaches to tunnel portals, bridge underpasses, restricted height parking structures, and at other covered structures such as those found at weigh stations and toll facilities. Two types of sensor technologies are used for overheight detection. The first is an optical system that detects overheight vehicles when they interrupt or break the transmission of the beam from transmitter to receiver as depicted in Figure 3-29. The second, a laser radar, functions in much the same manner, except that it transmits longer wavelength energy in the near-infrared spectrum.

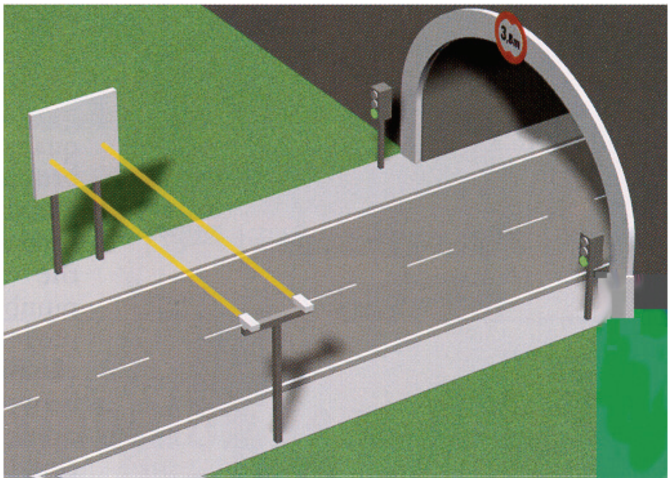


Figure 3-29. Overheight sensor configuration (Picture courtesy of MBB SensTech, Munich, Germany).

WEATHER SENSORS

A variety of data can be gathered from weather information systems mounted on or near the roadway, including wind speed and direction, air temperature and humidity, precipitation type and rate, visibility, roadway surface temperature, roadway wet/dry status, and road surface chemical analysis for determining whether application of ice-reducing chemicals is warranted. Many of these systems are solar powered and can transmit data over wireless communications links. Specialized forward-scatter sensors are available for measuring visibility through fog and using the information to control brightly lighted signs that warn drivers of the reduced visibility conditions ahead.[57] Weather monitoring systems, such as the one in Figure 3-30, integrate many sensors to provide specific data for an application. Still other systems provide computer interfaces that display video and data on a monitor as illustrated in Figure 3-31.

Environmental observation systems are beneficial for tracking the location and movement of fronts and storm cells and providing needed information to traffic management agencies. In order for roadway weather sensors to obtain accurate data, they must be calibrated regularly. Such calibrated sensor data could then also be used to support the needs of the National Weather Service and improve the accuracy of pavement temperature forecast models.

In an evaluation of research on road weather observations and predictions, the Desert Research Institute (DRI) and the Nevada Department of Transportation (NDOT) concluded that Road Weather Information Systems (RWIS) provide data that improve weather forecasts, but that challenges remain in forecasting when ice will form on roads.[70] In tests of models that predict ice formation as a function of meteorological parameters, DRI found a large variation in the minimum pavement temperature from 19 °F to 10 °F (−7 to −12 °C). Minimum pavement temperature appeared to be most sensitive to air temperature changes, total cloud cover, and precipitation. Inaccuracies in air temperature of 34 and 36 °F (1 and 2 °C) caused a change in minimum pavement temperature of more than 33 and 34 °F (0.5 and 1 °C), respectively. Significant underestimation of precipitation changed minimum pavement temperature by about 36 °F (2 °C). Inaccurate cloud cover predictions caused changes in minimum pavement temperature of more than 37 °F (3 °C).

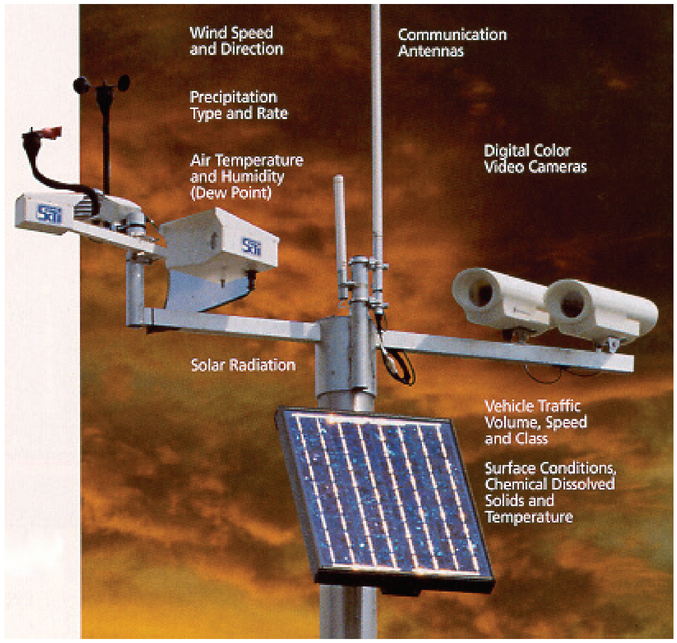


Figure 3-30. Integrated weather information system (Picture courtesy of Nu-metrics, Uniontown, PA).

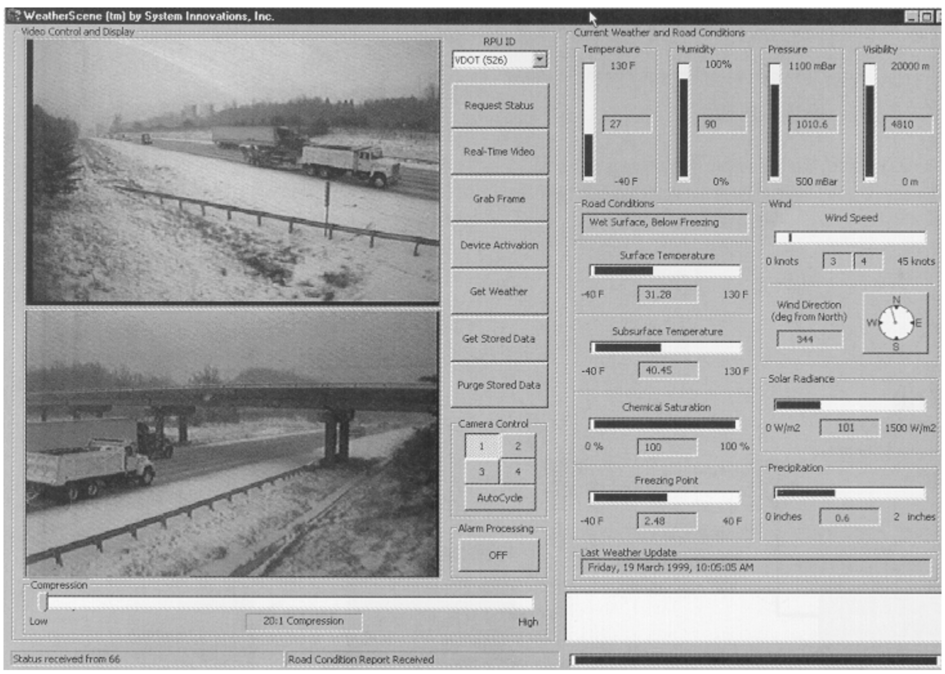


Figure 3-31. Weather imagery and data display (Picture courtesy of System Innovations, Fredericksburg, VA).

Based on a 2004 National Academy of Sciences report that described the need for a robust, integrated weather observational and data management system, FHWA responded with an initiative called Clarus—the Nationwide Surface Transportation Weather Observing and Forecasting System.(71,72) The initiative's overall objective is to reduce the impact of adverse weather for all road and transit users and operators. The projected six-year Clarus program has been structured to include:

- Developing partnerships between the surface transportation and weather communities to leverage and share resources for both research and operations.
- Strengthening of ties among Federal agencies with similar objectives such as FHWA and the National Oceanic and Atmospheric Administration (NOAA).
- Demonstrating a framework to collect the Nation's current and future surface transportation weather and road condition observations, and provide quality assured data as input to advanced weather models, as the basis for informative value-added products that can contribute to a safer and more efficient surface transportation system.
- Establishing an instrumented corridor test bed to host new cutting-edge technologies for fixed, mobile, and remote sensing.
- Establishing an Interagency Coordination Committee.

VEHICLE-MOUNTED SENSORS THAT ENHANCE SAFE OPERATION

The requirement for increased vehicle safety on arterials and freeways has led to the development of vehicle-mounted sensors that support advanced cruise control functions such as headway and proximity detection, and issue warnings when lane divider markings are crossed.

Table 3-6 gives examples of vehicle-mounted sensors that support various collision and hazard avoidance strategies and prevent specific types of crashes.(73) Headway detection advises the driver of an imminent crash with an obstacle or to maintain a safe distance when following another vehicle. Proximity detection gives the driver information about adjacent lane vehicles or obstacles while backing up. The lane position monitor advises the driver if the vehicle is drifting out of its travel lane. In-vehicle signing conveys posted or dynamic information to the driver from the traffic management infrastructure. A gap acceptance aid advises the driver when it is safe to cross or turn at intersections. Vision enhancement gives the driver a clearer image of the roadway and environment ahead during reduced visibility conditions such as heavy rain, fog, and nighttime.

However, sensors are not the only component of a hazard warning system. Human response time, the devices used to transmit the sensor information to the driver, and the dynamics of the encounter with a particular hazard are other factors that must be analyzed in designing a hazard warning system. A systems analysis approach reveals that uncertainty in human response time is the major factor affecting performance of an alerting system.(74) Thus, a system designed to protect slowly responding drivers may have a high unnecessary alert rate for rapidly responding drivers. Likewise, systems designed to have few unnecessary alerts also have relatively high crash rates. The way to combat this tradeoff is to decrease the response time uncertainty through driver training, improved display and audio alerts, or tuning each system to the driver using it. Although the systems methodology has been successful in modeling the performance of relatively simple systems such as a rear-end collision alert system, its extension to other applications requires more detailed dynamic models, probability distributions for a collision given an alert (obtained through Monte Carlo simulation), and inclusion of representative hazard variables such as driving speed and road condition.

Table 3-6. Collision avoidance strategies and preventable crash types that are supported by vehicle-mounted sensors.

Strategy	Vehicle-mounted sensor	Preventable crash
Headway detection	Millimeter-wave radar, laser radar, video	Rear-end
Proximity detection	Millimeter-wave radar, ultrasonic	Backing, lane change and merge
Lane position monitor	Laser radar, video, passive reflected light detector	Road departure, opposite direction
In-vehicle signing	Video, infrared communications, microwave transponder	Road departure, opposite direction, intersection, and crossing path
Gap acceptance aid	Video	Intersection and crossing path
Vision enhancement	Millimeter-wave radar, passive far infrared, intensified CCD cameras	Reduced visibility

IN-VEHICLE SENSORS FOR DISTANCE WARNING, CRUISE CONTROL, AUTOMATIC HIGHWAY SYSTEM FUNCTIONS, AND PRECRASH DETECTION

Automobile, truck, and bus use of radar sensors is projected to grow from worldwide revenues of $35 million (U.S.) in 1998 to between $300 and $500 million (U.S.) by 2003.(75) Automotive distance warning systems are a large part of this market, including front, side, and back radar to monitor obstacles. These systems calculate the distance from objects and their speed, if they are moving, and alert drivers if they are too close to another object or vehicle. Table 3-7 lists composite requirements for automotive radar distance warning systems.(76-78)

Radars designed for advanced adaptive cruise control provide the input to systems that automatically reduce vehicle speed and apply the brakes as the separation between approaching vehicles diminishes. Other smart cruise systems rely on highway-based sensors and appropriate in-vehicle systems to inform the driver of the presence and location of roadway obstacles, vehicles in an intersection, road surface condition, and pedestrians crossing streets.(79) A still more futuristic look at automatic vehicle control and guidance includes automatic highway systems. Here sensors are required to:

- Validate proper operation of the automotive systems (e.g., braking, steering, throttling, and communications).
- Support automatic guidance (lateral and longitudinal), acceleration, and deceleration of the vehicle on the automated highway.
- Provide coupling (and uncoupling) of the vehicle into (and out of) a platoon.
- Detect and report the presence of obstacles on the highway.(80,81)

Table 3-7. Automotive radar requirements for distance warning systems.

Parameter	Dimensioning criteria	Value
Center frequency	Compatible with government frequency allocation for the application	24 or 76 to 77 GHz
Waveform	Capable of detecting presence and measuring the relative speed of vehicles within the range measurement interval	Frequency modulated, continuous wave (FMCW)
Azimuth field-of-view	Large enough to detect a vehicles at sufficient range when the following vehicle is on a curved road section	±8.5 degrees
Antenna horizontal beamwidth	Obtain a sufficiently high probability that all relevant vehicles will be resolved from each other	1.5 degrees
Range measurement interval	Sufficient to include the distance to a car stopping ahead to the distance needed to reduce speed comfortably when a car is overtaken	1 to 200 m (3.3 to 656 ft)
Relative velocity of vehicles	Ability to measure relative speed of approaching and departing vehicles	-50 to +100 m/s (-112 to +224 mi/h)
Update rate	Fast enough to match the vehicle longitudinal control servo bandwidth with the reaction times	10 to 20 Hz
Mounting position	Obtain view of critical areas with radar waveform able to penetrate vehicle body materials, such as by placing antenna behind a plastic bumper or plastic grill	—

Radar sensors operating over shorter distances are finding applications in park distance control and airbag precrash detection, especially for side airbags. The frequencies being considered for these radars are 24 and 77 GHz. Laser radars can also potentially be used for the precrash detection application. Ultrasonic sensors are available from original equipment and aftermarket manufacturers for park distance control and obstacle detection.

IN-VEHICLE OPTICAL SENSORS FOR PASSENGER DETECTION, VISION ENHANCEMENT, AND LANE CONTROL

In-vehicle sensing functions are likely to be performed by optical sensors, either in the visible or infrared wavelength bands. Three-dimensional camera chips for passenger detection are in development for this purpose. In-vehicle sensing is proposed to control inflation of the smart airbag based on the shape and position of the occupant on each seat. Occupant data could also support automatic validation of high occupancy vehicle lane use and calculation of fees for occupant-based services.

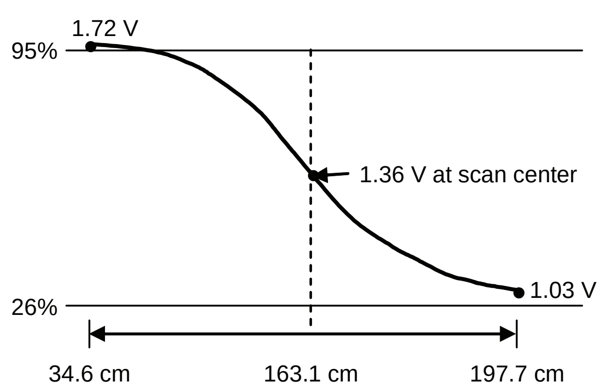
Night vision systems that utilize uncooled infrared detectors such as barium strontium titanate are scheduled for insertion into high-end passenger vehicles beginning with year 2000 models.(82) Lane tracking systems, such as Autovue shown in Figure 3-32, aid drivers of trucks and passenger vehicles in staying within prescribed lanes. This in-vehicle camera detects painted lines or other markers at the lane edges and sounds an alarm to alert the driver when the vehicle deviates from the center of the lane, but before the vehicle crosses the lane marker.

Another lane deviation sensing system utilize a laser diode to scan the roadway for paint that serves as a lane demarcator. In addition to the small glass spheres embedded in the paint, an external coating of spheres is often applied to a painted stripe to enhance the retroreflectivity of the marking. Since the road surface generally has low reflectivity, the energy backscattered from the paint and road surface is different and detectable by a scanning laser sensor.[83]

Figure 3-32. Autovue lane tracking sensor (Photograph courtesy of Iteris, Inc., Anaheim, CA).

A third method proposed for lane position monitoring utilizes a passive sensor, available light, and current road striping practice. The sensor consists of a cylindrical lens and a silicon crystal detector. Figure 3-33 shows the voltage generated as a prototype sensor scans across the lane centerline. The voltage is proportional to the location of the line's centroid.[84] The scan voltage is monotonic, smooth, and fairly linear in the central 50 percent of the plot. In this example, the detector output voltage is nonzero when the lane is physically centered on the detector. This indicates that overall background light falling on the detector is nonuniform, in this case produced by external sources in the sensor's field of view. Baffling is under consideration to eliminate the voltage offset.

Voltage is proportional to centerline location in the sensor's field of view. The contrast between the lane lines and the background in reflected sunlight was 0.57. The scan length was approximately 64 inches (163 cm) and the sensor declination was 21 degrees.



1 cm = 0.4 inch

Figure 3-33. Output voltage generated by passive lane positioning sensor.

REFERENCES

1. Kell, J.H., and I.J. Fullerton. *Manual of Traffic Signal Design*, Institute of Transportation Engineers. Prentice-Hall, Englewood Cliffs, NJ. 1982.
2. Gordon, R.L., R.A. Reiss, H. Haenel, E.R. Case, R.L. French, A. Mohaddes, and R. Wolcott. *Traffic Control Systems Handbook*, FHWA-SA-95-032. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. February 1996.
3. Neudorff, L.G., J.E. Randall, R. Reiss, and R. Gordon. *Freeway Management and Operations Handbook*, FHWA-OP-04-003. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. September 2003. http://ops.fhwa.dot.gov/Travel/traffic/freeway_management_handbook/notice.htm (Accessed Nov. 26, 2004.)
4. Wyman, J.H., and R.N. Lyles. "Evaluation of Speed Monitoring Systems," Maine Department of Transportation. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. August 1980.
5. *Highway Capacity Manual* 2000. Transportation Research Board, National Research Council, Washington, DC. 2000).
6. Rowe, E. "The Los Angeles Automated Traffic Surveillance and Control (ATSAC) System," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1. February 1991. pp. 16–20.
7. Gartner, N.H., C. Stamatiadis, and P.J. Tarnoff. "Advanced Traffic Control Strategies for Intelligent Vehicle Highway Systems," *Proc. SPIE*, vol. 2344. Bellingham, WA. November 1994, pp. 73–80.
8. Gordon, R.L. *Systems Engineering Processes for Developing Traffic Signal Systems: A Synthesis of Highway Practice*, NCHRP Synthesis 307. Transportation Research Board, Washington, DC. 2003.
9. Robertson, D.I. "Optimum Control of an Intersection for Any Known Sequence of Vehicle Arrivals," *Proceedings of the Second IFAC/-IFIP/IFORS Symposium on Traffic Control and Transportation Systems.* North-Holland, Amsterdam, Netherlands. 1974. pp. 3–17.
10. Robertson, D.I. "Traffic Models and Optimum Strategies of Control: A Review," *Proceedings of the International Symposium on Traffic Control Systems.* 1979. pp. 262–288.
11. Hunt, P.B., D.I. Robertson, R.D. Bretherton, and R.I. Winton. *SCOOT—A traffic responsive method of coordinating signals*, TRRL Report 1014. Transport and Road Research Laboratory, Crowthorne, Berkshire, England. 1981.
12. Hunt, P.B., and T.R. Holland. "The Effect of an Incident in a SCOOT System in Simulation and on Street," *Traffic Engineering and Control.* February 1985.
13. Robertson, D.I. "Research on the TRANSYT and SCOOT Methods of Signal Coordination," *ITE Journal.* January 1986. pp. 36–40.
14. Robertson, G.D. "Handling Congestion with SCOOT," *Traffic Engineering and Control.* 1987. pp. 228–230.

15. Robertson, D.I., and R.D. Bretherton. "Optimizing Networks of Traffic Signals in Real Time—The SCOOT Method," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1. February 1991. pp. 11–15.

16. Sims, A.G. "The Sydney Coordinated Adaptive Traffic System," in Urban Transport Division of *ASCE Proceedings*, Engineering Foundation Conference on Research Priorities in Computer Control of Urban Traffic-Systems. 1979. pp. 12–27.

17. Luk, J.Y.K. "Two Traffic-Responsive Area Traffic Control Methods: SCATS and SCOOT," *Traffic Engineering and Control.* 1984. pp. 14–20.

18. Gartner, N.H., M.H. Kaltenbach, and M.M. Miyarnoto. "Demand-Responsive Decentralized Urban Traffic Control, Part I: Single Intersection Policies (Feb. 1982), Part II: Network Extensions (July 1983)." U.S. Department of Transportation, Office of University Research, Washington, DC. 1983.

19. Lees, J.J. "The TRAC System," ARR 160, Australian Road Research Board, in L.Y.K. Luk, Ed., *Proceedings—Urban Traffic Control Workshop*, 14th ARRB Conference, Canberra, Australia. August 1988. pp. 35–53.

20. Ceder, A., and A. Shmilovits. *A Centralized Traffic Signalization Control System Coordinated with Supporting Units*, Working Paper. Technion—Israel Institute of Technology, Civil Engineering Department, Transportation Research Institute, Haifa, Israel. 1991.

21. Khoudour, L., J-B. Lesort, J-L. Farges, "PRODYN—Three Years of Trials in the ZELT Experimental Zone," *Recherche Transports Sécurité*, English issue no. 6*, Special Traffic Management.* 1991. pp. 89–98.

22. Vincent, R.A., and C.P. Young. "Self-Optimizing Traffic Signal Control Using Microprocessors—The TRRL 'MOVA' Strategy for Isolated Intersections*," Traffic Engineering and Control*, vol. 27, no. 7–8, July–August 1986. pp. 385–387.

23. Vincent, R.A., and J.R. Pierce. *MOVA Traffic Responsive Self Optimizing Signal Control for Isolated Intersections*, Research Report 170. Transport and Road Research Laboratory, Crowthorne, Berkshire, England. 1988.

24. Pierce, J.R., and P.J. Webb. *MOVA*: *The 20 Site Trial*, Research Report RR279. Transport and Road Research Laboratory, Crowthorne, Berkshire, England. 1990.

25. Kronborg, P., and F. Davidsson. "MOVA and LHOVRA: Traffic signal control for isolated intersections," *Traffic Engineering & Control* 34. 1993. pp. 195–200.

26. Bretherton, D., K. Wood, and N. Raha. "Traffic Monitoring and Congestion Management in the SCOOT UTC System," Transportation Research Board, 77th Annual Meeting. Transportation Research Board, National Research Council, Washington, DC. Janary 1998.

27. Wolshon, P.B., and W.C. Taylor. “Analysis of Intersection Delay Under Real-Time Adaptive Signal Control,” 77th Annual Meeting, Transportation Research Board. Transportation Research Board, National Research Council, Washington, DC. January 1998.

28. Abdel-Rahim, A., W.C. Taylor, and A. Bangia, “The SCATS effect,” *Traffic Technology International.* February/March 1998. pp. 57–60.

29. Nguyen, V.N. *SCATS Traffic Control and Air Pollution*, Technical Report RTA/TCC/NN. 9201a. Roads and Traffic Authority, Sydney, New South Wales, Australia. August 1992.

30. Pooran, F.J., C. Andrews, and N. Gartner. “Implementation of the OPAC Adaptive Control System in RT-TRACS.” 6th World Congress on Intelligent Transport Systems, Toronto, Canada. November 8–12, 1999.

31. Head, K.L., P.B. Mirchandani, and D. Sheppard. “A Hierarchical Framework for Real-Time Traffic Control,” *Transportation Research Record* 1360. Transportation Research Board, National Research Council, Washington, DC. 1992. pp. 82–88.

32. Head, K.L., P.B. Mirchandani, and S. Shelby. “The RHODES Prototype: A Description and Some Results,” 77th Annual Meeting, Transportation Research Board. Transportation Research Board, National Research Council, Washington, DC. January 1998.

33. Owen, L.E., C.M. Stallard, and D.M. Curtis. “An Evaluation of Real-Time Traffic-Adaptive Signal Control Prototypes,” 76th Annual Meeting, Transportation Research Board. Transportation Research Board, National Research Council, Washington, DC. January 1997.

34. Lindley, J.A. *Quantification of Urban Freeway Congestion and Analysis of Remedial Measures*, FHWA-RD-87-052. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. October 1986.

35. *Incident Management.* Cambridge Systematics. October 1990.

36. *Intelligent Vehicle Highway Systems: The State of the Art.* Prepared for Massachusetts Department of Highways, JHK and Associates, New York, NY. March 1993.

37. *Twin Cities Ramp Meter Evaluation: Executive Summary.* Cambridge Systematics, Inc., with SRF Consulting Group, Inc., and N.K. Friedrichs Consulting, Inc., Oakland, CA. February 1, 2001.

38. Endo, G., and V. Janoyan. “The Santa Monica Freeway Smart Corridor Project,” *Compendium of Technical Papers*, District 6, 44th Annual Meeting. Institute of Transportation Engineers, 1991. pp. 21–32.

39. Blumentritt, C.W., et al. *Guidelines for Selection of Ramp Control Systems*, National Cooperative Highway Research Program Report 232. Transportation Research Board, National Research Council, Washington, DC. May 1981.

40. McDermott, J.M., S.J. Kolenko, and R.J. Wojcik. *Chicago Area Expressway Surveillance and Control: Final Report*, Report Number FHWA-IL-ES-27. Illinois Department of Transportation, Oak Park, IL. March 1979.

41. Papageorgiou, M., H. Haj-Salem, and J.-M. Blosseville. "ALINEA: A local feedback control law for on-ramp metering," *Transportation Research Record* 1320. Transportation Research Board, National Research Council, Washington, DC. 1991. pp. 58–64.

42. Papageorgiou, M., H. Haj-Salem, and F. Middelham. "ALINEA local ramp metering: Summary of field results," *Transportation Research Record* 1603. Transportation Research Board, National Research Council, Washington, DC. 1997. pp. 90–98.

43. Mahmassani, H.S., T.Y. Hu, S. Peeta, and A. Ziliaskopoulos. *Development and Testing of Dynamic Traffic Assignment and Simulation Procedures for ATIS/ATMS Applications*, Technical Report DTFH61-90-C-00074-FG. University of Texas, Center for Transportation Research, Austin, TX. June 1994.

44. Taale, H., J. Slager, and J. Rosloot. "The Assessment of Ramp Metering Based on Fuzzy Logic," *Proceedings, 3rd Annual World Congress on Intelligent Transport Systems*. Orlando, FL. 1996.

45. Jacobson, L., K. Henry, and O. Mehyar. "Real-Time Metering Algorithm for Centralized Control," *Transportation Research Record* 1232. Transportation Research Board, National Research Council, Washington, DC. 1988.

46. O'Brien, A. "New Ramp Metering Algorithm Improves Systemwide Travel Time." *TR News*, vol. 209, July–August 2000. pp. 38–39.

47. Taylor, C., D. Meldrum, and L. Jacobson. "Fuzzy Ramp Metering: Design Overview and Simulation Results," *Transportation Research Record* 1634. Transportation Research Board, National Research Council, Washington, DC. 1998.

48. Taylor, C., D. Meldrum, and L. Jacobson. "Results of the On-line Implementation and Testing of a Fuzzy Ramp Metering Algorithm," 79th Annual Meeting, Transportation Research Board. Transportation Research Board, National Research Council, Washington, DC. January 2000.

49. Klein, L.A. *Sensor Technologies and Data Requirements for ITS* Artech House, Norwood, MA. 2001.

50. Gettman, D.M., K.L. Head, P.B. Mirchandani. "A multi-objective integrated large-scale optimized ramp metering control system (MILOS) for freeway traffic management," 78th Annual Meeting, Transportation Research Board. Transportation Research Board, National Research Council, Washington, DC. 1999.

51. NET. Presentation to Transportation Research Board Committee on Freeway Operations, Midyear Meeting. Marina del Rey, CA. June 1999.

52. Nelson, J.D., and M.G.H. Bell. "The Use of Origin-Destination Information to Support Ramp Metering Strategies," *Proceedings, PTRC Summer Annual Meeting*, vol. P-350. 1991. pp. 147–158.

53. Chen, L., and A.D. May. "Freeway Ramp Control Using Fuzzy Set Theory for Inexact Reasoning," *Transportation Research-A*, 24A(1). 1990. pp. 15–25.

54. "Design of a Highway Speed Indicator System," ERL Report no. 1480. Electronics Research Laboratory, Montana State University. November 1980.

55. Lee, C.E., D.W. Borchardt, and Q. Fei. *Truck-Monitoring and Warning Systems for Freeway to Freeway Connections*, Report Number 2915-1. University of Texas, Texas Transportation Institute and Center for Transportation Research, Austin, TX. October 1, 1999.

56. Shepard, F.D. *Synthesis of Highway Practice 228: Reduced Visibility Due to Fog on the Highway*. Transportation Research Board, National Academy Press, Washington, DC. 1996.

57. MacCarley, C.A., B.M. Hemme, and L.A. Klein. *Advanced Image Sensing Methods for Traffic Surveillance and Detection*, Final Report, document no. TEL5859F98 (rev. 1). California Polytechnic University. June 30, 1998.

58. *ITS Data Archiving: Five-Year Program Description*. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. March 2000.

59. *Intelligent Transportation Systems Impact Assessment Framework: Final Report*. U.S. Department of Transportation, Volpe National Transportation Systems Center, September 1995.

60. Lindley, J.A., and D.G. Capelle. "Freeway Surveillance and Control," in *Transportation Engineering Handbook*, 4th Edition, Institute of Transportation Engineers. Prentice Hall. 1992.

61. Ricci, R.C., and W.E. Gazda. "Transportation Modeling: An Essential Tool for Evaluation of IVHS," *IVHS Review*, Winter/Spring 1994. pp. 43–52.

62. Klein, L.A., and M.R. Kelley. *Detection Technology for IVHS*: *Final Report*, Report No. FHWA-RD-95-100. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 1995.

63. *What Have We Learned About Intelligent Transportation Systems*? U.S. Department of Transportation, Federal Highway Administration, Washington, DC. December 2000.

64. Jellison, L., A. Lee, and R. Macaluso. *Summary Report: California Field Operational Test Program Advanced Traffic Management Systems*. Caltrans New Technology and Research Program, Sacramento, CA. June 2000.

65. Parviainen, J.A. *An Overview of Available and Developing Highway Vehicle Electronic Technologies*. Ministry of Transportation, Transportation Technology and Energy Branch, Ontario, Canada. 1990.

66. Nør, M., and K. Strand. "A public priority," *Traffic Technology International*, August–September 1998. pp. 81–83.

67. *Model IVS-2000 Installation and Operation Manual*, Rev. 1.53. Intersection Development Corp, Downey, CA. September 1997.

68. *Manual on Uniform Traffic Control Devices.* U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 2003. http://mutcd.fhwa.dot.gov/pdfs/2003/pdf-index.htm (Accessed Nov. 26, 2004.)

69. Gates, P.M., and R.K. Duley. "Current Automatic Traffic Counting Practice." Sarasota Engineering Co., Ltd., undated.

70. Johnson, V., D. Koracin, W. Gallus, D.R. Fitzjarrald, P. Knight, and J. Horel. *Collaborative Research on Road Weather Observations and Predictions by Universities, State DOTs and National Weather Service Forecast Offices*, Report FHWA-HRT-04-109. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. September 2004.

71. "Where the Weather Meets the Road: A Research Agenda for Improving Road Weather Services," National Academies Press, www.nap.edu/catalog/10893.html. Committee on Weather Research for Surface Transportation, The Roadway Environment, National Research Council. 2004.

72. Pisano, P.A., R.M. Alfelor, J.S. Pol, L.C. Goodwin, and A.D. Stern. "Clarus—The Nationwide Surface Transportation Weather Observing and Forecasting System," 85th American Meteorological Society Meeting containing 21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, paper 6.3. San Diego, CA. January 2005.

73. Najm, W.G. "Comparison of Alternative Crash Avoidance Sensor Technologies," *Proceedings, SPIE*, vol. 2344. Bellingham, WA. 1994. pp. 62–72.

74. Kuchar, J.K. "Collision Alerting System Evaluation Methodology for Ground Vehicles," *Proceedings, SPIE*, vol. 2902, Bellingham, WA, November 1996, pp. 272–283.

75. Heide, P. "Commercial Microwave Sensor Technology: An Emerging Business." *Microwave Journal.* 1999. pp. 348–352.

76. Lowbridge, P., P. Brigginshaw, and B. Kumar. "A Low Cost mm-Wave Cruise Control System for Automotive Applications," *Microwave Journal.* October 1993. pp. 24–36.

77. Eriksson, L.H. "High Performance Automotive Radar." *Microwave Journal*, October 1996. pp. 24–38.

78. Ganci, P., S. Potts, and F. Okurowski. "Forward Looking Automotive Radar Sensor," *Collision Avoidance and Automated Traffic Management Sensors*, *Proceedings, SPIE*, vol. 2592, Bellingham, WA, October 1995, pp. 60–65.

79. *Intelligent Transportation Systems Handbook 1999–2000*, Highway Industry Developmental Organization and Ministry of Construction, Japan.

80. Chan, C.Y., and B.B. Litkouhi. "Technological Challenges in the Development of AHS," *J. of Intelligent Transportation Systems.* 1997.

81. Nakamura, M., A. Arai, M.P. Bendett, R.W. Goudy, and Y. Suzuki. "Vehicle Lateral Control System Using Laser Radar," *Proceedings, SPIE*, vol. 2344. Bellingham, WA. November 1994. pp. 267–277.

82. "Cheaper Cameras," *Photonics Spectra.* March 1999, p. 34.

83. Schuler, C.A., F.M. Tapos, M.M. Alayleh, and W.D. Bachalo. "Aerometric's Laser-Based Lane-Tracker Sensor: Engineering and On-the-Road Evaluation of Advanced Prototypes," *Proceedings, SPIE*, vol. 2902. Bellingham, WA. November 1996, pp. 98–105.

84. Geary, J.M. "Passive Optical Lane Position Monitor," *Proceedings, SPIE*, vol. 2902. Bellingham, WA. November 1996, pp. 86–97.

85. Coughlin, Sean, Personal Communication, Caltrans District 4, August 24, 1999.

TABLE OF CONTENTS: CHAPTER 4.

LIST OF FIGURES: CHAPTER 4.

LIST OF TABLES: CHAPTER 4.

CHAPTER 4. IN-ROADWAY SENSOR DESIGN

This chapter discusses the design of vehicle detection systems that use inductive-loop detectors, magnetometers, and magnetic detectors for surface street and freeway traffic management, with primary emphasis on inductive loops. The guidelines provided should serve traffic engineers in developing plans and specifications for vehicle detection at local intersections, detectorization for traffic signal systems, and freeway surveillance and control. Above-roadway sensors, such as those described in chapter 2, may also be an option for many of the applications discussed in this chapter.

When an in-roadway sensor is required for vehicle detection in a particular application, the first step in the design process is to determine what type of in-roadway sensor will supply the needed data or provide the required functionality. Some agencies have predetermined policies or standard plans that assist in the sensor selection process. To be effective, the standard plans should be updated to reflect current state of the practice in sensor and data transmission technology. Where standard plans are not appropriate or are not available, the following criteria may be applied as guidelines for in-roadway sensor selection.

SELECTION CRITERIA

Sensor selection for a specific traffic management application depends on the data parameters, data accuracies, spatial resolution, detection area, appropriate data transmission media, location-specific installation requirements, initial sensor cost, and the acceptability of the maintenance burden it will impose. These criteria, separately and in combination, should be assessed as part of the sensor selection process.

SELECTION BASED ON SENSOR TECHNOLOGY

Sensor technology and operating theory described in chapter 2 indicate that the principal in-roadway sensors (inductive-loop, presence-detecting magnetometers, and passage-detecting magnetometers (also referred to as magnetic detectors)) are suitable for some applications, but unsuitable for others. For example, magnetometers cannot be used with NEMA controllers where delayed-call capability is required, since this feature is not available in magnetometers. In Type 170 controllers, the controller performs the timing operation; therefore, magnetometers can be used with 170 and the newer 2070 controllers. Search coil magnetometers must be excluded for operations requiring presence detection because they are only capable of detecting a moving vehicle (passage detection).

Inductive-loop detectors are not always appropriate for some traffic signalization applications. For example, long loops are not suitable for detection of oversaturated flow or long queues. Loops can be utilized for freeway surveillance and control if the area of the induced magnetic field is tailored to provide vehicle detection in the lane of interest and the time required for output, pick-up, and drop-out are predictable.

SELECTION BASED ON APPLICATION

Applying lessons learned from the extensive application of in-roadway sensors further narrows sensor selection. In theory, both the inductive-loop detector and one or more presence-detecting magnetometers are suitable for large area detection on an approach to a signalized intersection. The loop detector, however, is usually less expensive. Conversely, for an approach where it is not important to screen out false calls for the green (i.e., right turns on red) and only rudimentary traffic detection is needed, any of the three in-roadway sensor types may be installed. In this particular application, the passage-detecting magnetometer might be favored because of its ruggedness and low cost/useful life ratio.

SELECTION BASED ON EASE OF INSTALLATION

In the 1960s-1970s, it was common for large cities such as New York, NY, or Atlanta, GA, to select microwave radar or ultrasonic sensors because they could be installed on a suitable pole already in place and not disrupt traffic or break the pavement. A crew could install one of these sensors in about 45 minutes. Recent advances in over-roadway sensor design have encouraged a new interest in these sensors and other over-roadway sensor technologies such as video image processors and laser radars.

Today, agencies often look favorably on eliminating a saw cut or replacing it with a drilled hole. The pervasiveness of deteriorating pavements has produced more interest in installing preformed loops, microloops, or pavement slabs with sensors already in place. Another approach to eliminating saw cuts is to install preformed loops or prewound loops in conjunction with repaving. Chapter 5 describes some of the approaches that simplify installation of in-roadway sensors.

SELECTION BASED ON EASE OF MAINTENANCE

Most traffic engineering agencies are aware of the cost differences in maintaining sensors based on different operating technologies. For example, the passage-detecting magnetometer detector, with its limited application, has managed to retain some popularity largely because of its ruggedness and long life with minimum maintenance. The improved reliability and functionality in inductive-loop detector electronics units have shifted the main sources of loop detector failure to the wire loop in the pavement, its connection to the lead-in cable, and loop installation issues. The requirement for more rugged loop detector installations has highlighted the need for reduced maintenance costs in terms of frequency of failures and resulting maintenance calls. Chapter 6 discusses these issues further.

DESIGN CONSIDERATIONS

Sensor selection for vehicle detection at intersections is a function of the types of timing intervals generated by the controller and the corresponding data needed to compute the intervals. Therefore, the timing interval types should be selected early in the traffic signal control design process. Similar considerations apply to the design of freeway signal control strategies. The following discussion defines various timing parameters, effect of short-loop and long-loop configurations, and detection alternatives for low-speed and high-speed approaches.

TIMING PARAMETERS

An actuated phase normally has three timing parameters in addition to the yellow change and all-red clearance intervals. These are the minimum green (also known as initial interval), the passage time (also called the vehicle interval, extension interval, or unit extension), and the maximum interval.[1] Relationships among the intervals are shown in Figure 4-1. These intervals are timed depending on the type and configuration of the sensor installation found at the intersection.

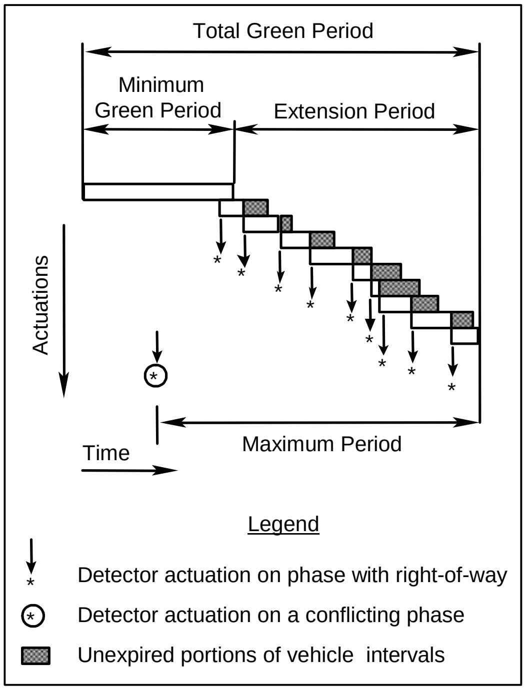


Figure 4-1. Actuated controller green phase intervals.

Minimum Green Interval

The majority of early vehicle sensors were *point detectors* consisting of treadles or pressure plates in the roadway. Today's 6- x 6-ft (1.8- x 1.8-m) inductive-loop detector also performs as a point detector. With point detection, the minimum green interval is set to allow vehicles stopped between the detection point and the stopline to start and move into the intersection. Table 4-1 shows the minimum green interval for various distances from the stopline. The data are based on an average vehicle headway distance of 20 ft (6 m) and the average times for vehicles from a queue to enter the intersection.

Table 4-1. Minimum green interval.

Distance between stopline and sensor feet	meters	Minimum green time seconds
0–40	0–12.2	8
41–60	12.5–18.3	10
61–80	18.6–24.4	12
81–100	24.7–30.5	14
101–120	30.8–36.6	16

A different timing approach is used to establish minimum green intervals for presence detection. When long loops (or a series of short loops) that terminate at the stopline are used, the initial interval is set as small as zero seconds to achieve “snappy” response for left-turn phases or possibly longer for through movements where drivers may expect a longer green. If the loop ends some distance from the stopline, this distance is used to calculate the interval time in the same way as for point detection.

Passage Time Interval

The passage time interval is calculated such that a vehicle can travel from the sensor to the intersection. This is particularly important where point detection is used and the sensors are located some distance from the stopline. The passage time also defines the maximum apparent time gap between vehicle actuations that can occur without losing the green indication to a call waiting. As long as the interval between vehicle actuations is shorter than the passage time, the green will be retained on that phase (subject to the maximum green interval described in the following section).

The apparent time gap for a single stream of vehicles passing over an inductive-loop detector is somewhat shorter than the actual time gap. A vehicle activates the loop upon entry over the loop, and the loop does not deactivate until the vehicle leaves the loop. The true time gap is reduced by the time it takes the vehicle to traverse the loop.

Another factor in determining an appropriate vehicle interval is the number of approach lanes containing sensors. Inductive-loop detectors for the same phase and function installed in adjacent lanes are often connected to the electronics unit by a single lead-in cable. This may introduce errors into the number of vehicles reported.

NEMA controllers identified as “advanced design,” “exceeds,” or “beyond” have a predefined minimum green time. If no further actuations occur, the minimum green is the total green. If there are further actuations, the passage time interval extends the green until a gap exceeds the passage time or the maximum green time is reached. The relationship of these times was illustrated in Figure 4-1.

When long loops are used in approaches, especially in left turn bays, the minimum green and the passage time intervals are generally set to zero or near zero. The long loop operates in the presence mode and the controller continuously extends the green as long as the loop is occupied. The critical gap is not a preset value, but a space gap equal to the length of the loop. Thus, the gap length is equivalent to no following vehicle entering the loop prior to the departure of the previous vehicle.

When a series of short loops is utilized, it acts as a single long loop, provided the space between loops is less than a vehicle length. If the spacing is greater than a vehicle length, a short vehicle interval can be used to provide the same effect as a single long loop.

Maximum Green Interval

The maximum green interval limits the time a phase can hold the green. The maximum green begins timing at the first call received from an opposing (or conflicting) phase. Ordinarily, maximum intervals for through movements are set between 30 and 60 seconds. When the signal is properly timed with appropriately short passage times (i.e., vehicle intervals), the maximum interval will not consistently time out unless the intersection is badly overloaded. Some actuated controllers are capable of providing two maximum intervals per phase. This allows longer maximums to be programmed during peak periods (or shorter maximums for selected phases) when very heavy traffic flows are expected on the major street.

Volume-Density Mode

Volume-density phases have more timing parameters than a standard actuated phase. For this type of operation, sensors are generally placed further back from the intersection, particularly on high-speed approaches. The minimum green interval can be increased to provide longer initial green times for those instances when the minimum green is not adequate to serve the actual traffic present. The length of the initial green interval is governed by three factors—*minimum green, seconds per actuation, and maximum initial*—as described by the timing procedure illustrated in Figure 4-2. The variable initial interval is described by NEMA Standards in the following paragraph:[2]

> In addition to MINIMUM GREEN, PASSAGE TIME, and MAXIMUM GREEN timing functions, phases provided with VOLUME DENSITY operation shall include VARIABLE INITIAL timings and GAP REDUCTION timings. The effect on the INITIAL timing shall be to increase the timing in a manner dependent upon the number of vehicle actuations stored on this phase while its signal is displaying YELLOW or RED. The effect on the extensible portion shall be to reduce the allowable gap between successive vehicle actuations by decreasing the extension time in a manner dependent upon the time waiting of vehicles on an opposing RED phase.

In volume-density phases, the extended green time (passage time) created by each new actuation after the initial green time has elapsed is normally based on the time required to travel from the sensor to the stopline. Because this distance can be relatively long, the passage time can be more than the desired allowable gap. The NEMA gap reduction procedure addresses this situation. It defines four time settings—*time before reduction, passage time, minimum gap, and time to reduce,* as illustrated in Figure 4-3. The time before reduction begins when there is a call on a conflicting phase. Once the time before reduction has expired, the allowable gap reduces linearly until the minimum gap is reached at the end of the time-to-reduce interval. The maximum green extension and yellow change and all-red clearance intervals are predetermined (precalculated) and programmed into the controller.

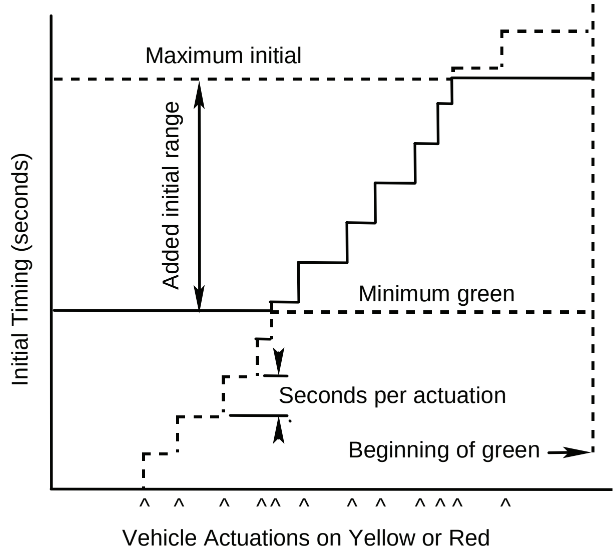


Figure 4-2. Variable initial NEMA timing for green signal phase.

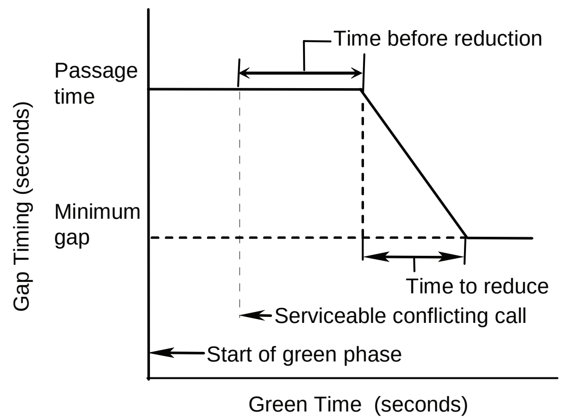


Figure 4-3. NEMA gap reduction procedure.

LOW-SPEED APPROACHES

Approaches with speeds of less than 35 mi/h (55 km/h) are considered low-speed approaches. The sensor design for a given approach depends on whether the controller phase has been set for *locking* or *nonlocking* detection memory. This is also referred to as *memory ON* or *memory OFF*. The locking feature means that a vehicle call for the green is remembered or held by the controller until the call has been satisfied by the display of the green indication, even if the calling vehicle has left the detection area (e.g., right turn on red). In the nonlocking mode, the controller drops a waiting call as soon as the vehicle leaves the detection area.

Locking Memory with Point Detection

Locking detection memory is associated with the use of small-area point sensors such as a 6- x 6-ft (1.8- x 1.8-m) inductive loop and is frequently referred to as conventional control. The minimum green interval (or initial interval) is preset to provide sufficient time to clear a standing queue between the sensor and the stopline. The passage time or unit extension sets a common value for both the allowable gap to hold the green and the travel time from sensor to stopline.

Since the allowable gap is usually 3 or 4 seconds, the sensor might be ideally located 3 or 4 seconds of travel time upstream from the intersection. This sensor position would appear to be the most efficient for accurately timing the end of green after passage of the last vehicle of a queue. However, a long minimum green (assured green) is created at approaches with speeds greater than 25 to 30 mi/h (40 to 50 km/h) because of the longer sensor setback. Therefore, the principle is amended to locate sensors 3 to 4 seconds of travel time from, but not more than 170 ft (52 m), from the stopline. Some agencies limit this distance to 120 ft (37 m). Table 4-2 displays the application of this principle to determining sensor location and associated timing parameters as a function of vehicle approach speed.

Table 4-2. Inductive-loop detector location and timing parameters.

Approach speed		Detector setback from stopline to leading edge of inductive-loop detector		Minimum green time	Passage time
mi/h	km/h	feet	meters	seconds	seconds
15	24	40	12	9	3.0
20	32	60	18	11	3.0
25	40	80	24	12	3.0
30	48	100	30	13	3.5
35	56	135	41	14	3.5
40	64	170	52	16	3.5
45+	72+	Volume-density or multiple inductive-loop detectors recommended.			

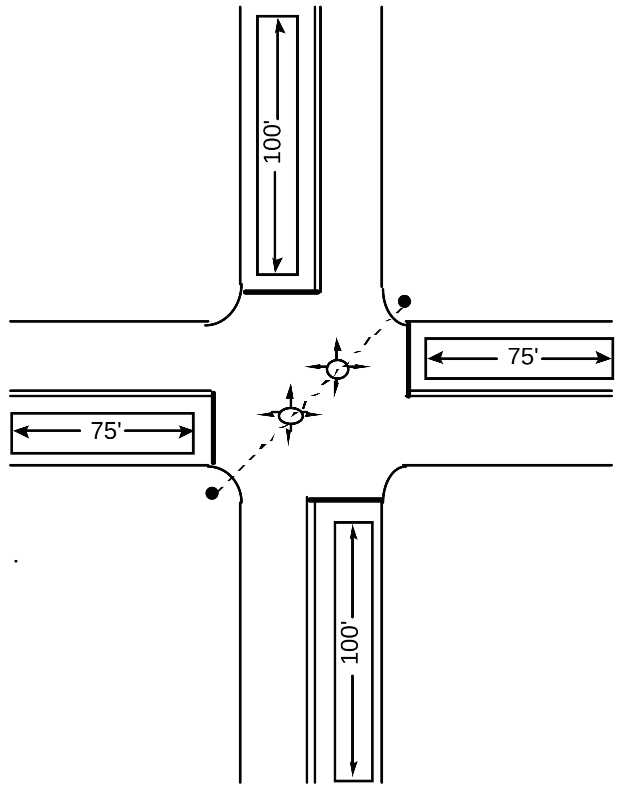
Note: Volume-density could be considered at speeds of 35 mi/h (56 km/h) or above.

The advantage of this single sensor approach is that the cost of installation is minimized. However, this type of control does not screen out false calls for green as occurs with right turns on red.

Nonlocking Memory with Larger-Area Presence Detection

Nonlocking detection memory is used with larger-area sensors such as a 6- x 50-ft (1.8- x 15.2-m) loop. Often called *loop occupancy control*, the technique requires a nominal loop configuration as shown in Figure 4-4. The vehicle presence information provided in the area near the intersection allows the elimination of many of the false calls for green and thus avoids unnecessary green signals in approaches with no waiting vehicles. The disadvantages of this traffic control method are the higher initial cost of sensor installation and the higher replacement and maintenance costs associated with larger inductive-loop detectors.

Nonlocking detection memory is particularly appropriate in left-turn lanes with separate signal control for the left turns. The green arrow is terminated as soon as the turning vehicle clears the loop. In addition, a call placed during the yellow change interval by a vehicle that clears on the yellow does not bring back the green to an empty approach. Another potential advantage occurs when the left turn is permitted. In this case, the left turn is allowed to filter across oncoming traffic on the circular green shown to the through movement. Accordingly, left turns may be serviced during the through phase and, therefore, do not require the display of a left arrow.



1 ft = 0.3 m

Figure 4-4. Intersection configured for loop occupancy control.

The left turn bay may use an electronics unit operating in a delayed-call mode, which is designed to send an output to the controller only if a vehicle is continuously detected beyond a preset period (e.g., 5 seconds). This allows the electronics unit and controller to ignore vehicles that are in transit over the loop if oncoming traffic is light enough to allow a left turn without the protective left-turn arrow. If oncoming traffic is so heavy that left turning vehicles queue up over the loop, then the green arrow is called. This type of controller operation is often used with lagging left-turn phasing.

A similar operation occurs on a single lane approach from the cross street where a right-turning vehicle approaches on the red. Again, the use of a delayed call electronics unit will avoid calling the green to the side street if the right turn on red can be made during the delay time set on the electronics unit.

Loop-occupancy control is also utilized for through lane control on low-speed approaches. The technique minimizes delay by allowing short passage times (unit extensions) in the range of 0 to 1.5 seconds. The length of the detection zone obviously depends on the approach speed and the controller unit time settings. Figure 4-5 gives the length of the long-loop presence detector for various passage time settings on the controller for approach speeds less than 30 mi/h (50 km/h). The figure is based on a desired allowable gap of 3 seconds and an average vehicle length of 18 ft (5.5 m). The formula for loop length is

$$L = 1.47\,S\,(3 - PT) - 18 \text{ in English units} \tag{4-1a}$$

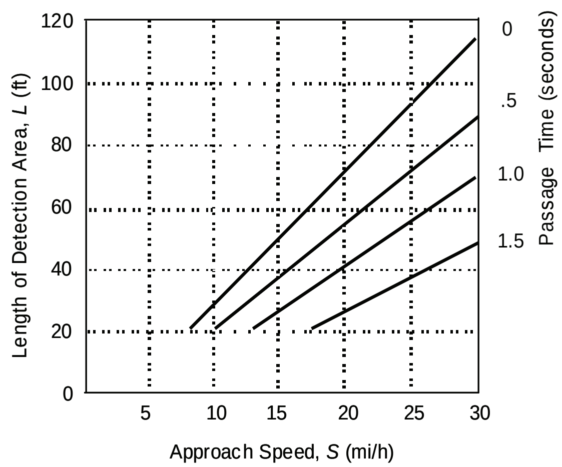
$$L = 0.277\,S\,(3 - PT) - 5.5 \text{ in metric units} \tag{4-1b}$$

where

L = length of detection area, ft (m)

S = approach speed, mi/h (km/h)

PT = passage time (unit extension),seconds.



1 mi/h = 1.6 km/h
1 ft = 0.3 m

Figure 4-5. Inductive-loop detector length for loop occupancy control.

High-Speed Approaches

Approaches with speeds in excess of 35 mi/h (56 km/h) are considered high-speed approaches. Several problems associated with high-speed approaches require special consideration in setting signal timing intervals. For example, it may be difficult for the driver to decide whether to stop or proceed when approaching a yellow change indication. An abrupt stop may result in a rear-end collision, while the decision to proceed through the intersection may cause a right-angle accident or a traffic violation.

The portion of the roadway in advance of the intersection in which the driver is indecisive (as to stopping or proceeding into and through the intersection at the onset of the yellow change interval) is called the *dilemma zone*. Some researchers have defined the dilemma zone as that area of the approach between a point where 90 percent of the drivers will stop on yellow and a point where 90 percent of the drivers will go (i.e., 10 percent will stop).[3,4] Table 4-3 shows these boundaries for various speeds.

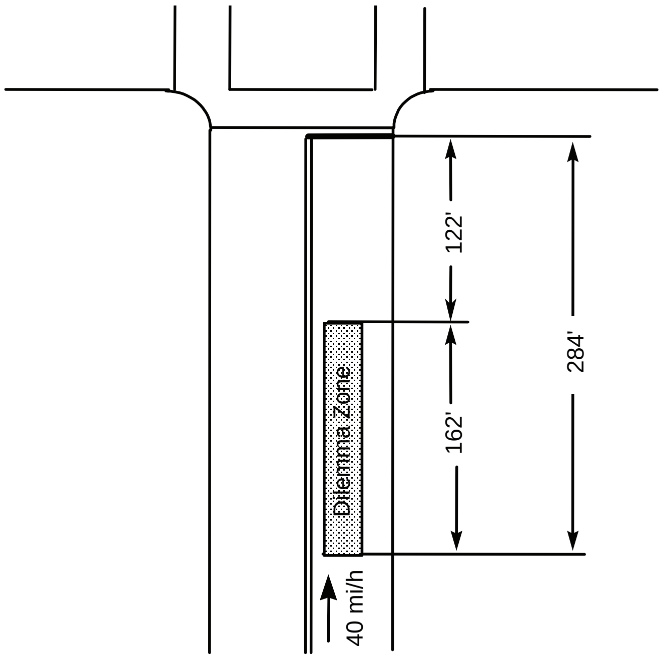
Table 4-3. Dilemma zone boundaries.

Approach speed		Distance from intersection for 90% probability of stopping		Distance from intersection for 10% probability of stopping	
mi/h	km/h	feet	meters	feet	meters
35	56	254	77	102	31
40	64	284	87	122	37
45	72	327	100	152	46
50	80	353	108	172	52
55	88	386	118	234	71

Figure 4-6 illustrates the dilemma zone for a vehicle approaching an intersection at 40 mi/h (64 km/h). To minimize the untimely display of yellow and thus the creation of a dilemma zone problem, a number of techniques have been devised for controllers with locking and nonlocking detection memory, basic actuated and volume-density controller circuitry, and various inductive-loop detector configurations.

The most straightforward conventional design for a high-speed approach utilizes a controller with a volume-density mode. This type of actuated operation can count waiting vehicles beyond the first because of the *added initial* feature. It also has a timing adjustment to reduce the allowable gap based on the time vehicles have waited on the red on a conflicting phase.

More efficient operation can be achieved with the volume-density mode than with fully actuated control because of the added initial and timing adjustment features and because detection is farther back on the approach—400 ft (120 m) is typical. A calling sensor near the stopline that operates only when the phase is red or yellow often supplements high-speed approach detection. This calling sensor is disabled when the signal turns green so that it cannot extend the green time inappropriately.



1 ft = 0.3 m

Figure 4-6. Dilemma zone for vehicle approaching an intersection at 40 mi/h (64 km/h).

Fully-actuated controllers utilizing an extended-call sensor just upstream of the dilemma zone have been in service since at least 1982. Designs for high-speed approaches using nonlocking detection memory include a long loop at the stopline as well as one or more small loops upstream. The long loop improves the controller's knowledge of traffic at the stopline, but tends to increase the allowable gap.

Inductive-loop designs for both normal fully actuated and volume-density controllers are available. Additional information concerning these solutions to the dilemma zone problem is presented later in this chapter.

Rural High-Speed Roadways

For isolated high-speed rural intersections, a "PREPARE TO STOP" extinguishable message sign is frequently deployed when the signal site undergoes periods of poor visibility caused by dense ground fog, orientation of the sun, or geometry that prevents signal visibility far enough in advance to ensure safety. These situations require vehicle sensors further in advance of the intersection than normal.

Using the *last car passage* feature of some density controllers, the gap in the traffic flow can be identified to allow the last car in the platoon to pass through the signal and presumably give the next vehicle sufficient time to stop. The message sign, such as the one depicted in Figure 4-7, would flash PREPARE TO STOP at the appropriate time, but would be blank or unreadable at other times.

When the controller is informed of a gap in the traffic, it does not change the signal until a preset time has elapsed to allow the last car to clear the intersection. The PREPARE TO STOP is illuminated when the gap is selected, so that the next vehicle following the platoon will see the sign. Thus, the driver will know he will be required to stop even though the signal ahead is still green.

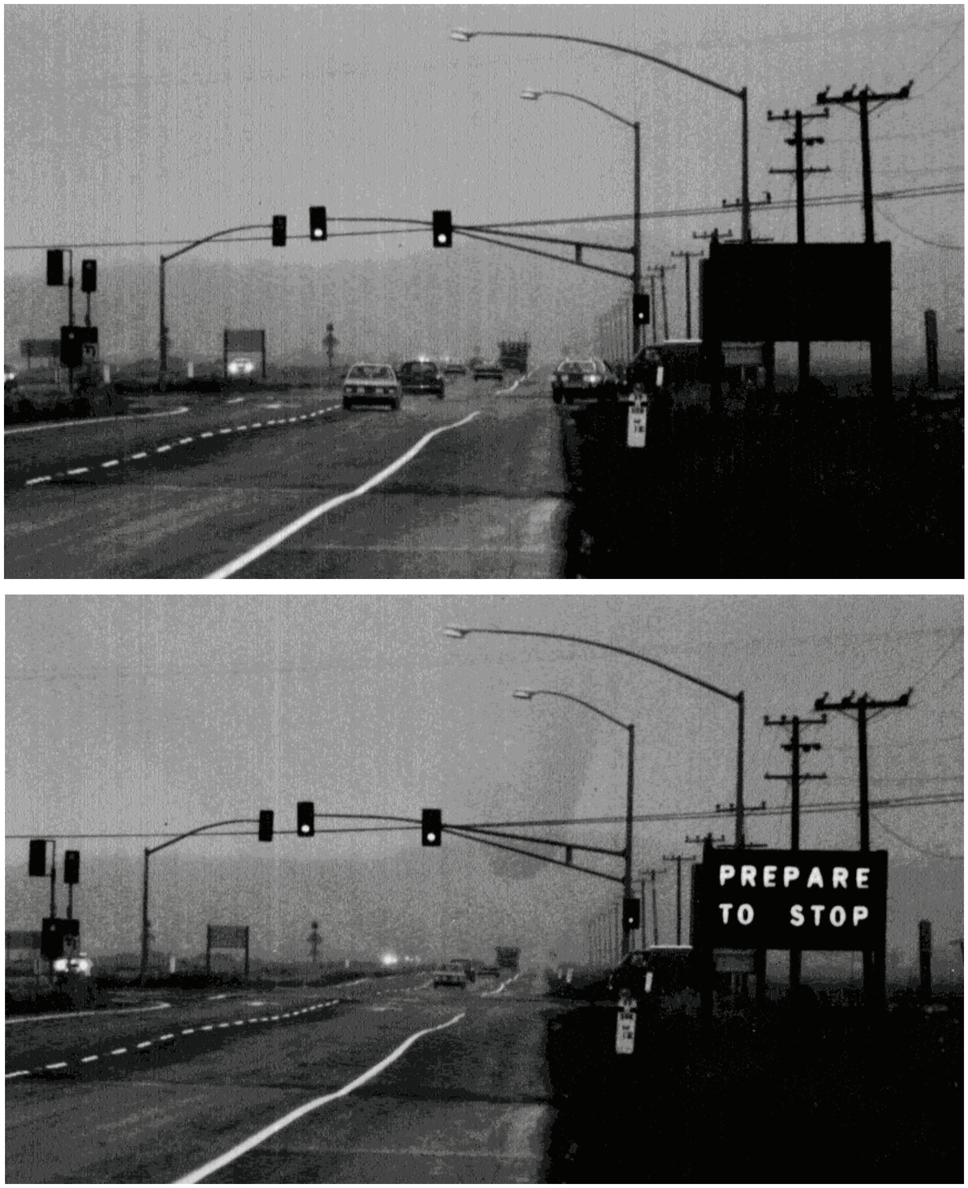


Figure 4-7. PREPARE TO STOP sign on an arterial in nonilluminated and illuminated states.

A typical inductive-loop detector and sign placement layout for a PREPARE TO STOP installation is shown in Figure 4-8. A display used by some jurisdictions consists of flashing beacons together with a diamond or rectangular sign with the message PREPARE TO STOP WHEN FLASHING.

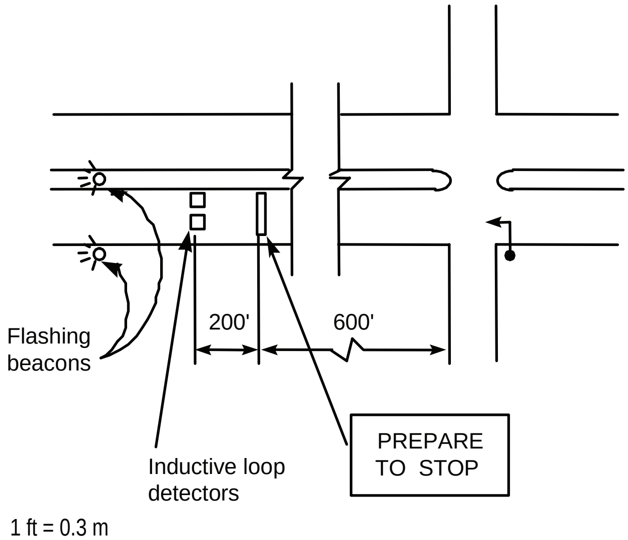


Figure 4-8. PREPARE TO STOP inductive-loop detector system.

REST-IN-RED SPEED CONTROL

The commuting driver frequently uses residential roadways as a time-saving route for reaching destinations, particularly when the residential street is parallel to a congested arterial. Residents generally perceive the added commuter traffic as a threat to the safety of children, pets, and their ability to safely exit their driveways, particularly where speeding occurs.

To control traffic speeds in residential areas, traffic engineers have used a variety of traffic control devices such as stop signs, warning signs, speed bumps, and coordinated traffic signals with vigorous enforcement of posted speed limits. While these measures are often successful, there are drawbacks associated with their use.

For example, the use of unwarranted stop signs to control vehicular speeds imposes delay penalties on all drivers, and it only affects speeds within 200 ft (60 m) of the stop sign. Unwarranted stop signs may also encourage drivers to ignore the stop sign, which is even more dangerous.

One approach for slowing speeding vehicles is to install two speed-measuring inductive-loop detectors (or other sensor types) approximately 180 ft (55 m) in advance of the intersection. The advance loops measure the speed (through measurement of the elapsed time to travel between the loops) of an approaching vehicle and register a call on the controller only if the vehicle is traveling at or below the speed limit. Assuming the signal is resting in four-way red and the vehicle is not speeding, a green is displayed and the driver may proceed through the intersection without being delayed. If, on the other hand, the approaching vehicle is exceeding the speed limit, no call is placed and the driver must slow down until he reaches a loop near the intersection. A call will then be placed and the green interval activated.

Consequently, the timing cycle to initiate a call begins when a vehicle crosses the first advance loop. If the vehicle speed is low enough, the predetermined interval will time out before the vehicle reaches the second loop. Then the call request will be passed to the controller. When the vehicle reaches the second loop, the timing device is reset, and any call being held at the timing device is cancelled. Thus, a vehicle exceeding the speed limit is never detected by the advance loop, and each succeeding vehicle is timed independently. This method is simple, economical, adjustable, and not dependent on vehicle size or length.

The spacing between the initiating and resetting advance loops is approximately 1 second of travel distance at the speed limit. The distance from the advanced loop to the first intersection loop is predetermined from the lowest travel speed to be accommodated (normally 20 mi/h (32 km/h)) and the maximum desirable passage time interval (4 seconds). A comfortable reaction time and the stopping distance at the design speed determine the minimum distance from a stopline to the advance detector loop.

INDUCTIVE-LOOP DETECTOR DESIGN ALTERNATIVES

Small-area (for point or passage detection) and larger-area (for area or presence detection) inductive-loop detector design is described in this section. Passage-detecting search coil magnetometers (magnetic detectors) can only be used for point detection because they are sensitive to vehicles in small areas and require vehicle motion (passage) for activation. Presence-detecting two-axis fluxgate magnetometers are also point detectors, but can be used as area detectors by using multiple sensors to cover a larger area.

Typical design configurations for sensor locations in through lanes and in left-turn lanes are also presented. Both simple and complex designs are described along with the type of controller operation, if appropriate. Treatments to alleviate the dilemma zone problem are also discussed.

SMALL-AREA DETECTION

Small-area detection is commonly implemented with a single short inductive loop. Although the literature defines short loops as being up to 20 ft (6.1 m) in length, by far the most common short-loop application is the 6- x 6-ft (1.8- x 1.8-m) loop in a 12-ft (3.6-m) lane. For narrower lanes, 5- x 5-ft (1.5- x 1.5-m) loops should be used to avoid adjacent lane pickup (splashover). Smaller loops are not recommended in areas where high-bed vehicles must be continuously detected.

The short loop is intended to detect a vehicle upstream of the stopline. When a vehicle passes over the loop, a call is output by the electronics unit to the controller. Timing of the green interval is commonly based on preset controller settings, not by the length of time the detection area is occupied by vehicles approaching the intersection. In most cases, the controller operates with locking detection memory circuits to insure calling the appropriate phase.

Short-loop detectors may be used in a variety of ways and may be located at varying distances from the stopline, depending on the operational requirements. A typical application may consist of one or more short loops near the stopline on the actuated approach of a low-speed intersection. Another typical application is to space a number of these loops well back of the stopline to act as extension sensors for higher-speed approaches.

Loop shapes were the subject of a great amount of research during the l970s and 1980s. Subsequently, many loop configurations were designed to detect the various sizes and shapes of vehicles that travel on the Nation's roadways, from bicycles and motorcycles to high-bed trailer trucks, while avoiding detection of vehicles in adjacent lanes. Each design purports to have advantages over other designs. Examples of short loop shapes are illustrated in Figure 4-9. Some of these configurations are common, while others are found at either a site-specific location or when particular types of vehicles require detection.

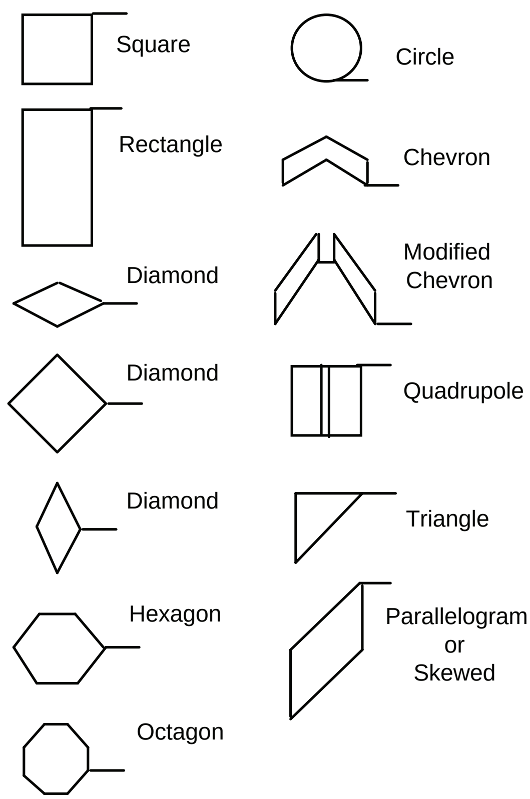


Figure 4-9. Small loop shapes.

A number of agencies and universities have conducted tests to determine an optimum loop shape.[5,6,7] These projects typically involved installing several different loop shapes and then testing and comparing the sensitivity of the loops in detecting several types of test vehicles. None of these projects tested all of the loop designs currently in use. In some cases, one loop design would test better when compared to one or two different loop designs. In most instances, the difference in sensitivity among loops was not significant, given the state of the art in electronics units. It is therefore difficult to cite one particular design as superior to all others. However, it is generally accepted that some loop designs are better suited than others for detecting small vehicles or high-bed trucks, as discussed in later sections.

Many States specify the acceptable loop shapes for use in their jurisdiction. An example is Caltrans's specified shapes shown in Figure 4-10. In this particular example, each unique shape is given a letter designation (e.g., type A is the conventional 6- x 6-ft (1.8- x 1.8-m) loop).

LARGER-AREA DETECTION

Larger-area detection normally contains a detection zone covering an area of at least 20 ft (6 m) or more in a traffic lane. It is primarily used for presence detection because the detection zone registers the presence of a vehicle as long as the zone is occupied. This concept originally used a single loop encompassing the entire detection zone. However, the long loop, as a single entity, is being supplanted by a sequence of short loops, which emulate the long loop. In this *Handbook*, the term *long loop* means either a single long loop or multiple short loops functioning as a single long loop.

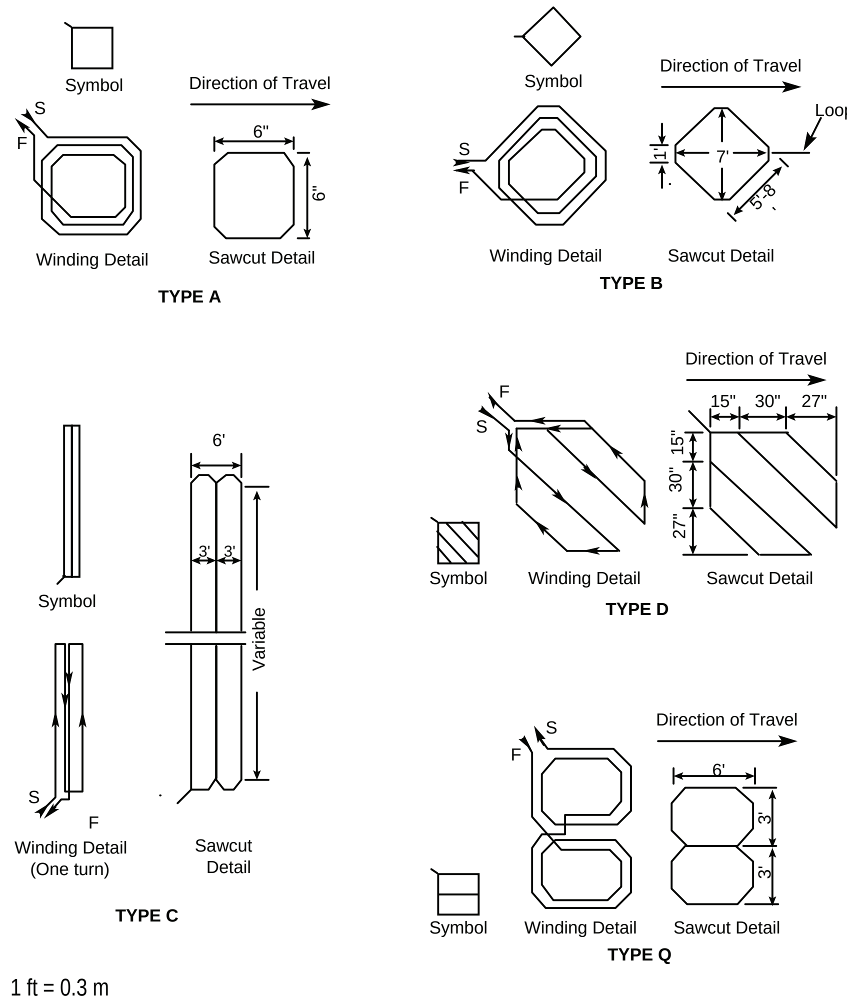


Figure 4-10. Caltrans-specified loop shapes.

Long Loops

Figure 4-11 illustrates the traditional long loop (i.e., a single loop 6 ft (1.8 m) wide by 20 to 80 ft (6 to 24 meters) long or longer) and other long-loop shapes. These long loops generally have only one or two turns of wire. If the rectangular, powerhead or trapezoidal loop needs to reliably detect all roadway vehicles, the sensitivity level must be set high which, in turn, causes detection of adjacent lane vehicles (splashover). The quadrupole loop is an appropriate design to eliminate this problem. However, due to its limited field height, it may have difficulty in continuously detecting high-bed vehicles. Quadrupoles are excellent wheel and axle sensors. The lengths associated with long loops increase the vulnerability to failure caused by pavement cracks and joint movement. In response to these problems, many agencies are installing sequential short loops.

The long loop normally provides input to the controller for operating in the loop occupancy mode. In this control mode, the minimum green interval (or initial interval in older controllers) is set to zero or near zero, and the passage time or vehicle interval is set to a small value, as was shown in Figure 4-5. When the green interval appears for the subject phase, it remains green as long as the loop is occupied (subject, of course, to the maximum green). As soon as the inductive-loop detector is cleared, the passage time is measured and, if no further actuation occurs during the passage time, the yellow change interval appears.

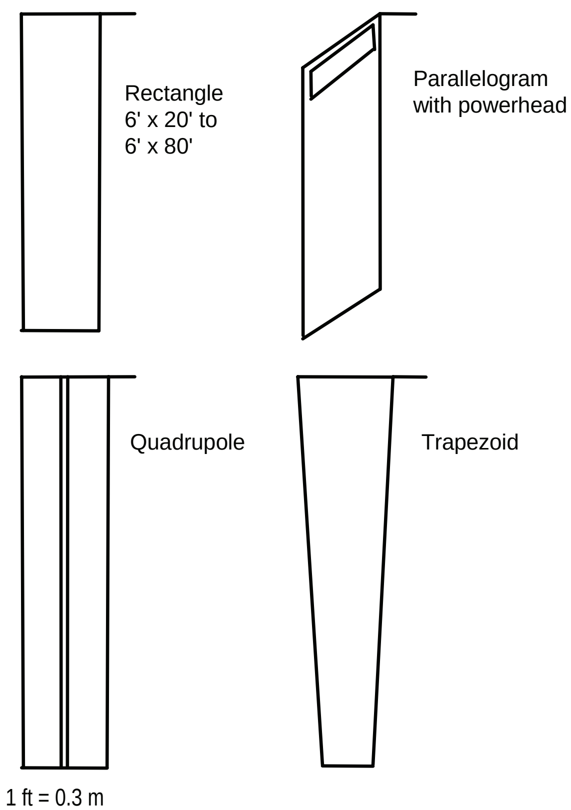


Figure 4-11. Long loop shapes

The effective time gap is equal to the travel time required to traverse the length of the loop plus one vehicle length plus the passage time. Thus, the length of the loop is a critical measure for providing appropriate operation. The length must be sufficient for a following car to come to a stop if the yellow interval appears just before the following vehicle reaches the loop or, conversely, to allow the vehicle to proceed through the intersection on the yellow.

If heavy trucks are included in the traffic stream, there may be a start-up issue if a long queue exists. Passenger cars in front of the truck may accelerate and clear the inductive-loop detector before the truck can accelerate and reach the detector.

One researcher examined the relationship between inductive-loop detector length and the time settings of vehicle interval and maximum green for intersections where vehicle approach speeds were less than 35 mi/h (56 km/h).[(8)] The purpose of the study was to determine the optimal combinations of loop length, vehicle interval, and maximum green for a wide range of flow conditions (i.e., flow rate per lane, distribution of traffic among lanes, and temporal variations in flow rates). Both two- and four-phase operation of presence mode control were analyzed for each flow pattern.

Optimal vehicle intervals are a function of inductive-loop detector length and flow rate. The study suggests that for loops 30 ft (9 m) long, the use of 2-second vehicle intervals can lead to the best signal performance over a wide range of operating conditions. For 50-ft (15-m) loops, 1-second vehicle intervals are desirable under a variety of flow conditions. When loops 80 ft (25 m) long are used, 0-second vehicle intervals can minimize delays. Longer vehicle intervals for such loop lengths are not desirable unless the combined critical flow at an intersection exceeds 1,400 v/h.

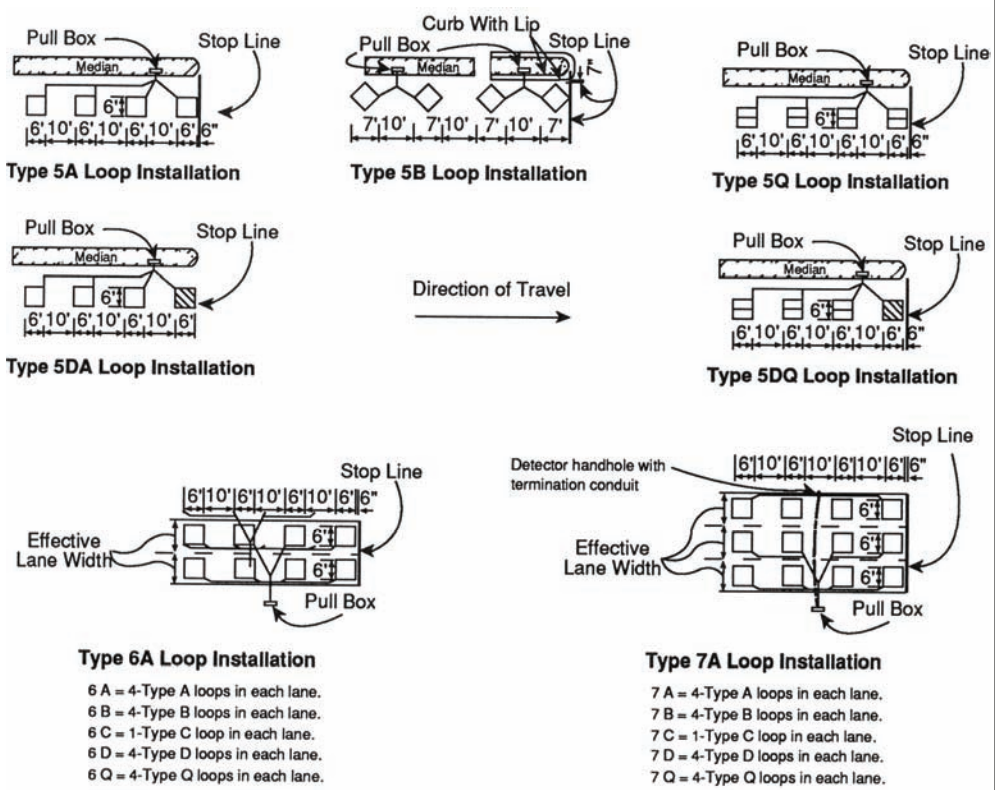
The study concluded that maximum green for presence-mode control is generally longer than optimal green durations for pretimed control. Flow patterns characterized by a larger concentration of traffic in short periods of time need longer maximum greens. The optimal maximum greens for hourly flow patterns with a peaking factor of 1.0 (a uniform flow rate) are about 10 seconds longer than the corresponding optimal pretimed greens. With a peaking factor of 0.85 (a larger concentration of traffic in a short period), the optimal maximum greens are approximately 80 percent longer than the corresponding optimal pretimed greens.

The study further concluded that loops 80 ft (25 m) long could consistently produce the best signal performance. However, 65-ft (20-m) loops gave comparable performance when the combined critical flow was less than 1,100 v/h. When the combined critical flow was less than 900 v/h, 50-ft (15-m) loops rather than 80-ft (25-m) loops incurred a delay of up to 2 seconds per vehicle. For a combined critical flow of less than 600 v/h, 30-ft (9-m) loops may be used instead of 80-ft (25-m) loops without incurring undue delays.

Sequential Short Loops

Use of sequential short loops to emulate a long loop is the preferred treatment in many agencies. The advantages of this configuration result primarily from fewer failures because of the loop's shorter length. Thus, they are less vulnerable to problems caused by crossing pavement cracks and joints and to adjacent lane pickup (splashover). Long loops are more subject to adjacent lane splashover since the entire length of the vehicle is exposed to the side of the long loop (approximately 17 ft (5 m)) as compared to less than a third of the vehicle length of about 6 ft (1.8 m) for a short loop. The short loops also provide superior detection of small vehicles, as explained later.

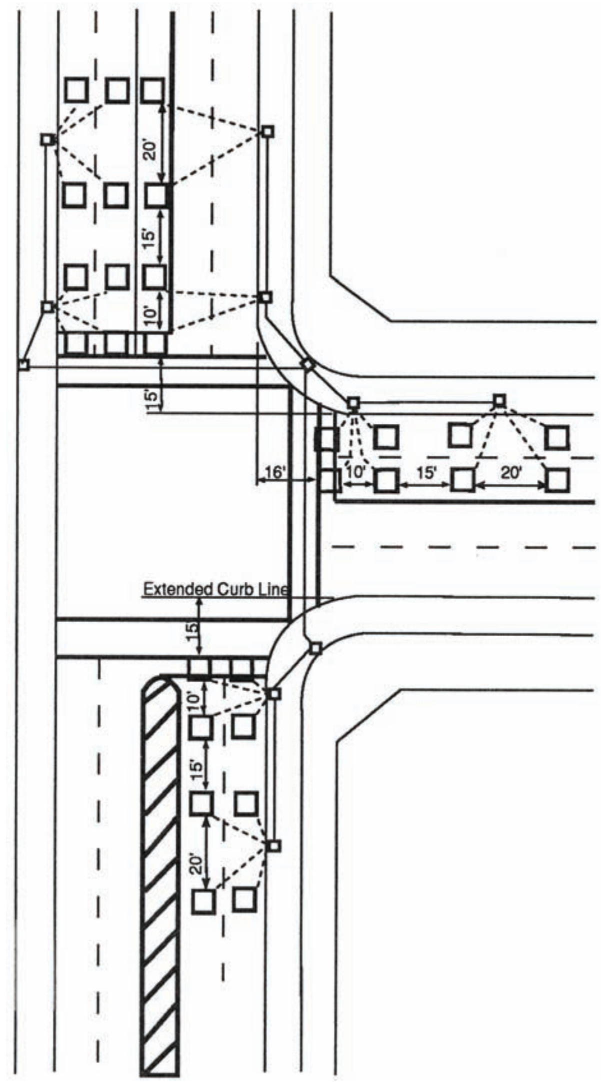
Sequential short loops usually consist of four 6- x 6-ft (1.8- x 1.8-m) square or diamond loops separated by 9 or 10 ft (2.7 or 3.0 m). This configuration is equivalent to a 51- or 54-ft- (15.3- or 16.2-m-) long loop. Figure 4-12 shows various configurations of sequential loops used by Caltrans. This standard employs loop-type designations defined in Figure 4-10.



1 ft = 0.3 m

Figure 4-12. Caltrans standard sequential inductive-loop configurations.

Figure 4-13 demonstrates a different spacing pattern the Pennsylvania DOT (PennDOT) uses for installations of sequential short loops. This spacing normally requires a passage time or vehicle interval greater than zero to provide proper signal operation. Other spacing configurations are used by agencies around the Nation.

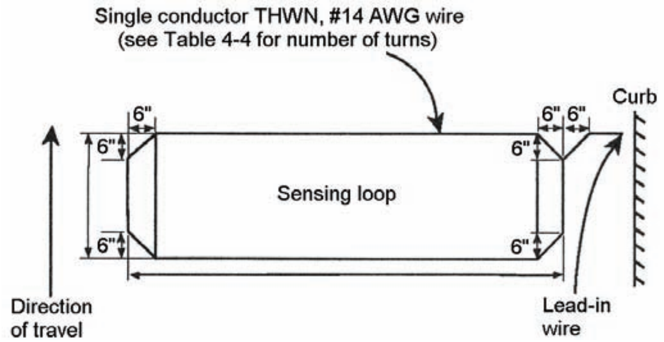


1 ft = 0.3 m

Figure 4-13. PennDOT short inductive-loop configurations.

Wide Loops

Some agencies use wide loops to cover wide lane or multiple lane approaches. These loops are normally 6 ft (1.8 m) in length in the direction of traffic flow and up to 46 ft (14 m) in width for a four-lane approach. The basic loop configuration for a wide lane is shown in Figure 4-14. The number of turns of wire varies according to the number of lanes covered. Table 4-4 gives the number of turns and the dimensions for the loop. Wide loops are generally not recommended nor are permitted by many agencies because they are subject to more frequent failure from crossing pavement joints and fractures. A failure anywhere on the perimeter takes the entire loop out of operation, which in turn removes all detection capability for that approach. Separate loops in each lane are less susceptible to failure and, even if a failure in one loop occurs, the remaining loops can provide approach detection.



1 inch = 2.54 cm

Figure 4-14. Wide inductive-loop detector layout.

Table 4-4. Wide inductive-loop detector dimensions.

Number of lanes	Total approach width of lanes		Loop width		Loop length		Turns of wire
	feet	meters	feet	meters	feet	meters	
1	12–18	3.7–5.6	6–12	1.8–3.7	6	1.8	4
2	19–32	5.7–9.8	13–26	3.7–8.1	6	1.8	3
3	33–45	9.9–13.6	27–39	8.2–12.1	6	1.8	2
4	46–52	13.7–15.8	40–46	12.2–13.7	6	1.8	1

Large loops of up to 30 ft (9 m) in width and 50 to 60 ft (15 to 18 m) in length have been installed to extend green time when occupancy increases to a saturation level in a given direction. The inductive-loop detector electronics unit is adjusted to be sensitive to more than some specified number of vehicles in the loop. Thus, the electronics unit only responds to a saturated condition. No additional green extension is given to the approach unless there is congestion. If no extensions are present (i.e., there is no saturation or congestion), the opposing street green receives the excess time. This application cannot be used for call initiation and is intended for use only in locations where unpredictable and extreme fluctuations of traffic are present, such as shopping center exits, some freeway exits onto main street flow, and industrial plant parking lot exits.

LEFT-TURN LANE DETECTION

Vehicle sensors in left-turn lanes can affect the capacity of an intersection by reducing unnecessary green time and left turn arrow indications. When the last vehicle proceeds a block or so past the signal before the conflicting phase begins, the travel time represents lost green time, which could more appropriately be used to increase the green time available to other phases. The design of left-turn detection is generally based on the premises discussed below.

From 3 to 5 seconds are normally required at the start of the green indication for the first vehicle in a queue to start up, with an average headway of 2 to 3 seconds between following vehicles. Longer startup and headway times are accounted for by providing an appropriate loop length to maintain the green for these slower vehicles. Moreover, trucks and other slow vehicles may require a still longer startup time, which frequently results in a three or four car-length gap ahead of them. Loop length also needs to account for these gaps.

Because green time is based on vehicular demand, only a short green time is needed for one or two vehicles. For example, a rapidly starting single vehicle can clear the turn lane with a green time of less than 5 seconds. A driver of a following vehicle just entering the left turn lane may be confused by the short green. The length of the inductive loop should allow the following car to reach the loop in time to enter the intersection on the green indication or brake to a stop. This length is based on the equation for maximum deceleration rates, which indicates that a vehicle traveling at 30 mi/h (48 km/h) can stop in 83 ft (25 m).

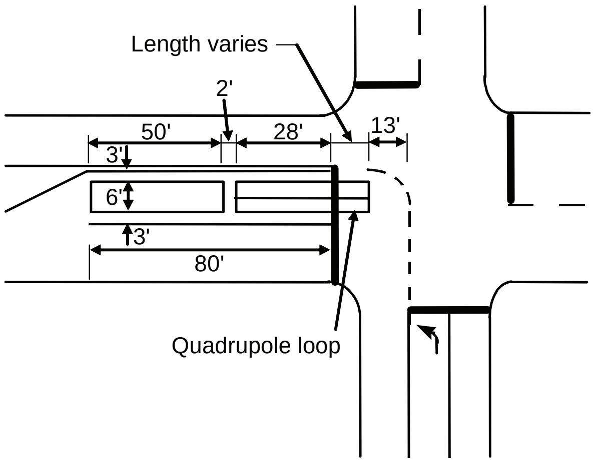
To accommodate these conditions, a loop length of 80 ft (24 m) from the stopline and a controller passage time of 1 second are frequently used. Adding more passage time on the controller compensates for passage over shorter loops. Controller passage time is the time a controller holds the green after actuation. A passage time of 1.0 second permits most motorists to almost complete their turning radius before the onset of the yellow change interval is displayed.

Another problem occurs when vehicles are permitted to turn on the circular green (green ball) indication. Drivers will usually proceed past the stopline and wait for a gap in the opposing traffic. If a gap does not occur or a vehicle ahead prevents the turn, the driver may be left stranded beyond a detection zone that ends at the stopline. In this case, the controller may skip the turn arrow in the next cycle because the vehicle is positioned ahead of the sensor's detection zone. Some agencies think it a good practice to extend the loop beyond the stopline 1 to 6 ft (0.3 to 2 m) to prevent this situation.

Small-vehicle (e.g., motorcycle) detection using short inductive loops requires a high sensitivity setting on the electronics unit. However, high sensitivity will frequently cause detection of vehicles in adjacent lanes (splashover). Many agencies use quadrupole loops to avoid splashover. As the quadrupole requires an additional sawcut equal to the length of the loop, it is desirable to limit quadrupole installation to the area near the stopline. Quadrupole design is discussed further under the topic of "Detection of Small Vehicles."

Figure 4-15 illustrates a minimum-length quadrupole left-turn loop designed by the Illinois DOT (IDOT) using the procedure below:[6]

- Locate the stopline in relation to cross-street turning radius.
- Measure back 80 ft (24 m) from the stopline to establish the back loop.
- Measure 50 ft (15 m) toward the stopline to establish front edge of the back loop.
- Allow 2-ft (0.6-m) gaps between loops and measure 28 ft (8.4 m) to the stopline.
- The front of the front loop should be within 13 ft (4 m) of the edge of the cross street traffic lane.
- Design the front loop as a quadrupole to detect small vehicles.



1 ft = 0.3 m

Figure 4-15. Left turn detection inductive-loop configuration used by IDOT.

When the left-turn demand requires 150 ft (45 m) or more of storage length or when higher approach speed requires long deceleration lanes, the loop layout should include an advanced detection loop. The advanced vehicle sensor with a call extension feature will extend the effective detection zone to accommodate heavy traffic volumes or high-speed traffic.

THROUGH-LANE DETECTION

Detection of vehicles in through lanes depends on the approach speed and the type of controller operation being used. Single-point detection, long-loop occupancy detection, a combination of long and short loops, or a sequence of short loops can be deployed. Each of these is discussed below.

Single-Point Detection

This is the simplest type of vehicle detection used for actuated controllers. It is primarily found on low-speed approaches where speeds are less than 35 mi/h (56 km/h). It may also be used on side-street approaches, with another form of detection on the major street.

A point vehicle sensor (e.g., a 6- x 6-ft (1.8- x 1.8-m) inductive loop) is located 2 to 4 seconds of travel time in advance of the stopline. As this is the only sensor in the approach lane, controller timing must be appropriately set to use the information. The actual distance divided by 25 ft (7.6 m) (which approximates the length of a vehicle plus the space between vehicles) indicates the number of vehicles that can be queued between an inductive-loop detector and the stopline when the light turns green. This number establishes the minimum green interval for the controller. The distance between the inductive loop and the stopline divided by the 15th percentile speed provides a good initial estimate of the passage time. The passage time is also the allowable gap that causes the green indication to drop. If this setting is too long or short to be acceptable as an allowable gap, the position of the loop should be moved to ensure an appropriate gap size.

Long-Loop-Occupancy Detection

Loop-occupancy detection is generally used on low-speed approaches. It normally consists of a single loop 50 ft (15 m) or more in length or a sequence of short loops (usually four) located immediately upstream of the stopline. Loop-occupancy timing is used on the controller as described earlier. This type of operation is most effective when speeds are 25 mi/h (40 km/h) or less. Even at this speed, there is some potential for the signal turning yellow just before an approaching vehicle reaches the loop. In this case, the vehicle will probably cross the intersection on the yellow indication.

As speeds increase, the detection zone must lengthen to accommodate the increased stopping distance. One jurisdiction uses the combination of approach speeds and loop lengths shown in Table 4-5. These loop lengths appear to be excessively long, resulting in long minimum gaps.

The 120-ft (37-m) detection area is measured from the stopline and consists of two 56-ft (17-m) loops. Where greater detection areas are required, either additional long loops or small loops may be used. If additional small loops are chosen, they must be connected to separate electronics units with the extension time programmed into the unit. The long loops are set to presence mode.

Table 4-5. Loop lengths for long-loop-occupancy detection.

Speed		Loop length	
mi/h	km/h	feet	meters
30	48	120	37
35	56	160	49
40	64	200	61
45	72	250	76

High-Speed Point Detection

For high-speed approaches (those with speeds greater than 48 km/h (30 mi/h)), detection becomes more complex. Volume-density control is one technique used that relies on the controller functions rather than extensive detectorization. Normally only one sensor is installed in each lane. This point sensor is usually placed at least 5 seconds and as much as 8 to 10 seconds from the stopline, which is more than the 2 to 4 seconds of travel time required for normal actuated operation.

The sensor is active at all times rather than just during the green interval. During the red interval, each actuation increments the variable initial timing period. Once the variable initial timing period exceeds the minimum green, each additional actuation adds an additional user specified time increment to the initial interval. If the variable initial timing does not exceed the initial interval, then that duration is used instead for the first portion of the green interval. During the green interval, the sensor is used to extend the green. At first the extension is equal to the passage time, but after a conflicting phase has registered a call, the extension is reduced, eventually reaching a minimum gap.

This type of signal control has the potential for trapping a vehicle in the dilemma zone at the onset of yellow. The following section discusses dilemma zones and describes how multiple sensors are used to alleviate the dilemma zone problem.

DETECTION FOR DILEMMA ZONES

Signalized intersections where speeds of approaching traffic are greater than 30 mi/h (48 km/h) have long been of concern to designers and operations and safety engineers. Drivers approaching at these higher speeds are frequently confronted with a dilemma—whether to stop or proceed through the intersection at the onset of the yellow change interval. The placement of sensors to ameliorate this problem has received serious consideration and research. This section defines the many variables that affect the dilemma zone problem and describes several sensor placement schemes that have proven effective.

Definition of the Dilemma Zone Problem

When a vehicle traveling at a constant speed S approaches an intersection and is positioned at distance X from the intersection at the beginning of the yellow change interval, the driver is faced with a decision. He may decelerate and stop the vehicle before entering the intersection or continue and enter the intersection, accelerating if necessary before the red interval begins. In some States, the driver is required to clear the intersection before the red appears. Depending on the distance from the intersection and the speed of travel, drivers may not be certain that they can stop in time, or they may be unsure that they can clear the intersection before conflicting vehicles enter. This creates the dilemma. Some drivers will opt to stop, while others may accelerate and continue through the intersection.

If the choice is to stop, the driver will decelerate after a short perception and reaction time. The distance the vehicle travels after the beginning of the yellow change indication includes the distance traveled during the perception and reaction time t and the distance traveled during deceleration. The inequality that must be maintained to ensure a safe and complete stop is given by

$$X \geq St + \frac{S^2}{2d} \qquad (4\text{-}2)$$

where

- X = distance from stopline at start of yellow interval, ft (m)
- S = approach speed, ft/s (feet per second) (m/s (meter per second)
- t = perception and reaction time (typically, 1 second)
- d = constant deceleration rate, ft/s^2 (m/s^2).

Safety and comfort require that the vehicle's deceleration rate not exceed one-third to one-half the acceleration of gravity. Using d^* to represent a critical deceleration rate under prevailing roadway conditions, a stopping distance X_s may be defined as

$$X_s \geq St + \frac{S^2}{2d^*} \qquad (4\text{-}3)$$

where X_s is the minimum distance from the stopline in which the vehicle can come to a complete stop after the beginning of the yellow interval. Thus, if a vehicle is closer to the stopline than X_s when the yellow begins, the driver will be unable to stop safely or comfortably before the intersection. The area between the stopline and X_s is an area in which drivers should not be expected to stop or cannot stop as shown in the upper portion of Figure 4-16.

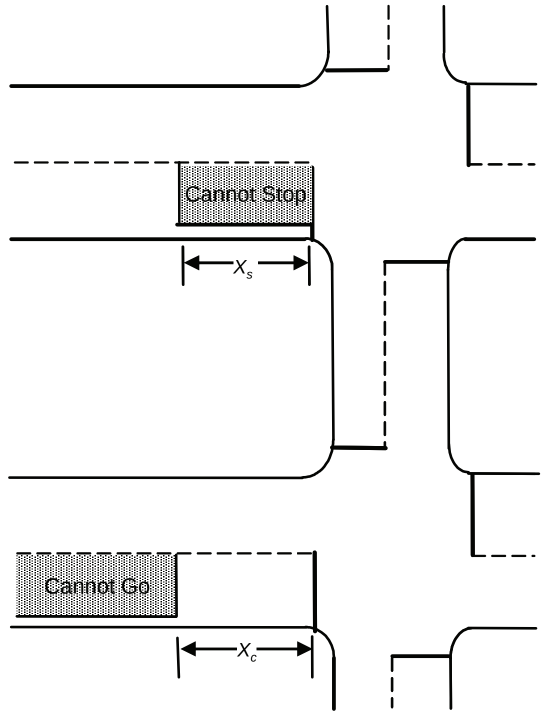


Figure 4-16. "Cannot Stop" and "Cannot Go" regions.

If the driver decides to accelerate and pass through the intersection, a clearance distance X_c must be maintained as defined by the inequality:

Equation 4-4 is based on the general relationship that distance traveled $X_c = X_0 + St + 0.5at^2$, where X_0 is some initial distance.

$$X_c \le -(W+L) + S(Y+R) + \frac{a(Y+R-t)^2}{2} \qquad (4\text{-}4)$$

where

X_c = clearance distance, ft (m)

t = perception and reaction time (typically 1 second)

a = acceleration rate, ft/s² (m/s²)

Y = yellow change interval, seconds

R = red clearance interval, seconds

W = effective width of intersection, ft (m)

L = length of vehicle, ft (m)

S = approach speed, ft/s (m/s)

$W+L$ = correction to stopbar distance to bring the rear of the car beyond the intersection.

Implementing Equation 4-4 permits the driver to completely clear the intersection before the appearance of the red signal. Many traffic engineers do not believe that a driver must clear the intersection on the yellow. In fact, most State vehicle codes do not require the vehicle to clear the intersection prior to the onset of the red indication, but merely to have entered prior to the red. If this interpretation applies, engineers may eliminate the $(W+L)$ term from the equation or may use 0.5 or 0.25 of the value of this term. In some cases, the red clearance interval is increased to assure clearance when needed rather than include the $(W+L)$ term in the equation.

The constant acceleration rate a available to the driver in Equation 4-4 may be estimated from Gazi's equation as

$$a = 4.9 - 0.213S \text{ in metric units} \tag{4-5a}$$

or

$$a = 16.0 - 0.213S \text{ in English units.} \tag{4-5b}$$

These equations show that larger acceleration rates can be attained when a vehicle is traveling at lower speeds. Clearance distance X_c is the maximum distance from the stopline at which a vehicle can clear the intersection as defined by

$$X_c = -(W+L) + S(Y+R) + \frac{a(Y+R-t)^2}{2}\,. \tag{4-6}$$

Since X_c is the maximum distance upstream of the stopline from which a vehicle can clear the intersection during the yellow interval, any vehicle positioned at a point beyond X_c (i.e., further upstream) would not be expected to clear the intersection during the yellow interval, and is thus in a region in which the driver "cannot go" without violating the red indication (see lower portion of Figure 4-16).

As both X_s and X_c are measured distances from the stopline, the relationship of these two quantities is defined by one of the following conditions:

- $X_s > X_c$.
- $X_s = X_c$.
- $X_s < X_c$.

When $X_s > X_c$, the dilemma zone is the overlapping area in which a vehicle can neither stop nor go if faced with a yellow indication as indicated in Figure 4-17. In this case, the driver of a vehicle within the dilemma zone at the onset of yellow has to accelerate or decelerate at an unsafe rate and consequently is vulnerable to a right-angle or rear-end accident.

When $X_s = X_c$, Figure 4-18 shows that the dilemma zone and its associated problems disappear. A driver in the "cannot go" region is able to stop safely, whereas a driver in the "cannot stop" region can successfully accelerate through the intersection.

Finally, when $X_s < X_c$, a driver in the area between X_s and X_c may either stop or go safely. Therefore, this region is considered an optional zone as depicted in Figure 4-19.

This relatively simply analysis indicates that a dilemma zone is only formed when $X_s > X_c$. Equation 4-3 shows that X_s is a function of speed, perception and reaction time, and deceleration rate, while Equation 4-4 shows that X_c is a function of speed, perception and reaction time, acceleration rate, yellow and red interval times, and effective width of the intersection.

Tables 4-6 (English units) and 4-7 (metric units) contain stopping distance X_s and clearance distance X_c for deceleration rates of 10 and 16 ft/s^2 (3.0 and 4.9 m/s^2) for an intersection width of 48 ft (15 m). Tables 4-8 (English units) and 4-9 (metric units) repeat the stopping and clearance distances for the same deceleration rates for an intersection width of 76 ft (23 m).

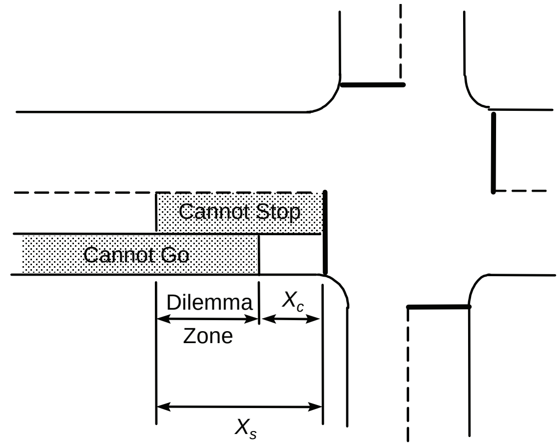


Figure 4-17. Dilemma zone ($X_s > X_c$).

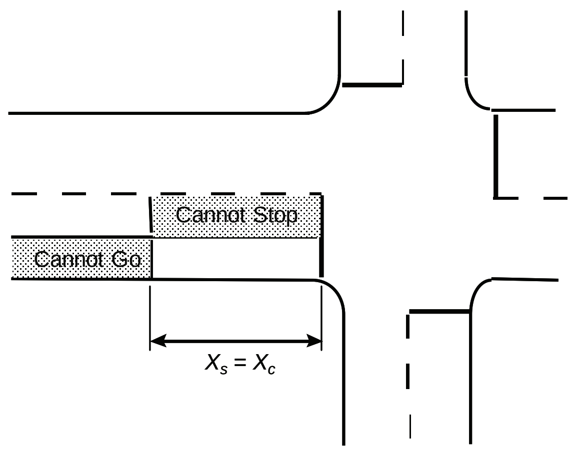


Figure 4-18. Dilemma zone removal ($X_s = X_c$).

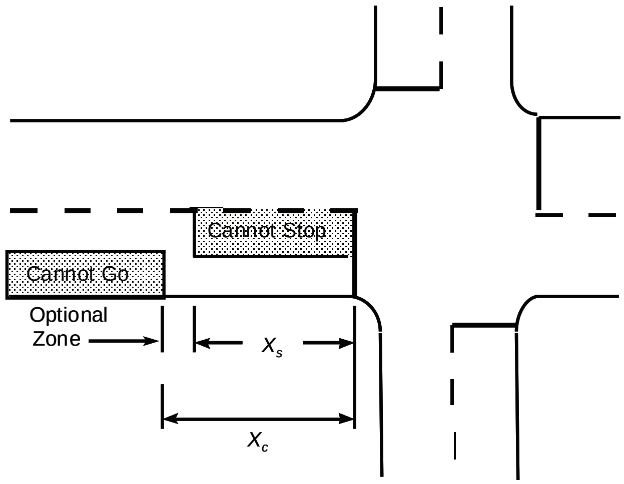


Figure 4-19. Optional zone creation ($X_s < X_c$).

Tables 4-16 through 4-19 are based on the following assumptions:

- All distances are measured from the stopline in the upstream direction.
- Vehicle length is 20 ft (6 m).
- Acceleration rate $a = 16.0 – 0.213$ x speed in units of ft/sec^2 or $a = 4.9 – 0.213$ x speed in units of m/s^2.
- Perception plus reaction time equals 1 second.
- Red clearance time equals 0 second.

Several general conclusions from the analysis of the dilemma zone problem are:

- The dilemma zone size increases as speed increases for a given yellow interval.

Table 4-6. Stopping and clearance distances for intersection width of 48 feet.

Decel. rate	Speed	Stopping distance	Clearance distance (ft) for yellow interval equal to:		
ft/sec²	mi/h	feet	3 sec	4 sec	5 sec
	20	73	39.5	93.2	156.6
	25	104	58.3	115.5	180.8
	30	141	77.2	137.8	205.0
	35	184	96.1	160.1	229.1
10	40	232	115.0	182.4	253.3
	45	285	133.8	204.7	277.5
	50	344	152.7	227.0	301.7
	55	408	174.0	254.6	335.3
	60	477	196.0	284.0	372.0
	20	56	39.5	93.2	156.6
	25	79	58.3	115.5	180.8
	30	105	77.2	137.8	205.0
	35	134	96.1	160.1	229.1
16	40	167	115.0	182.4	253.3
	45	203	133.8	204.7	277.5
	50	242	152.7	227.0	301.7
	55	285	174.0	254.6	335.3
	60	331	196.0	284.0	372.0

Table 4-7. Stopping and clearance distances for intersection width of 15 meters.

Decel. rate	Speed	Stopping distance	Clearance distance (m) for yellow interval equal to:		
m/sec²	km/h	meters	3 sec	4 sec	5 sec
	20	73	10.3	26.4	45.7
	25	104	13.8	30.6	50.2
	30	141	17.4	34.8	54.8
	35	184	24.6	43.3	64.0
3	40	232	31.7	51.7	73.1
	45	285	38.9	60.2	82.3
	50	344	42.4	64.4	86.9
	55	408	46.0	68.6	91.4
	60	477	54.0	79.0	104.0
	20	56	10.3	26.4	45.7
	25	79	13.8	30.6	50.2
	30	105	17.4	34.8	54.8
	35	134	24.6	43.3	64.0
5	40	167	31.7	51.7	73.1
	45	203	38.9	60.2	82.3
	50	242	42.4	64.4	86.9
	55	285	46.0	68.6	91.4
	60	331	54.0	79.0	104.0

Table 4-8. Stopping and clearance distances for intersection width of 76 feet.

Decel. rate ft/sec²	Speed mi/h	Stopping distance feet	Clearance distance (ft) for yellow interval equal to: 3 sec	4 sec	5 sec
10	20	73	11.5	65.2	128.7
	25	104	30.4	87.5	152.9
	30	141	49.3	109.8	177.0
	35	184	68.1	132.1	201.2
	40	232	87.0	154.4	225.4
	45	285	105.9	176.7	249.5
	50	344	124.8	199.0	273.7
	55	408	146.0	226.7	307.3
	60	477	168.0	256.0	344.0
16	20	56	11.5	65.2	128.7
	25	79	30.4	87.5	152.9
	30	105	49.3	109.8	177.0
	35	134	68.1	132.1	201.2
	40	167	87.0	154.4	225.4
	45	203	105.9	176.7	249.5
	50	242	124.8	199.0	273.7
	55	285	146.0	226.7	307.3
	60	331	168.0	256.0	344.0

Table 4-9. Stopping and clearance distances for intersection width of 23 meters.

Decel. rate m/sec²	Speed km/h	Stopping distance Meters	Clearance distance (m) for yellow interval equal to: 3 sec	4 sec	5 sec
3	30	19.9	2.3	18.4	37.7
	35	25.5	5.8	22.6	42.2
	40	31.7	9.4	26.8	46.8
	50	46.0	16.6	35.3	56.0
	60	63.0	23.7	43.7	65.1
	70	82.5	30.9	52.2	74.3
	75	93.2	34.4	56.4	78.9
	80	104.5	38.0	60.6	83.4
	90	129.2	46.0	71.0	96.0
5	30	15.3	2.3	18.4	37.7
	35	19.2	5.8	22.6	42.2
	40	23.5	9.4	26.8	46.8
	50	33.2	16.6	35.3	56.0
	60	44.4	23.7	43.7	65.1
	70	57.3	30.9	52.2	74.3
	75	64.2	34.4	56.4	78.9
	80	71.6	38.0	60.6	83.4
	90	87.5	46.0	71.0	96.0

- For a given speed and yellow interval, increases in the deceleration or acceleration rate cause a reduction in the length of the dilemma zone.
- Increases in the effective width of the intersection directly increase the length of the dilemma zone if the total width of the intersection is included in the calculation.

In pretimed signal control, the appropriate strategies for controlling the dilemma zone problem consist of providing a consistent yellow change interval and incorporating an appropriate red clearance interval. This strategy will, however, increase vehicular delay.

In actuated signal controlled intersections, the most appropriate strategy for resolving the dilemma zone problem involves sensor placement before, within, and after the dilemma zone in such a way as to reduce the probability of entrapment of a vehicle in the dilemma zone at the onset of the yellow interval. The various methods of sensor placement for dilemma zones are discussed below.

Multiple-Point Detection

The dilemma zone problem can be ameliorated by the strategic placement of multiple sensors at high-speed approaches to intersections controlled by actuated controllers. The mitigation methods described below assume the use of inductive-loop detector systems operating in the presence mode. As inventive as the procedures are, vehicles will still get caught in the dilemma zone because of maximum greens, force-offs, etc. Consequently, adequate change intervals (yellow and all-red displays) must be provided to ensure motorist safety.

The three sensor placement strategies in general use for multiple-point detection are:

- Green extension systems (for semiactuated controllers).
- Extended call detection systems (for basic controllers).
- Multiple point detection systems.

Green Extension System

A green extension system consists of an assembly of extended call sensors and auxiliary logic, which enable vehicles to be detected before entering the dilemma zone and provide the controller with data to extend the green until the vehicles clear the dilemma zone.[9] The logic monitors the signal display, enables or disables selected call sensors, and holds the controller in green. Although two loops are normally employed, three may be used at high-speed intersections. Figure 4-20 shows loop placement for vehicles traveling at the 85th percentile speed.[10]

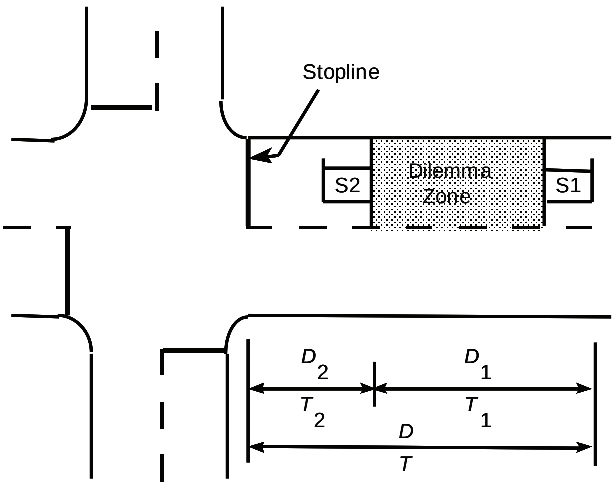


Figure 4-20. Green extension system using two inductive-loop detectors.

The appropriate distances for placing the loops are calculated using

$$D = 1.47St + \frac{S^2}{30f} \qquad (4\text{-}7)$$

$$D_2 = 1.47S\left(\frac{S}{30} + 1\right) \qquad (4\text{-}8)$$

$$D_1 = D - D_2 \qquad (4\text{-}9)$$

where

S = 85th percentile speed, mi/h (km/h)

t = perception and reaction time, seconds

f = coefficient of friction

D = stopping distance, ft (m)

D_2 = clearing distance, ft (m)

D_1 = separation between loops, ft (m).

With the loops positioned as shown, a vehicle passing over loop S1 actuates an electronic timer, which extends the green for the vehicle to reach loop S2 in time T_1. Similarly, when the vehicle passes over loop S2, a second timer maintains the green while the vehicle proceeds toward the intersection. This design does not insure that vehicles traveling at speeds less than the 85th percentile speed would not be trapped in the dilemma zone.

Extended-Call Detection System

This concept uses a 70-ft (21-m) presence loop extending upstream from the stopline and a small extended call loop 250 to 500 ft (75 to 150 m) upstream of the stopline, as shown in Figure 4-21. The magnitude of D is calculated from Equation 4-7, using the speed limit or the 85th percentile speed. D_2 is set equal to 70 ft (21 m). Time T_1 is calculated as D_1 (found from Equation 4-9) divided by a lower-limit approach speed, which is generally assumed equal to the 15th percentile speed. This time is programmed into the extended-call electronics unit. The controller is operated in the loop occupancy control mode.

Proper design of an extended-call detection system ensures that the last vehicle and those vehicles traveling below the speed limit are not trapped in the dilemma zone. A trailing vehicle may be trapped at the end of the maximum-extension limit (maximum green time after an actuation on an opposing phase). The 70-ft (21-m) presence loop guarantees that stopped vehicles queued behind the stopline will move forward and enter the intersection without triggering a premature gap out.

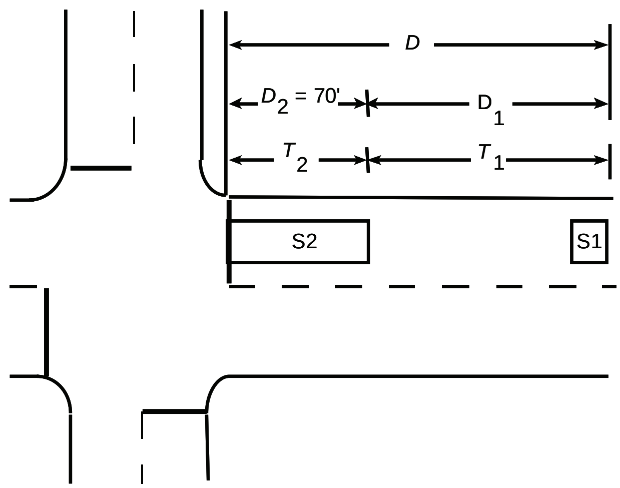


Figure 4-21. Extended call inductive-loop detector system.

Multiple-Point Detection Methods

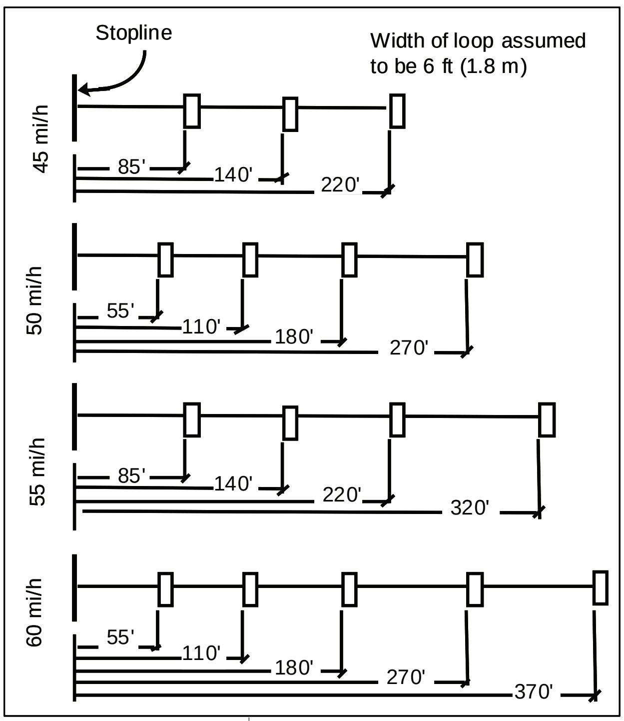
The green extension and the extended-call detection systems used two or (at most) three sensors. These techniques can be used effectively at intersections with relatively low speeds. However, when speeds are high, the dilemma zone becomes longer, and more sensors are needed to accommodate the large range of approach speeds that are generated. Three techniques are commonly used for determining the placement of the required sensors. These methods are identified by the developer, agency, or organization that pioneered the technique.

Beierle Method: This method (originated by Harvey Beierle of the Texas DOT[11] (TxDOT) uses a 1-second vehicle interval setting on a controller operating with locking detection memory. The sensors are 6- x 6-ft (1.8- x 1.8-m) presence mode inductive loops.

The outermost sensor upstream of the intersection is placed at a safe stopping distance from the intersection for highest normal approach speed. Safe stopping distances are based on a 1-second perception and reaction time plus braking distances resulting from coefficients of friction between 0.41 and 0.54 for speeds between 55 and 20 mi/h (90 and 30 km/h). The next sensor is tentatively located at a safe stopping distance for a vehicle traveling 10 mi/h (16 km/h) less than that assumed for the first sensor. If the travel time between the two sensors is greater than 1 second, the downstream sensor is relocated to allow the vehicle to reach the second sensor within the 1-second vehicle interval set on the controller.

This location procedure is repeated for each successive sensor until the last loop is within 75 ft (23 m) of the stopline, each time subtracting 10 mi/h (16 km/h) from the maximum considered speed. The minimum assured green time is set on the controller to permit vehicles stopped between the last sensor and the intersection to enter the intersection.

TxDOT examined a modification to the Beierle procedure, which uses American Association of State Highway and Transportation Officials (AASHTO) stopping distance criteria. Figure 4-22 illustrates the sensor spacing for speeds common to Texas.



1 ft =0.3 ft
1 mi/h = 1.6 km/h

Figure 4-22. Multiple inductive-loop detector placement (TxDOT) for alleviating the effects of dilemma zones.

Winston-Salem Method: The second method of multiple sensor placement was developed by Donald Holloman for that agency.[9] It is basically the same as the Beierle Method. The Winston-Salem Method differs by using a slightly shorter stopping distance for the outermost and innermost sensor and incorporates speeds up to 60 mi/h (97 km/h).

SSITE Method: The third method was developed by the Southern Section of the Institute of Transportation Engineers (SSITE).[12,13] It uses an iterative process and engineering judgment in locating the sensors. The outermost loop is positioned to provide safe stopping distance, as determined by data collected by the Southern Section. The differences with respect to the other multiple sensor methods are:

- Spacing between successive loops is 2 seconds.
- Innermost loop is located at the stopline.
- Allowable gap is set between 5 and 7 seconds, which is greater than in the other methods.

Tables 4-10 and 4-11 summarize the characteristics of the extended call and multiple point detection methods with respect to the number of loops and their spacing. Since the length of the dilemma zone becomes larger as speed increases, more sensors are required to track the vehicle through the dilemma zone. Moreover, larger spacing between sensors implies detection of only larger vehicle intervals and less efficient controller operation.

In general, it appears that the multiple point detection methods are more appropriate for use with high mean approach speed and high flow variability.

Table 4-10. Inductive-loop detector placement in extension or extended-call systems used to ameliorate effects of dilemma zones.

	Green extension systems for the semiactuated controllers method	Extended-call detection systems for basic controllers method
Parameter	Value	Value
Controller memory	Nonlocking	Nonlocking
Sensor type	Presence	Presence
Speed range	S = 85th percentile	S = 85th percentile
Outermost loop[a]	$D = 1.47St + \frac{S^2}{30f}$	$D = 1.47St + \frac{S^2}{30f}$
Innermost loop	$D_1 = 1.47S \; \frac{S}{30} + 1$	0
Spacing between loops[b]	$\frac{D - D_1}{S} > 2\,\text{sec}$	$\frac{D - 70}{S_{low\ limit}} > 2\,\text{sec}$
Number of Loops	2 (or 3)	2
Allowable gap	5 to 6 Seconds	5 to 6 Seconds

a. Distance D is measured from the stopline.
b. $S_{low\ limit}$ = Low speed limit; for example, 15th percentile speed.

Table 4-11. Inductive-loop detector placement in multiple-point detection systems used to ameliorate effects of dilemma zones.

	Beierle Method	Winston-Salem Method	SSITE Method
Controller memory	Locking	Nonlocking	Nonlocking
Sensor type	Presence	Presence	Presence
Speed range	$S \leq 50$	$S \leq 60$	$S \leq 60$
Outermost loop[a]	Use stopping distance from Beierle, Ref. 11	Use stopping distance from Sackman, Ref. 9	Use Southern Section ITE Report, Ref. 12
Innermost loop	Within 75 ft (23 m) of approach stopline	86 ft (26 m)	00
Spacing between loops[b]	1 second	1 second	2 seconds
Number of Loops	$\left\lvert \frac{S}{10} \right\rvert - 1$	$\left\lvert \frac{S}{10} \right\rvert - 2$	≤ 6
Allowable gap	2 to 5 seconds	2 to 5 seconds	5 to 7 seconds

a. Distance *D* is measured from the stopline.
b. $S_{low\ limit}$ = Low speed limit; for example, use 15th percentile speed.
c. $\lvert S/10 \rvert$ represents the integer part of S/10; for example $\lvert 37/10 \rvert = \lvert 3.7 \rvert = 3$

Dilemma Zone Design Options

To simplify their sensor design process, PennDOT defined seven basic design options together with an evaluation of the characteristics of each option. The following excerpt describes these options from the PennDOT *Traffic Signal Design Handbook*.[14]

> Option 1: Consists of a long loop 6 x 50 ft (1.8 x 15 m) maximum for each approach lane. This enables individual lane detection in the presence mode. Although it requires more loop wire than the other options, its initial cost is the lowest as less lead wire and fewer pull boxes are required. Construction cost is lowest of all options. The disadvantages include: all detection for a lane is lost should the loop break and long loops are the least sensitive of all loop configurations. When the sensitivity is increased, the loop becomes more susceptible to detecting vehicles in adjacent lanes.
>
> Option 2: Consists of sequential short loops for individual lane detection in either the pulse or presence mode. They may be wired either in series or parallel; however, best results are achieved when alternate loops are paired and wired in parallel to separate input channels. There is an added safety feature inherent to this option in that should one loop fail, detection is not completely lost. Although the initial cost is higher than that for long loops, maintenance is easier, as only a small loop need be replaced in case of damage.
>
> Like the long loop, the short loops are susceptible to detecting vehicles in adjacent lanes; however, they are more sensitive and are better suited for sensing small vehicles.

Option 3: Consists of a long quadrupole loop for each approach lane. Its operation is identical to that of Option 1. The major advantages of Option 3 are increased sensitivity for detecting bicycles and small motorcycles, coupled with its ability to reject detection of vehicles in adjacent lanes. Like Option 1, detection for an entire lane is lost if the loop be severed. Construction cost is approximately 20 percent greater than for Option 1.

Option 4: Consists of one short loop per lane located in advance of the intersection based on normal approach speeds. This option, which operates in pulse mode only, is best suited for providing extension intervals on roads with higher travel speeds.

The loops installed for Option 4 can also be used for individual lane counting and gap determination. There are two disadvantages with this option. First, should the loop fail, detectorization for the approach is lost. Second, since there are no loops near the stopline, any vehicle entering the approach from a driveway between the loop and the stopline is not detected and has to wait for another vehicle to place a call for the necessary phase unless calling detectors are installed.

Option 5: Basically the same as Option 4, except a single wide loop is used for multilane detection instead of individual lane sensors. Construction is less expensive than Option 4; however, should breakage occur, detection on that approach is completely lost.

Option 6: Consists of a single short loop per approach lane for use where a driveway is located between the intersection and the area of detection for Options 4 or 5. Traffic generated by the driveway is unable to actuate its phase without the additional loops placed near the stopline. A 6-ft x 6-ft (1.8-m x 1.8-m) calling inductive loop is used in these cases.

SPECIAL FUNCTION APPLICATIONS

Several applications require inductive-loop detector systems to perform unique functions. These include small-vehicle and bicycle detection, detection of long vehicles or large trucks, queue detection at freeway off-ramps, vehicle counting, and special safety applications to prevent accidents and reduce speeds.

Summaries of various loop shapes and operating characteristics are given below in terms of the functional requirements determined by the operating agency. These requirements are typically based on the type of application (e.g., intersection control and freeway surveillance and control), traffic or vehicle mix, climate, and other site-specific conditions.

DETECTION OF SMALL VEHICLES

Increased fuel costs tend to accelerate the proclivity for smaller, fuel-conserving vehicles. These range from small compact automobiles to l00-cc motorcycles, mopeds, and lightweight bicycles. The increasing number of these small vehicles and their behavior patterns often necessitates their detection with existing standard inductive-loop detector configurations.

A presence sensor should be able to detect a small motorcycle and hold its call until the display of a green signal. If the sensor drops a call prematurely, the motorcyclist could be trapped on the red phase. The required hold time should at least match the shortest cycle time observed at the intersection. The NEMA Standards (see appendix J) specify a minimum hold period of 3 minutes.

Calls may be dropped prematurely in some older inductive-loop electronics units that include the ability to compensate for environmental drift, primarily due to changes in temperature and moisture. This circuitry will frequently neutralize a weak detection from a small vehicle within a period of less than a minute. Newer electronics units do not have this problem and all meet the NEMA Standards and the Type 170 Specification, which both require a minimum hold time of 3 minutes.

In California and other temperate areas, the bicycle has become a common mode of transportation. As such, properly signalized intersection operation and safety require detection of bicycles. The inherent problems associated with bicycle detection include:

- Locating the loop on the street to assure the rider will be within the detection zone. A separate bike lane is ideal, but not always possible.
- Sequencing the traffic signal to accommodate a detected bicycle. This cannot be done with some control techniques.
- Providing sufficient signal timing to avoid trapping the bicyclist in the intersection. This also can be a problem with some intersection designs.

In response to these problems, it has been suggested that the inductive- loop electronics unit have extension timing and delay features. In such a system, one loop is located about 100 ft (30 m) from the stopline, and the second loop is located at the stopline. When a bicycle is detected at the first loop, the extension time is provided to hold the green to allow the bicycle to reach the loop at the stopline.

When the detection is made at the stopline, extension time is provided to allow the bicycle to move far enough into the intersection to safely clear before the end of the yellow indication. If the detection occurs when the light is red, the minimum timing feature assures that when the light turns green, the minimum green time will allow safe crossing of the intersection. This type of operation works best in a bike lane. The loop in the bike lane with a standard electronics unit could be wired to call the pedestrian timing, which would allow adequate time for the cyclist to cross the intersection.

A delay feature is used where vehicles merge into the bike lane to turn right. The detection is not immediately registered so that the vehicle may complete a right turn without creating delay for other traffic including bicycles.

A number of factors determine the most effective inductive-loop configuration needed to satisfy policies that dictate detection of small vehicles, including bicycles. Important criteria include shape of the loop, width of the lane, and loop placement within that lane. Loop configurations that enhance the ability to detect small vehicles are described below.

Multiple, Interconnected Small Loops

A configuration frequently used for the detection of small vehicles is multiple, interconnected small loops or, as it is often referred to, sequential short loops. The sensitivity can be better controlled with the multiple loops than with the conventional single long loop, whose sensitivity must be set so high for small-vehicle detection that false calls from adjacent lanes (splashover) result. With the small loops, total-loop inductance very close to optimum can be achieved by connecting the loops in series. In the past, series/parallel connections were mandated; however, newer electronics units have obviated this need. Moreover, the small vehicle will be detected when it reaches the first small loop rather than at the stopline, as is the case with the long loop with a powerhead described later. Also, should one loop fail, some detection capability still remains in that lane.

Quadrupole-Loop Configuration

The quadrupole-loop configuration was first used in the early 1970s. As shown in Figure 4-23, this configuration adds a longitudinal sawcut in the center of the lane. The loops are wired in a figure-eight pattern so that the center wires have current flowing in the same direction. Their fields reinforce each other, improving the capability to detect small vehicles. The center wires counteract the fields of the outer wires, which have their current flowing in the opposite direction from the center wires. The influence of the outer fields is diminished, thereby reducing the possibility of splashover.

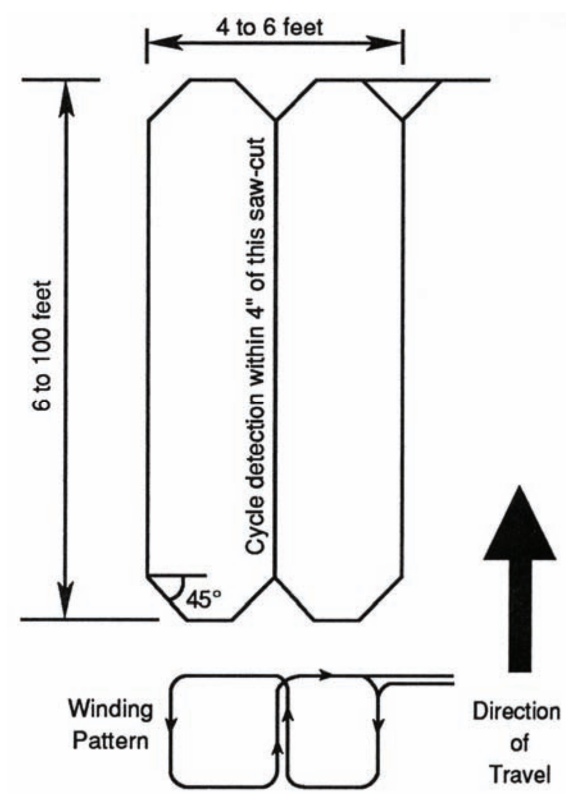


Figure 4-23. Quadrupole loop configuration.

The single-wire configuration ("1-2-1" with one layer in the perimeter slots and two layers in the center slot), shown in Figure 4-23, is used for the detection of automobiles, trucks, and the larger motorcycles. A double layer design ("2-4-2") is recommended for detecting small motorcycles and bicycles. Some agencies wind the 2-4-2 as two, 2-turn loops rather than the traditional, figure-eight winding pattern. No definitive tests were found to favor either method.

Loop placement in the lane is another important consideration. Installation in the center of the lane may fail to detect the small motorcycle if the travel path is outside the quadrupole field. For example, motorcyclists waiting to turn left will usually stop on the left side of the lane and thus may be outside of the quadrupole field. Where detection is required for a left-turn lane, it is recommended that the left edge of the quadrupole be located no further than 2 ft (61 cm) from the left edge of the left-turn lane.

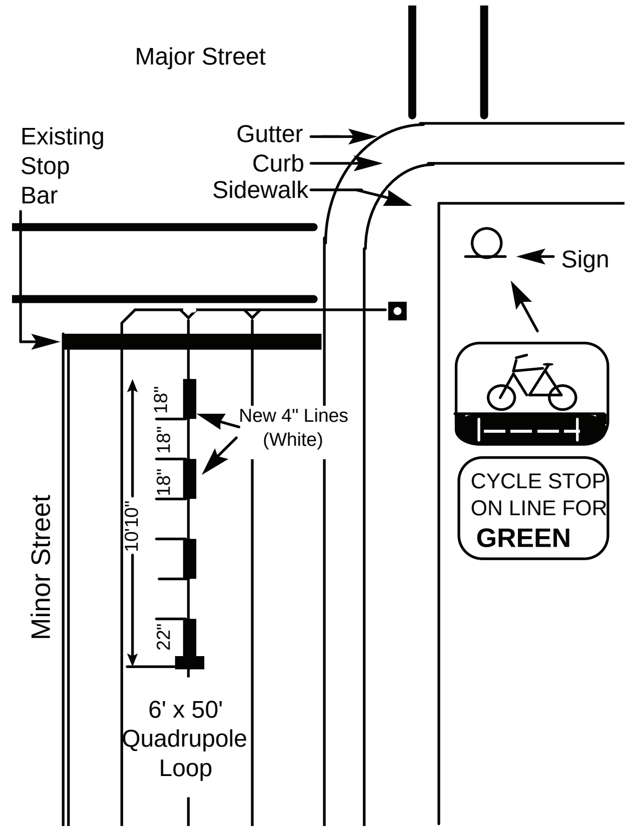
The quadrupole is used as a short or long loop, a single element, and in combinations such as series or series/parallel. The series/parallel configuration is used effectively across the country not only to detect small vehicles, but also to eliminate the problem of adjacent-lane detection (splashover) in high-sensitivity inductive-loop systems.

Short quadrupoles (those less than 30 ft (9 m)) tend to lose high-bed sections of trucks. Since the quadrupole is really two loops whose field detection height is approximately two-thirds of the short leg of the loop, the approximate detection height is 2 ft (0.6 m). This detection height is based on the dimension of two 3-ft- (0.9-m-) wide side-by-side quadrupole loops. With longer loops (those greater than 30 ft (9 m)), there are always one or more wheels or axles over the loop.

It is generally agreed that the 6-ft (1.8-m) quadrupole loop detects bicycles better than most other loop configurations. The major problem with this configuration for small-bicycle detection is the need for the bicyclists to ride close enough to a wire within the quadrupole to be detected. A number of unique pavement markings and signs have been developed to assist this application. The pavement marking identifies the location of the wire and the sign explains the markings to the cyclist.

Figure 4-24 illustrates a marking system used in Clarke County, GA. A pattern of 4- x 18-inch (10- x 46-cm) white stripes is placed 18 inches (46 cm) apart, starting at the stopline. The length is kept short since only one actuation is necessary to call the green indication.

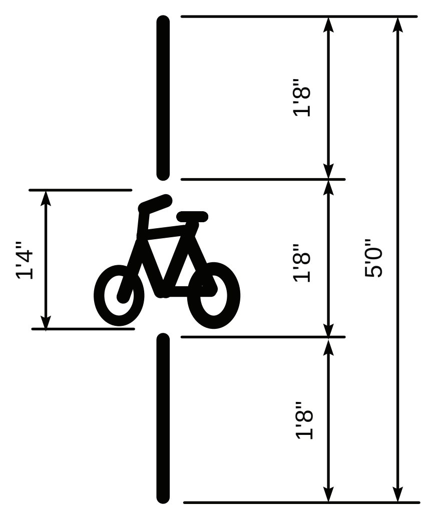
In the City of San Luis Obispo, CA, an aggressive public information program was mounted to inform bicyclists that they could cause the red light to turn green by traveling over the bicycle-shaped pavement marking shown in Figure 4-25. These markings were painted on all appropriate through, right-, and left-turn lanes at all signalized intersections that contained loop sensors.



1 ft =0.3 m

1 inch = 2.5 cm

Figure 4-24. Bicycle sensor sign and markings used in Clarke County, GA.



1 ft =0.3 m

1 inch = 2.5 cm

Figure 4-25. Special bicycle pavement marking used in San Luis Obispo, CA.

Chevron-Loop Configuration

Small vehicles can be detected with the chevron configuration shown in Figure 4-26. It consists of one or more four-turn parallelogram loop(s) with the short section in the direction of traffic and the long section at an angle of 30° with the short section. The long sides of the loop sections are 27.5 inches (70 cm) apart. Adjacent ends of successive loop sections may be in a single slot or separated by 2 ft (61 cm). This alignment allows a vehicle to cut the lines of flux more efficiently.(15)

Loop sections are wound alternately clockwise and counterclockwise so that currents in adjacent loop ends are always in the same direction. Successive sections wound in the same direction would produce dead paths where the sections are joined.

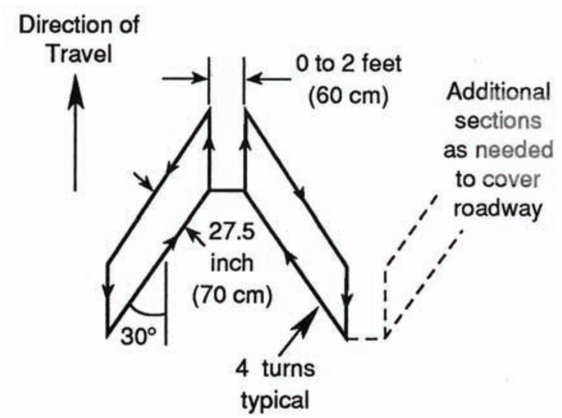


Figure 4-26. Chevron-loop configuration.

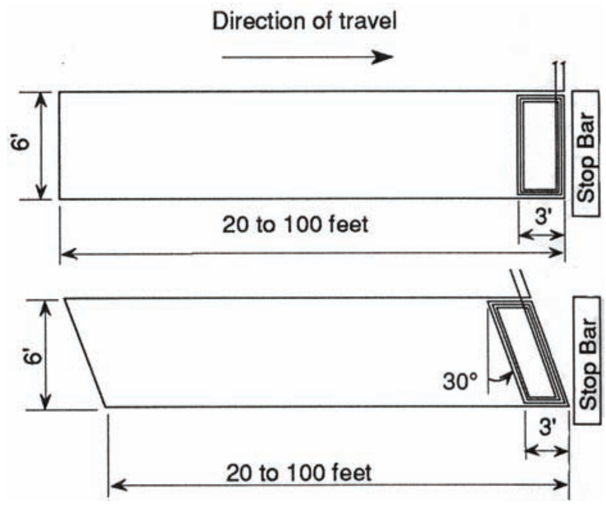
Long Loop with Powerhead

Frequently, a small motorcycle or bicycle will not produce a sufficient shift of inductance in a one-turn 6 x 20 ft (1.8 x 6.1 m) or longer loop. Some inductive-loop electronics units will detect these small vehicles with two or more turns of wire although there are dead areas in the center of the loop.

One approach is to use a small powerhead at the stopline with the long loop. This configuration is shown in the upper portion of Figure 4-27. The standard powerhead can be improved by angling the transverse wires as shown in the lower portion of Figure 4-27. The angling will cause the small vehicle to cut the lines of flux more efficiently, thereby increasing the signal by as much as 25 percent. The disadvantage is that the vehicle may not stop on the powerhead unless the detection area is clearly indicated by paint or signs.

Some traffic engineers have concerns about liability because small vehicles are not detected throughout the detection zone. Although these vehicles are sensed at the stopline, the controller operation is based on the detection of vehicles in the approach. Other engineers feel that the important consideration is to detect the small vehicles to ensure they receive the green signal, not to ensure that the signal operation is optimized for these vehicles. They do not believe that any liability is incurred.

Use of sequential short loops rather than a single long loop avoids nondetection of small vehicles anywhere in the detection zone. Also, modern electronics units sense small vehicles without the resultant splashover that occurs when using the single long loop with high-sensitivity settings. These two factors have reduced the need for the powerhead design.

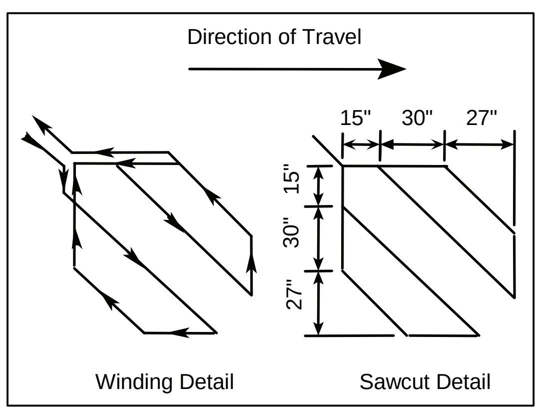


1 ft =0.3 m

Figure 4-27. Long loop with powerhead.

Inductive-Loop Bicycle Detectors

Caltrans developed a Type D loop configuration to better detect bicycles.(16) This configuration, shown in Figure 4-28, is a palm-shaped loop that fits into a 6-ft (1.8-m) square. The loop has three turns of wire when a single Type D loop is connected to an electronics unit channel, and five turns if one Type D loop is connected in series with three 6- x 6-ft (1.8- x 1.8-m) loops on an electronics unit channel (see Type 5DA or 5DQ installations in Figure 4-12). This configuration requires that the wire be protected where it bends around the acute corners of the cut. Drilling a hole in the corner or chipping out the inner angle to provide a radius will prevent kinking of the wire. This loop may be used in either vehicular or bicycle lanes.



1 inch = 2.5 cm

Figure 4-28. Type D loop configuration (Caltrans).

Figure 4-29. Type D loop installation (Photograph courtesy Darcy Bullock, School of Civil Engineering, Purdue University, West Lafayette, IN).

Figure 4-29 shows a Type D loop installation for detecting bicycles on a sidewalk or bicycle path. The wheel on the pole is used to push the wire into the sawcut and then to push foam backer firmly on top of the wire. There is debate about placing foam over the entire wire or simply using short foam pieces every foot or so. The first method supports easier installation from the contractor's perspective; the second provides better encapsulation by the sealant and less wire movement.

In some applications, it is desirable to detect the presence of a bicycle across a greater portion of a full-width traffic lane. Figure 4-30 shows one loop configuration suitable for this purpose, an 8-ft (2.4-m) square with three diagonal saw cuts traversing the square. Two layers of wire are used and are wound so that the current flow is in the same direction for both layers. This results in four layers of wire in each diagonal. The acute-angle corners are rounded to prevent damage to the wires. This is a special configuration that is installed in areas of heavy bicycle traffic such as near the University of California at Davis.

Figure 4-31 illustrates another method of reliable bicycle detection with two quadrupole loops placed side by side within the traveled area of a bike lane. An adjustable timed-call extension is generated to hold the call to the controller long enough for the bicycle to clear the intersection when operating in a loop occupancy mode. The 4.5- by 6-ft (1.4- by 1.8-m) configuration provides assurance of detecting all bicycles and complete adjacent lane rejection.

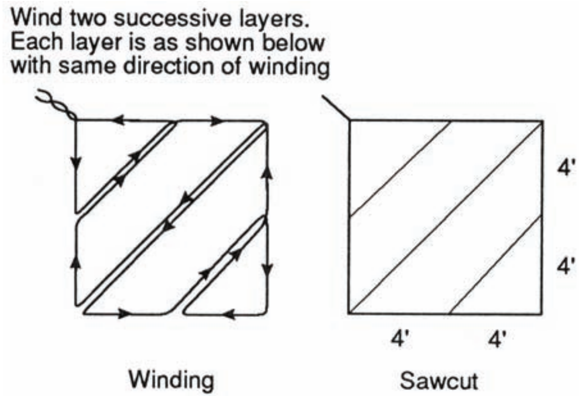
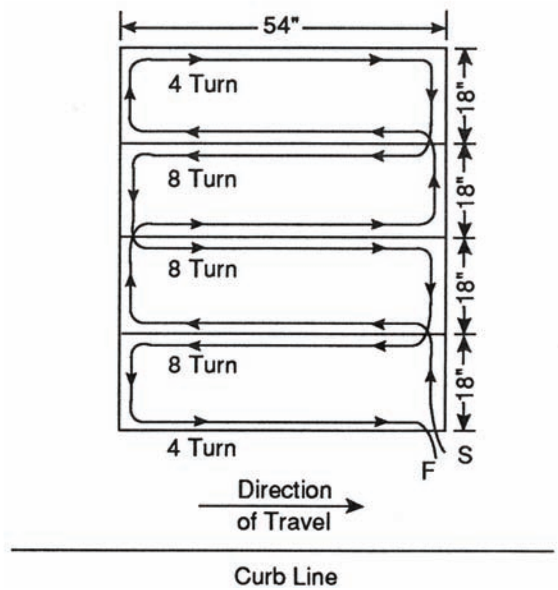


Figure 4-30. Wide coverage bicycle loop.



1 ft =0.3 m

1 inch = 2.5 cm

Figure 4-31. Bicycle lane loop layout with side-by-side quadrupoles.

DETECTION OF LONG, HIGH-BED VEHICLES

It is generally advisable to allow long, high bed vehicles (e.g., tandem trucks, semitrailer trucks, and cars pulling trailers) to travel through the intersection without stopping. There are three good reasons why long vehicles should not be stopped:

- Jackknifing of truck-trailers tends to occur under heavy braking conditions.
- After stopping, a large vehicle requires a longer startup time delaying following traffic.
- Increased noise and air pollution are associated with heavy truck startup.

A detection alternative for trucks consists of two loops spaced 30 ft (9 m) apart and located 302 ft (91 m) from the stopline.[17] This is the distance a loaded semitrailer vehicle traveling 45 mi/h (72 km/h) needs to stop safely. The sensor logic requires that the second loop be activated before the first loop is vacated.

The Detroit, MI, freeway program required that all vehicles be detected as a single entity, including high-bed trucks, semitrailers, and tanker trucks, as well as conventional vehicles. Their test of three-turn 6- x 6-ft (1.8- x 1.8-m) loops did not adequately satisfy their requirement. After numerous trials, the agency determined that a seven-turn 6- x 6-ft (1.8- x 1.8-m) loop rotated 45 degrees to form a diamond shape was superior particularly in its accuracy in detecting trucks. The diamond shape was further refined by carefully adjusting the angles of the diamond to avoid splashover.[25]

One manufacturer states that reliable truck detection from loop configuration change is not due to the diamond shape of the loop as suggested above. Rather it is due to the increased number of turns, which increases the inductance of the loop. The amount of signal received by the electronics unit is dependent on the loop inductance to lead-in cable inductance ratio. When the loop inductance to lead-in cable inductance ratio equals one, then the amount of change seen by the electronics units is one-half the change occurring in the loop. By increasing the loop inductance (by increasing the number of turns to seven), Detroit has greatly increased the amount of change received at the electronics unit, thereby resulting in more reliable detection of high-bed trucks.

QUEUE DETECTION

Freeways that operate under congested conditions are periodically likely to experience heavy volumes on off-ramps. When these off-ramps terminate at a signalized intersection, backups can extend to the freeway lanes, causing even more congestion. Depending upon agency policy, it may be desirable to detect and discharge such queues before they become a freeway problem.

One solution uses an actuated controller with a queue-discharge system consisting of a queue sensor with a preset delay time.[13] As shown in Figure 4-32, the queue-detection loop is located at a strategic position at the upstream end of the off-ramp. A timer starts when a vehicle enters the detection zone of the queue sensor and resets to zero when the vehicle exits the zone of detection. If the system counts a predetermined number of seconds, i.e., the preset delay time, the electronics unit's normal output relay is energized.

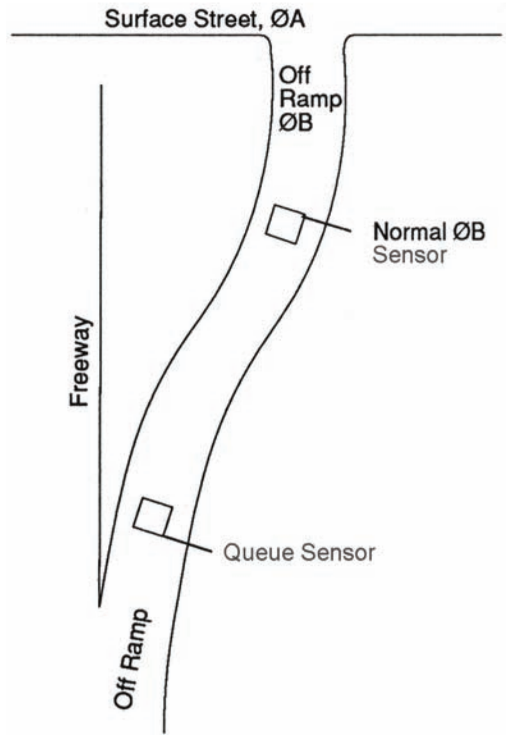


Figure 4-32. Queue discharge system.

If the queue of vehicles waiting at the red signal stopline extends upstream to the queue sensor, a vehicle will be located over the loop longer than the selected delay time. When the delay timer reaches the preset time, the sensor logic issues a signal to discharge the queue. The green signal remains on until all vehicles are moving with gaps longer than the loop itself.

The inductive loop must be long enough to span the distance between standing vehicles. Concomitantly, it must be shorter than the shortest gap in moving traffic so that the breaks between moving vehicles will cause the delay timer to reset. This latter consideration can be critical when a inductive-loop queue-sensor covers two or more lanes. A loop length of 30 ft (9 m) will generally satisfy this criterion.

INDUCTIVE-LOOP DETECTOR DESIGN FOR TRAFFIC SIGNAL CONTROL SYSTEMS

The sensor requirements for area-wide traffic signal control systems are dependent on the type of control implemented. TOD control does not require detectorization as it operates as a time clock. First-generation traffic responsive control and other advanced control strategies require vehicle sensors capable of early identification of traffic trends within a system. The sensors must provide an early indication of a peak period for the beginning of heavy traffic. Thus, the sensors must be placed on heavily traveled links with traffic patterns representative of the significant flows within a section of the area-wide system.

The application of sensors to various arterial and freeway traffic management techniques was explored in chapter 3. Additional sensor design considerations for traffic-responsive traffic control systems are discussed below.

Accuracy Goals

When sensor systems are used in computerized traffic-control systems, the particular traffic management application determines the traffic flow parameters to be collected and their accuracy. As discussed in chapter 3, volume and occupancy can be measured effectively. Other measured parameters may only provide approximations to the required variables, depending on how the data are gathered and subsequently processed. In estimating link-specific volumes, three components of error limit the measurement accuracy of each of the control variables. These are:

- A measurement error in the data on which the predictor operates.
- A prediction error in estimating the underlying mean.
- A component reflecting the randomness of traffic.

The data error can be expressed as *X* percent probability that the error is within *Y* percent. A normal distribution is assumed for the mean value of a large sample. The count error for first generation UTCS critical intersection control was plus or minus 3 vehicles 90 percent of the time.[18]

Several filtering and smoothing techniques may be applied to the data to calculate the values needed to produce measures of effectiveness that relate to traffic flow. A filtering equation determines the difference between the previous time period smoothed value (such as volume) and the latest unsmoothed value. The difference is used to update the value of the smoothed data.

Another error is introduced for vehicle presence because the controller does not generally observe vehicle presence continuously, but instead samples. This results in an error in occupancy and speed computations. There will always be a sampling error that increases as vehicle speed increases. This error cannot be eliminated but can be decreased by increasing the sampling rate. The *Computerized Signal Systems Student Workbook*[19] describes how to compute the percent sampling rate and how to implement filtering and smoothing techniques.

Link Selection for Critical Intersection Control (CIC)

The location of sensors in a traffic-control system is a three-step process. First, links are selected for sensor installation. The lateral and longitudinal sensor placements are then determined. The link selection must consider each function in the sequence, beginning with intersections that are candidates for CIC. Candidate intersections are those that could take advantage of variable split, but which operate in an unsaturated condition. When installing sensors in an intersection to operate under CIC, all approaches served by phases with variable green times must be detectorized. The step-by-step process for determining CIC candidates is found in the *Locating Detectors for Advanced Traffic Control Strategies Handbook*.[20]

The need for measuring general traffic trends will probably be satisfied by the CIC sensors in area-wide surveillance systems. If this type of detectorization is not installed, sensors should be located on major collectors or distributors to the network to obtain system trends with a minimum of instrumentation. Another group of candidates for this type of detectorization are entrances and exits of large parking facilities that would have significant effects on local traffic conditions.

The location of sensors to evaluate system operation is, to a large extent, dependent upon the degree of accuracy required by the evaluation. If a general evaluation of changes in system operations is desired, the sensor placement for traffic responsive operation would be adequate. If a more detailed evaluation is required such that speed and delays can be determined, it is necessary to increase the number of sensors within the system. The cost of this amount of detectorization may suggest that other techniques be employed.

Lane Selection

Observations have shown that a single vehicle sensor in the lane carrying the maximum through volume will be the most representative of the traffic to which the signal must respond. Moreover, the signal should be timed for critical lane volumes. However, it is not possible to derive reliable total volumes from a single sensor. Yet multiple inductive-loop detectors located in noncritical lanes may introduce errors that exceed the value of the data they provide as they can also measure parking vehicles, turning vehicles, etc., unless individual lane detection is used. Thus, the most reliable data are derived from the critical lane. The critical lane (defined as the one carrying the greatest volume) is usually easy to identify by observing the lengths of queues at the intersection.

Sensors should be installed in multiple lanes at locations where the critical lane changes with time of day, as should the corresponding TOD factors used in the software to select the inductive loop currently measuring the volume in the critical lane.

Traffic volumes should be measured on each lane of an intersection approach.[21] Average lane volumes per cycle need to be computed and compared with the tentative critical lane. Four conditions require engineering judgment pertaining to project priorities and knowledge of individual link traffic. These conditions are:

- Approaches where one lane is always critical.
- Approaches where the critical lane shifts between two lanes, but the difference in volume is not great or the shift is infrequent.
- Approaches where shifts in critical lanes are significant.
- Approaches exhibiting specific critical lanes during peak hours, but, for various reasons, erratic shifts in critical lanes occur during nonpeak periods.

It is unusual for more than two lanes to require detectorization, and most often, only one lane will need a sensor. After the critical lanes have been identified for the links to be detectorized, it is necessary to determine the longitudinal placement of the sensors on the links.

Longitudinal Placement

There are two guidelines for the longitudinal placement of sensors for signal-system control. One relates to the upstream intersection, and the other refers to the downstream intersection. From the upstream intersection, a sensor should be downstream from the zone of acceleration of vehicles entering the link. A distance of approximately 230 ft (69 m) is recommended.

From the downstream intersection, the sensor should be upstream from the point beyond which standing queues of vehicles do not usually extend. Although this distance is a function of the cycle length, split, and offset, it is recommended that values of 200 to 250 ft (60 to 75 m) be used in urban grid areas and values of 300 to 350 ft (90 to 105 m) be used in suburban arterial systems. When both criteria cannot be met, the criterion based on typical queue size is considered the more critical.

An additional issue of longitudinal-sensor placement involves the location of traffic sinks and sources (e.g., parking facilities). Sensor placement research has shown that a sink or source has minimal effect on traffic measured in the critical lane when the facility is operated as a sink such as a parking garage during the morning peak. Turns into the garage are made from the curb lane, which is not usually the critical lane.

However, during the evening peak hour, when the garage functions as a source, there are measurable effects on the critical lane. An evaluation of the paths of vehicles entering the lane from a source showed that most vehicles wait for a sufficient gap to enter the specific lane within the link. It is suggested that a critical lane sensor be located at least 50 ft (15 m) downstream from the source, provided that the downstream intersection criterion is not violated. In general, unless the source contributes more than 40 vehicles per hour to the critical lane, the effects of a source on the link demand are not significant.

Summary of System Sensor Location Considerations

Sensor location information developed in earlier tasks assists in selecting links to be detectorized, lanes in which the sensor should be placed, and an approximate sensor location with respect to the upstream and downstream intersections. With this information indicated on a map, a field visit should be conducted for each link. A walk-through by the designer will permit the selection of final locations with consideration given to access to the control equipment, special driveway problems, or other roadway or parking conditions. Each sensor location must pass a reasonableness test as well as the analytical test.

Several general guidelines are suggested concerning the field location of individual inductive loops:

- A loop should be located in the center of the traffic flow, not necessarily in the center of the marked lane. The center of the traffic flow can usually be identified by the oil markings or tire tracks on the pavement.
- The loop should be located in the areas of stable traffic flow. Sections of a link with excessive weaving or heavily impacted by entering and exiting driveways should be avoided.

- Where a major driveway is located within a link, the loop should be located at least 50 ft (15 m) downstream from the driveway, provided that the loop is at least 200 ft (60 m) upstream of the stopline.
- Inductive loops should not be located within 10 ft (3 m) of any manhole, water valve, or other appurtenance located within the roadway. This distance is required to permit sufficient clearance for work on the manhole without disturbing the loop.

In summary, the final decision concerning the location of sensors for advanced traffic-control strategies is a blend of analytical procedures coupled with engineering judgment. Not all links can be instrumented to yield input measures within the accuracy required by the algorithms. Short links and links with extremely poor lane discipline are typical of those that are not compatible with accurate instrumentation.

INDUCTIVE-LOOP DETECTOR ELECTRONICS UNITS

The operational characteristics of various types of modern inductive-loop electronics units are discussed in detail in chapter 2. Requirements for inductive-loop electronics devices are included in the NEMA Standards and the Type 170 Specifications (see appendixes I and J). The NEMA Standards define a series of self-contained loop electronics units designed for shelf mounting. It also describes a card-type electronics unit designed to insert into a multicard housing rack. The revised NEMA Standards (TS-2) emphasize card-rack mounted units. The Type 170 Specifications define card-type modules only, which are designed for insertion into the input file of the cabinet system.

NEMA electronics units are available with one, two, or four independent inductive-loop detector channels per unit. Some agencies insist on using only single channel electronics units because they believe that failures in the unit can be more easily and inexpensively corrected by replacing a single channel unit rather than replacing a multiple channel unit for a single channel's failure. Current reliability of electronics units makes this argument obsolete. The rack-mounted modules used in the Type 170 system contain either two or four independent channels.

NEMA Standards also define optional timing features that allow for the delay or extension of the electronics unit output. In the delayed-call mode, the electronics unit will wait for a user-defined period after a vehicle enters the detection area before it starts the output signal. In the extended-call mode, the unit will extend the output after the vehicle leaves the detection area. The Type 170 electronics unit does not provide this capability, as extension is normally performed by software in the controller unit itself.

MAGNETOMETER CONFIGURATIONS

A magnetometer sensor system, such as the one shown earlier in Figure 2-39, consists of one or more magnetometer sensors, the magnetometer electronics unit, and the lead-in cable between the sensor and the electronics unit. The sensor probe installation is illustrated in Figure 4-33.

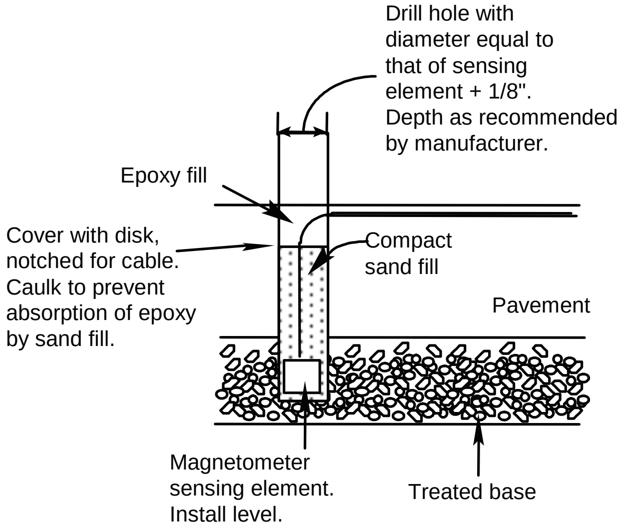


Figure 4-33. Conceptual magnetometer sensor installation.

Since the magnetometer is a passive device, there is no transmitted field or transmitted detection area. Modern magnetometer sensors detect changes in the horizontal and vertical components of the Earth's magnetic field caused by the passage or presence of a vehicle and, hence, can be used for either passage or presence detection as discussed in chapter 3.

Because the probes are buried in a drilled hole approximately 18 inches (0.5 m) below the surface, they find application in the northeastern United States, where the pavement deteriorates more rapidly due to thermal expansion and contraction and suffers damage from snow-removal equipment. They are also used in areas where loops cannot be cut in the road surface, such as on steel bridge decks.

The configuration for a magnetometer installation is dependent on a number of factors that must often be traded against one another when designing an optimum configuration for a specific location. The following section discusses these factors and their impact on the ultimate design.

SITE SELECTION

Magnetometers will detect a vehicle whenever a sufficient portion of its magnetic shadow (i.e., changes in the horizontal and vertical components of the Earth's magnetic field caused by the passage or presence of a vehicle) falls on a sensor probe. The dimensions of the magnetic shadow generally approximate the geometric dimension of the vehicle. In some cases, the magnetic shadow may be offset from the probe by a few feet in any direction.

SENSITIVITY

Electronics units for magnetometers provide two or more independent detection channels. Each channel supports 6 to 12 sensor probes, depending on the model. If only one probe per channel is installed in the street, the entire channel sensitivity is available to that single probe. However, as the number of probes per channel increases, the sensitivity is divided among the probes, thus decreasing the sensitivity of each individual probe. For example, when four probes are connected to a single channel, the total channel sensitivity remains the same, but the sensitivity at each probe will be reduced to one-fourth of the total channel sensitivity.

Therefore, it is necessary to determine what types of vehicles are to be detected in order to select the proper spacing between probes and to define the number of probes connected together per channel. The number of probes required per lane and their optimum cross-lane position are determined by the lane width and the size of the vehicles to be detected. Some portion of the vehicle must pass over a sensor probe to be detected. Some general guidelines for the number of probes for a given type of vehicle and the number of probes per channel are given below:

- **Auto, Trucks, and Buses:** Install probes at 5-ft (1.5-m) intervals. Six probes per channel, maximum.
- **Motorcycles (300 cubic centimeters (cc) and larger)** (300 cc = 18.31 cubic inches))**:** Install probes at 4-ft (1.2-m) intervals. Four probes per channel, maximum.
- **Motor Bikes (70–300 cc)** (4.27–18.31 cubic inches)**:** Install probes at 3-ft (0.9-m) intervals. Three probes per channel, maximum.
- **Bicycles:** Install probes at 3-ft (0.9-m) intervals. Two or three probes per channel.

Installation sites should be chosen to avoid conditions that adversely affect operation such as those adjacent to manholes or large pipes; near very high current transmission lines, trolley lines, or underground power lines; or within tunnels or other iron-structure enclosures.

Magnetometers are frequently the sensor of choice on bridge decks. Figure 4-34 illustrates a typical bridge deck installation. The presence of the steel deck over or under the probe has little effect on system performance. However, vertical structural steel members may detract from performance by reducing the intensity of the adjacent ambient magnetic field. As with loops, the most appropriate location for the probes is at the maximum distance from the steel supports or columns as shown in the illustration.

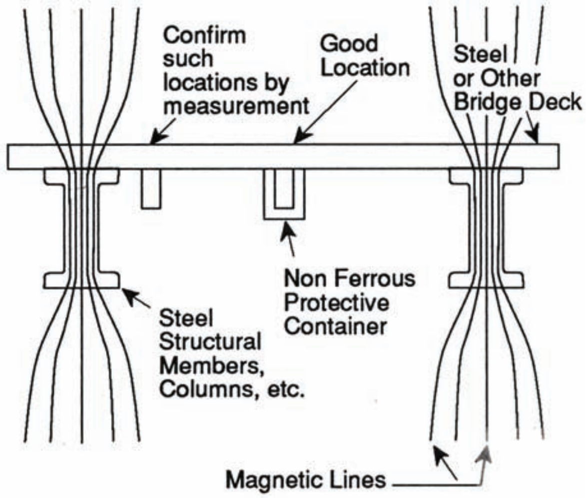


Figure 4-34. Typical installation of magnetometers on a bridge deck.

A magnetic field analyzer, such as the model shown in Figure 4-35, should be used to measure the intensity of the magnetic environment at the selected location. This instrument measures geomagnetic field intensity, magnetic noise, and ac magnetic field strength. It is especially useful at locations where the use of magnetometers is questionable (e.g., within tunnels, near large electrical devices, etc).

Situations exist where manmade magnetic noise is of sufficient intensity to impair magnetometer performance. These occur at sites where nonvehicular induced magnetic field perturbations exceed 5 milliOersteds such as near streetcar lines, some trolley bus lines, subway trains, or even elevators. Few sources are of sufficient intensity to affect a sensor probe located more than 30 ft (9 m) from the source. However, magnetometers that transmit data via a radio frequency link to the electronics unit in the controller cabinet are subject to interference from communications devices operating at the same frequency. Therefore, a frequency survey should be performed to determine a clear channel on which to transmit these magnetometer data.

PROBE PLACEMENT

The optimum lateral placement of probes in a lane is determined by such factors as the width of the narrowest vehicle to be detected, the lane width, and the detection quality required. Some part of the magnetic shadow of the vehicle must fall on one or more of the probes for detection to occur. Consequently, the maximum probe spacing is equal to the width of the narrowest vehicle to be detected. Since most autos, trucks, and buses are wider than 5 ft (1.5 m), a single probe centered in a 10-ft (3-m) lane provides adequate performance. In a 12-ft (3.7-m) lane, a single probe may fail to detect some small vehicles traveling on the edge of the lane.

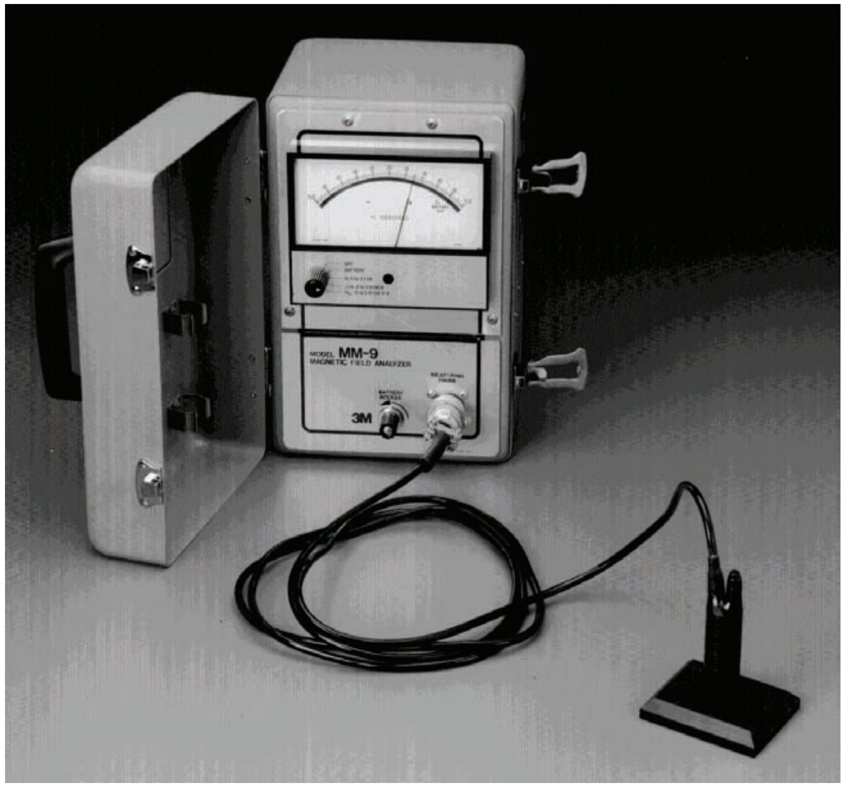


Figure 4-35. Magnetic field analyzer.

By increasing the number of probes to two per lane, virtually all four-wheeled vehicles within a 12-ft (3.7-m) lane would be detected. In this type of configuration, probes should be placed no further than 5 ft (1.5 m) from each lane boundary. If bicycles are to be detected, more probes per lane may be needed.

In general, the quality of detection improves as the number of probes per lane is increased because of the spatial averaging, which results when the magnetic shadow of a vehicle falls on several probes judiciously spaced. Reduction in field intensity at one lateral position may be compensated for by increases at other positions.

PROBE DEPTH

Vertical placement of probes is an important determinant of system performance properties. Deep placement such as 18–24 inches (45–60 cm) provides good single-count vehicle-presence detection, but results in a lower signal level. Conversely, shallow placement (e.g., 6 inches (15 cm)), provides higher signal levels, but with an increased incidence of multiple counts per vehicle. Multiple counts with shallow placement result primarily from the passage of major components of the vehicle such as engine, transmission, and differential, each of which may produce a separate magnetic perturbation. By increasing the depth to 18–24 inches (45–60 cm), most vehicles yield a single perturbation of the magnetic field because the deeply buried probes sense the overall magnetic bulk of the vehicle rather than details of the understructure.

In summary, for a detection application where only passage information is required and where multiple counts are not detrimental, probes should be located near the surface. Similarly, in constant speed applications where the time-extension feature of the extendable-presence mode can be used effectively, probes may be located near the surface. Small, two-wheeled vehicles such as bicycles or motorbikes develop narrow, low-intensity shadows. Their detection requires shallow placement of probes, as well.

SUGGESTED CONFIGURATIONS

Different probe placements and connections to the electronics unit are used to support various traffic management functions and data acquisition needs as depicted in Figure 4-36. Several typical probe configurations are presented in Appendix L. These illustrate the tradeoffs made in probe depth and lateral distance between probes. Configurations are included for single-lane, two-lane, and three-lane detection; wide-lane detection; and two-wheeled vehicle detection. Also shown are several suggested configurations for left-turn lanes with detection zones ranging from 30 to 70 ft (9 to 21 m).

WIRE SIZE AND CABLE SELECTION

Cable length and number of probes per channel determine minimum wire size. The probe excitation circuit provides current at a constant 125 mA peak-to-peak, but is limited to a 15-volt peak-to-peak swing. Maximum allowable probe-cable assembly resistance is therefore 125 ohms. At resistance levels below this limit, system operation is normal. At higher resistance levels, instability and performance degradation will occur.

Effective resistance of each probe is approximately 4 ohms. Probes on a channel are connected in series electronically, so their resistance is additive. Allowable cable resistance is 120 ohms, less the sum of the probe resistances. AWG #22 copper wire pairs have a loop resistance (going and returning total) of 35 ohms per l,000 ft (300 m), AWG #20 pairs have 20 ohms, and AWG #18 pairs have 13 ohms. Because of the changes in wire resistance with temperature, it is recommended that #18 AWG wire be used on cable runs exceeding 2,000 ft (600 m) in climates where extreme temperature ranges are typical.

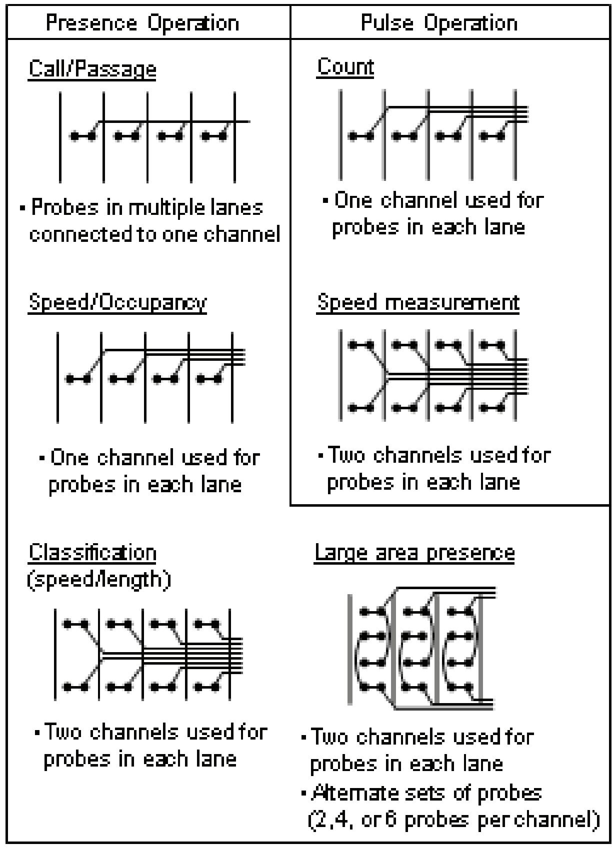


Figure 4-36. Magnetometer-probe placement in support of several data-gathering functions.

Connection of the magnetometer electronics unit and probes requires two pair of conductors per channel. One pair supplies the probe excitation current and the other carries the return signal. Properties of the cable become especially significant in long transmission distances of over 2,000 ft (600 m). The cables should combine distance capability with high noise immunity and environmental tolerance.

Although up to 12 probes per channel can be installed, the allowable lead-in length decreases as the number of probes increases. Generally, the lead-in length should not exceed 4,000 ft (1,200 m) for a 12-probe-per-channel arrangement. If 6 probes per channel are used, the lead-in length should not exceed 5,000 ft (1,500 m).

MAGNETIC-DETECTOR CONFIGURATION

Magnetic detectors first developed in the 1930s, i.e., sensors whose operation is based on an induction magnetometer (also referred to as a search coil magnetometer), are still in use today, particularly where deteriorated pavement or frost activity contributes to the failure of inductive-loop detector wires. Magnetic detectors are also applied where vehicle detection is required without placing sawcuts in the pavement. While magnetic detectors are inexpensive, reliable, and simple, most generate only pulse outputs, which are suitable in support of traffic actuated signal control and traffic volume counting. The exception is the Model 702 microloop probe from 3M, which can be connected in rows of three to detect stopped vehicles with application-specific software purchased from the microloop manufacturer.

The magnetic detector contains a coil of wire wound around a highly permeable core, which is placed below the surface of a roadway. When a ferrous metal vehicle comes near or passes over the coil, the lines of flux produced by the Earth's magnetic field that pass through the coil are perturbed by the vehicle, thus inducing a voltage in the coil. A high-gain amplifier boosts this voltage to operate a relay, which sends a signal to the controller that a vehicle has been detected.[22]

The following discussion is limited to the induction magnetometer or magnetic detector. The model shown in Figure 2-42 is cylindrical in shape, 2.25 inches (57 mm) in diameter and 21 inches (533 mm) long.

MAGNETIC-DETECTOR PROBE PLACEMENT

Probe placement is critical for the proper operation of this sensor system. For the probe to detect a change in the Earth's magnetic field, a vehicle must be moving at a speed greater than approximately 5 mi/h (8 km/h). This implies that the probe must be placed far enough back from the stopline where vehicles are normally in motion (generally at least 50 ft (15 m)). It must also be placed in the most appropriate location across the lane to generate a sufficiently strong signal to register a vehicle detection.

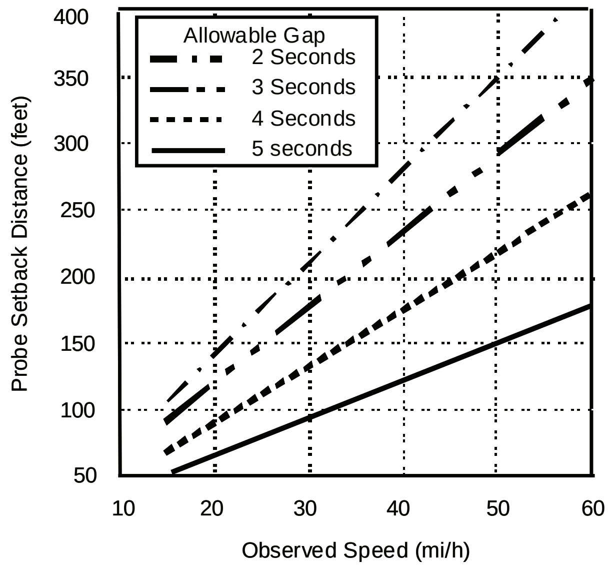
Distance from the Stopline

The setback distance of the probe from the stopline is based on the desirable allowable gap. Assuming an allowable gap (or equivalently, a passage time) of 3 seconds, the probe would be located 132 ft (40 m) from the stopline on a street with vehicles traveling at 30 mi/h (48 km/h). The graphs shown in Figure 4-37 identify the proper location of the probe relative to the stopline based on allowable gaps of 2, 3, 4, or 5 seconds.

Lateral Placement of Probe

In a single-lane approach to a stopline, the optimal location of the probe is under the path normally followed by the right wheels of the vehicle, as shown in Figure 4-38. On a two-lane approach where a single probe is used, the probe is placed between the lanes to provide satisfactory coverage. If the right lane of a multilane approach is designated for right-turning traffic, the probe is located in the middle of the through lane to minimize the effect of vehicles in the right-turn lane.

A common practice for multilane approaches is to place a probe in each lane. In such designs, the probe is placed under the right wheel track in each lane. Up to three probes can be placed in a conduit. When magnetic probes are placed in a new roadway, multiple conduits should be used, one probe per conduit. In an existing road, conduit runs should be kept to a minimum; thus, more than one probe may be placed in a conduit.



1 mi/h = 1.6 km/h

Figure 4-37. Setback of magnetic probe from stopline as a function of vehicle speed and allowable gap.

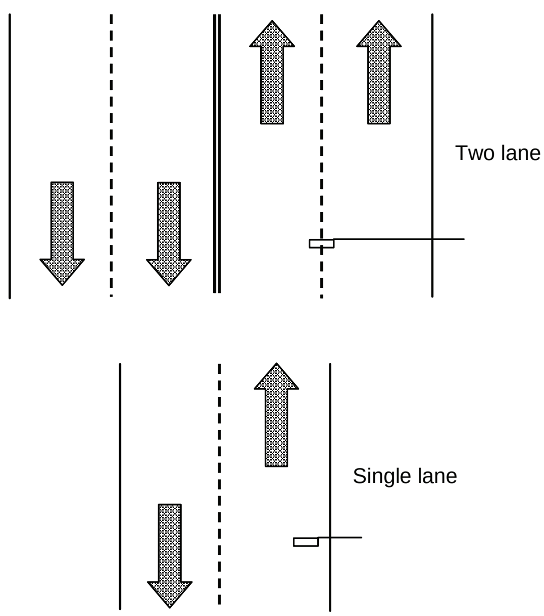


Figure 4-38. Magnetic probe placement for a single-lane and two-lane approach to a stopline.

SENSITIVITY

Two or more probes used on different approaches of the same traffic phase should be placed in similar locations (if possible) so that the area monitored by each probe is the same. This permits the sensitivity of all sensors to be adjusted by the single sensitivity-adjusting knob.

A sensor probe generates an impulse in response to any change in the magnetic conditions surrounding it. The impulse strength is proportional to the change in the magnetic field. Changes in the magnetic field can be caused by movement of iron or steel objects in the vicinity of the probe or by changes in the current flowing through power lines. With the probe placed as close as possible to the path of the vehicles, the strength of the impulses caused by these vehicles will be maximized. Impulses caused by traffic moving in other lanes or by current changes in nearby power lines will be minimized. The sensitivity of the sensor can be decreased until the unwanted impulses are reduced in strength so that they do not actuate the relay, while the impulses from vehicles in the proper lanes remain strong enough to cause actuation.

REFERENCES

1. Kell, J.H., and I.J. Fullerton. *Manual of Traffic Signal Design*, Institute of Transportation Engineers. Prentice-Hall, Englewood Cliffs, NJ. 1982.
2. *NEMA Standards Publication No. TS-1: Traffic Control Systems.* National Electrical Manufacturers Association, Washington, DC. 1989.
3. Zegeer, C.V. *Effectiveness of Green Extension Systems at High Speed Intersections*, Research Report 472. Kentucky Department of Transportation, Bureau of Highways, Division of Research, Lexington, KY. 1977.
4. "Large Area Detection at Intersection Approaches," Report of Technical Committee No. 17, *Traffic Engineering.* Southern Section ITE. 1976.
5. Courage, K.G., et al. *Inductive Loop Detector Configuration Study—Final Report*, Transportation Research Center, prepared by University of Florida. Florida Department of Transportation, Tallahassee, FL. 1985.
6. *Traffic Signal Detector Loop Location Design Installation.* Illinois Department of Transportation, District 6, Springfield, IL. 1988.
7. Woods, D.L. *Texas Traffic Signal Detector Manual*, Report No. FHWA/TX-90/1163-3F. Texas Transportation Institute, College Station, TX. 1992.
8. Lin, F.-B. "Optimal Timing Settings and Detector Lengths of Presence Mode Full-Actuated Control," *Traffic Control Devices and Rail-Highway Crossings*, Transportation Research Record 1010. Transportation Research Board, National Research Council, Washington, DC. 1985.
9. Sackman, H., et al. *Vehicle Detector Placement for High-Speed, Isolated Traffic-Actuated Control, Volume 2: Manual of Theory and Practice*, FHWA –RD-77-32. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 1977.
10. Wu, C.-S., et al. *Detector Configuration and Location at Signalized Intersections*, Center for Transportation Research, Research Report 259-1F, prepared by University of Texas. Texas State Department of Highways and Public Transportation, Austin, TX. 1983.

11. Beierle, H. "A Method of Detector Placement," *IMSA Signal Magazine*. Texas Highway Department District No. 15. June 1974.
12. "Large-Area Detection at Intersection Approaches," a section technical report, Technical Committee 17, *Traffic Engineering*. Southern Section ITE. 1976.
13. "Small Area Detection at Intersection Approaches," a section technical report, Technical Committee 18, *Traffic Engineering*. Southern Section ITE. 1974.
14. *Traffic Signal Design Handbook*, Pub. 149. Commonwealth of Pennsylvania, Department of Transportation, Bureau of Maintenance and Operations, Traffic Engineering and Operations Division, Harrisburg, PA. 1983.
15. *Applications Note 3217*, Traffic Control Devices Department. 3M, Canoga Controls, Traffic Control Products, St. Paul, MN. No date.
16. *Standard Plans*. California Department of Transportation, Sacramento, CA. 1988.
17. Blikken, W.A. *The Adjustable Diamond Loop for Vehicle Detection*. Michigan Department of Transportation, Detroit Freeway Operations Unit, Detroit, MI. 1979.
18. Capelle, D.G. *Freeway Traffic Management*. Transportation Research Board, National Research Council, Washington, DC. 1979.
19. *Computerized Signal Systems, Student Workbook*. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 1979.
20. Kay, J.L., et al. *Locating Detectors for Advanced Traffic Control Strategies Handbook*. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 1975.
21. *Traffic Control Systems Handbook*, Federal Highway Administration Implementation Package FHWA-IP-85-11. U.S. Department of Transportation, Federal Highway Administration, Washington, DC. 1985. Also published as ITE Publication No. LP-123. Institute of Transportation Engineers, Washington, DC. 1985.
22. Klein, L.A. *Sensor Technologies and Data Requirements for ITS*. Artech House, Norwood, MA. 2001.