Transport Engineering

Makenzie Mccutcheon

First Edition, 2012

ISBN 978-81-323-4408-7

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Introduction

The engineering of this roundabout in Bristol, England, attempts to make traffic flow free-moving

Transportation engineering is the application of technology and scientific principles to the planning, functional design, operation and management of facilities for any mode of transportation in order to provide for the safe, rapid, comfortable, convenient, economical, and environmentally compatible movement of people and goods (transport). It is a sub-discipline of civil engineering.Transportation engineering is a major component of the civil engineering discipline. The importance of transportation engineering within the civil engineering profession can be judged by the number of divisions in ASCE (American Society of Civil Engineers) that are directly related to transportation. There are six such divisions (Aerospace; Air Transportation; Highway; Pipeline; Waterway, Port, Coastal and Ocean; and Urban Transportation) representing one-third of the total 18 technical divisions within the ASCE (1987).

The planning aspects of transport engineering relate to urban planning, and involve technical forecasting decisions and political factors. Technical forecasting of passenger travel usually involves an urban transportation planning model, requiring the estimation of trip generation (how many trips for what purpose), trip distribution (destination choice, where is the traveler going), mode choice (what mode is being taken), and route assignment (which streets or routes are being used). More sophisticated forecasting can include other aspects of traveler decisions, including auto ownership, trip chaining (the decision to link individual trips together in a tour) and the choice of residential or business location (known as land use forecasting). Passenger trips are the focus of transport engineering because they often represent the peak of demand on any transportation system.

A review of descriptions of the scope of various committees indicates that while facility planning and design continue to be the core of the transportation engineering field, such areas as operations planning, logistics, network analysis, financing, and policy analysis are also important to civil engineers, particularly to those working in highway and urban transportation. The National Council of Examiners for Engineering and Surveying (NCEES) list online the safety protocols, geometric design requirements, and signal timing.

Transportation engineering, as practiced by civil engineers, primarily involves planning, design, construction, maintenance, and operation of transportation facilities. The facilities support air, highway, railroad, pipeline, water, and even space transportation. The design aspects of transport engineering include the sizing of transportation facilities (how many lanes or how much capacity the facility has), determining the materials and thickness used in pavement designing the geometry (vertical and horizontal alignment) of the roadway (or track).

Before any planning occurs the Engineer must take what is known as an inventory of the area or if it is appropriate, the previous system in place. This inventory or database must include information on (1)population, (2)land use, (3)economic activity, (4)transportation facilities and services, (5)travel patterns and volumes, (6)laws and ordinances, (7)regional financial resources, (8)community values and expectations. These inventories help the engineer create business models to complete accurate forecasts of the future conditions of the systemReview.

Operations and management involve traffic engineering, so that vehicles move smoothly on the road or track. Older techniques include signs, signals, markings, and tolling. Newer technologies involve intelligent transportation systems, including advanced traveler information systems (such as variable message signs), advanced traffic control systems (such as ramp meters), and vehicle infrastructure integration. Human factors are an aspect of transport engineering, particularly concerning driver-vehicle interface and user interface of road signs, signals, and markings.

Highway engineering

Engineers in this specialization:

- Handle the planning, design, construction, and operation of highways, roads, and other vehicular facilities as well as their related bicycle and pedestrian realms.
- Estimate the transportation needs of the public and then secure the funding for the project.
- Analyze locations of high traffic volumes and high collisions for safety and capacity.
- Use civil engineering principles to improve the transportation system.
- Utilizes the three design controls which are the drivers, the vehicles, and the roadways themselves.

Railroad engineering

Railway engineers handle the design, construction, and operation of railroads and mass transit systems that use a fixed guideway (such as light rail or even monorails). Typical tasks would include determining horizontal and vertical alignment design, station location and design, and construction cost estimating. Railroad engineers can also move into the specialized field of train dispatching which focuses on train movement control.

Railway engineers also work to build a cleaner and safer transportation network by reinvesting and revitalizing the rail system to meet future demands. In the United States, railway engineers work with elected officials in Washington, D.C. on rail transportation issues to make sure that the rail system meets the country's transportation needs.

Port and harbor engineering

Port and harbor engineers handle the design, construction, and operation of ports, harbors, canals, and other maritime facilities. This is not to be confused with marine engineering.

Airport engineering

Airport engineers design and construct airports. Airport engineers must account for the impacts and demands of aircraft in their design of airport facilities. These engineers must use the analysis of predominant wind direction to determine runway orientation, determine the size of runway border and safety areas, different wing tip to wing tip clearances for all gates and must designate the clear zones in the entire port.

Chapter 1

Bicycle Transportation Engineering

Engineering for bicycles - From no provision in hostile environments.

Engineering for bicycles - to dedicated cycle facilities.

Engineering for bicycles - to providing environments in which cyclists and others can share the road safely.

Bicycle transportation engineering is the study of transportation engineering as it affects bicycles and cycling. It includes the design of dedicated transport facilities for cyclists, but also the study of how mixed-mode environments (i.e. where cyclists are not segregated from other traffic) can be made to work safely.

Roads

Various methods of altering or reallocating the roadway right-of-way to facilitate bicycling and create bikeways have been added to many of the manuals used by transport planners and engineers.

Sidepath or Shared use path

A shared use path is a bikeway that is physically separated from motorized vehicle traffic by an open space (generally landscaping) or some form of barrier. While often perceived as safe and pleasant by cyclists, this type of facility can create additional hazards and problems of it's own, especially at driveways, side streets, and due to the need for extra land required compared to less extensive facilities. Depending on design and jurisdictional rules regarding crossing side streets, these paths may also be much slower for the rider, making them less attractive for sports or commutter cyclists.

Cycle track

Cycle tracks, also called *separated cycle lanes*, are bicycle exclusive facilities that provide physical separation from motorized vehicle traffic within the road right-of-way, generally by means of kerbs or other barriers. Cycle tracks can either incorporate bicycle-only signal phases at intersections (for 100% separation) or utilize “mixing zones” to merge bicycle and motor vehicle traffic. A cycle track combines the user experience of a separated path with the on-road infrastructure of a bike lane.

Bicycle lane

A designated bicycle lane, according to the 1998 United States Manual on Uniform Traffic Control Devices, is:

- a portion of a roadway or shoulder which has been designated for use by bicyclists
- marked by a white (usually solid) stripe painted on the pavement
- significantly narrower than traffic lanes
- found at the side of the traffic lanes

Cycle lanes require less space compared with the option of providing side paths or separate tracks, and generally do not have the legal / priority issues that facilities separated from the general traffic may have (depending on the locally applicable road laws). However, if traffic is heavy or fast, the lack of physical separation from motor vehicles may make cycle lanes unattractive to less confident cyclists. The lack of separation also means they are more apt to be blocked by (illegally) parked vehicles.

Non-segregated facilities

Wide outside lane (WOL)

Wide outside through lanes help motorists pass slower cyclists without having to decrease speed or change lanes. This method is held by some to be particularly important on roads with a high proportion of wide vehicles, such as buses or heavy trucks. These lanes also provide more room for cyclists to filter past queues of other vehicles in congested conditions. The use of such lanes is specifically endorsed by *Cycling: the way ahead for towns and cities*, the European Commission policy document on cycle promotion.

Cycle friendly infrastructure argues for a marked lane width of 4.25 m (14 ft). It is argued that, on undivided roads, this width provides cyclists with adequate clearance from passing wide vehicles while being sufficiently narrow to deter motorists from attempting to “double up” and form two lanes. This “doubling up” effect may be related to junctions. At non-junction locations, greater width might be preferable if this effect can be avoided.

Shared bus lane

A bus and cycle lane in Mannheim, Germany

A shared bus lane is a bus lane that allows cyclists to use it. Depending on the width of the lane, the speeds and number of buses, and other local factors, the safety and popularity of this arrangement vary.

Guidance produced for Cycling England endorses bus lanes as providing cyclists with a *direct and barrier free route into town centres* and as avoiding the difficulties associated with other provisions such as shared-use footways. According to a French survey 42% of cyclists described themselves as "enthusiasts" for shared bus bike lanes versus 33% who were of mixed opinion and 27% who were opposed. Many cycling activists view these as being more attractive than cycle paths, while others object to being in close proximity to bus exhausts. The sharing of bus lanes has been described as "generally very popular" with cyclists in London.

As of 2003, mixed bus/cycle lanes accounted for 118 km of the 260 km of cycling facilities in Paris. The French city of Bordeaux has 40 km of shared bus cycle lanes. It is reported that that in the city of Bristol, a showcase bus priority corridor, where road space was re-allocated along a 14 km stretch also resulted in more space for cyclists and had the effect of increasing cycling. The reverse effect has also been suggested, a review carried out in London reports that cycling levels fell across Kew bridge following the removal of a bus lane - this was despite a general increase in cycling level in the city generally. In addition, it is arguably easier, politically speaking, to argue for funding of joint facilities rather than the additional expense of both segregated cycling facilities and bus-only lanes.

In some instances. bus lane proposals have run into vehement opposition from cycling representatives - a typical theme is the perceived generation of conflict due to the narrowing of other lanes already shared by cars/cyclists so as to create space for the bus lanes The TRL reports that cyclists and bus drivers tend to have low opinions of each other There have been reports in Dublin of conflict as cyclists choose to cycle in the bus lanes and a bus driver apparently expected them to use adjacent cycle tracks instead. In some other cities the arrangements seem to work successfully, with bus companies and cyclists' groups taking active steps to ensure that understanding is improved between the two groups of road users.

Streets

Shared space

Shared space schemes, which are characterized by the removal of road markings, signs and signals, give all street users equal priority and equal responsibility for each others safety. Experiences where these schemes are in use, show that street users, particularly motorists, undirected by signs, curbs or road markings, reduce their speed and establish eye contact with other users. Results from the thousands of such implementations worldwide all show casualty reductions and most also show reduced journey times. Following the partial conversion of London's Kensington High Street to *shared space*, accidents were reduced there by 44% (the London average was 17%).

Bicycle boulevards

A bicycle boulevard is a low speed street which has been designed to discourage cut-through motor vehicle traffic and to give priority to to cyclists as through-going traffic.

Bike paths

Bike paths are bicycle routes that follow an independent right-of-way. There are two distinct types of bike paths: those used exclusively by bicycles and those shared with pedestrians. In the United States almost all bike paths are shared with pedestrians.

Bike paths are often found on old railroad right-of-ways, called rail trails.

Traffic lights

How traffic signals are designed and implemented directly impacts cyclists. For instance, poorly adjusted vehicle detector systems, used to trigger signal changes, may not correctly detect cyclists. Traffic managers in Copenhagen link cyclist-specific traffic signals on a major arterial bike lane to provide green waves for rush hour cycle-traffic. Cycling-specific measures that can be applied at traffic signals include the use of advanced stop lines and/or bypasses.

The frequency with which lights change is important to cyclists who may conserve energy by anticipating green lights ahead i.e. the shorter the interval the better for cyclists. Intersection clearance times and green wave timing may also be relevant.

Road surface

Bicycle tires being narrow, road surface is more important than for other transport, for both comfort and safety. The type and placement of storm drains, manholes, surface markings, and the general road surface quality should all be taken into account by a bicycle transportation engineer. Drain grates, for example, must not catch wheels.

Parking

Efficient two tiered Dutch Bicycle Rack

Bicycle parking is another important part of Bicycle Transportation Engineering. In most of the United States, bicycle parking facilities are scarce, or are so inadequate that nearby trees or parking meters are used. The hitching post type of bicycle stand is an improvement over the old type that had a slot for the front wheel but only allow for two bicycles per post. The Netherlands, where bicycles are much in use, has two-tiered bicycle racks giving high density (the handlebars overlap) and security (the bicycle is held well and is easy to lock).

Improved parking meter

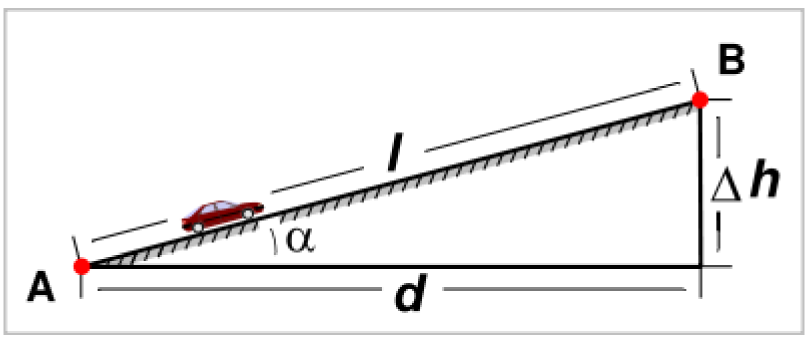
Secure bicycle parking is argued to be a key factor influencing the decision to cycle. To be considered secure, the parking must be of a suitable design: allowing the bicycle to be locked via the frame. A readily observable location can also permit so-called passive security from passers-by. Weather protection is also desirable. As a rule, where cycling is encouraged as an alternative to motoring, efforts are made to make bicycle parking more convenient and attractive to use than nearby car parking arrangements. This usually means providing a wide distribution of visible, clearly designated parking spots, close to the entrances of destinations being served.

Storage rooms or bicycle lockers may also be provided. In some cases large concentrations of bike parking may be more appropriate, sometimes being supervised and sometimes charging a fee. Examples include large bike parking facilities at public transport interchanges such as railway, subway, tram, bus stations, or at bike stations where they may be useful in Mixed-mode commuting.

Conversely, where cycling is seen as an unwelcome or inappropriate activity, or due to lack of knowledge about best practices, bicycle parking may simply not be provided or else placed at awkward, distant, and out-of-sight locations. Cyclists may be expressly forbidden from parking their bicycles at the most convenient locations. In April 2007, the authorities at the University of California's Santa Barbara campus started confiscating bicycles not parked at the allegedly inconvenient official bike racks. Often, property owners display signage on fencing to discourage bicyclists from locking their bicycles.

Chapter 2

Grade (Slope)



d = run
Δh = rise
l = slope length
α = angle of inclination

The **grade** (also called **slope**, **incline**, **gradient**, **pitch** or **rise**) of a physical feature, topographic landform or constructed element, refers to the amount of inclination of that surface to the horizontal. It is a special case of the gradient in calculus where zero indicates *gravitational level*. A larger number indicates higher or steeper degree of "tilt". Often slope is calculated as a ratio of "rise" to "run", or as a fraction ("rise over run") in which *run* is the horizontal distance and *rise* is the vertical distance.

Grade or slope is applied to measuring existing physical features (such as canyon and hillsides, stream and river banks and beds), or in designing and engineering new elements for construction (such as roads, landscape and garden grading, roof pitches, railroads, aqueducts, and pedestrian-handicapped-bicycle circulation routes).

Expression nomenclature

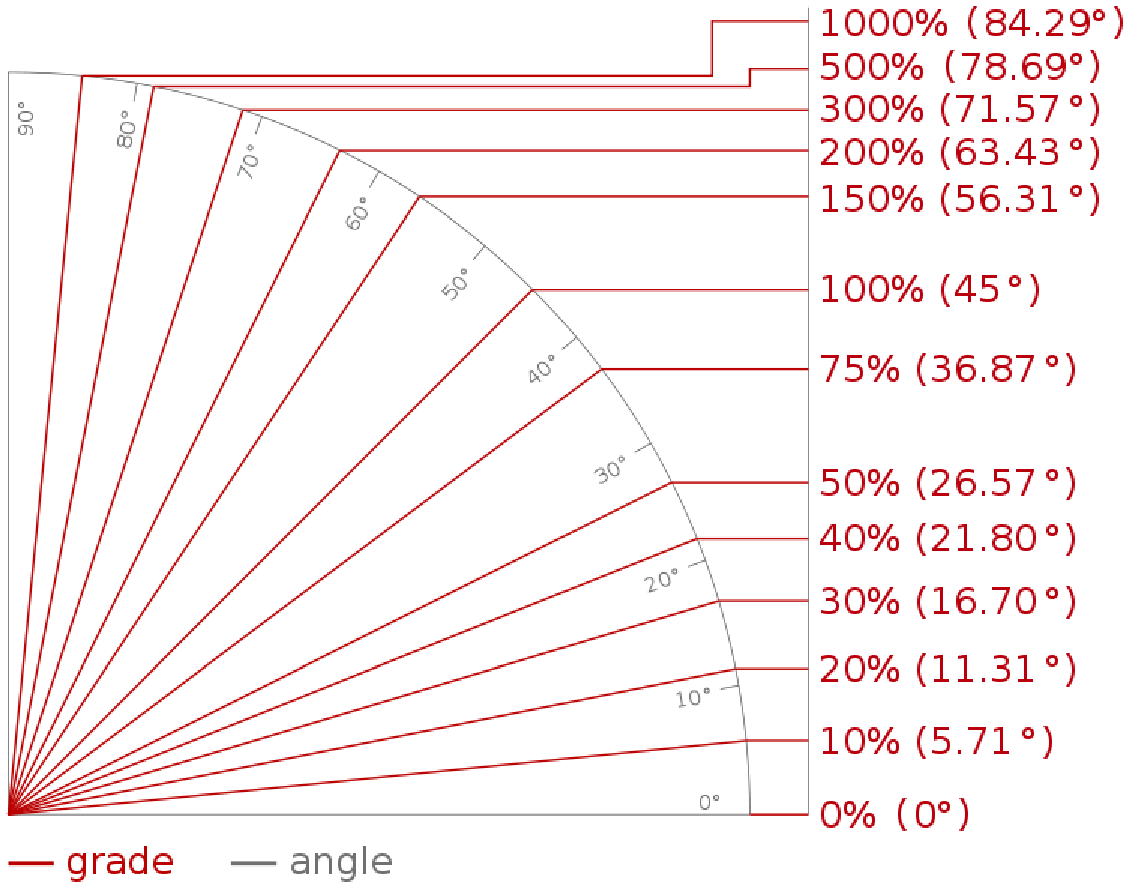


Illustration of grades in percent and angles in degrees

There are several systems for expressing slope:

1. as an *angle* of inclination to the horizontal. (This is the angle α opposite the "rise" side of a triangle with a right angle between vertical rise and horizontal run.)
2. as a *percentage*, the formula for which is $100\frac{\text{rise}}{\text{run}}$ which could also be expressed as the tangent of the angle of inclination times 100. In the U.S., this percentage "grade" is the most commonly used unit for communicating slopes in transportation, surveying, construction, and civil engineering.
3. as a *per mille* figure, the formula for which is $1000\frac{\text{rise}}{\text{run}}$ which could also be expressed as the tangent of the angle of inclination times 1000. This is commonly used in Europe to denote the incline of a railway.
4. as a *ratio* of one part rise to so many parts run. For example, a slope that has a rise of 5 feet for every 100 feet of run would have a slope ratio of 1 in 20. (The word "in" is normally used rather than the mathematical ratio notation of "1:20")

Any one of these expressions may be used interchangeably to express the characteristics of a slope. Grade is usually expressed as a percentage, but this may easily be converted to the angle α from horizontal since that carries the same information.

There is a method in which slope may be expressed when the horizontal run is not known: rise divided by the hypotenuse (the slope length). This is *not* a usual way to measure slope. This follows the sine function rather than the tangent function and this method diverges from the "rise over run" method as angles start getting larger.

Many of the mathematical principles of slope that follow from the definition are applicable in topographic practice. In the UK, for road signs, maps and construction work, the gradient was traditionally expressed as a ratio such as 1 in 12, but signs showing gradient expressed as a percentage are becoming more common.

In civil engineering applications and physical geography, the slope is calculated along a particular direction of interest which is normally the route of a highway or railway road bed.

Mathematical equations

Grades can be related using the following equations with symbols from the figure at top.

Tangent as a ratio

$$\tan \alpha = \frac{\Delta h}{d}$$

This ratio can also be expressed as a percentage by multiplying by 100.

Angle from a tangent gradient

$$\alpha = \arctan \frac{\Delta h}{d}$$

If the tangent is expressed as a percentage, the angle can be determined as:

$$\alpha = \arctan \frac{\%\,\mathrm{slope}}{100}$$

If the angle is expressed as a ratio *(1 in n)* then:

$$\alpha = \arctan \frac{1}{n}$$

Roads

In vehicular engineering, various land-based designs (cars, SUVs, trucks, trains, etc.) are rated for their ability to ascend terrain. (Trains typically rate much lower than cars.) The highest grade a vehicle can ascend while maintaining a particular speed is sometimes termed that vehicle's "gradeability" (or, less often, "grade ability"). The lateral slopes of a highway geometry are sometimes called fills or cuts where these techniques have been used to create them.

slope warning sign, Netherlands

25% slope sign, Wales

A 1371-metre long stretch of railroad with a 20‰ (2%) slope, Czech Republic

Slope warning sign, 30 % over 1500 m. La route des Crêtes, Cassis, France

Environmental design

Grade, pitch, and slope are important components in landscape design, garden design, landscape architecture, and architecture; for engineering and aesthetic design factors. Drainage, slope stability, circulation of people and vehicles, complying with building codes, and design integration are aspects of slope considerations in environmental design.

Railways

Grade indicator near Bellville, Western Cape, South Africa, showing 1:150 and 1:88 grades.

Steep gradients limit the size of load that a locomotive can haul, including the weight of the locomotive itself. A 1% gradient (1 in 100) halves the load. Early railways in the United Kingdom were laid out with very gentle gradients, such as 0.05% (1 in 2000), because the early locomotives (and their brakes) were so feeble. Steep gradients were concentrated in short sections of lines where it was convenient to employ assistant engines or cable haulage, such as from Euston to Camden Town, about 1.2 km. Extremely steep gradients need the help of cables, or some kind of rack railway.

The steepest non-rack railway lines include:

- 13.5 % - Lisbon tram, Portugal
- 11.6 % - Pöstlingbergbahn, Linz, Austria
- 11 % Cass Scenic Railway USA (former logging line)
- 9.0 % - Ligne de Saint Gervais - Vallorcine, France
- 7% - Bernina Railway, Switzerland
- 5.6% (1 in 18) - Flåm, Norway.
- 5.1% - Saluda Grade, North Carolina, United States
- 4.0% - Cologne-Frankfurt high-speed rail line
- 4.0% (1 in 25) - Tarana - Oberon branch, New South Wales, Australia.

- 4.0% (1 in 25) - Matheran Light Railway, India
- 3.7% (1 in 27) - Ecclesbourne Valley Railway, Heritage Line, Wirksworth, Derbyshire, UK

It is customary for civil engineers to refer to the steepest grade on a section of rail line as the ruling grade for that section. Civil engineering works such as cuttings, embankments and tunnels are employed to reduce this maximum grade.

Compensation for curvature

Gradients on sharp curves are effectively a bit steeper than the same gradient on straight track, so to "compensate" for this and make the ruling grade uniform throughout, the gradient on those sharp curves should be reduced slightly.

Effects of grade

The greater a grade, the more power an animal or a machine requires to climb it; therefore routes with lower grades are preferred, so long as they do not have other disadvantages, such as causing the overall travel distance to increase significantly.

Vehicles proceeding up a steep grade demand more fuel consumption with typically increased air pollution generation. Sound level increases are also produced by motor vehicles travelling upgrade.

Chapter 3

Geometric Design of Roads

The **geometric design of roadways** deals with the portioning of the physical elements of the roadway according to standards and constraints. The basic objective in geometric design is to provide a smooth-flowing, crash-free facility. The American Association of State Highway and Transportation Officials (AASHTO) has established guidelines relating to the design of roadways. Geometric roadway design can be broken into two main parts: vertical curves and horizontal curves. Combined, vertical and horizontal curves provide a three-dimensional layout for a roadway. Using guidelines provided by AASHTO, an engineer can design a roadway that is comfortable, safe, and appealing to the eye. The AASHTO guidelines take into account speed, vehicle type, road grade (slope), view obstructions, and stopping distance.

Vertical Curves

Vertical curves are used to provide a gradual change from one road slope to another, so that vehicles may smoothly navigate grade changes as they travel.

Sag vertical curves are those that have a tangent slope at the end of the curve that is higher than that of the beginning of the curve. When driving on a road, a sag curve would appear as a valley, with the vehicle first going downhill before reaching the bottom of the curve and continuing uphill or level.

Crest vertical curves are those that have a tangent slope at the end of the curve that is lower than that of the beginning of the curve. When driving on a crest curve, the road appears as a hill, with the vehicle first going uphill before reaching the top of the curve and continuing downhill.

Terminology

G1 = initial roadway (tangent)slope

G2 = final roadway (tangent)slope

A = absolute value of the difference in grades (initial minus final, expressed in percent)

L = curve length (along the x-axis)

BVC = begin of vertical curve

PVI = point of vertical interception (intersection of initial and final grades)

EVT = end of vertical tangent

x = distance from PTC/BVC

tangent elevation = elevation of a point along the initial tangent

Y (offset) = vertical distance from the initial tangent to the curve

Y' = curve elevation = tangent elevation - offset

Sag Curves

Sag vertical curves are curves which have a negative A-value - that is, the tangent slope at the end of the curve is greater than the tangent slope at the beginning of the curve. The most important design criteria for these curves is headlight sight distance. When a driver is driving on a sag curve at night, the sight distance is limited by the higher grade in front of the vehicle. This distance must be long enough that the driver can see an obstruction in the road and stop the vehicle within the headlight sight distance. The headlight sight distance (S) is determined by the angle of the headlight and angle of the tangent slope at the end of the curve. By first finding the headlight sight distance (S) and then solving for the curve length (L) in each of the equations below, the correct curve length can be determined. If the S<L curve length is greater than the headlight sight distance, then this number can be used. If it is smaller, this value cannot be used. Similarly, if the S>L curve length is smaller than the headlight sight distance, then this number can be used. If it is larger, this value cannot be used.

Sight Distance < Curve Length (S<L)

$L = AS^2 / (400 + 3.5S)$

Sight Distance > Curve Length (S>L)

$L = 2S - (400 + 3.5S) / A$

Crest Curves

Crest vertical curves are curves which have a positive A-value - that is, the tangent slope at the end of the curve is less than the tangent slope at the beginning of the curve. The most important design criteria for these curves is stopping sight distance. As a vehicle comes over the top of a crest vertical curve, the driver must be able to see any obstructions that lie ahead within a safe stopping distance. If a car on the other side of the hill is stalled or there is an animal in the vehicle's lane, the driver must see the obstruction and be able to stop the car within the stopping sight distance. The stopping sight distance (S) is determined by the speed limit on a road. By first finding the stopping sight distance (S) and then solving for the curve length (L) in each of the equations below, the correct curve length can be determined. If the S<L curve length is greater than the stopping sight distance, then this number can be used. If it is smaller, this value cannot be used. Similarly, if the S>L curve length is smaller than the stopping sight distance, then this number can be used. If it is larger, this value cannot be used.

Sight Distance > Curve Length (S>L)

$L = 2S - (2158 / A)$

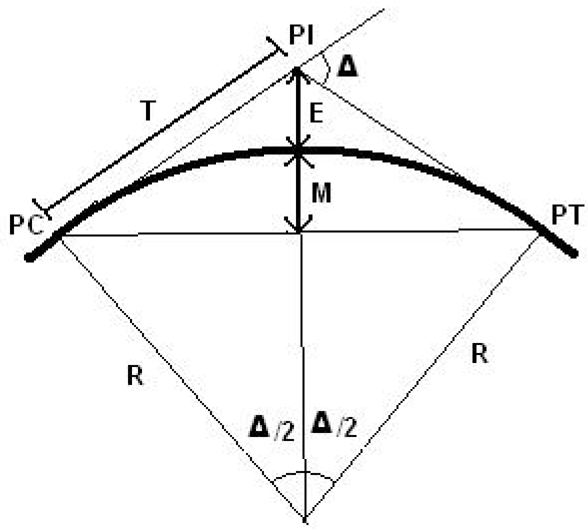
Sight Distance < Curve Length (S<L)

$L = AS^2 / 2158$

Horizontal Curves

Horizontal alignment in road design consists of straight sections of road, known as tangents, connected by horizontal curves. The design of a horizontal curve entails the determination of a minimum radius (based on speed limit), curve length, and objects obstructing the view of the driver. Using AASHTO standards, an engineer works to design a road that is safe and comfortable. If a horizontal curve has a high speed and a small radius of curvature, an increased superelevation (bank) is needed in order to assure safety. If there is an object obstructing the view around a corner or curve, the engineer must work to ensure that drivers can see far enough to stop to avoid an accident or accelerate to join traffic.

Terminology



R = Radius

PC = Point of Curvature (point at which the curve begins)

PT = Point of Tangent (point at which the curve ends)

PI = Point of Intersection (point at which the two tangents intersect)

T = Tangent Length

C = Long Chord Length (straight line between PC and PT)

L = Curve Length

M = Middle Ordinate, now known as HSO - Horizontal Sightline Offset (distance from sight-obstructing object to the middle of the outside lane)

e = Rate of Superelevation

fs= Coefficient of Side Friction

u = Vehicle Speed

Δ = Deflection Angle

Geometry

$T = R\tan(\Delta / 2)$

$C = 2R\sin(\Delta / 2)$

$L = R\Delta\pi / 180$

Sight Distance

$M = R(1 - \cos((28.65S) / R))$

Superelevation

This concept deals with the horizontal or lateral(cross) slope of a roadway. Superelevation of a roadway is used in order to accomplish two objectives: to provide adequate drainage flow for water on the roadway surface and to aid in steering of motor vehicles on curved roadways.

In a tangent section, a common superelevation or cross slope of 1-2% is applied in order to achieve drainage flow of surface water off of the subject roadway. Cross slopes of this magnitude especially when applied in both directions of travel with a crown point along the centerline of a roadway are commonly referred to as "normal crown" and are generally imperceptible to traveling motorists. In a curved section, calculated superelevations generally reaching a maximum of 6 or 8% (depending on selected design criteria) are applied in order to aid motorists in safely traversing these sections while maintaining entry speed of the vehicle along the length of the curve.

The equation for the desired radius of a curve is found below taking into account factors for speed and superelevation rates (e) of a given roadway section. This equation can be rearranged algebraically to provide instead for desired rates of superelevation based on the design speed of a roadway and the radius to be used in a curved section.

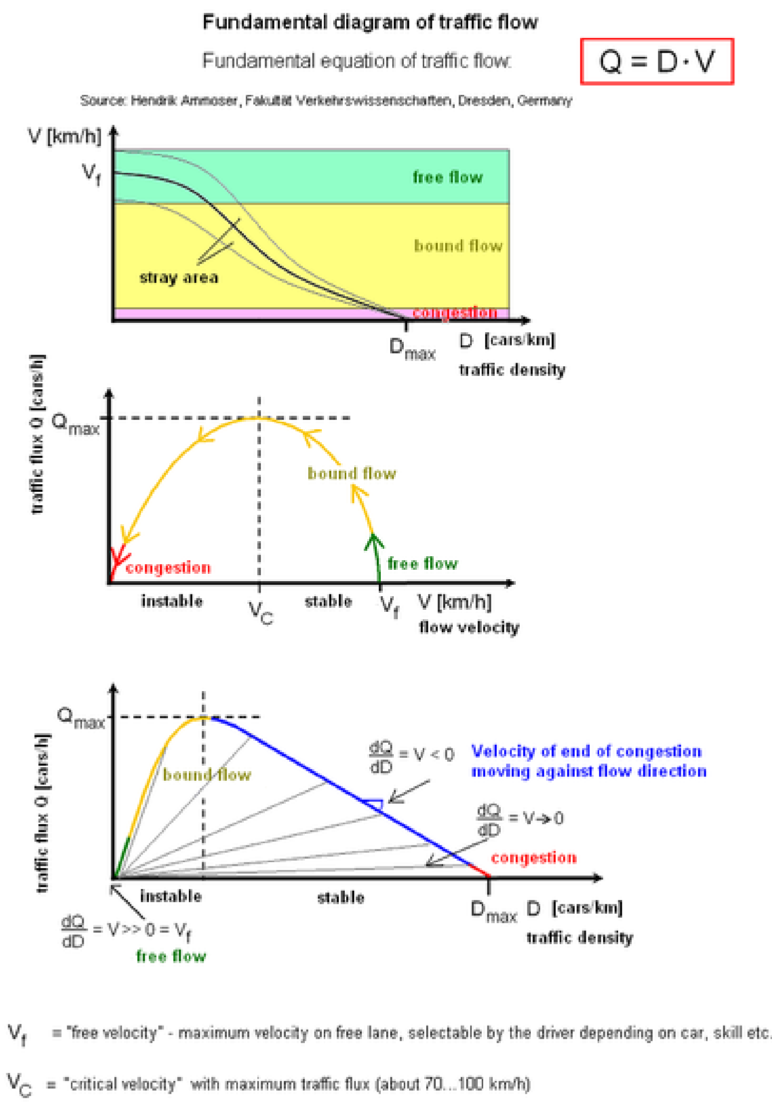
$R = u^2 / (15(e + fs))$

The American Association of State Highway and Transportation officials (AASHTO) provides a table from which desired superelevation rates can be easily interpolated based on the prescribed magnitudes for both design speed and radius of a curved section of roadway. This table can also be seen reprinted in many state roadway design guides and manuals in the U.S.

Chapter 4

Fundamental Diagram of Traffic Flow

The **fundamental diagram of traffic flow** is a diagram that gives a relation between the traffic flux (vehicles/hour) and the traffic density (vehicles/km). A macroscopic traffic model involving traffic flux, traffic density and velocity forms the basis of the fundamental diagram. It can be used to predict the capability of a road system, or its behaviour when applying inflow regulation or speed limits.



Fundamental Diagram of traffic flow

Basic statements

- There is a connection between traffic density and vehicle velocity: The more vehicles are on a road, the slower their velocity will be.
- To prevent congestion and to keep traffic flow stable, the number of vehicles entering the control zone has to be smaller or equal to the number of vehicles leaving the zone in the same time.
- At a critical traffic density and a corresponding critical velocity the state of flow will change from stable to unstable.
- If one of the vehicles brakes in unstable flow regime the flow will collapse.

The primary tool for graphically displaying information in the study traffic flow is the fundamental diagram. Fundamental diagrams consist of 3 different graphs: flow-density, speed-flow, and speed-density. The graphs are two dimensional graphs. All the graphs are related by the equation "flow = speed * density"; this equation is the essential equation in traffic flow. The fundamental diagrams were derived by the plotting of field data points and giving these data points a best fit curve. With the fundamental diagrams researchers can explore the relationship between speed, flow , and density of traffic.

Speed-Density

The speed-density relationship is linear with a negative slope; therefore, as the density increases the speed of the roadway decreases. The line crosses the speed axis, y, at the freeflow speed, and the line crosses the density axis, x, at the jam density. Here the speed approaches freeflow speed as the density approaches zero. As the density increases, the speed of the vehicles on the roadway decreases. The speed reaches zero when the density equals the jam density.

Flow -Density

In the study of traffic flow theory, the flow-density diagram is used to determine the traffic state of a roadway. Currently, there are two types of flow density graphs. The first is the parabolic shaped flow-density curve, and the second is the triangular shaped flow density curve. Academia views the triangular shaped flow-density curve as more the accurate representation of real world events. The triangular shaped curve consists of two vectors. The first vector is the freeflow side of the curve. This vector is created by placing the freeflow velocity vector of a roadway at the origin of the flow-density graph. The second vector is the congested branch, which is created by placing the vector of the shock wave speed at zero flow and jam density. The congested branch has a negative slope, which implies that the higher the density on the congested branch the lower the flow; therefore, even though there are more cars on the road, the number of cars passing a single point is less than if there were fewer cars on the road. The intersection of freeflow and congested vectors is the apex of the curve and is considered the capacity of the roadway, which is the traffic condition at which the maximum number of vehicles can pass by a point in a given time period. The flow and capacity at which this point occurs is the optimum flow and optimum density, respectively. The flow density diagram is used to give the traffic condition of a roadway. With the traffic conditions, time-space diagrams can be created to give travel time, delay, and queue lengths of a road segment.

Speed-Flow

Speed flow diagrams are used to determine the speed at which the optimum flow occurs. There are currently two shapes of the speed-flow curve. The speed-flow curve also consists of two branches, the freeflow and congested branches. The diagram is not a function, allowing the flow variable to exist at two different speeds. The flow variable existing at two different speeds occurs when the speed is higher and the density is lower or when the speed is lower and the density is higher, which allows for the same flow rate. In the first speed-flow diagram, the freeflow branch is a horizontal line, which shows that the roadway is at freeflow speed until the optimum flow is reached. Once the optimum flow is reached, the diagram switches to the congested branch, which is a parabolic shape. The second speed flow diagram is a parabola. The parabola suggests that the only time there is freeflow speed is when the density approaches zero; it also suggests that as the flow increases the speed decreases. This parabolic graph also contains an optimum flow. The optimum also flow divides the freeflow and congested branches on the parabolic graph.

Chapter 5

Banked Turn

A **banked turn** is a turn or change of direction in which the vehicle banks or inclines, usually towards the inside of the turn. The bank angle is the angle at which the vehicle is inclined about its longitudinal axis with respect to its path.

Turn on flat surfaces

If the bank angle is zero the surface is flat, simplifying the calculations, and the vehicle is just driving in a circle. The normal force is vertically upwards, so the only force keeping the vehicle turning on its path is friction, or traction. This must be large enough to provide the centripetal acceleration, a relationship which can be expressed as an inequality,

$$\mu mg > \frac{mv^2}{r}.$$

The expression on the right hand side is the centripetal acceleration multiplied by mass, so the force required to turn the vehicle. The left hand side is the maximum frictional force, which equals the coefficient of friction μ multiplied by the normal force. Rearranging the maximum cornering speed is

$$v < \sqrt{r\mu g}.$$

Note that μ can be the coefficient for static or dynamic friction. In the latter case, where the vehicle is skidding around a bend, the friction is at its limit and the inequalities becomes equations. This also ignores effects such as downforce which can increase the normal force and cornering speed.

Normal reaction in a banked turn

In the case of a car being parked on a banked turn, the Normal force would simply be:

$mg\cos\theta$

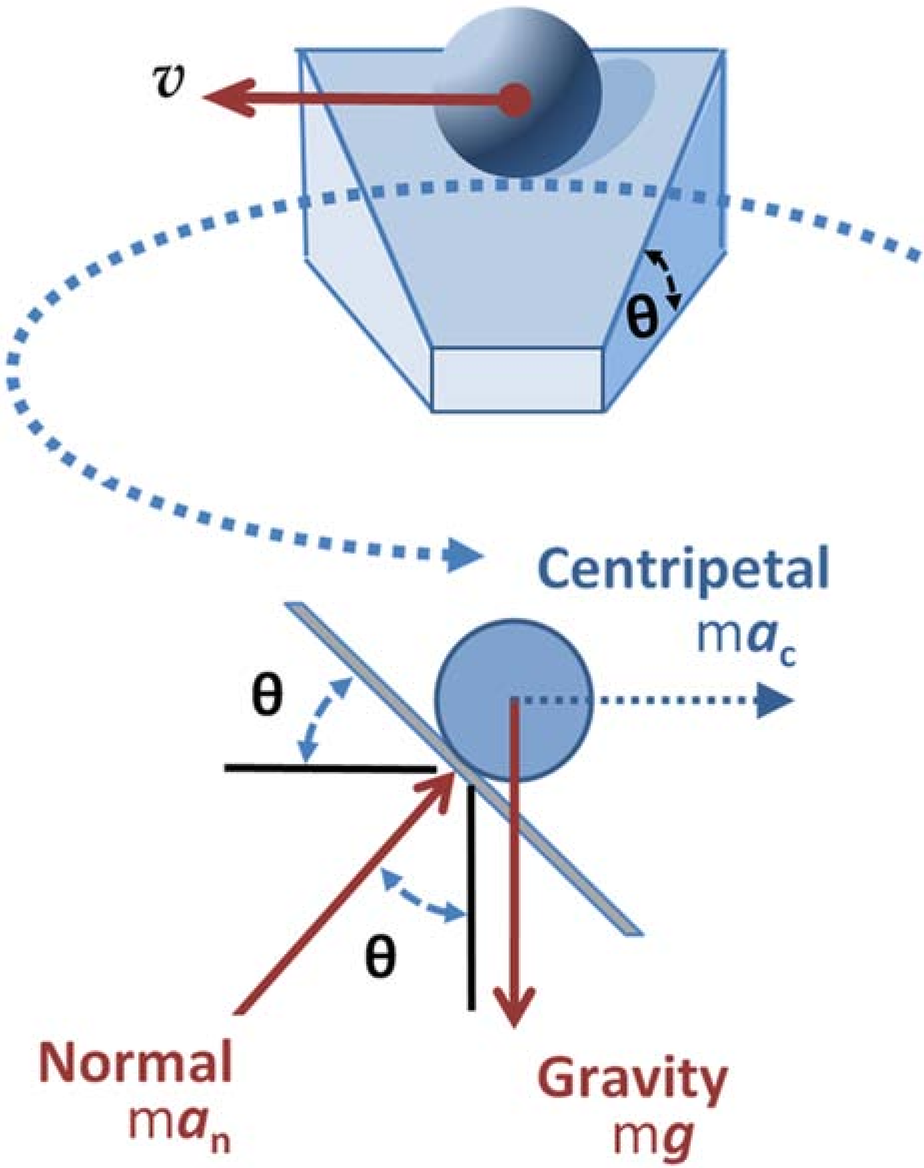
But once the car starts to move on a banked turn, it 'collides' with the turn itself, the turn feels this force, and returns it in the normal, causing the car to move circularly.

The Normal is thus greater than simply its gravitational component.

$$N = mg\cos\theta + \frac{mv^2}{r}\sin\theta$$

If the Normal were simply the component to gravity, one could not say $mg = N\cos\theta = mg\cos^2\theta$ when the vertical acceleration is 0.

Frictionless banked turn



Upper panel: Ball on a banked circular track moving with constant speed *v*; Lower panel: Forces on the ball. The resultant or net force on the ball found by vector addition of the normal force exerted by the road and vertical force due to gravity must equal the required force for centripetal acceleration dictated by the need to travel a circular path.

As opposed to a car riding along a flat circle, inclined edges add an additional force that keeps the car in its path and prevents it from being "dragged into" or "pushed out of" the circle. This force is the horizontal component of the car's normal force. In the absence of friction, the normal force is the only one acting on the car in the direction of the center of the circle. Therefore, as per Newton's second law, we can set the horizontal component of the normal force equal to mass multiplied by centripetal acceleration:

$$N \sin\theta = \frac{mv^2}{r}$$

Because there is no motion in the vertical direction, the sum of all vertical forces acting on the system must be zero. Therefore we can set the vertical component of the car's normal force equal to its weight:

$$N\cos\theta = mg$$

Solving the above equation for the normal force and substituting this value into our previous equation, we get:

$$\frac{mv^2}{r} = mg\tan\theta$$

Solving for velocity we have:

$$v = \sqrt{rg\tan\theta}$$

This provides the velocity that in the absence of friction and with a given angle of incline and radius of curvature, will ensure that the car will remain in its designated path. The magnitude of this velocity is also known as the "rated speed" of a turn or curve. Notice that the rated speed of the curve is the same for all massive objects, and a curve that is not inclined will have a rated speed of 0.

Banked turn with friction

When considering the effects of friction on the system, once again we need to note which way the friction force is pointing. When calculating a maximum velocity for our automobile, friction will point down the incline and towards the center of the circle. Therefore we must add the horizontal component of friction to that of the normal force. The sum of these two forces is our new net force in the centripetal direction:

$$\frac{mv^2}{r} = \mu_s N\cos\theta + N\sin\theta$$

Once again, there is no motion in the vertical direction, allowing us to set all opposing vertical forces equal to one another. These forces include the vertical component of the normal force pointing upwards and both the car's weight and vertical component of friction pointing downwards:

$$N\cos\theta = \mu_s N\sin\theta + mg$$

By solving the above equation for mass and substituting this value into our previous equation we get:

$$\frac{v^2\left(N\cos\theta-\mu_s N\sin\theta\right)}{rg}=\mu_s N\cos\theta+N\sin\theta$$

Solving for v we get:

$$v=\sqrt{\frac{rg\left(\sin\theta+\mu_s\cos\theta\right)}{\cos\theta-\mu_s\sin\theta}}$$

This equation provides the maximum velocity for the automobile with the given angle of incline, coefficient of static friction and radius of curvature. By a similar analysis of minimum velocity, the following equation is rendered:

$$v=\sqrt{\frac{rg\left(\sin\theta-\mu_s\cos\theta\right)}{\cos\theta+\mu_s\sin\theta}}$$

The difference in the latter analysis comes when considering the direction of friction for the minimum velocity of the automobile (towards the outside of the circle). Consequently opposite operations are performed when inserting friction into equations for forces in the centripetal and vertical directions.

Improperly banked road curves increase the risk of run-off-road and head-on crashes. A 2% deficiency in superelevation (say, 4% superelevation on a curve that should have 6%) can be expected to increase crash frequency by 6%, and a 5% deficiency will increase it by 15%. Up until now, highway engineers have been without efficient tools to identify improperly banked curves and to design relevant mitigating road actions. A modern profilograph can provide data of both road curvature and cross slope (angle of incline).

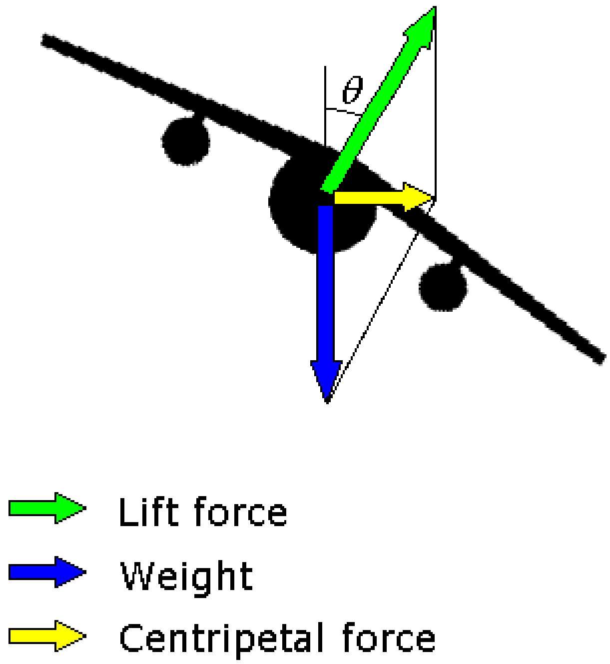
Banked turn in aeronautics

Douglas DC-3 banking to make a left turn

When a fixed-wing aircraft is making a turn (changing its direction) the aircraft must roll to a banked position so that its wings are angled towards the desired direction of the turn. When the turn has been completed the aircraft must roll back to the wings-level position in order to resume straight flight.

When any moving vehicle is making a turn, it is necessary for the forces acting on the vehicle to add up to a net inward force, to cause centripetal acceleration. In the case of an aircraft making a turn, the force causing centripetal acceleration is the horizontal component of the lift acting on the aircraft.

In straight, level flight, the lift acting on the aircraft acts vertically upwards to counteract the weight of the aircraft which acts downwards. During a balanced turn where the angle of bank is θ the lift acts at an angle θ away from the vertical. It is useful to resolve the lift into a vertical component and a horizontal component. If the aircraft is to continue in level flight (i.e. at constant altitude), the vertical component must continue to equal the weight of the aircraft. The horizontal component is unbalanced, and is thus the net force causing the aircraft to accelerate inward and execute the turn.



Vector diagram showing lift, weight and centripetal force acting on a fixed-wing aircraft during a banked turn.

During a banked turn in level flight the lift on the aircraft must support the weight of the aircraft, as well as provide the necessary component of horizontal force to cause centripetal acceleration. Consequently, the lift required in a banked turn is greater than that one required in straight, level flight and can be achieved either by increasing the angle of attack of the wing (typically by pulling on the elevator control) or by deploying flaps. The maneuver is usually complemented by an increase in power, in order to maintain airspeed.

Because centripetal acceleration is:

$$a = \frac{v^2}{r}$$

Newton's second law in the horizontal direction can be expressed mathematically as:

$$L \sin \theta = \frac{mv^2}{r}$$

where:

L is the lift acting on the aircraft
θ is the angle of bank of the aircraft
m is the mass of the aircraft
v is the true airspeed of the aircraft
r is the radius of the turn

In straight flight, lift is approximately equal to the aircraft weight. In turning flight the lift exceeds the aircraft weight, and is equal to the weight of the aircraft (mg) divided by the cosine of the angle of bank:

$$L = \frac{mg}{\cos\theta}$$

where g is the gravitational field strength.

The radius of the turn can now be calculated:

$$r = \frac{v^2}{g\tan\theta}$$

This formula shows that the radius of turn is proportional to the square of the aircraft's true airspeed. With a higher airspeed the radius of turn is larger, and with a lower airspeed the radius is smaller.

This formula also shows that the radius of turn is inversely proportional to the angle of bank. With a higher angle of bank the radius of turn is smaller, and with a lower angle of bank the radius is greater.

The angle of bank is the sole determinant of the aircraft's load factor during the turn.

Chapter 6

Red Light Camera

A red light camera in Chicago.

A red-light camera in use in Beaverton, Oregon, USA

A **red light camera** is a traffic camera which captures an image of a vehicle moving through a signalized intersection traffic moving in the direction of that vehicle has a red signal. Red light cameras enforce traffic laws and deter violation to traffic laws by photographing vehicles passing through signalized intersections illegally. Red light cameras automatically photograph vehicles whose drivers run red lights. The cameras are connected to the traffic signal and to sensors that monitor traffic flow just before the crosswalk or stop line. The system continuously monitors the traffic signal, and the camera captures any vehicle that doesn't stop during the red phase. Many red light camera programs provide motorists with grace periods of up to half a second after the light switches to red.

Red light cameras make great engineering tools too. The red light running data they collect expose engineering failures such as short yellows, dilemma zone problems and over-sensitive right-turn-on-red actuators. As the red light camera data shows, most red light running is caused by engineering defects, including the defect of the dilemma zone which departments of transportation build into every intersection in the world. This defect guarantees a steady stream of cars running red lights. Today's usage of red light cameras is as an enforcement device. The usage is based on the premise that red light running is only caused by bad driving, as opposed to bad engineering.

Depending on the particular technology, a series of photographs and/or a video clip show the red light prior to a car entering the intersection on a red signal, as well as the vehicle's progression through the intersection. Cameras record the date, time of day, time elapsed since the beginning of the red signal, vehicle speed, and license plate. Tickets typically are mailed to owners of violating vehicles, based on a review of photographic evidence.

Fines carried out against drivers for red light camera-caught violations in the U.S. are regulated at the State and City level and range from $50 in North Carolina to over $500 in California.

Studies

In the United States of America, a number of studies have addressed the question of whether, on balance, red-light cameras produce a safety benefit.

Most recently, the Insurance Institute for Highway Safety found a reduction in the rate of fatal crashes at signalized intersections both with and without red light cameras in cities where red light cameras are in use. After controlling for population density and land area, the rate of fatal red light running crashes during 2004-08 for cities with camera programs was an estimated 24 percent lower than what would have been expected without cameras. However Raleigh, North Carolina and Bakersfield, California showed increases. The rate of all fatal crashes at signalized intersections during 2004-08 for cities with camera programs was an estimated 17 percent lower than what would have been expected without cameras.

A July 2004 study of 17,271 crashes by Mark Burkey, PhD., and Kofi Obeng, PhD of North Carolina A & T University showed that the presence of red light cameras only increased the overall number of crashes by 40%. "The results do not support the view that red light cameras reduce crashes. Instead, we find that RLCs are associated with higher levels of many types of severity categories of crashes." However, this research received no scientific peer review and exhibited major research flaws.

A U.S. study , for example, found that red light cameras led to a decrease in right-angle crashes and an increase in the number of rear-end collisions. The total number of collisions remained essentially unchanged. The study applied estimates from a 1997 study of the cost of accidents based on severity to conclude the cameras yielded a positive overall cost benefit from a reduction in more expensive right-angle injury collisions.

This FHWA study was criticized in two follow-up studies. In addition to a study co-director having performed research for the Insurance Institute for Highway Safety, which represents an industry that profits significantly from red light camera surcharges, the follow up studies found critical methodological flaws and that the FHWA study completely missed an increase in fatalities associated with red light camera use. Additionally, "the authors spotlight the statistical difficulties of including the cost of fatalities, while ignoring the practical implications of such events", assuming that each angle injury crash had a societal cost of $64,468, when in fact the cost was $82,816 before camera use and $100,176 after camera use.

A 2005 Virginia Department of Transportation study of the long-term effect of camera enforcement in the state found a decrease in the number of right-angle crashes, but an increase in rear-end crashes and an overall increase in the number of accidents causing injuries .

A 2004 Texas Transportation Institute study found, "crashes decrease with an increase in yellow interval duration and a reduction in speed limit." After 1.0 second was added to the yellow signal timing at test intersections, accidents dropped by 35 to 40%. This compares with a 6.4% reduction for "area-wide officer enforcement of intersection traffic control devices... during the time of the enforcement activity" .

A 2005 study of the Raleigh, North Carolina red light camera program conducted by the Institute for Transportation Research and Education at North Carolina State University compared "before" and "after" red-light camera intersection data and found right-angle crashes dropped by 42 percent, rear-end crashes dropped by 25 percent and total accidents dropped by 17 percent . However, this contradicts a later report done by the Institute of Highway Safety (February 2011) which showed that crashes and fatalities increased in Raleigh. Also, the study contradicts the experience of several other cities such as:

- Albuquerque, NM, which found that crashes increased at camera-monitored intersections while declining at unmonitored intersections.
- Aurora, CO, which after starting camera enforcement at 4 intersections, saw 100% crash increases at two, 175% crash increase at one, and a 60% crash decrease at one intersection.
- Chicago, IL, where "there is no evidence that the red light camera have had a significant safety benefit".
- Los Angeles, CA, where cameras "failed to adequately demonstrate an improvement in safety".
- Minneapolis, MN, which saw a 11% reduction in crashes after disabling its red light camera system.
- Winnipeg, Canada, where crashes significantly increased in the years following red light camera introduction.

A 2005 meta analysis compared the results of 10 controlled before-after studies of red light cameras in the United States, Australia and Singapore. "Five studies found that use of red-light cameras cut the number of crashes in which there were injuries. In the best conducted of these studies, the reduction was nearly 30%... The evidence is less conclusive on total collisions, specific casualty collision types and violations, where reductions achieved could be explained by the play of chance. Most evaluations did not adjust for RTM or spillover, affecting their accuracy. Larger and better controlled studies are needed."

Usage

As of May 2009, red light cameras are used in more than 400 communities in the United States. Major cities that use red light cameras include Albuquerque, Atlanta, Austin, Baltimore, Baton Rouge, Chicago, Dallas, Denver, Houston, Los Angeles, New Orleans, New York City, Philadelphia, Phoenix, Raleigh, San Diego, San Francisco, Seattle, Tucson and Washington, DC.

On November 2, 2010, four years after the original installation, Houston, Texas voted 53% to 47% against the continued use of red cameras.

Some states have chosen to prohibit the use of red light cameras. These include Connecticut, Nevada, New Hampshire, West Virginia, Wisconsin, and Mississippi with other states, such as Louisiana considering a ban. In some states such as Wisconsin, the ban comes from decisions by the state supreme court declaring that type of device unconstitutional.

Some countries in other parts of the world that use this technology extensively include Australia, Austria, Belgium, Canada, Germany, Israel, Malta, the Netherlands, Singapore, South Africa, India, Switzerland, Taiwan and the United Kingdom.

Issues

While some public officials support the use of red-light cameras, various groups and individuals oppose them. They believe that the use of these devices raises legal issues and violates the privacy of citizens. They also question the effectiveness of red light cameras and if they really help traffic safety. In a few U.S. states (including California), the cameras *are* set up to get a "face photo" of the driver;. This has been done because in those states, red light camera tickets are criminal violations, and criminal charges must always name the actual violator. In California, that need to identify the actual violator has led to the creation of a unique investigatory tool.

In some areas, red light enforcement cameras are installed and maintained by private firms such as Affiliated Computer Services, American Traffic Solutions, Inc., and Redflex Traffic Systems. Many people disagree with this privatization of a police function. In Texas, red light violators caught by a red light camera are served with a civil citation rather than a criminal citation. The civil infraction (civil fine of $75, no traffic points) conflicts with the same criminal infraction (fines of $1 to $200, and traffic points). A December 2008 lawsuit against Dallas County's program, challenging a private camera operating company's right to hand out citations, was dismissed in March 2009.

Throughout the USA, red light cameras were used in 380 individual communities as of October 2008 and speed cameras were used in 45 communities plus Arizona on state roads and Illinois in work zones as of September 2008. Arkansas, Nebraska, Nevada, New Jersey, Utah, West Virginia and Wisconsin have also enacted various prohibitions on photo enforcement.

On February 19, 2010, Arizona completed a study of their statewide 76 photo enforcement cameras and decided that they would not renew their contract with Redflex in 2011. Less than expected revenue, mixed public acceptance and mixed accident data were cited, among other findings.

In New York State, red light cameras are allowed in New York City, and since 2009, Nassau County, Yonkers, Buffalo, Rochester, and Suffolk County,

On January 31, 2008, then New York State Assembly member Ivan Lafayette introduced New York State Assembly Bill A09877 with intent to prohibit the sale or use of a product that alerts of the presence of a red light camera under a civil penalty up to 500 USD. The memo of the bill claimed New York City's Red-Light Camera Program to be a revenue-neutral success to reduce the running of red lights and improve safety. It was referred to consumer affairs and protection on the same day.

Chapter 7

Three-Phase Traffic Theory

Three-phase traffic theory is an alternative theory of traffic flow developed by Boris Kerner between 1996 and 2002. It focuses mainly on the explanation of the physics of traffic breakdown and resulting congested traffic on highways. Kerner describes three phases of traffic, while the classical theories based on the fundamental diagram of traffic flow have two phases: *free flow* and *congested traffic*. Kerner's theory divides congested traffic into two distinct phases, *synchronized flow* and *wide moving jam*, bringing the total number of phases to three:

Synchronized flow

Jam

1. Free flow (*F*)
2. Synchronized flow (*S*)
3. Wide moving jam (*J*)

A **phase** is defined as a *state in space and time.*

*Free flow (*F*)*

In free traffic flow, empirical data show a positive correlation between the flow rate *q* (in vehicles per time interval) and vehicle density *k* (in vehicles per kilometres). This relation has an upper boundary at the maximal point of free flow at the flow rate at the corresponding critical density k_{crit} (Figure 1).

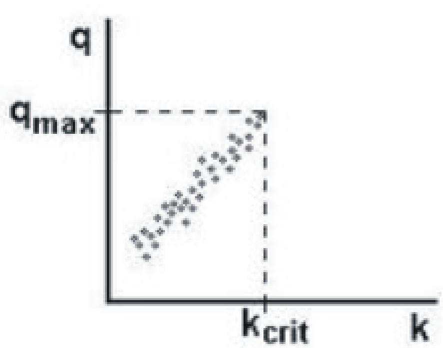


Figure 1:Measured data of flow rate related to vehicle density in free flow

Congested traffic

In congested traffic, the vehicle speed is lower than the minimal possible vehicle speed v_{free}^{min} in free flow, i.e., the line with the slope of the minimal speed $v_{\text{free}}^{\text{min}} = \frac{q_{\text{max}}}{k_{\text{crit}}}$ in free flow (dotted line in Figure 2) divides all empirical data on the flow-density plane into two regions: on the left side the data of free flow and on the right side the data of the congested traffic.

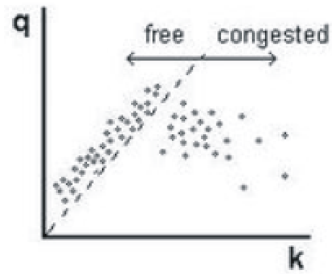


Figure 2: Measured data of flow rate related to vehicle density in free flow and congested traffic

Definitions of the phases J *and* S *in congested traffic*

Kerner's phase definitions [*J*] and [*S*] in congested traffic are the result of common spatial-temporal features of real traffic data. These phase definitions are as follows:

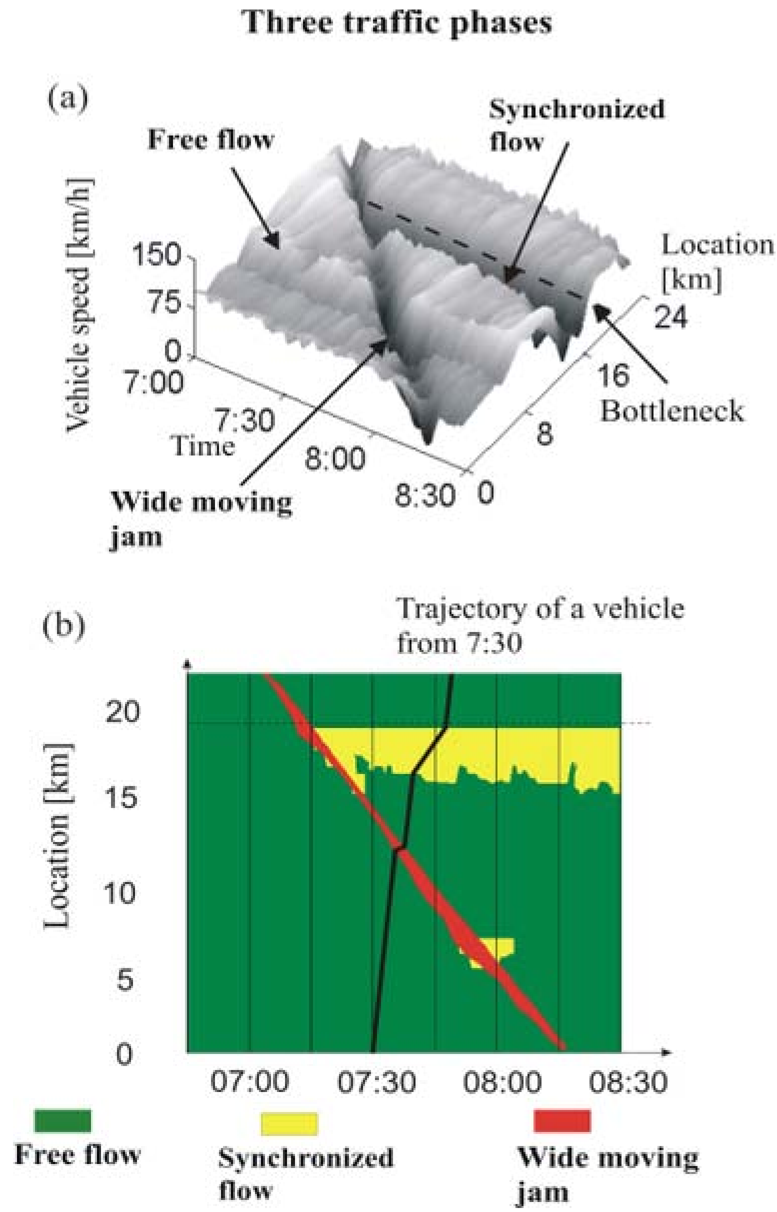


Figure 3: Measured data of speed in time and space (a) and its representation on the time-space plane (b)

The traffic phase "wide moving jam" (*J*)

A wide moving jam moves through a highway bottleneck while maintaining the mean velocity v_g of the jam downstream front at which vehicles escaping the jam accelerate either to free flow or synchronized flow. This is the characteristic feature of the wide moving jam.

The traffic phase "synchronized flow" (S)

The downstream front of synchronized flow, where the vehicles accelerate to free flow, does not show this characteristic feature of the wide moving jam. Specifically, the downstream front of synchronized flow is mostly fixed at the bottleneck.

Explanation of the traffic phase definitions based on measured traffic data

Measured data of the averaged vehicle speeds (Figure 3 (a)) illustrate the definitions [J] and [S]. There are two spatial-temporal patterns of congested traffic with low vehicle speeds in Figure 3 (a). One pattern of congested traffic propagates upstream with almost constant velocity of the downstream pattern front through the freeway bottleneck. According to the definition [J] this pattern of congested traffic belongs to the "wide moving jam" traffic phase. In contrast, the downstream front of the other pattern of the congested traffic is fixed at the bottleneck. According to the definition [S] this pattern of congested traffic belongs to the "synchronized flow" traffic phase (Figure 3 (a) and (b)).

The fundamental hypothesis of Kerner's three-phase traffic theory

The fundamental hypothesis of Kerner's three-phase traffic theory is associated with steady states of synchronized flow. A steady state of synchronized flow is a *hypothetical* state of synchronized flow of identical vehicles and drivers in which all vehicles move with the same time-independent speed and have the same space gaps (a space gap is a net distance between two following each other vehicles), i.e., this synchronized flow is homogeneous in time and space. The fundamental hypothesis is as follows: Steady states of synchronized flow cover a two-dimensional (2D) region in the flow-density plane (2D-region S in Figure 4(a)). The multitudes of free flow states (F) overlap steady states of synchronized flow in the vehicle density. The free flow states on a multi-lane road and steady states of synchronized flow are separated by a gap in the flow rate and, therefore, by a gap in the speed at a given density: at each given density the synchronized flow speed is lower than the free flow speed. In accordance with the fundamental hypothesis of Kerner's three-phase theory, at a given speed in synchronized flow, the driver can make an *arbitrary choice* in the space gap to the preceding vehicle within a finite range of space gaps associated with the 2D-region of steady states of synchronized flow (Figure 4(b)): the driver accepts different space gaps at different times and does not control a fixed space gap to the preceding vehicle.

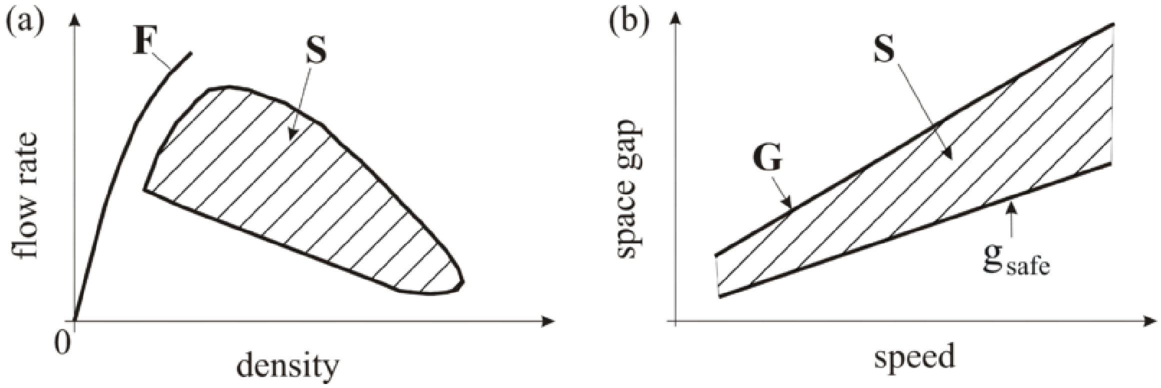


Figure 4: Fundamental hypothesis of Kerner' three-phase traffic theory: (a) Qualitative representation of free flow states (F) and 2D-region of steady states of synchronized flow (dashed region S) on a multi-lane road in the flow-density plane. (b) A part of the 2D-region of steady states of synchronized flow shown in (a) in the space-gap-speed plane (dashed region S). In (b), G and g_{safe}, are respectively a synchronization space gap and safe space gap between two vehicles following each other.

The fundamental hypothesis of Kerner's three-phase traffic theory contradicts the hypothesis of earlier traffic flow theories about the fundamental diagram of traffic flow that is a 1D-relationship between the vehicle density and flow rate.

Car following in three-phase traffic theory

In Kerner's three-phase theory, a vehicle accelerates when the space gap g to the preceding vehicle is greater than a synchronization space gap G, i.e., at $g > G$ (labelled by *acceleration* in Figure 5); the vehicle decelerates when the gap g is smaller than a safe space gap g_{safe}, i.e., at $g < g_{\text{safe}}$ (labelled by *deceleration* in Figure 5).

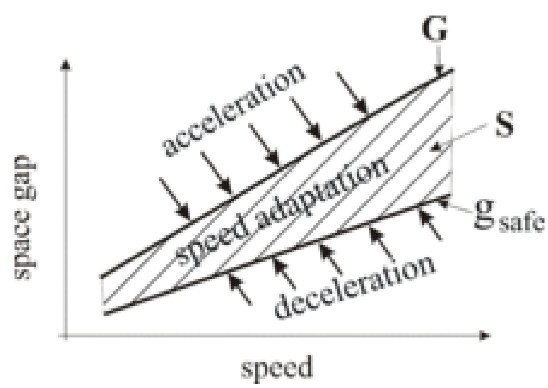


Figure 5: Qualitative explanation of car-following in Kerner's three-phase traffic theory: A vehicle accelerates at a space gap $g > G$ and decelerates at space gaps $g < g_{\text{safe}}$, whereas under condition $g_{\mathbf{safe}} \leq g \leq G$ the vehicle adapts its speed to the speed of the preceding vehicle without caring what the precise space gap is. A dashed region of synchronized flow is taken from Figure 4(b).

The synchronization space gap G is a space gap g between the vehicle and the preceding vehicle within which the vehicle tends to adapt the speed to the speed of the preceding vehicle without caring, what the precise space gap is, as long as this space gap is not smaller than the safe space gap g_{safe} (labelled by *speed adaptation* in Figure 5). Thus the space gap g in car following in the framework of Kerner's three-phase theory can be any space gap within the space gap range $g_{\mathrm{safe}} \leq g \leq G$.

Traffic breakdown - a F → S *phase transition*

In measured data, congested traffic most often occurs at highway bottlenecks, e.g., at on-ramps, off-ramps or at roadwork. Such a transition of free flow to congested traffic is known as traffic breakdown. In Kerner's three-phase traffic theory such a traffic breakdown is explained by a F → S phase transition. This explanation is supported by available measurements, because in measured traffic data after a traffic breakdown at a bottleneck the downstream front of the congested traffic is fixed at the bottleneck. Therefore, the emerging congested traffic after a traffic breakdown fulfils the definition [S] of the "synchronized flow" traffic phase.

Spontaneous and induced *F* → *S* transitions

Kerner states by using empirical measured data that synchronized flow can emerge in free flow spontaneously (spontaneous F → S phase transition) or externally induced (induced F → S phase transition). A spontaneous F → S phase transition means that the traffic breakdown occurs in the case that there has been free flow before at the bottleneck as well as both up- and downstream of the bottleneck. This implies that a spontaneous F → S phase transition occurs through the growth of an internal disturbance in free flow in a neighbourhood of a bottleneck. Alternately, an induced F → S phase transition occurs through disturbance of traffic flow that has initially emerged at a different road location than the bottleneck location. Normally, this correlates with an upstream propagation of a synchronized flow region or a wide moving jam. An empirical example of an induced breakdown at a bottleneck leading to synchronized flow can be seen in Figure 3: synchronized flow emerges through the upstream propagation of a wide moving jam.

Physical explanation of traffic breakdown

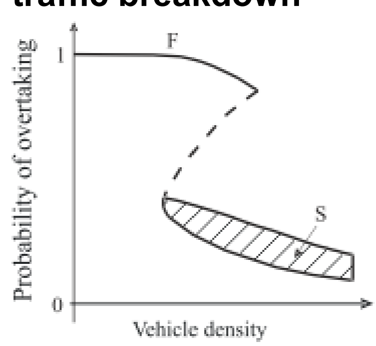


Figure 6: Explanation of traffic breakdown by a Z-like non-linear interrupted function of the probability of overtaking in Kerner's three-phase traffic theory. The dotted curve illustrates the critical probability of overtaking as function of traffic density.

Kerner explains the nature of the F → S phase transitions by a spatial-temporal competition of the vehicle acceleration through the overtaking of a slower vehicle ahead and the vehicle deceleration to the speed of the slower moving vehicle ahead ("speed adaptation"). The overtaking supports the further existence of free flow. In contrast, "speed adaptation" leads to synchronized flow. Such a speed adaptation will occur, if an overtaking is not possible. Kerner states that the probability of overtaking is an *interrupted function of the vehicle density* (Figure 6): at a given vehicle density, the probability of overtaking in free flow is much higher than in synchronized flow.

Infinite number of highway capacities

Maximum and minimum highway capacities

The spontaneous traffic breakdown, i.e., a spontaneous F → S phase transition, might occur in a wide range of flow rates in free flow. Kerner states based on empirical traffic data that because of the possibility of spontaneous or induced traffic breakdowns at the same freeway bottleneck there are an infinite number of highway capacities of free flow at the bottleneck. This infinite number of freeway capacities is between a minimum capacity q_{th} and a maximum capacity q_{max} of free flow (Figure 7).

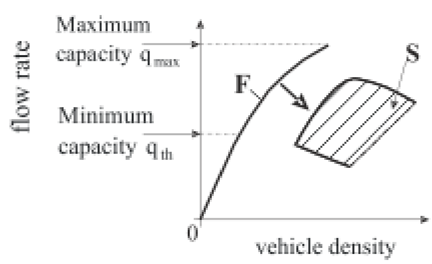


Figure 7: Maximum and minimum highway capacities in Kerner's three-phase traffic theory

Highway capacities and metastability of free flow

Already small disturbances in free flow at the bottleneck will lead to a spontaneous F → S phase transition, if the flow rate in free flow is close to the maximum capacity q_{max}. On the other hand, only very large disturbances in free flow at the bottleneck will lead to a spontaneous F → S phase transition, if the flow rate is close to the minimum capacity q_{th}. The probability of a smaller disturbance in free flow is much higher than of a larger disturbance. Therefore, the higher the flow rate in free flow at a bottleneck, the higher the probability of the spontaneous F → S phase transition. If the flow rate in free flow is lower than the minimum capacity q_{th}, there will be no traffic breakdown (no F → S phase transition) at the bottleneck.

The infinite number of highway capacities at a bottleneck can be illustrated by the meta-stability of free flow at the flow rates q with

$$q_{th} \leq q < q_{\max}.$$

Meta-stability of free flow means, that for small disturbances, free flow can remain stable (free flow remains), but at larger disturbances, the free flow becomes unstable and a F → S phase transition to synchronized flow occurs.

Discussion of capacity definitions

The infinite number of highway capacities at a bottleneck of Kerner's three-phase traffic theory contradicts fundamentally the classical traffic theories and methods for traffic management and traffic control, which at any time assume the existence of a *particular* highway capacity. In contrast, in Kerner's three-phase traffic theory *at any time* there are an infinite number of highway capacities, which are within the abovementioned flow rate range between the minimum capacity q_{th} and maximum capacity q_{max}. The values q_{th} and q_{max} can depend considerably on traffic parameters (the percentage of long vehicles in traffic flow, weather, bottleneck characteristics, etc).

Wide moving jams (J)

A wide moving jam will be called "wide", if the width of the wide moving jam (in direction of the flow) clearly exceeds the width of the jam fronts. The average vehicle speed within the wide moving jams is much lower than the average speed in free flow. At the downstream front the vehicles might accelerate to the free flow speed. At the upstream jam front the vehicles come from free flow or synchronized flow and must reduce their speed. According to the definition [J] the wide moving jam keeps its mean velocity of the downstream front v_g, even if the jam propagates through other traffic phases or bottlenecks. The flow rate is sharply reduced within a wide moving jam.

Characteristic parameters of wide moving jams

Kerner's empirical results show, that some characteristic features of wide moving jams are independent of the traffic volume and bottleneck features (e.g. where and when the jam has emerged). However, these characteristic features are dependent on weather conditions, road conditions, vehicle technology, percentage of long vehicles, etc.. The velocity of the downstream front of a wide moving jam v_g (in direction upstream) is a characteristic parameter, as well as the flow rate downstream of the downstream jam front q_{out} (with free flow at this location, see Figure 8). This means, that several wide moving jams have similar features under similar conditions. These parameters are relatively predictable due to theses reasons. The movement of the downstream jam front can be illustrated in the flow-density plane by a line, which is called "Line J" (Line J in Figure 8). The slope of the Line J is the velocity of the downstream jam front v_g, because the co-ordinate of the Line J at the flow rate zero is the vehicle density within the jam, k_{max}.

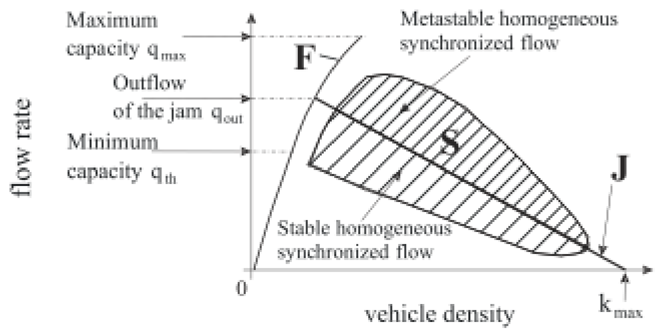


Figure 8: Three traffic phases on the flow-density plane in Kerner's three-phase traffic theory

Minimum highway capacity and outflow from wide moving jam

Kerner emphasizes, that the minimum capacity q_{th} and the outflow of a wide moving jam q_{out} describe two *qualitatively different features* of free flow: the minimum capacity q_{th} characterizes an F → S phase transition at a bottleneck, i.e., a traffic breakdown. The outflow of a wide moving jam q_{out} characterizes conditions of the existence of the jam, i.e., the traffic phase J. Depending on traffic parameters like weather, percentage of long vehicles, etc., and of bottleneck characteristics, where the F → S phase transition can occur, the minimum capacity q_{th} might be smaller (as in Figure 8), or greater than the jam's outflow q_{out}.

*Synchronized flow phase (*S*)*

In contrast to wide moving jams, both the flow rate and vehicle speed might vary significantly in the synchronized flow phase. The downstream front of synchronized flow is often spatially fixed, normally at the bottleneck at a certain road location. The flow rate in this phase could remain similar to the one in free flow, even if the vehicle speeds are sharply reduced.

Because the synchronized flow phase does not have the characteristic features of the wide moving jam phase J, Kerner's three-phase traffic theory assumes that hypothetical homogeneous states of synchronized flow cover a two-dimensional region in the flow-density plane (dashed regions in Figure 8).

S → J *phase transition*

Wide moving jams do not emerge spontaneously in free flow, but they can emerge in regions of synchronized flow. This phase transition is called a S → J phase transition.

"Jam without obvious reason" - F→S→J phase transitions

Therefore, the emergence of wide moving jam in free flow is observed as a cascade of F → S → J phase transitions: first, a region of synchronized flow emerges in a region of free flow. As explained above, such an F → S phase transition occurs mostly at a bottleneck. Within the synchronized flow phase a further "self-compression" occurs and vehicle density increases while vehicle speed decreases. This self-compression is called "pinch effect". In "pinch" regions of synchronized flow, narrow moving jams emerge. If these narrow moving jams grow, wide moving jams will emerge. Kerner notes that the frequency of the emergence of wide moving jams increases, if the density in synchronized flow increases. The wide moving jams propagate further upstream, even if they propagate through regions of synchronized flow or bottlenecks.

The physics of *S* → *J* transition

To further illustrate S → J phase transitions, it should be noted that in Kerner's three-phase traffic theory the Line J divides all homogeneous states of synchronized flow (Figure 8). States of homogeneous synchronized flow above Line J are meta-stable. States of homogeneous synchronized flow below Line J are stable in which no S → J phase transition can occur. Meta-stable homogeneous synchronized flow means that for small disturbances, traffic state remains stable. However, when larger disturbances occur, synchronized flow becomes unstable, and a S → J phase transition occurs.

Traffic patterns of S *and* J

Caused by F → S and S → J phase transitions, very complex congested patterns can be observed.

Classification of synchronized flow traffic patterns (SP)

A congestion pattern of synchronized flow (Synchronized Flow Pattern (SP)) with a fixed downstream and a not continuously propagating upstream front is called Localised Synchronized Flow Pattern (LSP).

Frequently the upstream front of a SP propagates upstream. If only the upstream front propagates upstream, the related SP is called Widening Synchronised Flow Pattern (WSP). The downstream front remains at the bottleneck location and the width of the SP increases.

It is possible that both upstream and downstream front propagate upstream. The downstream front is no longer located at the bottleneck. This pattern has been called Moving Synchronised Flow Pattern (MSP).

Catch effect of synchronized flow at highway bottleneck

The difference between the SP and the wide moving jam becomes visible when the WSP or the MSP reaches an upstream bottleneck: the so-called "catch-effect" occurs. The SP will be caught at the bottleneck and as a result a new congested pattern emerges. A wide moving jam will not be caught at a bottleneck and moves further upstream. In contrast to wide moving jams, the synchronized flow, even as an MSP has no characteristic parameters. As an example, the velocity of the downstream front of the MSP might vary significantly and can be different for different MSPs. These features of SP and wide moving jams are consequences of the phase definitions [S] and [J].

General congested traffic pattern (GP)

A very typical congested pattern contains both congested phases [S] and [J]. Such a pattern with [S] and [J] is called General Pattern (GP).

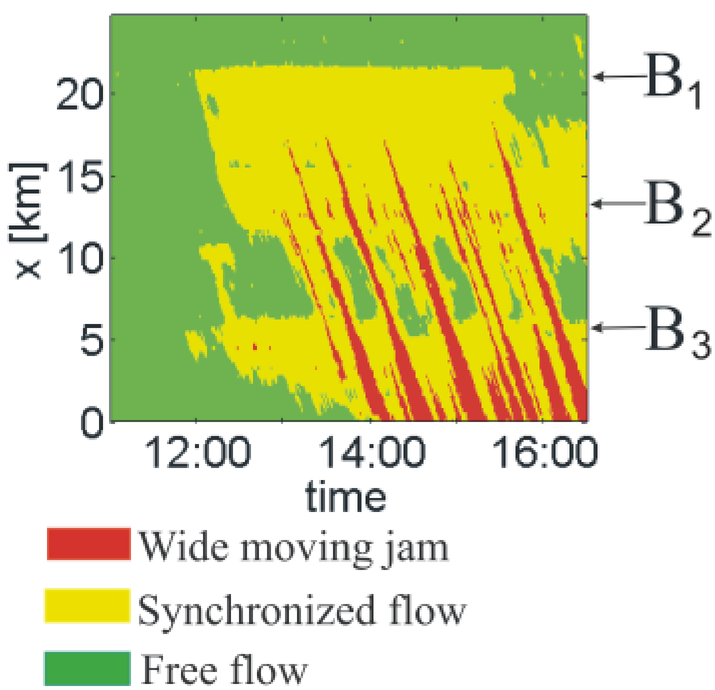


Figure 9: Measured EGP at three bottlenecks B_1, B_2 and B_3

In many freeway infrastructures bottlenecks are very close to each other. A congested pattern whose synchronized flow covering two or more bottlenecks is called Expanded Pattern (EP). An EP could contain synchronized flow only (called ESP: Expanded Synchronized Flow Pattern)), but normally wide moving jams are emerging in the synchronized flow. In those cases the EP is called EGP (Expanded General Pattern) (see Figure 9).

Applications of three-phase traffic theory in transportation engineering

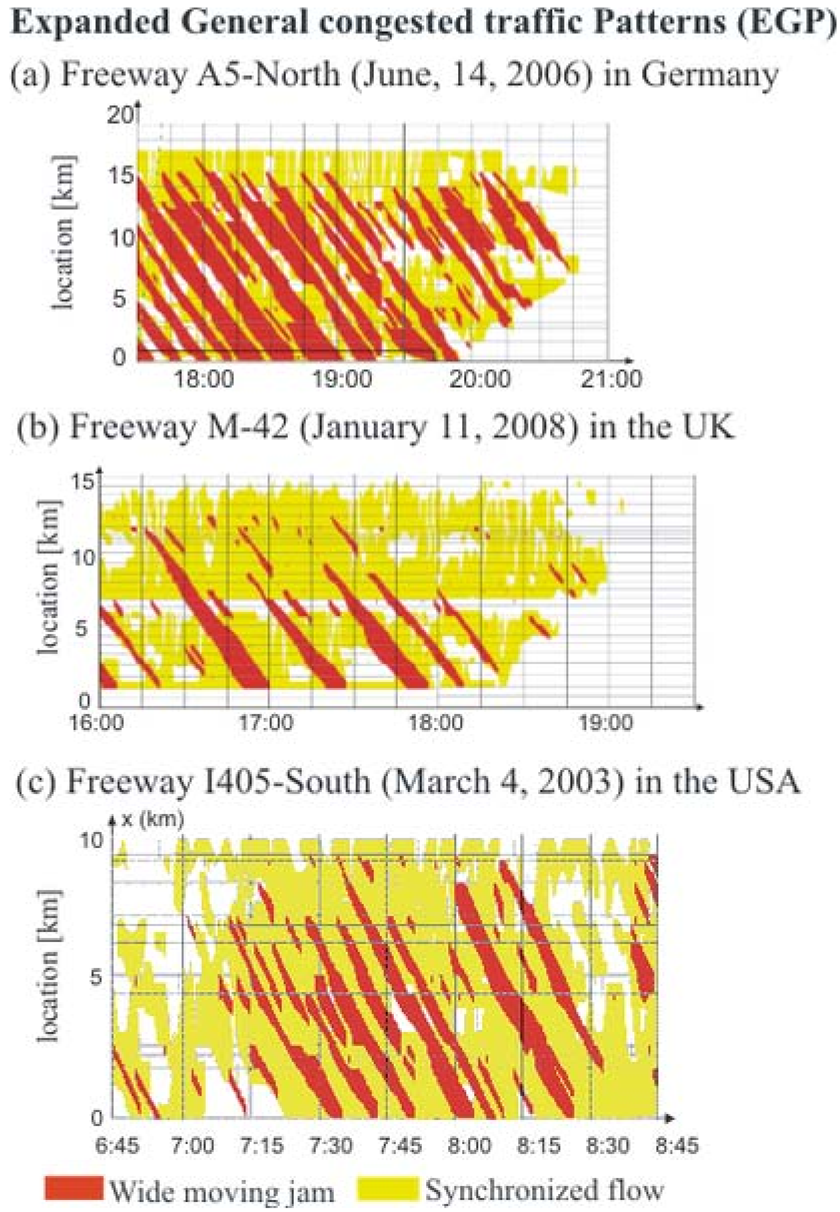


Figure 10: Traffic patterns in the ASDA/FOTO application in three countries

One of the applications of Kerner's three-phase traffic theory is the methods called ASDA/FOTO (**A**utomatische **S**tau**D**ynamik**A**nalyse (Automatic tracking of wide moving jams) and **F**orecasting **O**f **T**raffic **O**bjects). ASDA/FOTO is a software tool able to process large traffic data volumes quickly and efficiently on freeway networks. ASDA/FOTO works in an online traffic management system based on measured traffic data. Recognition, tracking and prediction of [S] and [J] are performed using the features of Kerner's three-phase traffic theory.

Further applications of the theory are seen in the development of traffic simulation models, ramp metering system (ANCONA), collective traffic control, traffic assistance and traffic state detection, as described in the books by Kerner.

Chapter 8

Traffic Barrier

Traffic barriers keep vehicles within their roadway and prevent vehicles from colliding with dangerous obstacles. Traffic barriers installed at the road side also prevent errant vehicles from traversing steep (non-recoverable) slopes. Traffic barriers installed at the medians of divided highways are also referred to as median barriers. The latter also prevent errant vehicles from entering the opposing carriageway of traffic and help to prevent head-on collisions.

Need and placement

Traffic barriers constitute hazards themselves and should only be used when the obstacle poses a greater threat than the barrier itself. In all cases, roadside hazards must be assessed for the danger they pose to traveling motorists based on size, rigidity and distance from the edge of travelway. For instance, small roadside signs and some large signs (ground-mounted breakaway post) often do not merit roadside protection as the barrier itself may pose a greater threat to general health and well-being of the public than the obstacle it intends to protect. In many regions of the world, the concept of clearzone is taken into account when examining the distance of an obstacle or hazard from the edge of travelway.

Clearzone also known as clear recovery area or horizontal clearance is defined (through study) as a lateral distance in which a motorist on a recoverable slope may travel outside of the travelway and return their vehicle safely to the roadway. This distance is commonly determined as the 85th percentile in a study comparable to the method of determining speed limits on roadways through speed studies and varies based on the classification of a roadway. In order to provide for adequate safety in roadside conditions, hazardous elements, whether they be obstacles or steep slopes can be placed outside of the clearzone in order to reduce or eliminate the need for roadside protection.

Common sites for installation of traffic barrier:

- Bridge ends
- Near steep slopes from roadway limits

- At drainage crossings or culverts where steep or vertical drops are present
- Near large signs/illumination poles or other roadside elements which may pose hazards

When barrier is needed, careful calculations are completed to determine length of need which take into account the aforementioned factors. Specifically, the traffic volumes and therefore, the classification of the roadway in addition to the distance of the hazard from the edge of travelway and the distance or offset of the barrier to be placed or installed from the edge of travelway. It is the case in current times, that barrier or rail that is to be used in construction and maintenance operations has undergone extensive testing in both government and private research facilities in order to determine proper 'crash-worthiness' and effectiveness in conditions which are prescribed for its use. In particular, most roadside protection, whether it be a concrete barrier or rail, or a metal beam fence will perform properly only when placed in adequate proximity to the travelway so as to prevent vehicle impacts at large (obtuse) angles. The method in which a barrier protects motorists from roadside hazards is in how it dissipates the energy of an impact.

Barrier types and performance

Traffic barriers are categorized in two ways: by the function they serve, and by how much they deflect when a vehicle crashes into them.

Barrier functions

Roadside barriers are used to protect traffic from roadside obstacles or hazards, such as slopes steep enough to cause rollover crashes, fixed objects like bridge piers, and bodies of water. Roadside barriers can also be used wide medians, to prevent vehicles from colliding with hazards within the median.

Median barriers are used to prevent vehicles from crossing over a median and striking an oncoming vehicle in a head-on crash. Unlike roadside barriers, they must be designed to be struck from either side.

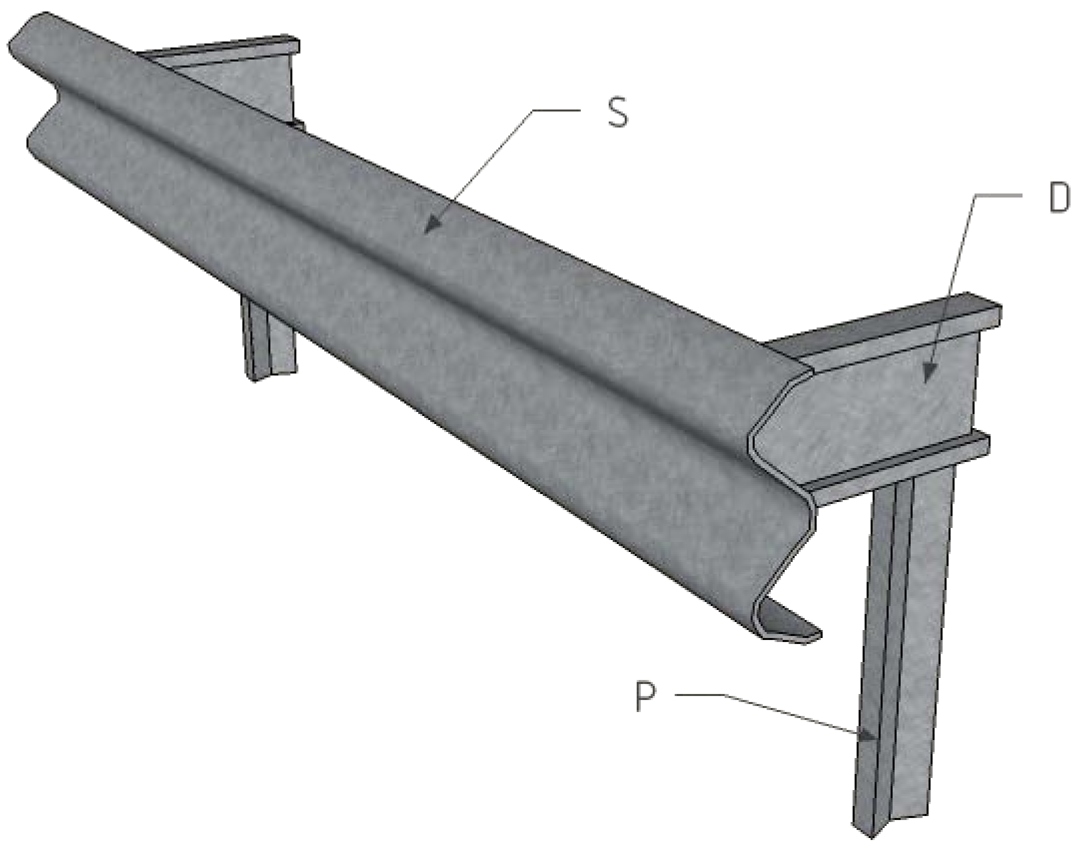
Bridge barrier is designed to restrain vehicles from crashing off the side of a bridge and falling onto the roadway, river or railroad below. It is usually higher than roadside barrier, to prevent trucks, buses, pedestrians and cyclists from vaulting or rolling over the barrier and falling over the side of the structure. Bridge rails are usually multi-rail tubular steel barriers or reinforced concrete parapets and barriers

Work zone barriers are used to protect traffic form hazards in work zones. Their distinguishing feature is they can be relocated as conditions change in the road works. Two common types are used: temporary concrete barrier and water-filled barrier. The latter is composed of steel-reinforced plasic boxes that are put in place where needed, linked together to form a longitudinal barrier, then ballasted with water. These have an advantage in that they can be assembled without heavy lifting equipment, but they cannot be used in freezing weather.

Barrier stiffness

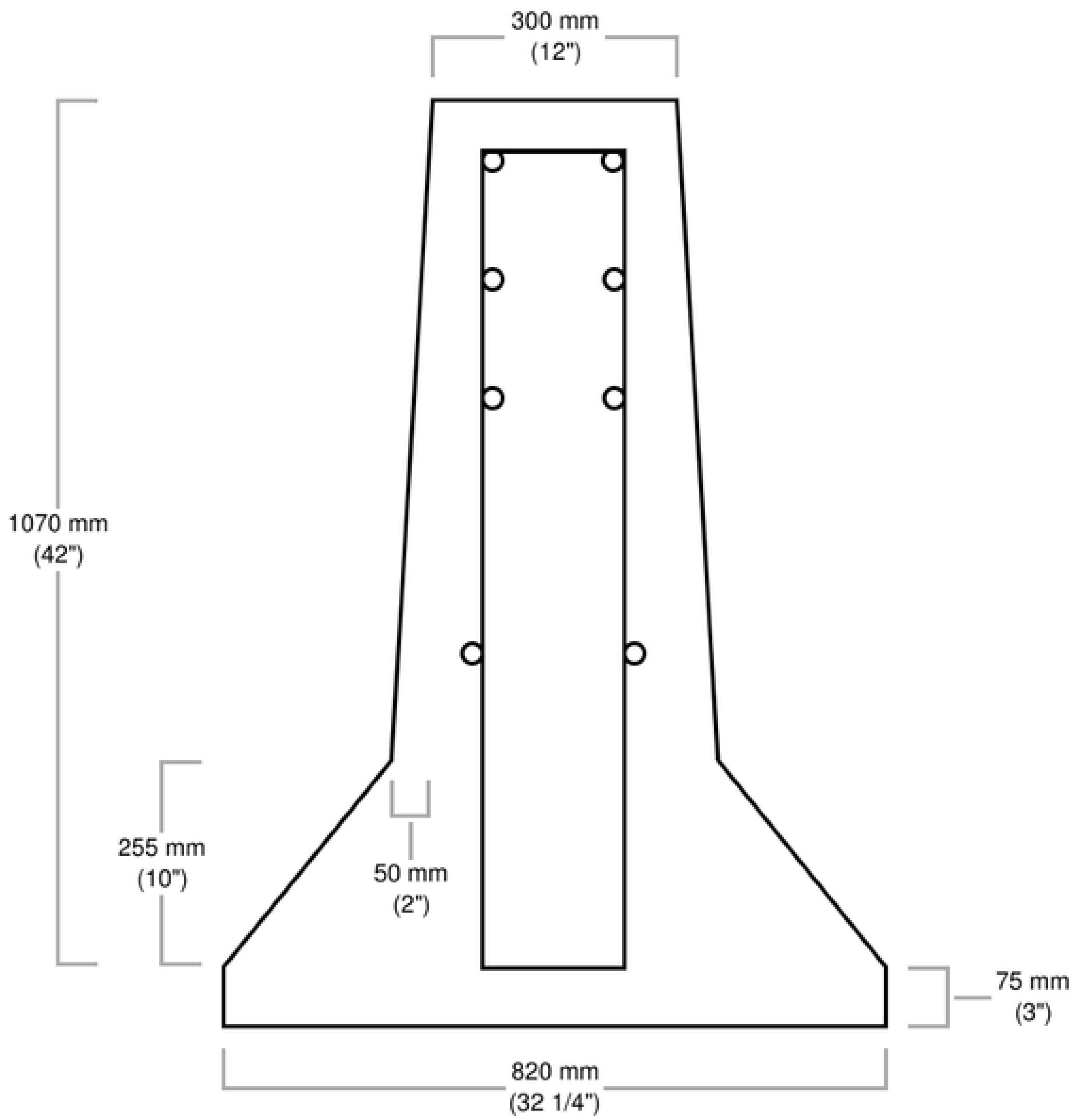
Barriers are divided into three groups, based on the amount they deflect when struck by a vehicle and the mechanism the barrier uses to resist the impact forces. In the United States, traffic barriers are tested and classified according to the AASHTO Manual for Assessing Safety Hardware (MASH) standards, which recently superseded Federal Highway Administration NCHRP Report 350. Barrier deflections listed below are results from crash tests with a 2000 kg pickup truck traveling 100 km/h, colliding with the rail at a 25 degree angle.

Flexible barriers include cable barriers and weak post corrugated guide rail systems. These are referred to as flexible barriers because they will deflect 1.6 m to 2.6 m when struck by a typical passenger car or light truck. Impact energy is dissipated through tension in the rail elements, deformation of the rail elements, posts, soil and vehicle bodywork, and friction between the rail and vehicle.



Components of a standard guardrail (A-profile): S – guardrail, D – distance piece/spacer, P – sigma post

Semi-rigid barriers include box beam guide rail, heavy post blocked out corrugated guide rail and thrie-beam guide rail. They deflect three to six feet: more than rigid barriers, but less than flexible barriers. Impact energy is dissipated through deformation of the rail elements, posts, soil and vehicle bodywork, and friction between the rail and vehicle. Box beam systems also spread the impact force over a number of posts due to the stiffness of the steel tube. These barrier systems deflect up to 1.5 m.



42 inches (110 cm)-high version of the Jersey barrier for deflecting automobiles and semi-trailer trucks.

Rigid barriers are usually constructed of reinforced concrete. A permanent concrete barrier will only deflect a negligible amount when struck by a vehicle. Instead, the shape of a concrete barrier is designed to redirect a vehicle into a path parallel to the barrier. This means they can be used to protect traffic from hazards very close behind the barrier, and generally require very little maintenance. Impact energy is dissipated through redirection and deformation of the vehicle itself. Jersey barriers and F-shape barriers also lift the vehicle as the tires ride up on the angled lower section. For low-speed or low-angle impacts on these barriers, that may be sufficient to redirect the vehicle without damaging the bodywork. The disadvantage is there is a higher likelihood of rollover with a small car than the single slope or step barriers. Impact forces are resisted by a combination of the rigidity and mass of the barrier. Deflection is usually negligible.

An early concrete barrier design was developed by the New Jersey State Highway Department. This led to the term Jersey barrier being used as a generic term, although technically it applies to a specific shape of concrete barrier. Other types include constant slope barriers, concrete step barriers and F-shape barriers.

Concrete barriers usually have smooth finishes. At some impact angles, coarse finishes allow the drive wheel of front wheel drive vehicles to climb the barrier, potentially causing the vehicle to roll over. However, along parkways and other areas where aesthetics are considered important, reinforced concrete walls with stone veneers or faux stone finishes are sometimes used. These barrier walls usually have vertical faces to prevent vehicles from climbing the barrier.

Barrier end treatments

Early traffic barrier designs often paid little attention to the ends of the barriers. Vehicles that struck blunt ends often stop abruptly or have steel rail sections penetrate into the passenger compartment. This often results in a severe injuries or fatalities.

As a result, barrier terminals were developed that brought the end of the terminal down to ground level. This prevented the rail from penetrating the vehicle, but could vault a vehicle into the air or cause it to roll over, since the barrier end formed a ramp.

To address the vaulting and rollover crashes, energy absorbing terminals were developed. These have a large steel impact head that engages the frame or bumper of the vehicle. The impact head is driven back along the guide rail, dissipating the vehicle's kinetic energy by bending or tearing the steel in the guide rail sections.

The final way to terminate a guide rail is bend it back to the point that the terminal is unlikely to be hit end-on, and, if possible, embed the end in a hillside or cut slope.

Chapter 9

Traffic Enforcement Camera

Gatso speed camera

A **traffic enforcement camera** (also **road safety camera**, **road rule camera**, **photo radar**, **speed camera**, **Gatso**) is an automated ticketing machine. It may include a camera which may be mounted beside on over a highway or installed in an enforcement vehicle to detect traffic regulation violations, including speeding, vehicles going through a red traffic light, unauthorized use of a bus lane, for recording vehicles inside a congestion charge area and others.

The latest automatic number plate recognition systems can be used for the detection of average speeds and raise concerns over loss of privacy and the potentially for governments to establish mass surveillance of vehicle movements and therefore by association also the movement of the vehicle's owner. Vehicles owners are often required by law to identify the driver of the vehicle and a case was taken to the European Court of Human Rights who found that the Human Rights Act 1998 was not being breached. Some groups, such as the National Motorists Association in the USA, claim that systems "encourage ... revenue-driven enforcement" rather than the declared objectives.

Types

Automatic speed enforcement gantry or "*Lombada Eletrônica*" with ground sensors at Brasilia, D.F.

Gatso Mobile Speed Camera, used in Victoria, Australia. The camera is mounted on the passenger side dash, whilst the black box on the front is the radar unit.

Bus lane enforcement

Some bus lane enforcement cameras use a sensor in the road which triggers a number plate recognition camera which compares the vehicle registation plate with a list of approved vehicles and records images of other vehicles. Other systems use a camera mounted on the bus, for example in London where they monitor Red routes on which stopping is not allowed for any purpose (other than taxis and disabled parking permit holders).

On Monday, February 23, 2009, New York City announced testing camera enforcement of bus lanes on 34th Street in Midtown Manhattan where a New York City taxi illegally using the bus lanes would face a fine of $150 adjudicated by the New York City Taxi and Limousine Commission.

Red light enforcement

Red light camera in Springfield, Ohio, USA.

A red light camera is a traffic camera that takes an image of a vehicle that goes through an intersection where the light is red. The system continuously monitors the traffic signal and the camera is triggered by any vehicle entering the intersection above a preset minimum speed and following a specified time after the signal has turned red.

Speed limit enforcement

Speed enforcement cameras are used to monitor compliance with speed limits which may use Doppler, LIDAR or Automatic number plate recognition. Other speed enforcement systems are also used which are not camera based.

Fixed or mobile speed camera systems that measure the time taken by a vehicle to travel between two or more fairly distant sites (from several hundred metres to several hundred kilometres apart) are called automatic number plate recognition (ANPR) cameras. These cameras time vehicles over a known fixed distance, then calculate the vehicle's average speed for the journey. The name derives from the fact that the technology uses infrared cameras linked to a computer to "read" a vehicle's registration number and identify it in real-time.

Number plate recognition systems

Automatic number plate recognition can be used for purposes unrelated to enforcement of traffic rules. In principle any agency or person with access to data either from traffic cameras or cameras installed for other purposes can track the movement of vehicles for any purpose.

In Australia's SAFE-T-CAM system, ANPR technology is used to monitor long distance truck drivers to detect avoidance of legally prescribed driver rest periods.

The United Kingdom's police ANPR system logs all the vehicles passing particular points in the national road network, allowing authorities to track the movement of vehicles and individuals across the country.

In the UK an 80-year-old pensioner John Catt and his daughter Linda (with no criminal record between them) were stopped by City of London Police while driving in London, UK in 2005, had their vehicle searched under section 44 of the Terrorism Act 2000 and were threatened with arrest if they refused to answer questions. After they complained formally, it was discovered they were stopped when their car was picked up by roadside ANPR CCTV cameras; it had been flagged in the Police National Computer database when they were seen near EDO MBM demonstrations in Brighton. Critics point out that the Catts had been suspected of no crime, however the UK's mass surveillance infrastructure allowed them to be targeted due to their association.

Other

- Congestion charge cameras to detect vehicles inside the chargeable area which have not paid the appropriate fee
- High-occupancy vehicle lane cameras to identify vehicles violating occupancy requirements.
- Level crossing cameras to identifying vehicles crossing railways at grade
- Noise pollution cameras that record evidence of heavy vehicles that break noise regulations by using engine braking
- Parking cameras which issue citations to vehicles which are illegally parked or which were not moved from a street at posted times.
- Toll-booth cameras to identify vehicles proceeding through a toll booth without paying the toll
- Turn cameras at intersections where specific turns are prohibited on red. This type of camera is mostly used in cities or heavy populated areas.
- Automatic number plate recognition systems can be used for multiple purposes, including identifying untaxed and uninsured vehicles, stolen cars and potentially mass surveillance of motorists.

Fixed camera systems can mounted in boxes or on poles beside the road or attached to gantries over the road, or to overpasses or bridges. Cameras can be concealed, for example in garbage bins.

Mobile speed cameras may be hand-held, tripod mounted, or vehicle-mounted. In vehicle-mounted systems, detection equipment and cameras can be mounted to the vehicle itself, or simply tripod mounted inside the vehicle and deployed out a window or door. If the camera is fixed to the vehicle, the enforcement vehicle does not necessarily have to be stationary, and can be moved either with or against the flow of traffic. In the latter case, depending on the direction of travel, the target vehicle's relative speed is either added or subtracted from the enforcement vehicle's own speed to obtain its actual speed. The speedometer of the camera vehicle needs to be accurately calibrated.

Some number plate recognition systems can be used from vehicles.

Controversy

Legal issues

There are a number of legal issues which arise as a result depending on local laws and the procedures used by the enforcing bodies. Various legal issues arise from such cameras and the laws involved in how cameras can be placed and what evidence is necessary to prosecute a driver varies considerably in different legal systems.

One issue is the potential conflict of interest when private contractors are paid a commission based on the number of tickets they are able to issue. Pictures from the San Diego red light camera systems were ruled inadmissible as court evidence in September 2001. The judge said that the "total lack of oversight" and "method of compensation" made evidence from the cameras "so untrustworthy and unreliable that it should not be admitted".

Some U.S. states and provinces of Canada such as Alberta operate "owner liability" where it is the registered owner of the vehicle who must pay all such fines regardless of whether he was driving at the time of the offense, although they do release the owner from liability if he signs a form identifying the actual driver and that individual pays the fine. These states do not issue demerit points for camera infractions which has has been criticized by some as giving a "license to speed" to those who can more easily afford speeding fines.

In Albuquerque, New Mexico, the city government attempted to bypass the legal issue of a defendant's right to cross-examine his accuser, as well as the issue of verifying the driver's identity. Automated red-light and speeding offenses are classed as public nuisances and fined to the vehicle's registered owner as civil violations, not as criminal offenses.

In April 2000 two motorists who were caught speeding in the United Kingdom challenged the Road Traffic Act 1988 which required the keeper of a driver to identify the driver at a particular time as being in contradiction to the Human Rights Act 1998 on the grounds that it amounted to a 'compulsory confession', also that since the camera partnerships included the police, local authorities, Magistrates Courts Service (MCS) and Crown Prosecution Service (CPS) which had a financial interest in the fine revenue that they would not get a fair trial. Their plea was initially granted by a judge then overturned but was the heard by the European Court of Human Rights (ECtHR), and the European Court of Justice (ECJ). In 2007 the European Court of Human Rights found there was no breach of article 6 in requiring the keepers of cars caught speeding on camera to provide the name of the driver.

Surveillance

- Police and government have been accused of "Big Brother tactics" in over-monitoring of public roads, and of "revenue raising" in applying cameras in deceptive ways to increase government revenue rather than improve road safety.

Revenue not safety

- In 2010 a campaign was set up against a speed camera on a dual carriageway in Poole, Dorset in a 30 mph area in the United Kingdom. which had generated £1.3m of fines every year since 1999. The initial Freedom of information request was refused and the information was only released after an appeal to the Information Commissioner.
- In May 2010 the new Coalition government said that the 'Labour's 13-year war on the motorist is over' and that the new government 'pledged to scrap public funding for speed cameras' In July Mike Penning, the Road safety minister reduced the Road Safety Grant for the current year to Local Authorities from £95 million to £57 million saying that local authorities had relied too heavily on safety cameras for far too long and that he was pleased that some councils were now focusing on other road safety measures. It is estimated that the as a result the Treasury is now distributing £40 million less in Road Safety Grant than is raised from fines in the year. Dorset and Essex announced plans to review camera provision with a view to possibly ending the scheme in their counties, however Dorset strongly affirmed its support for the scheme, albeit reducing financial contributions in line with the reduction in government grant. Seven counties also announced plans to turn off some or all of their cameras, amidst warnings from the country's most senior traffic policeman that this would result in an increase in deaths and injuries. Gloucestershire cancelled plans to update cameras and has reduced or cancelled maintenance contracts.

Unpopularity

Use of cameras is opposed by some motorists and motoring organisations. They have also be rejected in some places by referendum.

- The first speed camera systems in the USA was in Friendswood, Texas in 1986 and La Marque, Texas in 1987. Neither program lasted more than a few months before public pressure forced them to be dropped.
- In 1991 cameras have been rejected by voters in referenda in Peoria, Arizona voters were the first to reject cameras by a 2-1 margin. Speed cameras have since been installed on the highways in the Phoenix area since 2007.
- In 1992 cameras have been rejected by voters in referenda in Batavia, Illinois.
- Anchorage, Alaska rejected cameras in a 1997 referendum
- In 2002 the state of Hawaii experimented with speed limit enforcement vans but they were withdrawn months later due to public outcry.
- In 2005, the Virginia legislature declined to reauthorize its red light camera enforcement law after a study questioned their effectiveness, only to reverse itself in 2007 and allow cameras to return to any city with a population greater than 10,000.
- Steubenville, Ohio rejected cameras in a 2006 referendum.
- In 2009, a petition was started in the town of College Station, Texas which requested that all red light cameras be dismantled and removed from all of the town's intersections. Enough signatures were captured to put the measure on the November 2009 general election ballot. After an extensive battle between the College Station city council and the opposing sides, both for and against red light cameras, the voters voted to eliminate the red light cameras throughout the entire city. By the end of November the red light cameras were taken down. However, all citations issued are still valid and must be paid by the offenders.
- On May 4, 2010 an ordinance authorizing the use of speed cameras in the town of Sykesville, Maryland was put to a referendum, in which 321 out of 529 voters (60.4%) voted against the cameras. The turnout for this vote was greater than the number of voters in the previous local Sykesville election for mayor where 523 residents voted.
- Arizona decided to not renew their contract with Redflex in 2011 following a study of their statewide 76 photo enforcement cameras. Reasons given included less than expected revenue due to improved compliance, mixed public acceptance and mixed accident data.

Effectiveness

- The town of Swindon abandoned the use of fixed cameras in 2009, questioning their cost effectiveness with the cameras being replaced by vehicle activated warning signs and enforcement by police using mobile speed cameras: in the nine months following the switch-off there was a small reduction in accident rates which had changed slightly in similar periods before and after the switch off (Before: 1 fatal, 1 serious and 13 slight accidents. Afterwards: no fatalities, 2 serious and 12 slight accidents). The journalist George Monbiot claimed that the results were not statistically significant highlighting earlier findings across the whole of Wiltshire that there had been a 33% reduction in the number of people killed and seriously injured generally and a 68% reduction at camera sites during the previous 3 years.

- In January 2011 Edmonton, Alberta cancelled all 100,000 "Speed On Green" tickets issued in the previous 14 months due to concerns about camera reliability.

Avoidance/evasion

A GPS map showing speed camera POI information overlaid onto it

To avoid detection or prosecution drivers may:

- Brake just before a camera in order to travel past its sensor below the speed limit. This is however a cause of collisions.
- Use GPS navigation devices which contain databases of known camera locations to alert them in advance. These databases may in some cases be update in near-realtime. The use of GPS devices to locate speed cameras is illegal in some jurisdictions.
- Install passive laser detectors or radar detectors that detect when the vehicle's speed is being monitored and warn the driver. Use of these devices may be illegal in some jurisdictions.
- Install active laser jammer or radar jammer devices which actively transmit signals that interfere with the measuring device. These devices are illegal in many jurisdictions.
- Remove, falsify, obscure or modify vehicle license plate. Tampering with number plates is illegal in many jurisdictions.

In August 2010 a fast driving Swedish driver reportedly avoided several older model speed cameras, but was detected by a new model, as traveling at 186 mph (300 km/h), resulting in the world's largest speeding fine to date.

- In the past it was possible to avoid detection by changing lanes when SPECS average speed cameras were in use as they measured a vehicle's speed over distance in one lane only. As of 2011 the cameras are type approved to cover multiple lanes.

History

Older traffic enforcement camera in Ludwigsburg, Germany

The concept of the speed camera can be dated back to at least 1905; Popular Mechanics reports on a patent for a "Time Recording Camera for Trapping Motorists" that enabled the operator to take time-stamped images of a vehicle moving across the start and endpoints of a measured section of road. The timestamps enabled the speed to be calculated, and the photo enabled identification of the driver.

The Dutch company *Gatsometer BV*, which was founded in 1958 by rally driver Maurice Gatsonides, produced the 'Gatsometer'. Gatsonides wished to better monitor his average speed on a race track and invented the device in order to improve his lap times. The company later started supplying these devices as police speed enforcement tools. The first systems introduced in the late 1960s used film cameras to take their pictures. Gatsometer introduced the first red light camera in 1965, the first radar for use with road traffic in 1971 and the first mobile speed traffic camera in 1982;

From the late 1990s, digital cameras began to be introduced. Digital cameras can be fitted with a network connection to transfer images to a central processing location automatically, so they have advantages over film cameras in speed of issuing fines, maintenance and operational monitoring. However, film-based systems may provide superior image quality in the variety of lighting conditions encountered on roads, and are required by courts in some jurisdictions. New film-based systems are still being sold, but digital pictures are providing greater versatility and lower maintenance and are now more popular with law enforcement agencies.

A red-light and speed camera in Darwin, Northern Territory, Australia

Dazzle camouflaged speed camera in Loipersdorf, Austria

A red-light camera in use in Beaverton, Oregon, USA

Chapter 10

Intelligent Transportation System

The term ***intelligent transport system*** (ITS) refers to efforts to add information and communications technology to transport infrastructure and vehicles in an effort to manage factors that typically are at odds with each other, such as vehicles, loads, and routes to improve safety and reduce vehicle wear, transportation times, and fuel consumption.

Background

Interest in ITS comes from the problems caused by traffic congestion and a synergy of new information technology for simulation, real-time control, and communications networks. Traffic congestion has been increasing worldwide as a result of increased motorization, urbanization, population growth, and changes in population density. Congestion reduces efficiency of transportation infrastructure and increases travel time, air pollution, and fuel consumption.

The United States, for example, saw large increases in both motorization and urbanization starting in the 1920s that led to migration of the population from the sparsely populated rural areas and the densely packed urban areas into suburbs. The industrial economy replaced the agricultural economy, leading the population to move from rural locations into urban centers. At the same time, motorization was causing cities to expand because motorized transportation could not support the population density that the existing mass transit systems could. Suburbs provided a reasonable compromise between population density and access to a wide variety of employment, goods, and services that were available in the more densely populated urban centers. Further, suburban infrastructure could be built quickly, supporting a rapid transition from a rural/agricultural economy to an industrial/urban economy.

Recent governmental activity in the area of ITS – specifically in the United States – is further motivated by the perceived need for homeland security. Many of the proposed ITS systems also involve surveillance of the roadways, which is a priority of homeland security. Funding of many systems comes either directly through homeland security organizations or with their approval. Further, ITS can play a role in the rapid mass evacuation of people in urban centers after large casualty events such as a result of a natural disaster or threat. Much of the infrastructure and planning involved with ITS parallels the need for homeland security systems.

In the developing world, the migration of people from rural to urbanized habitats has progressed differently. Many areas of the developing world have urbanized without significant motorization and the formation of suburbs. In areas like Santiago, Chile, a high population density is supported by a multimodal system of walking, bicycle transportation, motorcycles, buses, and trains. A small portion of the population can afford automobiles, but the automobiles greatly increase the congestion in these multimodal transportation systems. They also produce a considerable amount of air pollution, pose a significant safety risk, and exacerbate feelings of inequities in the society.

Other parts of the developing world, such as China, remain largely rural but are rapidly urbanizing and industrializing. In these areas a motorized infrastructure is being developed alongside motorization of the population. Great disparity of wealth means that only a fraction of the population can motorize, and therefore the highly dense multimodal transportation system for the poor is cross-cut by the highly motorized transportation system for the rich. The urban infrastructure is being rapidly developed, providing an opportunity to build new systems that incorporate ITS at early stages.

Intelligent transport technologies

Intelligent transport systems vary in technologies applied, from basic management systems such as car navigation; traffic signal control systems; container management systems; variable message signs; automatic number plate recognition or speed cameras to monitor applications, such as security CCTV systems; and to more advanced applications that integrate live data and feedback from a number of other sources, such as parking guidance and information systems; weather information; bridge deicing systems; and the like. Additionally, predictive techniques are being developed to allow advanced modeling and comparison with historical baseline data. Some of the constituent technologies typically implemented in ITS are described in the following sections.

Wireless communications

Various forms of wireless communications technologies have been proposed for intelligent transportation systems.

Radio modem communication on UHF and VHF frequencies are widely used for short and long range communication within ITS.

Short-range communications (less than 500 yards) can be accomplished using IEEE 802.11 protocols, specifically WAVE or the Dedicated Short Range Communications standard being promoted by the Intelligent Transportation Society of America and the

United States Department of Transportation. Theoretically, the range of these protocols can be extended using Mobile ad-hoc networks or Mesh networking.

Longer range communications have been proposed using infrastructure networks such as WiMAX (IEEE 802.16), Global System for Mobile Communications (GSM), or 3G. Long-range communications using these methods are well established, but, unlike the short-range protocols, these methods require extensive and very expensive infrastructure deployment. There is lack of consensus as to what business model should support this infrastructure.

Computational technologies

Recent advances in vehicle electronics have led to a move toward fewer, more capable computer processors on a vehicle. A typical vehicle in the early 2000s would have between 20 and 100 individual networked microcontroller/Programmable logic controller modules with non-real-time operating systems. The current trend is toward fewer, more costly microprocessor modules with hardware memory management and Real-Time Operating Systems. The new embedded system platforms allow for more sophisticated software applications to be implemented, including model-based process control, artificial intelligence, and ubiquitous computing. Perhaps the most important of these for Intelligent Transportation Systems is artificial intelligence.

Floating car data/floating cellular data

"Floating car" or "probe" data collection is a set of relatively low-cost methods for obtaining travel time and speed data for vehicles traveling along streets, highways, freeways, and other transportation routes. Broadly speaking, three methods have been used to obtain the raw data:

- **Triangulation Method.** In developed countries a high proportion of cars contain one or more mobile phones. The phones periodically transmit their presence information to the mobile phone network, even when no voice connection is established. In the mid 2000s, attempts were made to use mobile phones as anonymous traffic probes. As a car moves, so does the signal of any mobile phones that are inside the vehicle. By measuring and analyzing network data using triangulation, pattern matching or cell-sector statistics (in an anonymous format), the data was converted into traffic flow information. With more congestion, there are more cars, more phones, and thus, more probes. In metropolitan areas, the distance between antennas is shorter and in theory accuracy increases. An advantage of this method is that no infrastructure needs to be built along the road; only the mobile phone network is leveraged. But in practice the triangulation method can be complicated, especially in areas where the same mobile phone towers serve two or more parallel routes (such as a freeway with a frontage road, a freeway and a commuter rail line, two or more parallel streets, or a street that is also a bus line). By the early 2010s, the popularity of the triangulation method was declining.

- **Vehicle Re-Identification.** Vehicle re-identification methods require sets of detectors mounted along the road. In this technique, a unique serial number for a device in the vehicle is detected at one location and then detected again (re-identified) further down the road. Travel times and speed are calculated by comparing the time at which a specific device is detected by pairs of sensors. This can be done using the MAC (Machine Access Control) addresses from Bluetooth devices, or using the RFID serial numbers from Electronic Toll Collection (ETC) transponders (also called "toll tags").
- **GPS Based Methods.** An increasing number of vehicles are equipped with in-vehicle GPS (sattelite navigation) systems that have two-way communication with a traffic data provider. Position readings from these vehicles are used to compute vehicle speeds.

Floating car data technology provides advantages over other methods of traffic measurement:

- Less expensive than sensors or cameras
- More coverage (potentially including all locations and streets)
- Faster to set up and less maintenance
- Works in all weather conditions, including heavy rain

Sensing technologies

Technological advances in telecommunications and information technology, coupled with state-of-the-art microchip, RFID (Radio Frequency Identification), and inexpensive intelligent beacon sensing technologies, have enhanced the technical capabilities that will facilitate motorist safety benefits for intelligent transportation systems globally. Sensing systems for ITS are vehicle- and infrastructure-based networked systems, i.e., Intelligent vehicle technologies. Infrastructure sensors are indestructible (such as in-road reflectors) devices that are installed or embedded in the road or surrounding the road (e.g., on buildings, posts, and signs), as required, and may be manually disseminated during preventive road construction maintenance or by sensor injection machinery for rapid deployment. Vehicle-sensing systems include deployment of infrastructure-to-vehicle and vehicle-to-infrastructure electronic beacons for identification communications and may also employ video automatic number plate recognition or vehicle magnetic signature detection technologies at desired intervals to increase sustained monitoring of vehicles operating in critical zones.

Inductive loop detection

Inductive loops can be placed in a roadbed to detect vehicles as they pass through the loop's magnetic field. The simplest detectors simply count the number of vehicles during a unit of time (typically 60 seconds in the United States) that pass over the loop, while more sophisticated sensors estimate the speed, length, and weight of vehicles and the distance between them. Loops can be placed in a single lane or across multiple lanes, and they work with very slow or stopped vehicles as well as vehicles moving at high-speed.

Video vehicle detection

Traffic flow measurement and automatic incident detection using video cameras is another form of vehicle detection. Since video detection systems such as those used in automatic number plate recognition do not involve installing any components directly into the road surface or roadbed, this type of system is known as a "non-intrusive" method of traffic detection. Video from black-and-white or color cameras is fed into processors that analyze the changing characteristics of the video image as vehicles pass. The cameras are typically mounted on poles or structures above or adjacent to the roadway. Most video detection systems require some initial configuration to "teach" the processor the baseline background image. This usually involves inputting known measurements such as the distance between lane lines or the height of the camera above the roadway. A single video detection processor can detect traffic simultaneously from one to eight cameras, depending on the brand and model. The typical output from a video detection system is lane-by-lane vehicle speeds, counts, and lane occupancy readings. Some systems provide additional outputs including gap, headway, stopped-vehicle detection, and wrong-way vehicle alarms.

Intelligent transport applications

Emergency vehicle notification systems

The in-vehicle eCall is an emergency call generated either manually by the vehicle occupants or automatically via activation of in-vehicle sensors after an accident. When activated, the in-vehicle eCall device will establish an emergency call carrying both voice and data directly to the nearest emergency point (normally the nearest E1-1-2 Public-safety answering point, PSAP). The voice call enables the vehicle occupant to communicate with the trained eCall operator. At the same time, a minimum set of data will be sent to the eCall operator receiving the voice call.

The minimum set of data contains information about the incident, including time, precise location, the direction the vehicle was traveling, and vehicle identification. The pan-European eCall aims to be operative for all new type-approved vehicles as a standard option. Depending on the manufacturer of the eCall system, it could be mobile phone based (Bluetooth connection to an in-vehicle interface), an integrated eCall device, or a functionality of a broader system like navigation, Telematics device, or tolling device. eCall is expected to be offered, at earliest, by the end of 2010, pending standardization by the European Telecommunications Standards Institute and commitment from large EU member states such as France and the United Kingdom.

Congestion pricing gantry at North Bridge Road, Singapore.

Automatic road enforcement

Automatic speed enforcement gantry or "*Lombada Eletrônica*" with ground sensors at Brasilia, D.F.

A traffic enforcement camera system, consisting of a camera and a vehicle-monitoring device, is used to detect and identify vehicles disobeying a speed limit or some other road legal requirement and automatically ticket offenders based on the license plate number. Traffic tickets are sent by mail. Applications include:

- Speed cameras that identify vehicles traveling over the legal speed limit. Many such devices use radar to detect a vehicle's speed or electromagnetic loops buried in each lane of the road.
- Red light cameras that detect vehicles that cross a stop line or designated stopping place while a red traffic light is showing.
- Bus lane cameras that identify vehicles traveling in lanes reserved for buses. In some jurisdictions, bus lanes can also be used by taxis or vehicles engaged in car pooling.
- Level crossing cameras that identify vehicles crossing railways at grade illegally.
- Double white line cameras that identify vehicles crossing these lines.
- High-occupancy vehicle lane cameras for that identify vehicles violating HOV requirements.
- Turn cameras at intersections where specific turns are prohibited on red. This type of camera is mostly used in cities or heavy populated areas.

Variable speed limits

Example variable speed limit sign in the United States.

Recently some jurisdictions have begun experimenting with variable speed limits that change with road congestion and other factors. Typically such speed limits only change to decline during poor conditions, rather than being improved in good ones. One example is on Britain's M25 motorway, which circumnavigates London. On the most heavily-traveled 14-mile (23 km) section (junction 10 to 16) of the M25 variable speed limits combined with automated enforcement have been in force since 1995. Initial results indicated savings in journey times, smoother-flowing traffic, and a fall in the number of accidents, so the implementation was made permanent in 1997. Further trials on the M25 have been thus far proved inconclusive.

Collision avoidance systems

Japan has installed sensors on its highways to notify motorists that a car is stalled ahead.

Dynamic Traffic Light Sequence

Intelligent RFID traffic control has been developed for dynamic traffic light sequence. It circumvents or avoids problems that usually arise with systems that use image processing and beam interruption techniques. RFID technology with appropriate algorithm and database were applied to a multi vehicle, multi lane and multi road junction area to provide an efficient time management scheme. A dynamic time schedule was worked out for the passage of each column. The simulation has shown that, the dynamic sequence algorithm has the ability to intelligently adjust itself even with the presence of some extreme cases. The real time operation of the system able to emulate the judgment of a traffic policeman on duty, by considering the number of vehicles in each column and the routing proprieties.

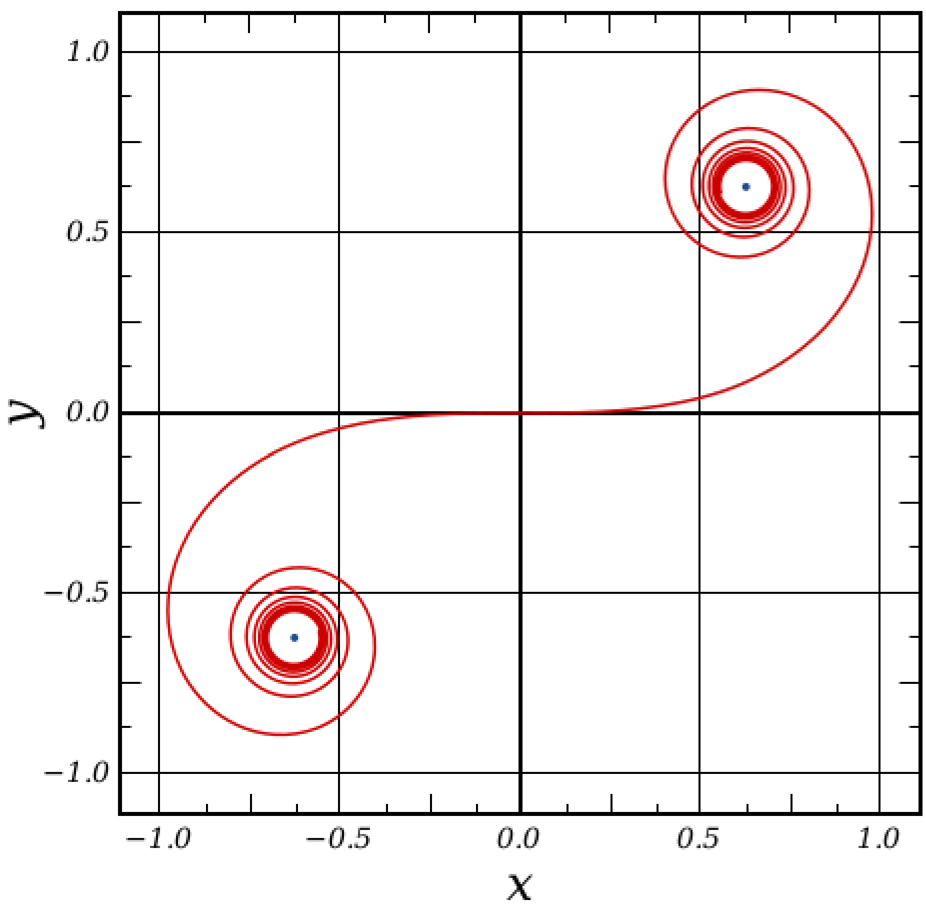
Cooperative systems on the road

Communication cooperation on the road includes car-to-car, car-to-infrastructure, and vice versa. Data available from vehicles is acquired and transmitted to a server for central fusion and processing. This data can be used to detect events such as rain (wiper activity) and congestion (frequent braking activities). The server processes a driving recommendation dedicated to a single or a specific group of drivers and transmits it wirelessly to vehicles. The goal of cooperative systems is to use and plan communication and sensor infrastructure to increase road safety. The definition of cooperative systems in road traffic is according to the European Commission:

> "Road operators, infrastructure, vehicles, their drivers and other road users will cooperate to deliver the most efficient, safe, secure and comfortable journey. The vehicle-vehicle and vehicle-infrastructure co-operative systems will contribute to these objectives beyond the improvements achievable with stand-alone systems."

Chapter 11

Euler Spiral



A double-end Euler spiral.

An **Euler spiral** is a curve whose curvature changes linearly with its curve length (the curvature of a circular curve is equal to the reciprocal of the radius). Euler spirals are also commonly referred to as **spiros**, **clothoids** or **Cornu spirals**.

Euler spirals have applications to diffraction computations. They are also widely used as transition curve in railroad/highway engineering for connecting and transiting the geometry between a tangent and a circular curve. The principle of linear variation of the curvature of the transition curve between a tangent and a circular curve defines the geometry of the Euler spiral:

- Its curvature begins with zero at the straight section (the tangent) and increases linearly with its curve length.
- Where the Euler spiral meets the circular curve, its curvature becomes equal to that of the latter.

Applications

Track transition curve

An object traveling on a circular path experiences a centripetal acceleration. When a vehicle traveling on a straight path approaches a circular path, it experiences a sudden centripetal acceleration starting at the tangent point; and thus centripetal force acts instantly causing much discomfort.

On early railroads this instant application of lateral force was not an issue since low speeds and wide-radius curves were employed (lateral forces on the passengers and the lateral sway was small and tolerable). As speeds of rail vehicles increased over the years, it became obvious that an easement is necessary so that the centripetal acceleration increases linearly with the traveled distance. Given the expression of centripetal acceleration V^2 / R, the obvious solution is to provide an easement curve whose curvature, $1 / R$, increases linearly with the traveled distance. This geometry is an Euler spiral.

Unaware of the solution of the geometry by Leonhard Euler, Rankine cited the cubic curve (a polynomial curve of degree 3), which is an approximation of the Euler spiral for small angular changes in the same way that a parabola is an approximation to a circular curve.

Marie Alfred Cornu (and later some civil engineers) also solved the calculus of Euler spiral independently. Euler spirals are now widely used in rail and highway engineering for providing a transition or an easement between a tangent and a horizontal circular curve.

Optics

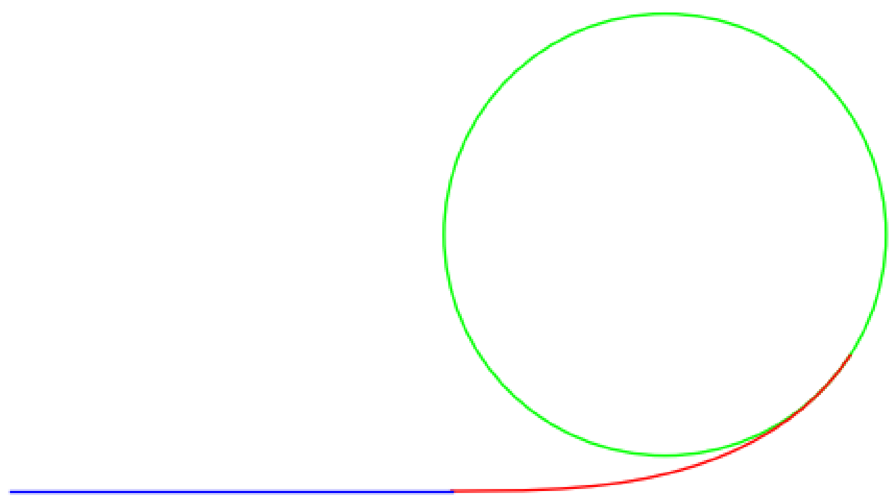
The Cornu spiral can be used to describe a diffraction pattern .

Formulation

Symbols

R	Radius of curvature
R_c	Radius of Circular curve at the end of the spiral
θ	Angle of curve from beginning of spiral (infinite *R*) to a particular point on the spiral.
	This can also be measured as the angle between the initial tangent and the tangent at the concerned point.
θ_s	Angle of full spiral curve
L, s	Length measured along the spiral curve from its initial position
L_s, s_o	Length of spiral curve

Derivation



The graph on the right illustrates an Euler spiral used as an easement (transition) curve between two given curves, in this case a straight line (the negative *x* axis) and a circle. The spiral starts at the origin in the positive *x* direction and gradually turns anticlockwise to osculate the circle.

The spiral is a small segment of the above double-end Euler spiral in the first quadrant.

From the definition of the curvature,

$$\frac{1}{R} = \frac{d\theta}{dL} \propto L$$

i.e.

$$RL = \text{constant} = R_c L_s$$

$$\frac{d\theta}{dL} = \frac{L}{R_c L_s}$$

We write in the format,

$$\frac{d\theta}{dL} = 2a^2 L$$

Where

$$2a^2 = \frac{1}{R_c L_s}$$

Or

$$a = \frac{1}{\sqrt{2R_c L_s}}$$

Thus

$$\theta = (aL)^2$$

Now

$$x = \int_0^L \cos\theta \, ds$$

$$= \int_0^L \cos\left[(as)^2\right] ds$$

If

$$L' = aL$$

Then

$$dL = \frac{dL'}{a}$$

Thus

$$x = \frac{1}{a} \int_0^{L'} \cos s^2 ds$$

$$y = \int_0^L \sin\theta \, ds$$

$$= \int_0^L \sin\left[(as)^2\right] ds$$

$$= \frac{1}{a} \int_0^{L'} \sin s^2 \, ds$$

Expansion of Fresnel integral

If a = 1, which is the case for normalized Euler curve, then the Cartesian coordinates are given by Fresnel integrals (or Euler integrals):

$$C(L) = \int_0^L \cos s^2 \, ds$$
$$S(L) = \int_0^L \sin s^2 \, ds$$

Expand C(L) according to power series expansion of cosine:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots$$
$$C(L) = \int_0^L \cos s^2 \, ds$$
$$= \int_0^L (1 - \frac{s^4}{2!} + \frac{s^8}{4!} - \frac{s^{12}}{6!} + \cdots) \, ds$$
$$= L - \frac{L^5}{5 \times 2!} + \frac{L^9}{9 \times 4!} - \frac{L^{13}}{13 \times 6!} + \cdots$$

Expand S(L) according to power series expansion of sine:

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots$$
$$S(L) = \int_0^L \sin s^2 \, ds$$
$$= \int_0^L (s^2 - \frac{s^6}{3!} + \frac{s^{10}}{5!} - \frac{s^{14}}{7!} + \cdots) \, ds$$
$$= \frac{L^3}{3} - \frac{L^7}{7 \times 3!} + \frac{L^{11}}{11 \times 5!} - \frac{L^{15}}{15 \times 7!} + \cdots$$

Normalization and conclusion

For a given Euler curve with:

$$2RL = 2R_cL_s = \frac{1}{a^2}$$

or

$$\frac{1}{R} = \frac{L}{R_cL_s} = 2a^2L$$

then

$$x = \frac{1}{a} \int_0^{L'} \cos s^2 \, ds$$
$$y = \frac{1}{a} \int_0^{L'} \sin s^2 \, ds$$

where $L' = aL$ and $a = \frac{1}{\sqrt{2R_c L_s}}$.

The process of obtaining solution of (x, y) of an Euler spiral can thus be described as:

- Map L of the original Euler spiral by multiplying with factor *a* to *L'* of the normalized Euler spiral;
- Find (x', y') from the Fresnel integrals; and
- Map (x', y') to (x, y) by scaling up (denormalize) with factor $1 / a$. Note that $1 / a > 1$.

In the normalization process,

$$R'_c = \frac{R_c}{\sqrt{2R_c L_s}} = \sqrt{\frac{R_c}{2L_s}}$$
$$L'_s = \frac{L_s}{\sqrt{2R_c L_s}} = \sqrt{\frac{L_s}{2R_c}}$$

Then

$$2R'_c L'_s = 2\sqrt{\frac{R_c}{2L_s}}\sqrt{\frac{L_s}{2R_c}} = \tfrac{2}{2} = 1$$

Generally the normalization reduces L' to a small value (<1) and results in good converging characteristics of the Fresnel integral manageable with only a few terms.

Illustration

Given:

$$R_c = 300\text{m}$$
$$L_s = 100\text{m}$$

Then

$$\begin{aligned}\theta_s &= \frac{L_s}{2R_c} \\ &= \frac{100}{2 \times 300} \\ &= 0.1667 \text{ radian}\end{aligned}$$

And

$$2R_cL_s = 60,000$$

We scale down the Euler spiral by √60,000, i.e.100√6 to normalized Euler spiral that has:

$$\begin{aligned}R'_c &= \tfrac{3}{\sqrt{6}}\text{m} \\ L'_s &= \tfrac{1}{\sqrt{6}}\text{m} \\ 2R'_cL'_s &= 2 \times \tfrac{3}{\sqrt{6}} \times \tfrac{1}{\sqrt{6}} \\ &= 1\end{aligned}$$

And

$$\begin{aligned}\theta_s &= \frac{L'_s}{2R'_c} \\ &= \frac{\frac{1}{\sqrt{6}}}{2 \times \frac{3}{\sqrt{6}}} \\ &= 0.1667 \text{ radian}\end{aligned}$$

The two angles θ_s are the same. This thus confirm that the original and normalized Euler spirals are having geometric similarity. The locus of the normalized curve can be determined from Fresnel Integral, while the locus of the original Euler spiral can be obtained by scaling back / up or denormalizing.

Other properties of normalized Euler spiral

Normalized Euler spiral can be expressed as:

$$x = \int_0^L \cos s^2 ds$$
$$y = \int_0^L \sin s^2 ds$$

Normalized Euler spiral has the following properties:

$$2R_c L_s = 1$$
$$\theta_s = \frac{L_s}{2R_c} = L_s^2$$

And

$$\theta = \theta_s \cdot \frac{L^2}{L_s^2} = L^2$$
$$\frac{1}{R} = \frac{d\theta}{dL} = 2L.$$

Note that $2R_cL_s = 1$ also means that $1 / R_c = 2L_s$, in agreement with the above.

Mathematical codes producing an Euler spiral

The following Sage code produces the second graph above. The first four lines express the Euler spiral component. Fresnel functions could not be found. Instead, the integrals of two expanded Taylor series are adopted. The remaining code expresses respectively the tangent and the circle, including the computation for the center coordinates.

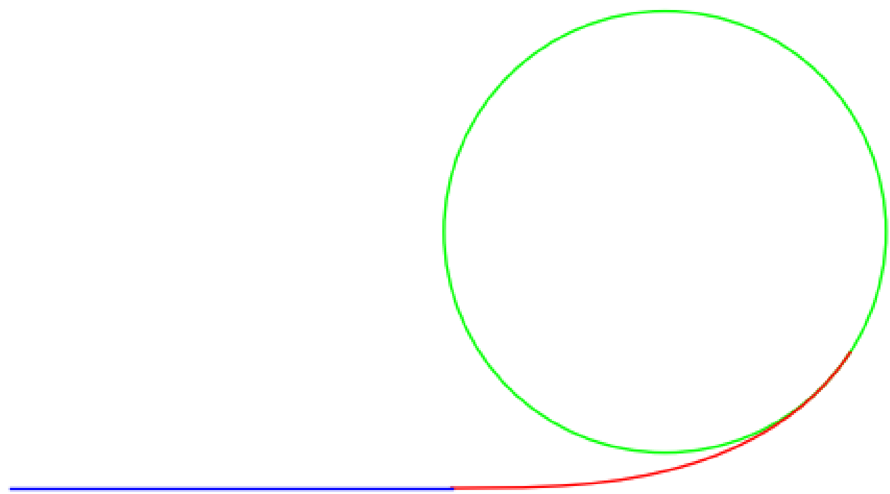
```
var('L')
p = integral(taylor(cos(L^2), L, 0, 12), L)
q = integral(taylor(sin(L^2), L, 0, 12), L)
r1 = parametric_plot([p, q], (L, 0, 1), color = 'red')

r2 = line([(-1.0, 0), (0,0)], rgbcolor = 'blue')

x1 = p.subs(L = 1)
y1 = q.subs(L = 1)
R = 0.5
x2 = x1 - R*sin(1.0)
y2 = y1 + R*cos(1.0)
r3 = circle((x2, y2), R, rgbcolor = 'green')
show(r1 + r2 + r3, aspect_ratio = 1, axes=false)
```

The following is Mathematica code for the Euler spiral component:

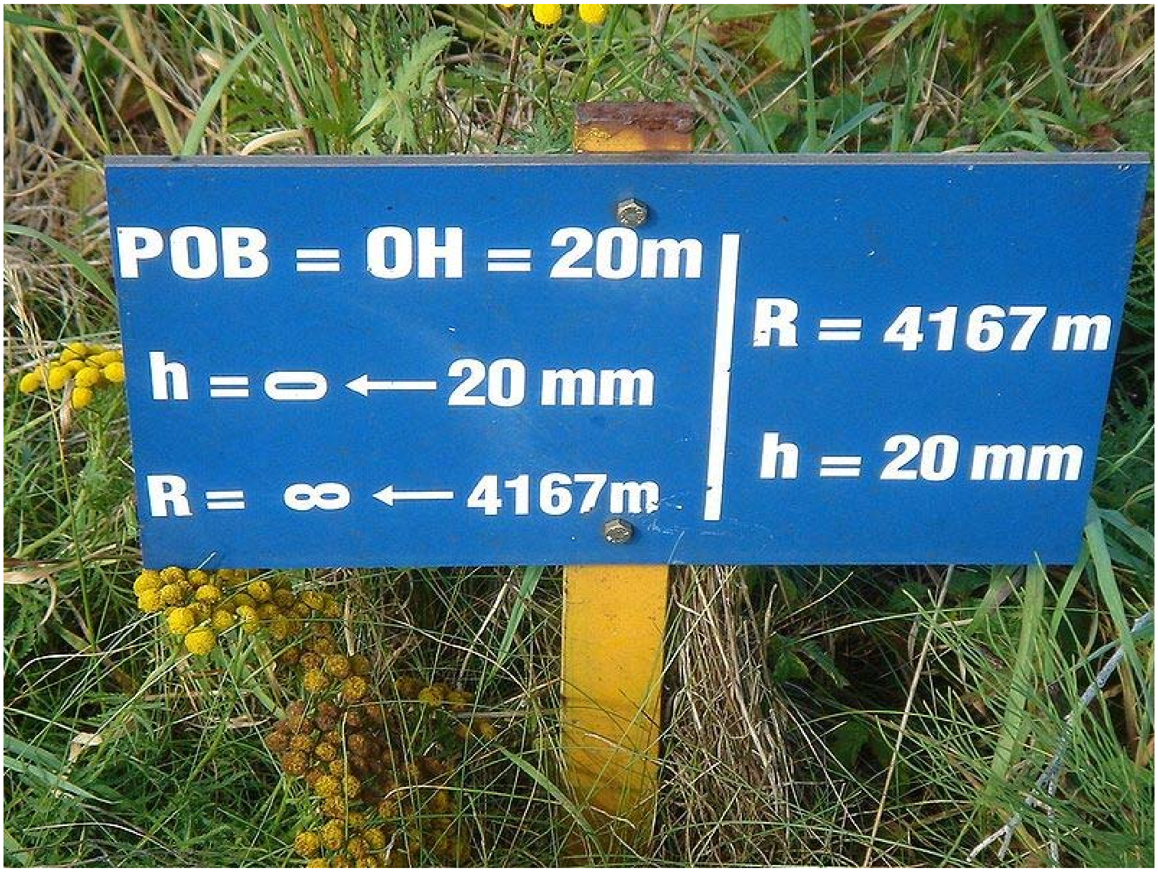
```
ParametricPlot[
 {FresnelC[Sqrt[2/\[Pi]] t]/Sqrt[2/\[Pi]],
  FresnelS[Sqrt[2/\[Pi]] t]/Sqrt[2/\[Pi]]},
 {t, -10, 10}]
```

Chapter 12

Track Transition Curve



The red **Euler spiral** is an example of an easement curve between a blue straight line and a circular arc, which shall be a segment of the green circle.



This sign aside a railroad (between Ghent and Bruges) indicates the start of the transition curve. A parabolic curve (*POB*) is used.

A **Track transition curve**, or **spiral easement**, is a mathematically calculated curve on a section of highway, or railroad track, where a straight section changes into a curve. It is designed to reduce the effects of centrifugal force experienced by users. In plan (i.e., the horizontal curve) the start of the transition is at infinite radius and at the end of the transition it has the same radius as the curve itself, thus forming a very broad spiral. At the same time, in the vertical plane, the outside of the curve is gradually raised until the correct degree of bank is reached.

If such easement were not applied, the lateral acceleration of a rail vehicle would change abruptly at one point – the tangent point where the straight track meets the curve – with undesirable results. With a road vehicle the driver naturally applies the steering alteration in a gradual manner and the curve is designed to permit this, using the same principle.

History

On early railroads, because of the low speeds and wide-radius curves employed, the surveyors were able to ignore any form of easement, but during the 19th century, as speeds increased, the need for a track curve with gradually increasing curvature became apparent. Rankine's 1862 "Civil Engineering" cites several such curves, including an

1828 or 1829 proposal based on the "curve of sines" by William Gravatt, and the *curve of adjustment* by William Froude around 1842 approximating the elastic curve. The actual equation given in Rankine is that of a cubic curve, which is a polynomial curve of degree 3. This was also known as cubic parabola at that time.

In the UK, only from 1845 when legislation and land costs began to constrain the laying out of rail routes and tighter curves were necessary, did the principles start to be applied in practice.

The "true spiral", where the curvature is exactly linear in arclength, requires more sophisticated mathematics (in particular, the ability to integrate its intrinsic equation) to compute than the proposals cited by Rankine. Several late-19th century civil engineers seem to have derived the equation for this curve independently (all unaware of the original characterization of this curve by Leonhard Euler in 1744). Charles Crandall gives credit to one Ellis Holbrook, in the Railroad Gazette, Dec. 3, 1880, for the first accurate description of the curve. Another early publication was *The Railway Transition Spiral* by Arthur N. Talbot, originally published in 1890. Some early 20th century authors call the curve "Glover's spiral", attributing it to James Glover's 1900 publication.

The equivalence of the railroad transition spiral and the clothoid seems to have been first published in 1922 by Arthur Lovat Higgins. Since then, "clothoid" is the most common name given the curve, even though the correct name (following standards of academic attribution) is "the Euler spiral".

Geometry

While railroad track geometry is intrinsically three-dimensional, for practical purposes the vertical and horizontal components of track geometry are usually treated separately.

The overall design pattern for the vertical geometry is typically a sequence of constant grade segments connected by vertical transition curves in which the local grade varies linearly with distance and in which the elevation therefore varies quadratically with distance. Here grade refers to the tangent of the angle of rise of the track. The design pattern for horizontal geometry is typically a sequence of straight line (i.e., a tangent) and curve (i.e. a circular arc) segments connected by transition curves.

In a tangent segment the track bed roll angle is typically zero. In the case of railroad track the track roll angle (*cant* or *camber*) is typically expressed as the difference in elevation of the two rails, a quantity referred to as the superelevation. A track segment with constant non-zero curvature will typically be superelevated in order to have the component of gravity in the plane of the track provide a majority of the centripetal acceleration inherent in the motion of a vehicle along the curved path so that only a small part of that acceleration needs to be accomplished by lateral force applied to vehicles and passengers or lading. The change of superelevation from zero in a tangent segment to the value selected for the body of a following curve occurs over the length of a transition curve that connects the tangent and the curve proper. Over the length of the transition the curvature of the track will also vary from zero at the end abutting the tangent segment to the value of curvature of the curve body, which is numerically equal to one over the radius of the curve body.

The simplest and most commonly used form of transition curve is that in which the superelevation and horizontal curvature both vary linearly with distance along the track. Cartesian coordinates of points along this spiral are given by the Fresnel integrals. The resulting shape matches a portion of an Euler spiral, which is also commonly referred to as a clothoid, and sometimes Cornu spiral.

A transition curve can connect a track segment of constant non-zero curvature to another segment with constant curvature that is zero or non-zero of either sign. Successive curves in the same direction are sometimes called progressive curves and successive curves in opposite directions are called reverse curves.

The Euler spiral has two advantages. One is that it is easy for surveyors because the coordinates can be looked up in Fresnel integral tables. The other is that it provides the shortest transition subject to a given limit on the rate of change of the track superelevation (i.e. the twist of the track). However, as has been recognized for a long time, it has undesirable dynamic characteristics due to the large (conceptually infinite) roll acceleration and rate of change of centripetal acceleration at each end. Because of the capabilities of personal computers it is now practical to employ spirals that have dynamics better than those of the Euler spiral.

Chapter 13

Level of Service

Level of service (LOS) is a measure used by traffic engineers to determine the effectiveness of elements of transportation infrastructure. LOS is most commonly used to analyze highways, but the concept has also been applied to intersections, transit, and water supply.

Level-of-Service in North America

The transportation LOS system uses the letters A through F, with A being best and F being worst. LOS A is the best, described as conditions where traffic flows at or above the posted speed limit and all motorists have complete mobility between lanes. LOS A occurs late at night in urban areas, frequently in rural areas, and generally in car advertisements.

B is slightly more congested, with some impingement of maneuverability; two motorists might be forced to drive side by side, limiting lane changes. LOS B does not reduce speed from LOS A.

LOS C has more congestion than B, where ability to pass or change lanes is not always assured. LOS C is the target for urban highways in some places, and for rural highways in many places. At LOS C most experienced drivers are comfortable, roads remain safely below but efficiently close to capacity, and posted speed is maintained.

LOS D is perhaps the level of service of a busy shopping corridor in the middle of a weekday, or a functional urban highway during commuting hours: speeds are somewhat reduced, motorists are hemmed in by other cars and trucks. LOS D is a common goal for urban streets during peak hours, as attaining LOS C would require a prohibitive cost and societal impact in bypass roads and lane additions.

LOS E is a marginal service state. Flow becomes irregular and speed varies rapidly, but rarely reaches the posted limit. On highways this is consistent with a road at or approaching its designed capacity. LOS E is a common standard in larger urban areas, where some roadway congestion is inevitable.

LOS F is the lowest measurement of efficiency for a road's performance. Flow is forced; every vehicle moves in lockstep with the vehicle in front of it, with frequent slowing required. Technically, a road in a constant traffic jam would be at LOS F. This is because LOS does not describe an instant state, but rather an average or typical service. For example, a highway might operate at LOS D for the AM peak hour, but have traffic consistent with LOS C some days, LOS E or F others, and come to a halt once every few weeks. However, LOS F describes a road for which the travel time cannot be predicted. Facilities operating at LOS F generally have more demand than capacity.

The Highway Capacity Manual and AASHTO Geometric Design of Highways and Streets ("Green Book") list the following levels of service:
A= Free flow
B=Reasonably free flow
C=Stable flow
D=Approaching unstable flow
E=Unstable flow
F=Forced or breakdown flow

LOS for At-Grade Intersections

The Highway Capacity Manual defines level-of-service for signalized and unsignalized intersections as a function of the average vehicle control delay. LOS may be calculated per-movement or per-approach for any intersection configuration; however, LOS for the intersection as a whole is only defined for signalized and all-way stop configurations.

LOS	**Signalized Intersection**	**Unsignalized Intersection**
A	≤10 sec	≤10 sec
B	10-20 sec	10-15 sec
C	20-35 sec	15-25 sec
D	35-55 sec	25-35 sec
E	55-80 sec	35-50 sec
F	≥80 sec	≥50 sec

When analyzing unsignalized intersections that are not all-way stop-controlled, each possible movement is considered individually. Each movement has a rank. *Rank 1* movements have priority over *rank 2* movements, which have priority over *rank 3* movements, which have priority over *rank 4* movements. The rank of each movement is as follows, with the *minor road* being the road that is controlled by the stop signs and the *major road* being the road whose through movement moves freely. As for vehicular movements that conflict with pedestrian movements of the same rank, pedestrians have priority:

1. Movements of this rank are the through movements on the major road, parallel pedestrian movements, and right turns from the major road. LOS for movements of this rank is trivial, because LOS is determined by control delay. These are "free" movements, and as such the control delay is always zero.
2. Movements of this rank include left turns from the major road.
3. Movements of this rank include through movements on the minor road, parallel pedestrian movements, and right turns from the minor road.
4. Movements of this rank include left turns from the minor road.

Movements are analyzed in order of rank (highest rank first), and any capacity that is left over from one rank devolves onto the next rank below. Because of this pecking order, depending on intersection volumes, there may be no capacity for lower ranked movements.

Modern Roundabouts

The 2000 Highway Capacity Manual provides skeleton coverage of modern roundabouts, but does not define level-of-service at this time. Instead, the measure-of-effectiveness is the quotient of the volume to the capacity. A *modern roundabout* in the United States is a roundabout in which traffic inside the circle always has priority. Entering traffic is controlled by a yield sign.

Level-of-Service in Other Transportation Network Elements

Performance of other transportation network elements can also be communicated by LOS. Among them are:

- Two-lane roadways (uninterrupted flow)
- Multilane roadways (4 or more lanes) (uninterrupted flow)
- Open freeway segments
- Freeway entrances (merges), exits (diverges), and weaving lanes
- Bicycle facilities (measure-of-effectiveness: events per hour; events include meeting an oncoming bicyclist or overtaking a bicyclist traveling in the same direction)
- Pedestrian facilities (HCM measure-of-effectiveness: pedestrians per unit area)

Theoretical Considerations

The level of service concept was first developed for highways in an era of rapid expansion in the use and availability of the private motor car. The primary concern was congestion, and it was commonly held that only the rapid expansion of the freeway network would keep congestion in check.

Since then, some professors in urban planning schools have proposed measurements of levels of service that take public transportation into account. Such systems would include wait time, frequency of service, time it takes to pay fares, quality of the ride itself, accessibility of depots, and, perhaps, other criteria as well.

LOS can also be applied to surface streets, to describe major signalized intersections. A crowded four-way intersection where the major traffic movements were conflicting turns might have an LOS of D or E. At intersections, queuing time can be used as a rubric to measure LOS; computer models given the full movement data can spit out a good estimate of LOS.

While it may be tempting to aim for an "A" Level of Service, this is unrealistic in urban areas. Urban areas more typically adopt standards varying between "C" and "E", depending on the area's size and characteristics, while "F" is sometimes allowed in areas with improved pedestrian, bicycle, or transit alternatives. More stringent Level of Service standards (particularly in urban areas) tend to necessitate the widening of roads to accommodate development, thus discouraging use by these alternatives. Because of this, some planners recommend increasing population density in towns, narrowing streets, managing car use in some areas, providing sidewalks and safe pedestrian and bicycle facilities, and making the scenery interesting for pedestrians.

A level of service standard has been developed by John J. Fruin, PhD., for pedestrian facilities. The standard uses American units and applies to pedestrian queues, walkways, and stairwells. It should be noted that this standard is not considered a good measure of pedestrian facilities by the planning or engineering professions, because it rates undesirable (and hence unused) sidewalks with a Level of Service "A", while pedestrians tend to prefer active, interesting sidewalks, where people prefer to walk (but rate a worse Level of Service on this scale). To rectify this and other issues, The National Cooperative Highway Research Program (NCHRP) is conducting a project to enhance methods to determine Levels of Service for automobile, transit, bicycle, and pedestrian modes on urban streets, with particular consideration to intermodal interactions.

LOS	Walking	Queueing	Stairs
LOS A	35 sq ft (3.3 m^2) per person or greater	13 sq ft (1.2 m^2). per person or greater	20 sq ft (1.9 m^2) per person or greater
LOS B	25 sq ft (2.3 m^2)-35 sq ft per person	10-13 sq ft. per person	15-20 sq ft. per person
LOS C	15 sq ft (1.4 m^2)-25 sq ft per person	7-10 sq ft. per person	10-15 sq ft. per person
LOS D	10 sq ft (0.93 m^2)-15 sq ft per person	3-7 sq ft. per person	7-10 sq ft. per person
LOS E	5 sq ft (0.46 m^2)-10 sq ft per person	2-3 sq ft. per person	4-7 sq ft. per person
LOS F	5 sq ft (0.46 m^2) per person or less	2 sq ft (0.19 m^2). per person or less	4 sq ft (0.37 m^2). per person or less

The A to F scale described above deals only with delays and service reliability. These delays are typically caused by congestion, breakdowns or infrequent service. It assumes there is a service in place that people can use. It also implies that poor levels of service can be solved by increased capacity such as additional lanes, overcoming bottlenecks, and in the case of transit, more buses or trains. It does not deal for instance with cases where there is no bridge across a river, no bus or train services, no sidewalks, or no bike-lanes.

An expanded level of service might look like: 0 - No Service exists. Latent demand may exist. 1 - Service is poor, unsafe or discouraging. Demand is suppressed below socially desirable levels. A-F - As per existing level-of-service scale. G - Further expansion of capacity is limited. H - No expansion is possible. Radical or innovative solutions are required

Level of Service in the UK

The Level of Service measure is much more suited to American Roads than roads in Europe and the UK, however the Highway Capacity Manual is used in the UK. The technique does find its way into UK textbooks, however in practice it is sparingly used in transportation analysis. The individual countries of the UK have different bodies for each areas roads, and as a result detailed techniques and applications vary in Scotland, England and Wales, however in general the practice is the same.

In the UK rural and urban roads are in general much busier than in the U.S, and as such service levels tend to be to the higher end of the scale, especially in the peak commuting periods. It is acceptable for roads to operate at 85% capacity, which equates to D and E LOS.

In general the principle the UK uses is to take the volume of traffic in one hour on the road and divide by the appropriate capacity of the road type to get a v/c rating. This v/c rating can be cross-referenced to the textbooks which publish tables of v/c ratings and their equivalent LOS ratings. The lack of definitive categories towards the D, E and F LOS ratings limit the use on UK roads, as an D or E category on an urban road, would be acceptable in the UK.

In certain circumstances the UK shortens the LOS categories to just A-D. The first 2 categories indicate free-movement of traffic (i.e. under 85% capacity), the C category indicates reaching capacity 85%-100%, whilst D indicates over capacity. Little reference to this can be found in textbooks and it may just be an 'unwritten engineering practice', agreed with certain authorities.

Level of Service in Australia

In Australia levels of service are an integral component of Asset Management Plans. In that context they can be defined as the service quality for a given activity. Levels of service are often documented as a commitment to carryout a given action or actions within a specified time frame in response to an event or asset condition data.

Refer Austroads Guide to Traffic Engineering Practice Part 2 for a good explanation.

Chapter 14

Signal Timing

Signal timing is the technique which traffic engineers use to determine who has the right-of-way at an intersection. Signal timing involves deciding how much green time the traffic lights shall provide at an intersection approach, how long the pedestrian WALK signal should be, and many numerous other factors.

Basic signal timing operation

To understand basic signal timing fundamentals, one must also understand the different modes of operation for the traffic signal controller. Many intersections have some sort of mechanism for detecting vehicles as they approach the intersection. Most common are induction loops. These are buried in the roadway and detect vehicles by changes in their magnetic field by the metal in passing vehicles. Other common methods are video detection which uses pixelation, microwave detection, and infrared detection among others. An intersection equipped with detection is said to be *actuated*. An intersection without detection is said to be *fixed*.

There are different categories of actuated signals. To save money on maintenance, some agencies opt to design an intersection as semi-actuated. Semi-actuated means the intersection has detection on the minor street approaches and major street left turns only. The major street is then programmed to operate a fixed time every cycle, but the controller will service the other movements only when there is demand. In signal coordination, most signals operate in a semi-actuated mode.

In fixed operation, a controller has a set programmed time to service all movements every cycle. The controller will service all movements whether or not there is vehicle demand. When a detector at an actuated signal breaks, that movement will then have to operate as fixed until the detector is repaired.

There are three general ways for a traffic signal to operate, FREE, COORD, and FLASH operation. In FREE operation, the signal is running based on its own demand and timing parameters based on the information provided by its detectors. It is not operating under any background cycle length. In COORD operation, short for coordination, the signal is running a background cycle length. Non-major street movements are usually still actuated, and the controller will rest on the major street until the background cycle length is fulfilled. The final mode is FLASH operation. When the volume of vehicles at an intersection no longer warrants the signal to be active, the signal can switch to FLASH mode. When volume picks up again the signal switches back into either FREE or COORD operation. The daily operation of a signal may involve it in FLASH early in the morning, COORD during the day, FREE in the evening, and back to FLASH late at night.

Basic timing functions

There are several basic timing functions that need to be programmed for the traffic controller to operate.

MIN time determines the minimum duration of the green interval for each movement. Left turns, minor streets, major streets, usually have different **MIN** times. Left turns and minor side street intervals are often in the range from 4 to 10 seconds while major streets often go higher than 15.

Gap, extension, or **passage time** determines the extendable portion of the green time for a movement. The movement remains in the extendable portion as long as an actuation is present and the **passage** timer has not expired. If the interval is set as three seconds and there is not vehicle present after that three seconds, the movement will terminate.

MAX time limits the maximum time of the green interval. If there are no conflicting demands on the intersection, the controller will ignore the **MAX** and rest in the major street movement.

Yellow Clearance determines the yellow time for the associated movement.

Red Clearance determines the all-red time for the associated movement.

Walk time provides the length of the walk indication.

Flashing Don't Walk is the duration of the flashing pedestrian clearance. This is timed as the length of the crosswalk divided by a speed of 3.5 feet per second.

Coordination

Coordination refers to the timing of the signals so that a "platoon" of cars traveling on a street arrives at a succession of green lights and proceeds through multiple intersections without stopping. A well coordinated signal system can enhance traffic flow, reduce delay and minimize pollution. However, it is not always possible to retain progression throughout a network of signals. It is also difficult to maintain signal progression on a two-way street. An early traffic engineer Henry Barnes, who served as Commissioner of Traffic in many cities including Baltimore, Maryland and New York, New York, developed coordinated traffic signal timings, so that large amounts of traffic could be accommodated on major traffic arterials.

Traffic signal timing is a very complex topic. For example timing a 'WALK' signal for a wide pedestrian crossing and slower pedestrians (for example the elderly) could result in very long waits for vehicles, and thus increases the likelihood of cars running the light, which could potentially cause accidents. Therefore, optimizing the safety of intersections involves multiple factors like street width, lane width, number of intersecting streets, availability of electricity for a signal, number of cars per unit of time and even/uneven nature of flow, number and type of pedestrians, and many other factors.

Traffic signals can be programmed to have different signal timing plans, depending on the time of day.

Chapter 15

Traffic Congestion: Reconstruction with Kerner's Three-Phase Theory

Vehicular traffic can be either free or congested. Traffic occurs in time and space, i.e., it is a spatiotemporal process. However, usually traffic can be measured only at some road locations (for example, via road detectors, video-cameras, probe vehicle data, or phone data). For efficient traffic control and other intelligent transportation systems, the reconstruction of traffic congestion is necessary at all other road locations at which traffic measurements are not available. Traffic congestion can be reconstructed in space and time (Fig. 1) based on Boris Kerner's three-phase traffic theory with the use of the ASDA and FOTO models introduced by Kerner. Kerner's three-phase traffic theory and, respectively, the ASDA/FOTO models are based on some *common* spatiotemporal features of traffic congestion observed in measured traffic data.

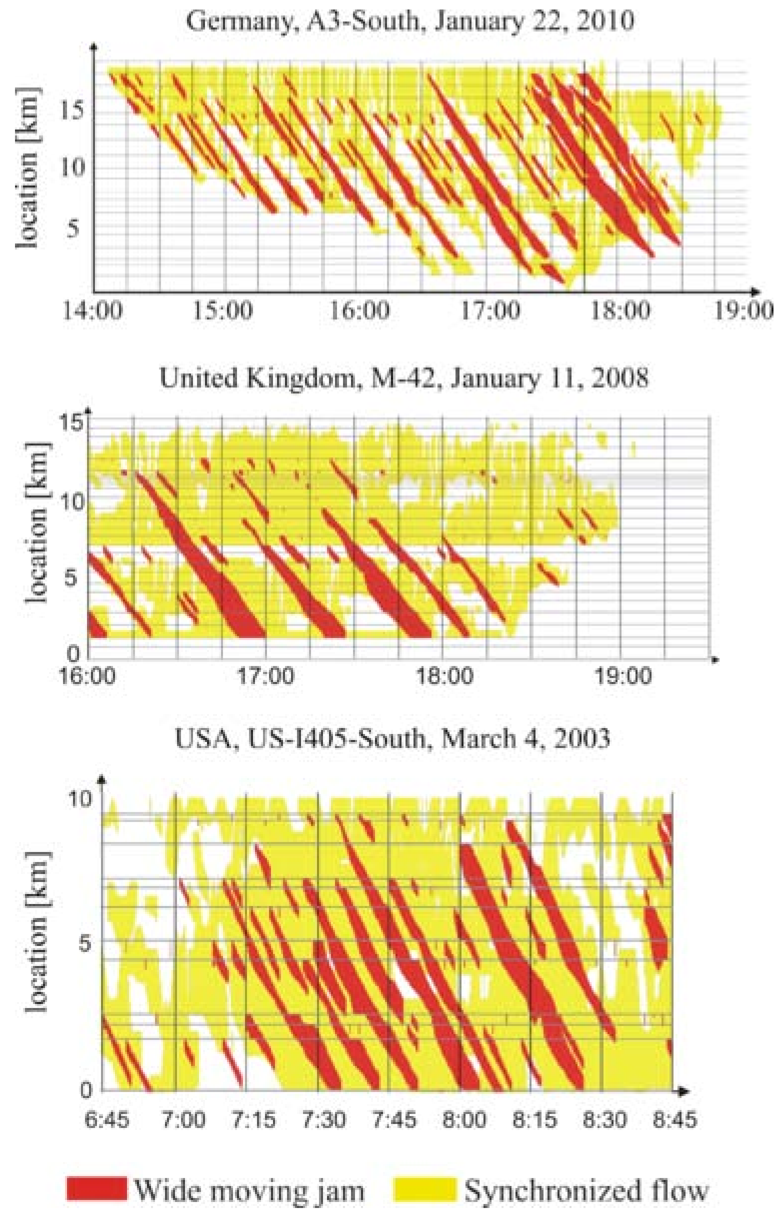


Fig. 1. Empirical examples of traffic congestion reconstructed by the ASDA/FOTO models using raw data measured by road detectors on different highways in the United Kingdom, Germany, and the USA. Representation of traffic congestion in space-time plane through regions associated with two qualitatively different traffic phases in congested traffic: 1. Wide moving jam (red regions). 2. Synchronized flow (yellow regions). White regions – free flow.

Common spatiotemporal empirical features of traffic congestion

Definition

Common spatiotemporal empirical features of traffic congestion are those spatiotemporal features of traffic congestion, which are qualitatively the same for different highways in different countries measured during years of traffic observations. In particular, common features of traffic congestion are independent on weather, road conditions and road infrastructure, vehicular technology, driver characteristics, day time, etc.

Kerner's definitions [S] and [J], respectively, for the synchronized flow and wide moving jam phases in congested traffic are examples of common spatiotemporal empirical features of traffic congestion.

Propagation of wide moving jams through highway bottlenecks

In empirical observations, traffic congestion occurs usually at a highway bottleneck as a result of traffic breakdown in an initially free flow at the bottleneck. A highway bottleneck can result from on- and off-ramps, road curves and gradients, road works, etc.

In congested traffic (this is a synonym term to traffic congestion), a phenomenon of the propagation of a moving traffic jam (moving jam for short) is often observed. A moving jam is a local region of low speed and great density that propagates upstream as a whole localized structure. The jam is limited spatially by two jam fronts. At the downstream jam front, vehicles accelerate to a higher speed downstream of the jam. At the upstream jam front, vehicles decelerate while approaching the jam.

A wide moving jam is a moving jam that exhibits the characteristic jam feature [J], which is a *common* spatiotemporal empirical feature of traffic congestion. The jam feature [J] defines the wide moving jam traffic phase in congested traffic as follows.

Definition [J] for wide moving jam

A wide moving jam is a moving traffic jam, which exhibits the characteristic jam feature [J] to propagate through any bottlenecks while maintaining the mean velocity of the downstream jam front denoted by v_g.

Kerner's jam feature [J] can be explained as follows. The motion of the downstream jam front results from acceleration of drivers from a standstill within the jam to traffic flow downstream of the jam. After a vehicle has begun to accelerate escaping from the jam, to satisfy safety driving, the following vehicle begins to accelerate with a time delay. We denote the mean value of this time delay in vehicle acceleration at the downstream jam front by $\tau_{del,jam}^{(a)}$. Because the average distance between vehicles within the jam, including average vehicle length, equals to $\frac{1}{\rho_{max}}$ (where ρ_{max} is the average vehicle density within the jam), the mean velocity of the downstream jam front v_g is

$$v_g = -\frac{1}{\rho_{max}\tau_{del,jam}^{(a)}} \qquad (1)$$

.

When traffic parameters (percentage of long vehicles, weather, driver characteristics, etc.) do not change over time, $\tau^{(a)}_{del,jam}$ and ρ_{max} are constant in time. This explains why the mean velocity of the downstream jam front v_g (1) is the characteristic parameter that does not depend of the flow rates and densities upstream and downstream of the jam.

Catch effect: pinning of downstream front of synchronized flow at bottleneck

In contrast with the jam feature [J], the mean velocity of the downstream front of synchronized flow is not self-maintained during the front propagation. This is the *common* feature of synchronized flow that is one of the two phases of traffic congestion.

A particular case of this common feature of synchronized flow is that the downstream synchronized flow front is usually caught at a highway bottleneck. This pinning of the downstream front of synchronized flow at the bottleneck is called the *catch effect*. Note that at this downstream front of synchronized flow, vehicles accelerate from a lower speed within synchronized flow upstream of the front to a higher speed in free flow downstream of the front.

Definition [S] for synchronized flow

Synchronized flow is defined as congested traffic that does not exhibit the jam feature [J]; in particular, the downstream front of synchronized flow is often fixed at the bottleneck.

Thus Kerner's definitions [J] and [S] for the wide moving jam and synchronized flow phases of his three-phase traffic theory are indeed associated with common empirical features of traffic congestion.

Empirical example of wide moving jam and synchronized flow

Vehicle speeds measured with road detectors (1 min averaged data) illustrate Kerner's definitions [J] and [S] (Fig. 2 (a, b)). There are two spatiotemporal patterns of congested traffic with low vehicle speeds in Fig. 2 (a). One pattern of congested traffic propagates upstream with almost constant mean velocity of the downstream pattern front through the freeway bottleneck. According to the definition [J] this pattern of congested traffic belongs to the "wide moving jam" traffic phase. In contrast, the downstream front of the other pattern of the congested traffic is fixed at the bottleneck. According to the definition [S] this pattern of congested traffic belongs to the "synchronized flow" traffic phase (Fig. 2 (a) and (b)).

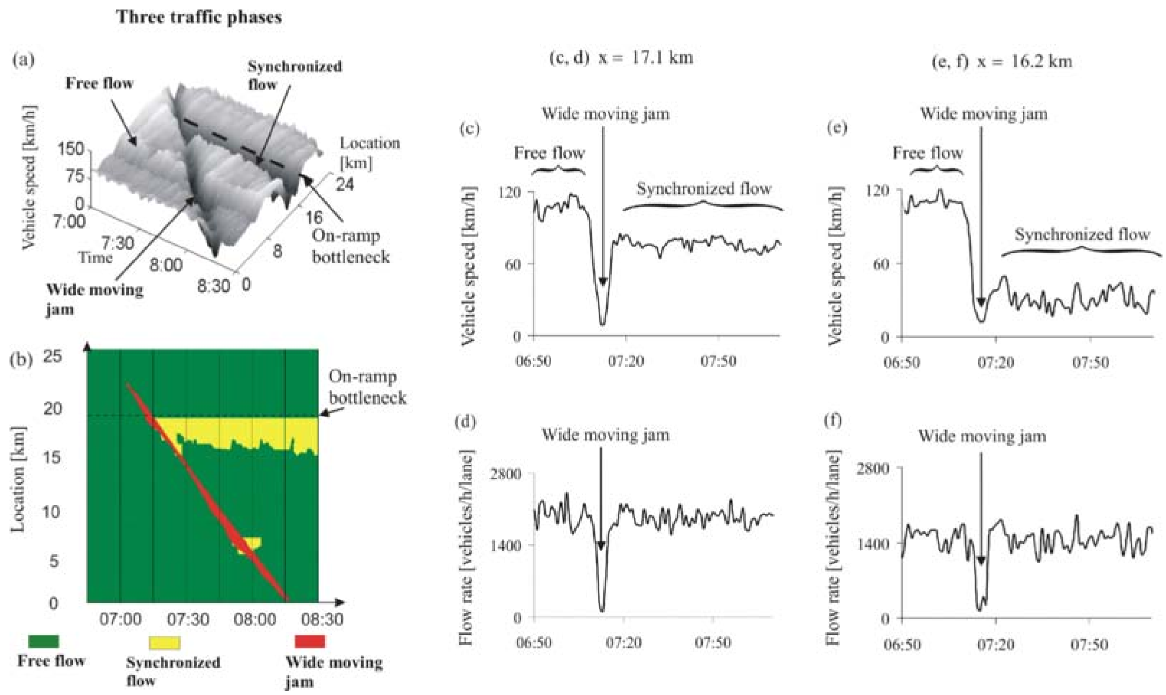


Fig.2. Empirical spatiotemporal common features of traffic congestion and the associated traffic phase definitions in Kerner's theory: (a) Measured data of average vehicle speed in time and space. (b) Representation of speed data in (a) on the time-space plane. (c-f) Time-dependences of speed (c, e) and flow rate (d, f) at two different locations within traffic congestion shown in (a, b); the data in (c, d) and (e, f) are measured respectively at location 17.1 km (c, d) (about 100 meter upstream of road location labelled "On-ramp bottleneck" in (a, b)) and at location 16.2 km (e, f). At location 17.1 the flow rate (d) in free and synchronized flows is greater in comparison with that at location 16.2 (f) due to on-ramp inflow at the bottleneck.

ASDA and FOTO models

The FOTO (**F**orecasting **o**f **t**raffic **o**bjects) model reconstructs and tracks regions of synchronized flow in space and time. The ASDA (**A**utomatische **S**tau**d**ynamik**a**nalyse: Automatic Tracking of Moving Jams) model reconstructs and tracks wide moving jams.

General features

Firstly, in accordance with common empirical features of traffic congestion, the ASDA/FOTO models identify the synchronized flow and wide moving jam phases in measured data of congested traffic. In addition with the common features, the following empirical result has been used in the ASDA/FOTO models for traffic phase identification: Within a wide moving jam, both the speed and flow rate are very small (Fig. 2 (c-f)). In contrast, whereas the speed with the synchronized flow phase is considerably lower than in free flow (Fig. 2 (c, e)), the flow rate in synchronized flow can be as great as in free flow (Fig. 2 (d, f)).

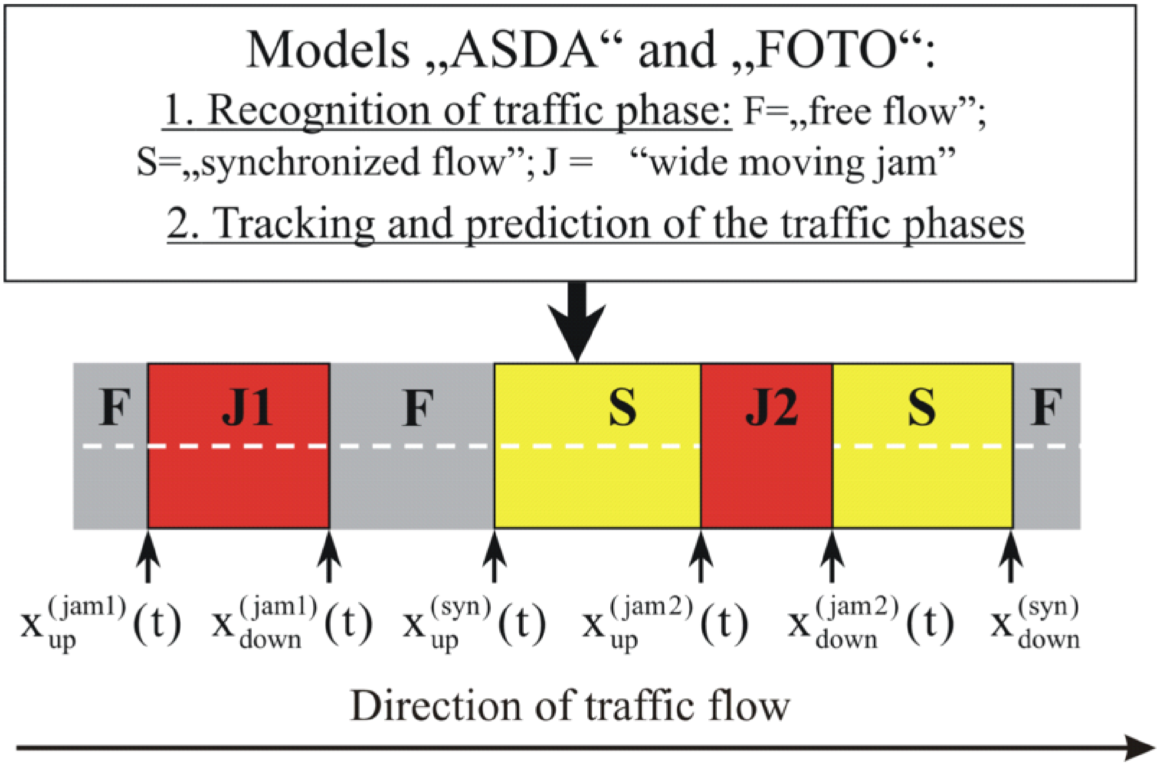


Fig. 3. Explanation of ASDA/FOTO models. Superscripts "jam 1", "jam 2" are related to two different wide moving jams. Superscripts "syn" are associated with synchronized flows. Subscripts "up" and "down" are related respectively to the upstream and downstream fronts of synchronized flow and wide moving jams.

Secondly, based on the abovementioned common features of wide moving jams and synchronized flow, the FOTO model tracks the downstream and upstream fronts of synchronized flow denoted by $x^{(syn)}_{down}(t)$, $x^{(syn)}_{up}(t)$, where *t* is time (Fig. 3). The ADSA model tracks the downstream and upstream fronts of wide moving jams denoted by $x^{(jam)}_{down}(t)$, $x^{(jam)}_{up}(t)$(Fig. 3). This tracking is carried out between road locations at which the traffic phases have initially been identified in measured data, i.e., when synchronized flow and wide moving jams cannot be measured.

In other words, the tracking of synchronized flow by the FOTO model and wide moving jams by the ASDA model is performed at road locations at which *no* traffic measurements are available, i.e., the ASDA/FOTO models make the forecasting of the front locations of the traffic phases in time. The ASDA/FOTO models enable us to predict the merging and/or the dissolution of one or more initially different synchronized flow regions and of one or more initially different wide moving jams that occur between measurement locations.

ASDA/FOTO models for data measured by road detectors

Cumulative flow approach for FOTO

While the downstream front of synchronized flow at which vehicles accelerate to free flow is usually fixed at the bottleneck, the upstream front of synchronized flow at which vehicles moving initially in free flow must decelerate approaching synchronized flow can propagate upstream. In empirical (i.e., measured) traffic data, the velocity of the upstream front of synchronized flow depends usually considerably both on traffic variables within synchronized flow downstream of the front and within free flow just upstream of this front. A good correspondence with empirical data is achieved, if a time-dependence of the location of the synchronized flow front is calculated by the FOTO model with the use of a so-called cumulative flow approach:

$$x_{up}^{(syn)}(t) = \mu \frac{1}{n} \int_{t_0}^{t} (q_S(t) - q_F(t))dt, \quad t \geq t_0 \qquad (2)$$

where q_F and q_S [vehicles/h] are respectively the flow rates upstream and downstream of the synchronized flow front, μ is a model parameter [m/vehicles], and n is the number of road lanes.

Two approaches for jam tracking with ASDA

There are two main approaches for the tracking of wide moving jams with the ASDA model:

1. The use of the Stokes-shock-wave formula.
2. The use of a characteristic velocity of wide moving jams.

The use of the Stokes-shock-wave formula in ASDA

The current velocity $v^{(jam)}$ of a front of a wide moving jam is calculated though the use of the shock-wave formula derived by Stokes in 1848:

$$v^{(jam)}(t) = \frac{q_2(t) - q_1(t)}{\rho_2(t) - \rho_1(t)} \qquad (3)$$

,

where q_1 and ρ_1 the flow rate and density upstream of the jam front that velocity should be found; q_2 and ρ_2 are the flow rate and density downstream of this jam front. In (3) *no any* relationship, in particular, *no* fundamental diagram is used between the flow rates q_1, q_2 and vehicle densities ρ_1, ρ_2 found from measured data independent of each other.

The use of a characteristic velocity of wide moving jams

If measured data are not available for the tracking of the downstream jam front with the Stokes-shock-wave formula (3), the formula

$$v_{down}^{(jam)} = v_g \qquad (4)$$

is used in which v_g is the characteristic velocity of the downstream jam front associated with Kerner's jam feature [J] discussed above. This means that after the downstream front of a wide moving jam has been identified at a time instant $t = t_1$, the location of the downstream front of the jam can be estimated with formula

$$x(t) - x(t_1) = v_g(t - t_1), \quad t > t_1 \qquad (5)$$

The characteristic jam velocity is illustrated in Fig. 4. Two wide moving jams propagate upstream while maintaining the mean velocity of their downstream fronts. There are two jams following each other in this empirical example.

However, in contrast with the mean velocity of the downstream jam front, the mean velocity of the upstream jam front depends on the flow rate and density in traffic flow upstream of the jam. Therefore, in a general case the use of formula (5) can lead to a great error by the estimation of the mean velocity of the upstream jam front.

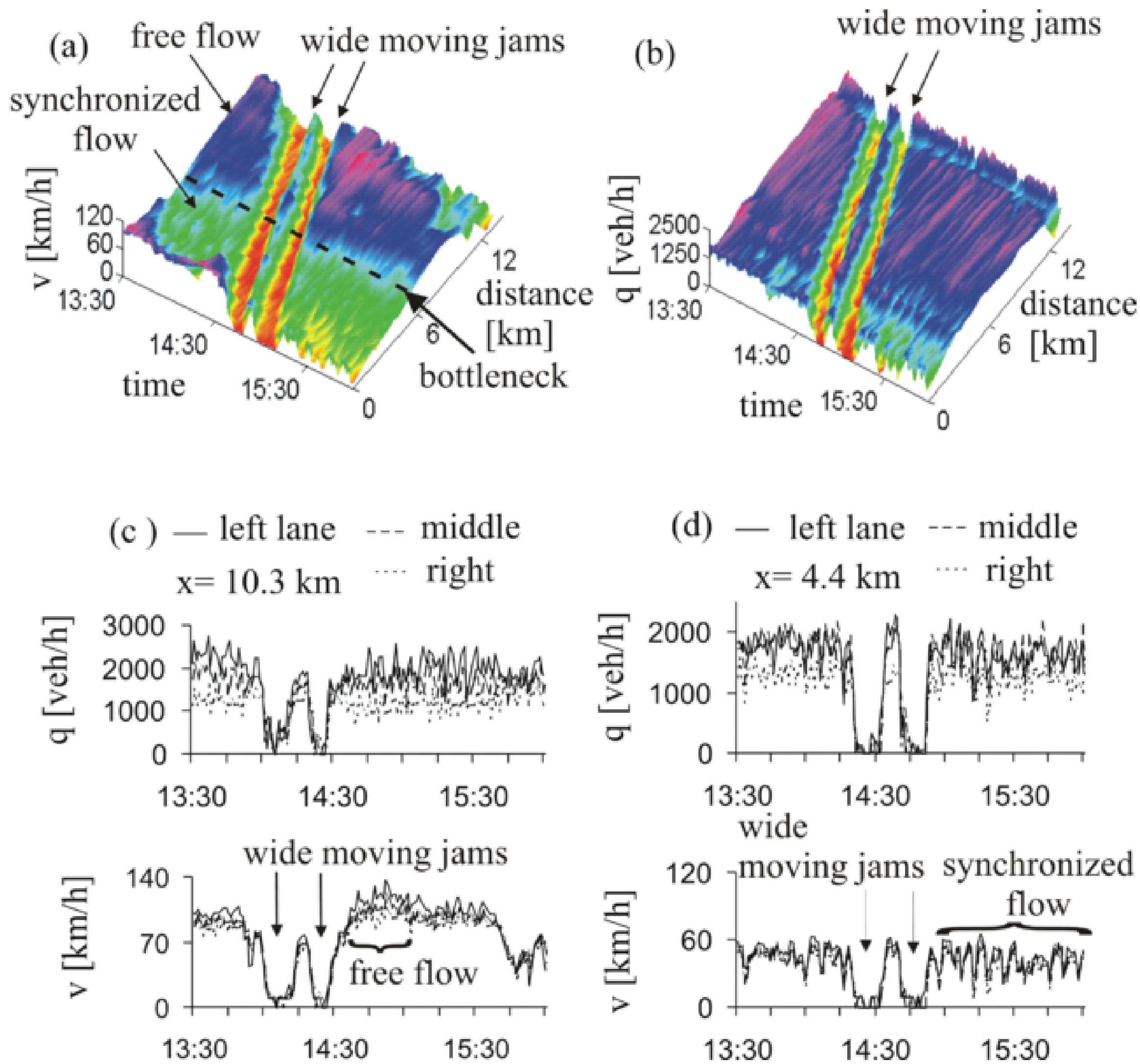


Fig. 4: Measured traffic data that illustrates the characteristic jam feature [J]: (a, b) Average speed denoted by v km/h (a) and flow rate denoted by q [vehicles/h] (b) in space and time. (c, d) Time-dependences of flow rate and speed within traffic congestion in (a, b) at two different road locations shown for each of the three road lanes.

In many data measured on German highways has been found $v_g \approx -15\text{km/h}$. However, although the mean velocity v_g of the downstream jam front is independent of the flow rates and densities upstream and downstream of the jam, v_g can depend considerably on traffic parameters like the percentage of long vehicles in traffic, weather, driver characteristics, etc. As a result, the mean velocity v_g found in different data measured over years of observations varies approximately within the range $-12 > v_g > -20$km/h.

On-line applications of ASDA/FOTO models in traffic control centres

Reconstruction and tracking of spatiotemporal congested patterns with the ASDA/FOTO models is done today online permanently in the traffic control centre of the federal state Hessen (Germany) for 1200 km of freeway network. Since April 2004 measured data of nearly 2500 detectors are automatically analyzed by ASDA/FOTO. The resulting spatiotemporal traffic patterns are illustrated in a space-time diagram showing congested pattern features like Fig. 5. The online system has also been installed in 2007 for North-Rhine Westphalia freeways. The raw traffic data are transferred to WDR, the major public radio broadcasting station from North-Rhine Westphalia in Cologne, who offers traffic messages to the end customer (e. g., radio listener or driver) via broadcast channel RDS. The application covers a part of the whole freeway network with 1900 km of freeway and more than 1000 double loop detectors. In addition, since 2009 ASDA/FOTO models are online in the northern part of Bavaria.

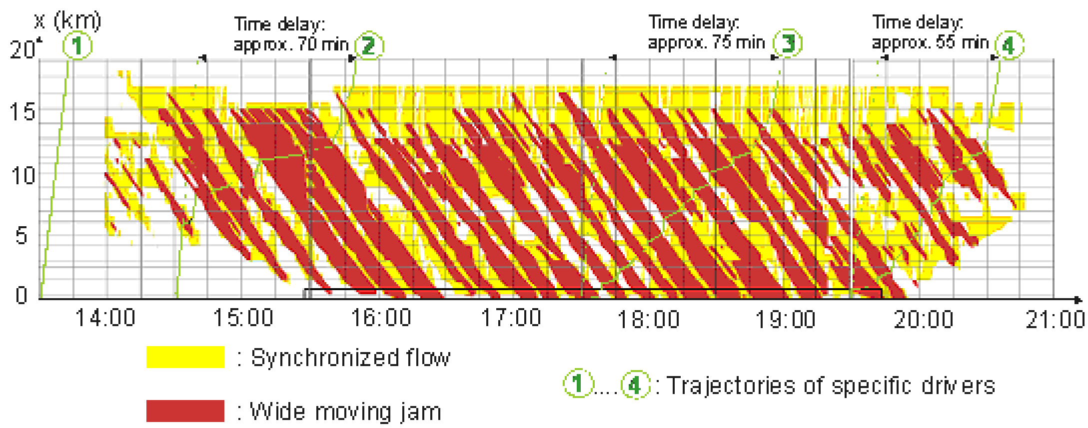


Fig. 5: Congested traffic pattern reconstructed by FOTO and ASDA models: space-time diagram with vehicle trajectories 1-4 and related travel delay times. Road detector data as input for ASDA/FOTO models is measured on freeway A5-North in Hessen, Germany, 14th June, 2006

Average traffic flow characteristics and travel time

In addition to spatiotemporal reconstruction of traffic congestion (Figs. 1 and 5), the ASDA/FOTO models can provide average traffic flow characteristics within synchronized flow and wide moving jams. In turn, this permits the estimation of either travel time on a road section or travel time along any vehicle trajectory (see examples of trajectories 1-4 in Fig. 5).

ASDA/FOTO models for data measured by probe vehicles

Firstly, the ASDA and FOTO models identify transition points for phase transitions along the trajectory of a probe vehicle. Each of the transition points is associated to the front separating spatially two of the three different traffic phases each other (free flow (F), synchronized flow (S), wide moving jam (J)). After the transition points have been found, the ASDA/FOTO models reconstruct regions of synchronized flow and wide moving jams in space and time with the use of empirical features of these traffic phases discussed above (see Figs. 2 and 4).