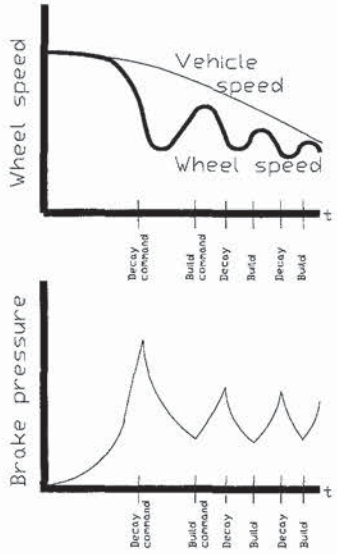


FIGURE 15.8 Typical antilock brake cycling.

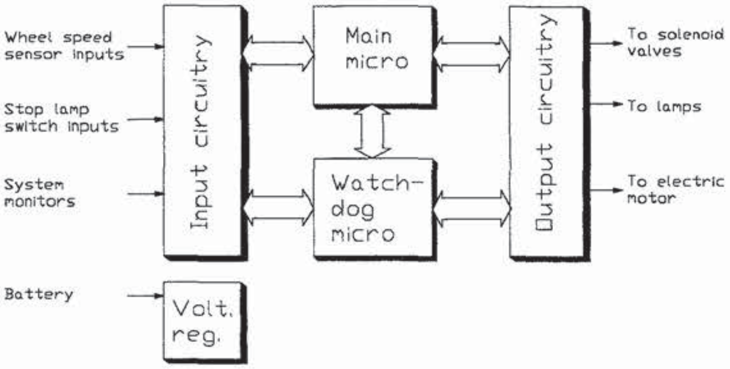


FIGURE 15.9 Electronic control unit block diagram.

decreasing microcontroller costs have made microcontroller-based electronic control units the norm rather than the exception. Although the control units can be either engine compartment-mounted or passenger compartment-mounted, reduced wiring costs favors the former. Also, for enhanced reliability, electronic control units may be either attached to or integrated with the hydraulic modulator.

15.3.3 Safety Considerations

Standard automotive brake systems have been developed and refined over the years to be highly reliable and safe. Because of its ability to decrease pressure in brakes, an antilock system must be designed using a disciplined methodology and must be rigorously tested prior to release for production.

Failure Mode and Effects Analyses/Fault Tree Analyses. Failure mode and effects analyses (or fault tree analyses) are essential to the proper design of antilock systems. Both system-level and subsystem-level analyses need to be performed and fault effects and detection assumptions must be tested. No single failure can result in an unsafe condition and, if a fault is undetectable in the field, that fault in conjunction with any other fault must not result in an unsafe condition. Because of the complexity of the electronic control units, simulation techniques are used to test those fault effects in which bench or field testing is impractical.

Common Design Techniques to Improve Safety. One of the most common techniques used to improve safety in antilock systems is to include extensive built-in-test within the electronic control unit. Typically, all inputs to the electronic control unit and outputs to the other components of the antilock system are tested for proper signals and loads, respectively, and all functions internal to the electronic control unit are extensively tested.

In addition, redundant processing is commonly used to insure the proper internal working of a microcontroller. This may take the form either of identical microcontrollers or of a main and a watchdog microcontroller that can inhibit operation.

In order to ensure inhibition of faulty antilock operation, antilock systems employ a relay function to remove actuation power from the output actuators; this function may take the form of a discrete relay or it may be a transistorized circuit. This relay function is a key element of the design since it affords a secondary method in which to inhibit energization of valves or the motor/pump and, therefore, a second level of safety relative to improper antilock operation.

Figure 15.9 is a typical electronic control unit block diagram.[3] Inputs are filtered and buffered prior to being presented to the microcontrollers for processing. Likewise, the microcontroller outputs are buffered/amplified and filtered prior to exiting the electronic control unit. In the diagram shown, the main microcontroller is responsible for the majority of processing and control of the outputs; the watchdog microcontroller, as its name implies, is responsible for monitoring for proper operation and inhibiting antilock if faults are indicated. A characteristic of modern antilock electronic control units is bidirectional communication between functional blocks; this is a result of the high level of built-in-test designed into the control units. For example, the output circuitry may be commanded to test the solenoid valves for proper current draw and convey the test results to the microcontrollers; similarly, the input circuitry may be commanded to perform tests on the sensors and other antilock components external to the electronic control unit, and convey the test results to the microcontrollers.

15.3.4 Antilock Control Logic Fundamentals

Due to the complexity of antilock braking and the requirements of stability and steerability as well as good stopping distance, the brake control algorithm is more easily represented as a state-space diagram than as a classical proportional-integrative-derivative control scheme.

A simplified state diagram for a single-channel antilock system is shown in Fig. 15.10. In this diagram, a vehicle not braking or decelerating would be in the NORMAL BRAKING state. If antilock action is warranted, it is because the brake pressure on a given channel has caused the wheel to begin to lock; the first action would be to decrease the brake pressure (DECAY state) in an effort to permit the locking wheel to reaccelerate. Fine control of the brake pressure is indicated by the states labeled HOLD OR BUILD/DECAY and SLOW BUILD and course control is indicated by the FAST BUILD state. (The course control is typically used during rapidly changing road surface conditions such as ice-to-asphalt transitions.) During the antilock cycle the state will change, as needed, to attain the type of brake pressure and resulting wheel speed activity as shown in Fig. 15.8. Once the need for antilock action has ended, the END ANTILOCK state is entered, the pump motor is de-energized, the valves are de-energized, and the system can return to the original NORMAL BRAKING state.[4]

How this state-space approach is integrated into a typical microcontroller flowchart is shown in Fig. 15.11. After RESET and INITIALIZATION, a microcontroller enters into a

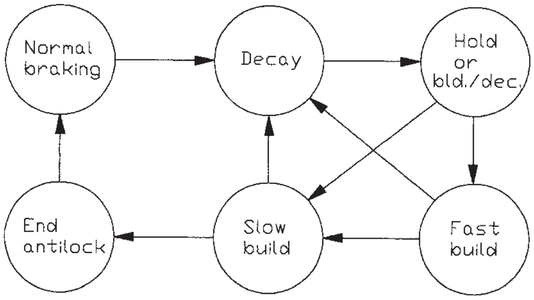


FIGURE 15.10 Simplified single-channel state diagram.

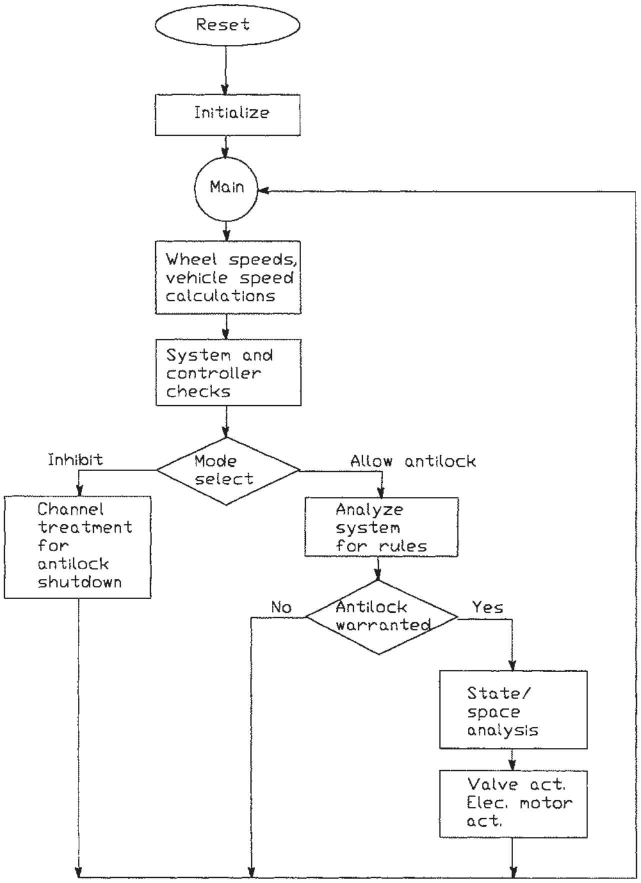


FIGURE 15.11 Simplified antilock flowchart.

MAIN loop that includes extensive system and electronic control unit checks as well as calculations of wheel speeds, prediction of vehicle speed, analysis of conditions warranting antilock action/state-space control law, and valve and motor/pump actuations.

Calculation of wheel speeds consists of scaling the wheel speed sensor inputs to a more usable form and possibly filtering noise due to axle deflection, brake squeal, other electrical systems, etc. A consideration is that the bandwidth of wheel acceleration and deceleration is large—50 g's may be attainable.

The vehicle velocity prediction is critical to many control schemes because wheel velocity relative to vehicle velocity, as well as wheel slip, may be used as a factor in determining appropriate valve action. Vehicle velocity prediction becomes difficult once the wheels begin to lock because the sensors will no longer be reliable indicators of vehicle speed. The methods used to predict vehicle velocity once the wheels have begun to lock consist of a set of rules that have been developed by antilock manufacturers through years of experience to ensure a prediction that has a high degree of accuracy to true vehicle speed.

The antilock system checks typically consist of sensor and valve/motor continuity tests and system voltage range tests. In addition, the checks normally include tests internal to the electronic control unit, such as inter-microcontroller communication.

Once it is determined that conditions are such that antilock action can be safely invoked if warranted, the wheel speed conditions are analyzed to establish the appropriate state for that channel. Primary indicators for most antilock control schemes are wheel slip and wheel deceleration. Another factor considered is the effect on vehicle stability if a particular state is commanded.

Actuation of the valves or electric motor actuators is a direct result of the decisions made in the analysis/state-space logic. Other than actuators requiring pulse-width modulation drives, the actuators normally will remain in the commanded state until the microcontroller loops back through the code (usually a few milliseconds).

15.3.5 Antilock System Testing

Antilock vehicle testing has evolved over the years to include the following most common tests:

- Straight-line stopping
- Braking in a turn
- Split coefficient stopping with associated stability criteria
- Transitional road surface testing including checkerboard and low/high and high/low coefficient surfaces
- Lane change maneuver

All of these tests may be performed on a variety of surfaces, at a variety of speeds, and with lightly loaded and heavily loaded vehicles.

15.4 FUTURE VEHICLE BRAKING SYSTEMS

A trend that will impact braking systems is the industry's desire to reduce vehicle wiring through the use of multiplexing techniques. As increasing numbers of vehicles are outfitted with antilock, this trend is expected to result in an increased number of antilock systems communicating with other vehicle systems through a multiplex link. In addition to the wheel speed/vehicle velocity information available from the antilock system, the antilock electronic control unit could benefit from this technology by being able to receive engine, transmission, steering angle, and other subsystem information.

Another trend in advanced electronically controlled braking systems is vehicle dynamics control during nonbraking maneuvers, as well as during braking. This is accomplished through use of the traction control actuators normally integrated in antilock hydraulic modulators, the addition of sensors to more accurately determine the dynamic state of the vehicle, and communication links with the drivetrain electronic controllers. Vehicle dynamic control holds the promise of safer vehicle operation through improved stability in all maneuvers.

The vehicle brake systems engineering community also is investigating the addition of radar to individual vehicles. This addition could lead to semiautomatic or automatic braking in emergency situations as the brake system anticipates the potential problem and aids the operator in safely applying the vehicle brakes in time to avoid a collision. This concept also lends itself to automatic braking in nonemergency situations to maintain safe distances between vehicles at high speeds.

Continuing interest in electric vehicles and the need for regenerative braking in these vehicles likely will significantly impact future braking systems. It is expected that the regenerative braking function will not be sufficient to provide adequate braking deceleration under all conditions and to provide operators the comfort and safety obtainable with conventional friction brake systems augmented by antilock systems. It is expected that a more complex electronic control system will be used in conjunction with electric vehicles to afford optimum power regeneration without sacrificing braking stopping distance, stability, or steerability.

These trends point to a continued use of friction brake systems through the end of the century and a significant expansion of the role of electronics in these systems.

GLOSSARY

Antilock (or **ABS** or **Antiskid**) A system designed to prevent wheel lock during overbraking.

Antilock hydraulic modulator A hydraulic brake pressure modulation actuator used in antilock systems.

Booster A brake pedal force amplifier, typically vacuum or hydraulically powered.

Booster crack point The brake pedal/push rod travel point initiating booster force amplification.

Booster runout A condition in which the brake booster can no longer provide the gain required due to high input forces and the input force/output force slope becomes less positive.

Brake caliper A part of a disc brake that contains the brake pads and the brake cylinder.

Braking force A force tending to stop a moving vehicle. Usually applied to the force resulting from brake torque being applied to a wheel of a moving vehicle.

Braking maneuver Any vehicle braking action intended to decelerate a moving vehicle, including partial as well as full stops.

Diagonal split brake system A brake system configuration in which a front brake and its opposing rear brake are included in the same brake channel. This technique is used to allow braking on one front wheel in the case of catastrophic failure of the other brake channel.

Disk brake A type of brake characterized by force being applied to both sides of a rotor, thereby creating braking torque.

Drum brake A type of brake characterized by brake force being applied to the inner surface of a drum, thereby creating braking torque.

Dynamic load transfer The characteristic of weight shift during deceleration that places more weight on the front wheels and reduces weight on the rear wheels.

Electric motor/pump The typical hydraulic power source used in antilock systems; an electric motor driving a hydraulic pump.

Lateral force Force perpendicular to the direction of travel.

Longitudinal slip Relative slip between the wheels and the road surface in the direction of travel.

Master cylinder A two-chambered hydraulic cylinder operated by the driver through actuation of the brake pedal.

Proportioning valve A hydraulic valve designed to reduce pressure to the rear brakes relative to the front brakes once a crack point is reached. The valve may be fixed or load-sensing.

Regenerative braking A type of braking used in electric vehicles in which the drive motor is used as a generator during braking, and it serves as the load to brake the vehicle. This technique is used to reclaim a portion of the energy expended during vehicle motion.

Vertical split brake system A brake system configuration in which both front brakes are on one channel and both rear brakes are on the other channel.

Wheel slip The difference between tangential wheel speed and road speed. A rolling tire with no braking torque on it exhibits 0 percent slip; a nonrotating tire on a moving vehicle exhibits 100 percent slip.

REFERENCES

1. J. Y. Wong, *Theory of Ground Vehicles,* John Wiley & Sons, New York, 1978.
2. T. D. Gillespie, *Fundamentals of Vehicle Dynamics,* Society of Automotive Engineers, Inc., Warrendale, Pa., 1992.
3. J. L. Cage, "The Bendix ABS underhood electronic control unit," *Automotive Technology International '90,* Sterling Publications International Limited, London, 1990.
4. J. L. Cage, and J. T. Hargenrader, "Safe electronic modules for antilock applications," *Automotive Technology International '91,* Sterling Publications International Limited, London, 1991.

ABOUT THE AUTHOR

Jerry L. Cage is the manager of electrical/electronics engineering, AlliedSignal Braking Systems. He has over 20 years experience in the design and development of automotive and aerospace control systems and harsh environment electronics. He has been awarded five U.S. patents in electronics and electrohydraulics.

CHAPTER 16
TRACTION CONTROL

Armin Czinczel
Development Engineer
Robert Bosch GmbH

16.1 INTRODUCTION

Traction control systems designed to prevent the drive wheels from spinning in response to application of excess throttle have been on the market since 1987. Vehicles with powerful engines are particularly susceptible to drive-wheel slip under acceleration from standstill and/or on low-traction road surfaces. The results include attenuated steering response on front-wheel-drive (FWD) vehicles, diminished vehicle stability on rear-wheel-drive (RWD) cars, and loss of effective accelerative force.

Large mutual discrepancies in left- and right-side traction levels engender early drive-wheel slip on slick surfaces. Under these conditions, the effective accelerative forces at both drive wheels are limited to a level corresponding to the adhesion available at the low-traction side. The traction control system inhibits wheelspin, allowing the wheel on the high-traction surface to apply maximum accelerative force to the road.

16.1.1 Optimizing Stability (Steering Control)

The essential requirement for systems designed to optimize vehicle stability (with RWD) and steering control (with FWD) is to maintain adequate lateral traction. The most basic arrangements achieve this end by controlling engine torque alone. Both drive wheels transmit the same level of motive force, dosed in accordance with the adhesion available at the low-traction wheel and thereby providing particularly large lateral-traction reserves at the wheel with the greater adhesion. When the traction levels are roughly equal at both drive wheels, the system enhances vehicle stability (steering control) while providing a certain increase in available effective accelerative force beyond that available on an uncontrolled vehicle with slipping wheels.

16.1.2 Optimizing Traction

The optimization of traction becomes a top priority when the motive force must be transmitted to surfaces on which the adhesion varies substantially between sides.

The typical passenger car features a differential unit at the drive axle; this unit allows virtually loss-free differences in wheel speed (for instance, in corners) in combination with uniform torque distribution to the drive wheels. This layout generally provides favorable dynamic vehicle response, as the equal distribution of drive torque inhibits vehicle yaw. However, a difference in the force-transmission potentials at the drive wheels can combine with demands for maximum traction to expose basic liabilities in the design principle.

Figure 16.1 illustrates the system dynamics of drive shaft, differential, and drive wheels on road surfaces affording differing levels of traction with adhesion coefficients μ_H, μ_L (high wheel, low wheel). The torque emanating from the driveshaft is distributed equally between the drive wheels. The low wheel responds to inadequate adhesion potential by spinning during brief wheel acceleration. The accelerative force transmitted through the high wheel then corresponds to the sum of the accelerative force at the low wheel plus its inertia $\Theta_R\,[\dot{\Omega}]/R$. Once the low wheel reaches its terminal speed, the accelerative force available at both wheels is limited to the maximum at the low wheel.

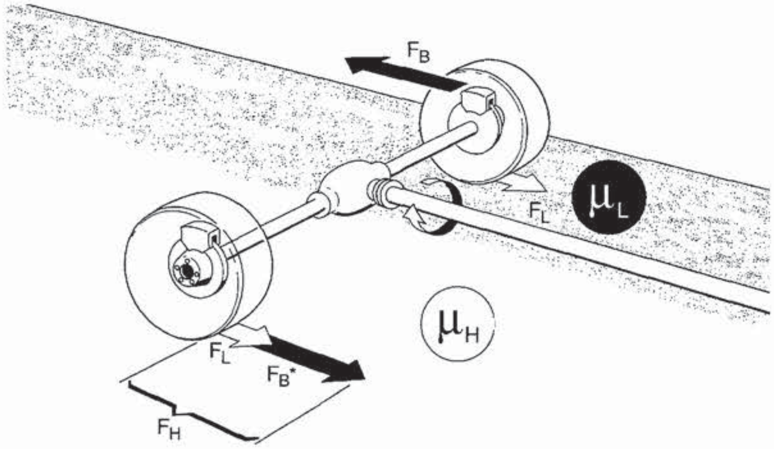


FIGURE 16.1 Braking intervention to limit differential slip.

The only way to increase the accelerative force at the high wheel is to prevent the low wheel from spinning. The first option, application of the wheel brake, is illustrated in Fig. 16.1. The application of braking force F_B at the low wheel prevents it from spinning. This makes the additional accelerative force F_B^* (the product of F_B multiplied by the ratio of effective braking radius to wheel radius) available at the high wheel.

A second option for maximum exploitation of traction potential is represented by the application of fixed, variable, or controlled differential-slip limitation mechanisms. These provide fixed coupling to ensure equal slippage rates at the drive wheels, thereby allowing them to develop maximum accelerative force.

During cornering at high rates of lateral acceleration, lateral variations in drive-wheel load occur, again producing a difference in acceleration potential. Brakes and limited-slip differential arrangements can also be applied to assist in ensuring maximum traction under these conditions.

16.1.3 Optimizing Stability and Traction

Traction-control systems incorporating engine-torque control and supplementary braking intervention (or controlled differentials) can be applied simultaneously to ensure consistent vehicle stability (steering control) and optimal acceleration within the limits imposed by physical constraints. Engine-torque control is the preferred method on road surfaces affording uniform adhesion, while application of braking force (or differential control) provides optimal acceleration at both drive wheels for dealing with surfaces displaying lateral variations in traction.

16.2 FORCES AFFECTING WHEEL TRACTION: FUNDAMENTAL CONCEPTS

The dynamic forces that define the tires' braking response on straights and during cornering are already familiar from the technical literature. The transmission of accelerative force in straight-line operation and in curves is subject to the same qualitative principles that apply during braking. The slip ratio which applies for braking.

$$\lambda_B = \frac{V_F - \Theta_R R}{V_F}$$

is replaced by the ratio

$$\lambda_A = \frac{\Theta_R r - V_F}{V_F} \quad \text{with} \quad \Theta_R r \geq V_F$$

Acceleration slip rates can range all the way from 0 to the very high numbers used to describe the conditions that can occur when the drive wheels spin freely during attempts to accelerate from rest.

Figures 16.2 to 16.4 show acceleration and side-force coefficients as a function of the acceleration slip. Figure 16.2 applies for acceleration during straight-line operation. The demand for reserves in lateral adhesion is fairly diminutive under these conditions (including, for instance, compensation for side winds); thus, traction remains the salient factor.

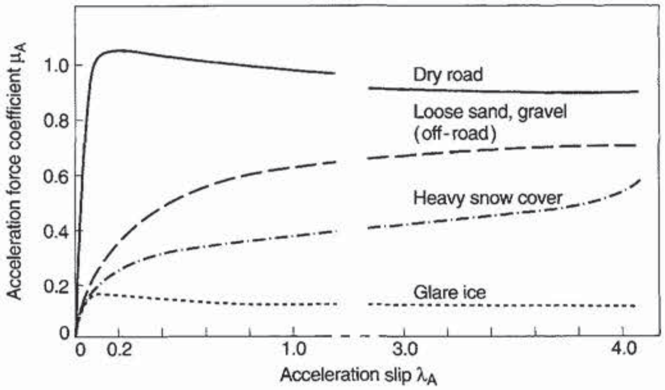


FIGURE 16.2 Adhesion coefficient for acceleration μ_A as a function of acceleration slip λ_A.

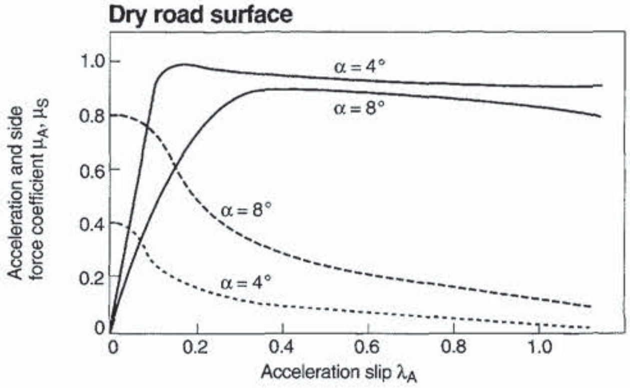


FIGURE 16.3 Acceleration and lateral traction coefficients at different slip angles α.

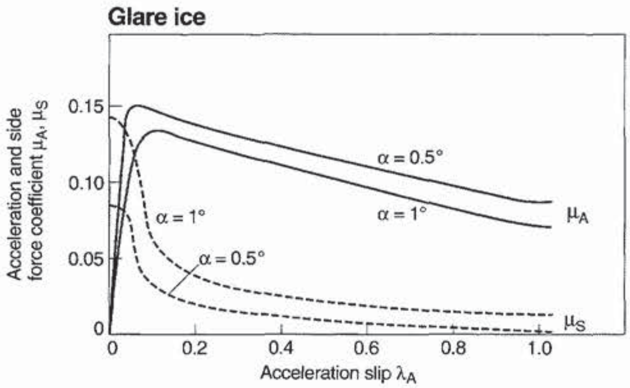


FIGURE 16.4 Acceleration and side-force coefficients at different slip angles α.

On dry road surfaces, maximum accelerative force is available at slip rates of 10 to 30 percent, with traction enhancements of 5 to 10 percent possible relative to spinning drive wheels.

On glare ice, maximum traction is achieved at extremely diminutive acceleration slip levels (2 to 5 percent). On loose sand and gravel and in deep snow (especially in combination with snow chains), the coefficient of acceleration force increases continually along with the slip rate, with the respective maxima only being reached somewhere beyond 60 percent. Thus, the slip rates of 2 to 20 percent found within the ASR's operating range will not provide adequate traction under all operating conditions.

For this reason, all known ASR systems incorporate slip-threshold switches or ASR deactivation switches, which allow the vehicle operator to either reset the ASR slip-control threshold to substantially higher levels, or to switch the system off entirely should the need arise.

Figures 16.3 and 16.4 apply to acceleration during cornering; under these conditions the drive wheels are subject to various degrees of lateral force as a function of the vehicle's rate of lateral acceleration. Increasing acceleration slip (and increasing accelerative forces) cause a drop in the lateral forces, which then respond to still higher slip rates by collapsing to small residual levels.

Figure 16.3 represents the response pattern on a dry road surface. The curve starts at a rate of acceleration slip of zero. Initially, the side-force coefficient displays a moderate downward trend. However, continuing increases in the coefficient of acceleration force induce a substantial fall in the side-force coefficient. The figure shows that the accelerative force must be limited to a fraction of its ultimate potential if sufficient lateral forces are to be maintained.

On glare ice (Fig. 16.4), the extremely limited friction potential means that vehicle stability under acceleration remains available only at relatively small slip angles (ca. ≤2°). Relatively diminutive slip angles (≤0.05°) will be sufficient to induce a radical drop in the side-force coefficient. This makes it clear that an extremely precise and sensitive slip control is required on glare ice (and other low-friction road surfaces). The traction-control system must thus exhibit a high degree of monitoring accuracy, while signal processing and actuation of the final-control elements must be rapid and precise.

16.3 CONTROLLED VARIABLES

The four wheel speeds used for the ABS supply the following closed-loop control parameters for the ASR traction control system: the acceleration slip from the lateral variation in the rotation speeds of the driven and nondriven wheels, and the angular acceleration of the driven wheels.

The following secondary control parameters are also calculated: vehicle velocity and acceleration based on the speeds of the nondriven wheels, and curve recognition, derived from comparisons of the speeds of the nondriven wheels.

The target value for acceleration slip is defined as the mean rotational velocity of the nondriven wheels plus a specified speed difference known as the slip threshold setpoint. The main goal of regulating acceleration slip can thus be divided into two subsidiary objectives: closed-loop control of acceleration slip to maintain slip rates at the specified levels with maximum precision, and calculation of optimal slip setpoints for different operating conditions and their implementation as control objectives.

Depending on the final-control strategy being used, various control concepts can be employed to meet the first objective. With throttle-valve control, a setpoint calculated from a number of signals is adopted for regulation as soon as the closed-loop control enters operation. The subsequent control process basically corresponds to that of a PI controller. When the brakes are used, arrangements are necessary to compensate for the nonlinear pressure-volume curve which governs the response in the brake calipers. The first stage of the closed-loop control program thus employs a sensing pulse corresponding to a relatively large volume; this compensates for compliance in the brake caliper. In the next stage, the system responds to positive deviations from the setpoint with graduated pressure increases; the rate of increase corresponds to the degree of divergence. A subsequent drop below the control setpoint initiates a pressure-relief stage (sequence of defined pressure-relief and holding phases). This impulse series, in which the length of the pressure-relief phases increases continually, is followed by termination of braking intervention.

Ignition and fuel-injection intervention essentially conform to the D controller closed-loop control concept. The difficulty associated with determining satisfactory setpoint values results from the fact that optimal acceleration and lateral forces cannot be achieved simultaneously. The ASR control algorithm must therefore meet varying operator demands for linear traction and lateral adhesion by using priority-control strategies and adaptive response patterns.

High vehicle speeds are accompanied by lower operator requirements for traction, especially with low coefficients of adhesion. At the same time, reductions in vehicle stability and steering response are not acceptable. The control strategy is thus designed to provide progressively lower slip threshold setpoints as the vehicle speed increases, with priority being shifted from linear traction to lateral adhesion.

The vehicle's acceleration rate and the regulated level of engine output provide the basis for reliable conclusions regarding the coefficient of friction. Thus, another important strategy takes into account the coefficient of friction at the road surface. The slip threshold setpoint is raised in response to higher friction coefficients. This ensures that an ASR system designed for optimum performance on low-friction surfaces will not intervene prematurely on high-traction surfaces.

Yet another important control strategy is based on the cornering detection mentioned previously. This system employs the difference in the wheel speeds of the nondriven wheels as a basis for reductions in the slip setpoint to enhance stability in curves. This speed differential can be used to calculate the vehicle's rate of lateral acceleration. A large discrepancy indicates a high rate of lateral acceleration, meaning that a high coefficient of friction may also be assumed. In this case, the slip setpoint should not be reduced, but rather increased.

16.4 CONTROL MODES

16.4.1 Modulation of Engine Torque

Various control intervention procedures can be employed, either singularly or in combination, to regulate engine torque:

- Throttle-valve control with the assistance of the electronic performance control or an automatic throttle-valve actuator (ADS)
- Adjustment of the ignition-advance angle
- Selective ignition cutout, combined with suppression of fuel injection
- Fuel injection suppression alone

Slippage at the drive wheels generally occurs in response to an excess of torque relative to the coefficient of friction available at the road surface. Controlled reduction of engine torque is thus a logical step. It is always the most suitable method in cases where virtually identical adhesion is present at both drive wheels. At the same time, the response times for the individual engine controls must be considered if adequate vehicle stability is to be ensured (see Fig. 16.5).

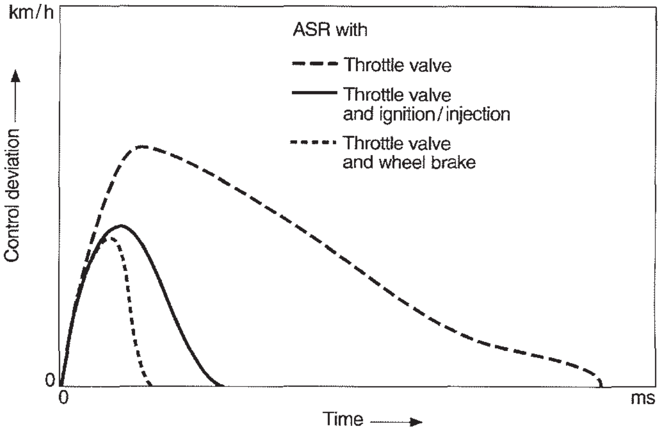


FIGURE 16.5 Deviation of the controlled variable during first control cycle with different actuators.

If control is restricted to the throttle-valve position alone, the throttle valve's response time, response delays within the intake tract, inertial forces in the engine, and drivetrain compliance will all result in palpable wheel slippage continuing for a relatively long period of time. Throttle regulation alone cannot ensure adequate vehicle stability on rear-wheel-drive vehicles. This qualification is particularly applicable to vehicles with a high power-to-weight ratio. On FWD and 4WD vehicles, throttle-valve control alone can be sufficiently effective if response delays are minimized.

Arrangements combining throttle-valve regulation with interruptions in the fuel injection produce substantial reductions in the amplitude and duration of wheel slip. Thus, this concept can be used to guarantee good vehicle stability regardless of which axles are driven.

In principle, it is also possible to design an ASR system based solely on regulation of the ignition and injection systems. This concept employs a system with sequential fuel injection. It alternately cuts out individual cylinders, while the ignition is also adjusted for the duration of the control process. Although this concept can be employed to ensure adequate vehicle stability regardless of drive configuration, certain sacrifices in comfort are unavoidable, especially during operation on ice and during the warm-up phase.

16.4.2 Brake Torque Control

The brakes at the drive wheels are capable of converting large amounts of kinetic energy into heat, at least for limited periods of time. In addition, the response times can be held extremely short, making it possible to limit slippage increases to very low levels.

ASR systems relying exclusively on braking intervention appear suitable for regulating spin at the drive wheels.

Traction enhancements during starts and under acceleration on road surfaces affording varying levels of adhesion at the left and right sides are especially significant with this system.

The ASR hydraulic unit used to generate the braking forces employs components which are already present for the ABS. Cost considerations make it important that ASR hydraulic systems require an absolute minimum in additional components beyond those already available for the ABS.

The hydraulic concepts can be classified in two categories, according to whether stored hydraulic energy is employed or not. A dual-strategy system including rapid braking intervention with stored hydraulic energy is always to be recommended where the engine torque control is based entirely on throttle valve adjustments with their relatively long response times.

Figures 16.6 and 16.7 illustrate two examples of ASR braking intervention using stored energy.

Brake Torque Control with Stored Energy. In this ASR system, designed for RWD vehicles in the upper price range, engine torque is regulated exclusively by the engine performance control (EPS) unit. Rapid braking intervention is required for enhanced regulation of vehicle stability; this also ensures optimal traction at both drive wheels, especially with lateral variations in adhesion potential. This system offers optimal vehicle stability and traction combined with a high level of comfort. Figure 16.6 illustrates the design of the integrated ABS/ASR hydraulic unit.

The primer pump draws brake fluid from the master-cylinder reservoir and supplies it to the ABS return pump under small positive pressure. To meet the requirements of the ABS system, the dual-circuit ABS return pump is expanded to include a third plunger, which assumes responsibility for the ASR storage charge. For "rapid" braking intervention, the pressure in the wheel cylinder must escalate from 0 to 50 bar in less than 200 ms. A hydropump, with its limited supply capacity, is not capable of providing this order of pressure build-up in the required time, necessitating the use of a high-pressure accumulator to supply the required brake fluid with adequate alacrity.

When the traction control system is activated, the switchover valve moves to the third position and brake fluid from the reservoir is supplied to the two ABS/ASR control valves. While

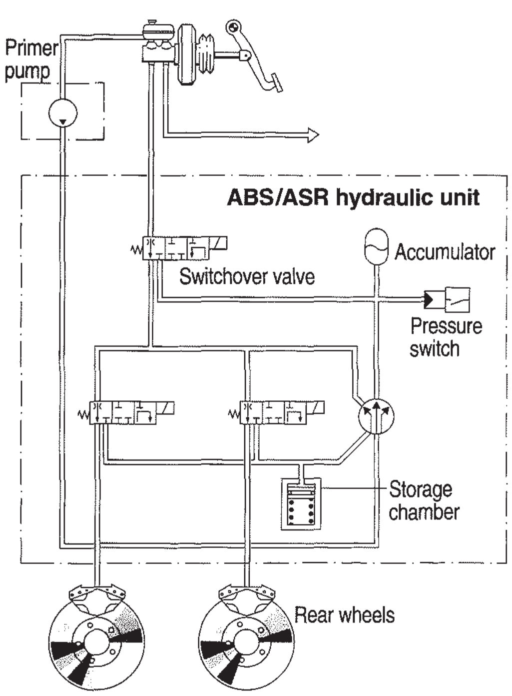


FIGURE 16.6 Diagram showing the ABS/ASR hydarulic circuit using braking force with stored energy (integrated system).

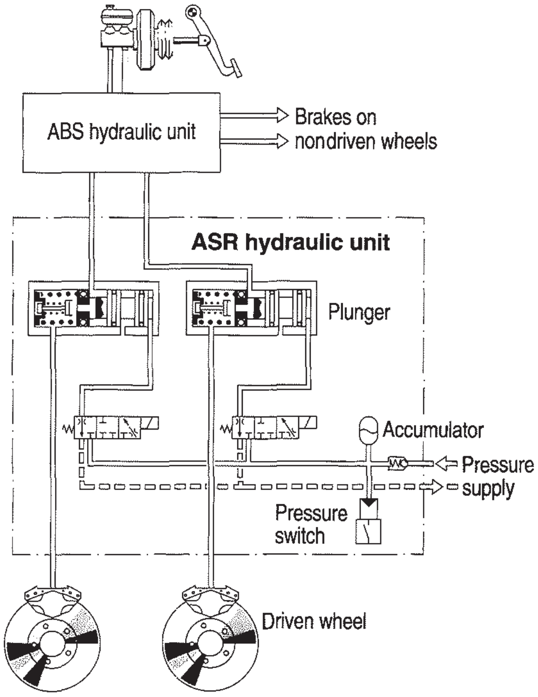


FIGURE 16.7 Diagram showing hydraulic unit for ASR system with braking intervention using stored energy via plunger.

an ABS solenoid valve can satisfy the requirements of the ABS hydraulic circuit, two solenoid valves are needed for the ABS/ASR hydraulic unit in order to allow individual braking control at the drive wheels.

These two solenoid valves regulate brake pressures at the driven wheels in accordance with the ASR control commands. Accumulator charging is controlled by the pressure switch and continues during normal vehicle operation. A special low-noise pump design is employed to meet stringent customer specifications.

ABS/ASR System with Brake Intervention Using Stored Energy (separate design). One version employs a high-pressure accumulator to provide hydraulic energy for ASR braking. An engine-driven pump (level control) supplies hydraulic fluid to the accumulator. The system includes a plunger for each driven-wheel brake. An additional 3/3 ABS valve is also installed to govern the supply of hydraulic fluid from the reservoir to the primary side of the plunger and to separate this medium from the brake fluid present on the secondary side. The solenoid valves respond to the position commands from the ABS/ASR control unit to control pressure accumulation, maintenance, and release operations to maintain the required pressure level on the plungers' primary sides. The plungers move to the left. Initially, a central valve closes to block the connection with the master cylinder. The plunger's displacement pressurizes the brake fluid on the secondary side to produce the desired pressure level in the wheel cylinder.

ASR designs relying on stored energy to activate plungers are always worthy of consideration in those cases where the vehicle is already equipped with a hydraulic energy supply for other purposes (e.g., power steering, level control).

Brake Torque Control without Stored Energy. Each of the ASR hydraulic units described requires a high-pressure accumulator to ensure that the braking energy can be provided quickly enough. This means additional design complication with attendant expenditure.

Another system differs from those already described by using the supply circuit of the ABS return pump exclusively to regulate the braking force at the drive wheels. The return pump forms part of a self-priming circuit, thus employing the (already installed) ABS return pump as an inexpensive source of energy for braking. The ASR braking function can thus be achieved with a minimum of additional design complication.

Figure 16.8 shows the hydraulic circuit for a passenger car with a K-pattern brake circuit and front-wheel-drive. When the intake valve opens, the self-priming pump extracts brake fluid from the reservoir and draws it into its circuit before supplying it directly to the ABS/ASR control valve at the drive wheel. The control valve regulates braking force by building up, maintaining, and releasing pressure in accordance with the respective position commands from the control unit. The fluid bled by the control valve in the pressure-release mode is returned to the pump's intake side. A switchover valve connects the high-pressure side of the self-priming return pump with the second brake circuit (governing the other front-wheel brake) and the pressure-relief valve responsible for regulating the ASR system pressure. With the exception of the suction line, the second brake circuit features a symmetrical layout. The passage to the master cylinder is also closed.

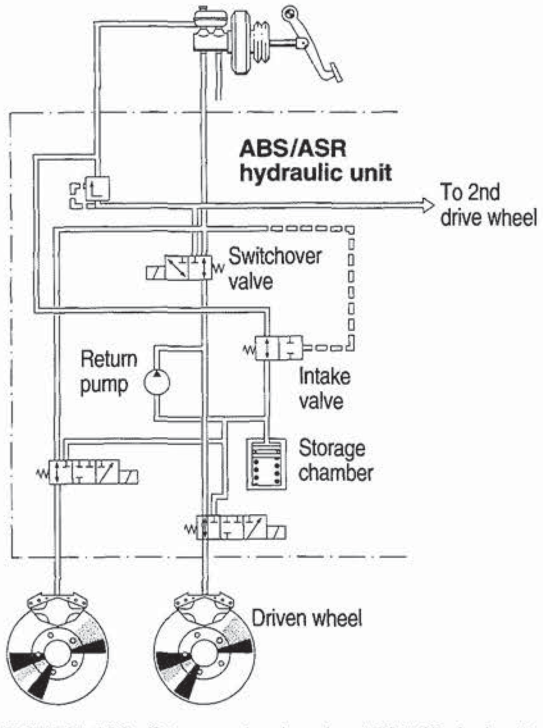


FIGURE 16.8 Diagram showing the ABS/ASR hydraulic unit employed for braking intervention without stored energy.

This ASR system can be installed in rear-wheel-drive vehicles with TT brake circuit configurations with even less expenditure: because both drive-wheel brakes are then in the same circuit, only a single switchover valve is required.

New Design for Brake Torque Control without Stored Energy. The principle of braking intervention without stored energy has been developed further for a new generation of ASR systems entering production in 1993. (See Fig. 16.9.)

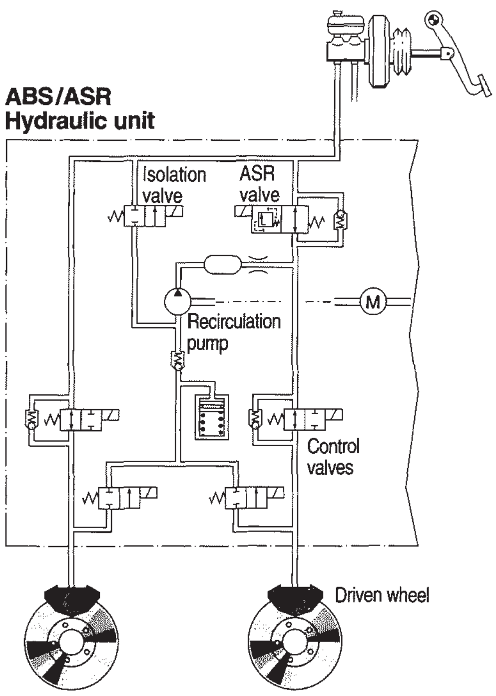


FIGURE 16.9 New design for brake torque control without stored energy.

The self-priming recirculation pump draws fluid from the master cylinder through an electrically controlled isolation valve, obviating the requirement for an additional intake line. Meanwhile, the pressure-relief valve can be integrated within the ASR valve. This configuration allows creation of a self-contained system combining simplified installation with enhanced safety. In addition, the reduction in the number of components improves reliability.

The 3/3 valves employed on earlier versions have been replaced by extremely small 2/2 solenoid valves.

16.4.3 Differential Slip Modulation

ABS/ASR System Regulating Differential Slip on Rear-Wheel-Drive Vehicles. This system controls the differential's lateral slip to improve traction for starting off and for simultaneous acceleration and cornering on road surfaces affording different levels of traction from left to right. The slip-limitation mode remains active until a specific vehicle speed is attained, and is deactivated completely at higher speeds.

When the vehicle starts off, the rotation speeds of the wheels on the outside of the curve are subjected to a mutual comparison. The lock is activated once a specific difference in the two speeds is exceeded. There then follows a comparison of the rotation speeds of the driven wheels. The lock is deactivated as soon as the difference in rotation speeds drops below a specified level.

ABS/ASR System for Controlling Rear-Axle and Interaxle Differential Slip Rates on 4WD Passenger Cars. This version incorporates an additional interaxle control feature.

Four-wheel drive passenger cars employ a specific fixed front-to-rear distribution of engine output to provide optimum vehicle characteristics within the stable range (that is, with limited amounts of acceleration slip). One or both wheels at either axle can respond to throttle application on low-traction surfaces with immoderate wheelspin. This is where the interaxle slip limiter is activated to adapt the distribution of engine torque to the traction available at the respective axles, thereby improving traction while also enhancing vehicle stability and steering response.

Figure 16.10 illustrates the design of the hydraulic unit used to control the lateral and interaxle slip-limitation mechanisms. An electric pump supplies a high-pressure accumulator. The accumulator, in turn, provides pressure to a 3/3 solenoid valve for control of the lateral and interaxle locking mechanisms. The return volume from the pressure-relief phase is conducted to a separate reservoir.

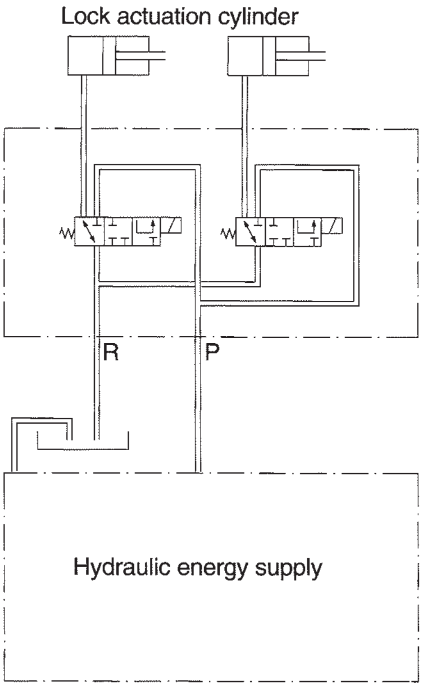


FIGURE 16.10 Hydraulic unit for control of rear-axle and interaxle differential locks.

16.5 TRACTION CONTROL COMPONENTS

The following is a selection of the components used in Bosch traction control systems.

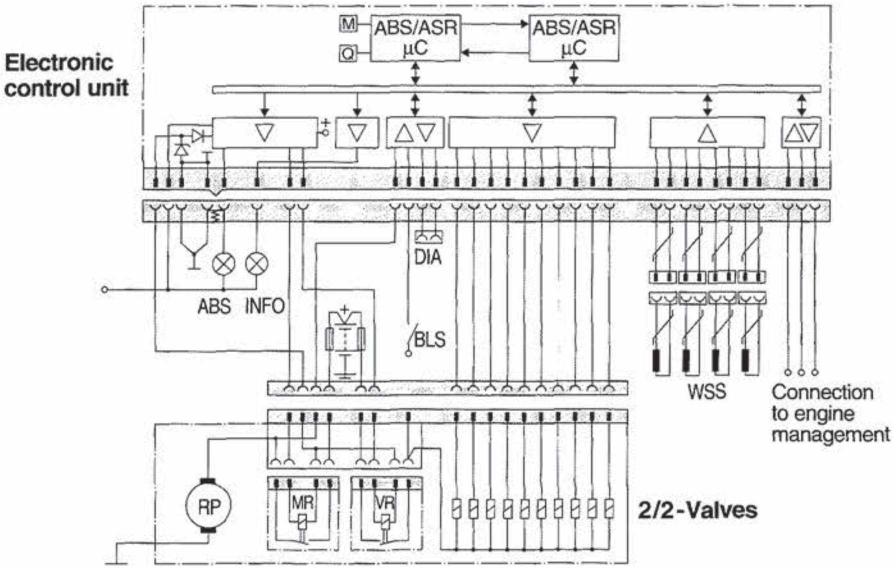
16.5.1 Wheel-Speed Sensors

The system employs the same wheel-speed sensors that provide the information for the antilock braking system.

16.5.2 Electronic Control Unit

Figure 16.11 shows an ASR circuit diagram. An input amplifier IC receives the signals from the wheel-speed sensor; the signal frequency indicates the wheel speed. Two microcontrollers then process the signals to determine the wheel speed and acceleration rate. These data, in turn, provide the basis for calculations to determine the actual and desired values for slip control. Overall signal processing, the control algorithm, and the monitoring software are present in each of these microcontrollers to provide the system with backup capabilities.

Three output amplifiers control the solenoid valves, the ABS and ASR indicator lamps, the driver-information lamp, and the motor and valve relay. An additional IC is required to monitor the braking requirement and pump-motor voltage and for diagnosis.



M :	memory
Q :	quartz
ABS/ASR μC :	ABS/ASR microcomputer
ABS :	ABS warning lamp
INFO :	information lamp
DIA :	diagnosis connector
BLS :	brake light switch
WSS :	wheel speed sensors
RP :	recirculation pump motor
MR :	motor relay
VR :	valve relay

FIGURE 16.11 Operation diagram for the complete system.

16.5.3 Hydraulic Unit

Four examples of ASR hydraulic units were given in Sec. 16.4.2.

16.5.4 Electronic Throttle Control Actuator

See Sec. 10.3.1.

16.5.5 Simplified Throttle Control Actuator

See Sec. 10.3.1.

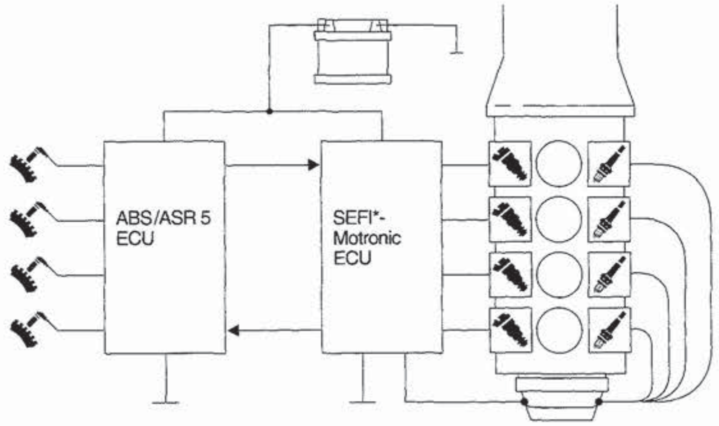
16.5.6 Fuel Injection and Ignition Control

This system reduces engine output by suppressing the fuel injection process.

Complete fuel injection suppression would lead to a total loss of engine output—a smooth, graduated response would be impossible with this kind of arrangement. In contrast, selective suppression of the injection process at individual cylinders can be employed to achieve a good compromise between quick response and a graduated reduction of engine power. This is the design principle behind the new concept.

With suppression according to individual cylinders, the number of control increments is the same as the number of cylinders. Because this limited number of control stages is still inadequate for a (as an example) four-cylinder engine, a supplementary strategy is employed: this is referred to as alternating injection suppression. It consists of varying the number of active cylinders by one after every two crankshaft rotations to produce a mean torque lying between the torques produced at the two cylinder stages. This method doubles the number of control stages to achieve an acceptable level of driving comfort, while complementary reductions in ignition advance can be employed to provide additional incremental adjustments.

In cases where the excess torque is substantial, injection suppression can be supplemented by short-term ignition cutout to provide extremely rapid output reductions. Figure 16.12 shows the design of the system. In addition to the modest expense, this system also offers vehicle manufacturers an additional advantage in the form of space savings (no additional space required) and simplicity (limited amount of extra wiring).



* SEFI = Sequential Electronic Fuel Injection

FIGURE 16.12 ASR EZ system.

16.6 APPLICATIONS ON HEAVY COMMERCIAL VEHICLES

Heavy commercial vehicles are used in a highly variegated range of applications. In principle, the ASR installed on these vehicles employs the same control strategies used for passenger cars: engine-output control and braking intervention.

Differences in vehicle application can make it necessary to employ ASR systems in various levels of complexity (for instance, relying on braking control or engine output control exclusively). The control unit recognizes the design stage and carries out its control functions accordingly. This makes it possible to employ the most economical ASR system for each vehicle type and particular application.

Figure 16.13 features a schematic diagram showing a typical 4×2 vehicle equipped with a top-of-the-line ABS/ASR system. During acceleration on μ-split surfaces, the brake-force regulator limits slip between the drive wheels. The ASR valve (4) is activated on the side with the spinning wheel, while the ABS pressure-control valve (3) allows graduated increases in pressure at the wheel cylinder.

If both drive wheels start to spin on a road surface affording equal traction on both sides, the engine-output controller responds by reducing the drive slip to optimum values. In this example, the electric performance control (EPC) (10) assumes the role of final-control ele-

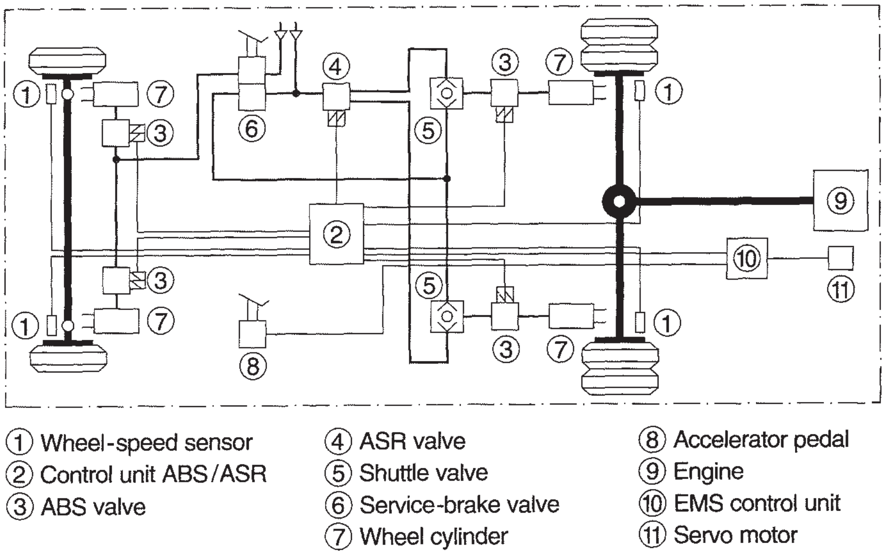


FIGURE 16.13 Four-circuit ABS/ASR, 4 × 2 vehicle.

ment. The ABS/ASR control unit (2) transmits the desired increment of reduction to the EPC control unit via interface, and also controls the electric motor at the injection pump.

16.7 FUTURE TRENDS

Ensuring driving stability is the most important task of ASR with rear-wheel-drive cars. This task can be achieved by a fast engine torque control or a combination of throttle control and a fast brake control.

ASR systems with a fast engine torque control (ignition and injection intervention) and brake control without stored energy will be widely used.

Although the application of controllable differential locks also offers efficient ASR control, the higher costs of this system will prevent wide usage.

The predominant demand on ASR systems for front-wheel-drive cars is that of traction optimization. Therefore, an ASR system with brake torque control is needed. The combination of brake torque control and engine torque control results in a complex, efficient system.

In the future, only ASR systems with brake torque control will be widely used with front-wheel-drive cars. Especially the combination of brake torque control and engine torque control with ignition and injection intervention will be widely used.

GLOSSARY

ABS return pump A piston pump that draws back brake fluid to the master cylinder.

Alternating injection suppression A variation of the number of active cylinders by one after every two crankshaft rotations in order to modulate the engine torque.

ASR deactivation switch A device to switch off ASR on sand and loose gravel in order to achieve maximum traction on these surfaces.

Automatic throttle valve actuator A simple actuator for automatic throttle angle reduction in case of excessive acceleration slip.

Braking intervention Automatic brake application at drive wheels in case of excessive acceleration slip.

D controller A controller with differentiating characteristics.

Electronic performance control Electronic accelerator control.

Engine torque control An actuator to modulate engine torque in case of excessive acceleration slip.

PI controller Controller with proportional and integral characteristics.

Slip threshold switch Switch to increase desired slip threshold on sand and loose gravel.

Switchover valve A valve to switch hydraulic performance from normal braking to ASR performance.

Throttle valve control ASR actuator to modulate the throttle angle.

3/3 ABS valve A valve with three connections and three positions for ABS wheel pressure modulation.

BIBLIOGRAPHY

Maisch, Wolfgang, Jonner Wolf-Dieter, and Alfred Sigl, "Traction control—a logical addition to ABS," *SAE International Congress and Exposition,* Detroit, Feb. 23–27, 1987.

Kolberg, Gerhard, "Elektronische Motorsteuerung für Kraftfahrzeuge," *Motortechnische Zeitschrift,* Germany, 1985.

Demel, H., and A. Czinczel, "ABS-task, design, and function," *I Mech E Conference,* London, 1988-6, pp. 89–94.

Buschmann, G., N. Ocvirk, and P. Volz, "Integrated traction control function—the consequent extension of highly sophisticated anti-lock systems," *I Mech E Conference,* 1988-6, pp. 103–112.

Klein, Hans-Christof, "PkW-ABV-Bremssysteme mit weiteren integrierten Funktionen," *Automobilindustrie,* May 1989, pp. 659–673.

Kupper, R., and N. Hover, "Anti-slip-control system with high safety and comfort," *I Mech E,* C382/113, 1989, pp. 523–531.

Sigl, A., and H. Demel, "ASR-traction control, state of the art and some prospects," SAE 900.24.

Sigl, A., and A. Czinczel, "ABS/ASR5—the new ABS/ASR system for passenger cars," *I Mech E Conference,* London, 1993, pp. 81–87.

ABOUT THE AUTHOR

Dipl. Ing. Armin Czinczel received a degree in mechanical engineering at the Technical University in Hanover, Germany. He started his professional career with development work on navigation systems. He joined Bosch in 1968, working on ABS development until his retirement in 1994.

CHAPTER 17
SUSPENSION CONTROL

Akatsu Yohsuke
Manager, Vehicle Research Laboratory
Nissan Motor Co., Ltd.

The function of a suspension system in an automobile is to improve ride comfort and stability. An important consideration in suspension design is how to obtain both improved ride comfort and stability, since they are normally in conflict. Advances in electronic control technology, applied to the automobile, can resolve this conflict.

17.1 SHOCK ABSORBER CONTROL SYSTEM

During the past 20 years, many different damping control systems have been studied. The main purpose of all these systems is to select the optimum damping force for various driving conditions. The first function of a shock absorber is to control vehicle movement against inertial forces, such as roll when the vehicle turns and pitch when the vehicle is braked. The second function is to prevent vehicle vibration caused by road surface inputs. To satisfy both functions it is necessary to control damping forces.

There are three basic parts of a damping control system: a damping control device (actuator), sensors, and software (control strategy). Optimum damping forces should be set for various running conditions in order to improve ride comfort and handling stability.

17.1.1 System Configuration

One of the damping control system configurations is shown in Fig. 17.1. This system uses five sensors, including a supersonic road sensor, to detect running conditions. Control signals are sent to adjust the damping force of the variable shock absorbers to optimum values. A main advantage to this type of system is that, through the use of a road sensor, it can provide optimum control in accordance with the actual road conditions. This system incorporates three discrete damper characteristics. Sensors used are: a vehicle speed sensor, a steering angle sensor, an acceleration and deceleration sensor, a brake sensor, and a supersonic sensor to detect road conditions.

One system uses four piezo sensors and four piezo actuators on each wheel in order to change the damping forces as quickly as possible. This system incorporates two discrete damper characteristics. Sensors used are: four piezo sensors on each wheel, a stop lamp switch, a steering sensor, and a vehicle speed sensor.

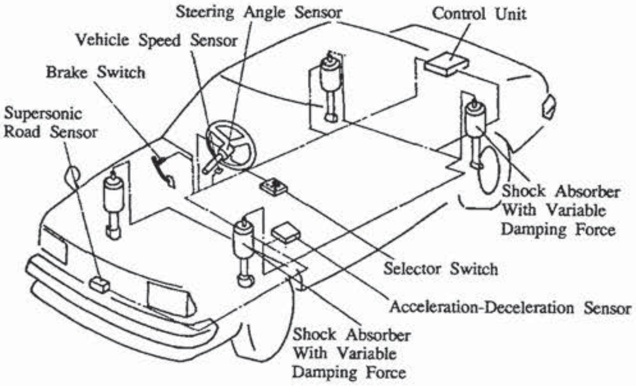


FIGURE 17.1 Principal components of a damping control system.

17.1.2 Actuator and Sensor

Actuator. Discrete damping control actuators often use a built-in motor to change the damping force. This motor turns a rotary valve to select the orifice diameter for three different damping levels: soft, medium, and hard. The stopping position of the rotary valve is controlled by encoder signals.

Another actuator is a piezo actuator consisting of 88 piezo elements, and it is installed in the piston rod of a shock absorber. When a high voltage (500 V) is applied to the piezo actuator, it expands about 50 μm with reverse piezoelectric effect. Elongation in the piezo actuator causes the plunger pin to be pushed out through the displacement hydraulic coupling unit. As a result, the plunger pin moves down to open the bypass of the damping force switching valve. The result is a soft damping force. Figure 17.2 shows this valve.

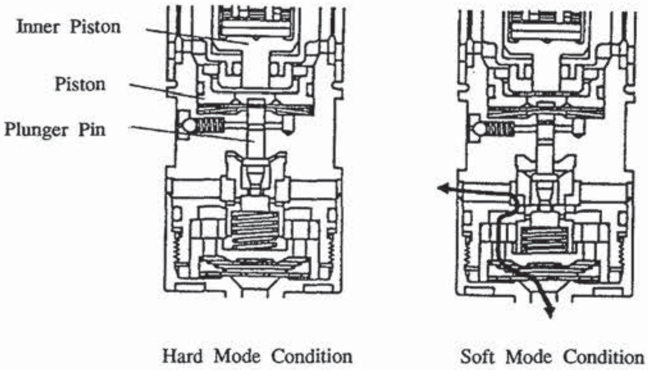


FIGURE 17.2 Damping force switching valve.

Sensor. A sensor using a supersonic wave to detect the road surface is shown in Fig. 17.3. The vehicle height from road to body is calculated on the basis of reflection time T. Judgment of the road condition is made by analyzing the pattern of change in vehicle height.

A sensor using the piezoelectric effect is shown in Fig. 17.4. Installed in the piston rod of a shock absorber, the sensor generates an electric charge in accordance with axial force from the road surface.

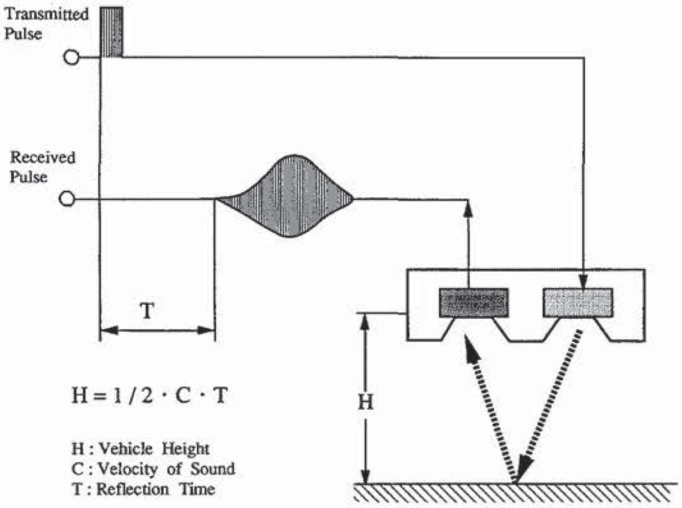


FIGURE 17.3 Supersonic sensor.

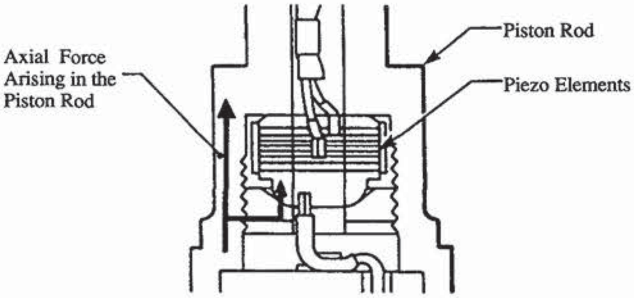


FIGURE 17.4 Piezoelectric sensor.

17.1.3 Control System

The control system of Fig. 17.1 is shown in Table 17.1. The occurrence of roll and pitch can be predicted from various sensors. Single bumps or dips are detected from changes in vehicle height. The outline of the road surface judgment logic is shown in Table 17.2.

The block diagram of the system using four piezo sensors and four piezo actuators is shown in Fig. 17.5. The damping force arising in an absorber increases instantaneously as the tire or wheel goes up or down in accordance with unevenness of the road surface. Each piezo sensor outputs to the electronic control unit continuous signals indicating damping force differential. If the value of damping force differential exceeds a predetermined level, the system switches from firm mode to soft mode. The system is designed so that the firm mode re-engages immediately after the vibrations due to poor road surface cease.

Many of the limitations and disadvantages of the conventional shock absorber can be eliminated by the damping control systems. They include semiactive suspension systems which are capable of providing both ride comfort and good handling.

TABLE 17.1 Damping Force Control System

Control objectives		Sensors used					Damping force
		Vehicle speed	Steering angle	Accel/ decel	Brake	Road condition	Front Rear
Roll	Roll reduction for quick steering operation	○	○				Hard Hard
Pitch	Reduction of nose diving by braking				○	○	Hard Hard
	Reduction of pitching when accelerating and decelerating	○		○			Medium Medium
Bouncing	Reduction of light, bouncy vibrations in bottoming	○				○	Medium Medium
	Reduction of light, bouncy vibrations in bouncing on a heaving road	○				○	Medium Medium
Road holding performance	Road holding performance improvement when running on rough roads	○				○	Medium Medium
Others	Stability improvement at high speed	○					Medium Soft
	Prevention of shaking when stopping and rocking when passengers exit or enter	○					Hard Hard

TABLE 17.2 Road Condition Judgment Logic

High-frequency components	Low-frequency components	
	Small	Large
Small	Smooth road damping force control unnecessary	Heaving road
Large		Rough road

17.2 HYDROPNEUMATIC SUSPENSION CONTROL SYSTEM

A hermetically sealed quantity of gas is used in the hydropneumatic suspension control system. The gas and hydraulic oil are separated by a rubber diaphragm, as shown in Fig. 17.6*a*. The mechanical springs are replaced by gas. The shock absorber damping mechanism is achieved by the orifice fitted with valves.

17.2.1 System Configuration

As shown in Fig. 17.6*b*, by adding an additional sphere to the hydropneumatic system, a controllable hydropneumatic system can be realized. If the regulator is closed, the system is in a firm mode. If the regulator is open, the spring constant of the suspension system becomes lower by increasing the total volume of the sphere, and the total damping force is reduced.

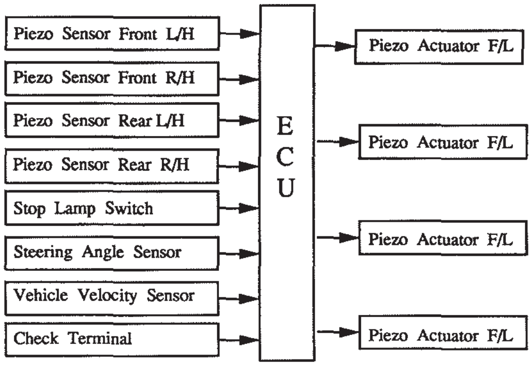


FIGURE 17.5 Block diagram of damping control system.

Depending on the sensors used, which detect vehicle driving and road surface conditions, this system can change the regulator characteristics in order to achieve both good ride comfort and handling stability.

17.3 ELECTRONIC LEVELING CONTROL SYSTEM

In a pneumatic and hydropneumatic suspension, the vehicle body can be maintained at a constant height from the road surface, keeping a low spring constant. The advantages of an electronic control system are:

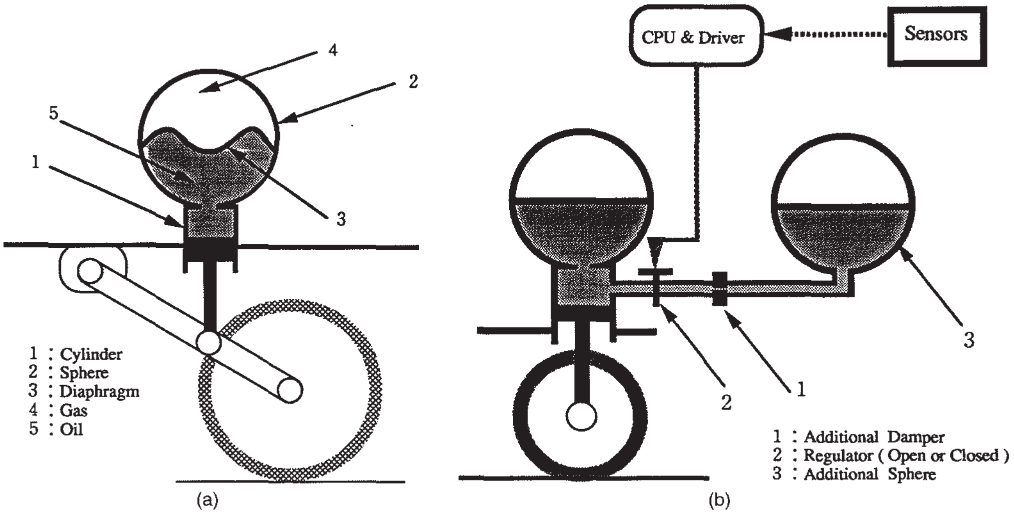


FIGURE 17.6 (*a*) Hydropneumatic suspension system; (*b*) Controllable hydropneumatic suspension system.

- Keeping a low spring rate to achieve good ride comfort independent of load conditions
- Increase in vehicle body height on rough road surfaces
- Changing spring rate and damping force in accordance with driving conditions and road surfaces

17.3.1 System Configuration

The system is shown in Fig. 17.7. It consists of eight sensors, a mode select switch, air spring/shock absorber units on four wheels, actuators to operate the changing valves in the unit, a compressor unit and five height control valves for air springs, and an electronic control unit (ECU). The system configuration is shown in Fig. 17.8.

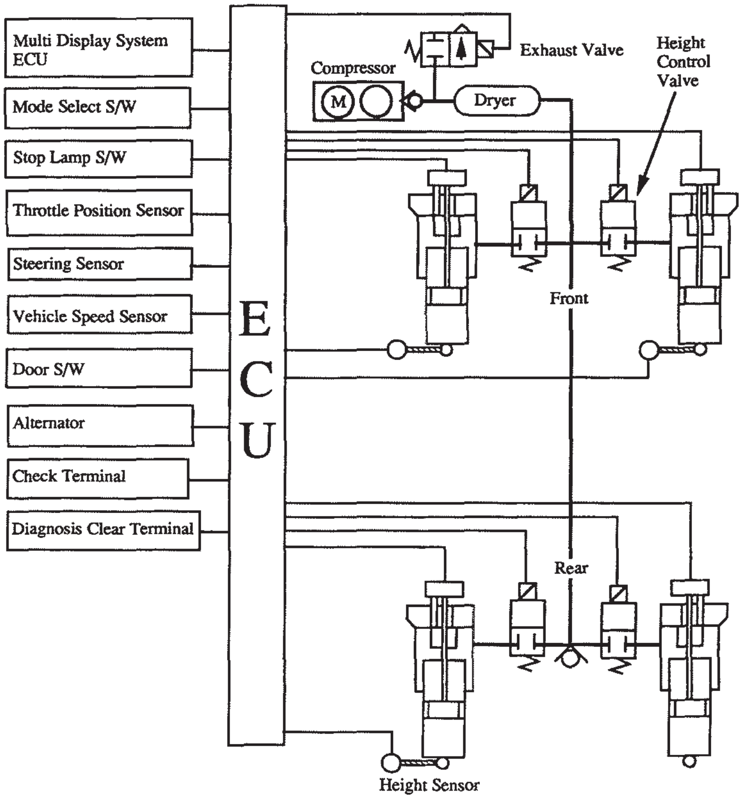


FIGURE 17.7 Principal components of an air suspension system.

17.3.2 Components

The structure of the air suspension unit consists of a shock absorber, a pneumatic piston surrounding the shock absorber, main and sub-air chambers, a rolling diaphragm, and valves which change the suspension stroke.

The actuator uses a dc motor, which has two shafts to operate the valves for the air spring and the shock absorber. The rotation of the motor is reduced by the sector gear and operates

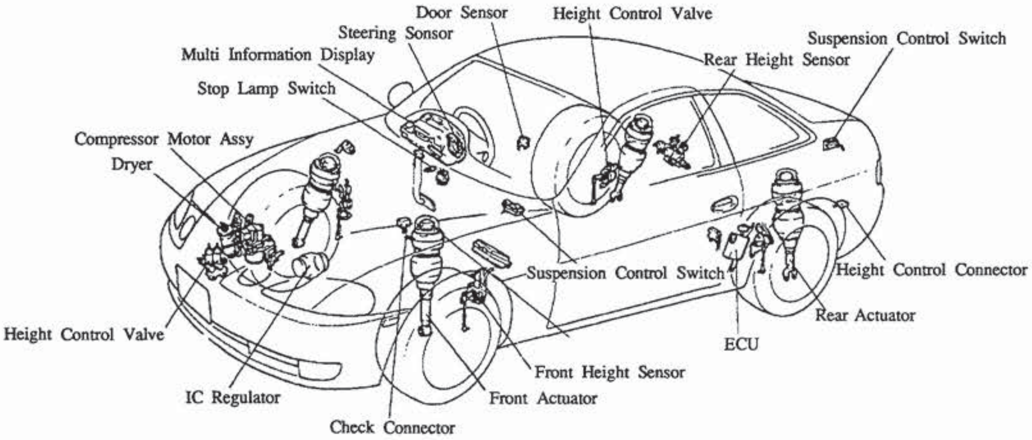


FIGURE 17.8 Configuration of an air suspension system.

the rotary valve to change the damping force. At the same time, another gear engaged with the sector gear operates the air valve to change the spring rate.

17.3.3 Control System

This system can change the spring rate and the damping force into three levels and vehicle height levels of low, normal, or high can be selected. One of the control logics is shown in Table 17.3. This is a control which changes the suspension characteristics in response to vehicle speed and road conditions. The spring rate/damping force and the vehicle height are controlled independently according to each control logic.

Electronic leveling control systems do not need much energy to control vehicle height. They control both spring rate and damping force. As a result of keeping the low spring rate, electronic leveling control systems can provide both good ride comfort and handling stability.

TABLE 17.3 Basic Control Logic of Air Suspension System

		Spring rate, damping force					
		Soft mode			Medium mode		
Function	Operating condition	Soft	Medium	Firm	Soft	Medium	Firm
Antiroll	Rapid steering	○ →		○		○ →	○
			○ →				
Antidive	Braking at V* > 60 km/h	○ →		○		○ →	○
			○ →				
Antisquat	Rapid starting at V < 20 km/h	○ →		○		○ →	○

* V: Vehicle speed.

17.4 ACTIVE SUSPENSION

Suspension control systems for passenger cars have evolved through several stages over the years. Work in the field began with the air suspension for controlling vehicle height and then progressed to the hydropneumatic suspension and suspensions with variable damping force and spring rate control. Now efforts are underway to develop an active suspension. It is defined as one that has the following features:

- Energy is constantly supplied to the suspension and the force generated by that energy is continuously controlled
- The suspension incorporates various types of sensors and a unit for processing their signals that generates forces that are a function of the signal outputs

17.4.1 System Configurations

Basic Configuration. Hydraulic active suspension can be divided into two large systems: the hydraulic system and the control system, as shown in Fig. 17.9. The hydraulic pressure of each of the actuators located on each wheel is controlled in accordance with the output values from the G sensors to suppress changes in vehicle body position (bounce, pitch, roll) and reduce vibration from the road surface.

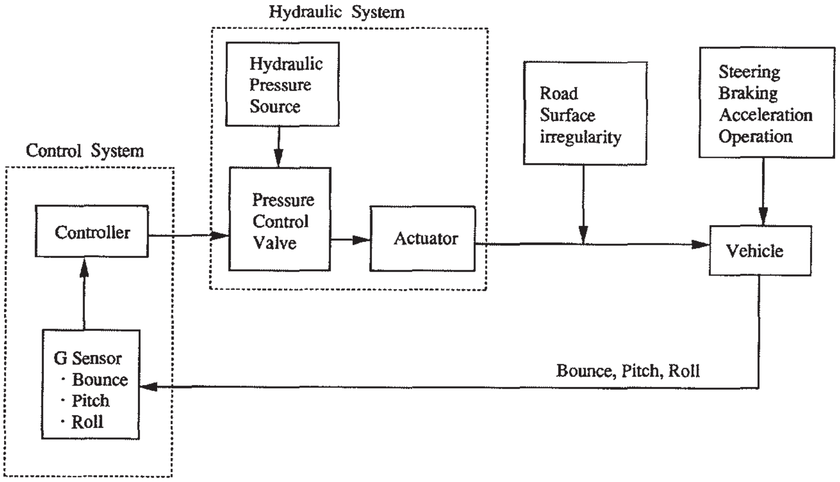


FIGURE 17.9 Hydraulic and control systems for an active suspension.

Hydraulic and Control System Configuration. A basic overview of the system is shown in Fig. 17.10. The functions of the main units of the hydraulic system are shown in Table 17.4.

As shown in Fig. 17.11, the control system contains the controller and all of the sensors including the vertical G sensors, lateral G sensors, fore and aft G sensors, and vehicle height sensors.

17.4.2 Components

Oil Pump. The oil pump has seven cylinders arranged around the circumference and can output a maximum oil flow of 12 liters per minute. The pump is connected in tandem with the power steering vane pump.

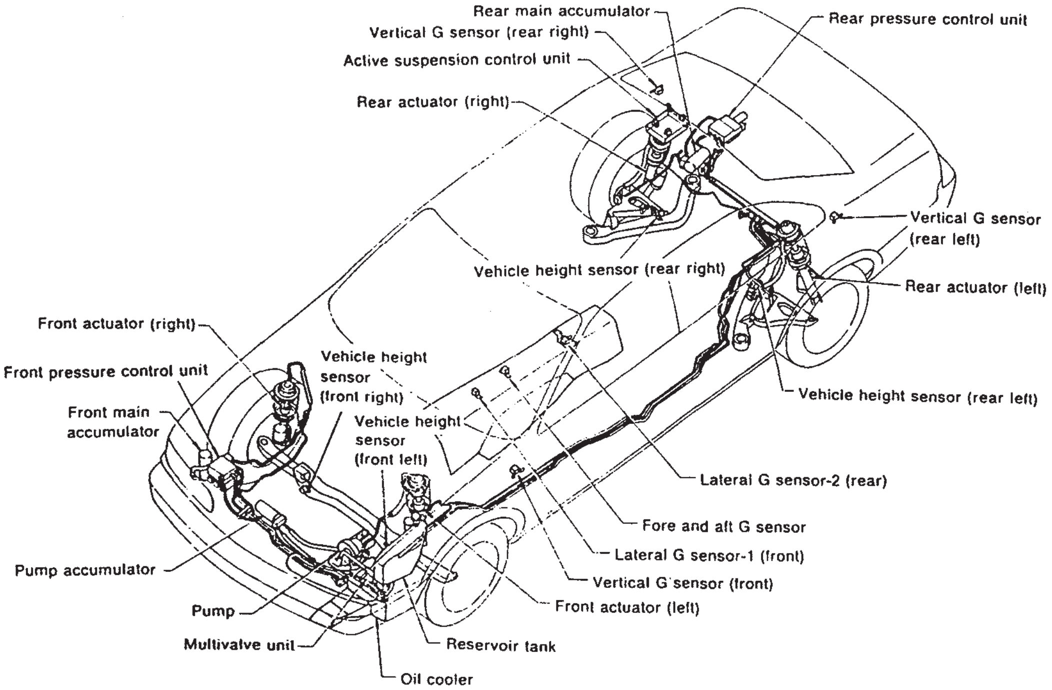


FIGURE 17.10 Principal components of an active suspension system.

Pump Accumulator. To dampen pulsating hydraulic pressure generated by the oil pump, pump accumulators are installed in the hydraulic supply unit, including one on the side of the oil pump. To dampen the high-frequency pulsations, a metal-bellows-type accumulator is used.

Multivalve Unit. As shown in Fig. 17.12, the multivalve unit contains valves for many different functions. The main purpose of the multivalve unit is the basic control of hydraulic pressure for the whole system. The function of the multivalve unit is shown in Table 17.5.

Main Accumulator. The main accumulator is positioned at both the front and rear axles. The main accumulator has two principal functions: it stores oil from the multivalve unit and provides extra flow to the actuators when necessary, and it preserves vehicle height when the engine is turned off.

TABLE 17.4 The Functions of Main Units of the Hydraulic Systems

	Main basic function
Oil pump	Supplies the necessary oil for system operation (power supply)
Pump accumulator	Removes the pulsating action from pressurized oil supplied by the oil pump
Multivalve unit	Controls the supply of pressurized oil, failsafe function, etc.
Main accumulator	Maintains oil pressure, compensates when large amount of flow is required and preserves vehicle body height
Pressure control unit	Controls the hydraulics for the actuators on each wheel according to signals received from the control unit
Actuator	Controls vehicle attitude and absorbs external forces from the road surface

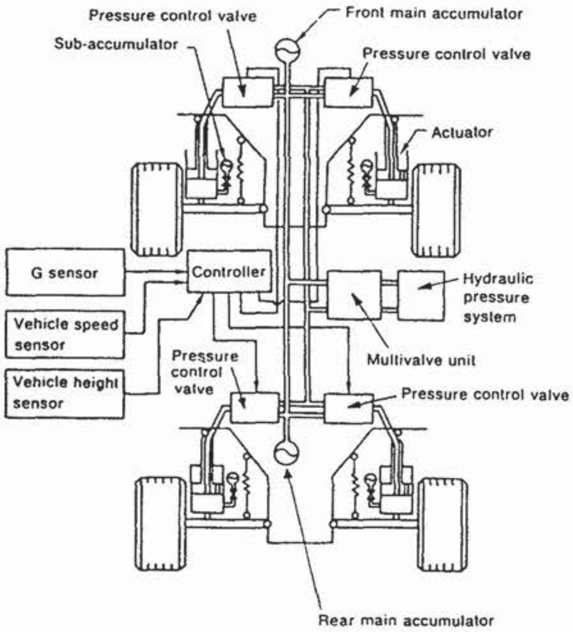


FIGURE 17.11 Controls for an active suspension system.

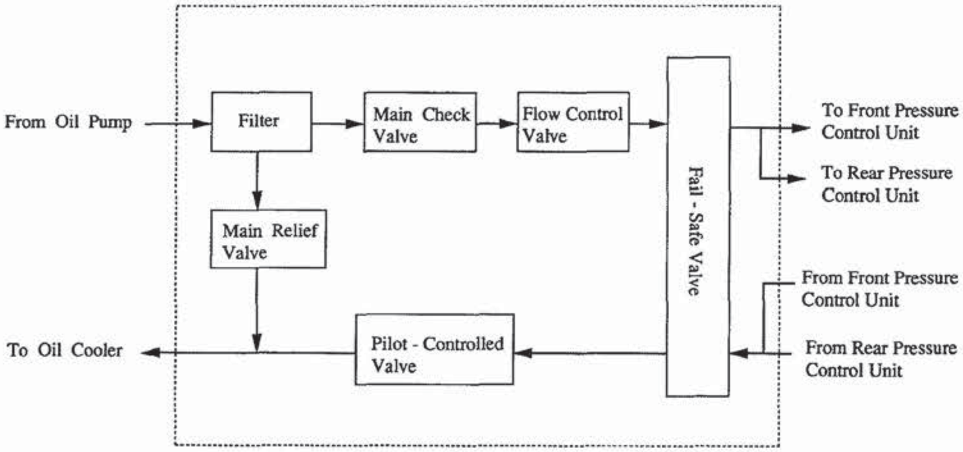


FIGURE 17.12 Multivalve unit configuration.

Pressure Control Unit. As shown in Fig. 17.13, the pressure control unit controls the hydraulic movement of the actuator of each wheel in accordance with instructions received from the control unit. Table 17.6 shows the valve's construction and operating principle.

The electrohydraulic pressure control system consists of a pressure actuator and a pressure control valve. The actuator is of the single acting type and is provided with a damping valve and an accumulator below the cylinder. The pressure control valve is built with three ports

TABLE 17.5 Multivalve Unit Functions

Function	Valve	Outline
Pressure supply management function	Main relief valve	When the oil pressure exceeds a constant value, the main relief valve will return some of the oil flow. This keeps the oil supply pressure at a constant pressure.
Vehicle height maintenance function	Main check valve Pilot-controlled check valve	The main check valve is a nonreturn valve that controls the flow from the line filter and directs it to the flow control valve. The pilot controlled check valve is a supply-pressure-reaction-type open/closed valve. When the hydraulic pressure exceeds a constant value, the valve opens and when the hydraulic pressure falls below that level, it closes. In addition, it maintains the hydraulic pressure at a constant level when the engine is turned off.
Vehicle height control function	Flow control valve	When the engine is turned off, the flow control valve closes the main passage and directs the flow through the bypass passage orifice, slowly increasing the hydraulic pressure, after which the main passage is opened. This prevents any sudden changes in vehicle height when starting the engine.
Failsafe function	Failsafe valve	When any irregularities occur in the electronic system, it changes the hydraulic passage, preventing any sudden changes in vehicle height.

and employs a pilot type proportional electromagnetic control valve. This pressure control valve has two main functions:

- It controls the pressure of the actuator according to the control input. This is accomplished by driving the solenoid so that it adjusts the pilot valve, causing the spool to move.
- Feedback control is applied to move the spool in response to fluctuations in actuator pressure caused by road surface inputs; the action of the spool works to keep the actuator pressure at a certain fixed level.

Actuator. As shown in Fig. 17.14, the actuator consists of the hydraulic power cylinder, subaccumulator, damping valve, etc. Auxiliary coil springs are also employed to reduce the pressure necessary for the overall system and to reduce the amount of horsepower expended. The subaccumulator and damping valve at the bottom of the hydraulic power cylinder absorb and damp the high-frequency vibration from the road surface.

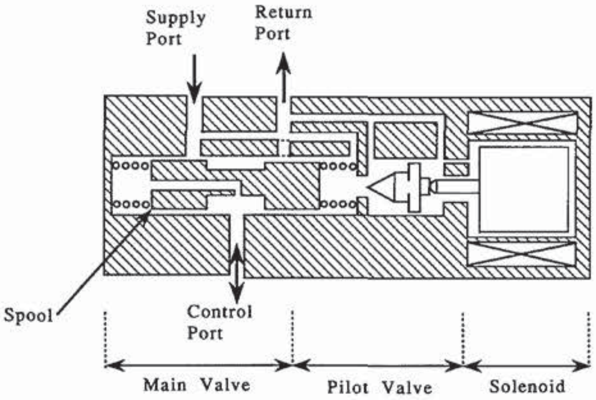


FIGURE 17.13 Pressure control valve construction.

TABLE 17.6 Pressure Control Valve Functions

Active control function	The pressure in the control port (actuator) is controlled in response to the electric current applied to the solenoid, thus controlling the vehicle attitude.
Passive damping function	When various pressure levels are caused in the interior of the actuator by road surface forces, this pressure passes through the control port, causing feedback on the spool and the generation of appropriate damping forces.

Controller. As shown in Fig. 17.15, the controller is constructed using two 16-bit microcomputers, MCU1 and MCU2.

MCU1 processes signals from the G sensors and then sends attitude control signals to the pressure control valve solenoid drive circuit. MCU2 processes signals from the vehicle height sensors and then sends attitude control signals to the solenoid drive circuit.

MCU1 and MCU2 normally perform mutual transmission, but should an irregularity occur, the signal will be sent to the failsafe circuit, causing the failsafe valve to operate and thus guarantee a high degree of safety.

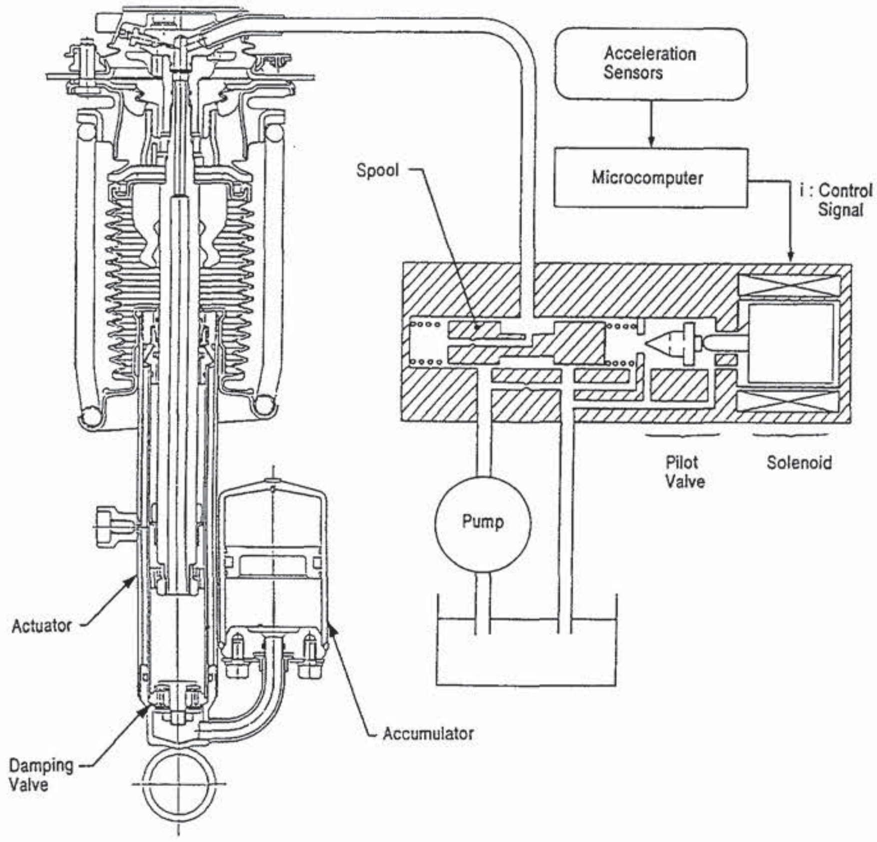


FIGURE 17.14 Schematic diagram of electrohydraulic pressure control system.

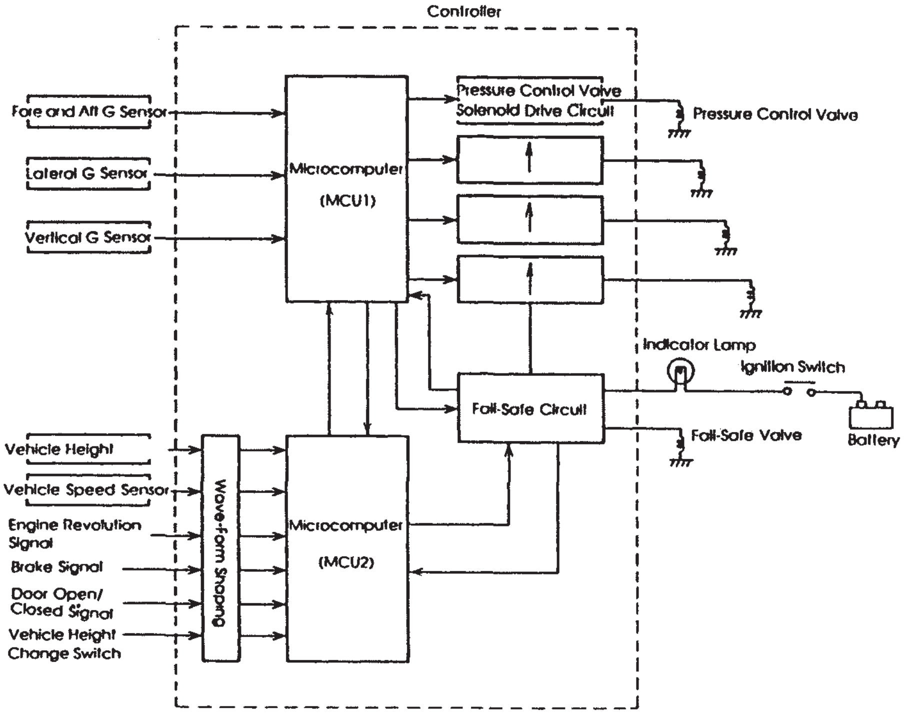


FIGURE 17.15 Interior construction of the electrohydraulic pressure controller.

G Sensors. The G sensors are ball position detection type sensors. They detect changes in the magnetic field caused by the position change of a steel ball as the result of acceleration.

17.4.3 Control System

Roll Control. The inertia force, which causes the car to roll, is detected by the lateral G sensor. Roll control is initiated by increasing the control pressure on the wheels on the outside of the turn and by decreasing the control pressure for the wheels on the inside of the turn. Figure 17.16 shows this operating principle.

The relation between the lateral G and the force generated by the actuator is:

$$F = m\alpha \tag{17.1}$$

$$Fh = \Delta F d$$

$$\Delta F = \frac{Fh}{d} = \frac{m\alpha h}{d}$$

Pitch Control. During braking, inertia is generated at the vehicle's center of gravity and causes pitching. The longitudinal G sensor detects this inertia and cancels it to suppress nose dive by increasing the control pressure to the front and decreasing control pressure to the rear, as shown in Fig. 17.17. The relationship between longitudinal G and the actuator generated force is:

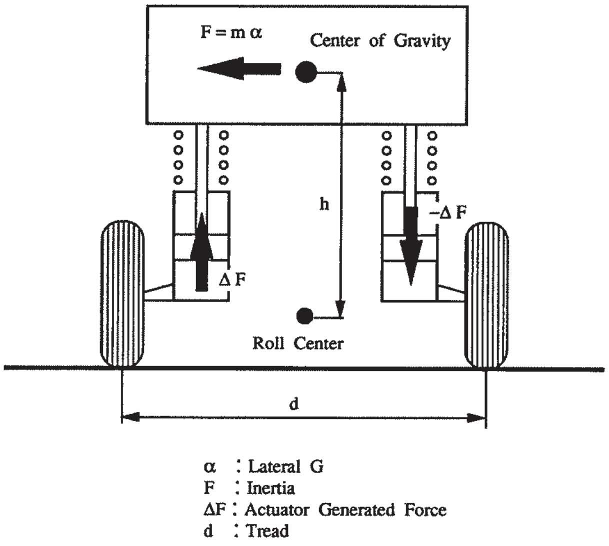


FIGURE 17.16 Operating principle for roll control.

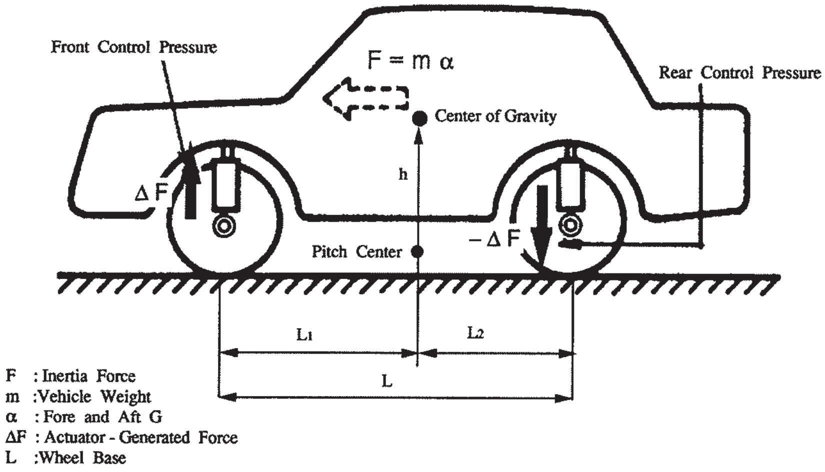


FIGURE 17.17 Pitch control parameters.

$$Fh = \Delta F\,(L_1 + L_2)$$

$$\Delta F = \frac{Fh}{(L_1 + L_2)} = \frac{m\alpha h}{L} \tag{17.2}$$

Bounce Control. The vertical G sensor attached to the vehicle body detects the value for vehicle body acceleration. By integration of the acceleration, the absolute velocity of the body is calculated. A force proportional to the absolute velocity is generated by the pressure control valve.

This control method, called skyhook damper control, is adapted. It dampens the motion of the car body regardless of any input from the road surface.

In the case of a passive damper, the vertical motion of the body relative to the road surface inputs can be given as:

$$\frac{X_2}{X_1} = \frac{2j\omega_2\xi_2\omega + \omega_2^2}{-\omega^2 + 2j\omega_2\xi_2\omega + \omega_2^2} \tag{17.3}$$

The vibration transmission rate at the resonant point is:

$$\left|\frac{X_2}{X_1}\right|_{\omega=\omega_2} = \sqrt{1 + \frac{1}{4\xi_2^2}} \tag{17.4}$$

and always has a value greater than one.

By contrast, the vibration characteristics of the skyhook damper are given as:

$$\frac{X_2}{X_1} = \frac{2j\omega_2\xi_2\omega + \omega_2^2}{-\omega^2 + 2j\omega_2(\xi_2 + \xi_s)\omega + \omega_2^2} \tag{17.5}$$

The vibration transmission ratio at the resonant point is:

$$\left|\frac{X_2}{X_1}\right|_{\omega=\omega_2} = \frac{\sqrt{4\xi_2^2 + 1}}{2(\xi_2 + \xi_s)} \tag{17.6}$$

hence,

$$\xi_s \geq \sqrt{\xi_2^2 + \frac{1}{4}} - \xi_2 \tag{17.7}$$

and it is possible to reduce the ratio to less than one.

The effects of the hydraulic active suspension are organized in Table 17.7 according to those related to the vehicle.

17.4.4 Effectiveness

Figure 17.18 shows the lateral G and angle of roll during cornering. Figure 17.19 shows the fore and aft G force, angle of nose dive, and squat angle during starts and stops. In either case, the car with the hydraulic active suspension outperformed the other cars with conventional suspensions.

TABLE 17.7 Effect of Active Suspension Control

Control	Vehicle-related effects
Roll control	During transient control of wheel loading, as when changing lanes, the steering characteristics of the vehicle can be optimally controlled. The tires are used to their utmost performance ability because there is minimal roll, minimal change in the camber of the tires to the ground, and because the tires are continually kept in square contact with the road.
Pitch control	Nose dive and tail lift are minimized during braking. Squats are minimized during starts and rapid acceleration.
Bounce control	Vertical vibration of the vehicle is reduced and continuity is improved. Vertical load fluctuation is minimal, and the contact of the tires with the road is greatly improved.

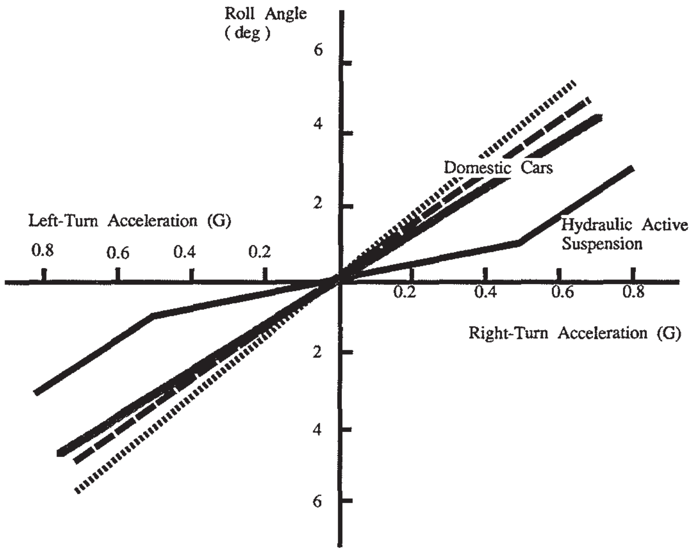


FIGURE 17.18 Comparison of roll angles during cornering.

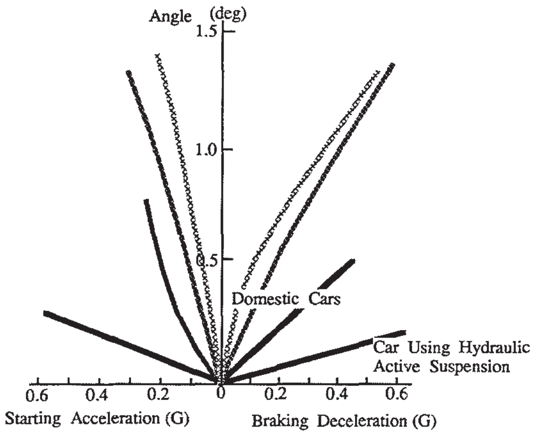


FIGURE 17.19 Comparison of nose dive and squat angles.

Figure 17.20 shows the effects of bounce control for the skyhook damper. Compared with cars using conventional suspension systems, the car with the hydraulic active suspension system exhibited superior performance and low vibration levels.

Figure 17.21 shows the ride characteristics and roll rate for various cars with some suspension systems. As the results clearly demonstrated, the hydraulic active suspension system, through advanced roll control and bounce control, provides a ride and a level of control far superior to that of other suspension systems.

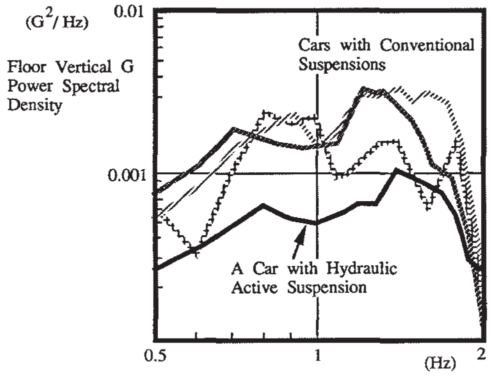


FIGURE 17.20 Comparison of vertical vibrations.

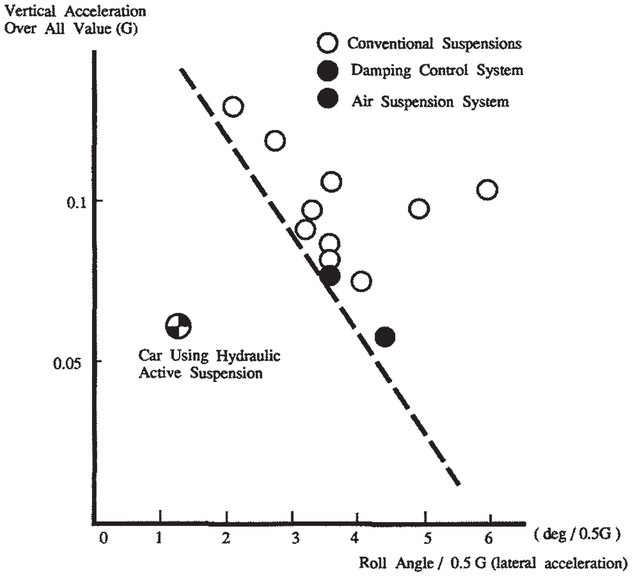


FIGURE 17.21 Ride characteristics and roll rate for various cars with different suspension systems.

17.5 CONCLUSION

The active suspension system provides outstanding levels of performance which are unobtainable with other suspension control systems and conventional passive suspensions. Evaluations made with actual vehicles confirmed the effectiveness of the active suspension system in improving ride comfort and handling properties.

GLOSSARY

Hydropneumatic suspension A suspension system using oil or air to support the car body.

Skyhook damper control The control law applied to control the vehicle as if it were fixed within absolute space suspended from the sky.

Supersonic sensor A sensor used to measure the distance between a car body and the road surface using supersonic waves.

NOMENCLATURE

C_2: passive damping coefficient
C_s: active damping coefficient
K_1: tire stiffness
K_2: spring stiffness
M_1: unsprung mass
M_2: sprung mass
$\omega_1 = (K_1/M_1)^{1/2}$ natural frequency of unsprung mass
$\omega_2 = (K_2/M_2)^{1/2}$ natural frequency of sprung mass
$\xi_2 = (C_2/2)\ (M_2K_2)^{1/2}$ active damping ratio
$\xi_S = (C_S/2)\ (M_2/K_2)^{1/2}$ passive damping ratio

BIBLIOGRAPHY

Decker, H., and W. Scramm, "An optimized approach to suspension control," SAE Paper 900661, 1990.

Decker, H., W. Scramm, and R. Kallenbach, "A modular concept for suspension control," *XXIII Fisita Congress,* Torino, Italy, 1990, 905124.

Decker, H., W. Scramm, and R. Kallenbach, "A practical approach toward advanced suspension systems," *Proc. IMechE International Conference on Advanced Suspensions,* London, 1988, S. 93–99.

Doi, S., E. Yasuda, and Y. Hayashi, "An experimental study of optimal vibration adjustment using adaptive control methods," *Proceedings of the Institution of Mechanical Engineers,* London, Oct. 24–25, 1988.

Guy, Y., M. B. Lizell, and M. W. Kerastas, "A solenoid-actuated pilot valve in a semi-active damping system," SAE Paper 881139, Aug. 1988.

Haget, K. H., et al., "Continuous adjustable shock absorbers for rapid-acting ride control systems (RCS)," *XXIII Fisita Congress,* Torino, Italy, 1990, 905125.

Hennecke, D., P. Baier, B. Jordan, and E. Walek, "EDCIII—the new variable damper system for BMW's top models—a further development of our adaptive, frequency-dependent damper control," SAE Paper 900662.

Hennecke, D., P. Baier, B. Jordan, and E. Walek, "Further market-oriented development of adaptive damper force control," *XXIII Fisita Congress,* Torino, Italy, 1990, 905143.

Hennecke, D., B. Jordan, and U. Ochner, "Electronic damper control—a fully automatic adaptive system for damping force adjustment on the BMW 635 CSi," *ATZ 89,* 9, 1987, pp. 471–479.

Hrovat, D., D. L. Margolis, and M. Hubbard, "An approach toward the optimal semi-active suspension," *Transactions of the ASME,* 288, vol. 110, Sept. 1988.

Karnop, D., and D. L. Margolis, "Adaptive suspension concepts for road vehicles," *Vehicle System Dynamics,* 13, 1984, pp. 145–160.

Karnop, D., M. J. Crosby, and R. A. Harwood, "Vibration control using semi-active force generators," *ASME Journal of Engineering for Industry Transactions of the ASME,* No. 96, 1974, S. 613–626.

Margolis, D. L., "Semi-active control of wheel hop in ground vehicles," *Vehicle System Dynamics,* 12, 1983, pp. 317–330.

Sharp, R. S., and S. A. Hassan, "Performance and design considerations for dissipative semi-active suspension systems for automobiles," *Proceedings Institution of Mechanical Engineers,* vol. 201, no. D2, pp. 149–153.

Sugasawa, F., et al., "Electronically controlled shock absorber system used as a road sensor which utilizes super sonic waves," SAE Paper 851652, 1985.

Tsutsumi, Y., H. Sato, H. Kawaguchi, and M. Hirose, "Development of Piezo TEMS (Toyota Electronic Modulated Suspension)," SAE Paper 901745, 1990.

Yokota, Y., K. Asami, and T. Hajima, "Toyota electronic modulated suspension system for the 1983 Soarer," SAE Paper 840341, 1984.

Hydropneumatic Suspension

Akatsu, Y., N. Fukushima, K. Takahashi, M. Satoh, and Y. Kawarazaki, "An active suspension employing an electrohydraulic pressure control system," *XXIII Fisita Congress,* Torino, Italy, 1990, 905123.

Carbonaro, O., "Hydractive suspension electronic control system," *XXIII Fisita Congress,* Torino, Italy, 1990, 905101.

Dominy, J., and D. N. Bulman, "An active suspension for a Formula One Grand Prix racing car," *Transactions ASME, Journal Dynamic Systems Measurement Control,* vol. 107, 1, 1985, pp. 73–78.

Hedrick, J. K., "Railway vehicle active suspensions," *Vehicle System Dynamics* **10**, 1981, pp. 267–283.

Henning, H., "Citroen SM V6," *ATZ 92,* 1990, pp. 23–27.

Hrovat, D., "Optimal active suspension structures for quarter-car vehicle models," *Automatica,* vol. 26, 5, 1990, pp. 845–860.

Karnop, D., "Active damping in road vehicle suspension systems," *Vehicle System Dynamics* **12**, 1983, pp. 291–316.

Sharp, R. S., and S. A. Hassan, "An evaluation of passive automotive suspension systems with variable stiffness and damping parameters," *Vehicle System Dynamics,* **15**(6):1986, pp. 335–350.

Tanahashi, H., K. Shindo, T. Nogami, and T. Oonuma, "Toyota electronic modulated air suspension for the 1986 Soarer," SAE Paper 870541, 1987.

Thompson, A. G., "An active suspension with optimal linear state feedback," *Vehicle System Dynamics* **5**, 1976, pp. 187–203.

Tobata, H., N. Fukushima, K. Fukuyama, and T. Kimura, "Advanced control method of active suspension," *Avec92,* Yokohama, Japan, 1992, 923023.

Wright, P., "The application of active suspension to high performance road vehicle," *I MechE,* C239/84, 1984.

ABOUT THE AUTHOR

Akatsu Yohsuke has been Vehicle Research Manager for Nissan Motor Co., Ltd., since 1980. He was awarded the B.E. degree from the Tokyo Institute of Technology in 1978 and the M.E. degree from there in 1980.

CHAPTER 18
STEERING CONTROL

Makoto Sato
Honda R&D Co., Ltd.

The application of electronic control technology to vehicle steering systems is still in the development stage. The structure and functioning of such systems are not yet clearly defined. Accordingly, the material that follows is intended primarily to introduce systems which have already been published, and it is devoted entirely to the electronic aspects of those systems and does not include descriptions of the basic operation of hydraulic power steering systems.

18.1 VARIABLE-ASSIST STEERING

18.1.1 Fundamentals of Electronically Controlled Power Steering

Electronically controlled power steering improves steering feel and power-saving effectiveness, and increases steering performance. It does so with control mechanisms that reduce the steering effort. An electronic control system, for example, may be added to the hydraulic booster or the whole system may be composed of electronic and electric components.

The intent of electronic controls, initially, was to reduce the steering effort when driving at low speeds and to supply feedback for the appropriate steering reaction force when driving at high speeds. In order to achieve those goals, devices such as vehicle speed sensors were used to detect vehicle speed in order to make smooth and continuous changes in the steering assist rate under conditions ranging from steering maneuvers at zero speed to those at high speeds. However, as vehicles became equipped with electrohydraulic systems and fully electronic and electric systems, the emphasis for these systems started to include reduction in power requirements and higher performance.

The main functions required for electronically controlled power steering are listed in Table 18.1.

18.1.2 Types of Electronically Controlled Power Steering

Electronically controlled power steering systems presently available commercially can be classified according to their basic structure and basic principles into three types: hydraulic, hybrid, and full electric systems, as shown in Table 18.2. Detailed explanations of these systems are given as follows.

TABLE 18.1 Functions Required for Electronically Controlled Power Steering

Reduction of driver's burden when turning the steering wheel and improvement in the steering feel	• Reduction in steering effort • Smoothness of steering operation • Feedback of appropriate steering reaction forces • Reduction of kickback[1] • Improvement in convergence[2] • Creation of other new functions
Power saving	
Failsafe	• Maintaining of manual steering function in the event of any malfunctions

TABLE 18.2 Classification of Electronically Controlled Power Steering System

			Sensors					Major effects	
Basic structure	Control method	Control objects	Vehicle speed	Steering torque	Angular velocity	Current	Actuator	Steering force responsive to vehicle speed	Power saving
Electronically controlled hydraulic system	Flow	Flow supply to power cylinder	○			○	Solenoid	○	○
	Cylinder bypass	Effective actuation pressure given to cylinder	○			○	Solenoid	○	
	Valve characteristics	Pressure generated at control value	○			○	Solenoid	○	
	Hydraulic reaction force control	Pressure acting on the hydraulic reaction force mechanism	○			○	Solenoid	○	
Hybrid system	Flow	Flow supply to power cylinder	○		○	○	Motor	○	○
Full electric system	Current	Motor torque	○	○		○	Motor	○	○
	Voltage	Motor power	○	○	○	○	Motor	○	○

18.1.3 Explanations of Each System

Electronically Controlled Hydraulic System. This system consists of a linear solenoid valve, a vehicle speed sensor, and other electronic devices located in part of the hydraulic circuit of the hydraulic system. The opening of the solenoid valve is controlled based on signals from the vehicle speed sensor. The flow and pressure of the hydraulic fluid is controlled by means of the opening of the solenoid valve.

The assist rate is smoothly and continuously varied in response to the vehicle speed, so that when the vehicle is stationary, the opening of the solenoid valve is small to ensure that the steering effort is appropriately light. When the vehicle is moving at high speed, the opening of the solenoid valve is large to ensure that the steering effort is appropriately heavy.

Flow Control Method. In this method, a solenoid valve is located at the pump discharge port as shown in Fig. 18.1. The electronic control device regulates the solenoid valve opening at high vehicle speeds to reduce the pump discharge volume, thus increasing the required steering effort. By reducing the resistance of the circuit between the pump and the power cylinder, power requirements are reduced. Figure 18.2 shows the position of the solenoid

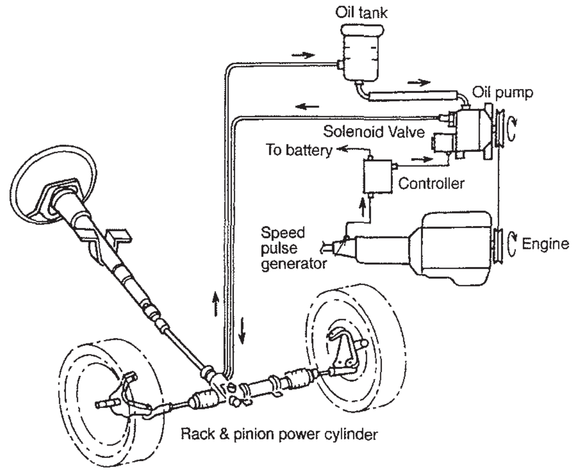
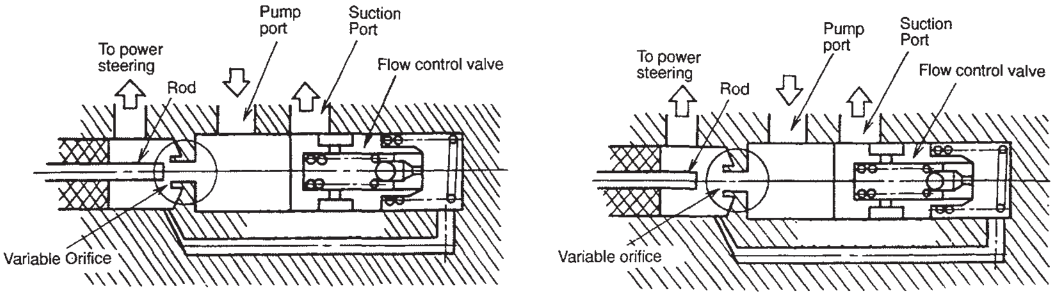


FIGURE 18.1 Vehicle speed-responsive pump discharge flow volume control type.[1]



Solenoid valve operation at high speeds

Solenoid valve operation at low speeds

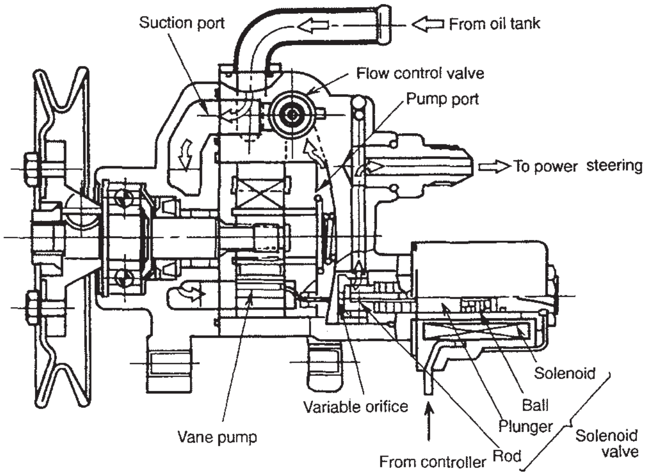


FIGURE 18.2 Structure and operation of pump with solenoid valve.[1]

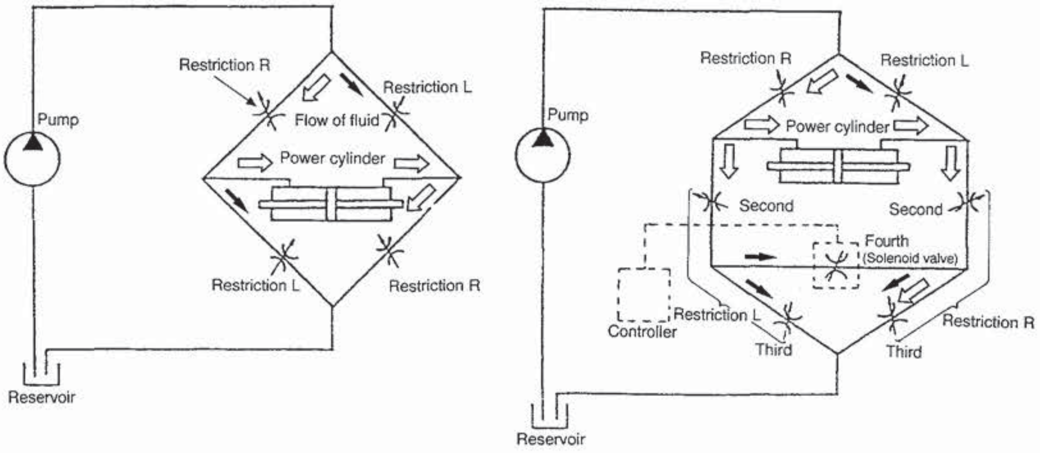
valve at the pump, with separate diagrams showing the operation at high and low vehicle speeds.

The flow of hydraulic fluid to the power cylinder is reduced when driving at high speeds, so that, for this method, the magnitude of the steering response rate and the steering reaction force are balanced at a point of equilibrium.

Cylinder Bypass Control Method. In this method, a solenoid valve and a bypass line are located between both chambers of the power cylinder.[2] The opening of the valve is extended by the electronic control equipment in accordance with increases in vehicle speed, thus reducing the hydraulic pressure in the power cylinder and increasing the steering effort. Like the flow control method, this system may also seek the equilibrium point for the steering response rate and the steering reaction force.

Valve Characteristics Control Method. In this method, the pressure control restrictions of the rotary valve (control valve) mechanism, which control the volume and pressure of the hydraulic fluid supplied to the power cylinder, are divided into second and third parts. A fourth part, controlled by means of the vehicle speed signal, is provided in the hydraulic line between the second and third parts as shown in Fig. 18.3. The structure of this system is shown in Fig. 18.4. The steering effort is controlled by carrying out variable control of the fourth part to vary the assist ratio. Because the structure is simple and the flow from the pump to the cylinder is supplied efficiently without waste, this system exhibits a good response rate. Figure 18.5 shows the hydraulic pressure characteristics in the valve characteristics control method with the driving current of the solenoid valve in accordance to the vehicle speed control signal as a parameter. When the current is 0.3 A, the valve is fully open, and this represents the high-speed driving condition.

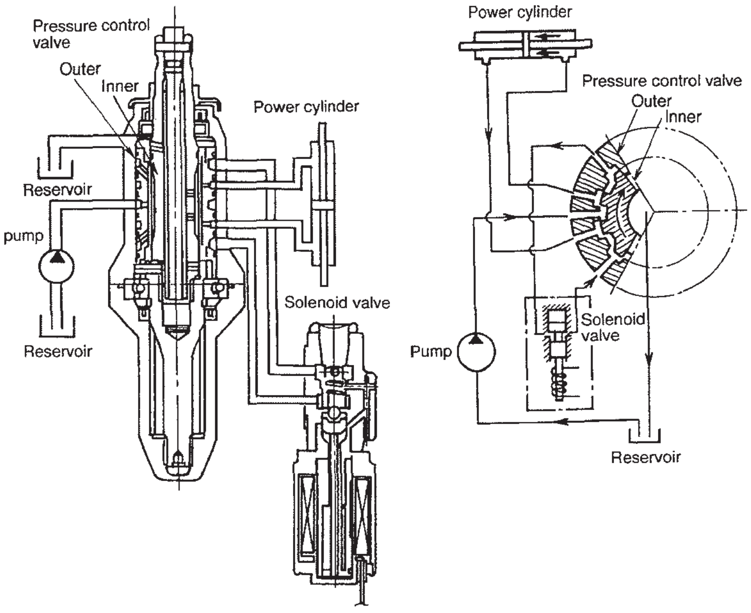
Hydraulic Reaction Force Control Method. In this method, the steering effort is controlled by means of a hydraulic reaction force mechanism, which is located at the rotary valve (control valve). A hydraulic reaction force control valve increases the hydraulic pressure (reaction pressure) introduced into the hydraulic reaction force chamber in accordance with increases in the vehicle speed. The rigidity of the reaction force mechanism (equivalent spring constant) is variably controlled so as to directly control the steering effort.



Bridge circuit for conventional power steering

Bridge circuit for electronically-controlled power steering

FIGURE 18.3 Closing bridge circuit for control valve.[2]



Cross-section of rotary valve structure

Hydraulic pressure circuit of rotary valve

FIGURE 18.4 Valve characteristics control method.[1]

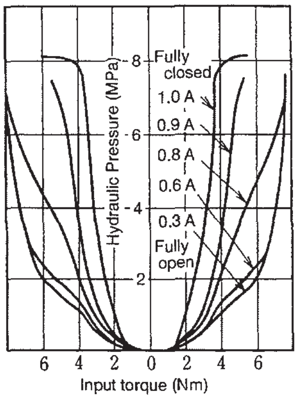


FIGURE 18.5 Valve characteristics in response to vehicle speed.[1]

This method requires the inclusion of a reaction force mechanism, which makes the structure of the control valve more complex, which in turn increases the cost. However, because the rigidity of the reaction force mechanism increases in accordance with increases in the vehicle speed, there is no vagueness in the steering feel in the area around the straightforward steering position. Because this method assigns the steering reaction force irrespective of the volume of hydraulic fluid supplied to the power cylinder, the magnitude of the steering reaction force can be set freely without the need to sacrifice any of the steering response rate.

18.1.4 Hybrid Systems

Hybrid systems utilize a flow control method in which the hydraulic power steering pump is driven by an electric motor. The steering effort is controlled by controlling the rotating speed of the pump (discharge flow).

The drive efficiency of the generator and motor are low compared to that of the hydraulic pump, which is driven by the vehicle engine. But because any residual flow is not discharged, the power loss is lower than that of the engine pump when driving at high speeds.

Because the pump is not driven by the vehicle engine, there is also a large degree of freedom in the selection of the mounting location for the pump.

Driving Mode Responsive Method. In this method, the control system consists of a vehicle speed sensor, steering angular velocity sensor, an electronic control unit, and a motor driven hydraulic pump, as shown in Fig. 18.6.

As is shown in Fig. 18.7, the driving conditions (such as driving in urban areas, country areas, winding regions, or highways) are automatically judged, and the pump flow rate is controlled in accordance with this condition in order to provide the appropriate steering effort for the driving conditions. Fine control adjustments are achieved by means of this method as compared with vehicle speed-responsive types mentioned previously.

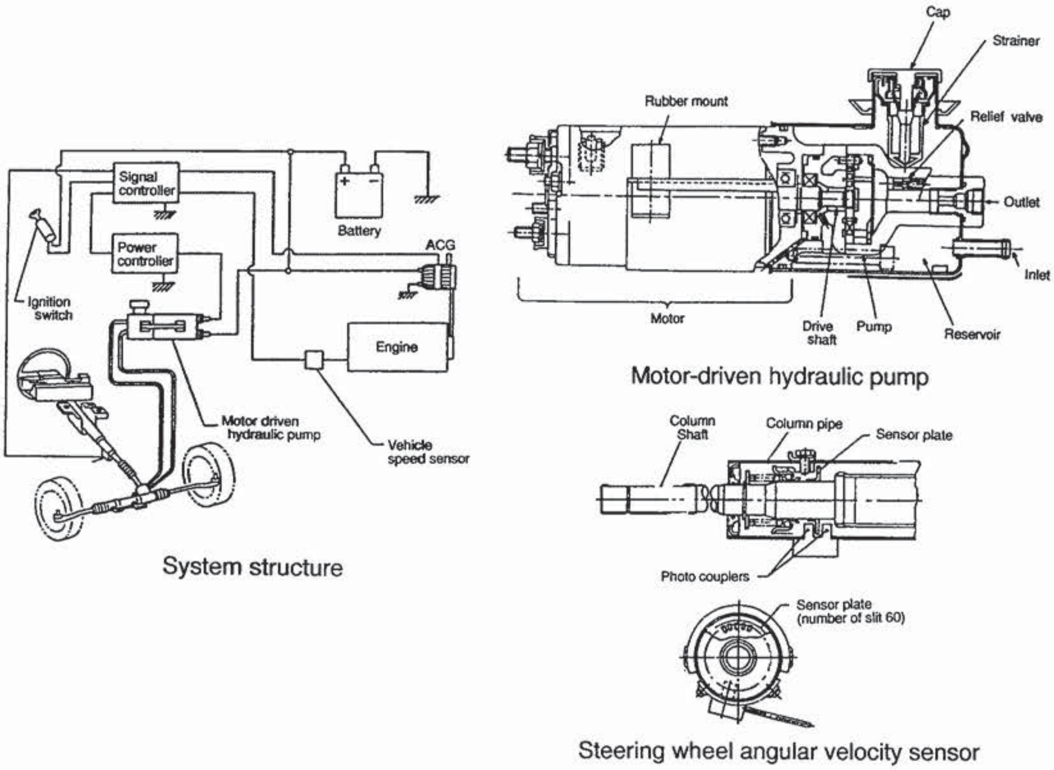
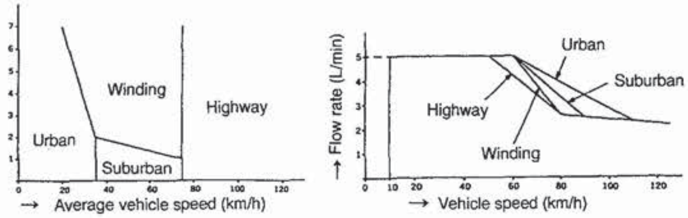


FIGURE 18.6 Driving mode responsive-type hybrid power steering.[3]

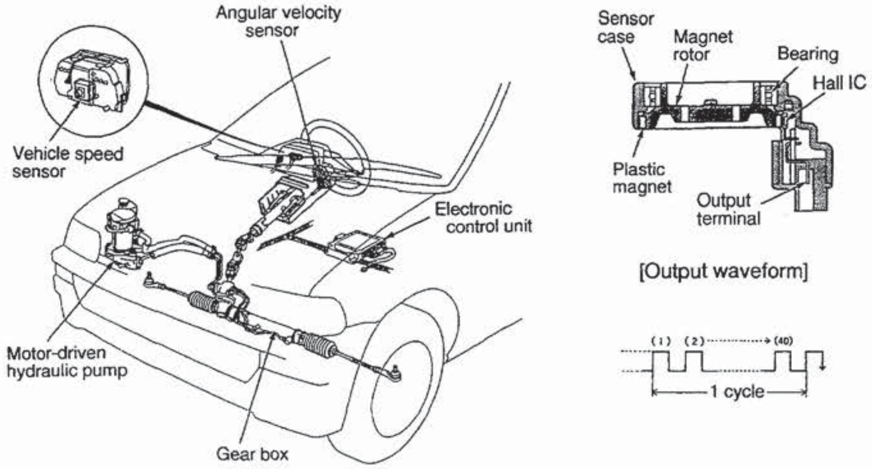


Driving mode map

Flow rate control map

FIGURE 18.7 Driving mode and pump flow rate.[3]

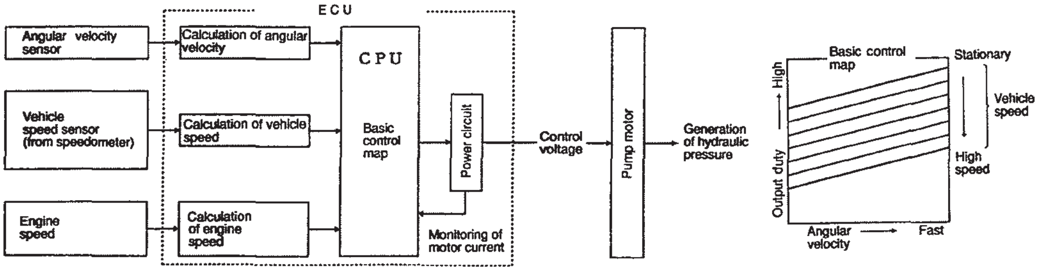
Steering Wheel Speed Responsive Method. For this method, the system consists of components such as a vehicle speed sensor, steering wheel angular velocity sensor, an electronic control unit, and a motor-driven hydraulic pump, as shown in Fig. 18.8. The pump flow volume is controlled in accordance with the angular velocity of the steering wheel and the vehicle speed as shown in Fig. 18.9. As mentioned previously, the discharge flow volume of the pump is reduced and the steering response drops when the vehicle is driven at high speeds. Therefore, in this system, the speed of the motor is increased in accordance with the detected angular velocity of the steering wheel in order to increase the discharge flow volume to solve the problem. Accordingly, losses in power resulting from the circulation of residual flow within the system are kept to the minimum possible level, and the magnitude of the reaction force can be controlled freely without sacrificing any of the steering response rate.



System structure

Structure and output waveform of steering wheel angular velocity sensor

FIGURE 18.8 Steering speed-responsive-type hybrid power steering.[4]



Control block

Control map

FIGURE 18.9 Basic control.[4]

18.1.5 Electric Power Steering

Electric power steering (EPS) is a fully electric system, which reduces the amount of steering effort by directly applying the output from an electric motor to the steering system. This system consists of vehicle speed sensors, a steering sensor (torque, angular velocity), an electronic control unit, a drive unit, and a motor, as shown in Fig. 18.10. Signal outputs from each sensor are input to the electronic control unit, where the necessary steering assistance is calculated and applied by the drive unit to control the operation of the motor.

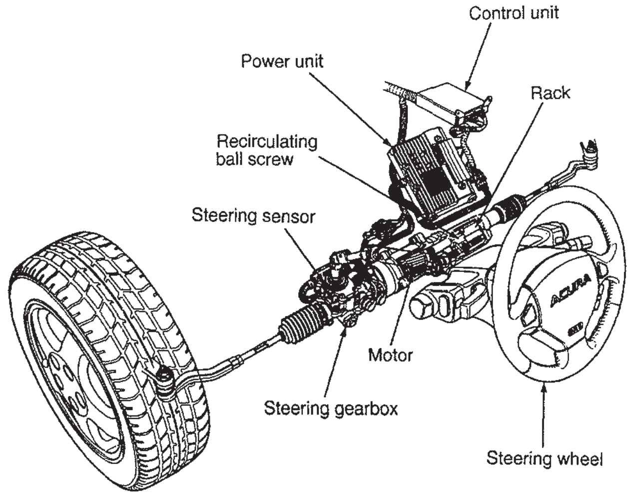


FIGURE 18.10 Structure of EPS System[5] (rack assist-type ball screw drive).

Because the motor output is controlled directly in this system, the setting range for the steering effort is large, and also because it is possible to supply only the amount of power that is necessary when the steering wheel is turned, a large reduction in power requirements can be effectively achieved with no power losses. This means, in contrast to hydraulic systems, that it is not necessary for the pump to keep operating continuously when the steering wheel is not being turned.

TABLE 18.3 Classification of Motor Drive Mechanism in EPS

Method	Motor drive method	Power transmission mechanism	Figure
Pinion assist	Column shaft drive	Motor→worm gear→column shaft→pinion shaft	18.11
	Pinion shaft drive	Motor→gear train→pinion shaft	18.12
Rack assist	Another shaft pinion drive	Motor→planetary gear train→another shaft pinion→rack shaft	
	Ball screw drive	Motor→ball screw→rack shaft	18.10

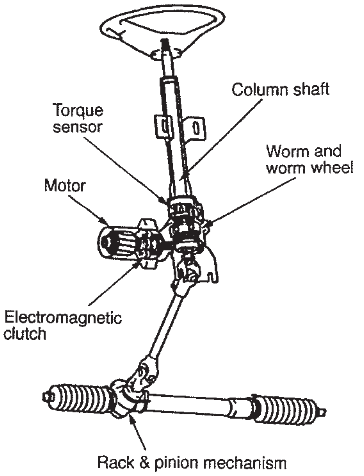


FIGURE 18.11 Column shaft drive method.[1]

In rack-and-pinion steering mechanisms, the EPS system applies the motor power to the pinion gear shaft or to the rack shaft. Several reduction gears are incorporated to amplify the torque of the motor. This system can be classified according to the drive method as given in Table 18.3 and Figs. 18.10 to 18.12. The maximum amount of assist, the smoothness of the steering feel, and the degree of noise occurring during steering are, by and large, determined by the power transmission systems in this table. In general, it is possible to obtain a greater amount of assist from the rack assist method than from the pinion assist method, and the rack assist method is optimal for vehicles in which the front axle load is high.

Details of the respective sensors, controls, and the results achieved thereby are given as follows under common headings.

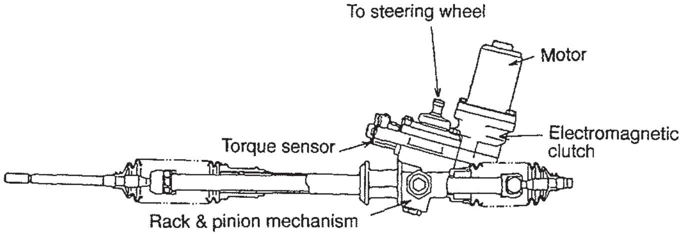


FIGURE 18.12 Pinion shaft drive method.[1]

Sensors. The EPS system utilizes a variety of sensors to control the motor. These sensors include a torque sensor, which detects the steering effort of the steering wheel; a steering wheel angular velocity sensor, which detects the angular velocity of the steering wheel; a battery sensor, which detects the battery voltage; a current sensor, which detects the motor current and the battery current; and a vehicle speed sensor. Of these sensors, the torque sensor and the steering wheel angular velocity sensor, which form the core of the EPS system, are described as follows. (Also see Chap. 3.)

Torque Sensor. The pinion shaft in the rack-and-pinion steering mechanism is divided into two sections of input shaft and pinion gear. The torque sensor comprises a torsion bar

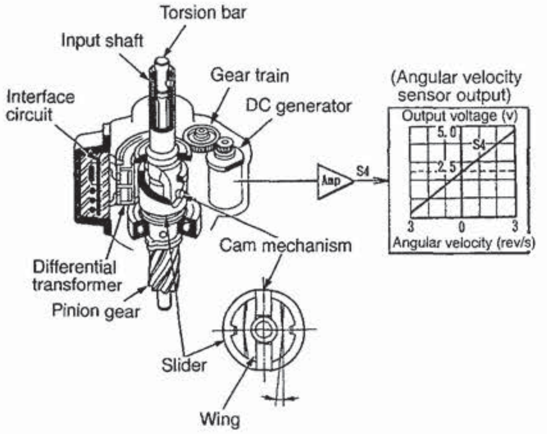


FIGURE 18.13 Torque sensor and angular velocity sensor.[5]

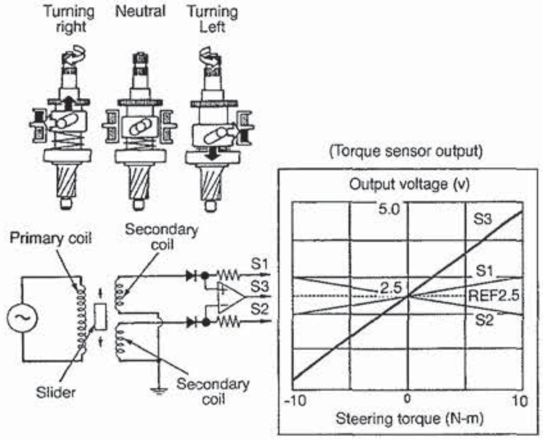


FIGURE 18.14 Diagram of torque sensor operation.[5]

that connects the two sections, a slider with a movable iron core, a cam mechanism that converts the twist torque of both sections of the shaft into an axial direction displacement, and a differential transformer that converts the axial direction displacement of the slider into an electric signal.

The structure is shown in Fig. 18.13. Figure 18.14 shows examples of the detection of the magnitude and direction of the slider displacement. Signal S3 is the output signal of torque and S1 and S2 are outputs for diagnosis use. The differential transformer-type torque sensor has a dual electric structure to provide differential outputs, so that a high sensing accuracy and good temperature characteristics can be obtained, and detection of failure can also be made accurately.

In addition to this type of torque sensor, there are other torque sensors that use a potentiometer instead of a differential transformer, and also types in which the relative displacement in the turning direction is measured in a noncontact manner using a coil, without there being a mechanically movable part such as the slider.

Angular Velocity Sensor. The angular velocity sensor consists of a gear train, which is located around the input shaft, and a dc generator, which is driven at an increase of speed by this gear train. The structure is shown in Fig. 18.13. The turning direction and angular velocity of the steering wheel are detected by the turning direction and angular velocity of this dc generator. Signal S4 indicates the output from the steering wheel angular velocity sensor.

Electronic Control Unit (ECU). The ECU consists of an interface circuit that coordinates the signals from the various sensors, an A/D converter and a PWM unit that are all built into an 8-bit one-chip microprocessor, a watchdog timer (WDT) circuit that monitors the operation of this microprocessor, and a PWM drive circuit that drives the power unit mentioned previously. The ECU conducts a search for data according to a table lookup method based on the signals input from each sensor and carries out a prescribed calculation using this data to obtain the assist force.

In addition, trouble diagnosis for the sensors and the microprocessor is also carried out. When a problem is detected, power to the motor is interrupted, an indicator lamp illuminates, and the problem condition is memorized. Then this problem mode flashes on a display as necessary.

Power Unit. The power unit comprises a power MOSFET (metal oxide semiconductor field effect transistor) bridge circuit which drives the motor in a forward or reverse direction, a drive circuit which controls the respective power MOSFET of this power MOSFET bridge circuit, a current sensor, and a relay which turns the motor current ON and OFF.

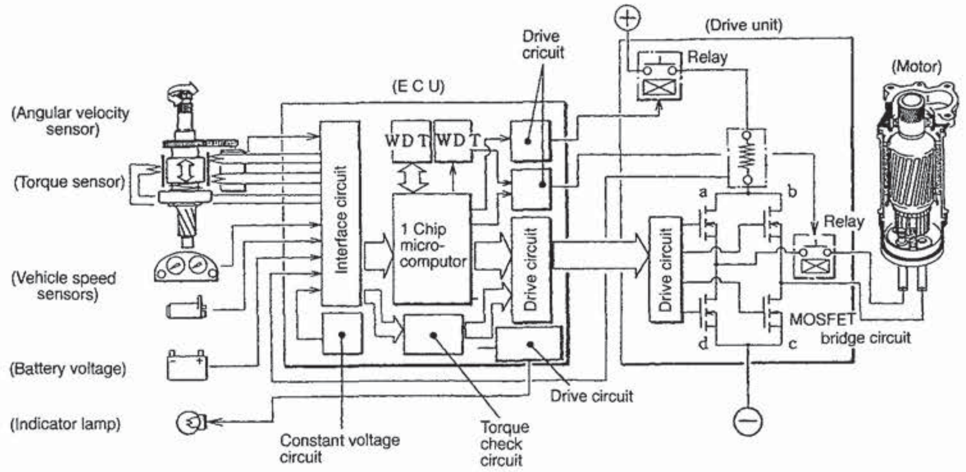


FIGURE 18.15 ECU and drive unit.[6]

The structure is shown in Fig. 18.15. The motor is driven based on instructions from the ECU. The current at this time is monitored by the ECU, and the power supplied to the motor is interrupted in the event of an abnormality.

Depending on the magnitude of the current, some systems are provided with an integrated ECU and power unit, while other systems have each section separate.

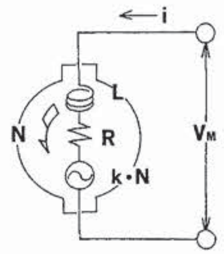


FIGURE 18.16 Motor equivalent circuit.[6]

Motor Control Methods. In the equivalent circuit of the motor, which is shown in Fig. 18.16, the relationship between the terminal voltage V_M, the impedance L, the resistance R, the induced voltage constant k, the revolution speed N, the current i, and the time t, is expressed by the following equation:

$$V_M = L(di/dt) + R \cdot i + k \cdot N \quad (18.1)$$

$$\doteqdot R \cdot i + k \cdot N \quad (18.2)$$

And it is known that the current **i** is proportional to the motor torque T_M.

As can be understood from Eq. (18.2), there are two control methods. In the motor current control method (refer to Fig. 18.17), the target motor current I_T, which is proportional to the motor assist torque T_M, is determined from the signal output T from the torque sensor, and control is performed so that there is no difference between this target current value I_T and the value detected through feedback from the current sensor I_M.

In the motor voltage control method (refer to Fig. 18.18), the motor voltage component ($V_{M1} = R \cdot i = k_T \cdot T_M$; k_T is a proportional constant) which corresponds to the motor assist torque as calculated from the output signal T from the torque sensor, and the motor voltage component ($V_{M2} = k \cdot N$) which corresponds to the motor speed as calculated from the output signal $\dot{\theta}_1$ from the steering wheel angular velocity sensor. These two voltage components are then added and output.

Current Control Method. In this method, the target value for the motor current, which corresponds to the motor assist torque, is set so that it is equal to the vehicle speed response type derived from the signal of the vehicle speed sensor.

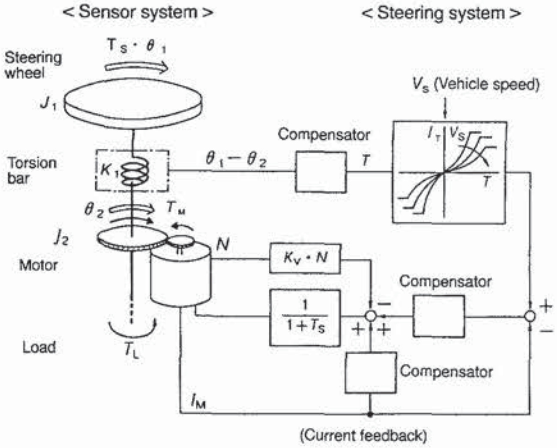


FIGURE 18.17 Diagram of principle of motor current control method.[7]

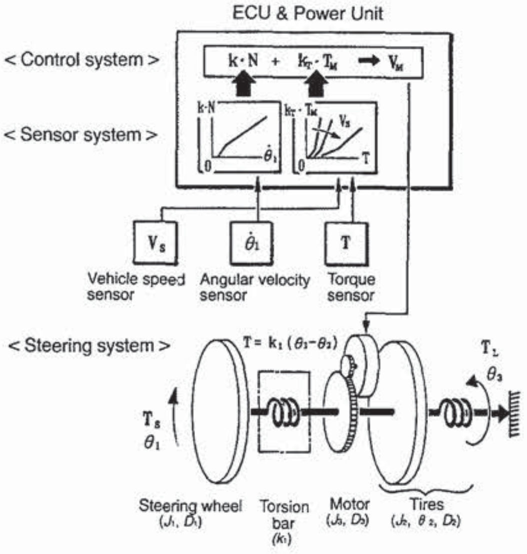


FIGURE 18.18 Diagram of principle of motor voltage control method.[8]

Voltage Control Method. In this method, both the motor torque and the motor speed can be controlled by the output from the torque sensor and the steering wheel angular velocity sensor. When the vehicle is traveling at low speeds, *normal control* is carried out. With this control, the value for $(R \cdot i + k \cdot N)$ in Eq. (18.2) is output to the motor to obtain a good motor response rate (steering response rate) and thus provide a comfortable steering performance. When the vehicle is traveling at high speeds, it is possible to carry out two types of control. In the first method, *return control,* the value for $(k \cdot N)$ is made smaller so that a damping torque which is proportional to the motor speed is generated. In the second method, *damper control,* motor torque is generated in the opposite direction of motor rotation with $V = 0$ when the steering wheel is released in turning.

Normal Control. This is a method of drive control for steering with a reduced steering effort and a good steering response rate. As is shown in Table 18.4, when the steering wheel is turned to the right, FET (a) is ON at the same time that FET (c) is carrying out PWM drive, and current flows to the FET bridge circuit as shown in Fig. 18.19 based on the value of V_M in Eq. (18.2). If the angular velocity of the steering wheel increases, the PWM duty also increases.

An example of the steering characteristics in an actual vehicle can be seen in Fig. 18.20. Figure 18.20*a* represents the assist characteristics during stationary swing, with the dotted line representing the steering characteristics when no assist force is provided. Figure 18.20*b* is the steering response rate; Fig. 18.20*c* is the steering characteristics corresponding to the vehicle speed; and Fig. 18.20*d* is the steering characteristics relative to the lateral acceleration of the vehicle.

TABLE 18.4 FET Drive During Normal Control[5]

Steering condition	FET (a)	FET (b)	FET (c)	FET (d)	Motor operation
Steering to right	ON	OFF	PWM	OFF	Operates in direction steering to the right
Straight ahead	OFF	OFF	OFF	OFF	Stops
Steering to left	OFF	ON	OFF	PWM	Operates in direction steering to the left

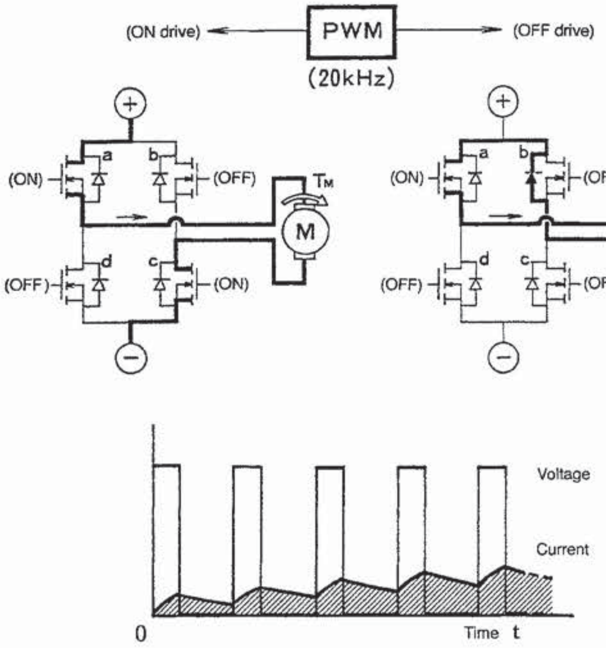
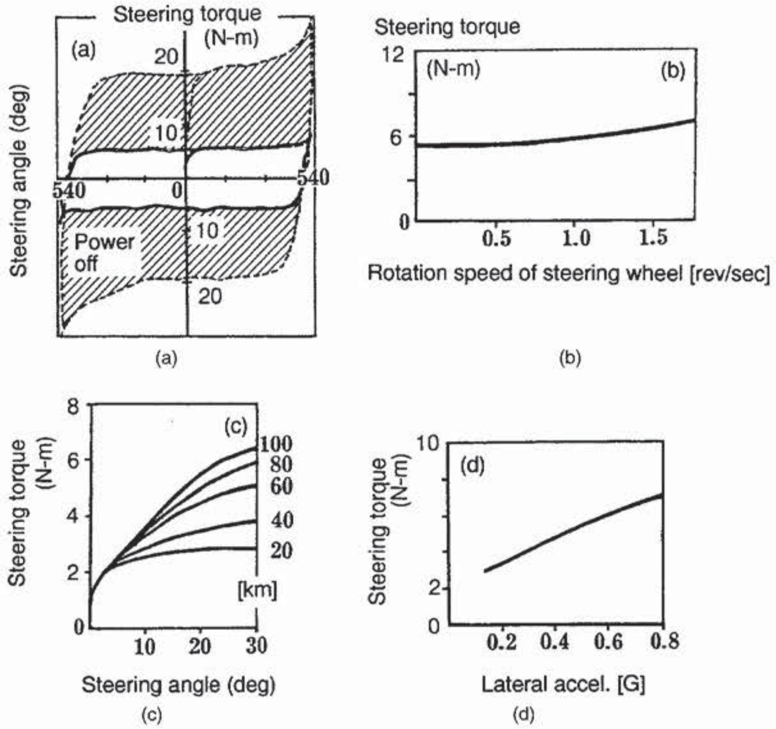


FIGURE 18.19 Diagram of bridge circuit operation.[6]



(a)

(b)

(c)

(d)

FIGURE 18.20 Steering characteristics.[8]

Return Control. This method of drive control is for varying the steering wheel return characteristics. When the driver is returning the steering wheel to the neutral position at low-speed driving, the motor current is immediately reduced in order to make the motor operate in the reverse direction to the torque generating directions. As a result, a good returnability of the steering wheel should be obtained. At high-speed driving, motor current is gradually reduced in order to suppress returnability and obtain more stable steering characteristics.

As shown in Table 18.5, when the steering wheel returns to the straightforward position after being turned to the right, FET (a) carries out PWM-r drive based on the signals from the steering wheel angular velocity sensor, and, at the same time, FET (c) also carries out PWM drive based on the signals from the torque sensor.

TABLE 18.5 FET Drive During Return Control[5]

Steering condition	FET (a)	FET (b)	FET (c)	FET (d)
Return from right steering to straight ahead	PWM-r	OFF	PWM	OFF
Return from left steering to straight ahead	OFF	PWM-r	OFF	PWM

The results obtained by means of this return control are shown in Fig. 18.21. Figure 18.21*a* represents the steering characteristics when return control is not applied, and Fig. 18.21*b* represents the steering characteristics when return control is operating. The hysteresis is greater and the returnability is weaker than in Fig. 18.21*a*.

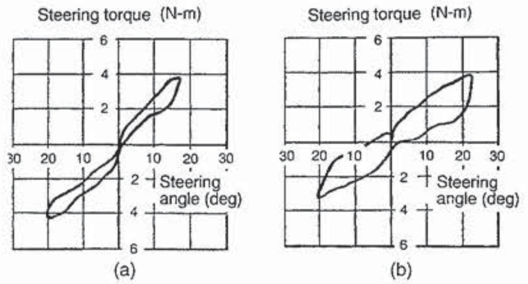


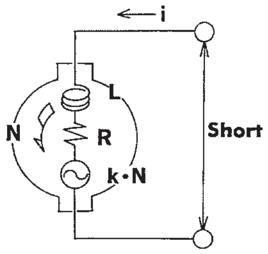
FIGURE 18.21 Comparison of steering wheel return characteristics.[6]

Damper Control. This method of drive control is for improving the convergence of the steering wheel when the vehicle is traveling at high speeds and for eliminating wandering of the steering wheel caused by the tire inputs.

As shown in Fig. 18.22, when the motor terminals are shorted, it is possible to generate motor torque in the reverse direction in proportion to the speed of the motor as is shown by the equation in Fig. 18.22, and this characteristic is utilized for control. Figure 18.23 shows the results of the application of damper control during a convergence test carried out at a vehicle speed of 120 km/h.

18.1.6 Power-Saving Effectiveness

Because the EPS system is a "power-on-demand" system, which supplies only the necessary amount of power at the necessary time, almost no power losses occur when the steering wheel is not being turned. Because of this, the system has extremely high fuel efficiency. Table 18.6 shows the results of measurements of fuel consumption during different driving modes to compare EPS with the hydraulic system. The LA-4 mode corresponds to urban road driving.



$$0 = k_T \cdot T_M + k \cdot N \quad \cdots\cdots\cdots(3)$$

$$\therefore T_M = -\left(\frac{k}{k_T}\right)N \quad \cdots\cdots\cdots(4)$$

FIGURE 18.22 Principle of damper control.

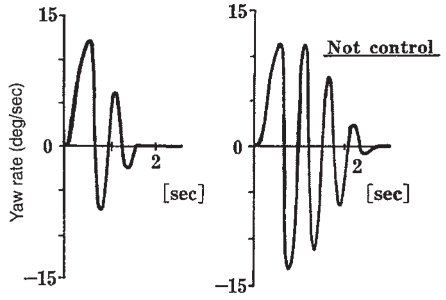


FIGURE 18.23 Effect of damper control.

18.1.7 Trends in Research and Development

The demands for faster speed, higher quality, and reduced power requirements in vehicles is continually increasing. In order to respond to these demands, research and development is under way on the application of electronic control systems with the aim of further improving functions and performance. Features that are being proposed include the introduction of fuzzy logic and the application of power steering, which responds to the driving environment by varying the assist amount in accordance with the traffic conditions or the road surface conditions in order to provide steering feel to fit the sensitivities of human operators. The most important of these is probably active reaction power steering, which provides feedback to the driver regarding the behavior of the vehicle in the form of steering reaction force. Such a system provides the driver with information regarding the operating conditions of the vehicle, for instance, the yaw velocity and/or lateral acceleration, as steering reaction forces. Not only would it improve the relationship between the driver and the vehicle to make it possible to achieve a steering feel that better suits the sensitivities of the driver, but a function that automatically compensates for irregularities in vehicle behavior caused by disturbances could also be expected.

Figure 18.24 is a system control block diagram in which the yaw rate is fed back as a steering reaction force.

Figure 18.25 shows an example of the effect of suppressing irregularities in vehicle behavior caused by side disturbances by comparing the amount of lateral removal when braking on a rut of road surface with the case for conventional power steering.

18.2 FOUR-WHEEL STEERING SYSTEMS (4WS)

For vehicles with extremely long wheel bases and vehicles which need to be operated in narrow places, the concept of a four-wheel steering system is attractive. In such systems, the rear wheels are turned in the opposite direction to the steering direction of the front wheels in

TABLE 18.6 Measurement Results on Mode Fuel Consumption[6]

			Improvement in fuel consumption	
Mode	EPS (mpg)	Hydraulic power steering (mpg)	Fuel consumption difference (mpg)	Improved rate (%)
LA-4	20.51	20.01	0.80	2.50
Highway	28.88	28.18	1.14	2.52

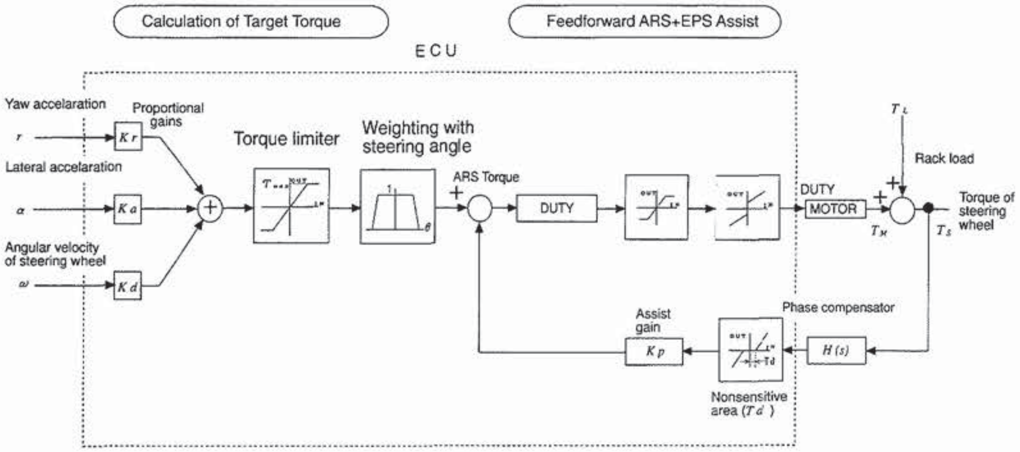


FIGURE 18.24 Block diagram of active reaction control.

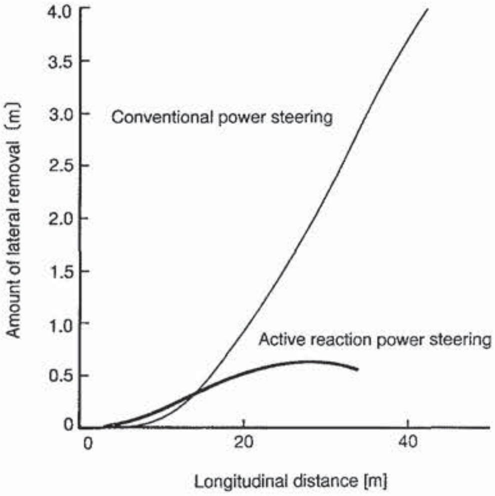


FIGURE 18.25 Effect of active reaction power steering.

order to make the turning radius as small as possible and to improve the handling ability. Such 4WS systems have been under development for some time. However, the concept of the system being used in passenger vehicles for the purpose of improving vehicle stability and steering response at medium to high speeds is relatively new.

A 4WS system for passenger cars has the following two aims:

- Reducing the turning motion (yawing) of the vehicle by steering the rear wheels in the same direction as the front wheels, thus improving the vehicle stability at high speeds
- Improving the steering response at medium speeds, while at the same time reducing the turning circle radius at low speed, by steering the rear wheels in the opposite direction to the front wheels

18.2.1 Basic Principles of 4WS

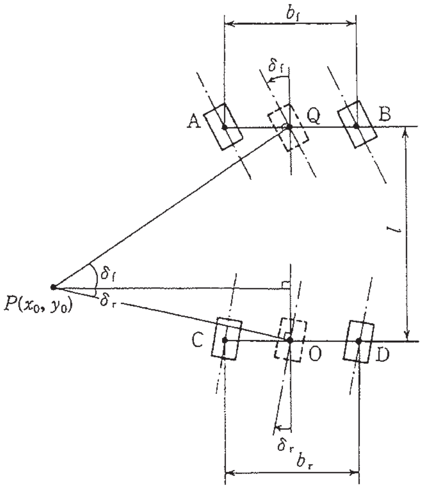


FIGURE 18.26 Half vehicle model of 4WS.[1]

Shortening the Minimum Turning Radius. As shown in Fig. 18.26, provided the origin point of the coordinate is set at the center of the rear tread, the coordinates of the turning center of the vehicle $P(X_0, Y_0)^v$ when the rear wheels are turned in the opposite direction from the front wheels are given in the following equations:

$$x_0 = \iota \,/\, (\tan \delta_f + \tan \delta_r) \tag{18.3}$$

$$y_0 = \iota\, (\iota \cdot \tan \delta_1) \,/\, (\tan \delta_f + \tan \delta_r) \tag{18.4}$$

And, if the turning radius for the front outer wheel is R and the difference between the turning radius of the front and rear outer wheels is ΔR, then:

$$R = \overline{BP} = \sqrt{\left(\frac{b_f}{2} + X_0\right)^2 + (\iota - y_0)^2} \cdot \sqrt{\left(\frac{b_f}{2} + \frac{\iota}{\tan \delta_f + \tan \delta_r}\right)^2 + \left(\frac{\iota \cdot \tan \delta_f}{\tan \delta_f + \tan \delta_r}\right)^2} \tag{18.5}$$

$$\begin{aligned}
\Delta R &= \overline{AP} - \overline{CP} \\
&= \sqrt{\left(-\frac{b_f}{2} + X_0\right)^2 + (\iota - y_0)^2} - \sqrt{\left(-\frac{b_r}{2} + X_0\right)^2 + y_0^2} \\
&= \sqrt{\left(-\frac{b_f}{2} + \frac{\iota}{\tan \delta_f + \tan \delta_r}\right)^2 + \left(\frac{\iota \cdot \tan \delta_f}{\tan \delta_f + \tan \delta_r}\right)^2} \\
&\quad - \sqrt{\left(-\frac{b_r}{2} + \frac{\iota}{\tan \delta_f + \tan \delta_r}\right)^2 + \left(\frac{\iota \cdot \tan \delta_f}{\tan \delta_f + \tan \delta_r}\right)^2}
\end{aligned} \tag{18.6}$$

In these equations, δ_f is the front wheel steering angle (average of left and right wheels), δ_r is the rear wheel steering angle (average of left and right wheels), b_f is the tread of the front wheels, b_r is the tread of the rear wheels, and ι is the wheelbase.

$$R = \overline{PQ} = \frac{\iota}{\sin\delta_f + \cos\delta_f + \tan\delta_f} \tag{18.7}$$

$$\begin{aligned}
\Delta R &= \overline{PQ} - \overline{PO} \\
&= \frac{\iota}{\sin\delta_f + \cos\delta_f + \tan\delta_f} \\
&\quad - \frac{\iota}{\sin\delta_f + \cos\delta_f + \tan\delta_f}
\end{aligned} \tag{18.8}$$

It can be seen from Eq. (18.7) that when the rear wheels steer in the opposite direction as the front wheels, the turning radius becomes smaller than when the rear wheels are not steered ($\delta_r = 0$). In addition, it can be seen from Eq. (18.8) that when the steering amount for the front and rear wheels is the same ($\delta_f = \delta_r$), then it becomes possible to obtain a difference of 0 between the turning radius of the front and rear wheels

Improvement in Stability and Maneuverability When Driving at Medium to High Speeds. The steering characteristics in yaw velocity and lateral acceleration of the vehicle with 4WS in which the rear wheels are steered in proportion to the front wheels steering angle in the same direction, are shown in Table 18.7, along with that for a vehicle yaw 2WS system. However, the half-car vehicle model shown in Fig. 18.26 is used as the vehicle model.

From Table 18.7, in a system in which the rear wheels are turned in the same direction as the front wheels, the stability factor K, the damping ratio ζ, and the natural oscillation of yawing ωn are not different from the values for conventional vehicles with 2WS, so that the intrinsic stability of the vehicle will not vary. On the other hand, because the lateral acceleration response delay will become smaller as the coefficients of numerators s and s^2 increase and the steady state gain in the yaw velocity will drop by the proportion to $(1 - k)$, the yawing movement which accompanies the lateral movement of the vehicle will become less and the stability within the range of practical use will be improved. If, however, the rear wheels are steered in the opposite direction to the front wheels, $k < 0$ and so $1 - k > 1$. This means that the steady state gain of the yaw velocity will increase, with the result that the steering response will be improved. The symbols used in Table 18.7 and Fig. 18.27 are explained in Table 18.8.

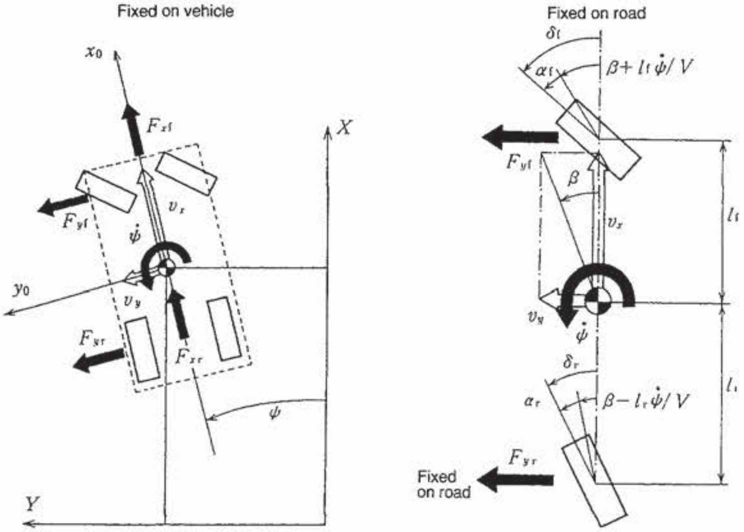


FIGURE 18.27 Coordinates fixed on road and vehicle in half vehicle model.[1]

18.2.2 Classification of 4WS

Four-wheel steering systems that are currently being implemented in vehicles are classified according to their functions and mechanisms. The aims and characteristics of each system are briefly explained in Tables 18.9 and 18.10. Because the fully mechanically controlled system shown in Table 18.10 has a low degree of control freedom, there is less tendency for it to be used. Regardless of its low cost, however, this system provides every basic function of 4WS, so that there is a possibility that it may be used in the future, primarily in smaller vehicles.

TABLE 18.7[1]

Symbol	Name	
K	Stability factor (s^2/m^2)	
A	Steady state gain in yaw velocity	(1/s)
B_f, B_r	Steady state gain vehicle side slip angle for front and rear	(s)
T_f, T_r	Time constant for front/rear yaw velocity	(s)
τ_f, τ_r	Time constant for front/rear side slip angle	(s)
ω_n	ω_n : Response frequency	(rad/s)
	$f_n(\omega_n/2\pi)$: Natural frequency	(Hz)
$\zeta \cdot \omega_n$	$\zeta \cdot \omega_n$:	(1/s)
	$\zeta \cdot \omega_n \times V$: Steering capacity	(m/s^2)
	ζ : Damping rate	

	Vehicle with 2WS	Vehicle with 4WS in which the rear steering angle is proportional to front
Stability factor	$\frac{m(l_r K_r - l_f K_f)}{2l^2 K_f K_r} = K$	K
Steady state gain in yaw velocity	$\frac{1}{1+KV^2} \cdot \frac{V}{l}$	$\frac{1}{1+KV^2} \cdot \frac{V}{l}$
Damping rate	$\frac{1}{2} \cdot \frac{(k_f + K_r I_t + (K_f l_f + K_r l_r^2)m}{\sqrt{m I_z K_f K_r l^2 (1+KV^2)}} = \zeta$	ζ
Response frequency	$\frac{2l}{V} \cdot \sqrt{\frac{K_f K_r (1+KV^2)}{I_z m}} = \omega_n$	ω_n
Steering capacity	$\frac{1}{V}\left(\frac{K_f + K_r}{m} + \frac{K_f l_f^2 + K_r l_r^2}{l_z}\right) = \zeta \cdot \omega_x$	$\zeta \cdot \omega_n$
$\frac{\dot{\varphi}}{\delta_H}$	$\frac{1}{i_s} G_\varphi(0) \frac{1 + T_f s}{1 + \frac{2\zeta}{\omega_n} s + \frac{1}{\omega_n^2} s^2}$ $G\varphi(0) = \frac{1}{1+KV^2} \cdot \frac{V}{l}$	$\frac{1-k}{i_s} G\varphi(0) \frac{1 + (1+\lambda\phi) T_f s}{1 + \frac{2\zeta}{\omega_n} s + \frac{1}{\omega_n^2} s^2}$ $G\phi(0) = \frac{1}{1+KV^2} \cdot \frac{V}{l}$ $\lambda\varphi = \frac{k}{1-k} \cdot \frac{T_f - T_r}{T_f}$
$\frac{\alpha_y}{\delta_H}$	$\frac{1}{i_s} G_{ay}(0) = \frac{1 + T_{ay1} s + T_{ay2} s^2}{1 + \frac{2\zeta}{\omega_n} s + \frac{1}{\omega_n^2} s^2}$ $G_{ay}(0) = \frac{1}{1+KV^2} \cdot \frac{V^2}{l}$ $T_{ay1} = \frac{l_r}{V}, \quad T_{ay2} = \frac{l_z}{2K_r l}$	$\frac{1-k}{i_s} G_{ay}(0) \cdot \frac{1 + (1+\lambda_{ay1}) T_{ay1} s + (1+\lambda_{ay2}) T_{ay2} s^2}{1 + \frac{2\zeta}{\omega_n} s + \frac{1}{\omega_n^2} s^2}$ $G_{ay}(0) = \frac{1}{1+KV^2} \cdot \frac{V^2}{l}$ $\lambda_{ay1} = \frac{k}{1-k} \cdot \frac{T_{ay1} + T'_{ay1}}{T_{ay1}}$ $\lambda_{ay2} = \frac{k}{1-k} \cdot \frac{T_{ay2} + T'_{ay2}}{T_{ay2}}$ $T_{ay1} = \frac{l_r}{V}, \quad T_{ay2} = \frac{l_f}{V}$ $T_{ay2} = \frac{l_z}{2K_r l}, \quad T'_{ay2} = \frac{l_z}{2K_f l}$

TABLE 18.8[1]

Symbol	Dimension	Name
m	kg	Mass of the vehicle
I_z	$\mathrm{kg} \cdot \mathrm{m}^2$	Moment of the inertia of vehicle
l	m	Wheel base
l_f, l_r	m	Distance between center of gravity and front/rear wheel shaft
V	m/s	Vehicle speed $V = \sqrt{v_x^2 + v_y^2}$
v_x, v_y	m/s	Longitudinal and lateral velocity of the center of gravity
a_x, a_y	$\mathrm{m/s^2}$	Longitudinal and lateral acceleration of the center of gravity
ψ, ϕ	rad	Yaw angle, roll angle
$\dot{\psi}, \dot{\phi}$	rad/s	Yaw velocity roll rate
β	rad	Side slip angle of the center of gravity
δ_f, δ_r	rad	Front/rear wheel steering angle
α_f, α_r	rad	Front/rear side slip angle
K_f, K_r	N/rad	Equivalent cornering power of front/rear
δ_H	rad	Steering wheel angle
i_s	I	Steering ratio in over all

TABLE 18.9[1]

Classification by functions	Aims
Small range of rear steer angle only controlled electronically	Improvement of steering response and vehicle stability in medium to high speed
Not only small range in medium to high speed but large range in low speed of rear steering angle are controlled electronically	In addition to the above, making the minimum turning radius small

TABLE 18.10[1]

Classification by mechanism	Feature
Full mechanical system	Simple mechanism
Electronic-hydraulic system	High degree for control freedom (compact actuator)
Electronic-mechanical-hydraulic system	High degree of freedom (mechanism is not simple)
Full electric system (electronic-electric system)	High degree of control freedom simple mechanism

18.2.3 Introduction to Each System

Fully Mechanical System. This was the first 4WS system used in passenger vehicles. It was adopted in the Honda Prelude in 1987 and in the Honda Accord in 1989. This system is a front-wheel steering angle responsive-type system in which the rear-wheel steering angle is determined wholly by the steering angle of the front wheels. The structure and fabrication of this system are shown in Figs. 18.28 and 18.29.

In rack-and-pinion type steering gearboxes for the front wheels, a rear-wheel steering pinion is provided in order to transmit the steering angle of the front wheels to the rear wheels. The displacement in the steering angle is transmitted to the rear steering gearbox via the center steering shaft. As shown in the figure, the rear steering gearbox consists of a combination of an eccentric shaft and a planetary gear. The input and output characteristics that can be obtained are shown in Fig. 18.30. The result of this is that when the steering angle for the front

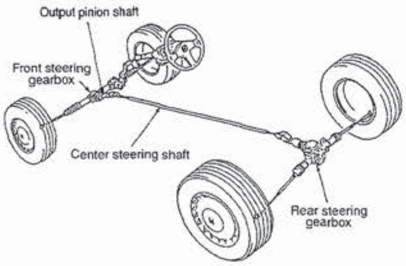


FIGURE 18.28 Fully mechanical 4WS.

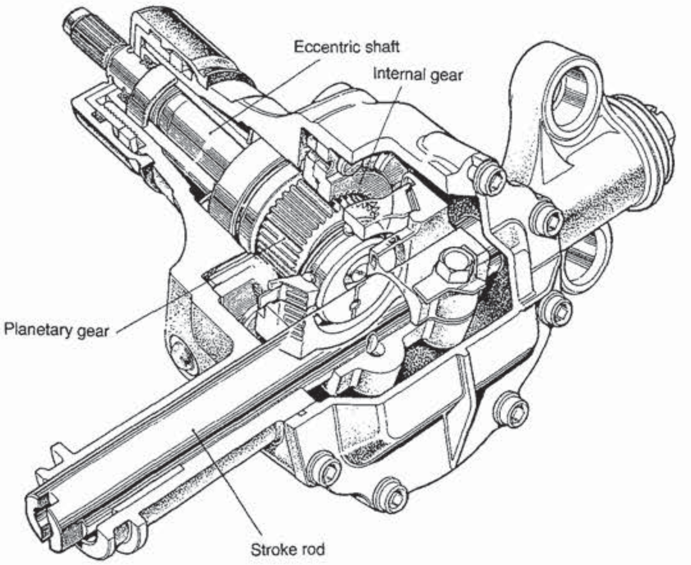


FIGURE 18.29 Rear steering actuator of full mechanical 4WS.

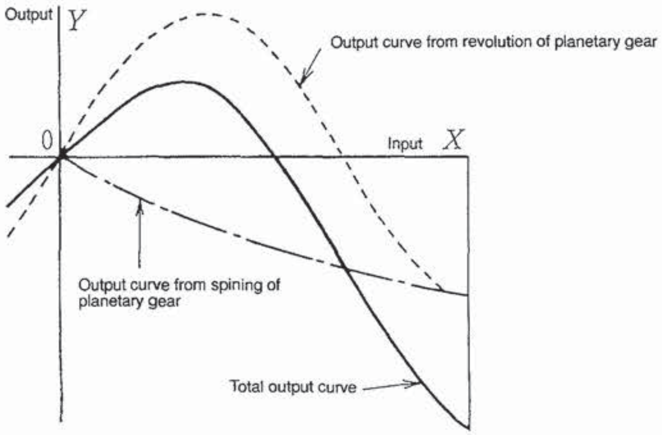


FIGURE 18.30 Input-output characteristics of rear steering gearbox.

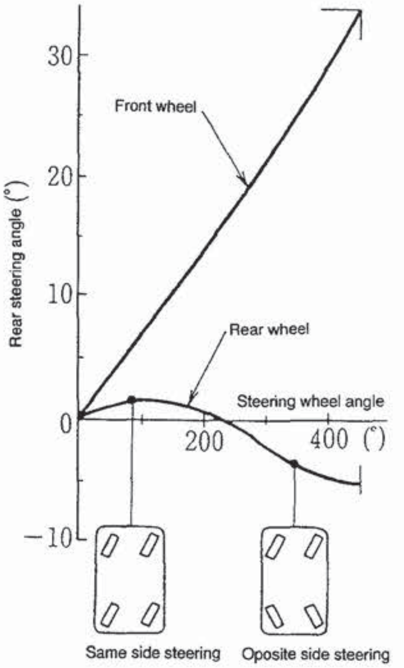


FIGURE 18.31 Characteristics of front steering speed-responsive 4WS.

wheels is small, the rear wheels turn in the same direction, but when the steering angle for the front wheels is large, the rear wheels turn in the opposite direction, as shown in Fig. 18.31. During high-speed driving, since the front wheels are only turned by very small amounts, the rear wheels thus turn in the same direction, and driving stability is improved. At extremely low speeds where the front wheels are steered through much larger angles, the rear wheels turn in the opposite direction and the turning radius becomes small, thus, the working ability should be improved.

Electronic-Hydraulic Control System

Vehicle Speed/Lateral Acceleration Responsive Type. This system was installed in vehicles like the 1986 Nissan Skyline. The structure of the system and the construction of the rear gearbox are shown in Figs. 18.32 and 18.33, respectively. In this system, a special hydraulic valve, which generates hydraulic pressure balanced with the reaction force from the front wheels in proportion to the lateral acceleration, is provided in the front-wheel power steering system, and this hydraulic pressure is transmitted to the actuator for the rear wheels. The rear-wheel actuator contains a high spring rate spring which allows displacement of the output rod to the position that balances the hydraulic pressure received. The rear wheels are steered in the same direction as the front wheels by means of the displacement of this rod. Accordingly, the relationship between the vehicle speed and the steering angle of the rear wheels varies in accordance with

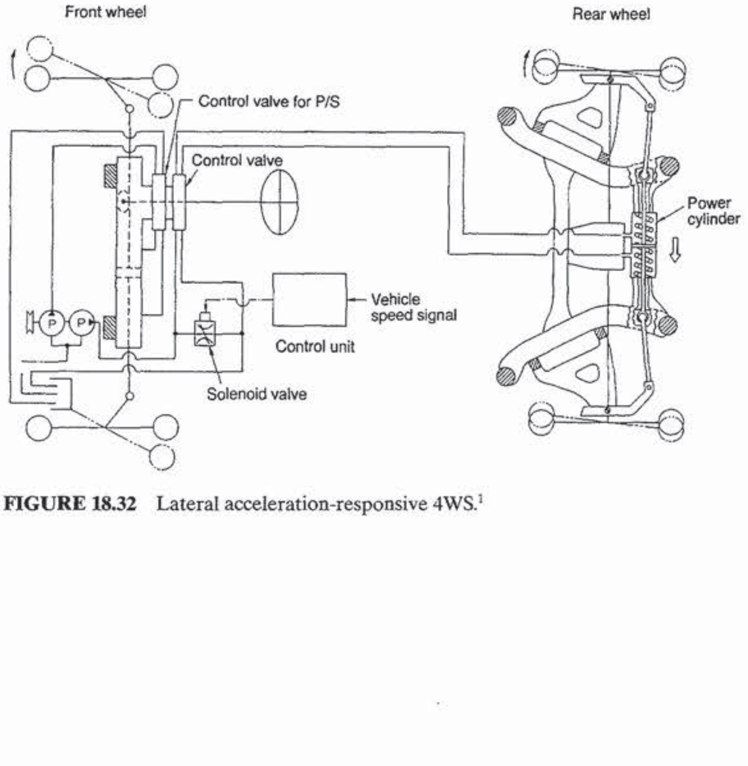


FIGURE 18.32 Lateral acceleration-responsive 4WS.[1]

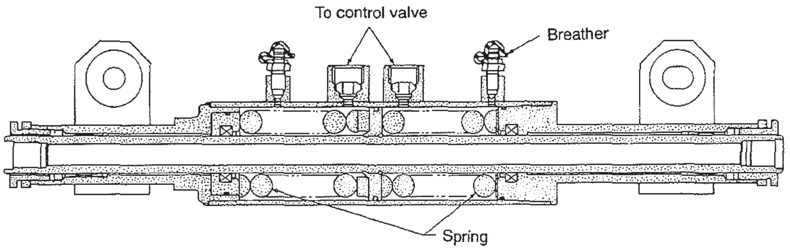


FIGURE 18.33 Rear steering actuator of lateral acceleration-responsive 4WS.[1]

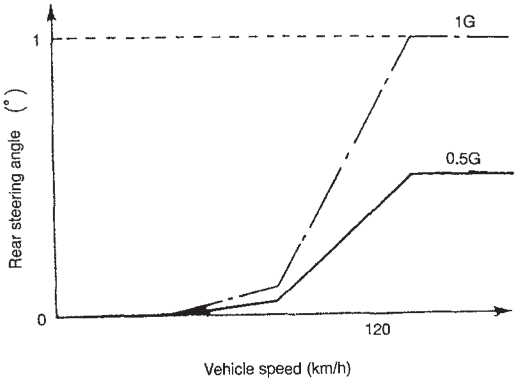


FIGURE 18.34 Characteristics of rear steering angle.[1]

the lateral acceleration as shown in Fig. 18.34. In this system, the maximum steering angle of the rear wheels is confined to a fairly low value, so that it is not intended to be a system for improving the minimum turning radius of the vehicle at very low speeds.

Vehicle Speed/Front-Wheel Steering Angle/Steering Wheel Velocity Responsive Type. This system was installed in vehicles such as the 1988 Nissan Sylvia, and was developed for the purpose of further improvement of vehicle stability and maneuverability during medium- to high-speed driving. The structure of the system and the construction of the rear-wheel actuator (solenoid servo valve) are shown in Figs. 18.35 and 18.36, respectively. In this system, the working fluid discharged from the hydraulic pump is directly introduced to the solenoid servo valve and controlled by instructions from an ECU (electronic control unit), after which it is sent to the power cylinder.

The ECU detects the front-wheel turning angle and the steering speed of the steering wheel by means of signals from the steering wheel angle sensor located in the steering wheel. The values thus calculated and the vehicle speed are used to determine the steering angle for the rear wheels. The maximum steering angle of the rear wheels is confined to a fairly low value in this system also, so that it was intended only to be a system for improving the vehicle stability and maneuverability at medium to high speeds. However, it does not improve only the stability by turning the rear wheels in the same direction, but also improves the maneuverability by momentarily turning the rear wheels in the opposite direction to the front wheels when the steering wheel is turned quickly at medium vehicle speeds. That is, the initial yawing movement (yaw velocity) of the vehicle is improved and steering response is thus improved. Figure 18.37 shows the steering angle pattern of the rear wheels with time expressed along the horizontal axis.

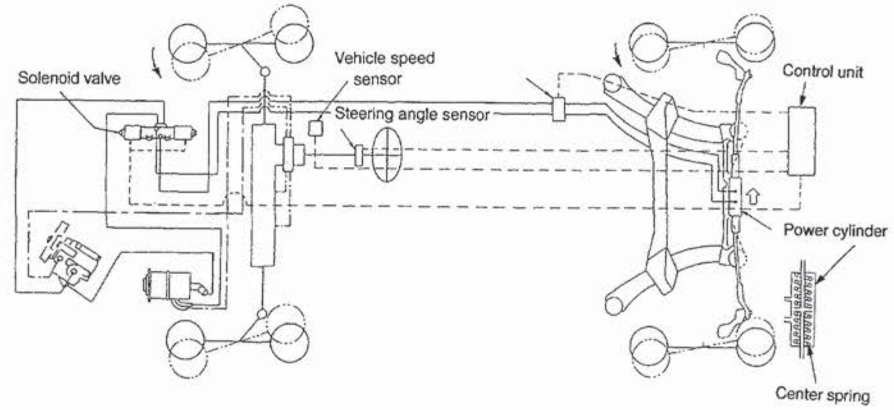


FIGURE 18.35 System construction of front-wheel-responsive 4WS.[1]

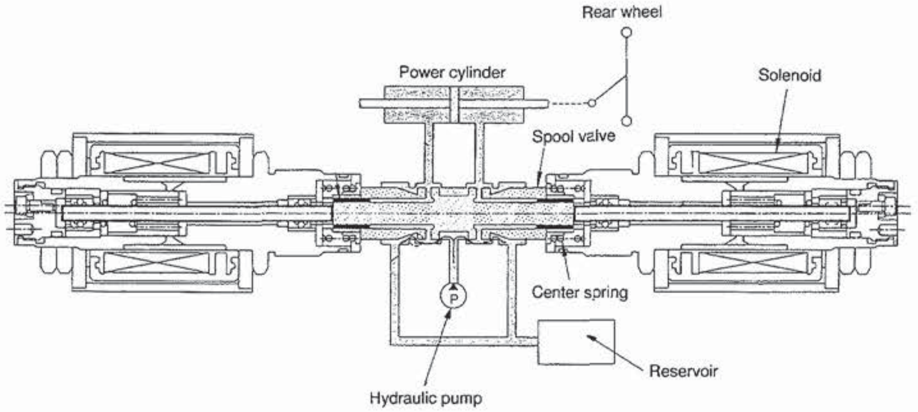


FIGURE 18.36 Solenoid servo valve of front steering angle-responsive 4WS.[1]

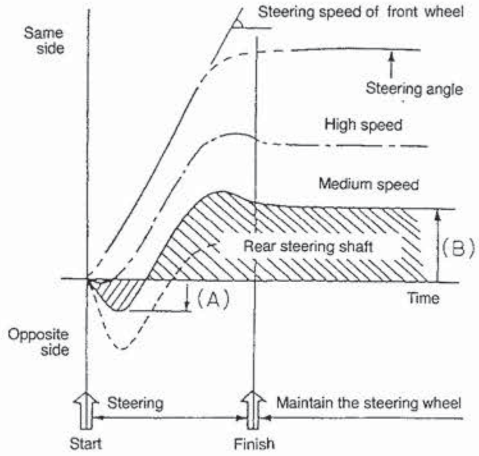


FIGURE 18.37 Characteristics of front steering angle-responsive 4WS.[1]

Electronic-Hydraulic-Mechanical Systems

Vehicle Speed/Front-Wheel Steering Angle Responsive Type. This system was introduced in the 1987 Mazda Capella, and was designed with the aim of being a system for improving both stability when driving at high speeds, like the fully mechanical system, and vehicle minimum turning radius at low speeds. The structure of the system and the construction of the rear actuator are shown in Figs. 18.38 and 18.39, respectively. This system can be broadly divided into the power assist section and the phase control section. The power assist section consists of a linear spool valve and a power piston, and utilizes hydraulic pressure as the dynamic force. The phase control section comprises a bevel gear which engages with the input shaft, a control yoke, and a control rod which is connected to a valve. The angle of the control yoke is driven by a stepping motor via a worm gear, thus controlling the turning direction of the rear wheels so that the rear wheels can be turned, not only in the same direction, but also in the opposite direction as the front wheels.

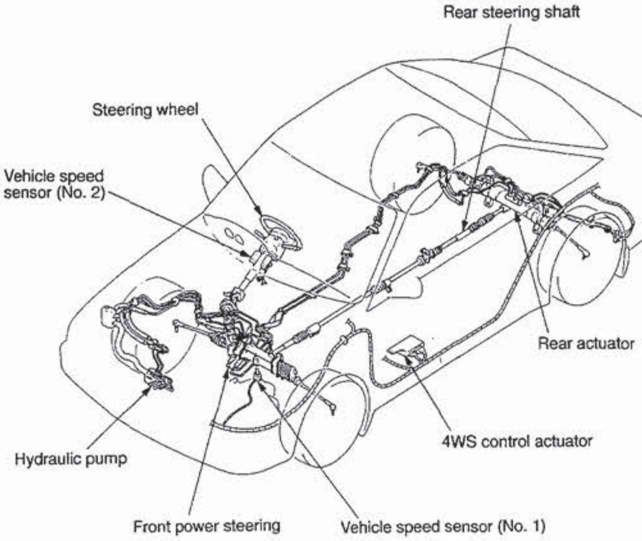


FIGURE 18.38 Front steering angle and vehicle speed-responsive 4WS.[11]

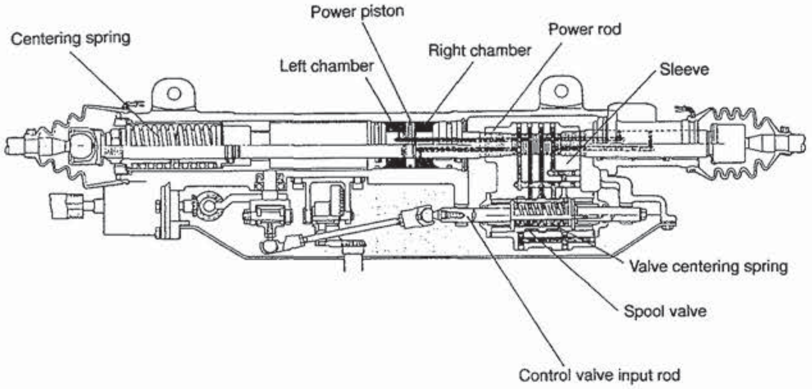


FIGURE 18.39 Rear steering actuator.[11]

In Fig. 18.40, *a* represents the situation in which the rear wheels are turned in the same direction, and *b* represents the situation in which the rear wheels are in the opposite direction. The direction of movement of the valve spool in relation to the turning direction of the input shaft is opposite for *a* and *b* as shown. Figure 18.41 shows the situation in which the ratio of the steering angle of the rear wheels with respect to the steering angle of the front wheels is continuously changing with respect to the vehicle speed.

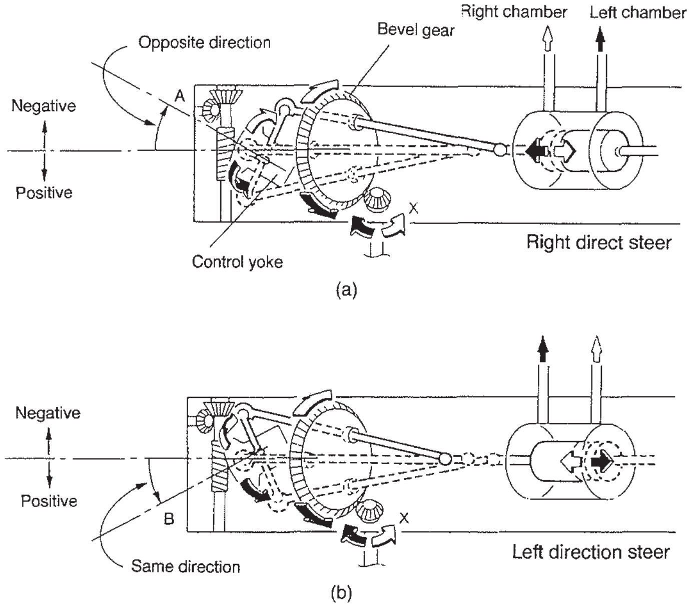


FIGURE 18.40 Explanation diagram of rear actuator.[11]

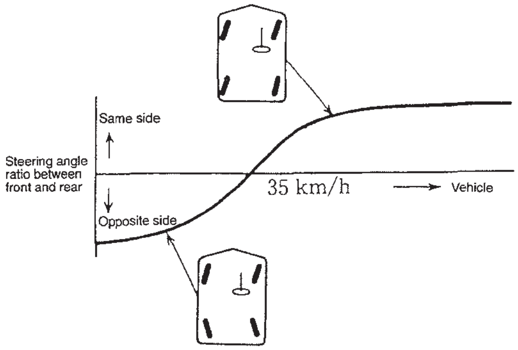


FIGURE 18.41 Characteristics of front steering angle and vehicle speed-responsive 4WS.[1]

Vehicle Speed/Front-Wheel Steering Angle/Yaw Velocity Responsive Type. This system was introduced in the 1992 Toyota Soarer. Figures 18.42 and 18.43 show the system structure and the construction of the rear-wheel steering actuator, respectively. The system consists of two sections: a *mechanical-hydraulic type* steering mechanism to steer the rear wheels by a

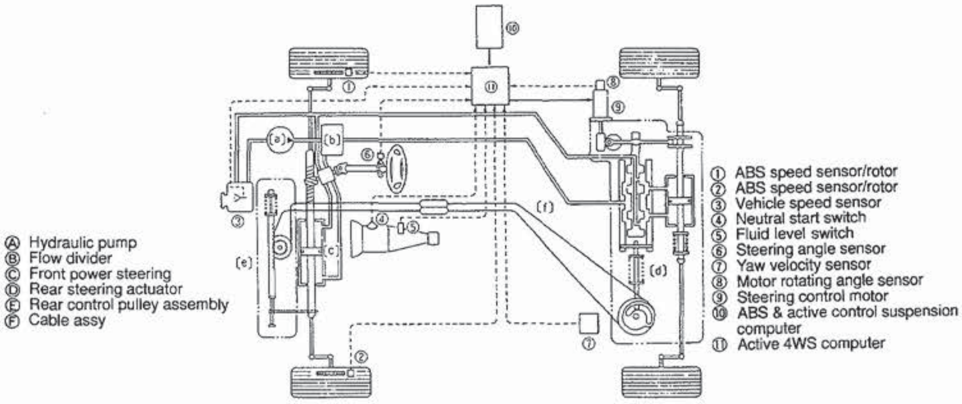


FIGURE 18.42 Active 4WS system.[10]

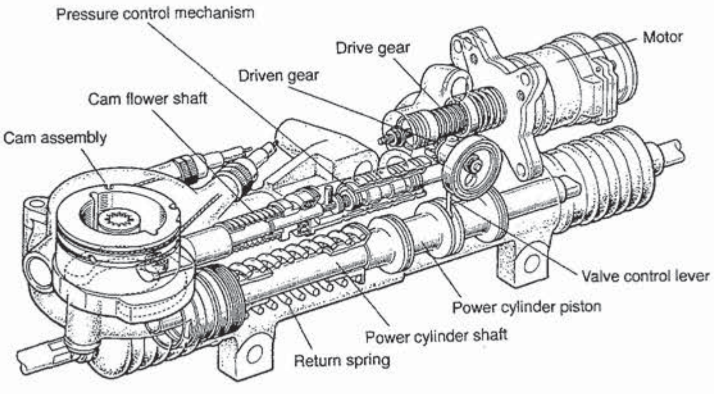


FIGURE 18.43 Rear steering actuator unit.[10]

considerably large steering angle in the opposite direction to the front wheels in order to make the turning radius small at low speeds, and an *electronic-hydraulic type* steering mechanism to steer the rear wheels by a considerably smaller value in order to not only improve the steering response and the stability during middle- to high-speed driving, but also to suppress the unexpected movements due to the outside disturbances.

Mechanical-Hydraulic Type Steering Mechanism. This mechanism consists of steering angle transmission cables, which transmit information concerning the steering angle of the front wheels to the rear-wheel steering actuator; a joint to connect the front and rear cables; a pulley assembly, which converts the displacement of the front-wheel steering rack into cable movement; a cam, which converts the movement of the cable into displacement of the sleeve valve inside the rear-wheel steering actuator; and a copy valve, which steers the rear wheels as far as the instruction given by the sleeve valve and maintains them in position. Furthermore, a dead zone is provided in the pulley assembly so that the rear wheels are not steered in the opposite direction to the front wheels when the vehicle is traveling at high speed.

Electronic-Hydraulic Type Steering Mechanism. This mechanism is designed to improve the steering response and the stability of the vehicle and to suppress the unexpected movements from outside disturbances when the vehicle is traveling at medium to high speeds. For this purpose, the system consists of five types of sensors for detecting the driving conditions and the steering situations, an ECU which determines the steering angle for the rear wheels based on the signals from these sensors, a pulse motor which drives the gear mechanism based on instructions from the ECU, a gear mechanism which converts the rotation of the pulse motor into a spool valve displacement, and a hydraulic copy valve which steers the rear wheels as far as the instruction given by the sleeve valve and maintains them in that position.

Control Method. The basic algorithm to determine the steering angle for the rear wheels is given in Fig. 18.44. When driving in the medium- to high-speed range, the front-wheel steering angle proportional gain for the rear wheels is set to 0; instead of that, the yaw velocity proportional gain in the direction which suppresses the yaw rate is set. Accordingly, in the medium- to high-speed range, the rear wheels are turned in the same direction as front wheels, with the same effect as that of the other system mentioned before. The feature of this system is that a yaw rate sensor is provided to detect the vehicle dynamics and utilizes its signal as a control parameter. The effect of this is the addition of a function which automatically compensates the confusion in vehicle behavior caused by outside disturbance.

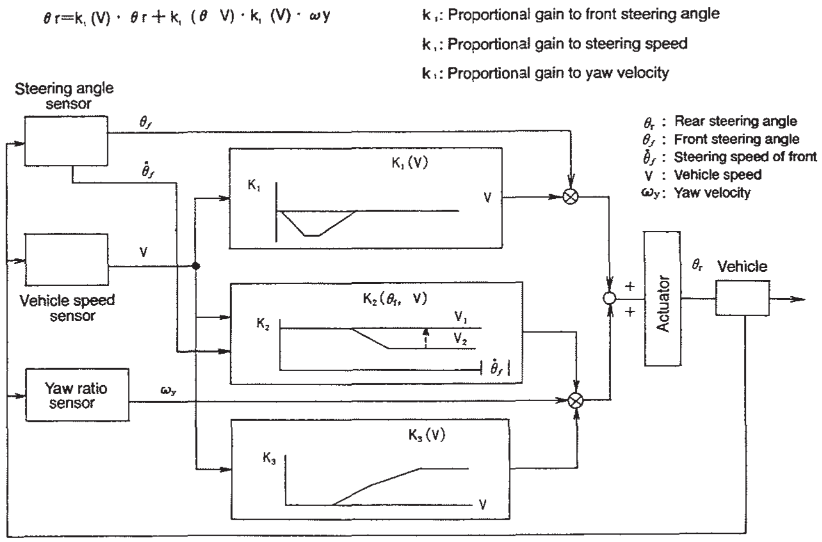


FIGURE 18.44 Control block diagram.[10]

18.2.4 Electronic-Electric Control Systems

Vehicle Speed/Front-Wheel Steering Angle/Steering Speed Responsive Type. This system was introduced in the 1991 Honda Prelude. The system structure is shown in Fig. 18.45, and the construction of the rear-wheel actuator is shown in Fig. 18.46. This system has a structure whereby the rotation of an electric motor that is controlled by the ECU is converted into linear motion by means of a ball screw to directly drive the rear wheels. The system is easy to install. Cross sections of the structures of the front-wheel steering angle sensor and the rear-wheel steering angle sensor are shown in Figs. 18.47 and Fig. 18.48, respectively.

Also in this system, the turning of the rear wheels is determined by the vehicle speed, the front steering angle, and the steering wheel velocity. The determining algorithm is similar to the system mentioned previously. In this system, improvements in the stability at high speeds, improvement in the steering response when the steering wheel is turned quickly, and improvement in the turning radius of the vehicle at low speeds have been realized simultaneously.

The most recent example of an electronic-electric system is the system[12] that has been installed in the 1993 Nissan Laurel. The approach to control in this system is the same as in the

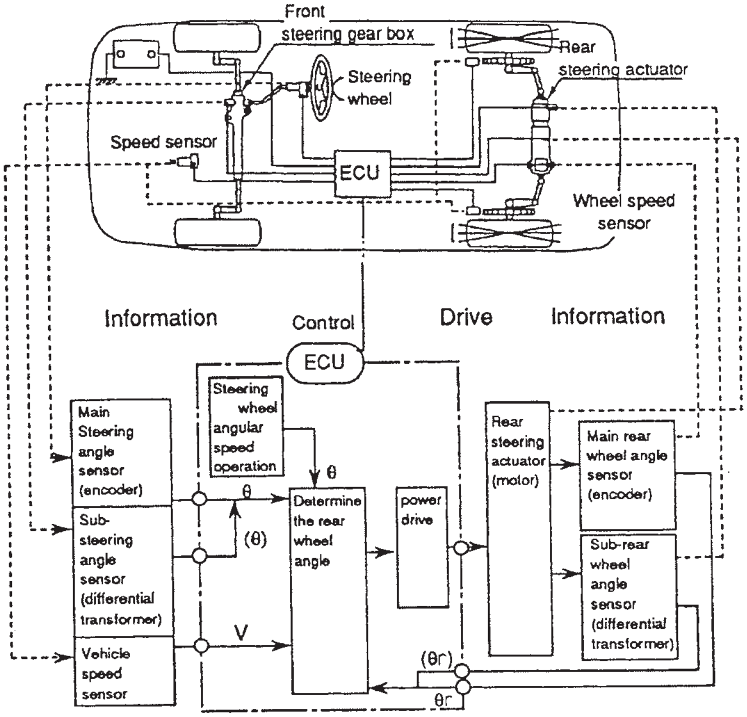


FIGURE 18.45 Block diagram of the full-electric 4WS.

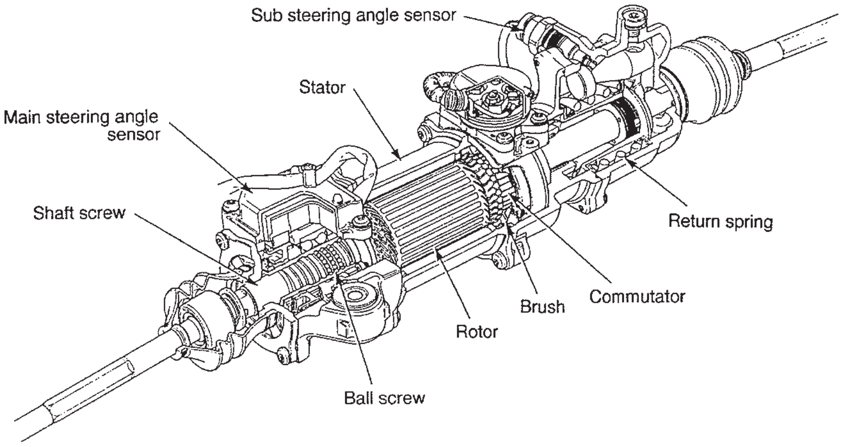


FIGURE 18.46 Rear steering actuator.

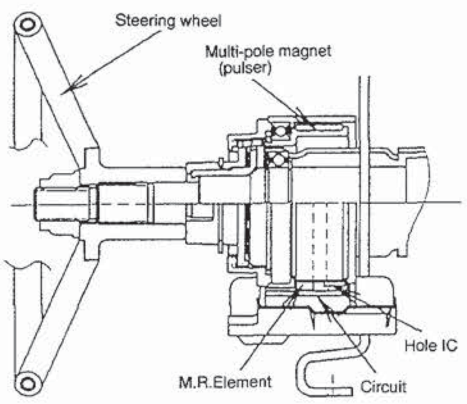


FIGURE 18.47 Main steering angle sensor.

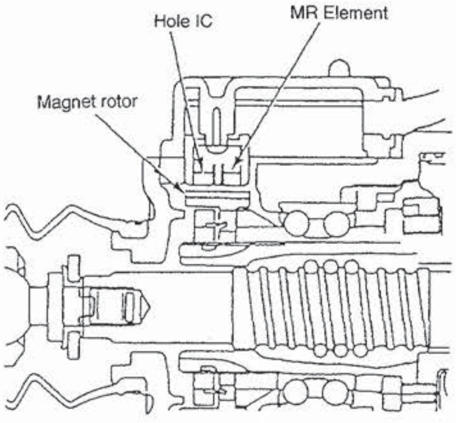


FIGURE 18.48 Rear steering angle sensor (main).

vehicle speed/front-wheel steering angle/steering wheel velocity responsive type system mentioned before, but by using an electric actuator, the space taken up by the system has been reduced and ease of installation improved.

Figure 18.49 shows the system structure, and Fig. 18.50 shows the steering angle generation modes for the rear wheels.

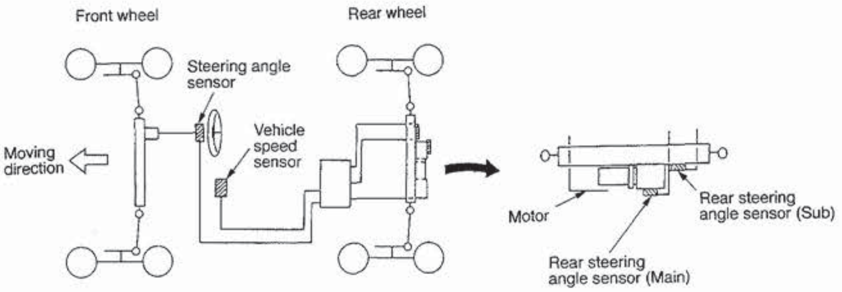


FIGURE 18.49 System construction of HICAS.[12]

18.2.5 Trends in Research and Development

Four-wheel steering systems that have been installed in current vehicles mainly adopt the program control technique in which the rear-wheel steering angle is programmably determined based on the beforehand scheduling relationship with the vehicle speed, the front steering angle, and steering wheel velocity. The aim of the system is the improvement of the insufficient performances in handling characteristics of the vehicle under special steering situations.

Recently, however, the system based on the new concept is being earnestly researched in order to drastically improve the handling characteristics in any region for practical use. Such technology is similar to that of CCV (control configured vehicle), which is used in airplanes, and named as *active control technology* in general. In this section, a basic introduction to the following two representative concepts concerning the active four-wheel steering system (A-4WS), which adopts the active control technology in the 4WS systems, will be given.

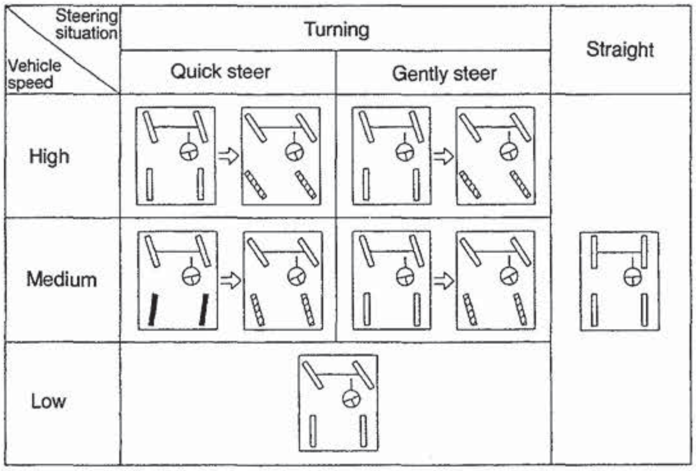


FIGURE 18.50 Rear steering angle generation mode.[12]

Vehicle's Slip Angle Control System. Generally, the moving direction of the center of gravity of the vehicle is not the same as the heading direction of the vehicle while the vehicle is turning, and the angle between these two directions is called the *side slip angle of the vehicle.* When the side slip angle of the vehicle is large, the driver is forced to drive with an oblique line of vision. This can be considered one cause of increased driving difficulties. From this point of view, the control concept has been proposed whereby the side slip angle of the vehicle while driving is always kept close to 0. In addition, the cause of the increasing of phase delay in lateral acceleration according to the vehicle speed is that the steady state gain in the sideslip angle is reduced together with the vehicle speed and becomes negative at high speed. In order to eliminate such phenomena, the concept which keeps the sideslip angle close to 0 may also be desirable.

The sideslip angle B can be expressed in general by the following equation:

$$\beta = \frac{1}{POL(s)}\{B_f(1 + \tau_f \cdot s)\delta_f + B_r(1 + \tau \cdot s)\delta_r\} \tag{18.9}$$

$$B_f\colon \frac{2l_r l k_r - l_f m v^2}{2lk_r V} \cdot A, B_r\colon \frac{2l_f l k_f - l_r m v^2}{2lk_f V} \cdot A$$

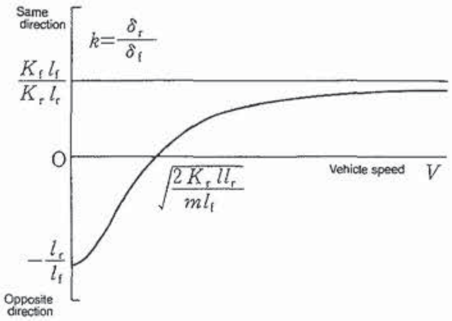


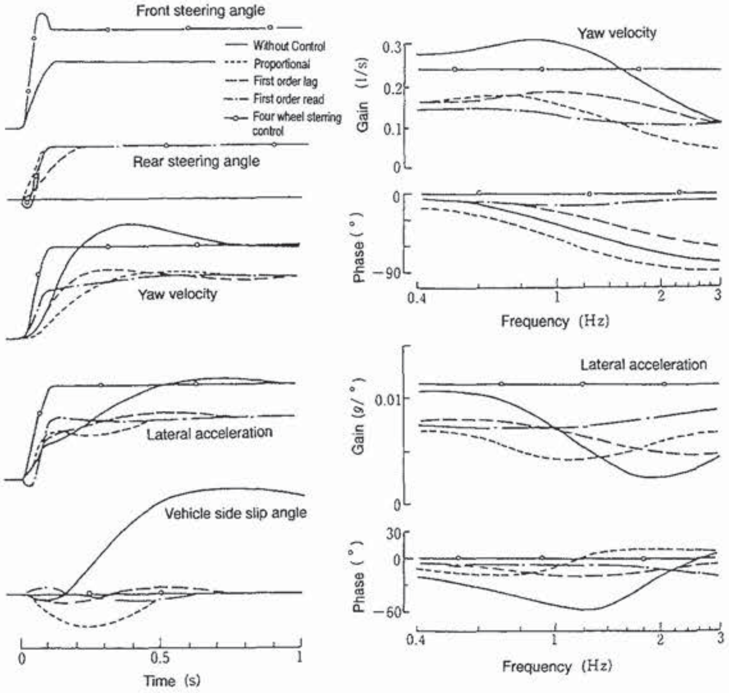
FIGURE 18.51 Steering angle ratio for making the vehicle sideslip angle equal to 0.[1]

Provided the numerator *s* in Eq. (18.9) = 0, the steering angle ratio between the front and rear wheels that is required for realizing a sideslip angle of 0 can be obtained from Eq. (18.10) as a function of the vehicle speed. This is shown graphically in Fig. 18.51.

$$\frac{\delta_r}{\delta^f} = -\frac{B_f - B_f \tau_{fs}}{{}_B r + B_r \tau_r s} \tag{18.10}$$

$$= \frac{-\iota_{f+} + (\iota_f m/2k_f\iota)V^2 - (I_z/2K_r\iota)V \cdot s}{\iota_{f+} + (\iota_r m/2k_f\iota)V^2 - (I_z/2K_f\iota)V \cdot s}$$

If the front- and rear-wheel control principle that includes the phase proceeding func-



Response Characteristics for Stepping State Steering Input

Response Characteristics for Steering Input Frequency

FIGURE 18.52 Affections in steering response characteristics by using the delay/advanced control means.[1]

tion for the front-wheel steering angle is employed, the vehicle characteristics can be set more freely. An example of the simulation analysis when the front and rear wheels are steered with a control principle such as the above is shown in Fig. 18.52.

Model Following Control System. In this system, the desirable vehicle behavior in accordance with the driver's steering operations is predetermined as a reference model, and the control is carried out so that the actual vehicle behavior fits this model. Generally, the lateral acceleration and the yaw velocity are used as the parameters for expressing the lateral movement of the vehicle. If the independent models are set for both of these characteristics, and make the actual vehicle behavior follow them, it is necessary to actively control both of the front and rear wheels. Research into systems such as this is currently being carried out, but in this section, a system for controlling the steering angle of the rear wheels that uses a range model for D^* which is defined by the linear combined Eq. (18.11) for the yaw velocity and the lateral acceleration will be introduced.

$$D^* = da_y + (1 - d)V\phi \tag{18.11}$$

If the weighted constant d in Eq. (18.11) equals 0, this control system becomes a yaw velocity model following control system, and if $d = 1$, then it is a lateral acceleration following control system.

GLOSSARY

Convergence Yaw stability of the vehicle when driver inputs a rapid steering wheel movement and releases the steering wheel.

Kickback Steering torque and angle arising from inverse input through the tires from uneven concrete road or other surface.

REFERENCES

1. "Steering," chap. 8, *Handbook of Society of Automotive Engineers of Japan* vol. 2 (Design), Society of Automotive Engineers of Japan, Inc., 1991.
2. Nissan Ciema, guidebook in Japan, 1991.
3. Shigeru Iga, et al., "Motor driven power steering for the maximum steering sensation in every driving situation," SAE paper 880705, 1988.
4. *Honda Today* service manual in Japan, 1991.
5. Acura NSX service manual, 1991.
6. Yasuo Shimizu, et al., "Development of electric power steering," SAE paper 910014.
7. Syunichi Wada, et al., "Electric power steering system," Mitsubishi Denk Giho, vol. 66, no. 9, 1992.
8. Yasuo Shimizu, et al., "Control of electric power steering (EPS)," *The 68th JSME Spring Annual Meeting,* vol. C 1605, 1991.
9. Naohiro Yuhara, et al., "Improvement of vehicle handling quality through active control of steering reaction torque," *International Symposium on Advanced Vehicle Control 1992,* no. 923073.
10. *Toyota Technical Review,* vol. 41, no. 1, 1989.
11. Mazda Motor Company Introduction to Capella (new car), 89-5-NM005, 1989.
12. Nissan Motor Company Introduction to Laurel (new car), F005705, 1989.

ABOUT THE AUTHOR

Makoto Sato, executive chief engineer of Honda R&D Co., Ltd., was graduated from Nagoya Institute of Technology in 1960 and began his career with Kokusan Electric Co. He moved to Honda R&D Co., Ltd. in 1968 and began working on vehicle dynamics control systems with electronic technologies. He is now researching new safety technologies with intelligent electronics, especially in the area of crash avoidance.

CHAPTER 19

LIGHTING, WIPERS, AIR CONDITIONING/HEATING

Richard Valentine
Motorola Inc.

19.1 LIGHTING CONTROLS

Controlling lamps with power electronics offers many advantages and a few disadvantages over conventional switches or relays. An important advantage includes easier diagnostics compared to the classical mechanical switch or relay approach, while a significant disadvantage includes the higher cost of the electronics. Because some lights, such as headlamps, turn signals, and brake lamps are safety related, a method to test the integrity of these lamps is an advantageous feature. The power electronic design can not only turn lights on or off, but it can vary the light's intensity and detect abnormal conditions such as open or shorted lamps. The cost tradeoff issue becomes more interesting when the power electronics load control is connected onto a data bus or multiplexed network (see Chap. 26).

The typical automotive lamp can range from small-wattage panel lamps to large 60-W or higher headlamps (refer to SAE ref HS-34/93 for auto lamp standards). Tungsten and halogen types prevail for many vehicular lamp designs. The prime difference between standard tungsten nonhalogen and halogen lamps is in light efficiency. Halogen types produce about 20 percent more light for each watt of energy consumed. Other types of lamps, such as light-emitting diodes (LEDs), fluorescents, or gas discharge, are somewhat adaptable for vehicular applications. The power electronics operating requirements vary for each lamp technology category.

19.1.1 Incandescent/Halogen

The single most important rule when switching normal incandescent lamps is to select a power transistor that can sustain a 10 to 15 times inrush current level. For example, a 2-A-rated stop lamp requires a 20- to 30-A-rated power transistor for an electronic switch design. When designing a lamp-driven circuit, other considerations besides inrush currents must be taken into account: maximum power transistor operating temperatures, shorted load protection, voltage supply transients, reverse voltage supply condition, open load detection, high or low side design topology, MCU interface, switching edge speed limits, and variable brightness control. In addition, the electronic lamp switch should not drop more than 0.3 V to ensure that the lamp's intensity is not degraded. Incandescent lamp light output is very sensitive to the lamp's operating voltage. A 20 percent drop in an incandescent lamp's nominal operating voltage can reduce its light output by 50 percent.

Inrush Current Effects. Lamp loads behave like high positive temperature coefficient resistors. When the lamp is cold, the resistance value of its filament is about 1/10 of its normal operating resistance. Lab tests show a definite relationship between the lamp's peak inrush current level and light output intensity with respect to time. If the lamp current is limited to its nominal value, it takes longer—one-half second or more—for the filament to reach its nominal temperature, but if the current supply is unlimited, the lamp heats up very quickly. The lamp response time and the available supply current must be considered for applications when an instantaneous light output is desired, such as with turn signals or brake lights in a vehicle. A 1-s delay at 89 km/h (55m/h) equals 24.6 m (81 ft). Four seconds of time is about how long some vehicles take to stop at a vehicle speed of 89 km/h. If the driver has exerted maximum braking, and his stop lamp waits for 1 s to convey this action to trailing vehicles, the driver will almost be halfway stopped before the trailing vehicle has a stop light indication. This is not acceptable, and the delay time in stop or possibly in turn signal indicators must be minimized to benefit the reaction time for other drivers.

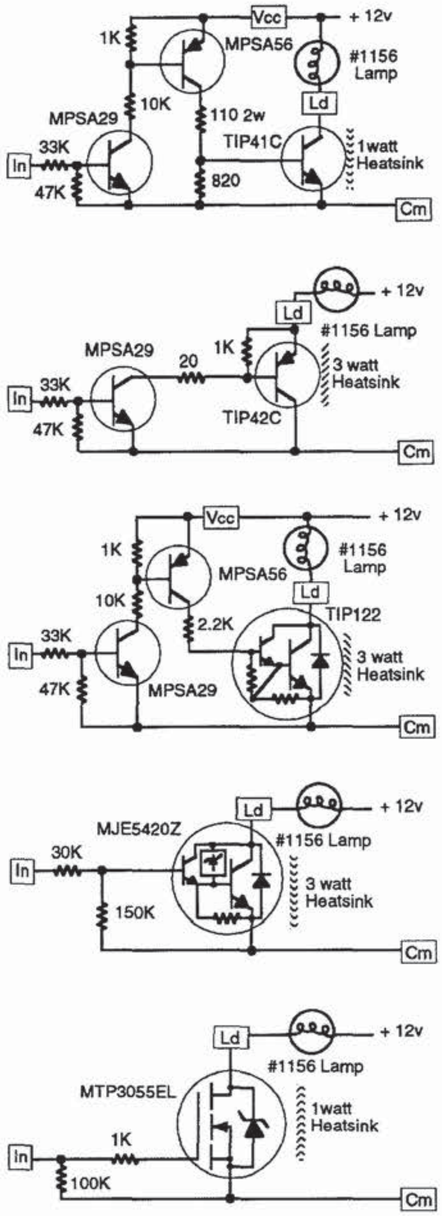
Limiting the lamp's inrush current will probably increase the lamp's longevity in high duty cycle applications, such as a flasher, because the mechanical stress on the filament is reduced. A limited inrush current design should decrease the chances of the lamp failing during the initial turn on. In vehicular applications when external induced filament vibration is high, the lamp reliability may not be improved enough to justify a current limiting design. A current limiting design does increase the power transistor's peak heat dissipation, and when operating with an energy source that is capable of supplying high peak currents, such as an automotive battery, the rationale to include current limiting would be to use a smaller, less expensive transistor or to minimize voltage sags in the wire bus when several lamps are switched on at the same time. A current limiting design may include a network to control the switching transition times from OFF to ON to OFF.

Setting the switching times to over 0.001 s would help minimize EMI effects. Most power transistors can switch in less than 1 μs, with some power FETs able to switch in 50 ns. These RF switching speeds are undesirable and will be detectable by other electronic systems in the vehicle, such as the AM radio.

Lamp Driver Design. The power electronics can drive the lamp in one of two simple ways. One is *low side switching* (Fig. 19.1) with the lamp connected between the power supply and the transistor, or *high side switching* (Fig. 19.2) with the lamp connected between the transistor and common. The main drawback to low side switching is that one side of the lamp is always connected directly to the battery supply, and if this lead is shorted to common, very high currents will occur. A high side switch disconnects the power supply from the load, so if the switch is off, the chances of an accidental power supply short when replacing or troubleshooting the lamp are minimal. On the other hand, in the high side switch design, a switch ON condition will cause very high current levels to flow through the power transistor in the event of a short. The selection between high or low side switching is affected by the lamp's environment.

In a low side design, leakage currents that may flow from the lamp's socket or connectors can be a concern. This is especially true if the lamp is operated in a salt-laden, high-moisture atmosphere. This problem occurs because of the constant voltage potential at the socket and is one reason why most exterior vehicular lighting systems use high side switching. Other applications, such as instrument backlighting or signal indicators that are usually located inside the vehicle, can be a low side design to take advantage of lower-cost NPN transistors or N-channel power FETs. N-channel power FETs can be used for high side switching by adding a charge pump as shown in Fig. 19.2*a*. Special power linear devices are also available that incorporate a charge pump for high side switching of small lamps. These power linear IC devices (MC3399, MC33091, or equivalent) will find more usage as their pricing and reliability become more attractive for automotive application.

Power Control Device Selection. When the drive topology has been decided upon, a power device can be chosen. There are numerous tradeoffs between bipolar and power FET transistors. The N-channel power FET device requires minimal drive power for a low side switch and



The NPN transistor runs in full saturation mode, but requires 1.5 watt base drive resistor.

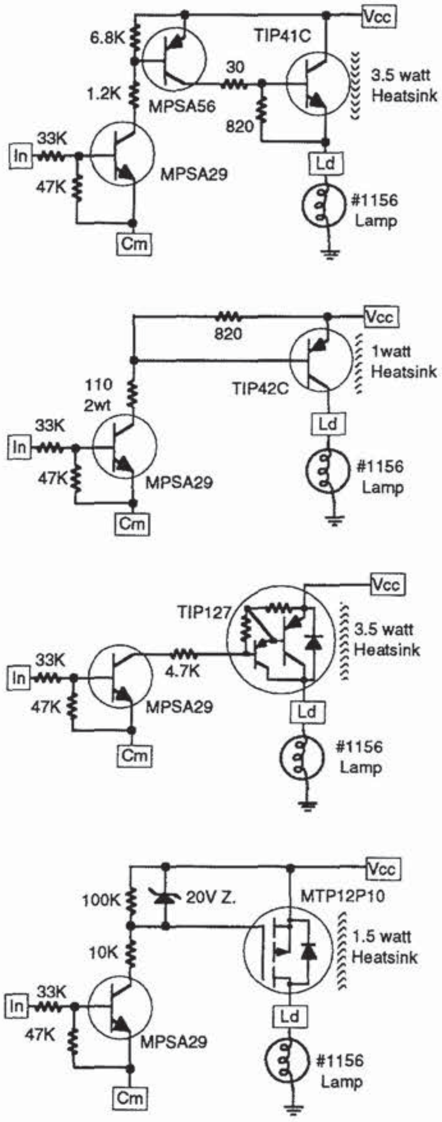
The PNP transistor dissipates about 3 watts since it does not operate in saturation mode.

NPN Darlington transistor dissipates about 3 watts since it does not operate in saturation mode.

NPN Darlington transistor has internal active zener element to clamp over voltage transients in forward SOA mode. Device dissipates about 3 watts since it does not operate in saturation mode.

N Channel MOSFET transistor dissipates about .8 watts, and can sustain low energy voltage spikes.

FIGURE 19.1 Low side switching.



The NPN transistor dissipates about 3 watts since it does not operate in saturation mode. Circuit requires two predrivers.

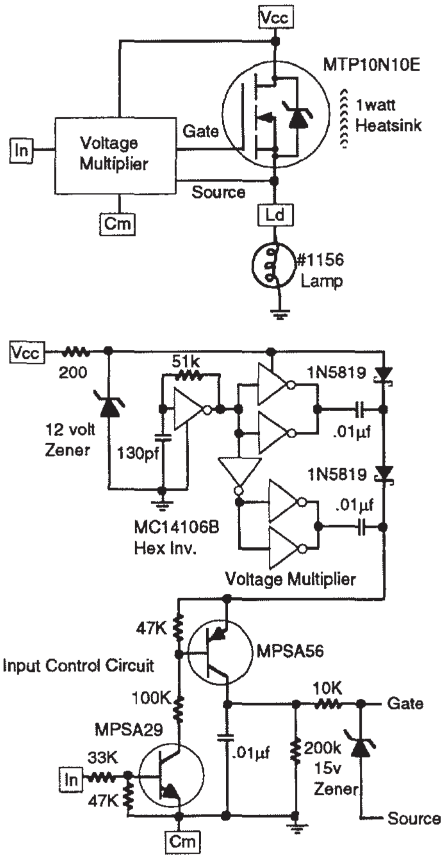
A PNP transistor can run in saturation mode, but dissipates about 1.5 watts in base drive resistor.

PNP Darlington requires small base current, but can't operate in saturation mode, and dissipates about 3 watts.

P Channel MOSFET offers best electrical performance, but the internal silicon die has to be twice as large as similar rated N-Channel MOSFET.

FIGURE 19.2 High side switching.

is cost-competitive with similar rated bipolars. Low side switching can easily use N-power FETs, whereas a high side switch would need more expensive P power FETs. Power FET transistors will usually be able to conduct the high peak inrush currents much more readily than bipolar transistors. This is because power FETs have higher transconductance than bipolars. Bipolars are usually biased for operation at the lamp's average load current and will therefore operate in a linear mode during the lamp's initial turn on, thereby offering some current-limiting at the expense of increased peak junction temperature. The power FET will usually con-



The N-Channel MOSFET requires a gate voltage bias above the source voltage. This voltage can be supplied from a separate power supply or from a voltage multiplier.

Voltage multiplier allows N-Channel gate to be biased above source voltage. A Schmidt Invertor is used as an oscillator to generate a square wave that charges capacitors with 180° out of phase signals. The capacitor charging voltage adds to the rectifier DC voltage, and totals up to 3x the Vcc bus minus rectifier drops and switching losses. Several N-Channel MOSFETs could be driven with additional input control circuits from one charge pump circuit.

FIGURE 19.2*a* N FET high side design.

duct many times its average current without going into a linear mode, provided the overcurrent condition time period is fairly short, less than 100 ms.

Lamp driver long-term reliability is determined mainly by the power transistor's heat dissipation and somewhat by the solder joint totals. Selecting P-power FET devices for the high side configuration because of the low number of parts, ease of drive interface, and low power dissipation appears best, but P-channel FET devices are usually 50 percent more costly. The N-channel devices win both for cost and reliability in a low side switch design. Table 19.1 summarizes the various lamp driver designs.

Short-Circuit Protection. Another consideration is shorted circuits. The bipolar transistor is typically biased for minimum base drive power loss. During a shorted load condition, the bipolar will pull out of saturation, enter into a linear operation mode, and may fail due to excessive junction temperatures, ordinarily before a fuse will blow. A properly sized power FET may be able to safely conduct enough shorted load current to blow the fusing element. Power FET devices of 50 V are available that exhibit less than 0.01 ohm forward ON resistance in a fingernail-sized surface-mount package such as a MTB75N05HD or the equivalent.

TABLE 19.1 Lamp Switch Design Comparison

Device type	Driver stage loss, W	Output device loss, W	Pwr. eff. in/out, %	Relative system cost	Parts count total, pcs	Solder joints total, pts	Relative reliability
			High side switch configurations for 2.1-A lamp load				
NPN	0.2	2.7	89	high	9	25	low
PNP	1.31	0.8	92	medium	6	18	medium
Darlington PNP	0.03	2.7	89	high	5	13	low
P-channel MOSFET	0.01	0.9	96	high	7	19	medium
N-channel MOSFET	0.05	0.7	97	highest	19	54	low
			Low side switch configurations for 2.1-A lamp load				
NPN	1.32	0.8	91	medium	9	23	medium
PNP	0.13	3.2	88	high	5	15	low
NPN Darlington	0.08	2.7	90	high	8	23	low
NPN Darlington w/Zener	0.0004	2.7	90	low	3	10	medium
N-channel MOSFET	0.0002	0.7	97	low	3	10	high

Power supply voltage = 12 V

This type of device could safely conduct 14 A continuously (assuming a 50 °C ambient temperature and 2 W constant power dissipation, $[I = (P/R)^{1/2}]$.

Another way of protecting a transistor against excessive current levels is to monitor the current value by the insertion of a low-value power resistor in series with the load. A comparator is used to detect the current sense resistor's voltage drop and will toggle if this voltage drop exceeds the comparator's reference voltage. The comparator output is tied to an R-S flip-flop logic network that latches off the input signal. Some power FET devices are now manufactured with an internal current mirror element and eliminate the need for an external large power resistor. A basic circuit is shown in Fig. 19.3 for an internal current-sensing power

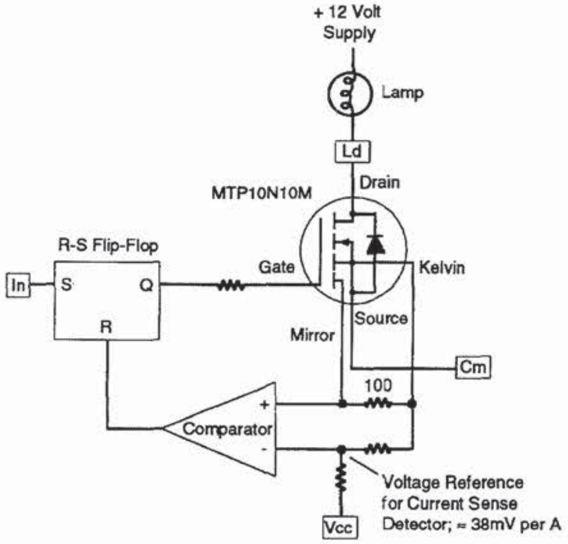


FIGURE 19.3 Internal current-sensing FET lamp drive.

FET lamp driver. This circuit uses an RS flip-flop that will shut off the gate voltage. When the RS flip-flop has been toggled by the comparator, the RS input (set line) has to be retoggled to turn on the current-sensing power FET again. When an overcurrent condition occurs, the current-sensing power FET is latched off until the input is retoggled.

The transistor's voltage rating will be determined by the maximum voltage transients on the power supply line. In vehicular battery-charging systems, the normal 12-V supply can reach over 85 V due to intermittent battery connections when the alternator is operating. If the transistor is not avalanche-rated and enters a second breakdown due to excessive power supply voltages, it will fail shorted. Higher-voltage systems require much higher-voltage-rated devices.

MCU-based Lamp Driver. The current-sensing power FET can be directly interfaced to a single-chip microcomputer. The use of a single-chip microcomputer can allow the switch control circuit to perform many tasks: open load detection, shorted load protection, variable intensity, and diagnostics. The MCU concept is especially attractive when several lamps are clustered together and driven from the same module. Figure 19.4 shows the use of a single-

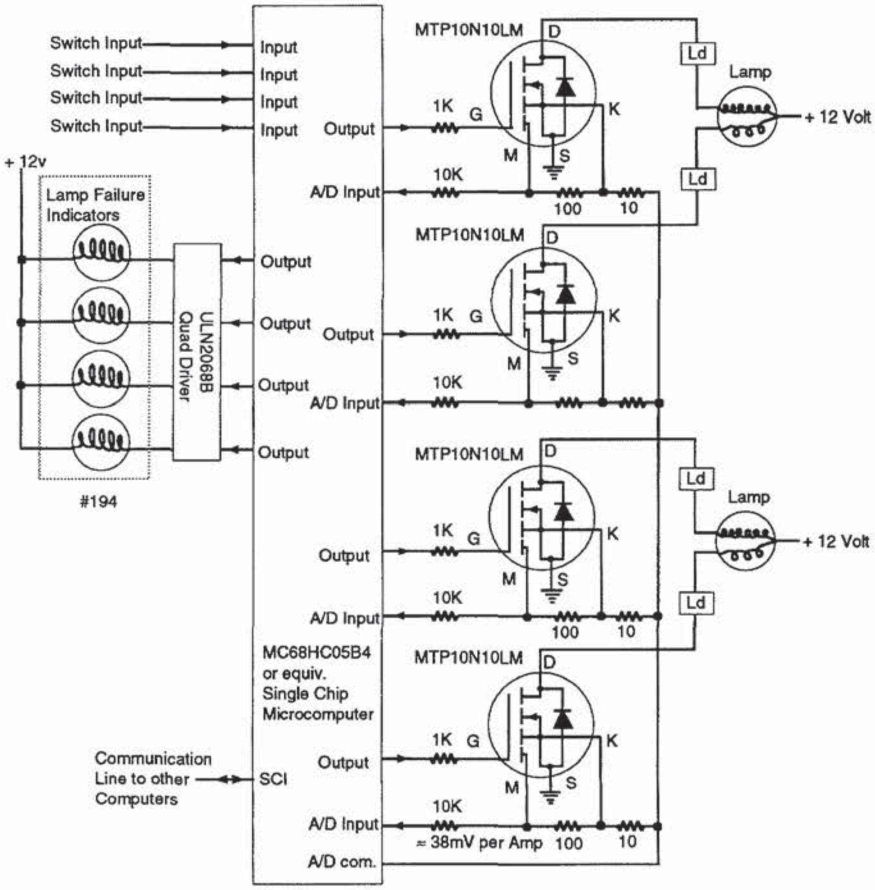


FIGURE 19.4 MCU lamp control.

chip MCU for controlling four lamp filaments. Many more lamps could easily be added to this design. Low side switching is used to accommodate the use of logic level N-channel current-sensing power FETs. The MCU's program is set to sustain a shorted load condition for less than 50 ms; this allows a cold lamp turn-on condition to be tolerated. If a shorted load occurs after the initial turn on, the program can shut off the load in less than 1 ms, wait for 1 s, and then try to renenergize the lamp. A diagnostic fault lamp is energized to alert the user that a fault has occurred. An open load can be detected when the lamp is energized by having the program monitor the A/D inputs for a zero value. An 8-bit A/D converter will resolve to about 0.02 V when referenced from a 5-V supply. Therefore, a normal load current of 2 A will generate about a 0.084-V level, which will be read as a 04-hex number by the program. A value of 00 to 01 H would be considered an open, and a value of greater than 06 H would constitute a shorted load. A distinction could be made in the lamp failure indicator between an open and shorted load by turning the failure indicator full on for an open, and then flashing it at a 1-Hz rate for a shorted condition. The failure indicators are 3-W lamps for bright visibility, such as is required in vehicular dashboards. High-efficiency LEDs could be used when the 3-W lamp's brilliance is not required, and the LEDs may be directly driven from the MCU's output ports when connected in a low side switch configuration.

Variable light intensity can be obtained by programming a 100-Hz frequency with 0 to 100 percent pulse-width modulation. The 8-bit MCU can vary the pulse width by 256 steps, which would appear as an infinite intensity control to the human observer. One difficulty with PWM is when its duty cycle is producing very narrow pulses, such as at 1 percent or 99 percent. These narrow pulses may not allow the power stage to completely switch on or off. One method to minimize narrow PWM pulses is to force the PWM to jump from 0 percent to about 5 percent, and from 95 to 100 percent. This can easily be accomplished in the software design.

An external communication link can be established via the MCU's serial data port. This would allow the light control functions to be remotely accessible by another computer.

In summary, standard discrete power FET or bipolar transistors can effectively control incandescent lamp loads. Power FETs are preferred because of their low drive power requirement and high transconductance. The addition of an MCU enables the lamp control design to include many desirable features, such as shorted load protection, and minimizes the need for more advanced lamp driver semiconductors such as smartpower or power IC devices. MCUs are available with a built-in MUX port to communicate with other load control modules for remote operation of several lamps or other loads.

19.1.2 Fluorescent Lamps

A fluorescent lamp is commonly defined as a light source that produces light by the interaction of an arc with the gas mixture inside a glass vacuum tube. When this gas mixture is ionized by means of a high-voltage pulse from the tube's end terminals, its internal resistance decreases from megohms to several hundreds of ohms. Fluorescent lamps are typically 2 to 3 times more power efficient than standard incandescent lamps. The automotive lamp applications for fluorescents may include inside lighting, instrument back panel lighting, and other areas where an unfocused and discontinuous light spectrum, but highly efficient light source, can be utilized. The main drawback to fluorescent lamps is that they require a fairly high voltage to initiate the arc and a means to limit the lamp current once the arc is established. In other words, the fluorescent design will cost much more than a similar incandescent-sized light. This is an important consideration when operating the fluorescent lamp from a 12-V supply.

A 12-V powered fluorescent driver design requires an 8- to 16-V input to about 300-V output convertor with current limiting. These types of circuits usually operate at high frequencies (20 to 100 kHz) to minimize the size of passive components, such as transformers and capacitors, and to minimize audible noise. This type of circuit can generate significant EMI and RFI if not properly shielded and filtered. Figure 19.5 shows a block diagram of a 12-V fluorescent driver.

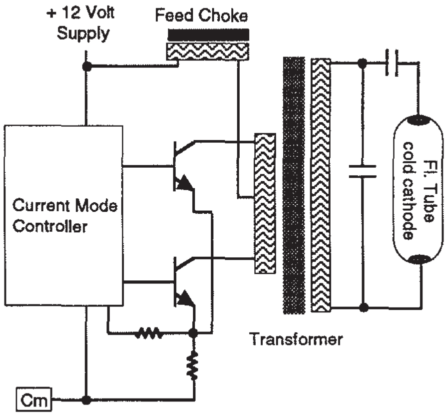


FIGURE 19.5 12-V fluorescent driver.

19.1.3 High-Intensity Gas Discharge Lamps

This class of lights includes mercury vapor, metal halide, and low- or high-pressure sodium. Some progress has been made to incorporate HID (high-intensity discharge) lamps in automotive applications. The HID metal halide type lamp offers high efficiency, up to four times more lumens per watt than incandescent, and high light output in a small package that can be utilized for headlamps. Again, the main drawback is the cost of the necessary power driver electronics. HID lamps require a circuit design similar to fluorescent drivers but at much higher power levels. The cost of the HID driver electronics limits their widespread use for automotive usage.

19.1.4 Electroluminescent (EL)

These low-intensity blue-green color light sources are used for decorative night lighting purposes and as backlighting for liquid crystal displays (LCD). EL lighting generally requires a low-current 100-V ac source. A 12-V dc to 100-V ac invertor is required for vehicular systems.

19.1.5 Light Emitting Diodes (LEDs)

Red LEDs may be adapted for stop light indicators. The LED produces a true red and does not require colored lenses. By combining several series LED strings together in one assembly, a stop light indicator is possible. A constant current source can be utilized to maintain a constant LED illumination level as shown in Fig. 19.6.

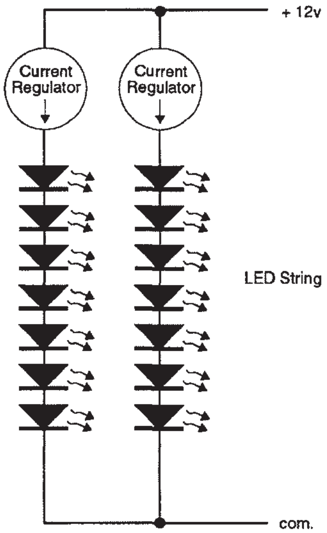


FIGURE 19.6 Constant LED illumination level design.

19.2 WINDSHIELD WIPER CONTROL

Windshield wiper (WW) systems are required by government regulations and therefore should be designed for high reliability. The motor control usually allows for low- and high-speed operation, or variable speed. The mechanical wiping motion is accomplished by a worm-gear design which gives torque multiplication and connects to the windshield wiper mechanical arm assembly. The key design specification is that the motor must sustain a "stalled" wiper condition without damage. The motor must also provide sufficient torque to run the wiper mechanism under worst-case conditions, such as blades frozen to the windshield. Electronic controls are used with WW systems, usually for intermittent operation plus variable speed. The driver adjusts a time interval control potentiometer for a one- to several-seconds wiper rate. The electronic circuit activates the motor with the appropriate time interval. A travel limit switch is normally used to insure that the wipe cycle is completed, and that the blades are returned to their nominal position.

Automatic windshield wiper designs are possible with today's technology that utilize a moisture detector to determine the degree of WW action required.

19.2.1 PM Motor Speed Control

The auto industry uses dc PM motors because they are economical to produce and ensure good performance. The motor's speed can be varied by either a simple voltage dropping passive resistor or an active linear voltage regulator. This simplistic method is widely used for vehicular motors requiring variable speed control. It does have a serious drawback in terms of power efficiency. For example, to control the speed of a motor that draws 20 A at full speed requires about 10 A at half speed. At full speed, the overall motor control system's efficiency will be around 80 percent. If the speed is reduced to half, the system's efficiency drops to 40 percent. This is because there is a heat loss of 70 W in the series voltage dropping element, in addition to the 14 W lost in the motor. A more efficient speed control system is therefore a desirable goal and can be accomplished by interrupting the motor's voltage at a variable duty cycle or by using a switching power supply. A switching power supply to control up to 300 W would cost more than the motor because of its high-frequency transformers and other components.

PWM Speed Control Design. Electronic permanent magnet motor speed designs can range from a single power transistor to turn the motor on or off to an "H" bridge microcontroller-based system with closed-loop speed control. The design of a unidirectional 280-W 20-A motor speed control will be examined that uses power FET semiconductor technologies.

Since the PM motor is a dynamic machine with the armature acting as a flywheel, the voltage interruption or chopping rate could be 1000 Hz or slower, before the motor's speed actually pulsates. A problem at 1000 Hz or other audible rates is noise generated from within the motor. This audible motor sound may not be tolerable for drivers or passengers. At higher frequencies—16 kHz or greater—the audible noise is minimized. Another noise issue is the significant electronic radio frequency interference (RFI) that radiates into the electrical system, including the vehicle's radio equipment. This RFI is generated by the fast switching edges of the PWM signal. The drawing in Fig. 19.7 illustrates slow and fast switching speeds and the resultant power transistor dissipation relationships. Slowing down the switching edges minimizes the radio frequency interference (RFI), but the switching speed parameters also play a crucial role in the overall controller system. High transitional speeds are desirable to minimize switching losses and to improve reliability, but very fast switching speeds become impractical beyond a certain point due to the inherent inductance in the wire and component leads. A compromise has to be made between the edge speeds and power device heat loss.

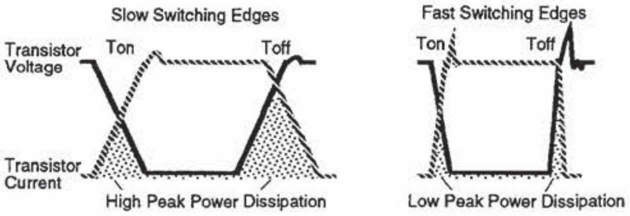


FIGURE 19.7 Slow vs. fast power loss.

Switching Speed Effects. Figure 19.8 graphs the switching edge speed effects upon the power transistor's power dissipation when switching a 20-A motor load from a 14.4-V power supply. Three different operating frequencies are plotted, one at 20 kHz, one at 16 kHz, and the last at 2 kHz. Note how the 2-kHz switching times are less critical than the higher 16-kHz and 20-kHz frequencies. In order to minimize audible noise, the 16- or 20-kHz frequency can be used, but, unfortunately, the 200- to 500-ns switching speeds required to minimize switching losses cause the intrinsic inductance in the hookup wiring, components, and motor to become critical. These faster edge speeds will generate significant voltage spikes across the hookup wiring and may lead to reliability problems unless precautions are made to minimize stray inductance in this wiring.

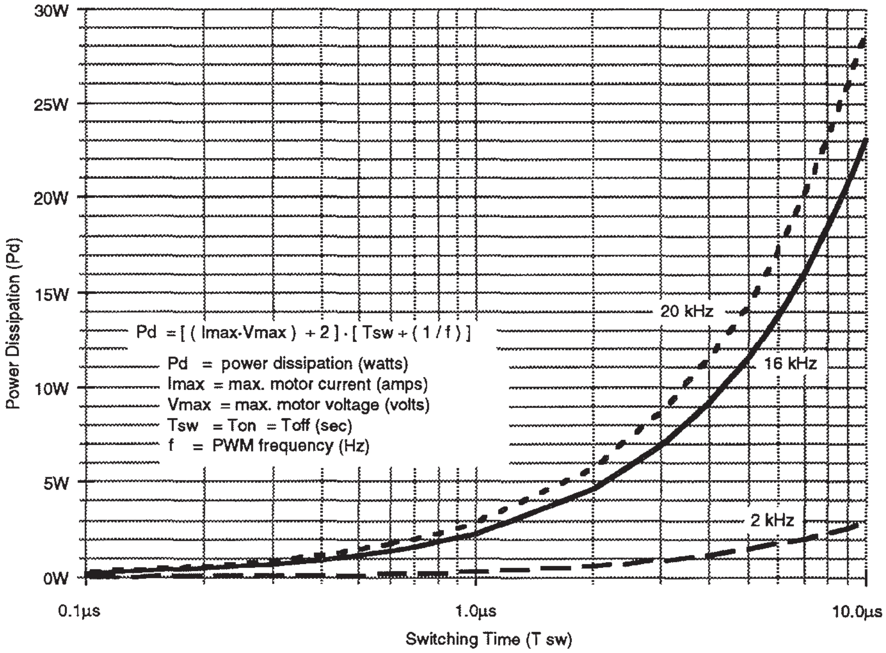


FIGURE 19.8 Power loss vs. switching speeds.

If the motor controller design has lead lengths that exceed 25 mm (1 in), their inductance values must be considered. An inductance value of 0.12 μH can be obtained in about 127 mm (5 in) of wire length with a #12 wire size, as shown in Fig. 19.9. This 0.12-μH value may not sound like much, but a plot (Fig. 19.10) of power line inductance versus switching transition times shows that just 0.1 μH of lead inductance can generate a 10-V voltage spike (at 20-A motor current) when switching edges are about 0.2 μs. Therefore, if fast switching edges are to be used, as with a 20-kHz operating frequency, a capacitor filter network is mandatory near the motor's power lead and controller. This filter network actually has to smooth out both the 20-kHz, and the much higher frequencies associated with the 200- to 1000-ns switching edges. To summarize, when fast switching edges are generated, filter networks are mandatory, or the hookup wiring will act as an antenna.

Motor Inductance Clamping. The kickback voltage from the motor's inductance also has to be contained, as the motor's inductance value will be dominant over the hookup wiring. This high-energy-laden voltage spike must be dealt with or the power transistor will enter an

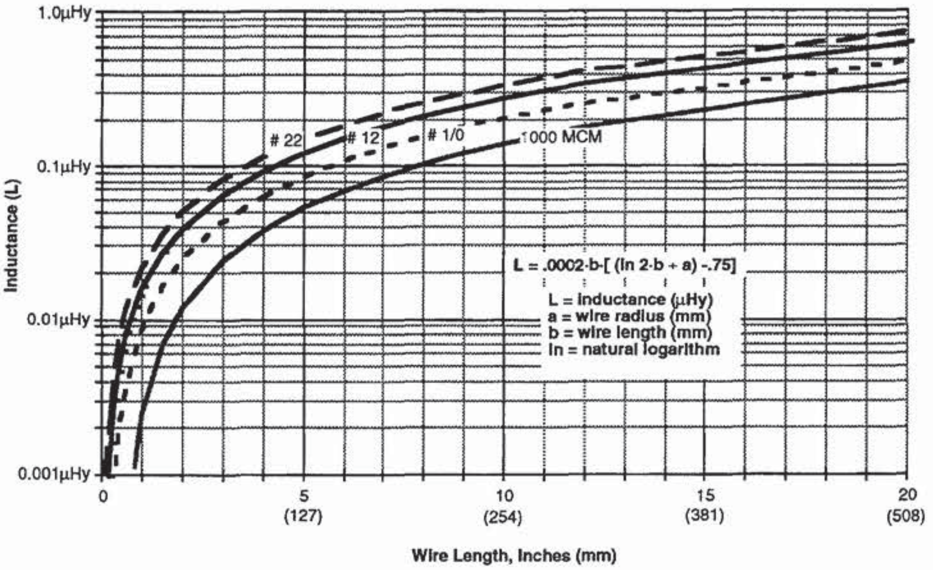


FIGURE 19.9 Line Z.

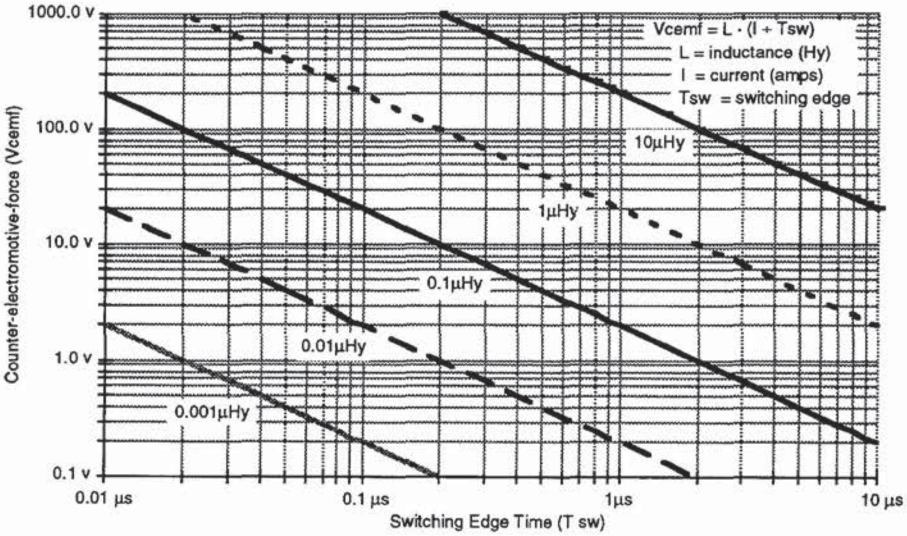


FIGURE 19.10 L di/dt effects.

avalanche or second breakdown mode, which is something to be avoided. The energy content of the motor's inductive kickback voltage is mainly determined by the motor's current, the motor's internal and external wire inductance value, and the rate at which the current is switched off. A high-current-rated freewheeling rectifier network can be connected across the controller's battery and motor terminals to clamp the motor's inductive kickback voltage spike. The reverse recovery times of these freewheeling rectifiers do affect the power transistor's switching performance.

When operating at a frequency of 20 kHz or a 50-µs repetition rate, the motor's inductance value is large enough to keep the freewheeling rectifier in a forward conduction mode until the next 50-µs cycle occurs. For example, if the PWM is set to a 25 percent duty cycle, the power transistor is on for the first 12.5 µs, and the freewheeling rectifier is in conduction for the remaining 37.5 µs. This means that when the next cycle occurs, the power transistor is switching on while the freewheeling rectifier is still in conduction. The result is that the power transistor and freewheeling rectifier conduct a high level "shoot-through" current spike. The magnitude and duration of this current spike are directly related to the rectifier's reverse recovery time rating, lead inductances, and the battery supply source impedance at the rectifier and power transistor location. Selecting freewheeling rectifiers with a "soft" reverse recovery characteristic will help minimize the shoot-through current problem.

The motor's back-emf, is produced from the rotating armature in a magnetic flux from the permanent magnets (PM). The PM motor's back-emf with a fixed mechanical load is normally lower than its power supply voltage and should therefore not exceed the voltage ratings of the power transistor.

Motor Current. When the motor's inductive generated transients are safely contained, one needs to deal with stalled and shorted motor conditions. Note that the motor's copper windings exhibit a positive temperature coefficient of 0.00393 per °C. Therefore, a 0.25-ohm motor resistance value at 25 °C room temperature would be about 0.18 ohm at –40 °C. Using the 20-A motor as the load, the maximum stalled or locked rotor current can be calculated as shown below.

$$I_{\text{max}} = E_{\text{max}} \div R_{\text{mtr}} = 14.4 \div 0.18 = 77\ \text{A}$$

where E_{max} = maximum power supply voltage
R_{mtr} = minimum motor resistance*

Power Transistor Specifications. When the maximum motor current has been established, the power transistor device specifications can be determined. In this case, the device needs an average current rating of at least 77 A, but the prime consideration for reliable power transistor operation is its worst-case heat dissipation.

The worst case scenario would include maximum values for the supply voltage, ambient temperature, and motor current. A maximum junction temperature of 150 °C for the power transistors is used as a maximum point. It should be noted that the reliability will increase by about one order of magnitude for each 10 °C drop in the power transistor's junction temperature. The following equations calculate the power transistor's maximum allowable heat dissipation for use in an 85 °C environment using a 2.7 °C/W heatsink unit and a 1 °C/W junction to case power FET thermal resistance.

$$\begin{aligned} PD_{\text{max}} &= (TJ_{\text{max}} - TA_{\text{max}}) \div (R\varnothing JC + R\varnothing CS + R\varnothing SA) \\ &= (150 - 85) \div (1 + .1 + 2.7) \\ &= 17.1\ \text{W} \end{aligned}$$

where TJ_{max} = maximum allowable junction temperature
TA_{max} = maximum ambient temperature
$R\varnothing JC$ = junction to case thermal resistance
$R\varnothing CS$ = case to heatsink interface thermal resistance
$R\varnothing SA$ = heatsink to ambient thermal resistance

---

* Formula for calculating the motor's cold resistance value is $R_{\text{mtrcold}} = [(\text{delta temp.} \times R_{\text{coef}})\ R_{\text{mtr}}] + R_{\text{mtr}} = [(-65 \times 0.00393) \times 0.25] + 0.25 = 0.18$.

After the maximum transistor power dissipation is known, the maximum forward voltage drop, VDS_{on}, and ON resistance can be calculated as shown:

$$\begin{aligned} VDS_{on} &= PD_{max} \div I_{max} \\ &= 17.1 \div 77 \\ &= 0.22\ \text{V at } 150\ ^\circ\text{C Tj} \\ RDS_{on} &= VdS_{on} \div I_{max} \\ &= 0.22 \div 77 \\ &= 0.003\ \text{ohm at } 150\ ^\circ\text{C Tj} \end{aligned}$$

In order to obtain a 0.003-ohm power FET ON resistance value at 150 °C Tj, several power FET devices will need to be paralleled. Bipolar devices were not even considered because of their high-current base drive requirements. Power FETs exhibit about a 75 percent ON resistance increase from a 25 to 150 °C rise in junction temperature. Therefore, six MTB75N05HD or equivalent devices in parallel would be required to achieve 0.003-ohm total ON resistance value when operating in an 85 °C ambient with a 2.7 °C/W heatsink. Six devices may not be economically attractive, so a larger heatsink may be more desirable. If the heatsink's thermal resistance is lowered to 1 °C/W, an ON resistance of 0.005 ohm would be required, which is three MTB75N05HD or equivalent paralleled devices. This is an example of the cost tradeoff between heatsink size and power FET ON resistance. It also points out that the power FETs ON resistance must be considered at the actual worst-case operating temperatures.

Short Current Protection. Some form of overcurrent protection would minimize power transistor failures in the event of a shorted motor, as might happen if the motor leads touched each other. Assuming motor's internal resistance will always present at least 90 percent of its nominal value, the shorted motor leads is the worst possible overcurrent condition. If three 75-A-rated power FETs are used in parallel, they would be capable of conducting well over 200 A. This current level may clear a 30-A fuse or activate a circuit breaker before the power FETs start to seriously overheat. In some wiring systems a *fuse link* is used to protect against a catastrophic wiring harness fire. The fuse link is usually just a short piece of copper wire that is encased in flame resistant insulation and is one wire-size smaller than the motor harness wire. The fuse link would probably not clear before the failed power transistors had shorted and blown open. If a fuse is selected that is slightly larger than the motor's normal run current, the power transistors may be capable of sustaining the shorted load current levels long enough to blow out the fuse before the transistors fail. A more reliable shorted load protective concept may still be necessary.

It may be desirable to latch off the power transistor during shorted load conditions. Load current sensing can be accomplished in three ways, as shown in Fig. 19.11. The voltage drop across the power transistor can be monitored, and, when a certain threshold is reached, the gate drive is latched off until the next operating cycle. Another method is to insert a series resistor in the source's path and monitor the voltage drop across the resistor. An internal current-sensing power FET device can also be substituted for one of the power FETs, with its current mirror voltage representing a portion of the total load current.

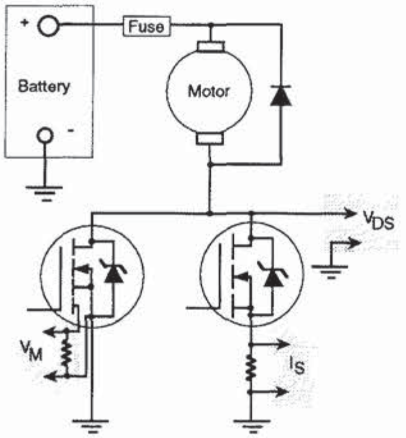


FIGURE 19.11 Load current sensing.

The method of measuring the transistor's forward ON voltage does have some inherent drawbacks. The drain-to-source measurement circuit has to be synchronized with the gate drive voltage, and the forward voltage will change by at least 2 to 1 at a nominal current level over a –40 to +175 °C junction temperature range. The temperature drift of 2 to 1 means that, at 175 °C, a 60-A overcurrent would read the same as 120 A at –40 °C. Therefore, at high temperatures, the overcurrent detector may falsely shut down with a stalled motor, and, at cold temperatures, the overcurrent may not shut down at all. The most accurate current measurement method is to use a series resistor added in the high current path.

A single-chip microcontroller could greatly enhance the operation of this design by allowing more functionality. For example, a stalled motor condition could be detected and a preset shutdown mode invoked to protect the motor against burnout. The MCU would include internal A/D for current detection directly from the current-sensing resistor or from the internal current-sensing power FET. The PWM signal would also be generated from the MCU. Figure 19.12 shows an MCU-based speed control conceptual design that uses an 8-bit single-chip microcomputer.

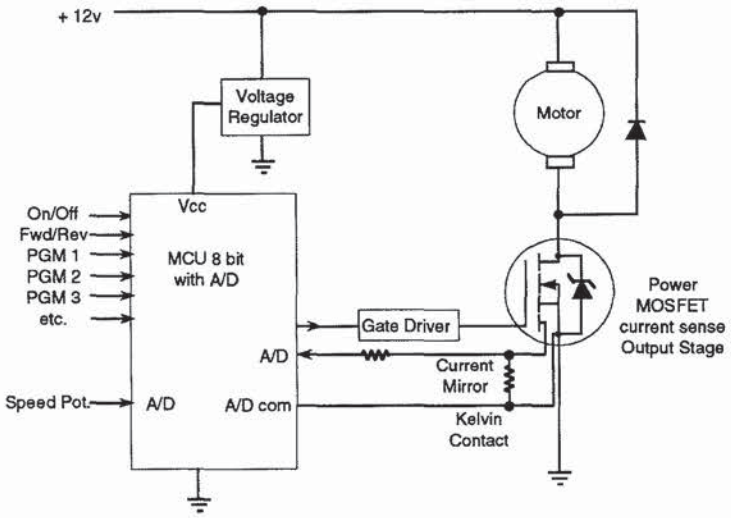


FIGURE 19.12 MCU motor control.

19.3 AIR CONDITIONER/HEATER CONTROL

Electronic controls allow an automatic climate control system to be designed for the comfort of the driver and passengers. An upscale automatic HVAC (heater and air conditioning) system, Fig. 19.13, may utilize an MCU to compute the most effective heat and ac flow rates. The heater and ac evaporator fan motor speed control designs can be similar to the WW variable speed design previously discussed if PM motors are selected. Brushless motors (BLMs) offer several features for this application. The BLM offers no brush-to-commutator noise. This feature alone is desirable in luxury vehicles, and, when one considers the increased mechanical reliability gained by the elimination of the brush-commutator assembly, the BLM cost-to-performance ratio seems more reasonable.

19.3.1 BLM System

If a motor-driven load requires closed-loop speed and direction control, the cost disadvantage of a BLM system becomes less significant. Figure 19.14 shows a typical permanent magnet brush motor compared with a BLM with both motors utilizing speed and directional control

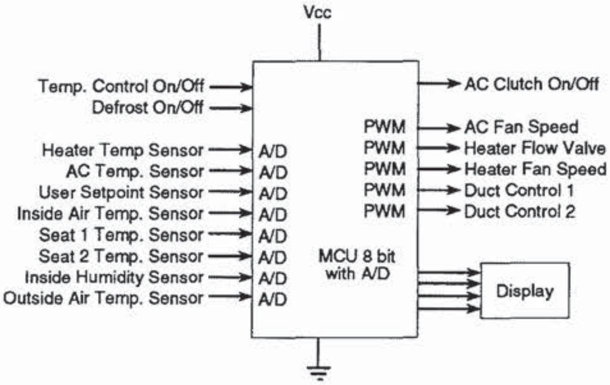


FIGURE 19.13 Climate control system.

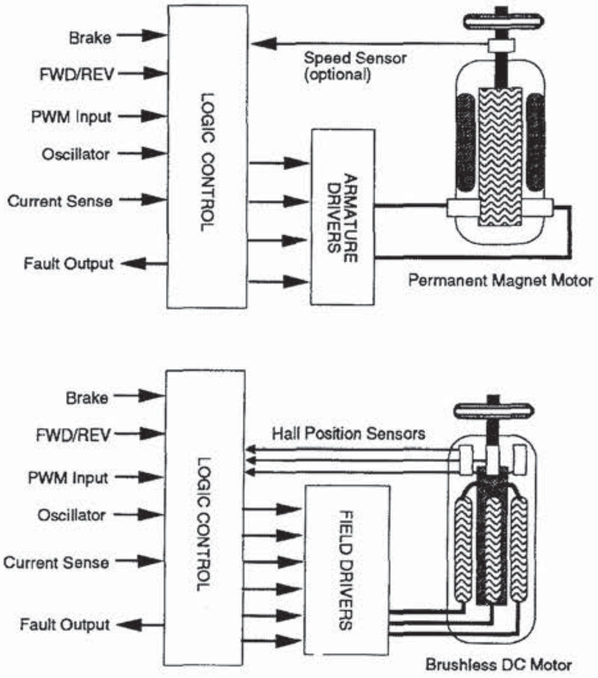


FIGURE 19.14 PM vs. BLM.

networks. Note that the BLM system uses only about one-third more drive electronics than the PM motor.

The BLM drive electronics may consist of a BLM linear bipolar integrated circuit, such as an MC33035 or similar device, that not only contains logic gates for commutation timing, but also includes internal driver stages capable of driving either power bipolar or power FET output transistors, or the MCU itself can be used to generate the necessary BLM control signals. One possible system may include the MC33035 linear bipolar IC and six power FETs. Three of the power FETs are internal current-sensing types. These current-sensing FET signals are conditioned and fed back to the MCU and to the linear IC as signals representing a motor normal condition or a motor overload condition. When the MCU detects a motor overload condition, a predetermined course of action takes place. First, the overload is counted as an event. The MCU timer allows the overload to continue for a length of time depending on how much time has elapsed from the previous overload. This is done to allow for the thermal heat buildup that will occur when the motor is overloaded or stalled on a repetitive basis, enabling the motor to have a high starting torque when cold, but to protect the windings, bearings, and drive electronics when the motor's temperature has become dangerously high due to repeated starts with a faulty load. The MCU's EEROM is used to remember user settings such as temperature preferences, as well as for interfacing into a multiplex data network.

19.3.2 Clutch Coil Control

The HVAC system also controls the ac clutch coil, plus other small duct motors or solenoids. Care must be taken when switching these highly inductive loads with power transistors. Figure 19.15 shows several possible voltage clamping networks to prevent damage to the power module when the inductive loads are switched off. The following formula should always be on hand and used when designing power-switching circuits that drive coils or solenoids.

$$V_{pk} = L\ (Di/Dt)$$

where V_{pk} = maximum inductive generated voltage
Di = coil's peak current level when switched off
Dt = switching time of the coil's current level

A typical air-conditioner clutch coil may exhibit 15 mH of inductance and 3 ohms resistance (or 4.8 A of current), and, when operating from a 14.4-V battery, would generate about 720 V when switched off in 100 µs ($V_{pk} = 0.015 \times (4.8/.0001)$). Needless to say, this would severely stress or destroy a typical 50-V-rated power transistor. Some form of voltage clamping is required.

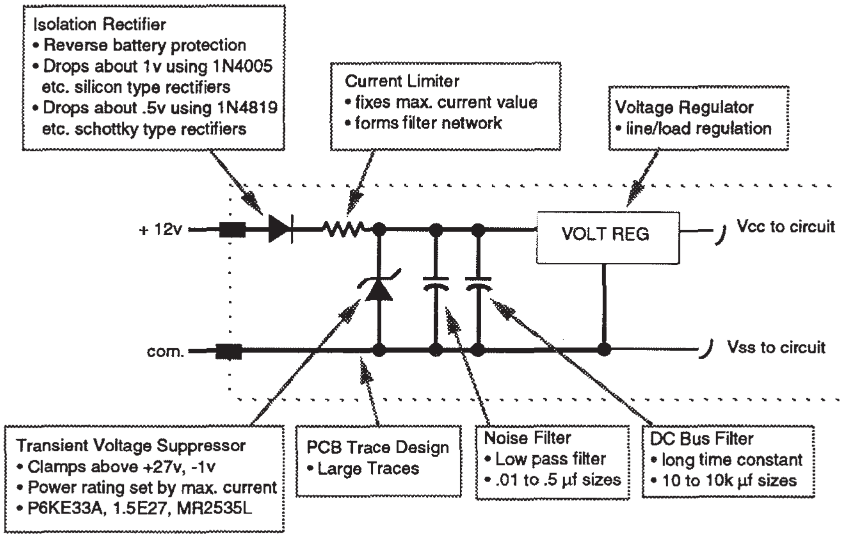
A fully automatic HVAC system can monitor the inside and outside air temperatures. The MCU program can monitor these inputs plus the user's temperature setpoint and then compute the heater and/or ac fan motor's speed plus heater core flow rate and/or ac clutch to achieve the desired inside temperature. The HVAC MCU software design can be implemented using fuzzy logic rules rather than standard control algorithms or massive lookup tables.

19.3.3 Fuzzy HVAC Control

Fuzzy logic design can be used to develop an HVAC control program. In general terms, a fuzzy logic design allows the software programming of an MCU-based controller to be simplified and is especially useful for complex nonlinear control dynamics. At this time, standard MCU devices can be used for fuzzy programming. Future MCU devices will probably be available that will enhance a fuzzy program. Programmers who can easily understand how to compose assembly language programs for converting various input signal conditions into the required output signals may not contemplate a fuzzy logic design, but, for many system designers, the fuzzy program method will shorten the software design and can allow the program design to be evaluated with the use of fuzzy development tools even before the hardware has been constructed.

A fuzzy logic implementation using an 8-, 16-, or 32-bit single-chip microcontroller (MC68HC11, MC68HC16, MC68332, etc.) requires the control system's behavior to be well defined from an intuitive sense. The designer defines the input conditions and desired output in a subjective manner. For example, if the auto's inside temperature is 25 °C, most people would probably find that value comfortable. In fuzzy terms, 25 would be 100 percent true for an input

ELECTRONIC MODULE PROTECTION



SIGNAL INPUT LINE PROTECTION

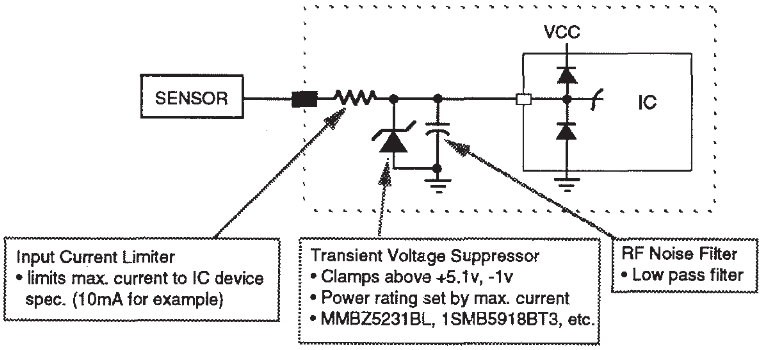


FIGURE 19.15 Voltage clamping networks.

membership called "comfortable," whereas 20 °C might be 0 percent and 30 °C might be 0 percent. Memberships could include "icycold," "cold," "cool," "comf.," "warm," and "hot." This thinking process fuzzifies the input temperature analog signal into numerical values that are attached to each membership. The user setpoint temperature is also divided into membership functions. The user setpoint temperature input is also fuzzified. Figure 19.16 shows the ambient and user setpoint memberships graphically. Other inputs, such as outside temperature, humidity, solar radiation, etc., could also be incorporated to produce a high-performance HVAC system. Once the input signals have been fuzzified, the first of three parts of a fuzzy design has been completed.

A second part of the fuzzy design is to assign membership functions as shown in Fig. 19.17 to the output control signals. In this example, only the heater side is shown. The output mem-

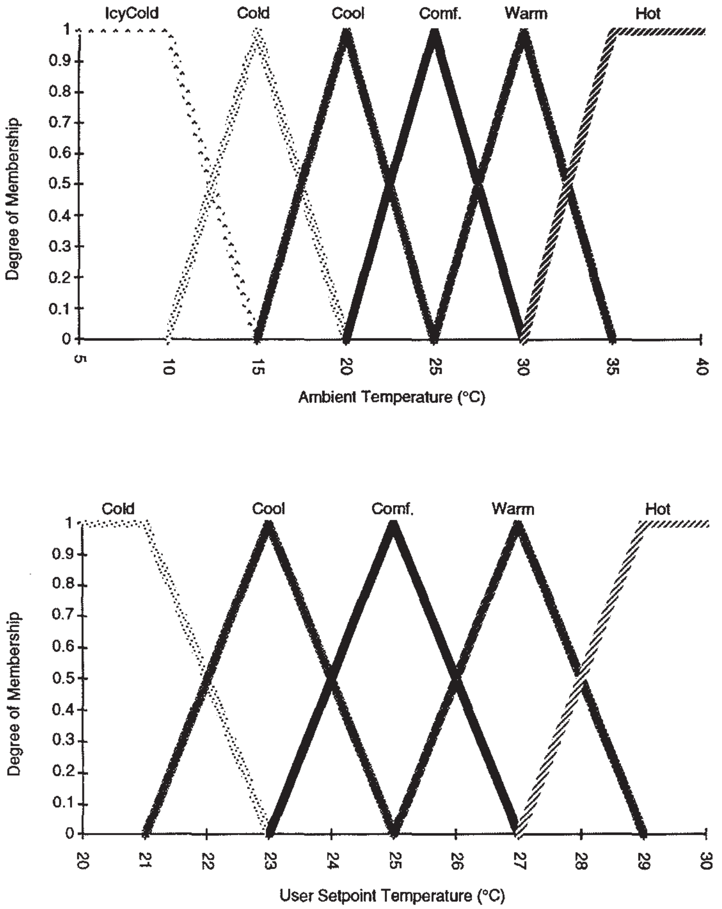


FIGURE 19.16 Fuzzy input memberships.

bership functions are classified as singletons, which only gives one value per membership. This approach trades precision for memory size.

The third part of a fuzzy design uses the min-max inference method to establish rules. A matrix representation of the input and output memberships and rules is helpful as shown in Fig. 19.18. The matrix size is determined by the input membership quantity, which is 5 for the user setpoint and 6 for the ambient temperature. This means a total of 30 rules are possible (5 × 6 = 30). The designer selects which output membership function occurs for each input membership function.

Fuzzy logic development tools are available that can simulate the fuzzy design. These programming design tools simplify testing the program's performance before production. A 3-D interactive graphic display allows the designer to visually examine the input to output relationships. Irregularities can be traced back and corrected before the first prototype is built. The availability of integrated fuzzy design software for personal computers allows the system engineer to evaluate the fuzzy design approach.

In summary, the fuzzy design method requires: the input data to be applied to the input membership functions or fuzzified, invoking the rule set to the fuzzified data, and, finally, generating an output or defuzzification based upon the rules and output function type. The fuzzy

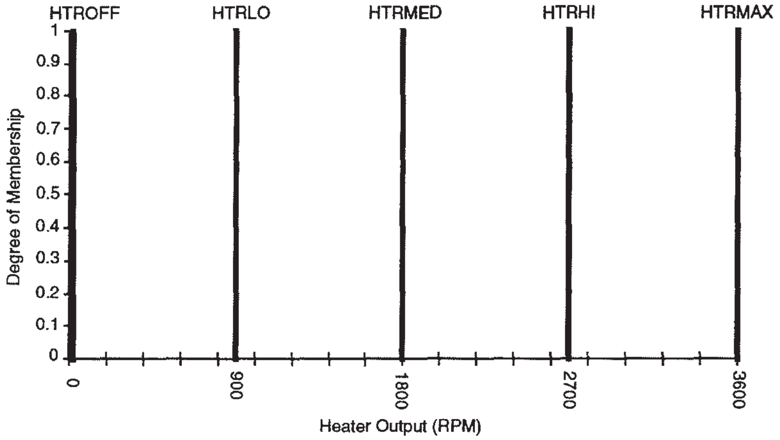


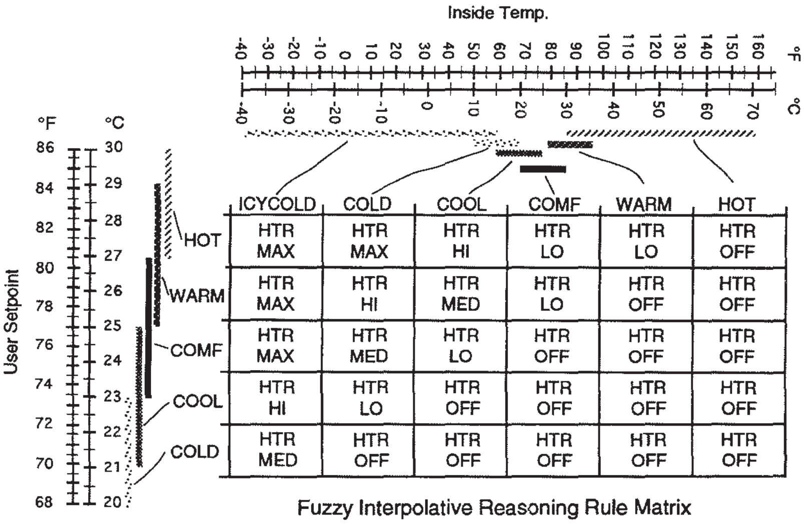
FIGURE 19.17 Fuzzy output memberships.

design can be modified with common-sense rules and rule strengths, whereas the conventional algorithm design method may require modifications at the most basic assembly code levels.

19.4 MISCELLANEOUS LOAD CONTROL REFERENCE

19.4.1 Current-Sensing Power FETs

Internal current-sensing power FETs can be used for current sensing, but their implementation may require some additional considerations. The current-sensing power FET device is a



	ICYCOLD	COLD	COOL	COMF	WARM	HOT
HOT	HTR MAX	HTR MAX	HTR HI	HTR LO	HTR LO	HTR OFF
WARM	HTR MAX	HTR HI	HTR MED	HTR LO	HTR OFF	HTR OFF
COMF	HTR MAX	HTR MED	HTR LO	HTR OFF	HTR OFF	HTR OFF
COOL	HTR HI	HTR LO	HTR OFF	HTR OFF	HTR OFF	HTR OFF
COLD	HTR MED	HTR OFF	HTR OFF	HTR OFF	HTR OFF	HTR OFF

Fuzzy Interpolative Reasoning Rule Matrix

FIGURE 19.18 Fuzzy rules.

normal power FET transistor with a very small portion of its active die area set aside for a current mirror. To use its current-sensing ability, an external resistor is connected between the current mirror output and the Kelvin output. The voltage drop across this resistor will be directly proportional to the device's current level when the device is operated in the active region. Choosing the correct value of the sense resistor is important to maintain an accurate drain current-to-sense voltage ratio over wide temperature ranges and when the unit is operated in the full ON or ohmic region.

When operating in the full ON region, the current mirror network changes from a constant current mode to a voltage source mode. This mode change occurs because the forward transconductance of the current mirror FET cells is decreasing due to the lower drain-to-source voltage. The result lowers the mirror current level to the sense resistor during full ON conditions when the load is connected between the drain and Vcc.

By selecting a sense resistance value that is less than 10 percent of the ohmic or rM(ON) value, the current mirror accuracy can be maintained. For example, if the current-sensing power FET device's rDS(ON) is 0.16 ohm and the current mirror ratio is 1800, the rM(ON) will be about 288 ohms ($0.16 \times 1800 = 288$). Therefore, a sense resistor value of less than 28 ohms is required. If a lower rDS(ON) current-sensing power FET device is used, for example, 0.026 ohm, and the current mirror ratio is 950, a sense resistor of less than 2.5 ohms should be used.

The use of an amplifier stage is still required to boost the current mirror's output signal level, which is similar to a traditional current-sensing network that uses very low resistance values (0.005 to 0.1 ohm). A low-cost operational amplifier (op-amp), LM324A, for example, will give fair performance over wide temperature ranges. The input offset voltage range of the op-amp will be a determining factor in the overall accuracy of the current-sensing power FET circuit design because of the low-millivolt current mirror signals. Instrument-grade op-amps can be used when higher accuracy is required. The op-amp circuit layout should be connected in close proximity to the current-sensing power FET's Kelvin and mirror output pins to minimize external noise pickup.

19.4.2 Overvoltage, "Load Dump," Reverse Battery

A hazard to any automotive electronics is reversed power supply connections. The power FET's internal rectifier will become forward-biased and will turn on the motor or lamp load. The FET's power dissipation will increase about 5 times normal since its voltage drop is now about 1 volt instead of the nominal 0.2 V. A standard bipolar transistor does not have the internal collector-to-emitter rectifier, but it still will suffer from a reversed power supply condition. The emitter-base junction avalanches at around 6 to 8 V, and the base-collector forms a rectifier. Therefore, the typical bipolar transistor is cooking itself into oblivion with about an 8-V drop while conducting about 1 A of lamp current. Most auto manufacturers specify that any electronic or electrical device must be able to withstand a reverse battery connection. The exact magnitude of the reverse voltage requirement varies per manufacturer, but the worst case seems to be –24 V for 10 min. This requirement is especially troublesome when the load requires an inductive kickback clamp rectifier, and either a power FET or bipolar transistor is used as the load driver. The power transistor's intrinsic rectifier and the load rectifier become forward-biased as shown in Fig. 19.19. The result? The clamp rectifier fails by shorting, the transistor shorts, and then the fuse may blow unless the rectifier or power transistor blows open first. A simple reverse battery isolator relay can be used to guard against incorrect battery polarity connections. This simple relay has been incorporated in the power supply line, which will open the line in the event of a power supply voltage reversal or if the power supply should exceed 32 V.

There are several other hazards for 12-V vehicular electronics that have to be considered. These can be divided into two groups: minimum/maximum operational voltages and voltage transients.

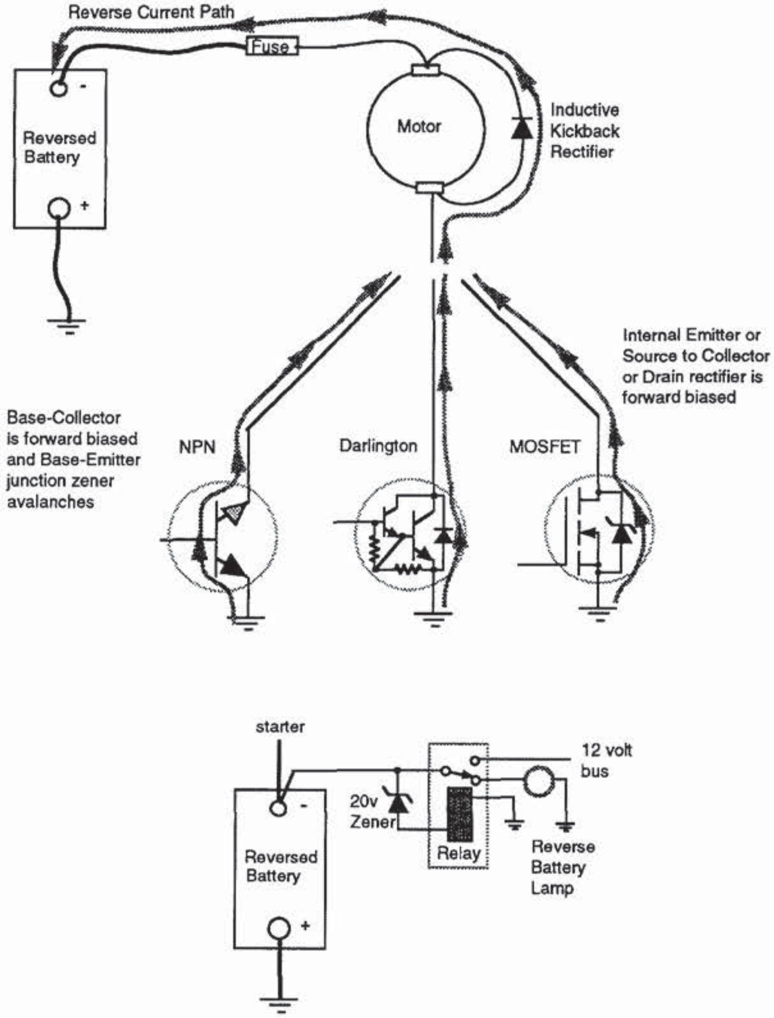


FIGURE 19.19 Reverse battery rectifier path.

The typical power supply voltage operating range for nonengine functions is 8 to 16 V. Most nonengine vehicular loads are either lamps or motors, and these types of loads will function in some proportion all the way down to near zero voltage when controlled by simple switches. When electronic control circuits are used with these loads, a power supply voltage threshold level occurs, and the load abruptly stops functioning. This threshold level normally takes place between 3 to 8 V and is dependent upon the control circuit's biasing networks and internal transistor technologies. The important criterion is that the electronic control should behave in a stable fashion when the power supply voltage varies between nominal levels and zero. A smooth transition during power supply on/off conditions in the electronic designs can be implemented with some form of hysteresis element for the analog portions and a low-voltage interrupt/reset for MCU/digital circuits.

Voltage transients tend to be either high-energy, up to 125-V levels, or low-energy, high-voltage spikes. The alternator "load dump" voltage spike ranges from about 85 to 125 V for a duration of approximately ½ second. This voltage spike is of a high-energy nature because of the alternator's low source impedance. By using Zener-type rectifiers in the alternator in place of the alternator's normal rectifiers, the load dump voltage can be clamped to under 40 V.

High-voltage transients are generated by the fast turnoff of high-current inductive loads, such as air-conditioning compressor clutches. The polarity of high-voltage inductive kickback spikes depends upon the control switch configuration. Adding a rectifier across the inductive load to clamp the turnoff voltage spike will affect the mechanical performance of some solenoids, actuators, or relays. The mechanical turnoff will slow down. As the solenoid's magnetic field collapses, a countervoltage is generated that, in turn, causes the magnetic field to reverse, which speeds up the solenoid's turnoff action. The insertion of a back-emf rectifier clamps or shunts the counter voltage, and therefore minimizes any significant reversal of the magnetic field. Using back-to-back Zener rectifiers across the inductive load that are selected will allow satisfactory mechanical turnoff action, and yet will protect the power electronics.

Another form of electrical stress that the motor control system has to sustain is ESD or electrostatic discharge. The ESD is normally encountered in the manufacturing cycle, but can occur in the application. To minimize ESD, the control shafts, switch handles, user faceplate, etc., should be designed to dissipate ESD through a ground path away from the electronic module.

19.4.3 Microcomputer I/O Line Protection

Voltage transients on MCU input or output (I/O) lines can cause the MCU to fail. Figure 19.20 shows three different methods to contain excessive MCU input or output line voltage spikes.

Single input line protection is accomplished by a 5.1-V Zener clamp to contain positive-going transients and a Schottky rectifier for the negative-going transients. The Schottky rectifier does have temperature problems at both ends. At 125 °C, its leakage current may reach 50 μA when the input line is at a 5-V level. Therefore, the input signal may lose 50 μA of current to the Schottky which may reduce its voltage level too much for the MCU input port to respond properly. Another problem occurs at cold temperatures. At –40 °C, the forward voltage rises to about 0.47 V, which is getting close to the –0.50-V maximum level specification for most HCMOS-type microcomputer chips. A simple RF low pass filter network is also a good idea to incorporate on each input line. The RC network will minimize problems from high-energy RF fields and will limit excessive current levels from flowing through the MCU's internal substrate.

Multiple input line protection uses two Schottky rectifiers per line for clamping positive- and negative-going voltage transients. A single 4.7-V Zener is used as a common clamp to all input lines.

MCU output line protection uses a resistor, capacitor, Zener, and Schottky rectifiers. The resistor is connected in series to the load, and will therefore have to be sized to drop about 5 to 10 percent of the nominal load voltage. The resistor's main function is to limit reverse current into the MCU in case the output line is raised above Vdd or below Vss, and no other protective devices are used. Adding a small value capacitor at the load output will help minimize RFI, at the expense of reducing switching time from the MCU. Adding a Schottky and Zener rectifier to the MCU output line will protect against both negative- and positive-going transients.

Adding program traps in the MCU software will help to restart the MCU in the advent of a software crash. Experiments have shown MCU-based electronic modules that fail (program crashes, and MCU locks up until hardware restarts it) due to high bursts EMI or RFI, will still perform, but at reduced throughput, when software traps are used. One method uses a trap consisting of three NOP instructions followed by a JMP to RESET instruction. These traps

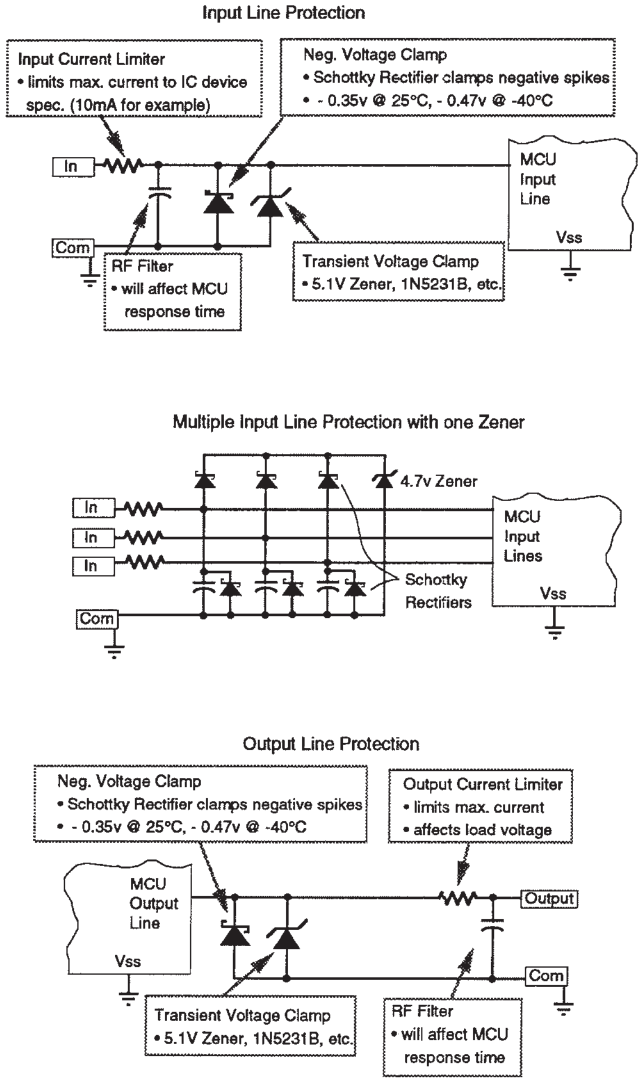


FIGURE 19.20 MCU I/O voltage spike clamping.

are used after loops or subroutines. These software traps, in addition to a watchdog timer if the MCU has this feature, will help in catching an out-of-control program.

The location of the watchdog timer reset instruction code is also important. If the watchdog timer is reset from within an interrupt routine, there is a possibility that the program could still be out of control, but that the interrupt vectors could still be working, and will therefore not reset. A good place to locate the watchdog update code is in a main program path rather than in an interrupt routine.

The PCB or thick-film design layout affects the MPU's reliability in terms of overvoltage transients. The clamping networks should be located physically close to the input socket pins and common. The MPU common traces should not be in the path of any voltage-clamping networks. It may also be a good idea to place an inexpensive metal shield around the MPU PCB area to protect against RFI and ESD problems.

19.4.4 Wire Size Reference

Most wire tables only give basic weight, dimensions, and dc resistance. A more complete version would include straight-line inductance as shown in Table 19.2. Unlike resistance which adds in a linear proportion to its length, inductance adds up on a logarithmic scale.

19.5 FUTURE LOAD CONTROL CONCEPTS

Higher integrated or more complex semiconductor power devices are available for specific load controls that can justify the extra cost of these smarter power types. The tradeoff is that the extra internal circuitry in the device usually limits its application to specific loads.

TABLE 19.2 Wire Gage Resistance and Inductance

Wire #	Dia., mm	Dia., in	Ohms per 305 mm (ft)	L of 25 mm (1 in), μH	L of 305 mm (1 ft), μH	L of 914 mm (1 yd), μH
1000 MCM	29.261	1.1520	0.000008	.003	.182	.746
500 MCM	20.650	0.8130	0.000016	.004	.203	.810
250 MCM	14.605	0.5750	0.000031	.006	.224	.873
0000	11.684	0.4600	0.000049	.007	.238	.914
000	10.404	0.4096	0.000062	.008	.245	.935
00	09.266	0.3648	0.000078	.008	.252	.956
0	08.252	0.3249	0.000098	.009	.259	.977
1	07.348	0.2893	0.000124	.010	.266	.999
2	06.553	0.2580	0.000156	.010	.273	1.019
4	05.189	0.2043	0.000248	.011	.287	1.062
6	04.115	0.1620	0.000395	.012	.301	1.105
8	03.264	0.1285	0.000628	.014	.315	1.147
10	02.588	0.1019	0.000999	.015	.329	1.189
12	02.052	0.0808	0.001589	.016	.344	1.232
14	01.628	0.0641	0.002526	.017	.358	1.274
16	01.291	0.0508	0.004016	.018	.372	1.317
18	01.024	0.0403	0.006386	.020	.386	1.359
20	00.812	0.0320	0.010154	.021	.400	1.401
22	00.644	0.0254	0.016139	.022	.414	1.444
24	00.511	0.0201	0.025671	.023	.428	1.486
26	00.405	0.0159	0.040818	.024	.443	1.529
28	00.321	0.0126	0.064914	.025	.457	1.571
30	00.255	0.0100	0.103093	.027	.471	1.613
32	00.202	0.0080	0.164095	.028	.485	1.656
34	00.160	0.0063	0.260891	.029	.499	1.698
36	00.127	0.0050	0.414849	.030	.513	1.741
38	00.101	0.0040	0.659695	.031	.527	1.783
40	00.080	0.0031	1.048548	.033	.542	1.825

$R_{ohm} = 4\,Lp/\pi D^2$. L = length in meters, D = length in mm, p = resistivity .017241Ω per μΩm (= Ω mm^2/m). (Ref: *Automotive Handbook,* 2d ed, Bosch, 1986, pp. 176–177.)

$L\mu H = (0.0002\,L)*[(IN(2*L/(D/2)) - 0.75)]$. L = length in mm, IN = nat. log., D = dia. in mm. (Ref: *ARRL Handbook,* 1988, pp. 2–18.)

Some form of more intelligent brake indicator system may help minimize the classic chain vehicle crashes that occur in poor driving conditions. One possible method using electronic switched stop lamps would involve a rate of closure detector system to determine if the vehicle's speed is safe for objects ahead of it. If the closure rate is unsafe, the stop lights could be activated to alert trailing (or tailgating) drivers to a pending accident.

19.5.1 Fully Integrated Power Devices

The concept of putting the MCU and power drivers on one piece of silicon is viable at limited power levels and breakdown voltages. The MC68HC05V8 MCU is but one of several semiconductor industry examples of this technology.

GLOSSARY

Charge pump A circuit that usually consists of an oscillator driving logic gates that yield out-of-phase signals that are applied to a rectifier voltage doubler network. The charge pump circuit can be integrated into MCUs and other power analog integrated circuits.

Current mirror A design that generates a signal whose level is directly proportional to another current level. The current mirror concept is used in current-sensing power MOSFETs.

Electromagnetic interference (EMI) Unwanted magnetically coupled voltages that affect the normal operation of electronic systems. EMI is a problem when high-energy circuits are in close proximity to high-impedance sensors, or any low-energy electronic control circuit. EMI also occurs in a wiring harness. One wire conducting a high-energy pulse in this harness will couple the pulse into the other harness wires. Twisted wire pairs are often used to minimize EMI susceptibility.

Freewheeling rectifier A rectifier that is connected across an inductive load or transistor switching device to suppress the voltage spike generated when the load is turned off.

H-bridge A design utilizing four power devices that are connected to reverse the voltage across both terminals of a load. H-bridges are used for bidirectional motor controls.

Half H-bridge A design using two power devices that are connected in series from the positive power supply bus to common, and a load terminal connected to the transistor's middle-connection. Two half H-bridges can make a full H-bridge. Three half H-bridges are commonly used in three-phase motor controls.

Intrinsic rectifier A rectifier that is inherently part of a semiconductor design. Most power MOSFET transistors have a drain-to-source rectifier that is formed by the nature of the MOSFET design structure. In battery-powered designs, the intrinsic rectifier is a concern during reversed battery hookups.

Load dump The effect of disconnecting the load from an alternator running at full power. In automotive equipment this condition can be caused by intermittent battery connections when the alternator is applying maximum current to the battery. When the load is abruptly disconnected, the alternator control circuitry shuts off the field current, but this current takes up to ¼ s to decay, and allows the alternator to produce a voltage much higher than its nominal value, up to 125 V in some cases. This abnormal voltage tracks the field current decay.

Pulse-width modulation A common control signal modulation method used in dc speed controls. The signal's frequency is fixed at a rate of 100 Hz to 20 kHz, but its pulse width is varied from 0 to 100 percent. The variable pulse width, in effect, acts like a variable voltage source when driving most automotive type loads.

P-channel MOSFET Usually a power transistor that requires a negative gate-to-source control voltage of 3 to 10 V. P-channel MOSFETs are normally connected between power supply positive bus and the load, and can operate from a positive voltage bus.

N-channel MOSFET Usually a power transistor that requires a positive gate-to-source control voltage of 3 to 10 V. N-channel MOSFETs are normally connected between common and the load, and operate from a positive voltage bus. The gate input characteristics are capacitive with gate power loss occurring during the switching edge transition period.

Radio frequency interference (RFI) High-frequency signals of sufficient magnitude to influence the normal operation of electronic systems. RFI sources from automotive equipment include spark plug arcing, dc motor brush arcing, electrical contact switching, and fast transitional voltage pulses from power electronic circuits.

Saturation A power transistor's forward ON region normally defined as when an increase in drive input voltage or current has little effect upon further reducing the transistors forward ON voltage. Saturation regions vary from less than 0.1 to over 3 V depending on the transistor technology and design.

Zener rectifier A semiconductor rectifier device that is designed to operate in a reverse bias mode. When the reverse voltage reaches the Zener voltage, the device abruptly starts conducting current. Zener rectifiers are used for shunt voltage regulators, overvoltage protection, and voltage level detectors. Zener rectifiers are also designed specifically for high-power voltage transient suppression. These devices are constructed to handle very high momentary current levels.

Note

Based in part on the article by Richard Valentine, "Don't underestimate transistor-based lamp driver design," *EDN,* June 7, 1990, pp. 119–124.

BIBLIOGRAPHY

Aptronix Inc., *FIDE Users Manual,* 1992.

Bairanzade, M., "The electronic control of fluorescent lamps," Motorola Semiconductor Application Note, AN1049, 1990.

Bannatyne, R., "Using fuzzy logic in practical applications," *Motorola Semiconductor Engineering Bulletin,* EB412, 1993.

Bacelis, O. W., et al., "The automotive primary power supply system," SAE Technical Paper 741208, 1974.

Berringer, Ken, "High-current DC motor drive uses low on-resistance surface mount MOSFETs," Motorola Semiconductor Application Note, AN1317, 1992.

Catherwood, M., "Designing for EMC compatibility with HCMOS MCUs," Motorola Semiconductor Application Note, AN1050, 1989.

Davis, B., et al., "The impact of higher system voltage on automotive semiconductors," SAE technical paper 911658, 1991.

Gauen, K., and J. Alberkrack, "Three piece solution for brushless motor controller design," Motorola Semiconductor Application Note, AN1046, 1989.

Graham, D., and J. Savage, "Brushless DC motor technology," *Second IAVD Congress on Vehicle Design & Components,* Mar. 1985, pp. c1–23.

Hollander, D., "The hidden dangers of electrostatic discharge," Motorola Semiconductor Application Note, AR300, 1987.

Huettl, T., and R. Valentine, "Simplified control for brushless motors," *Machine Design,* Feb. 21, 1985, pp. 236–241.

Mihin, M., and R. Valentine, "New rectifiers clamp alternator voltage levels," SAE technical paper 872002, 1987.

Motorola Inc., "Reducing A/D errors in microcontroller applications," Motorola Semiconductor Application Note, AN1058, 1990.

Motorola Inc., TMOS power MOSFET transistor data, Chap. 2, DL135 Rev. 4, 1992.

Schultz, W., "New components simplify brush DC motor drives," Motorola Semiconductor Application Note, AN1078, 1989.

Society of Automotive Engineers, *SAE Ground Vehicle Lighting Standards Manual,* SAE HS-34, 1993.

Tal, J., "Design and analysis of pulsewidth-modulated amplifiers for DC servo systems," *IEEE Transactions on IECI,* vol. 23, no. 1, Feb. 1976, pp. 47–55.

Valentine, Richard, "Brushless DC motor control," *Control Engineering,* July 1985, pp. 134–140.

Valentine, Richard, "Don't underestimate transistor-based lamp-driver design," *EDN,* June 7, 1990, pp. 119–124.

Zhang, W., "Temperature control 2," FIDE Application Note 005-920903, Aptronix Inc., 1992.

ABOUT THE AUTHOR

Richard J. Valentine is a principal staff engineer at Motorola SPS in Phoenix, Ariz. His present assignments include engineering evaluation of advanced semiconductor products for emerging automotive systems. He holds two patents and has published 29 technical articles during his 24 years at Motorola.

P · A · R · T · 4

DISPLAYS AND INFORMATION SYSTEMS

CHAPTER 20

INSTRUMENT PANEL DISPLAYS

Ronald K. Jurgen, Editor

20.1 THE EVOLUTION TO ELECTRONIC DISPLAYS

In the early automobile years, cars had analog displays that contained minimal information, usually just car speed and oil pressure. As the use of electronics increased in cars dramatically from the late 1970s on, however, the traditional mechanical or electromechanical analog displays with a circular dial face and a pointer began to be challenged by newer technologies. The major ones included gas discharge or plasma displays, vacuum fluorescent displays (VFDs), liquid crystal displays (LCDs), cathode-ray tubes (CRTs), and more recently, head-up displays (HUDs). These newer technologies made it possible for car makers to give drivers a broad spectrum of information, including sophisticated graphics.

In 1978, the first production electronic digital display, a gas plasma device, was used in the Cadillac Seville; in 1984, the first standard equipment production LCD cluster was used in the Chevrolet Corvette; and in 1985, the first full-color CRT was used in a production vehicle, the Toyota Soarer.[1]

The initial rush by car makers to use electronic displays has now abated somewhat. Drivers, in many instances, did not take kindly to overkill with electronic displays—any more than they did to electronically generated voice messages—and today's use of displays tends to be more conservative. In fact, one approach fast gaining popularity is electronic analog displays rather than strictly digital ones.

This chapter will describe typical electronic displays used in cars. It is not intended to be all-inclusive. That would be beyond the scope of this handbook. But the display examples given are representative of what can be found in cars today.

20.2 VACUUM FLUORESCENT DISPLAYS

One of the most widely used electronic displays is the vacuum fluorescent display. It was first used in automotive clocks and has since been applied in other ways including audio systems, air conditioning/heating, message centers, and head-up displays. Its popularity stems from features that include availability of a variety of colors, high luminance, low voltage operation, reliability, and long life.

The basic structure of a modern VFD, Fig. 20.1, consists of a cathode, grid, and anode.[2] The tungsten-wire cathode emits thermal electrons that are accelerated when a positive voltage is applied to the anode and grid. The phosphor-coated anode emits a blue-green light when struck by the emitted electrons. Other colors such as yellow-green, green-yellow, yellow-orange, orange, and red-orange colors, obtained through the use of optical filters, can be used independently within the same display. Thick- or thin-film screens are used to shape the phosphor pattern in any desired manner. VFDs produce maximum perceived brightness at low input power.

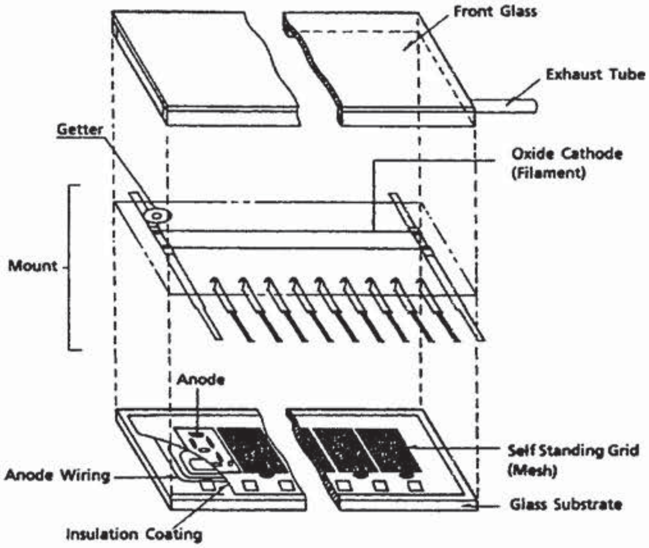


FIGURE 20.1 The basic structure of a vacuum fluorescent display. *(Reprinted with permission from SAE SP-858 ©1991 Society of Automotive Engineers, Inc.)*

Initial VFDs were single-digit displays. Multiple-digit displays followed and now single large-scale VFDs contain multiple functions such as speedometer, fuel and temperature gages, trip odometer, turn indicators, and warning signals.

A typical application of VFDs was in the electronic instrument panel on the 1985 Chrysler LeBaron.[3] The VFDs are controlled (Fig. 20.2) by a Motorola MC6805R3 8-bit, HMOS, single-chip microprocessor with internal analog-to-digital converters. Another microprocessor, a 4-bit device, is used to generate a high switching rate for the center display in the instrument panel.

20.3 LIQUID CRYSTAL DISPLAYS

LCDs are fluids consisting of organic compounds whose molecular orientation rotates polarized light by 90 degrees as light passes through the material.[4] When an electric field is applied to the fluid, its molecular orientation is altered as is its polarizing ability. The change is visible in reflected or transmitted light and can be used to form predetermined patterns or characters.

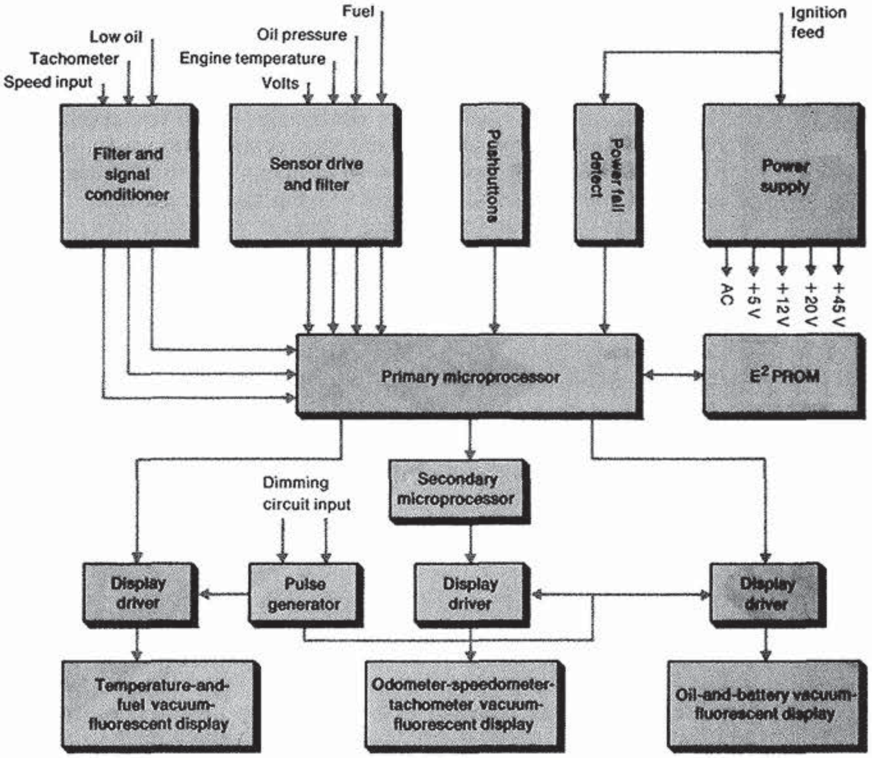


FIGURE 20.2 Vacuum fluorescent displays and their driving circuits as used in the 1985 Chrysler LeBaron. *(©1984 IEEE)*

There are three general classes of liquid crystals called *nematic, smectic,* and *cholesteric.*[5] In nematic liquid crystals, the molecules are lined up in one dimension but have random order in the other two directions. In smectic liquid crystals the ordering of the molecules is similar except that it is much higher. In cholesteric liquid crystals, the molecules stack in layers that have a twist and tilt that gives the molecules a spiral configuration.

Certain chemical substances applied on top of the two electrodes of a liquid crystal display orient the long axes of the molecules so that they line up parallel to the surface of the glass plate. The display cell is assembled so that the long axes of the molecules on the top plate are oriented at 90 degrees to the direction of the molecules on the bottom plate. The molecules between the plates assume a spiral configuration.

A polarizer is placed on top of the display cell. When light enters the liquid crystal region of the display, the direction of polarization is rotated as the light travels along the spiral configuration. A second polarizer behind the bottom plate is oriented to allow the light with its altered polarization to pass through. The display then appears bright to the observer. An electric field applied across the display cell, the molecules realign with their long axes parallel to the direction of the field. The polarized light entering the cell is no longer twisted and the second polarizer cuts off the light passing through the liquid crystal. The observer sees a dark screen.

In what is called a dichroic LCD, no polarizers are used, but a dichroic dye is combined with the liquid crystal material.[4] The dye molecules combine with the liquid crystal molecules to produce unique spectral absorption characteristics. With no voltage applied, the dye absorbs certain wavelengths of the incident ambient light and the observer sees a colored background. When the display is activated, the observer sees colorless display elements against the colored background.

In simple liquid crystal displays containing only a few seven-segment characters, each numeral is addressed by activating the appropriate segments to form a specific numeral. But for more complex displays, consisting of many picture elements (pixels) that are twisted nematic cells, this type of addressing is too cumbersome and matrix addressing is used instead. The pixels are turned on by applying voltage to both their row and their column electrodes rather than directly to the individual pixels.

20.4 CATHODE-RAY TUBE DISPLAYS

Although the cathode-ray tube (CRT) has found wide success in a variety of applications including television receivers, instruments, and radar and medical equipment, it has been far less successful in the automotive field. Some of the reasons for its lack of impact are the amount of space needed behind dashboards to house the CRT, possible washed-out images in sunlight, lack of instant-on capability, and possible safety hazards from x-ray exposure, implosion, and high voltages.[6]

Despite the difficulties that CRTs present to car makers, there have been limited applications in cars over the years in several General Motors cars and in some Ford Motor Company models. In 1986, for example, the General Motors Corp. introduced a CRT-based display called a Graphic Control Center in 1986-model Buick Rivieras.[7] The center combined diagnostic with control functions in an integrated information system and handled inputs from 7 to 10 microprocessors with the actual number dependent on the number of options on the car. The center could be used to call up information on the CRT in the basic areas of climate, radio, trip monitor, and gages, in addition to diagnostics information. The display for each area would appear as a "page" of information. In some cases, there would be more than one page, starting with the general and proceeding to the more specific.

When the ignition key was turned on, a summary page would appear displaying key information from all the basic areas. To access more detailed information, the driver simply touched an appropriate spot on the screen. A Mylar switch panel positioned over the screen used ultrathin wires encoded by rows and columns to send signals to the control circuits.

Although General Motors had CRT systems in the Riviera and subsequently the Buick Reatta and a somewhat similar system in the Oldsmobile Toronado and Trofeo, they were eventually dropped. One of the main drawbacks was that the driver had to take his or her eyes off the road in order, for example, to tune the radio. The Oldsmobile system improved on the Buick version in this regard by careful placement of some of the more commonly used controls so that the driver could activate them while still looking at the road. But car buyers, by and large, never became convinced that the CRT system was necessary or desirable.

A possible comeback for the CRT may be in the offing when automobile navigation systems take hold (see Chap. 29). Many of those systems use CRTs to provide the driver with useful trip information.

20.5 HEAD-UP DISPLAYS

A head-up display (HUD) is aptly named since it allows viewing of data superimposed on the driver's visual field with his or her head up. HUDs have been used in the military aircraft industry for over 20 years and more recently have had limited application in automobiles. A main advantage of a HUD in a car is that the driver need not constantly refocus his or her eyes as when switching them from the road to conventional dashboards and back again. The first applications in production automobiles were in the 1988 Nissan Silvia model and in special editions of the 1988 Oldsmobile Cutlass Supreme.

One of the first HUDs to appear in cars was in 1988 when the Oldsmobile Cutlass Supreme Indianapolis 500 Pace Car and its 54 replicas were equipped with the systems.[7] This was followed by offering the HUD as an option on the 1989 Oldsmobile Cutlass Supreme and on the Pontiac Grand Prix. Those systems were developed jointly by General Motors engineers from the Hughes Aircraft Corp., Delco Electronics Corp., and C-P-C Engineering. HUDs continue to be offered today by General Motors on some models.

The heart of the original GM system was an image source—a custom-designed, high-intensity, blue-green vacuum fluorescent tube made by Futaba Corp. of America. High brightness was made possible by keeping the cathode energized at all times and the electrodes at higher voltages than would be used in vacuum fluorescent tubes for conventional purposes.

The system works as follows. Speed and other sensor inputs are processed by an electronic module which then sends signals to the vacuum-fluorescent tube to activate segments of seven-segment numbers or graphic symbols in the tube. Optical elements then project the light from those energized segments onto the windshield of the car. The driver sees virtual images that seemingly float in space near the front end of the car (Fig. 20.3).

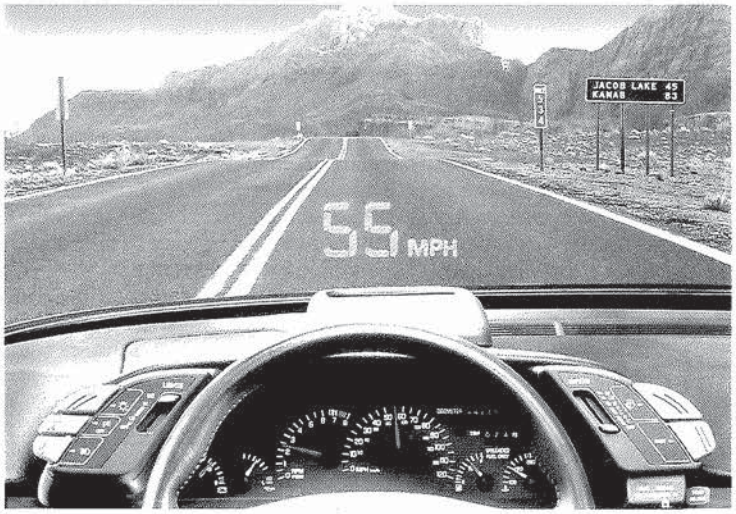


FIGURE 20.3 This head-up display, showing a virtual image of car speed, was standard on General Motors' Pontiac Division's 1993 Bonneville SSEi and is available as an option on the Bonneville SSE and on all Grand Prix models. Information that is projected onto the windshield includes vehicle speed, turn signal indicator, high-beam lights, check gages warning, and low-fuel alert.

The images can display car speed in either mi/h or km/h, left and right arrows for turn signal indications, a headlight symbol for high-beam indicator, and a gas pump for low-fuel warning.

As shown in Fig. 20.4, the image is projected from the image source onto the car's windshield by mirrors. A proprietary optical design keeps the image in the correct aspect ratio. The driver can adjust both the brightness and the vertical location of the image but the horizontal

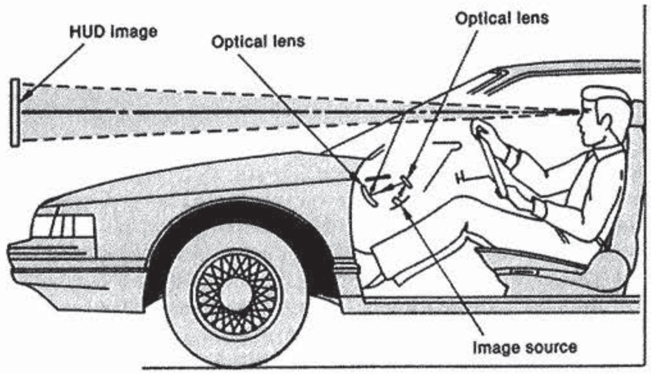


FIGURE 20.4 General Motors' head-up display uses an image projected onto the windshield of the car but the driver sees a virtual image of the projected information positioned at the front of the car. *(©1988 IEEE)*

location is fixed. Making it variable would require expensive variable-magnification changes in the system optics.

20.6 ELECTRONIC ANALOG DISPLAYS

Despite the proliferation of electronic digital displays in cars, many drivers still prefer analog displays. To satisfy this need while at the same time taking advantage of advances in electronics, many car makers have introduced electronic analog displays. These displays, despite their dependence on electronic circuits, present the driver with essentially the same type of display as in the electromechanical speedometer, for example.

Some car makers have seized the opportunity to offer their customers unique analog displays. One notable example is the high-precision electronic analog display in the Toyota Lexus LS 400.[8] Called a combination meter, Fig. 20.5, it uses a smoked-filter glass to cover the

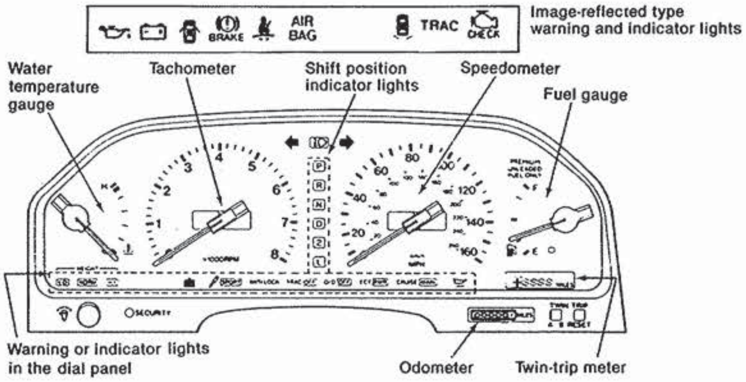


FIGURE 20.5 Toyota's Lexus LS 400 has a so-called combination meter that presents analog displays that are electronically activated. *(©1989 IEEE)*

entire display area that is blacked out when the ignition switch is turned off. Cold-cathode tubes containing mercury gas light the light-transmitting dial panel. When powered, the tubes emit a white light to illuminate the dials when the ignition switch is turned on.

The speedometer, tachometer, fuel gage, and water temperature gage all have self-light-emitting needles consisting of cold-cathode tubes with a sealed-in mixture of neon and xenon gases that emit white light when energized. A cableless speedometer driven by a moving coil indicates speed based on pulse signals from the transmission speed sensor.

20.7 RECONFIGURABLE DISPLAYS

What may be a future trend in displays was presented at the Society of Automotive Engineers annual meeting in Detroit, Mich., in 1993. A reconfigurable automotive display system was described that displays multifunctional information on a dot matrix fluorescent indicator panel.[9] An advanced 8-bit microcontroller with real-time processing capabilities is used in the system to perform display refreshing, keypad scanning, and serial communications.

The dot matrix display in the system is a vacuum fluorescent type composed of an 80 by 16 matrix of pixels in a 9.85-mm-high by 50.5-mm-long graphics area. Two lines of 13 characters each can be displayed using a 5 by 7 font. The system can display text messages of different sizes, locations, and font types as well as graphics symbols. The advantage is that several fixed segment display systems can be replaced with a single dot matrix display system to cut costs and packaging space.

REFERENCES

1. Eric M. Ethington, Gary A. Streelman, and David R. Clark, "What information does an operator expect in an automobile display?," *Automotive Information Systems and Electronic Displays: Recent Developments,* SP-770, Society of Automotive Engineers, Inc., Feb. 1989, pp. 51–54.
2. Nobuo Akiba; Robert J. Davis; Kato, Satoro; Tatiyoshi, Sadao; Torkai, Masahiro; and Tsunesumi, Satoshi, "Technological improvements of vacuum fluorescent displays for automotive applications," *Vehicle Information Systems & Electronic Display Technology,* SP-858, Society of Automotive Engineers, Inc., Feb. 1991, pp. 59–68.
3. Ronald K. Jurgen, "More electronics in Detroit's 1985 models," *IEEE Spectrum,* Oct. 1984, pp. 54–60.
4. Lawrence A. Lopez, "Display Technology," *Proceedings International Congress on Transportation Electronics,* P-111, Society of Automotive Engineers, Inc., Oct. 1982, pp. 191–197.
5. Charles M. Apt, "Perfecting the picture," *IEEE Spectrum,* July 1985, pp. 60–66.
6. Ronald K. Jurgen, "All electronic dashboards coming," *IEEE Spectrum,* June 1981, pp. 34–37.
7. Ronald K. Jurgen, "New frontiers for Detroit's Big Three," *IEEE Spectrum,* Oct. 1988, pp. 32–34.
8. Ronald K. Jurgen, "Global '90 cars: electronics-aided," *IEEE Spectrum,* Dec. 1989, pp. 45–49.
9. Gregory T. Gumkowski and Adnan Shaout, "Reconfigurable automotive display system," *Automotive Display Systems and IVHS,* Society of Automotive Engineers Inc., pp. 13–19.

CHAPTER 21
TRIP COMPUTERS

Ronald K. Jurgen, Editor

21.1 TRIP COMPUTER BASICS

Trip computers have evolved over the years from simple systems that estimate only the distance that can be traveled with the remaining fuel to sophisticated systems that also offer such features as instantaneous and average fuel economies, amount of fuel used, average speed, amount of fuel remaining, estimated time of arrival, oil life indicator, and diagnostic capabilities.

21.1.1 Basic System Configurations

A simple distance-to-empty system such as that shown in Fig. 21.1[1] has transducers that convert distance and fuel quantity into time-varying voltages or currents. Electronic signal pro-

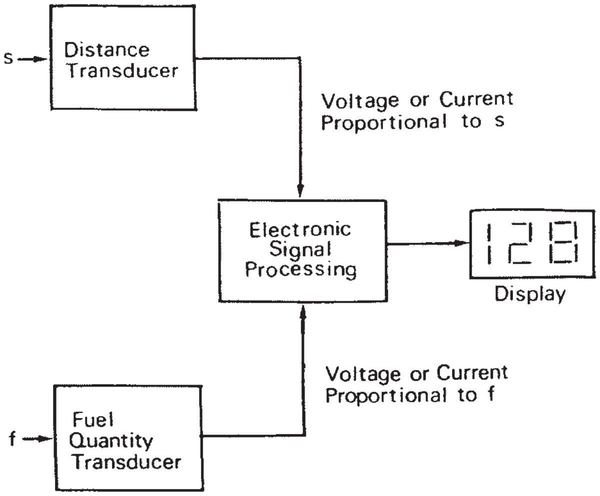


FIGURE 21.1 Elements of a basic distance-to-empty system. *(Reprinted with permission from SAE Technical Paper 800240 ©1980 Society of Automotive Engineers, Inc.)*

cessing operates on those voltages or currents to produce a distance-to-empty estimate that is formatted for transmission to a display device.

A basic trip computer, shown in Fig. 21.2,[2] in addition to distance-to-empty, computes instantaneous fuel consumption, average fuel consumption, and average cruising speed. Through use of a sequential selector button, the driver causes the readouts to appear in order on the display.

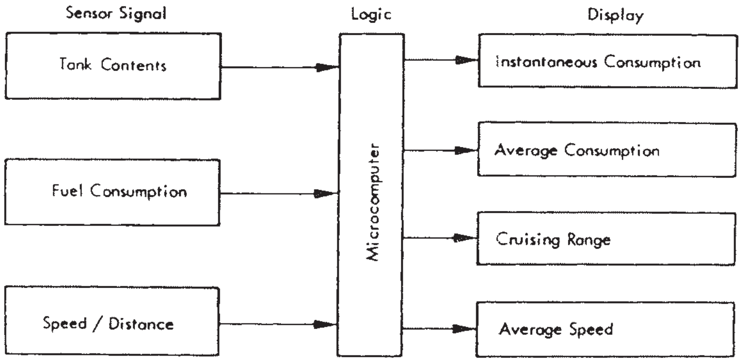


FIGURE 21.2 A simple version of a basic trip computer. *(Reprinted with permission from SAE Technical Paper 810302 ©1981 Society of Automotive Engineers, Inc.)*

21.1.2 A Full-Function Trip Computer

A full-function trip computer, shown in Fig. 21.3,[2] incorporates many more functions. The driver can feed information into the computer at the beginning of a trip—distance to destination, for example. During the trip, the driver can request specific information.

21.2 SPECIFIC TRIP COMPUTER DESIGNS

Different car makers have taken various approaches to trip computers for their car models over the years. Two selected examples follow.

21.2.1 The General Motors Trip Computer

The General Motors Trip Computer was first available as a high-cost option on the Cadillac Seville in 1978. The computer's principal four parts are a function-select keyboard, the central processing unit (CPU), the displays, and the interconnecting special wiring. The CPU translates the various engine and vehicle sensor inputs into the appropriate information needed by the driver and also provides constant speed and fuel displays, generator system information, and diagnostic features.

A block diagram of the bus-oriented CPU designed around the Motorola M6800 microcomputer family with N-channel, 8-bit parallel processing is shown in Fig. 21.4.[3] A 16-bit

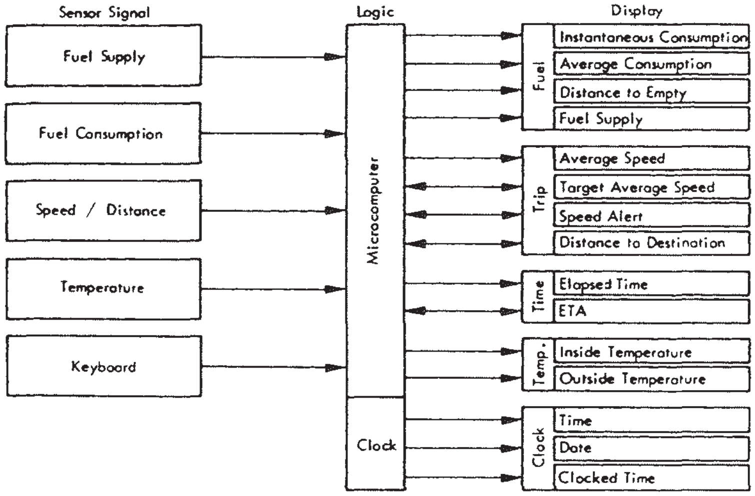


FIGURE 21.3 A full-size basic trip computer. *(Reprinted with permission from SAE Technical Paper 810302 ©1981 Society of Automotive Engineers, Inc.)*

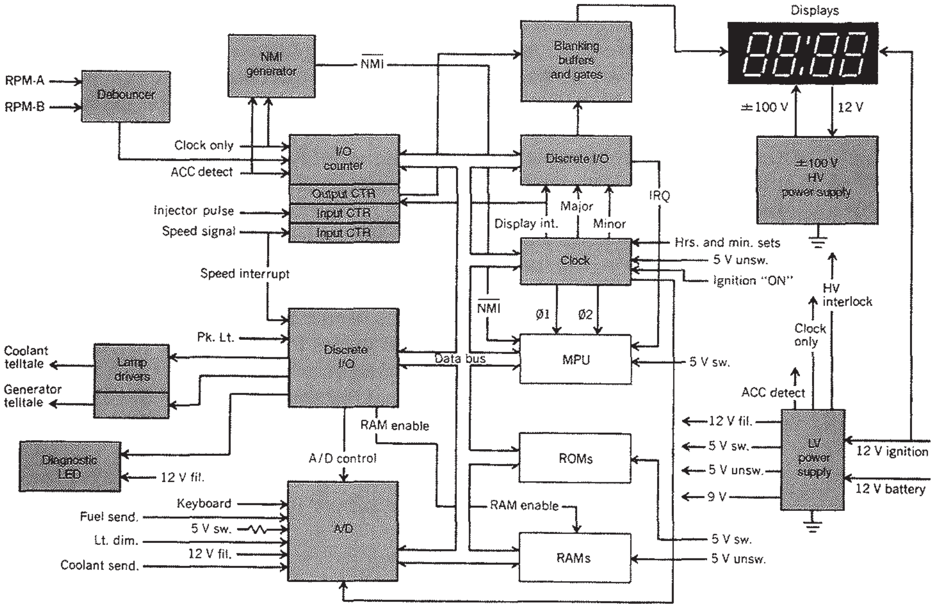


FIGURE 21.4 The central processing unit for the General Motors Trip Computer contains the following software routines in ROM: major loop (fuel-flow content and correction, distance and speed information, trip time and average speed time, fuel sense and display, time-of-day clock, clock colon control, keyboard requests, dimmer information, coolant sensor, system voltage, engine r/min, self-diagnostics); minor loop (keyboard call-up, diagnostic LED control); display-interrupt service (multiplexing, nonvolatile RAM information update, keyboard select saved for minor loop, engine r/min), speed-interrupt service (distance information maintenance for RAM, vehicle-speed information maintenance for RAM); initialization (displays and telltales, setup computer control registers, update trip parameters for "ignition off" time); nonmaskable interrupt service (display blanking, RAM disabling during power down). *(©1978 IEEE)*

address bus provides addressing capability of up to 65,536 word locations. The ROMs hold the instruction op code and the control program data. The RAMs provide software registers, software accumulators, and a stack area for the control program.

The trip computer contains self-diagnostics. When a light-emitting diode on the outside of the computer case is lit, it indicates certain failures. For example, if after sampling the critical +9 V of the A/D chip multiplexer channel, a failure is noted, it is flagged and saved in the RAM for servicing by the minor loop, Fig. 21.4, and then activates the LED on the computer case by way of the discrete I/O.

21.2.2 Ford's Second Generation Tripminder

Ford Motor Company's 1982 Tripminder was a second-generation vehicle trip computer. It performs computations on five input variables to generate the functions available to the driver. It was available in three versions—high, mid, and low series—for use in the Continental, Ford of Europe Granada, and Ford/Mercury Thunderbird/XR-7, respectively.

A block diagram of the Tripminder hardware is shown in Fig. 21.5*a* and software architecture in Fig. 21.5*b*.[4] The 8050 microprocessor could support 4K of internal ROM and 256 bytes of keep-alive RAM used for trip-log functions.

Computations are performed on five input variables. One of them, time, is produced by a precision crystal oscillator internal to the Tripminder. The others are sensed from the vehicle environment. The European and high-series versions have a distance-to-empty feature requiring a fuel-tank-level input. Driving range available is based on the amount of fuel left in the tank and the historical fuel economy of the car.

Fuel flow information is acceptable in two forms. In carbureted engines, fuel flow data is provided by a fuel flow sensor installed in the fuel line just ahead of the carburetor. The sensor's output of 48,000 pulses per gallon is recorded in an 8-bit event counter in the interface integrated circuit. In fuel injected engines, a buffered form of the fuel injector signal is sent to the event counter to determine the number of fuel injector firings that occurred during a sample period. The signal is also sent to an interval counter to determine the total on time of the fuel injector.

Speed information comes from either an electronic speedometer or from a variable reluctance speed sensor installed in line with the speedometer cable. The data consists of 8000 pulses per mile and is recorded in an event counter.

The TIC integrated circuit performs fundamental system control and timekeeping operations as well as the data conversion functions for the sensor inputs. It is a CMOS device that must be continuously powered. At midnight, when the time clock registers roll over, the TIC circuit automatically starts up the microprocessor to update the software calendars. The TIC circuit also continuously monitors the accessory input signal and the clock keyboard button. Sequencing on and off of the power supplies is controlled by the TIC circuit to make certain that the microprocessor is initialized correctly and that the keep-alive memory in RAM is not accidentally written into during power supply transients.

21.3 CONCLUSION

Trip computers vary from relatively simple versions to complex designs that give the driver a wealth of information. Some trip computers also provide a means for on-board diagnostics (see Chap. 22). What a trip computer can do depends on how many sensor inputs are fed to it. But the bottom line, as with any automotive component, is cost tied in with perceived value to the customer.

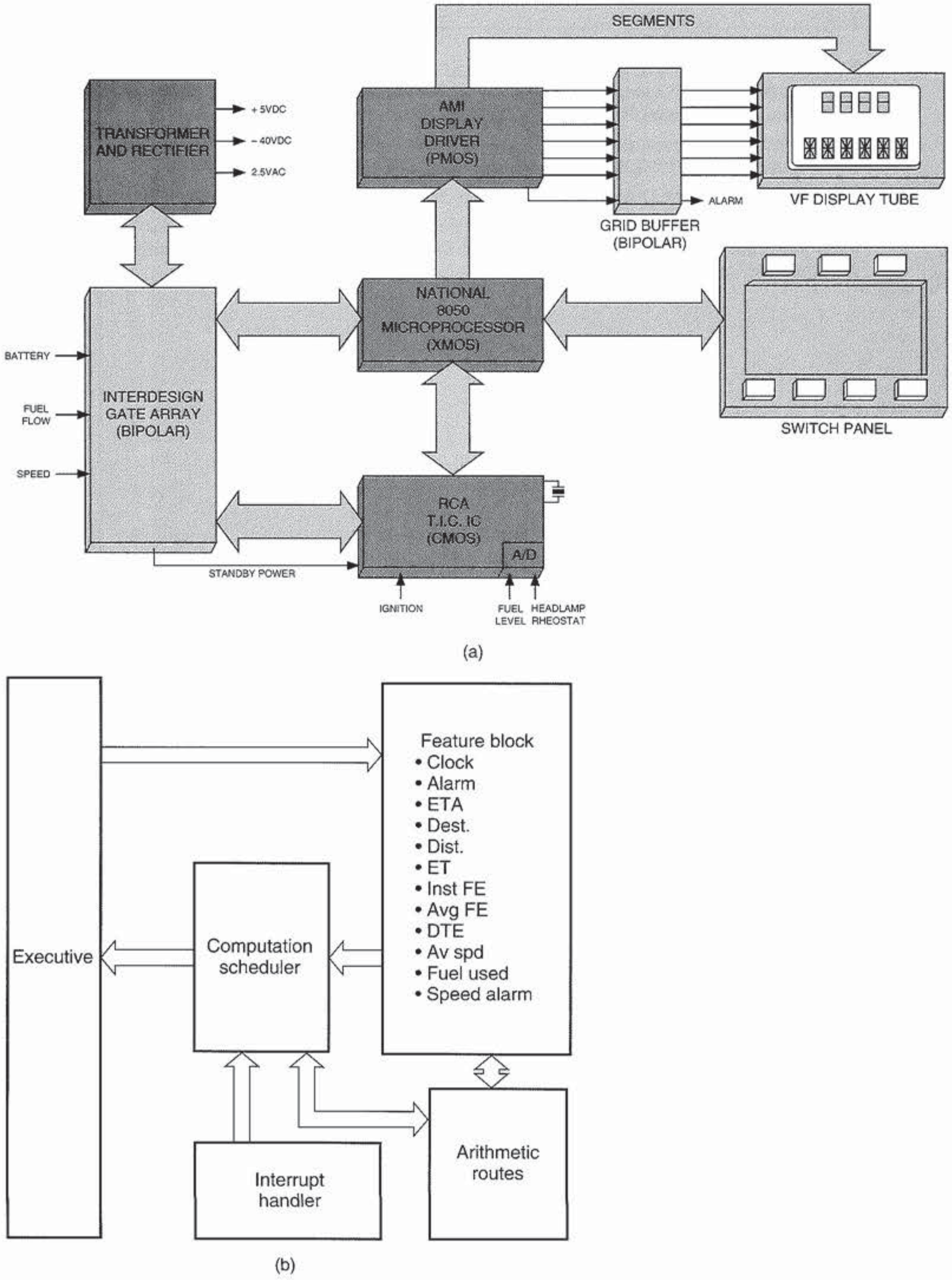


FIGURE 21.5 (*a*) Hardware architecture and (*b*) software architecture for Ford Motor Company's second generation Tripminder. *(Reprinted with permission from SAE Technical Paper 820107 ©1982 Society of Automotive Engineers, Inc.)*

REFERENCES

1. D. R. Sendelback and T. J. Wood, "Response and accuracy limitations for a fuel level based distance-to-empty calculation," *Automotive Electronic Instrumentation—Displays and Sensors,* SP-457, Society of Automotive Engineers, Feb. 1980, pp. 21–24.
2. W. H. Hucho, "The evolution towards a new driver information system," *Electronic Displays and Information Systems,* P-92, Society of Automotive Engineers, Feb. 1981, pp. 67–75.
3. Ronald K. Jurgen, "For Detroit: smaller and smarter chips," *IEEE Spectrum,* Nov. 1978, pp. 33–35.
4. G. Cilibraise, "The second generation family of Ford Tripcomputers—the Tripminder," *Electronic Displays and Information Systems and On-Board Electronics,* P-103, Society of Automotive Engineers, Inc., Feb. 1982, pp. 127–135.

CHAPTER 22
ON- AND OFF-BOARD DIAGNOSTICS

Wolfgang Bremer, Frieder Heintz, and Robert Hugel
Robert Bosch GmbH

22.1 WHY DIAGNOSTICS?

The desire for greater safety, driving comfort, and environmental compatibility is leading to a rapid increase in electronic control units and sensors in upper class, medium-sized, and compact vehicles. Additional functions and their corresponding equipment in today's cars create a bewildering tangle of cables and confusing functional connections. As a result, it has become more and more difficult to diagnose faults in such systems and to resolve them within a reasonable period.

22.1.1 Diagnostics in the Past and Today

On-board diagnosis has been limited thus far to a few error displays and fault storage achieved by relatively simple means. It has been left more or less to each manufacturer to decide to what extent diagnosis would be carried out. Diagnosis always means the working together of man and machine and consists essentially of three major components: registration of the actual condition, knowledge of the vehicle and its nominal condition, and strategy—how to find the smallest exchangeable deficient component by means of combining and comparing both the nominal and actual conditions.

All three points are inseparably connected. Only the means to the end have changed over time. The oldest and simplest method of diagnosis is that done with the help of our sense organs, but the limits of this kind of diagnosis are obvious. In fact, the objective in the development of diagnostic techniques is the extension of human abilities with the aid of diagnostic tools in order to be able to measure more precisely and more directly, to compare more objectively, and to draw definite conclusions.

The development of control techniques was essentially determined by the following items: the development of automotive engineering; the structure of workshops—that is, essentially the relation between the costs of labor and materials; and the development of electronics and data processing.

For a long time, motor diagnosis was limited to ignition control and timing. In the 1960s, new exhaust-gas measuring instruments for fuel injection adjustment were developed, but the mechanic still had to make the diagnosis. In the 1980s, the introduction of electronics in the vehicle was followed by a new generation of measuring instruments in the workshops. Not

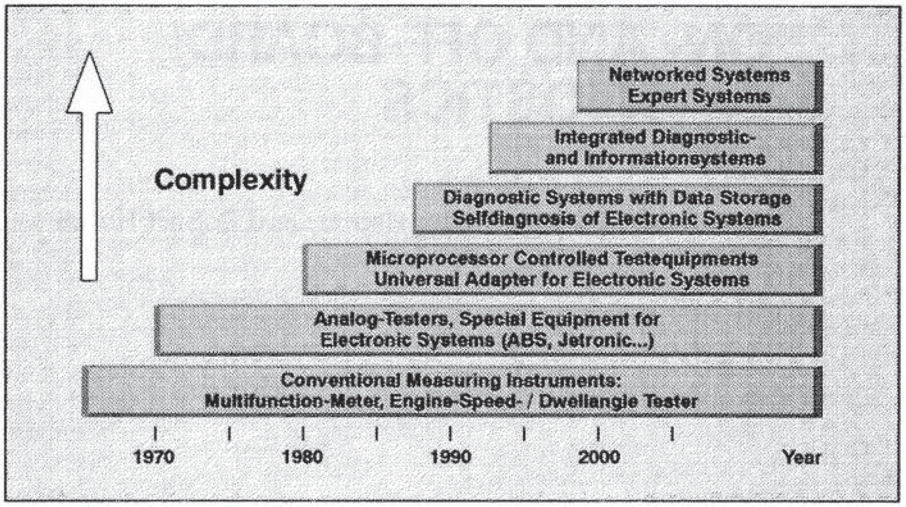


FIGURE 22.1 Evolution of diagnostic test equipment.

only were separate measurements combined with comprehensive test procedures, but also the information about the nominal condition of the vehicle was stored in a data memory.[1] A view of the development is shown in Fig. 22.1.

As more and more electronic systems were added to cars, the more difficult it became to determine the actual condition in case of a defect. Soon a multitude of connecting cables and adapters were required to reach the necessary measuring points. Moreover there was an increasing amount of information needed to make an effective diagnosis. In the majority of workshops, diagnosis is carried out as shown in Fig. 22.2. The most important test points of

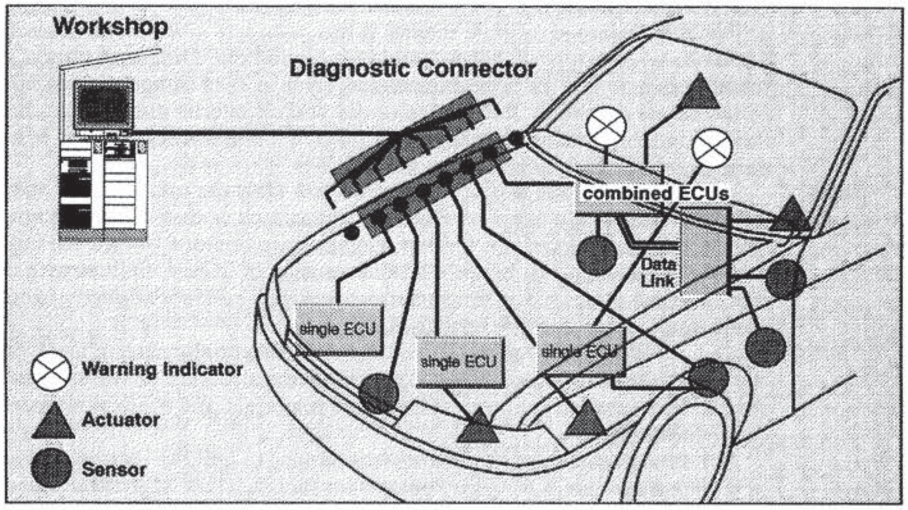


FIGURE 22.2 Present-day diagnostic connector installation in a vehicle.

control units and sensors are tied to a diagnostic connector which is plugged into the measuring instrument with a corresponding adapter for the respective vehicle. Because of the permanently increasing amount of electronic functions, it is necessary to develop connectors with more and more contacts. It is evident that this method soon will become too unwieldy.

Modern electronics in vehicles support diagnosis by comparing the registered actual values with the internally stored nominal values with the help of control units and their self-diagnosis, thus detecting faults. By interconnecting the measuring instruments, a detailed survey of the entire condition of the vehicle is available and an intelligent on-board diagnostic system is able to carry out a more precise and more definite localization of the defect.[2] With the help of an interconnection and standardization of the interface leading to the external tester, the many different complex and expensive adapters have become superfluous. Modern diagnosis will look like what is shown in Fig. 22.3.

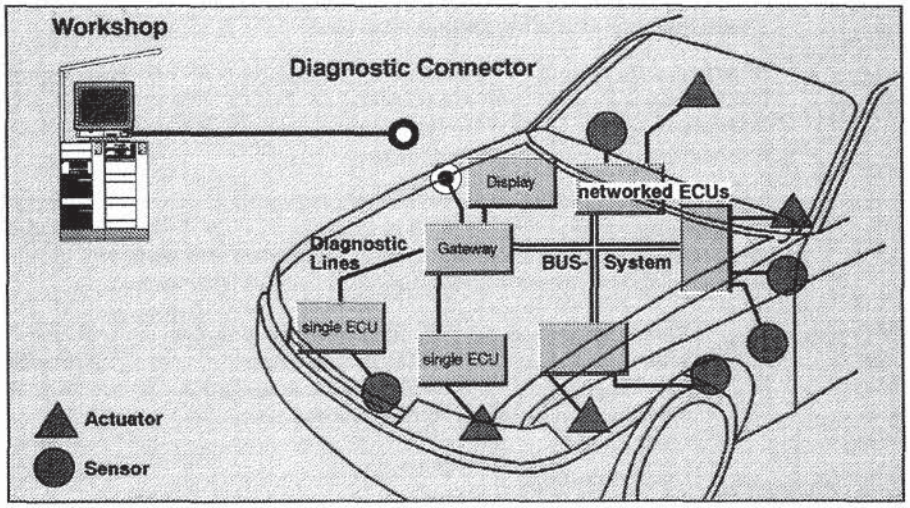


FIGURE 22.3 Future diagnostic connector installation in a vehicle.

Instead of a multiplicity of adapters there is only a single standardized interface, provided by the diagnostic processor. By means of interconnection, the diagnostic processor is provided with all available data and the condition of the vehicle is known. With the help of the diagnostic processor, the external measuring instrument has access to the measuring and diagnostic values of the sensors and is able to directly reach the actuator for measuring purposes.[3]

Such a diagnosis also demands a certain change in the functional structure of a vehicle. Corresponding hierarchical models have already been presented.[4]

22.1.2 Reasons for Diagnostics in Vehicles

Which are the most important reasons for diagnostics as demanded and desired in today's vehicles?

Existing Diagnostic Problems. A number of diagnostic problems must be resolved:

- Early diagnostic information was related only to single components and control units. In case of a defective comprehensive system, every unit, component, sensor, and connecting cable of the system had to be tested and controlled. This was a very time consuming and expensive process.

- Because of the single component and control unit checks, it was impossible to analyze all the additional data correlated with a particular defect.
- In the case of a defect in single sensors or units, the car was often inoperable. Taking into consideration all available information about the vehicle, it is possible to use alternative parameters and procedures in order to achieve at least a so-called limp-home function and sometimes continue the use of the vehicle under only slightly limited operating conditions.
- Usually there was only a global error display with an often ambiguous warning light available for the driver. Drivers desire more detailed information and especially guidelines for what procedures should be followed.
- The multitude of adapter cables, plugs, diagnostic equipment, and communication interfaces in a workshop has become so complex that the effectiveness decreased dramatically, with the repair costs increasing disproportionally.

New Legal Proposals. Worldwide new legal proposals and governmental regulations [e.g., California Air Resources Board (CARB), On Board Diagnostics II (OBDII), Environmental Protection Agency (EPA)] are forcing manufacturers and subcontractors to seek more profitable, effective, and convincing diagnosis of vehicles.

Serial Data Networks. New serial data networks for the connection of control units and vehicle body components, installed in the vehicle, offer the possibility of absolutely new optimum approaches and even anticipate maintenance and diagnosis up to the introduction of autodidactic data processing systems and external data bases.[5,6,7,8]

International Initiatives for Standardization. Initiated by legislative and governmental demands for better diagnostics in the area of emission control, initiatives for standardization in the entire diagnostic field in vehicles were launched during recent years to achieve worldwide standardization of tools, interfaces, connectors, and protocols.

22.1.3 Diagnostic Tasks in Vehicles

In order to minimize the number of defects or even to completely avoid them, a vehicle requires regular checks. In case of an inevitable defect, a clear and directed diagnosis is required and has to be followed by a prompt, reliable, and inexpensive repair. Therefore appropriate diagnostic systems are being developed considering the following targets: simplification of maintenance, fault indication in time, guidelines for the driver in case of a defect, and safer and faster repairs with the help of a specific fault indication.

In addition to technical considerations, environmental aspects are now being taken into consideration as reflected in the diagnostic concepts. In the future, only perfect systems will be accepted, in order to keep environmental pollution to a minimum. It is understandable, therefore, that legislators insist on increased monitoring standards, particularly for exhaust-related components.

As an example of the new monitoring standards, consider the requirements of CARB and EPA in the United States and the resulting consequences for diagnosis. At the moment, the extent of such a detailed monitoring has to be a compromise between the different requirements and the possible technical and economical solutions, but the environmental aspects will gain more and more importance. The increased amount of available data will certainly permit a considerably higher rate of in-depth fault localization and will also allow clear fault identification without interactive outside intervention. Having knowledge of the functional interrelationships and access to all essential data, a picture of the defect can be created with the help of individual pieces of information. The driver and the workshop can then be provided with appropriate instructions. In this context, on-board expert systems are being considered.

For an effective and successful diagnosis today and in the future the following tasks and targets can be defined.

Fault Storage with Boundary Conditions. A very important aspect of modern diagnosis is the clear and reliable analysis of the respective fault. During the self-diagnosis, it is absolutely necessary to store not only the respective fault information but also all relevant marginal parameters in the control unit, e.g., ambient temperature, velocity, engine speed, engine knock, and so on. The additional data can be stored when a defect occurs as well as during specified intervals around the moment of a defect. Such additional data is called "freeze frame" data.[9]

Fault Localization. Mechanics must be able to locate a defective control unit quickly and then determine which component of that control unit is at fault so that it can be replaced.

Data Correlation, Recognition of Imminent Faults. A large amount of data useful for the analysis of a vehicle is now available and even more will be available in the future. These data will have to be evaluated and compared with the help of modern data processing techniques, including fuzzy logic, neural networks, autodidactic systems, and expert systems. These techniques will not only enable the diagnosis of the actual condition of the vehicle but will also determine future maintenance needs. As a result, the reliability and availability of a vehicle will be increased and the possible consequences of a defect kept to a minimum. The driver can also be forewarned about imminent problems and can then take appropriate steps before starting on a trip.

Parameter Substitution. The breakdown of a sensor in modern diagnostic procedures is not necessarily followed by a lack of the respective information. After having diagnosed a fault, the diagnostic computer—with the aid of the available information—is often able to compute an auxiliary parameter to replace the original one. As a result, either a limp-home condition is possible or else the nominal function can be assured but under slightly limited conditions. Simple examples for such a calculated parameter are vehicle speed (considering the gear and the synchronous speed, or the antilock braking information, or the data of the navigation system), motor temperature (considering the outside temperature and the operating time), and the amount of remaining fuel (considering the last actual fuel content and the calculated consumption).

Providing Guidelines. As mentioned earlier, a diagnostic system has to provide clear information to the driver in case of a defect. A global warning indication is not sufficient. The driver needs to learn the extent of the defect and its consequences by appropriate text, graphics, or synthetic voice. In addition, the driver needs to be told the steps that have to be taken (e.g., "refill cooling water," "minimum speed to the next service station, risk of engine breakdown," "stop, brake system out of order").[10]

The diagnostic monitoring system can also be used, if there is no service station nearby, as a substitutional off-board system. The defect is then localized by an interactive working together of the indicating system and an appropriate input medium.

External Diagnostic Access. For off-board diagnosis, the diagnostic system of the vehicle has to provide a standardized access to all relevant components, control units, and stored information. This standardized access might also be used by the vehicle manufacturer, legislator, application engineer, and the end-of-the-line programmer. The access itself has to be controlled with the help of an appropriate mechanism to prevent possible abuse.[11]

Logbook Function. The control unit or the diagnostic computer of the vehicle is supposed to store every repair that has been carried out in the format of a logbook. It should contain the time and name of the workshop, every exchanged and newly installed element, every inspection carried out, and so forth.

22.2 ON-BOARD DIAGNOSTICS

The more complex automobiles became, the greater the number of electronic systems and the more difficult became the registration of the actual condition in case of a defect. To reach the necessary measuring points, many connecting cables and adapters were required. In addition, much data about the different systems and their working together was needed to allow a system-specific diagnosis. Modern electronics with self-diagnosis supports the service mechanic by registrating the actual values, comparing them with the nominal values, and diagnosing faults that are stored for repair purposes. Actually, the internal functions are checked whenever an ECU is turned on.

First, the checksum of the program memory is checked together with its function and the correct version. Then a read and write test of the RAM cells is performed. Special peripheral elements (e.g., AD converters) are also checked within this test cycle. During the entire operating time of the vehicle, the ECUs are constantly supervising the sensors they are connected to. With the help of an adequate interpretation of the hardware, controllers are able to determine whether a sensor has a short circuit to ground or battery voltage, or if a cable to the sensor is interrupted. By comparing the measured values and the stored technical data, a controller is able to determine whether the measured values exceed the limits, drift away, or are still within the tolerable limits. The combination of information provided by other sensors allows the monitoring for plausibleness of the measured values.

Sensors are tested similarly to the way actuators are monitored for short circuits or interruptions of cables. The check is carried out by measuring the electric current or reading the diagnostic output of intelligent driver circuits. The function of an actuator under certain conditions can be tested by powering the actuator and observing the corresponding reaction of the system. If discrepancies to the nominal values are diagnosed, the information is stored in an internal fault memory together with relevant outside parameters, e.g., the motor temperature or the engine speed. Thus, defects that appear once or under certain conditions can be diagnosed. If a fault occurs only once during several journeys, it is deleted. The fault memory can be read later in the workshop and provides valuable information for the mechanic.

In case of a detected defective sensor, the measured values are replaced by nominal values or an alternative value is formed using the information of other sensors to provide at least a limp-home function.

With the help of an appropriate interface, a tester can communicate with the ECUs, read the fault memory and the measured values, and send signals to the actuators. In order to be able to use self-diagnosis as universally as possible, manufacturers aim at the standardization of the interface and the determination of appropriate protocols for data exchange.

Another task of self-diagnosis is the indication of a defect to the driver. Faults are mostly indicated by one or more warning lights on the dashboard. Modern developments aim at more comprehensive information using displays for text and graphics, which provide priority-controlled information for the driver. Legal regulations concerning exhaust-gas gave rise to an essential extension of self diagnosis. The control units have to be able to control all exhaust-relevant functions and components and to clearly indicate a defective function or the exceeding of the permissible exhaust limits. Some of the demanded functions require an enormous amount of additional instructions; therefore, the extent of self-diagnosis already reaches up to 40 percent of the entire software of the control unit.

22.3 OFF-BOARD DIAGNOSTICS

The continual increase in the use of electronics within the broad range of different vehicles represents one of the major challenges for customer service and workshop operations. Modern diagnosis and information systems must cope with this challenge and manufacturers of test equipments must provide instruments that are flexible and easy to handle. Quick and reliable fault diagnosis in modern vehicles requires extensive technical knowledge, detailed vehicle information, and up-to-date testing systems.

Due to the different demands of the service providers, there are many different test equipments on the market. They can be subdivided into two main categories: handheld or portable instruments and stationary equipments. Handheld instruments are commonly used for the control of engine functions like ignition or fuel injection and the request of error codes of the electronic control units (ECUs). Stationary test equipment, on the other hand, covers the whole range of function and performance checks of the engine, gear, brakes, chassis, and exhaust monitoring.

Most of the common testers are used for the diagnosis of the engine. The Bosch MOT 250, for example, offers the following functions:

- Engine speed by means of the top dead center (TDC) transmitter, cylinder 1 or terminal 1 signal
- Ignition timing with TDC sensor or stroboscope
- Dwell angle in percent, degrees, or dwell time
- On/off-ratio in percent
- Injection timing or other times measured at the valve or other suitable measuring points
- Electric cylinder balance in absolute or relative terms
- Voltage to ground or floating potential including lambda-sensor voltages or dynamic voltage at terminal 1
- Current with two test adapters for maximum 20 A and 600 A
- Resistances from milliohms to megohms
- Temperature with oil-temperature sensor

For most variables, a maximum of four blocks of measured variables can be stored and recalled one after the other. Twelve blocks can be stored for the cylinder balance function. A digital storage oscilloscope records and stores up to 32 oscillograms of ignition voltages, alternator ripple, and current or voltage transients in the electric or electronic systems. Two RS232 interfaces are provided for documentation purposes and data exchange.

For repair, service, and maintenance, many different manuals and microfiches are stored in the workshops. It is a time-consuming task to collect all the necessary information, especially when vehicles of different makes have to be repaired. To avoid unnecessary paper, information and communication systems among workshop, dealer, and manufacturer are built up. The corresponding manuals have to be standardized and distributed on electronic data processing media, preferably on CD-ROMs.

Every garage or workshop, equipped with the appropriate data system (basically a tester connected to a PC), will receive servicing aids and updates via telephone line or by periodic receipt of updated CDs. A committee of the SAE is preparing rules for the standardization of manuals. There are already published draft international standards (DIS) for terms and definitions (J1930) used in the manuals, for diagnostic codes/messages (J2012), or electronic access/service information (J2008) (see the following). Most of the available test equipment is capable of storing operator manuals within its memory and offers menu-guided assistance to the service personnel. Automatic vehicle and component identification by the tester and the availability of corresponding data at the workbench eases troubleshooting and repairs.

22.4 LEGISLATION AND STANDARDIZATION

22.4.1 CARB, EPA, OBD II

The following is an abstract of the California Air Resource Board (CARB) Regulations for On-Board-Diagnosis two(OBDII):

> All 1994 and subsequent model-year passenger cars, light-duty trucks, and medium-duty vehicles shall be equipped with a malfunction indicator light (MIL) located on the instrument panel that will automatically inform the vehicle operator in the event of a malfunction of any power train component which can affect emission and which provide input to, or receive output from, the on-board computer(s) or of the malfunction of the on-board computer(s) itself. The MIL shall not be used for any other purpose.
>
> ....
>
> All 1994 and subsequent model-year passenger cars, light-duty trucks, and medium-duty vehicles required to have MIL pursuant to paragraph above shall also be equipped with an on-board diagnostic system capable of identifying the likely area of the malfunction by means of fault codes stored in the computer memory. These vehicles shall be equipped with a standardized electrical connector to provide access to the stored fault codes . . . Starting with model-year 1995, manufacturers of non-complying systems shall be subject to fines pursuant to section 43016 of the California Health and Safety Code for each deficiency identified, after the second, in a vehicle model. For the third deficiency and every deficiency thereafter identified in a vehicle model, the fines shall be in the amount of $50 per deficiency per vehicle for non-compliance with any of the monitoring requirements . . .

Systems to Be Monitored

OBD II Functions. These include catalyst monitoring, misfire monitoring, evaporative system monitoring, secondary air system monitoring, fuel systems monitoring, oxygen sensor monitoring, exhaust-gas-recirculation (EGR) system monitoring, and comprehensive component monitoring.

Catalyst. Legal requirements (CARB excerpt): "The diagnostic system shall individually monitor the front catalyst or catalysts which receive untreated engine out exhaust-gas for malfunction. A catalyst is regarded as malfunctioning when the average hydrocarbon conversion efficiency falls between 50 and 60 percent."

Technical solution: In addition to the oxygen sensor upstream the catalyst, another sensor is mounted downstream.

A properly working catalyst shows a storage effect so that the oscillation of the lambda-controller appears damped at the downstream lambda probe. A worn-out catalyst has a reduced damping effect and the signals of up- and downstream sensors are equivalent.

The ratio of the signal amplitudes is a measure of the conversion efficiency. The electronic system that controls the fuel injection monitors these signals together with other relevant engine conditions to derive the catalyst efficiency.

Misfire Detection. Legal requirements (CARB excerpt): "To avoid catalyst damage, the diagnostic system shall monitor engine misfire and identify the specific cylinder experiencing misfire."

Technical solution: Misfire can be caused by worn-out spark plugs or defective electrical wiring. Unburned fuel reaches the catalyst and may destroy it by overheating. Even the least amount of misfire rates influences the emission and therefore single misfire events must be detected.

The speed of the engine is measured very precisely. In case of misfire, the momentum, which is normally produced by the combustion, is lacking. Thus abnormal variations of speed-changes at steady state conditions may be considered as misfire. To distinguish clearly between misfire and other malfunctions, complicated calculations have to be carried out.

If a certain percentage of misfires within 200 or 1000 revolutions is detected, a fault code is stored in the control unit and the fault is indicated to the driver.

Oxygen Sensor. Legal requirements (CARB excerpt): "The diagnostic system shall monitor the output voltage, the response rate, and any other parameter which can affect emission and all fuel control oxygen sensors for malfunction."

Technical solution: The control unit has a special input circuit for detecting shorts or breaks and monitors the switching frequency of the control loop.

By means of a second lambda probe behind the catalyst, it is possible to monitor the lambda probe in front of the catalyst for its correct position. A lambda probe which is subject to an increased temperature for extensive periods may react slower on variations of the air/fuel mixture, thus increasing the period of the lambda-probe regulation. The diagnostic system of the control unit controls the regular frequency and indicates slow sensors to the driver by means of a warning light.

Heated sensors are monitored for correct heater current and voltage by hardware means within the control unit.

Evaporative System. Legal requirements (CARB excerpt): "The diagnostic system shall control the air flow of the complete evaporative system. In addition, the diagnostic system shall also monitor the complete evaporative system for the emission of HC vapor into the atmosphere by performing a pressure or vacuum check of the complete evaporative system. From time to time, manufacturers may occasionally turn off the evaporative purge system in order to carry out a check."

Technical solution: At idle position, the canister purge valve is activated and the lambda controller is monitored for its reaction. For leak detection of the evaporative system, the output to the active carbon filter is shut off and the canister pressure is decreased to about –1.5 kPa. Then the complete system is turned off and the pressure within the canister is monitored for variation with time. The pressure gradient, together with other parameters like the amount of fuel, may indicate possible leaks.

Secondary Air System. Legal requirements: "Any vehicle equipped with any form of a secondary air delivery system shall have the diagnostic system monitor the proper functioning of the secondary air delivery system and any air switching valve."

Technical solution: The lambda controller is monitored for correlated deviations when the secondary air flow is changed.

Fuel System. Legal requirements: "The diagnostic system shall monitor the fuel delivery system for its ability to provide compliance with emission standards."

Deviations of the stochiometric ratio which last for a longer time are stored within the adaptive mixture controller. If these values exceed defined limits, components of the fuel system obviously do not correspond to the specification.

Exhaust-Gas Recirculation (EGR) System. Legal requirement: "The diagnostic system shall monitor the EGR system on vehicles for low and high flow rate malfunctions."

Technical solution: (1) At overrun, the fuel is cut off and the EGR valve is completely opened. The flow of exhaust gas to the manifold raises the manifold pressure, which is recorded and allows statements about the function of the EGR valve. (2) Another possibility is to control the increase of the manifold intake temperature when the EGR valve is opened.

In a conclusion to the previously described OBD II requirements and technical solutions, we can define the following four quality demands for electronic control units:

- Guarantee for exhaust-gas-relevant components with repair costs >$300 for seven years or 70,000 miles for all 1990 and subsequent model-year vehicles (CARB).
- Guarantee for exhaust-gas-relevant components with repair costs >$200 for eight years or 80,000 miles for all 1994 and subsequent model-year vehicles (EPA/Clean Air Act).
- Guarantee protocols in case of a reclamation rate of exhaust-gas-relevant components higher than 1 percent (CARB).
- Recall of vehicles in case of a calculated reclamation rate of more than 20,000 ppm within a period of five years/50,000 miles (CARB).

22.4.2 International Standardizations

Because of the manifold requirements on modern diagnostics, the national and international standardization committees soon came to the conclusion that with the help of appropriate and, if possible, international agreements about protocols, connectors, tools and auxiliaries, the process of diagnosis can be standardized, thus reducing time and costs.

Figure 22.4 shows how, in a standardized graphic, control units and diagnostic tools are connected and diagnostic data exchanged.

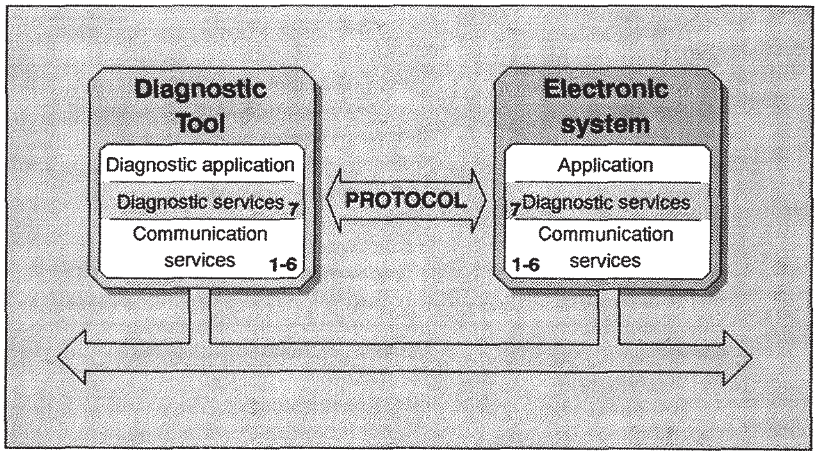


FIGURE 22.4 Standardized testing link according to the OSI model.

For data exchange, electronic systems are structured and described according to a seven-layer model (OSI model, open system interconnection) developed by the ISO (International Standardization Organization). Every unit connected to a data network can be structured with the help of this model—control units as well as diagnostic tools.

The diagnostic services that the controller may use during the diagnostic process are regulated in the seventh layer. Diagnostic service means definite instructions, which actuate determined and standardized diagnostic procedures, e.g. "start diagnostic session," "read diagnostic trouble codes," "read freeze frame data," and so on. There are different sequences of bits and bytes code for such instructions. On the hardware level (plugs, cables, potentials), the sequences are finally transmitted from unit to unit. The ISO and the SAE (Society of Automotive Engineers) developed corresponding standards in the area of service definition

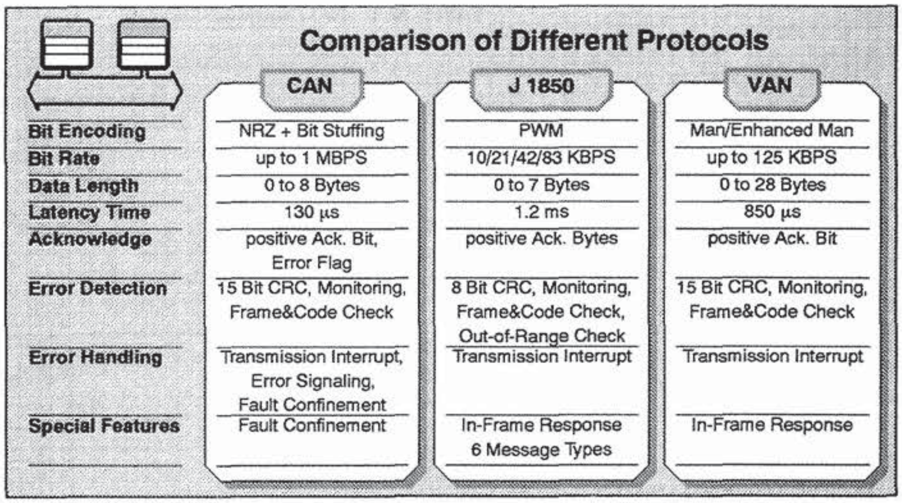
TABLE 22.1 ISO Diagnostic Services

Diagnostic management
StartDiagnosticSession
StopDiagnosticSession
SecurityAccess
TesterPresent
EcuReset
ReadEcuIdentification
DisableNormalMessageTransmission
EnableNormalMessageTransmission
Data transmission
ReadDataByLocalIdentifier
ReadDataByGlobalIdentifier
ReadMemoryByAddress
WriteDataByLocalIdentifier
WriteDataByGlobalIdentifier
WriteMemoryByAddress
SetDataRates
StopRepeatedDataTransmission
Input/output control
InputOutputControlByGlobalIdentifier
InputOutputControlByLocalIdentifier
Stored data transmission
ReadNumberOfDiagnosticTroubleCodes
ReadDiagnosticTroubleCode
ReadDiagnosticTroubleCodesByStatus
ReadStatusOfDiagnosticTroubleCodes
ReadFreezeFrameData
ClearDiagnosticInformation
Remote activation of routine
StartRoutineByLocalIdentifier
StartRoutineByAddress
StopRoutineByLocalIdentifier
StopRoutineByAddress
RequestRoutineResultsByLocalIdentifier
RequestRoutineResultsByAddress
Upload download
RequestDownload
RequestUpload
TransferData
RequestTransferExit

as well as in the area of communication. Table 22.1 shows the diagnostic services as proposed by the ISO.

Figure 22.5 presents the determined standards with some essential technical details as developed for the field of communication.

Unfortunately the whole spectrum of available standards has become very complex and difficult to use. The following explanations try to provide a unified system for the existing standards in the area of diagnosis.



Comparison of Different Protocols

	CAN	J 1850	VAN
Bit Encoding	NRZ + Bit Stuffing	PWM	Man/Enhanced Man
Bit Rate	up to 1 MBPS	10/21/42/83 KBPS	up to 125 KBPS
Data Length	0 to 8 Bytes	0 to 7 Bytes	0 to 28 Bytes
Latency Time	130 μs	1.2 ms	850 μs
Acknowledge	positive Ack. Bit, Error Flag	positive Ack. Bytes	positive Ack. Bit
Error Detection	15 Bit CRC, Monitoring, Frame&Code Check	8 Bit CRC, Monitoring, Frame&Code Check, Out-of-Range Check	15 Bit CRC, Monitoring, Frame&Code Check
Error Handling	Transmission Interrupt, Error Signaling, Fault Confinement	Transmission Interrupt	Transmission Interrupt
Special Features	Fault Confinement	In-Frame Response 6 Message Types	In-Frame Response

FIGURE 22.5 In-vehicle networks.

Figure 22.6 shows a general model for diagnostic concepts. The three main levels comprehensively describe the whole area of diagnostics. The three levels are hierarchically structured, closely linked together with flowing transition from one level to the other. Although there are certain similarities between this model and the seven-layer model of the OSI, both models do not correlate.

The upper level comprises the elements, which are essential for the user or generator of diagnostic applications. The term "user" includes the driver, the legislator, the mechanic, and the manufacturer. This upper level can be subdivided into three main fields of activities: user

FIGURE 22.6 Model for diagnostic concept.

interface, strategy, and diagnostic data. Although presented as layers, these activities do not correlate hierarchically, but each is associated with a service or group of services.

The "user interface" describes how information flows between the user and the diagnostic service. This includes a functional description of scan tools, handheld testers, monitoring systems, and so on.

The term "strategies" stands for strategic details, which are essential for the diagnosis or repair of a vehicle, including communication access, diagnostic data and information.

The term "diagnostic data" includes the data that are necessary for the diagnosis itself. The details concerning parameters, trouble codes, and so on are described here.

The intermediate level describes the diagnostic services, defining a set of services and a set of commands for general purpose, which allow the diagnosis of a vehicle. The set of commands is supposed to cover the needs of users concerning repair and maintenance as described by the strategies and diagnostic data.

The lowest level deals with the communication area. It describes every technical detail that is necessary for communication and provides the information about how to start communication (initialization). It also specifies the appropriate Baud rate, the suitable protocol, and the necessary hardware (connector, cable, and so on).

This model offers a general description of the essential fields of diagnostic interest and allows the categorization of all ISO and SAE standardization activities in the three main levels of the diagnostic concept model.

Figures 22.7 and 22.8 are presented in the same graphic form (three-level structure). They provide a summary of the concrete standardization activities of the SAE and ISO. Figure 22.7 shows the existing standards or drafts of automotive diagnosis for general purpose.

The user interface for general purposes is undefined. The SAE J2186 (Data Link Security) and the SAE J2008 (Electronic Access/Service Information) are strategic documents, though most strategies are not standardized and diagnostic data is described in documents SAE J2012 (Diagnostic Codes and Messages) and SAE J2190-2 (Parameters—in preparation).

On the level of diagnostic services, the standardization activities can be divided in two fields called service definition and service implementation. The term "service definition" describes a set of useful diagnostic services, which enable the user to run a diagnostic session

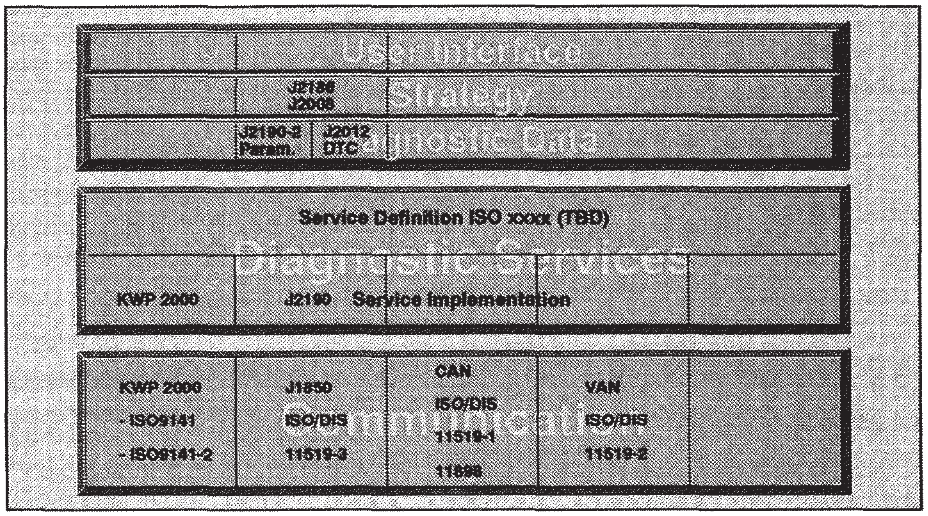


FIGURE 22.7 Realization for general automotive diagnosis.

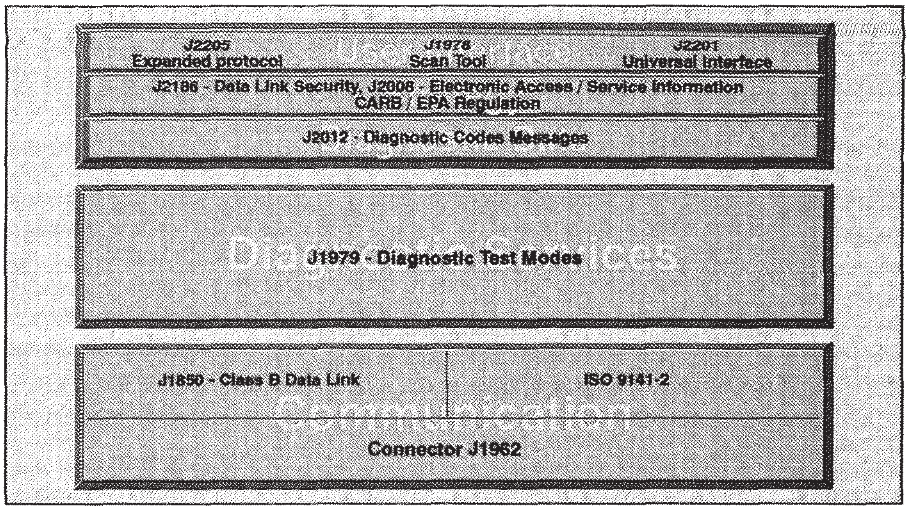


FIGURE 22.8 Realization for CARB and EPA requirements.

independently of the knowledge of any technical detail in the communication area as described in the level below.

This set of diagnostic services for general purpose can now be mapped on different protocols. Any bit representation of the different services can be built up. This is called service implementation. At the moment, there are two implementations available, the SAE J2190 (Diagnostic Test Modes) and the KWP 2000 (ISO Draft: Keyword Protocol 2000). The lowest level (the Communication level) shows the standardized details of communication such as the data formats and the physical layers; e.g., the KWP 2000 uses the physical layer of ISO 9141 or ISO 9141-2, the SAE J2190 uses the SAE J1850 Class B network (ISO/DIS 11519-3). It is shown that communication can also be built up with a CAN or a VAN network.

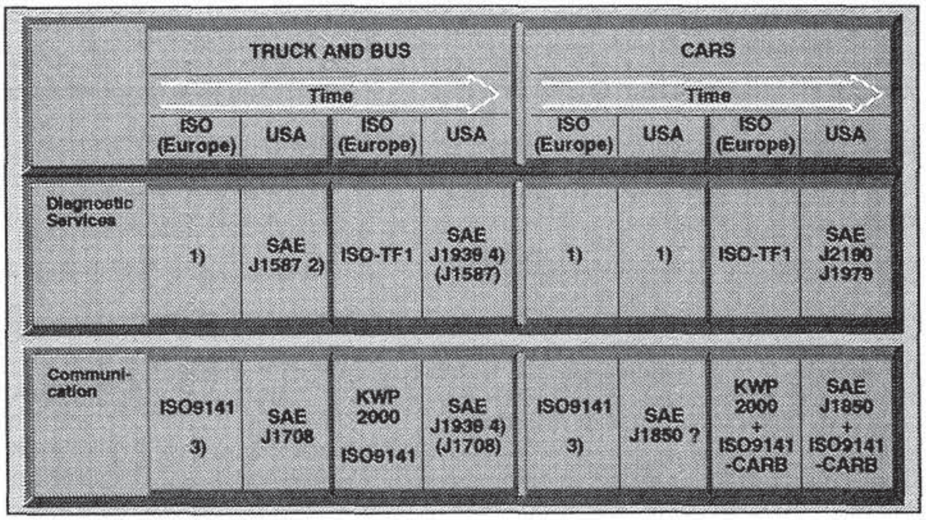
Figure 22.8 shows the standardization activities for the special requirements of the CARB and the EPA using the same three-level-concept.

The user interface, a generic scan tool, is standardized within the SAE J1978, including the SAE J2205 (Expanded Diagnostic Protocol) and the SAE J2201 (Universal Interface). Some aspects of the diagnostic strategy are described in the SAE J2186 (Data Link Security), the SAE J2008 (Electronic Access/Service Information), and some in the regulations. The diagnostic data is described in the document SAE J2012—Diagnostic Codes and Messages.

The level of diagnostic services defines one SAE J1979 standard—Diagnostic Test Modes. This standard is a closely linked combination of a service definition and a service implementation (referring to the SAE as "modes").

In the field of communication, the possible networks are described in the SAE J1850 (Class B Data Network) and the ISO 9141-2 (CARB Requirements for Interchange of Digital Information).

A standard for the physical connector (SAE J1962) has also been developed. Figure 22.9 shows the status of diagnostic standards for trucks and buses and for passenger cars in Europe and in the United States. It shows also a time schedule for the development of standards. A comparison of the communication and diagnostic services levels has already been realized. The titles of the different SAE and ISO numbers are shown in Tables 22.2 and 22.3, where all ISO and SAE papers, relevant for diagnostics, are listed. Table 22.3 offers a detailed list of trucks and bus activities (J1939).



	TRUCK AND BUS				CARS			
	Time				Time			
	ISO (Europe)	USA	ISO (Europe)	USA	ISO (Europe)	USA	ISO (Europe)	USA
Diagnostic Services	1)	SAE J1587 2)	ISO-TF1	SAE J1939 4) (J1587)	1)	1)	ISO-TF1	SAE J2190 J1979
Communication	ISO9141 3)	SAE J1708	KWP 2000 ISO9141	SAE J1939 4) (J1708)	ISO9141 3)	SAE J1850 ?	KWP 2000 + ISO9141 -CARB	SAE J1850 + ISO9141 -CARB

FIGURE 22.9 Status of diagnostic standards.

22.5 FUTURE DIAGNOSTIC CONCEPTS

As yet, most vehicle manufacturers have installed a diagnostic connector in the engine compartment in order to offer essential electric signals for diagnostic purposes. Due to the multitude of different equipments and philosophies of car makers, the connectors have different shapes and contact arrangements. Therefore, a workshop has to keep a lot of different expensive cables and adaptors in store.

For future diagnostic systems, the connection between control unit and vehicle is supposed to be realized with the help of a standardized connector. A connector for the legally demanded exhaust-gas diagnosis was defined by an SAE draft (J1962), concerning form, contact arrangement, and installation position. (Fig. 22.10)

With this connector and a so-called generic scan tool, anyone is able to read the fault-memory in regard to exhaust-gas-relevant defects. The interconnection of the control units allows the access to the entire electronics of the vehicle.

The necessary protocols are partly defined and developed further in standardization committees of the ISO. At the moment, there are two actual standards available:

1. *ISO 9141-2:* Determination of the requirements on hardware and communication protocols. The requirements on hardware are essentially determined by the maximum Baud rate of data transfer and the maximum number of control units simultaneously connected with the diagnostic cable.

   Communication is started by means of a trigger address, and is followed by a synchronization byte of the control unit(s), which is necessary for the automatic setting of the Baud rate. The trigger address calls either a particular control unit or a function, that may also address several control units.

   After transmission of the synchronization byte, the control unit waits for the tester to set the Baud rate, then sends two key-bytes that inform the tester about the suitable data transfer protocol. The tester responds with the last inverted key-byte, in order to confirm the correct receipt. The connection between tester and control unit is now established.

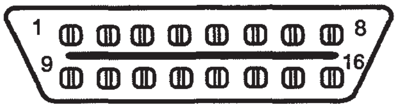
TABLE 22.2 ISO and SAE Documents

ISO 9141		Road Vehicles—Diagnostic System—Requirements for Interchange of Digital Information
ISO/DIS 9141-2		Road Vehicles—Diagnostic System—Part 2: CARB Requirements for Interchange of Digital Information
ISO/DIS 11519-1		Road Vehicles—Low-Speed Serial Data Communication—Part 1: General Definitions
ISO/DIS 11519-2		Road Vehicles—Low-Speed Serial Data Communication—Part 2: Low Speed Controller Area Network (CAN)
ISO/DIS 11519-3		Road Vehicles—Low-Speed Serial Data Communication—Part 3: Vehicle Area Network (VAN)
ISO/DIS 11519-4		Road Vehicles—Low-Speed Serial Data Communication—Part 4: Class B Data Communication Network Interface (J1850)
ISO/DIS 11898		Road Vehicles—Interchange of Digital Information—Controller Area Network (CAN) for High-Speed Communication
ISO/WD 14229		Diagnostic Systems—Diagnostic Services Specification
ISO/WD 14230		Diagnostic Systems—Keyword Protocol 2000 (3 parts: 1: Physical Layer, 2: Data Link Layer, 3: Implementation)
SAE J 1213/1	IR	Glossary of Vehicle Networks for Multiplexing and Data Communications
SAE J 1583	IR	Controller Area Network (CAN), An In-Vehicle Serial Communication Protocol
SAE J 1587	RP	Joint SAE/TMC Electronic Data Interchange Between Microcomputer Systems in Heavy-Duty Vehicle Applications
SAE J 1699	RP	J 1850 Verification Test Procedures
SAE J 1708	RP	Serial Data Communications Between Microcomputer Systems in Heavy-Duty Vehicle Application
SAE J 1724		Vehicle Electronic Identification (New Task Force)
SAE J 1850	RP	Class B Data Communication Network Interface
SAE J 1930	RP	Electrical/Electronic Systems Diagnostic Terms, Definitions, Abreviations and Acronyms
SAE J1939/xx		Truck + Bus, Details next page
SAE J 1962	RP	Diagnostic Connector
SAE J 1978	RP	OBD II Scan Tool
SAE J 1979	RP	E/E Diagnostic Test Modes
SAE J 2008	RP	Electronic Access/Service Information
SAE J 2012	RP	Diagnostic Trouble Code Definitions
SAE J 2037	IR	Off-Board Diagnostic Message Formats
SAE J 2054	IR	E/E Diagnostic Data Communications
SAE J 2056/1	RP	Class C Application Requirement Considerations (Part 2: IR: Survey of Known Protocols, Part 3: IR: Selection of Transmission Media)
SAE J 2057/1	IR	Class A Application/Definition (Part 3: IR: Class A Multiplexing Sensors, Part 4: IR: Class A Multiplexing Architecture Strategies)
SAE J 2106	IR	Token Slot Network for Automotive Control
SAE J 2112	IR	Diagnostic Technician Questionnaire Summary
SAE J 2178	RP	Class B Data Communication Network Messages (Part 1: Detailed Header Formats and Physical Address Assignments, Part 2: Data Parameter Definitions, Part 3: Frame Ids for Single Byte Forms of Headers, Part 4: Message Definition for Three Byte Headers)
SAE J 2186	RP	E/E Data Link Security
SAE J 2190	RP	Enhanced E/E Diagnostic Test Modes
SAE J 2201	RP	Universal Interface for OBD II Scan Tool
SAE J 2205	RP	Diagnostic Specific Functionality Protocol
SAE J 2216	RP	Application of the Clean Air Act Amendment of 1990 (Section 207, Paragraph M5)

RP = Recommended Practice, IR = Information Report

TABLE 22.3 SAE Truck and Bus Documents

SAE J 1939	RP	Serial Control and Communication Vehicle Network (Class C)
SAE J 1939/01		Truck and Bus Control and Communication Vehicle Network (Class C)
SAE J 1939/02		Agricultural Equipment Control and Communication Network
SAE J 1939/1x		Physical Layer, x refers to a specific version
SAE J 1939/11		Physical Layer, 250 kBaud, Twisted Shielded Pair
SAE J 1939/12		Physical Layer, 125 kBaud, Twisted Pair
SAE J 1939/13		Physical Layer, 250 kBaud, Twisted Pair with Ground
SAE J 1939/14		Physical Layer, 1 MBaud, Fiber Optic
SAE J 1939/15		Physical Layer, 50 kBaud, German Agricultural
SAE J 1939/21		CAN 29 Bit Identifier Data Link Layer
SAE J 1939/3x		Network Layer, x refers to a specific version
SAE J 1939/31		Truck + Bus Network Layer
SAE J 1939/4x		Transport Layer, x refers to a specific version
SAE J 1939/5x		Session Layer, x refers to a specific version
SAE J 1939/6x		Presentation Layer, x refers to a specific version
SAE J 1939/7x		Application Layer, x refers to a specific version
SAE J 1939/71		Truck, Bus, Agricultural and Construction Equipment Application Layer
SAE J 1939/72		Virtual Terminal
SAE J 1939/73		Application Layer—Diagnostics
SAE J 1939/81		Network Management
SAE J 1939/??		Tractor-Trailer-Interface



PIN #	Assignment
1	discretionary
2	BUS + Line of SAE J1850
3	discretionary
4	Chassis Ground
5	Signal Ground
6	discretionary
7	K Line of ISO 9141-2
8	discretionary
9	discretionary
10	BUS - Line of SAE J1850
11	discretionary
12	discretionary
13	discretionary
14	discretionary
15	L Line of ISO 9141-2
16	Unswitched Vehicle Battery Positive

Note: Assignment of pins 1, 3, 6, 8, 9, 11, 12, 13, and 14 is left to the discretion of the vehicle manufacturer

FIGURE 22.10 SAE J1962 diagnostic connector.

2. *Interface according to the SAE J1850 (Class B Data Communication Network Interface):* The SAE J1850 defines means and methods for serial data exchange for automotive application at the physical and data link layer of the OSI model. It is used for networked systems and for diagnostic purposes.

   Two implementations are characterized: pulse-width modulation (PWM) at 41.6 kbps transmitted on twisted pair wires, and variable pulse-width modulation (VPM) at 10.4 kbps, transmitted on a single wire.[12]

A generic scan tool, as mentioned, therefore, has to handle the three different interfaces.

A new protocol, *Keyword 2000,* is prepared by the ISO committees. It is supposed to combine the protocols that have been used up to now.

With the introduction of more and more diagnostic functions and networked systems in the vehicle, the functional structure will be modified (Fig. 22.11).

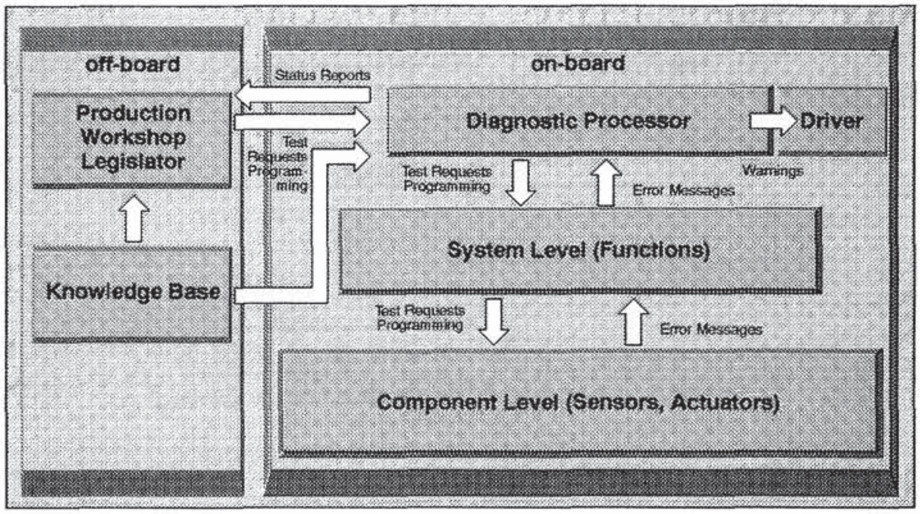


FIGURE 22.11 Logical structure for future diagnosis.

A diagnostic processor on top of a hierarchical structure of functions has access to every system via the network. It can request status information of the functions of the levels below, or of the sensors and actuators, and receives warning messages if problems are detected by the self-diagnosis of the different subsystems. The diagnostic processor serves as a man-machine interface to the driver and as a gate to the outside. It is the only secure access to the entire system of the vehicle.

GLOSSARY

CAN Controller Area Network (standardized protocol developed by Bosch for networked systems).

CARB California Air Resources Board.

CD-ROM Compact disk read only memory, a data storage medium.

DIS Draft International Standard.

ECU Electronic control unit.

EGR Exhaust-gas recirculation.

EPA Environmental Protection Agency.

Freeze frame Faults stored together with various related parameters.

HC Hydrocarbon.

ISO International Standardization Organization.

ISO 9141-2 Standardized protocol for data exchange between ECUs and testers.

Lambda controller Electronic system for controlling the air/fuel ratio.

Lambda sensor A sensor for air/fuel ratio (oxygen sensor).

MIL Malfunction indicator lamp (indicates emission-related faults to the driver).

OBDII On Board Diagnostics II.

Off-board diagnosis Diagnosis performed by means outside a vehicle.

On-board diagnosis Diagnosis performed by means within a vehicle.

OSI Open System Interconnection.

PC Personal computer.

PWM Pulse-width modulation.

RS 232 Standardized data link (hardware).

Scan tool Small tester that can be connected to the diagnostic connector to interrogate emission-related fault codes.

SAE Society of Automotive Engineers.

TDC Top dead center.

Terminal 1 Connection to a signal related to ignition timing.

VAN Vehicle Area Network (French proposal for network protocol).

VPM Variable pulse-width modulation.

REFERENCES

1. W. Bremer and Heintz, F., "Was bieten die Karosserie-Zentral-Elektronik mit Kabelbaum-Multiplex und die On-Board-Diagnose?," (Advantages of a central body-electronic with multiplex and on-board diagnosis), Bosch Customer Information, 1977.
2. W. Bremer, Heintz, F., and Hugel, R., "Diagnosis in Networked Systems," *International Conference Automotive Diagnostics,* London, 1990.
3. K. Dieterich and Unruh, J., "CAN—A Bus System for Serial Data Transfer," *23. FISITA Congress "The promise of new technology in the automotive industry,"* 1990.
4. T. Goelzer and Leonhard, R., "A new Architecture for Car Electronics", *International Symposium Vehicle Electronics Integration ATA-EL 91,* 1991.

5. W. Kremer and Kaminski, D., "Integrated Vehicle Electronics with CAN," *23. FISITA Congress "The promise of new technology in the automotive industry,"* 1990.
6. T. Kühner and Kaminski, D., "Structuring of vehicle electronics in cars for an effective on-board-diagnosis system," *VDI Reports "Electronic in Vehicles,"* Nr. 819, 1990.
7. E. Hipp, Jung, C., and Morizur, P., "On Board Diagnosis as the Central Interface for Modern Vehicle Electronics," *International Symposium Vehicle Electronics Integration ATA-EL 91,* 1991.
8. W. Botzenhardt, Litschel, M., and Unruh, J., "Bussysteme für Kfz-Sterugeräte" (Bus systems for vehicle electronic control units), *VDI Reports "Electronic in Vehicles,"* Nr. 612, 1986.
9. B. Przbylla, "Eigendiagnose von elektronischen Steuergeräten im Kraftfahrzeug" (Self-diagnosis of electronic control units in the vehicle), *VDI Reports "Electronic in Vehicles,"* Nr. 612, 1986.
10. W. Bremer, "Möglichkeiten der fahrzeugfesten Überwachung mit Kabelbaum-Multiplex," (Possibilities of On-Board Diagnostics with Multiplex), *Status-seminar of the German Ministry of Research and Technology,* 1978.
11. D. Nemec, "Möglichkeiten komfortabler Testgeräte zur Auswertung der Eigendiagnose von Steuergeräten" (Possibilities of comfortable test equipment for the evaluation of selfdiagnostic data of control systems), *VDI Reports "Electronic in Vehicles,"* Nr. 687, 1988.
12. A. W. Millsap, Lowden, M. T., Folkerts, M. A., Unruh, J., and Dais, S., "Mapping J1850 Messages into CAN Version 2.0," *SAE International Congress,* Paper 930437, 1993.

ABOUT THE AUTHORS

WOLFGANG BREMER studied electrical communication techniques from 1962 to 1968 at the University of Karlsruhe (Germany). Afterwards he was engaged with Siemens AG in Karlsruhe in the development of high-precision electronic balances. Since 1970, he has been working in the advanced engineering department of measurement and information techniques of Robert Bosch GmbH and is responsible for the development of serial communication and diagnostic systems in the chassis area of vehicles. He is working in several ISO and SAE committees and working groups in the field of diagnosis.

FRIEDER HEINTZ studied electrical communication techniques from 1954 to 1959 at the University of Karlsruhe. Afterwards he was engaged in research and development of process-control computers with Siemens AG in Karlsruhe, Munich, and New York. Since 1969, he has been the head of the advanced engineering department for measurement and information techniques at Robert Bosch GmbH. For 20 years, he has managed national and international working groups for diagnosis and serial data transfer in vehicles within the ISO (International Organization for Standardization).

ROBERT HUGEL studied physics at the University of Karlsruhe from 1967 to 1972. Then he was engaged for six years in the development, production, and sales of infrared spectrometers, two years in the development of laser-based measuring systems, and the last 13 years in developing various automotive electronics at Robert Bosch in the advanced engineering department for measurement and information techniques.

P · A · R · T · 5

SAFETY, CONVENIENCE, ENTERTAINMENT, AND OTHER SYSTEMS

CHAPTER 23
PASSENGER SAFETY AND CONVENIENCE

Bernhard K. Mattes
Electrical Engineer
Robert Bosch GmbH

23.1 PASSENGER SAFETY SYSTEMS

Electronically controlled passenger safety systems, such as air bags, seat belt tensioner, and rollover sensor systems, help to avoid injuries or to reduce injury severity in an accident. If the passenger compartment cell is not too heavily deformed, then these systems are effective in minimizing the acceleration or load forces acting on the vehicle occupants. During a rollover accident, the rollover protection system helps provide the necessary survival space. In an open vehicle, a rollover bar, as well as protection elements such as extending headrests behind the rear seats, considerably increase occupant safety.

Such occupant protection systems are also classified as passive restraint systems, since the protective function is independent of any active contribution by the passengers.

23.1.1 Air Bag and Seat Belt Tensioner Sensor Systems: Principle Sensor Function and Timing Sequences

The functional components of electronically controlled air bag and seat belt tensioner initiation consist of crash detection sensors, signal processing, squib triggering, and system monitoring. Whereas a vehicle in a crash situation is rapidly decelerated by contact with an external obstacle, the occupants continue to move forward until restrained or contact within the vehicle occurs (steering wheel, windshield, dashboard, or other rigid structural parts). Air bags and seat belt tensioners are designed to reduce impacts of passengers within the vehicle.

The igniter (squib) activates the pyrotechnical inflator used for both air bag deployment and seat belt tightening. The sensor range comprises an angular sensitivity of ±30° to the longitudinal axis. Normally, electronic sensor units contain one or more accelerometers with main sensitivity along the longitudinal axis.

Crash conditions demand a timely activation of restraining devices. It is equally important, however, that they are also released only when required.

Depending on the kind of collision (frontal, oblique, offset, pole, underride, etc.) the triggering time must be calculated correctly, so that the allowed forward displacement will not be exceeded by the time the air bag is inflated or the seat belts are tensioned.

In most cases, the allowable forward passenger travel with an air bag system is 12.5 cm (5-inches rule). For seat belt tensioning systems, the acceptable displacement drops to about 1 cm. Approximately 30 ms are required to inflate air bags, and the time required to tension a seat belt with a pyrotechnically activated seat belt retractor is approximately 10 ms. Thus, trigger commands must be given by the time the maximally allowable forward displacement will be reached minus the activation time of the respective restraining device. Effective injury prevention requires timely firing based on detection of acceleration or deceleration crash signals and their processing by the crash discrimination algorithm or electromechanical sensor.

Different crash-sensing schemes are shown in Fig. 23.1.

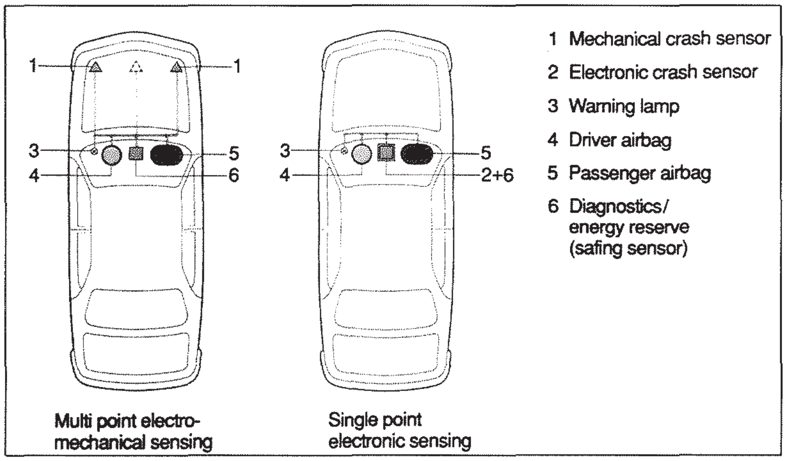


FIGURE 23.1 Crash-sensing schemes.

23.1.2 Multipoint Electromechanical Sensing Systems or Distributed Air Bag Systems

A distributed air bag sensing system or multipoint electromechanical sensing system consists of two to four electromechanical crash discrimination sensors, strategically mounted in the vehicle's crush zone, plus an additional arming sensor mounted in the passenger compartment inside an electronic control unit including diagnostics, energy reserve, and voltage converter.

Sensors. Two typical types of electromechanical sensors are based on the *ball-in-tube* respectively or the *rolamite* principle. The function of the ball-in-tube sensor is shown in Fig. 23.2. A ball held in position by a permanent magnet begins to move if the inertial force acting on it due to the vehicle's deceleration exceeds the magnetic restraining force. A narrow air gap between ball and tube results in viscous air damping and integration behavior of the relative motion. Upon reaching the other end of the tube, the gold-plated ball closes a mechanical switch.

Closure of at least one of the sensors located in the vehicle's crush zone and simultaneous closure of the arming switch inside the passenger compartment directly connect the igniters to battery plus and minus and initiate air bag deployment. The two-stage passenger inflator is ignited in series. A transistor turns on the ignition current for the second stage approximately 10 to 15 ms after the first stage. This delay time provides a softer passenger bag inflation and slower pressure increase inside the passenger compartment.

Such electromechanical spring mass sensors obey second-order differential equations of motion. Parameters like restraining magnetic force, damping factor, mass of the ball and

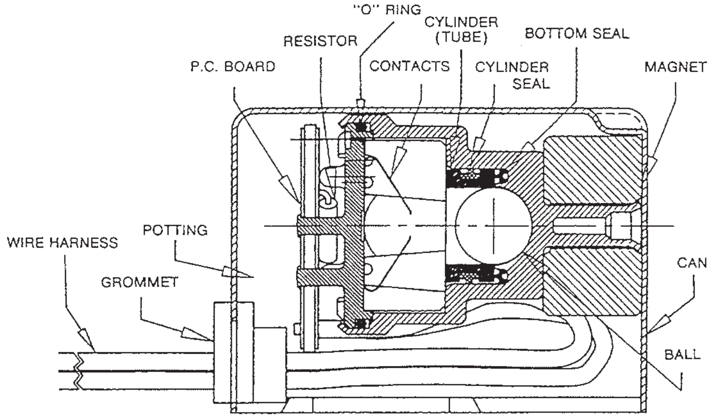


FIGURE 23.2 Ball-in-tube sensor. (*Courtesy Breed Technologies Inc.*)

travel distance determine the sensor's dynamic characteristics and have to be designed according to the vehicle's crash behavior.

The function of the rolamite sensor is as follows. A sheet metal spring band wound around a metal cylinder keeps the roller in its rest position as long as the retaining spring force is bigger than the inertial force acting on the roller cylinder. The rolling of the cylinder over the convex guide is thus kinematically constrained by the guide and the spring to translation along the guide's surface. After having reached or exceeded a specified travel distance, the firing contact is closed.

Normally, the arming sensor inside the diagnostics and energy reserve module is of the same type as the discriminating sensors used in the crush zone.

The advantage of such systems is that the discriminating sensors can be installed in front end positions where high acceleration amplitudes can be sensed in early stages of impact. Disadvantages range from no crash prediction capabilities, no sensor "stuck open" indication in the field, high cost and installation expense, as well as no seat belt tensioner function.

Diagnostics and Energy Reserve Module. A system readiness indicator is legally required for air bag-equipped vehicles (with the exception of all-mechanical systems). This means that air bag system-readiness has to be constantly monitored. Electronic diagnostics and energy reserve modules *periodically* perform the following diagnostic functions:

Monitor. This diagnostic function includes the following checks:

- All ignition loops for too high or too low resistances
- All ignition loops for leakages to battery plus or minus
- External crash discrimination sensors for continuity (there is a diagnostic resistor in parallel to the contact), for cable harness short circuits to battery plus or ground, and for too long contact closure (longer than 1 s)
- Internal arming sensor for continuity and for too long contact closure
- Warning lamp output for short circuits to battery plus or minus and for interruption
- Energy reserve capacitor(s) for correct voltage(s) and capacitance(s)

Control. This diagnostic function includes the following checks:

- Battery supply for too low and too high voltage
- Internal regulated voltage (normally = 5 V) for too low or too high level
- Diagnostics interface for short circuits

After the initialization phase at power turn-on, the following checks are done once:

- RAM, ROM, EEPROM-read check
- Watchdog check
- Output transistor check for 2d passenger stage (if it exists)

Each fault type is characterized by a special fault code and stored into EEPROM after being judged as faulty. There are different modes of fault assessment and different degrees of fault tolerance.

Fault Clock. State-of-the-art diagnostics and energy reserve modules have one fault clock counting the elapsed time of the faults in total. But there are also modules with individual fault clocks for each fault type.

Storage capacity of the fault time counter normally includes 50 to 100 h with a time resolution of 1 to 5 min.

Crash Recorder. The sequence of crash-relevant events like closure of discriminating sensors, arming sensor, battery voltage level, energy reserve voltage, turn-on of power stages (if existing), can be stored in the EEPROM. This can be done in the form of time-discrete "snapshots" of system conditions for approximately 10 to 20 ms before and approximately 30 to 50 ms after deployment. Advanced crash recorders also store acceleration and deceleration values of some time period before and after the firing moment. This is done to get information about impact energy of real-world crashes.

Serial Diagnostic Interface. EEPROM content (type of unit, fault codes, fault time, crash recorder) can be retrieved via a bidirectional diagnostics interface. Initiation of communication, diagnostics concept, and software depend on the car manufacturer's requirements.

Energy Reserve and Voltage Converter. If battery supply is lost in a crash, ignition function and crash recorder storage is maintained by the energy reserve. This backup supply is performed by one or more capacitors (acting as energy reservoir(s). For the ignition loops, the energy reserve is wired or with vehicle battery voltage. For the monitoring circuits part, the energy reserve is switched on to the voltage regulator input in case of too low battery. Survival time of such components ranges between 0.1 and 1 s.

There are systems with an individual energy reserve for each ignition loop as well as one for the monitoring circuit. In such systems, there is no loss of function for the rest of the ignition loops and the crash recorder if one loop gets short circuited in a crash and its energy reserve capacitor is discharged.

A step-up converter keeps the energy reserve(s) charged up to voltages nominally higher than battery voltage (e.g., $V_{ER} = 22$ V to $V_{ER} = 35$ V).

23.1.3 Single-Point Electronic Sensing Systems or Central Air Bag Systems

With single-point electronic sensing systems (or central air bag systems), the electronic control module is located in the passenger compartment. There are no external sensors in the crush zone. The electronic circuits include acceleration sensors, signal-processing algorithms, diagnostics, output stages, energy reserve, and voltage converter.

Historical Evolution. Production of single-point sensing electronic air bag units started in 1980. This first concept used a strain gage acceleration sensor and a mercury switch as "arming" or "safing" sensor. The signal processing was performed by analog integration of the acceleration signal and resulted in a value related to velocity change during the impact, a so-called Δv-value.

If Δv exceeded a vehicle-specifically set threshold and the mercury switch was closed, then the restraining devices were triggered. The first systems comprised three electronic components: sensor, analog circuitry, and diagnostic unit; voltage converter unit; and energy reserve unit.

In the beginning of 1987, another analog integrating system went into production and consisted of two electronic components: sensor and diagnostic unit, and voltage converter and energy reserve unit. This was the first air bag sensor system based on using a piezoelectric accelerometer, and contained a microcontroller only for monitoring functions. Storage of fault codes, fault clock, and crash recorder in an EEPROM were included as well. The unit was designed for triggering the driver air bag, as well as the driver and passenger belt tensioners.

Mid-1987 marked the production start of the first air bag sensor unit with a digital single-point sensing algorithm. With this system, all the functions could be integrated into one box. Figure 23.3 shows a block diagram of this unit.

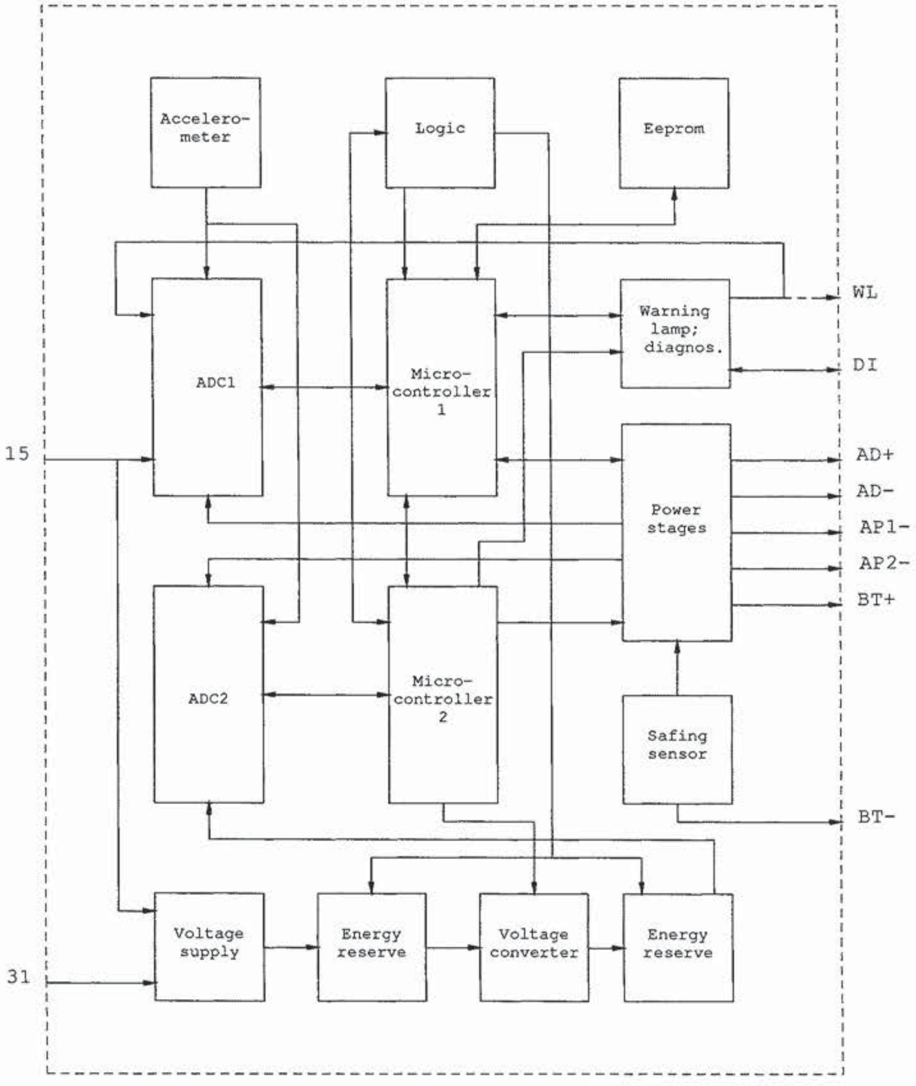


FIGURE 23.3 Block diagram AB 3-ECU.

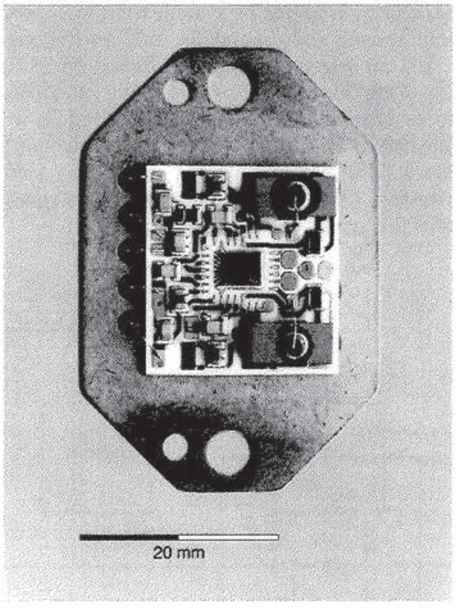


FIGURE 23.4 Piezoelectric accelerometer BSA 6.

For redundancy against inadvertent deployment, this unit was a two-microcontroller design (parallel processing) and contained a mercury switch as arming sensor. In 1989, this mercury switch was replaced by a reed switch.

Current Systems. The key features of a single-point air bag electronic control unit (ECU) are the accelerometer, the digital single-point sensing algorithm, the degree of function availability, the safety against inadvertent release, and the degree of fault tolerance of the system.

Electronic Accelerometers.

Piezoelectric accelerometers. Current accelerometers have been upgraded and are robust against electromagnetic interferences. Dual-channel accelerometers exist that deliver differential mode output signals and offer the possibility to design all-electronic air bag ECUs and remove the need for the untestable mechanical arming sensor. Inverse polarity of the sensor signals allows the unit to distinguish from common mode signals which can result from electrical disturbances. Figure 23.4 shows a dual-channel accelerometer.

Diagnosis of the electrical function of these sensors can be performed after initialization, and, thus, function reliability and system readiness are monitorable. Other piezoelectric accelerometers contain mechanically deflectable sensing elements, (e.g. by piezoelectric actuators), for verification of system integrity.

Micromachined accelerometers. Micromachined accelerometers can be mounted directly on the printed circuit board. The required interface circuitry is included on the same chip with the monolithic capacitive accelerometer. High linearity is guaranteed by closed-loop operation. This means that the movable beam is always electrostatically centered by a feedback voltage proportional to deflection (= acceleration/deceleration). The measurement range is ±50 g. For self-test, the functional beam is electrostatically deflected.

Single-point sensing algorithm. Different types of digital crash-sensing algorithms are currently in use. The sampling rate of the acceleration signal varies between 0.5 and 1 ms.

Mathematical manipulation of the crash signal (differentiation, multiplication, integration) and release threshold variation by software allows an early discrimination of different impact types (frontal, oblique, offset, pole, underride). Furthermore, application of the appropriate prediction model for forward displacement and determination of the correct trigger point is possible.

Digital sensing algorithms increase the possibility both to detect problem crashes such as underrides with perceived damage, and to distinguish between deployment and nondeployment impacts. This shows a distinct improvement over hardwired analog integration systems. The digital approach allows the end of line programming of sensitivity parameters for different car models. Thus sensitivity parameters are programmed into the EEPROM of the microcontroller which allow the use of the same ROM-mask for different car types.

Electronic control modules. Today, state-of-the-art ECUs use a one-channel acceleration sensor and an arming sensor in series to the output stages. Such units are a mix of electronic and mechanical control. The first *all-electronic* single-point sensing ECU went into production in mid-1992. This module incorporated a dual-channel electronic accelerometer and no mechanical safing switch. Figure 23.5 shows the block diagram of a state-of-the-art electronic control module.

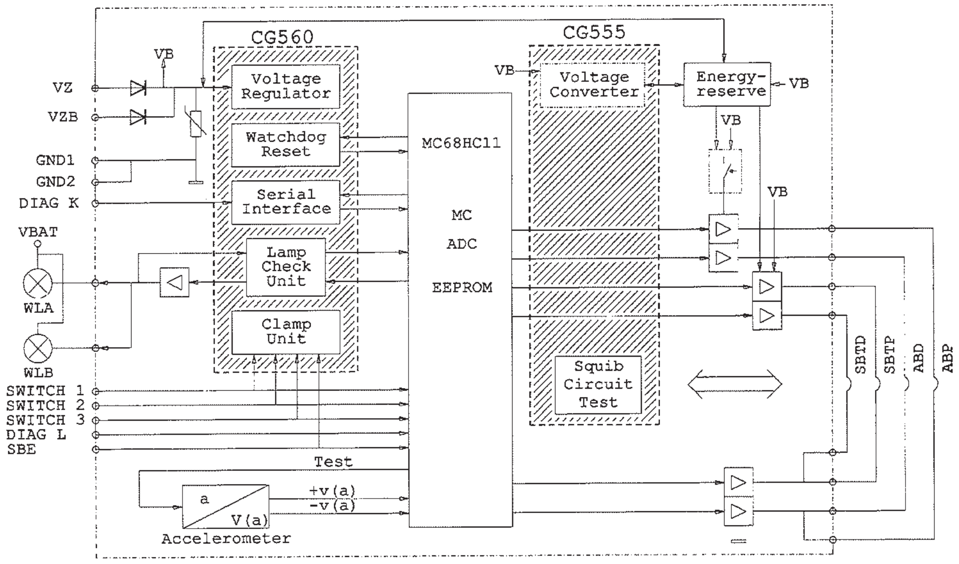


FIGURE 23.5 Block diagram AB6.3-ECU.

AC-firing. Normally squibs are fired with a dc pulse (dc-firing). The first unit with ac-firing went into production in mid-1993. With ac-firing, there is a capacitor inside the squib connector in series to the squib. This capacitor is periodically charged up and discharged so that a firing is only possible with alternating current (ac-firing).

AC-firing was introduced because of faults occurring in the vehicle periphery (outside the ECU). Typical faults consist of short circuits in the air bag module (= squib + inflator + bag + cover) by metal (splinters), defects in the contact unit for the driver air bag (= clock spring), and short circuits in the harness. Igniters are very sensitive to short current pulses (3 A for 60 μs can be sufficient to fire the air bag).

The main benefit of ac-firing is an immunity to all short circuits to dc-vehicle voltages with and without the ECU, as well as immunity from inadvertent deployments with all types of static ECU defects. The danger of false triggering is reduced because the push-pull output stages can only fire if activated with the correct asynchronous sequence of pulses by the microcontroller. A microcontroller disturbance inadvertently producing such a pulse sequence is highly unlikely.

A disadvantage of the current ac-firing system is the more complex ignition loop diagnostics required in the ECU. These diagnostics have to monitor the resistance and capacitance in the loop.

23.1.4 Rollover Sensors

Convertibles can overturn in almost any direction (although a backward turnover around the lateral axis is not very likely). Thus a rollover sensor unit has to sense in practically every direction in the horizontal plane. The kind of sensing and the signal evaluation times are the differences between rollover and air bag sensor units. A rollover accident has to be detected within approximately 200 ms; the activation time of the mechanical protection device takes approximately 300 ms.

The following rollover sensing concepts are used. The first concept (URS = *Überrollsensor*) went into production in 1989: omnidirectional tilt switch in conjunction with rear-axle switches, and longitudinal and lateral accelerometers. These are the same ones as used for air bag ECUs.

The tilt switch consists of a rocker cylinder with a permanent magnet mounted on top. A Hall-IC placed in the lid of the switch senses the rocker cylinders position. The discrimination threshold can be adjusted to the desired angle (e.g., 22°).

Rear-axle switches open if one or both half-axles of the vehicle are completely unloaded. This means that a potential vehicle rollover is indicated by an airborne wheel.

The system cannot release the rollover protection device (e.g., the rollover bar) if the respective rear wheel cannot move any more because of a fender deformed during impact. To avoid this, and for redundancy purposes, the sensor unit also senses the acceleration in every horizontal direction. It contains a longitudinal and a lateral piezoelectric accelerometer whose signals are squared and added. Exceeding the release threshold of approximately 5 g causes activation of the rollover protection element(s) located behind the rear seats. This sensor also unlocks the central locking system after a vehicle rollover.

Rear-axle switches are one way to provide a redundant signal for the rollover indication. Another system (ÜRSS = *Überroll-Schutz-System*) is based on a combination of three "spirit level" sensors, optically evaluated by LEDs and photo transistor-light barriers, and a vertical gravitation switch consisting of a magnet, a spring, and a reed contact.

Two of the level sensors are mounted ±52° to the horizontal plane and sense overturning around the vehicle's longitudinal axis. They also close if an acceleration greater than 1.3 g is exceeded for more than 80 ms. The third-level sensor is inclined by 80° to the horizontal plane and senses overturning around the lateral axis in the forward direction. This sensor closes for an acceleration exceeding 5.7 g for more than 80 ms.

The reed contact in the vertical acceleration switch is closed after the vehicle loses ground contact, since it is no longer possible to sense gravity in the vehicle if it is airborne. Thus, a spring with a restoring force of approximately 10 percent of the magnet's normal weight is enough to move the magnet over the reed contact and cause its closure. The rollover protection is then activated after a delay of approximately 300 ms. This activation can be initiated either by one of the level sensors or by the vertical acceleration switch.

A release of the rollover protection equipment in the case of a backward turnover around the vehicle's lateral axis would require an additional fourth-level sensor in the sensor unit.

23.1.5 Tire Pressure Control Systems

Tires must have the correct pressure in order to provide expected driving performance like the transfer of high breaking or lateral adhesion forces to the road surface. They must have good aquaplaning properties, which also depend on the tire pressure. In addition to safety considerations, improved economy can be expected since low tire pressure generally increases rubber abrasion and fuel consumption.

Electronic tire pressure control systems alert the driver in case of pressure loss. Low pressure indication is performed using a reference pressure switch mounted in the rim which compares the tire pressure with the reference pressure in a box hermetically sealed by a high-grade steel diaphragm. A contact opens if tire pressure falls below the reference pressure. Temperature dependancy of the pressure switch is negligible since the temperature of tire air is practically identical to temperature of reference gas.

A high-frequency sensor mounted on the axle detects if the pressure switch is closed. At each pass, the closed pressure switch contact activates a serial resonance circuit by energy absorption out of the axle-mounted high-frequency transmitter. So, rotation can only be sensed if the tire pressure is high enough.

An ECU for each wheel compares transmitter signal with wheel rotation signal of a rotation sensor. As there is a defined relation between these signals, any discrepancy within a specified driving distance can be detected, and the driver warned.

23.2 PASSENGER CONVENIENCE SYSTEMS

23.2.1 Electromechanical Window Drives

There are three different systems currently in use.

Lift Mechanism System. An electric gear motor drives a spur gear meshing in a gear segment. This segment acts on the window lift mechanism.

Flexible Cable System. An electric gear motor drives a system of flexible cables to lift the window.

Stiff Cable System. An electric gear motor is connected with the window via a tension and compression stiff cable. This cable is driven by a toothed wheel.

Window Lift Motors. The limited space inside the door requires window lift motors with a flat design. The lift mechanism consists of a self-locking worm gear system, which keeps its position if not intentionally activated and which must resist trials to open the window by force. An elastic claw clutch provides good damping behavior.

Window Lift Controls. Activation is performed using rocker switches.

Best driver comfort results from a combination of window drives and the central door-locking system. Here the windows are either automatically closed after leaving and locking the vehicle, or they are driven to a designated ventilating position. Such an automatic system requires the use of a driving force limiting or jamming protection. This protection must be effective during upward motion of the window within a range of 200 to 4 mm before the upper window stop.

Jamming force must not exceed 100 N, and the rise of force must remain less than 10 N/mm. Hall sensors monitor the rotation speed of the lift motors. If a speed decrease is detected, the rotation direction is immediately reversed. A complete window closure is enabled by turning off the jamming protection just before the window is driven into the door gasket. Thus, the motor can be driven just to blocking (if necessary). On this occasion, initialization of the window position is done.

An ECU for all window lifts can either be mounted centrally, or the electronic controls can be integrated into the window lift motors in order to save harness expense. Such drives would be preferable when used in conjunction with bus systems.

23.2.2 Electromechanical Sunroof Drives

State-of-the-art sunroof drives incorporate the functions of a lifting and sliding roof. Lifting or sliding motions are initiated by electronic or electromechanical controls. Electromechanical controls include mechanical interlocks of limit switches which allow the closed sunroof to be opened or lifted. Reversing the polarity of the drive motor lowers or closes an opened or lifted sunroof. In case the sunroof control is connected to a central locking system, then an electronic unit with jamming protection would be advantageous.

Additional functions such as preselectable position control, closing if demanded by a rain detector, and data bus connection can be realized with low extra expense.

The permanent magnet drive motor has a rated output of approximately 40 W and is coupled to a worm gear unit. Sunroof motors have a thermal overload protection implemented by thermal circuit breakers. The sunroof is moved either by flexible cable drives or by contraction/tension-stiff cables. It must be possible to close the sunroof using simple on-board tools in case there is a failure of the electrical system.

23.2.3 Electromechanical Seat and Steering Wheel Adjustment

Electrically adjustable seats are especially well suited for vehicles frequently driven by different persons of various size. After each driver change, the seat can be adjusted to best fit the new driver's requirements.

Up to five motors per front seat allow the following position changes: height adjustment of front edge of seat plane, height adjustment of rear edge of seat plane, longitudinal adjustment of seat, tilt adjustment of back rest, and height adjustment of headrest. Electric motors drive gear units via flexible shafts for longitudinal and vertical adjustments of the seat.

There are seat adjustment controls currently on the market with programmable nonvolatile memories, allowing a readjustment of the seat to positions stored in the EEPROM. Adjustments of the external mirror and the height of the seat belt guide can be programmed as well, according to individual demands. Furthermore, seating comfort in passenger cars is improved using electromechanically adjustable steering columns and tiltable steering wheels. The adjustment is performed via a self-locking gear unit with an electric motor and is integrated directly into the steering column. Adjustment can be achieved by manual activation of a position switch, or by connection to the programmable seat memory unit.

23.2.4 Central Locking Systems

Central locking of the doors, trunk lid, and tank cap is done either by pneumatic or electromotor driven actuators.

With the pneumatic system a bipressure pump provides the over- and underpressure to the system. To achieve this, the pump is driven by an electromotor in both rotation directions. The system can be operated from different locations: central position switch, driver door lock, passenger door lock, and trunk lid lock.

All-electric motor-driven central locking systems are based on the same concept. A small electric motor with reduction gear drives to position a bar for closing or opening the lock.

In case of power loss, it must be possible to open the doors with the mechanical key or with the internal door handle. On central locking systems with integrated theft alarm, opening is only possible with the mechanical vehicle key.

There are central locking systems which can be operated by ultrasonic remote control. These systems offer keyless closure or entry and thus high operational comfort.

23.2.5 Reverse Parking Warner

The system consists of an ECU, LCD, and sensors. It is equipped with ultrasonic transmitter/receiver sensors using the echo depth sounder method. Bursts of ultrasonic waves at approximately 30 kHz are transmitted for 30 ms. The travel time of the first echo signal gives information about the distance to the obstacle.

Upon shifting into the rear gear, the system is activated and completes a self-check by turning on all display characters for a specified short time period and checking the ultrasonic sensors. A space up to 0 to 160 cm behind the vehicle is observed and divided into different warning sectors. The most critical distance is an obstacle within 30 cm. In this case, the optical display flashes and the acoustical warner gives a permanent signal.

23.2.6 Emergency Light Flasher

An ECU turns on the blinking lights. Synchronous flashing of all blinkers of a vehicle must be possible with the ignition turned on or off. A specific control lamp to indicate the status of the emergency light flashers must be included inside the vehicle. Usually this lamp is integrated into the on/off switch.

GLOSSARY

Accelerometer Inertial deflection of this sensing element results in an electric signal proportional to acceleration or deceleration of the vehicle.

Airbag Gas-filled cushion for impact energy absorption of vehicle occupants during a crash.

Allowable forward displacement Forward displacement of the occupant which is allowed before the restraining device has reached its full protection capability.

Arming sensor An electromechanical acceleration switch integrated in the electronic control unit located inside the passenger compartment. Activation of the restraining device is only possible if the arming sensor completes the triggering circuit.

EEPROM Electrically erasable and programmable read only memory.

Energy reserve Backup power supply for the restraining system in case of battery loss.

Forward displacement Forward travel of the occupant during a crash which occurs until there is contact with the restraining device or vehicle interior.

Hall IC Integrated circuit containing a Hall sensor plus appropriate evaluation circuitry. The Hall sensor delivers an electric voltage proportional to the magnetic field strength by which it is affected.

Igniter Electrically activated initiation element for firing of an inflator.

Inflator Gas-producing unit for air bag inflation or activation of seat belt tensioner.

Piezoelectric Materials, e.g., special ceramics, which have the capability to convert mechanical deformation into electric signals (voltage and/or charge) and vice versa, are known as piezoelectric materials.

RAM Random access memory.

Reed contact Mechanical, magnetic-field-sensitive switch, hermetically sealed inside a glass tube.

Restraining device Air bag, seat belt, and seat belt tensioner.

ROM Read only memory.

Safing sensor Another expression for arming sensor.

Seat belt tensioner Pyrotechnically or spring-activated device to tighten the seat belt during a crash.

Squib Electrically activated initiation element for firing an inflator (another expression for igniter).

Voltage converter Electronic device to keep the energy reserve charged up to a voltage normally higher than battery voltage.

Watchdog Electronic circuit for checking the correct flow of the microcontroller program.

BIBLIOGRAPHY

Analog Devices Specifications for "Monolithic accelerometer with signal conditioning ADXL50,*" edition C1808-20-6/93. (*Patent pending.)

Bosch Automotive Handbook, 2d ed., 1986.

"Bosch Kraftfahrtechnisches Taschenbuch," 21. Auflage, July 31, 1991. Bosch internal design documentation of electronic control units for restraint systems.

Handout of lecture "Der BMW-Überrollsensor (ÜRSS)," given by M. Jost on 13. Conference "Elektronik im Kraftfahrzeug," Haus der Technik, Essen, Germany, June 15–16, 1993.

ABOUT THE AUTHOR

Bernhard Mattes has been employed with Robert Bosch GmbH since he was graduated from the Technical University of Stuttgart, Germany, as a Diplom-Ingenieur in 1968. He began work at Bosch as an electrical engineer in an automotive electronics design department and has been a section manager since 1984. His working fields at Bosch have included electronic gear box controls, cruise controls, crash-sensing systems for front and side impacts, sensing systems for rollovers, accelerometer and safing sensor development, and tilt and rotation sensor development.

CHAPTER 24
ANTITHEFT SYSTEMS

Shinichi Kato
Manager, Electronics Design Section No. 2
Nissan Motor Co., Ltd.

24.1 VEHICLE THEFT CIRCUMSTANCES

24.1.1 Vehicle Theft in the United States

In the United States, the number of thefts from vehicles exceeds the number of actual vehicles stolen. A much larger profile is given to stolen cars, however, because of the economic losses suffered by car owners and the tie-in with stolen cars and drug abuse and other crimes.

The number of vehicle thefts in the United States has increased continuously to a 1991 level of approximately 1.7 million. The ease with which thefts can be carried out, the light punishments given to car thieves, and the ready availability of car theft tools and information on how to use them, has led to a systematic network of organized car crimes and the establishment of black markets for both stolen cars and parts.

24.1.2 Vehicle Theft in Europe

By 1992, the number of vehicle thefts in France, Germany, and the United Kingdom totaled 1.08 million, a radical 64 percent increase since 1988. The breakdown of communism in eastern Europe, adding to the existing black market for vehicles and spare parts, the trend toward joyriding in stolen vehicles, and other social problems similar to those in the United States, have all been contributing factors in the increase.

24.1.3 Vehicle Theft Worldwide

The United Kingdom has the largest incidence of vehicle theft followed in order by Australia, Sweden, the United States, and France.

24.1.4 Reasons for Vehicle Theft

The reasons for vehicle theft vary from country to country and between different areas of the same country. In the United Kingdom and Australia, for example, joyriding is the cause of most auto thefts. In the United States, on the other hand, the theft of vehicle contents ranks highest.

24.1.5 Methods of Entry

In the United States, "slim-jims" or similar devices to unlock car doors and gain entry are the most popular among car thieves. Vehicle contents can then be removed or the steering lock broken and the vehicle hot-wired and driven away. The use of roughly ground ceramic has recently gained popularity over traditional methods of breaking windows such as hammers or spring-punches because of ceramic's ease of use and lack of noise when used.

With the advent of keyless entry systems, devices to "steal" the code while the driver operates the keyfob and transmit it back when the thief chooses, have been developed and readily available to the aspiring thief.

24.2 OVERVIEW OF ANTITHEFT REGULATIONS

The total cost to the public of vehicle thefts in the United States alone is estimated at $8.3 billion per year, taking into account the vehicle and insurance losses and the expense of police time. Due to these economic pressures and spiraling vehicle-theft-related crime, regulations are now being introduced to help protect the vehicle-buying public and insurance companies.

24.2.1 Vehicle Security Regulations in the United States

Three types of vehicle security regulations exist in the United States:

- The vehicle must be fitted with an audible warning that operates when the ignition key remains in the ignition and the driver's door is opened.
- As vehicle theft is not consistent among vehicle types and different areas of the country, 11 states (Florida, Illinois, Kentucky, Massachusetts, Michigan, New York, New Jersey, Pennsylvania, Rhode Island, Texas, and Washington) have introduced regulations to deduct 5 to 15 percent from insurance premiums if vehicles are equipped with certain antitheft systems.
- The Vehicle Theft Act (1986) makes it mandatory for vehicles with a particular theft rating or above to have specified parts of the vehicle clearly marked.

Although this places a cost burden on car makers, they do not need to mark parts on vehicles they equip with antitheft systems approved by the National Highway Traffic and Safety Administration.

24.2.2 Vehicle Safety Regulations in Europe

The United Kingdom. In an attempt to improve vehicle security, insurance companies in Europe define the effectiveness of individual antitheft systems. For example, in the United Kingdom, Thatcham (a survey division of the Association of British Insurers) classifies all vehicles with regard to their antitheft worthiness. This is done on a points basis, with points awarded as shown in Fig. 24.1.

Effective in 1994, Thatcham criteria indicate that the two most important attributes of a vehicle security system are a two-way mobilizer to prevent theft of the vehicle and interior sensors to deter unauthorized entry and theft from the vehicle. Any vehicle judged to have adequate antitheft devices gets significant savings on insurance premiums, thereby offering a powerful incentive to car buyers.

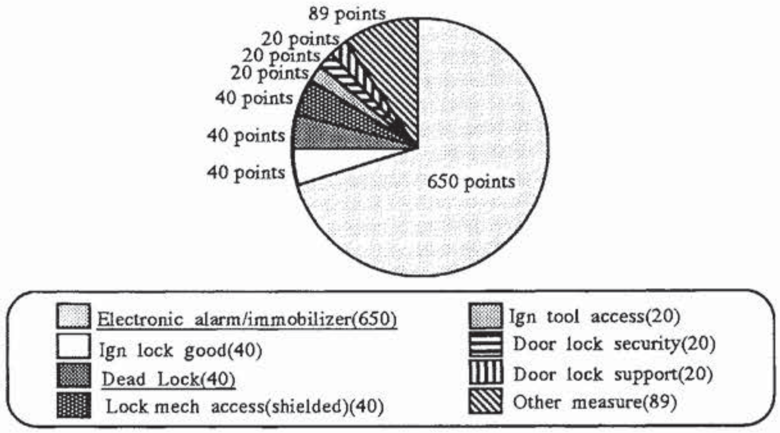


FIGURE 24.1 United Kingdom's Thatcham point system, which assigns points to various antitheft devices.

Germany. In Germany, the country's largest vehicle insurer is Allianz. It announced the policy detailed in Table 24.1 whereby heavy financial penalties are introduced to encourage car owners to ensure that their vehicles meet security standards. These penalties exist in the form of 10 percent premium increases and only 90 percent repayments from insurers in the case of a stolen vehicle that has no antitheft systems. As in the United Kingdom, the application of antitheft systems is likely to spread as more customers demand more secure vehicles from car makers.

TABLE 24.1 Allianz Insurance Company Antitheft Policy

Timing	Scope	Antitheft system	Impact noncompliance
07/01/93	Vehicle to be fitted with aftermarket system	Immobilizer step 1: interrupt at least 3 systems • Starter circuit • Fuel supply • Ignition system • Other system	In case of theft claim, customer only gets back 90 percent of car value
01/01/95	Standard fitment for all new registrations	Immobilizer step 2: engine management operation (electronic code)	↑

24.3 A BASIC ANTITHEFT SYSTEM

A basic antitheft system, Fig. 24.2, must do three things: sense unauthorized entry into the vehicle or unauthorized movement of the vehicle, detect unauthorized attempts to start the vehicle or to disarm the alarm system, and upon detecting any unauthorized act actuate an

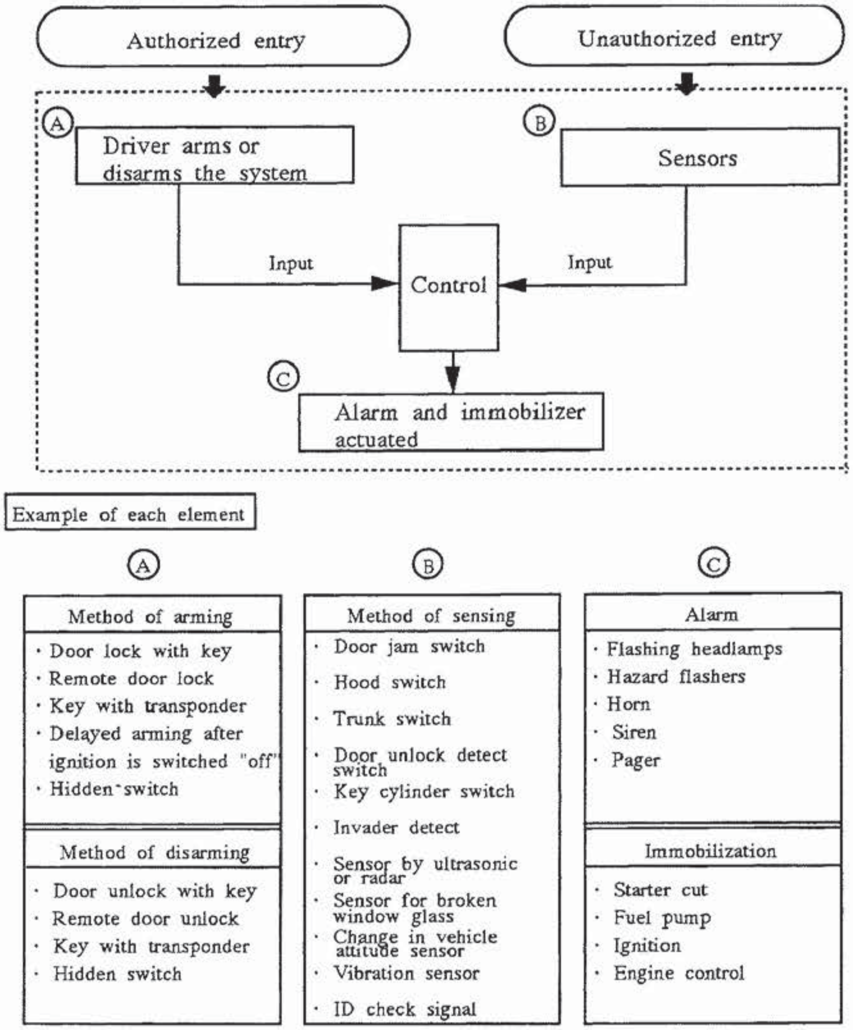


FIGURE 24.2 The basic structure of an antitheft system.

alarm and inhibit the vehicle from starting. Many vehicles now include visual indication in the form of stickers or flashing lights to indicate that they are equipped with an antitheft system.

Antitheft systems use active and passive arming systems. In an active system, the driver can arm the system with an ultrasonic or infrared keyfob or by throwing a hidden switch. In a passive arming system, operating the door key switch automatically arms the system when the driver leaves the vehicle. Insurers in the United States feel that the passive system is more effective and, consequently, give a larger insurance premium deduction for vehicles equipped with such systems. Two thirds of all vehicle security systems are passive, while almost all factory fit systems fall into this category.

24.3.1 Unauthorized Entry Detection

A tamper switch detects any abnormal forces applied to the door and trunk lock key cylinders, while make-break switches sense opening of all apertures. Ultrasonic beams inside the vehicle can detect the vibration from broken glass or movement within the car. The trigger level of these units is particularly sensitive, because too low a level can lead to false alarms.

24.3.2 Alarm and Immobilization

Alarm activation is indicated by the sounding of the car horn or siren combined with the flashing of the vehicle's headlights and taillights. A recent innovation has been the application of radio paging systems to indicate stolen vehicles. The vehicle is also immobilized by causing an interrupt in circuits vital to vehicle operation such as the starter motor, fuel pump, and ignition systems.

24.3.3 A Typical System Example

The following explanation relates to one of the more popular vehicle alarm configurations shown in Figs. 24.3 and Fig. 24.4. Abnormal conditions or unauthorized entries are detected as inputs from door, hood, and trunk switches, ignition switch, and ultrasonic sensors. When any of these conditions occurs, the horn sounds and the vehicle's lights flash for a predetermined period (usually one to three minutes).

A circuit example using a custom integrated circuit is shown in Fig. 24.4. A flow chart of the system operation is shown in Fig. 24.5.

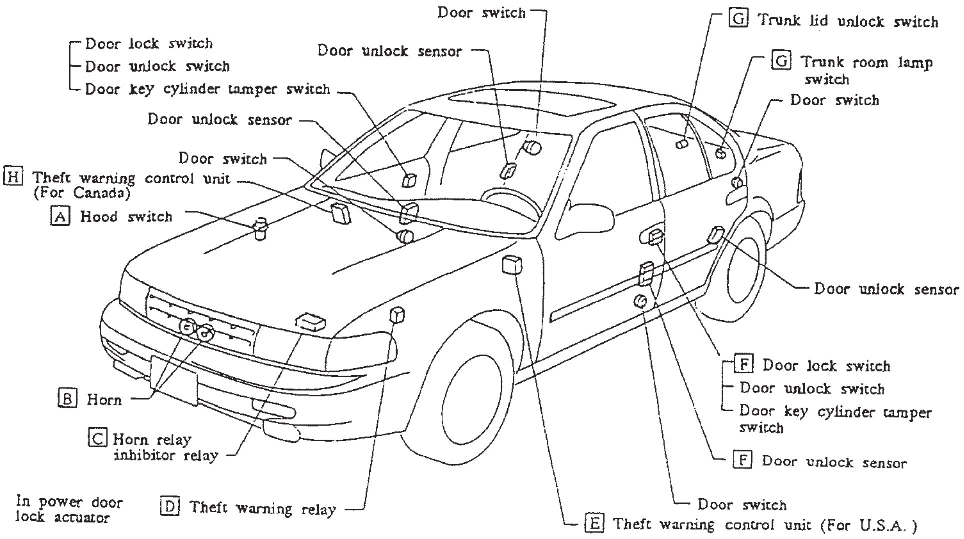


FIGURE 24.3 Component parts and harness connector location for an antitheft warning system.

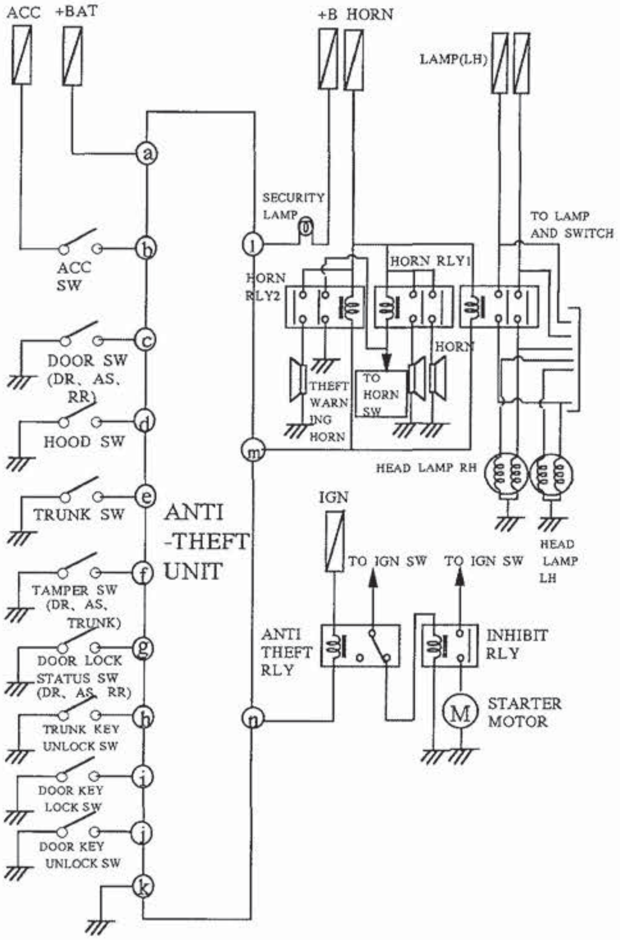


FIGURE 24.4 Circuit diagram of an antitheft system using a custom IC.

24.3.4 Immobilizer

Because of the frequency of false alarms activated by oversensitive invasion sensors, vehicles with alarms sounding are increasingly being ignored by people in the vicinity. Therefore, as a deterrent to vehicle theft, unauthorized entry detection systems alone are no longer considered effective by insurance companies. They prefer—and reduce premium reductions accordingly—to have, in addition, an immobilizer to electrically or mechanically disable the vehicle unless it is correctly disarmed.

Now that most vehicles are equipped with electronic engine control units, car makers feel that they provide the best means for immobilization. This is accomplished with three basic driver interfaces for arming and disarming (Fig. 24.6):

- An ID code input to the ECU by means of a keyboard, which may be separate or combined with another electronic dashboard device such as the radio

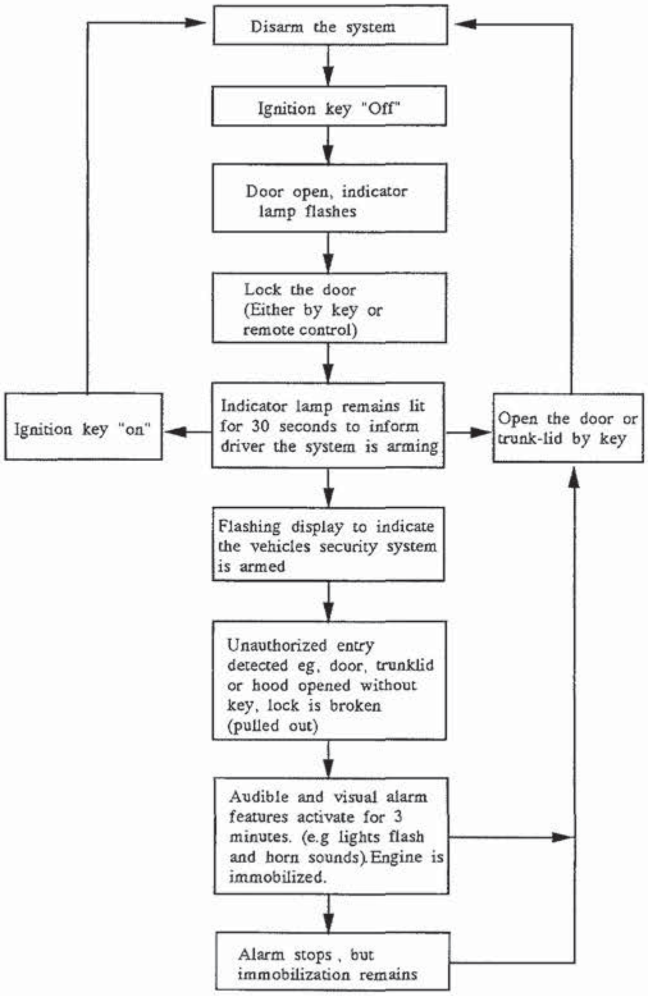


FIGURE 24.5 A flowchart for the system shown in Fig. 24.4.

- An ID signal transmitted by a keyfob to a receiver and decoding unit in the vehicle
- An ID signal received from a transponder built into the ignition key

The transponder built into the ignition key is the most recent innovation (Fig. 24.7). The ignition key contains a transponder device. A reader unit is installed in the vehicle close to the ignition key cylinder. A "power pulse" is transmitted by the reader's antenna shortly after the ignition key is inserted into the lock. This power pulse is received as energy by the transponder and stored in a capacitor. The power is then used to transmit an ID code signal back to the reader. The reader identifies the received signal as authorized or not and, if authorized, enables the ECU to start the vehicle.

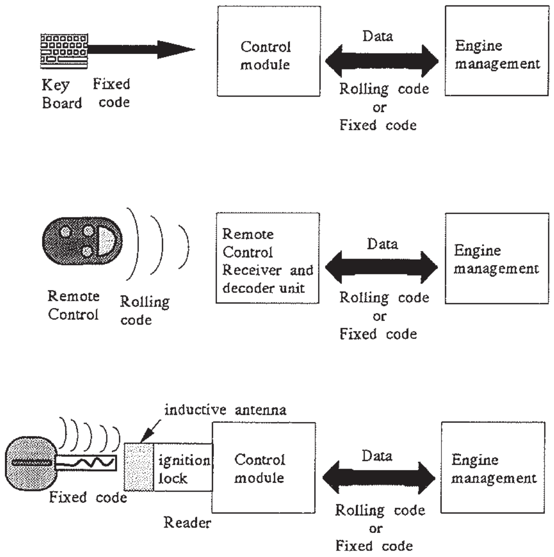


FIGURE 24.6 Three methods for arming and disarming an antitheft system.

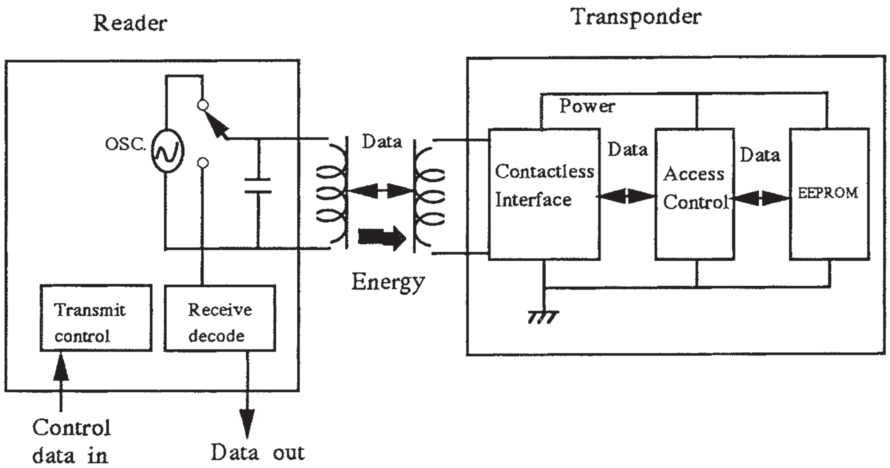


FIGURE 24.7 A block diagram of a typical antitheft system incorporating a transponder built into the ignition key and a reader unit installed near the ignition key cylinder.

ABOUT THE AUTHOR

Shinichi Kato was graduated from Tokyo Institute of Technology in 1973 and joined Nissan Motors Co. Ltd. that same year in charge of electric parts design. He was sent to Nissan Research and Development in Michigan in 1988, where he was in charge of the Ford and Nissan joint project as an electrical design engineer. He returned to Japan in 1992 as the manager of Nissan's electrical design department.

CHAPTER 25
ENTERTAINMENT PRODUCTS

Tom Chrapkiewicz
Philips Semiconductors

25.1 FUNDAMENTALS OF AUDIO SYSTEMS

The purpose of an automotive audio system is to present a realistic illusion of a musical event. While this chapter is concerned primarily with automotive audio systems, it is appropriate to cover some of the basics of audio systems and the human auditory system.

25.1.1 Characteristics of Sound

Sound has a dual nature: it may be considered a physical disturbance in a medium such as air, or it may be considered a psychophysical perception resulting from nerve impulses stimulating the acoustic cortex of the brain. In audio, we are vitally concerned with both. The ear itself determines the quality of sound signals, but the sound is carried to the ear through physical stimuli. Acoustics is the branch of physics that deals with this transmission of sound in a medium to the ear. The complex relationship process that relates stimulus and sensation is the field of psychoacoustics.

***Basic Acoustics*[1].** For sound to be transmitted from one place to another, a medium is required that has elasticity and inertia. Air has these characteristics, as do other materials such as water, steel, wood, etc. When an air particle is moved by something vibrating, it moves, passing on momentum to adjoining particles as it strikes them. The original air particle is then pulled back toward its equilibrium position by elastic forces residing in the air. Any particular air particle vibrates about its equilibrium position, receives momentum from collisions, and passes momentum on to other particles, which pass it on to others, and so on. Consider a wave traveling through a pond of water. The wave progresses through the water, but each individual water molecule remains at (relatively) the same location. The actual transmission of sound through a medium is a type of longitudinal wave propagation.

The velocity at which an acoustic wave moves through a medium is dependent on that medium. The velocity of sound in air at standard temperature and pressure is approximately 344 m/s. This is considerably slower than the speed of light (3×10^8 m/s). Table 25.1 gives the velocity of sound in different media.

Any audible sound is of an alternating character with a characteristic frequency and amplitude. As this sound moves through the air, it has a particular wavelength, determined by its frequency and propagation velocity. The wavelength is calculated as:

TABLE 25.1 The Speed of Sound in Various Media

Media	Speed of sound, m/s	Speed of sound, ft/s
Air, 21 °C	344	1,129
Fresh water	1,480	4,856
Salt water, 21 °C, 3.5% salinity	1,520	4,987
Plexiglass	1,800	5,906
Wood, soft	3,350	10,991
Fir timber	3,800	12,467
Concrete	3,400	11,155
Mild steel	5,050	16,568
Aluminum	5,150	16,896
Glass	5,200	17,060
Gypsum board	6,800	22,310

$$\lambda = \frac{c}{f}$$

Thus, a sound with a frequency of 1000 Hz has a wavelength of about 0.344 m.

25.1.2 Characteristics of Audible Sound

Amplitude Range of Sound. Obviously, sounds can be very loud or very quiet. Since the range of levels of sound is so large, a logarithmic scale is used. The unit of this measurement level is the Bel (named after Alexander Graham Bell). The threshold of hearing is about 20×10^{-6} Pa (1 Pascal = 1 Newton per square meter). The threshold of pain is over 100 Pa. This represents a ratio of over 5 million. In logarithmic terms, this ratio of 5 million is log (p1/p2) or about 6.7. Since one Bel represents a rather large difference, the decibel (or one-tenth of a Bel) is usually used. So, a ratio in decibels (or dB) is $10 \times \log (p1/p2)$. Also, since the square of sound pressure is proportional to acoustic power, we use $20 \times \log (p1/p2)$. Thus, the *dynamic range,* as it's called, of the human auditory system is: $20 \times \log (100/20 \times 10^{-6})$ or about 134 dB.

Frequency Range of Sound. As mentioned, any sound has a characteristic frequency. The average range of the human auditory system is from 20 Hz to 20 kHz. This represents a wavelength range of 1.72 cm at 20 kHz to 17.2 m at 20 Hz. A particular sound does not consist of all frequencies within this range, but only one or (usually) several frequencies (recall Fourier). Also, the human auditory system does not respond equally to all frequencies. The human auditory system is most sensitive at about 3 kHz. Some well-known frequency ranges are detailed in Fig. 25.1.

25.1.3 Basic Psychoacoustics[2]

The Human Hearing System. The human ear is commonly considered to have three parts: the outer, middle, and inner ear.[2,3] The sound is gathered and modified by the pinna, directed down the auditory canal, and terminates in the tympanic membrane (eardrum). The ear canal is about 0.7 cm in diameter and about 3 cm long. Notice that a quarter-wavelength of a wave 3 cm long is 2870 Hz, which is near the frequency at which the human auditory system is most sensitive. The other side of the tympanic membrane faces the middle ear. The middle ear is filled with air, and air pressure equalization takes place through the Eustachian tube opening into the pharynx (throat) so that static pressure is the same on both sides of the eardrum.

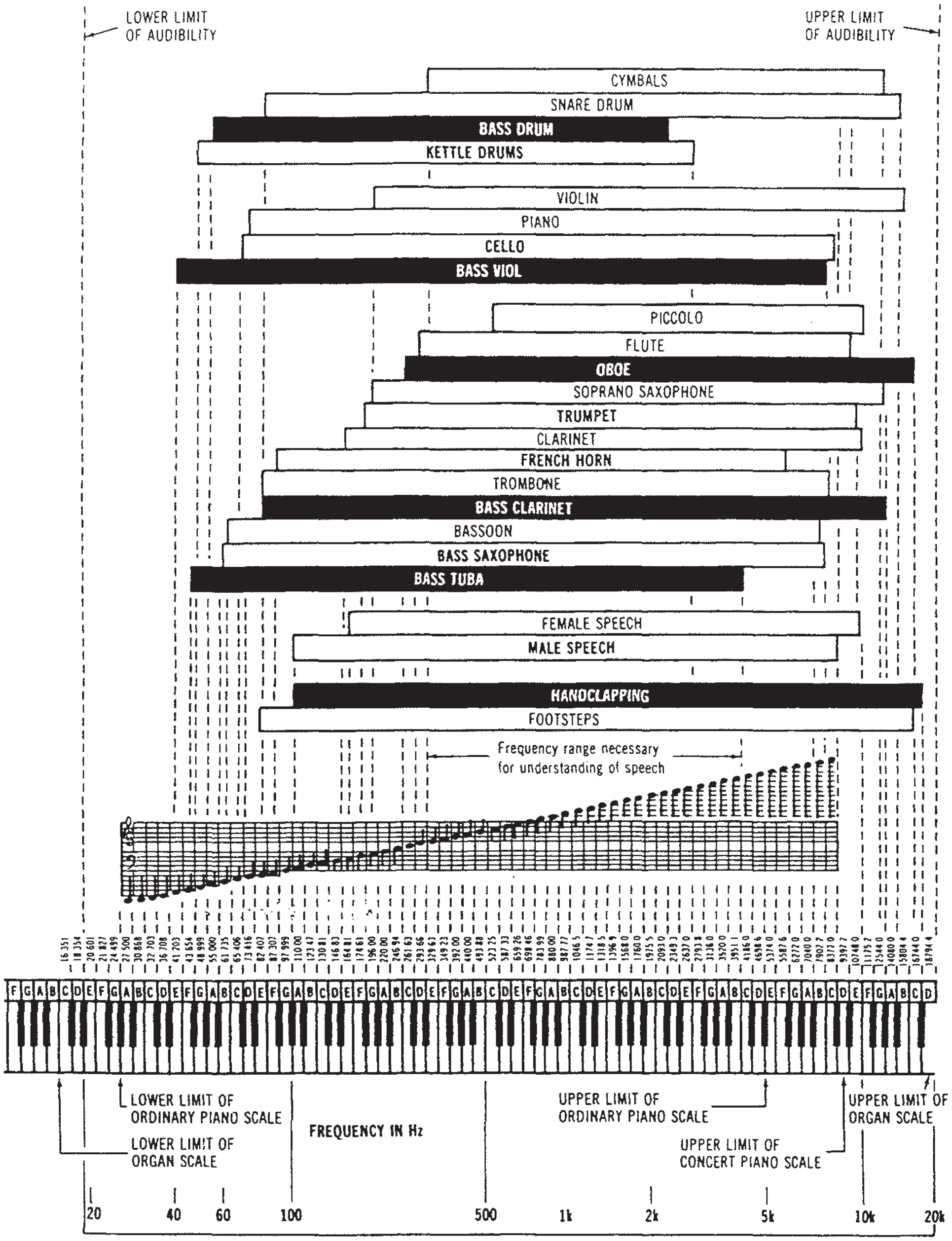


FIGURE 25.1 Audible frequency detail.

Fastened to the tympanic membrane is one of the ossicles. The three bones of the ossicles form a lever, the end of which excites the oval window of the cochlea. The cochlea is filled with an incompressible liquid which carries the sound waves. The cochlea is lined with hair cells (about 25,000) that send nerve impulses to the brain. The cochlea is basically a mechanical-to-electric transducer.

25.1.4 Psychoacoustic Phenomena

Localization. In localizing a sound, our ear and brain mechanism utilizes several types of angle-dependent data. Due to air absorption, there are the amplitude differences between the sounds arriving at the two ears. Since our ears are 8 or 9 inches apart, there is a time delay between sounds arriving at the two ears.

There are also differences in the spectrum of the sound entering the two ears. These differences are caused by reflections on the pinna and head diffraction. It has long been believed that both ears are necessary for localization, but more recent studies have shown that listeners can be trained to localize with only one ear.[4] Ear canal measurements have shown frequency nulls around 8.2 kHz. This implies a reflected wave with a delay of about 61 μs. In an audio system with multiple loudspeakers, the ear senses multiple sound sources. The auditory system uses the timing and spectral clues to localize the apparent sound sources. In addition to the real sources (the loudspeakers), the auditory system creates virtual sources between the real sources. This effect is known as imaging. The image (sometimes also referred to as the sound stage) is used by media producers to enhance the aural experience of the listener.

Echo Perception. As noted previously, time delays of less than hundreds of microseconds are utilized as localization clues in the hearing system. For delays much greater than this, but less than about 50 ms, the auditory system integrates the energy, and perceives a fuller sound than without the delays. The precedence effect refers to the fact that if a delayed signal reaches the listener from a different direction within 50 ms of the first, the apparent sound source is still in the direction of the original signal. For delays greater than 50 ms, the listener perceives a discrete echo. Additions of echoes are used in room or hall simulation (see Sec. 25.4.2).

25.2 A BRIEF HISTORY OF AUTOMOTIVE ENTERTAINMENT

It is most curious that the gestation of both radio broadcasting and the automobile occurred almost simultaneously. In 1885, Karl Benz produced the first automobile at Mannheim, Germany. It was powered by a spark-ignited gasoline engine. In 1886, Professor Heinrich Hertz of Karlsrule Polytechnic in Germany demonstrated the transmission of electromagnetic energy from a spark.

The first automotive entertainment system was installed in 1926. William M. Heina filed a patent Sept. 16, 1926, for a "Portable Radio Apparatus." It was issued as patent #1,626,464 on April 26, 1947. The patent made no revolutionary claims, but it served as a model for many future installations. Unsubstantiated reports indicate that Radio Auto Distributors and All American Mohawk had offered automotive radios in 1926.[5] Many milestones in radio broadcasting had already occurred. In 1910, spark-gap transmission tests of the Chalmers-Detroit Motor Car Company and the 1915 experiments of DeForest and Sarnoff led to the first broadcasts of the Detroit Station 8MK (later to become WWJ) on Aug. 20, 1920, and the broadcasts of KDKA on Nov. 2, 1920 (earlier known as 8XK).

The radio circuits of the day were based on vacuum-tube designs and consequently required high voltages and power consumption. By 1931, the dynamic loudspeaker (which was pioneered by Magnavox in 1928) was installed in automobiles. (This is the fundamentally the same type of loudspeaker used today) This offered quite an advantage, since earlier electrodynamic loudspeakers had quiescent current consumption of about 1 A. Tremendous growth had occurred leading to 1930. By that time many manufacturers of automotive radios were in existence: Automatic Radio Manufacturing Company, United American Bosch Corporation, Carteret Radio Laboratories, Inc., Crosley Radio Corporation, Charles Hoodwin Co., Galvin Mfg. Corporation (later to become Motorola), National Company, Philco-Transitone, Pilot Radio & Tube Corporation, Sparks-Withington Company, and The United States Radio & Television Corporation.

The predominant medium of transmission to the automobile was via the AM band. While FM had been in existence since 1940, it was not until 1957 that the first FM radios were installed in automobiles. The year 1966 gave us the first OEM installation of 8-track tape players after a joint effort among Ford Motor, Lear Manufacturing, Motorola, and RCA. This was the medium of choice through the early 1970s, only to be replaced by the more rugged compact cassette (see Sec. 25.3.1 on Media). There were other short-lived intermediate media such as vinyl records and other obscure tape cartridges. Their life and breadth of application were very limited, although they paved the way for the media of today. Modern media such as the compact disk have limited installation rates. The medium provides superior performance, yet is not nearly as popular as cassette installations. The near-future of media may rest with the digital compact cassette (DCC) or the MiniDisc (MD).

25.3 CONTEMPORARY AUDIO SYSTEMS

An audio system could more correctly be called a high-fidelity sound reproduction system. The goal of any audio system, be it in an automobile, home, or movie theater, is to reproduce an actual or intended acoustic event. Most commercially available recordings (cassette, compact disc, or LP record) are recordings of actual acoustic events—that is, all performers performing together interacting with each other. Modern multitrack recording technology, however, has removed this restriction from the artists. The intended or virtual event that is captured may have never really occurred; it may be an imaginary event in the mind of the producer. A duet may be performed with the singers separated by many miles and even years.[6] Modern digital signal-processing (DSP) technology has allowed the artist to create not only virtual events such as this, but also virtual performers that have not ever performed a piece.[7]

The first audio system was invented by Thomas Edison between August and December 1877.[8, 9] This hand-cranked system consisted of a grooved metal cylinder, a mouthpiece, and tin foil as the recording medium. The indentations in the foil cylinder were an analog reproduction of the sound waves impressed on the foil. These analog indentations were of a vertical or "hill-and-dale" nature. By 1896, Emile Berliner developed the disc recording system, in which the sound was stored on a flat disc as lateral grooves in the surface of the disc. This basic method has remained in use to this day (1994) on records otherwise known as LPs (although LPs are increasingly difficult to find). By 1926, the recording and playback of these discs had been transformed from a strictly acoustical system to an electrical system. In 1931, Alan Blumlein of Columbia Laboratories conducted experiments in stereo reproduction of sound (from the Greek *stereos,* meaning solid). These systems required two information channels and early attempts to do this on a disc failed. By 1957, commercially available stereo records became available. These had the two channels of information stored in the two walls of a single groove. It was not until the 1970s that stereo records became universal and the older, monaural records were no longer produced. During this same time period, the system of magnetic tape for audio storage was being developed, first in Germany in the 1940s and imported to the United States by the Ampex Corporation. This system stores one or more channels of information in analog form on an oxide- or metal-coated plastic tape. This system is still in use today in the form of the analog cassette (originally called the compact cassette). The compact cassette was developed in the late 1960s by Philips Corporation for use in dictation equipment and was shortly thereafter adapted to audio system usage. In the late 1970s, Philips and Sony Corporation began development of the compact disc system. This was the first consumer system to use a digital storage medium. The information is stored in digital form using an optical medium. By the late 1980s, the compact disc was the dominant format in dollar sales. In the early 1990s, there were several competing digital audio recording formats [digital audio tape (DAT), digital compact cassette (DCC), and recordable minicompact disc) for the home, although it is not yet clear which one will dominate the market.

A typical audio system controls a stereo signal which is obtained from any medium such as a radio broadcast tuner (AM, FM, etc.), turntable (record player), cassette tape, or compact disc. These media contain two discrete channels (left and right) of sound. Ideally, there is an infinite amount of separation between these two channels, but the human auditory system cannot distinguish separation beyond about 25 dB. In actual recordings, the information contained in these two channels is usually very similar, but different depending on the tastes of the musician and producer. The differences in the two channels is what gives recordings vastly different spatial characteristics.

Typically, these two channels are first processed by the volume, tone, and balance controls and are ultimately transformed to an acoustic signal at the loudspeakers. The analog signal levels in a modern audio system are about –10 dBu (where 0 dBu is 0.775 V), or about 0.3 V RMS with very little current capability (on the order of milliamperes). This is what is commonly referred to as a *line-level* signal. After the power amplifier stage, the signal is at a level of about 1 to 50 V with a current capability of up to 10 A. Recall that these analog signals have a 20-Hz to 20-kHz bandwidth and a 90-dB dynamic range. Digital signals in a consumer audio system (while currently rare but gaining popularity), are about 0.5 V peak-to-peak, requiring a bandwidth of about 4 megabits per second.

25.3.1 Media

AM and AM Stereo. Amplitude modulation is defined as the process of changing the amplitude of a high-frequency carrier in accordance with the amplitude of the signal to be transmitted. Commercial AM broadcasting in the United States is within the carrier band of 530 to 1600 kHz. When there is no signal to be transmitted, the carrier is of a constant frequency and amplitude. When an audio signal is to be transmitted, the amplitude of the carrier is modulated in accordance with the amplitude of the audio signal. Within the 530- to 1600-kHz band, each AM station is spaced 10 kHz apart. This allows for a maximum audio frequency of about 10 kHz to be transmitted. Thus, the AM media is limited to a maximum transmitted signal frequency of 10 kHz and a dynamic range of about 40 dB. Primary markets for AM are news, talk radio, sports, and limited music broadcasts. The nature of the AM system allows for easy corruption of the amplitude envelope from atmospheric noise, electromagnetic interference, and lightning.[10]

Experimentation as early as the 1950s and 1960s by Philco and CBS proposed AM stereo broadcasting.[11] These systems were not compatible with existing monaural systems and were consequently rejected by the Federal Communications Commission (FCC). By the late 1970s, several competing and compatible AM stereo systems were proposed. The most notable competitors were Harris, Motorola, and Kahn. The system that is in widespread use today is the Motorola C-Quam® System.[11]

FM and FM Multiplex Stereo. Frequency modulation is defined as the process of changing the frequency of a high-frequency carrier in accordance with the amplitude of the signal to be transmitted. Commercial FM broadcasting in the United States is within the carrier band of 88 to 108 MHz. When there is no signal to be transmitted, the carrier is of a constant frequency and amplitude. When an audio signal is transmitted, the frequency of the carrier is modulated in accordance with the amplitude of the audio signal. Notice that the amplitude of the transmitted modulated carrier is of a constant amplitude at all times. See Fig. 25.2 for an example of a typical signal and the resultant carrier in an FM system. Within the 88- to 108-MHz band, each FM station is spaced 200 kHz apart. This allows for a maximum audio frequency of about 15 kHz to be transmitted. Thus, the FM medium is limited to a maximum transmitted signal frequency of 15 kHz. Primary markets for FM have been music. The nature of the FM system provides for a very robust noise-free transmission, since any amplitude irregularities of the received signal can be ignored. Any interference which may cause amplitude irregularities can therefore be eliminated.

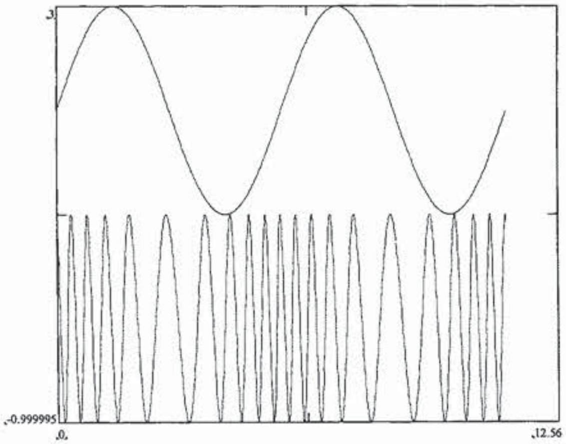


FIGURE 25.2 Signal and FM modulated carrier.

The Frequency Modulation Multiplex (FM MPX) system was developed in the early 1960s as an extension of the existing FM system. Channel-spacing is maintained at 200-kHz increments. A sum signal, the sum of the left and right channels (L + R), is frequency modulated on the carrier frequency exactly as in the monaural system. A 19-kHz pilot at 10 percent modulation level indicates the presence of a stereo signal. A difference signal (L-R) is modulated about a suppressed subcarrier of 38 kHz. The FM stereo tuner circuit demodulates and decodes these signals and outputs the left and right audio signals for further audio processing. The audio signals are limited to a 15-kHz bandwidth and a dynamic range of greater than 60 dB.[12]

FM Reception Difficulties. Due to the relatively high carrier frequency, FM broadcast reception is less susceptible to the interference and poor reception of AM systems. However, FM reception can be limited by signal cancellation which is commonly called multipath distortion.[13–15] Multipath distortion is the loss of FM signal encountered when a direct and reflected signal simultaneously arrive at the receiver out of phase and cancel, resulting in a loss of signal. Since in an automotive receiver, the receiver is constantly moving through these null zones, the familiar spitting sound results. A method for compensating for these signal losses is what is called a diversity antenna system. This is a system of two or more antennas in extreme locations on the vehicle. Since the antennas will be more than one half-wavelength apart, at least one of them will not be in a null zone. The tuner then monitors the signal content at all antennas and switches to the antenna with the best signal. Antenna switching can be accomplished in less than 30 μs of detecting multipath interference. To eliminate antenna switching noise, the switching can be synchronized with the zero crossings of the audio signal.[14–18]

The second most common FM reception problem is ignition noise.[13,14] This problem continues to increase as automotive ignition systems become increasingly complex. This problem is exacerbated by the fact that conventional methods of eliminating high-frequency radiation, such as resistance wires and grounding straps, do not eliminate the problems in the FM frequency band.

The circuit that detects and eliminates ignition noise is known by several different names: electronic ignition suppression (EIS), interference absorption circuit (IAC), and noise blanker (NB). Strategies for detecting ignition noise pulses vary greatly among manufacturers. The noise-blanking system may momentarily mute the audio signal or perform some type of piecewise linear approximation to the signal being repaired.

Analog Cassette. The analog compact cassette (commonly known today as the cassette) is based on conventional multiple-track stereo analog magnetic recording technology.[14, 19] The cassette as we know it today was developed in 1963 by Philips Corporation primarily as a dictation medium. Advances in magnetic tape formulations and the advent of the Dolby™ B Type Noise Reduction System transformed the cassette into a high-fidelity music storage medium. Modern cassette players can achieve a dynamic range of over 70 dB and a frequency response approaching 20 kHz. Through the early 1990s, the cassette continued to be the largest revenue-producing medium for prerecorded music. More recent advances such as improved noise reduction systems and improved tape formulations are continuing to extend the life of this robust magnetic medium. As a result of its maturity and ruggedness, the cassette continues to be the most popular music carrier for automotive applications.

Compact Disc. The fundamental technology that led to the development of the compact disc[14] began in the 1970s at Philips Laboratories. The compact disc format as it exists today was codeveloped in 1980 by Philips Corporation and Sony Corporation. The basic system is an optical, noncontact system, which is also used for video (laser disc), information storage (CD-ROM), interactive media (CD-I), photo storage (Photo-CD), and other applications (CD-V, CD-G).

The playback system for a compact disc is composed of two basic subsystems: the servo and control system and the audio data processing system (Fig. 25.3). The data is read from the disc via a laser diode pickup. The focus and tracking of the pickup is controlled via the closed loop servo control system. The audio information is stored in digital form at a sampling rate of 44.1 kHz (early proposals called for a rate of 44.056 kHz for compatibility with existing video storage formats) with a resolution of 16 bits. This provides a dynamic range of over 90 dB and frequency response to 20 kHz. This digital bitstream is stored as a continuous spiral

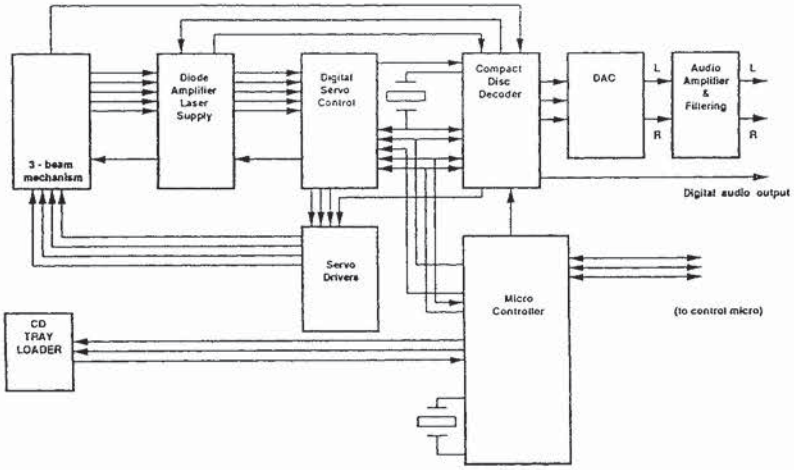


FIGURE 25.3 Typical compact disc playback system.

track beginning at the center of the disc. The data is interleaved throughout the disc to provide robustness to errors due to scratches or dust. To further minimize errors, a Cross Interleaved Reed-Solomon Code (CIRC) is used. Due to the overhead required for interleaving and error correction, the raw bit rate from a compact disc is about 4 megabits per second. The ultimate output of a CD system can be a stereo analog output and/or a digital bitstream output. This digital output has been error-corrected and can be in one of several forms. The most common are the IEC[20] serial format which is also known as SPDIF (Sony-Philips Digital Interface Format) or I^2S. The IEC format is transmitted via a single wire or possibly via a fiber optic cable. The I^2S format contains separate data, bit clock, and word (L/R) clock signals.[18]

25.3.2 Signal Processing

Frequency Response Modification: Tone Controls and Graphic Equalizers and Loudness Compensation. Personal taste is certainly an important aspect in the enjoyment of any audio system. Given that the listener is presented with a music source that is somewhat flat in the frequency domain, that listener may still prefer to accentuate (boost) or attenuate (cut) certain frequency bands. The bass and treble controls provided on most of even the simplest automotive radios provide the user with the flexibility to do this. The bass control can either boost or cut the low-frequency extreme of the audio spectrum; the treble control can boost or cut the high-frequency extreme of the audio spectrum. Figure 25.4 shows the response of a typical set of bass and treble controls. For further frequency content modification, a graphic equalizer provides greater flexibility and control over the audio spectrum. This device divides the audio spectrum into several bands and allows boost or cut within these bands. Figure 25.5 shows a composite response of a typical graphic equalizer. The type of frequency response alterations described here are user-adjustable. Many automotive audio systems, particularly the higher-end systems, employ fixed equaliza-

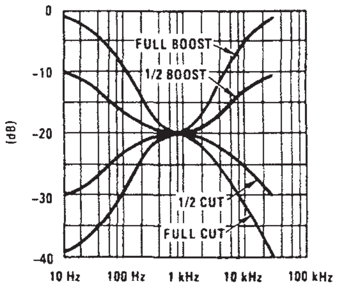


FIGURE 25.4 Bass/treble characteristic.

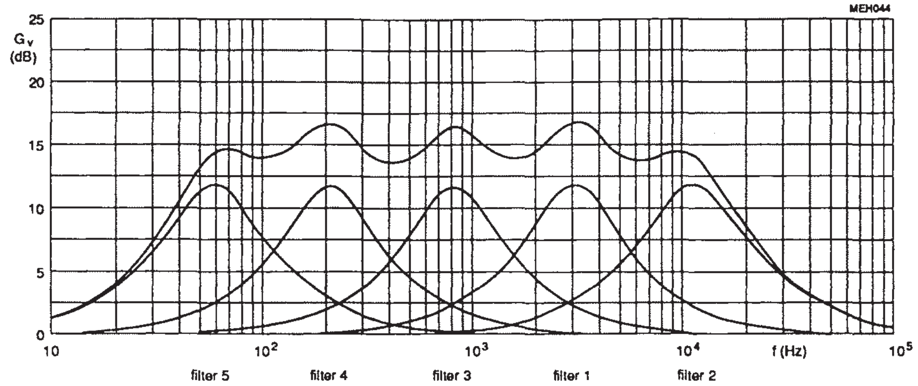


FIGURE 25.5 Graphic equalizer characteristic.

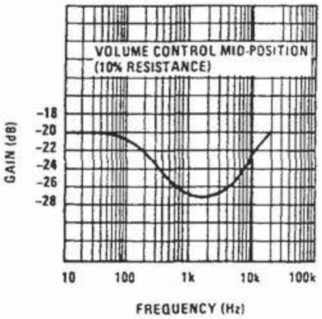


FIGURE 25.6 Loudness characteristic.

tion, which compensates for the acoustics of the vehicle interior. This type of vehicle-specific equalization is not adjustable by the user, but certainly provides an improved sound quality in the vehicle (see Vehicle Acoustics in Sec. 25.3.3).

Given the nonlinear response of the human ear to changes in level (recall the amplitude response curves), many users find it desirable to have the audio system automatically compensate for the decreasing sensitivity of the ear at low frequencies. The loudness feature is a type of automatic low-frequency boost. At lower settings of the volume control, the low-frequency output of the system is automatically boosted. This has the added benefit in an automotive environment in that a significant portion of the noise floor in an automobile cabin is low frequency (see Fig. 25.6 for a typical loudness characteristic). Some loudness characteristics also boost the high frequencies at lower volume settings. The loudness function is often hardwired into the audio system and cannot be defeated by the user.

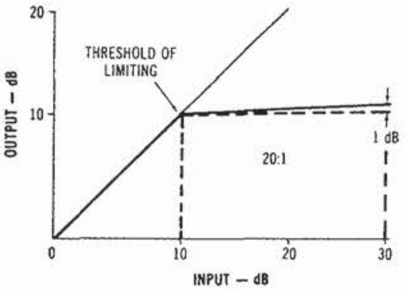


FIGURE 25.7 Limiter characteristic.

Dynamic Range Modification: Compressors and Limiters. Given the extreme capabilities of digital media such as the compact disk, compressors and/or limiters may be installed in the audio system. In order to limit these excursions, a limiting circuit may be applied (see Fig. 25.7 for the transfer function of a typical limiter). Note that in this example transfer function, for a signal above the threshold, a 20-dB change in input level will result in a 1-dB change in output level.

Wide dynamic range results in music with both extremely loud and extremely quiet passages. Given that the noise floor in an automobile environment is about 50 dB, quiet passages below that level will not be heard. A compressor will not only make the louder sounds quieter, but will also make the quieter sounds louder, thus bringing the low signal levels above the noise floor. Limiter and compressor functions may be user-switchable or hardwired on by the system manufacturer.

Power Amplifiers. For the listener to perceive sound, the air must be compressed. The transducer that accomplishes this is the loudspeaker. The power amplifiers in the audio system are responsible for providing the relatively high current drive to these loudspeakers. Modern automotive audio systems have six or more separate channel amplifiers, each of which may be required to provide up to 24 W. In order to achieve these power levels with an automobile voltage source of about 14 V, circuit techniques such as bridge-tied-load (BTL) are used.[14] Other techniques on the horizon promise more efficient delivery of power.

25.3.3 System Considerations

Signal Distribution. In the earliest days of automotive audio systems, the output of the radio was a simple loudspeaker signal which drove the loudspeaker directly. Modern systems, with multiple tuners and speakers, require new signal distribution for signal distribution. Further, the automotive audio system may have other essential portions of the system located outside of the traditional location in the instrument panel. For instance, recent systems locate the tuner directly at the antenna to reduce losses in the antenna wiring. Also, to reduce the I^2R losses in the speaker wiring and connectors, audio amplifiers will be located at or near the loudspeakers. This siting also minimizes the weight of lengthy high-gage wiring to the loudspeakers.

With these requirements, it is now required that system suppliers distribute audio and possibly RF signals throughout the vehicle. Depending on the noise environment of the vehicle, this may require something more elaborate than the old standard 14-gage lengths of stranded wire. Audio signals with a bandwidth of 20 kHz and nominal peak-to-peak amplitudes of 3 V must be distributed with great care. Alternatives are balanced signal drivers driving a twisted pair of wires which may have an integral shield. The twisted wire and shield minimize noise coupling into the wire. Then a receiving amplifier with good common mode rejection ratio is used to cancel out any noise pickup that may have occurred.

With increased digitization of the automotive audio system, it is worth considering distributing the signal in digital form[20] (see Sec. 25.4.1 on Media and sec. 25.4 on Future Trends). This provides a most robust method of distributing the signal without concern for noise pickup. The distribution media can be wire or fiber optic cable. If a wire is used, a shielded cable should be used to prevent radiation from the signal cable.

Standards are being proposed which may lead to control information being simultaneously transmitted along with the digital audio bitstream.[21] Thus, signal and control information share the same signal distribution media, further lessening the weight and cost impacts of multiple interconnects. This is an extension to the SPDIF standard which would provide for the automotive audio system components to be tied together via a "ring" not unlike a business computer network.

Vehicle Acoustics (Vehicle Cabin Characteristics). The interior cabin of a vehicle presents a most challenging acoustic environment. The small cabin dimensions produce standing waves which cause peaks and valleys in the low-frequency acoustic response of the vehicle interior. The highly absorptive materials (carpeting, seat material) combined with highly reflective glass, generate frequency response abberations across the audio frequency band. It should be noted that when making acoustic measurements on a vehicle, one must not rely on spot measurements. The response at any one point in the interior volume is quite irregular, and the response at various points throughout the interior volume can vary greatly. It is recommended that some type of averaged response be used to give an accurate view of what the response at various points in the interior truly is.[22]

Several techniques can be used to combat problems in vehicle acoustic problems. Acoustically transparent materials for speaker grilles are essential. Some common sense is easy to apply here: Carpeting is not acoustically transparent. Loudspeakers aimed at the listener's ears rather than pointed at a reflective surface (or within a map pocket) provide a more acoustically transparent response. Many premium or high-level audio systems provide fixed equalization, which is usually applied in the head unit or power amplifier signal processing sections. This equalization is ideally the mirror image of the vehicle cabin response, resulting in a combined response that is near flat.[22]

There is an easy way to demonstrate the frequency response irregularities caused by reflective surfaces: listen to someone talking with their hands cupped around both sides of their mouth, as they would do while shouting, or have that person hold two hardcover textbooks on either side of their mouth while talking.[23] The nasal type of sound experienced is due to the reflections within the hands or books.

A more subtle experience while listening to an automotive audio system is that of the virtual acoustic image. A note on imaging is appropriate here. Presumably, one has experienced the virtual image formed while listening to a home audio system. Any musical material that is produced equally from the left and right speakers of a properly set up audio system will appear to come from a virtual loudspeaker midway between the two real speakers. The ideal acoustic image within a vehicle would be front and center with the musical instruments spread across the vehicle interior. Due to the nonideal placement of the loudspeakers, the image in the vehicle may appear somewhere in the floor, the back seat, or some other nonideal location. A simple approach—although often difficult to execute—to combat this problem is to locate the loudspeakers at or close to ear level and direct them at the listener's ears. Furthermore, in a vehicle, the listener(s) are never located centrally between the loudspeakers. An interesting approach to fixing the noncentral location of the listener is to apply delays appropriate to center the image.[24]

25.4 FUTURE TRENDS

The future of automotive audio systems proves to be most exciting. Driven by the rapid increase of more powerful microcomputers, new communication techniques, improved media, and data storage techniques, the systems will change drastically through the end of this millennium.

25.4.1 Media

Radio Broadcast Data System (RBDS). The Radio Broadcast Data System is an extension of the Radio Data System (RDS) which has been in use in the European community since 1984. The system allows the broadcaster to transmit text information at the rate of about 1200 bits per second. The information is transmitted on a 57-kHz suppressed subcarrier as part of the FM MPX signal.[25,26]

RBDS was developed for the North American market by the National Radio Systems Committee (NRSC), a joint committee composed of the Electronic Industries Association (EIA) and the National Association of Broadcasters (NAB). The possibilities for applications of text transmission to the vehicle are numerous. For instance, song title and artist, traffic, accident and road hazard information, stock information, or weather. In emergency situations, the audio system can be enabled to interrupt the cassette, CD, or normal radio broadcast to alert the user, then return. Currently there are at least 50 FM stations in the United States transmitting and testing the RBDS system.[27] The system is not quite mature as of this writing, but it is anticipated to mature rapidly over the next five years.

Digital Audio Broadcast (DAB). Digital Audio Broadcasting is designed to provide high-quality, multiservice digital radio broadcasting for reception by stationary and automotive receivers. It is being designed to operate at any frequency up to 3 GHz. The system as being investigated has been developed by the Eureka 147 Consortium. This system is being demonstrated and extensively tested in Europe, Canada, and the United States. The system is a rugged, yet highly spectrum- and power-efficient sound and data broadcasting system. It uses advanced digital techniques to remove redundancy and perceptually irrevelant information from the audio source signal, then applies closely controlled redundancy to the transmitted signal for error correction (see section on recordable media that follows). The transmitted information is then spread in both the frequency and the time domains so a high-quality signal is obtained in the receiver, even under severe multipath propagation conditions. The feature of frequency reuse will permit broadcasters to be extended, virtually without limit, using additional transmitters, all operating on the same radiated frequency. A common worldwide frequency in the L band (around 1.5 GHz) is being considered, although some disagreement still exists.[28,29] The implemenation of DAB is inevitable. The only question is when.

Recordable Media. Recent advances in recording technologies and digital signal processing have driven the development of two competing recordable media formats: digital compact cassette (DCC) and MiniDisc. The digital audio tape (DAT) has not been forgotten, but its position as an automotive medium is dubious.

Digital Compact Cassette. The DCC was invented by Philips as the marriage of the compact cassette and compact disk quality digital audio. Market forces of existing libraries of analog cassettes were deciding factors in developing a physically compatible format. A DCC player can also play analog compact cassettes. Although very similar physically, construction and recording technique used in the DCC is vastly different from the analog compact cassette. The DCC is a stationary head magnetic recording system using nine parallel tracks, each 185 mm wide. To achieve these miniature dimensions, the DCC record/playback head assembly calls on the high-tech thin-film head technology already well proven in multichannel professional recording. In order to accommodate playback of analog compact cassettes, the head unit also provides analog heads (Fig. 25.8).

In order to achieve high-quality digital audio in a limited bandwidth, the DCC relies on the psychoacoustic properties of the human hearing system. The DCC system compresses the digital audio data according to the masking properties of the human hearing system. The threshold of hearing referred to earlier in this chapter is a static characteristic and is actually a dynamic threshold. The characteristics vary with the type of sound presented to the ear (Fig. 25.9). When the ear is presented with a relatively loud sound, it cannot detect a quieter one. This phenomenon depends on the relative frequency and amplitude of the signals involved. Philips has developed the precision adaptive sub-coding (PASC) technique which implements in high-speed digital signal processing (DSP) circuits the masking properties of the human hearing system. The system determines what portion of the audible spectrum the listener is able to hear and dedicates signal processing and compression algorithms to store that portion of the signal most efficiently.[30] Testing has shown that the DCC system is actually capable of storing digital audio data to a better resolution than the compact disc.[31]

Eight of the nine tracks contain all the PASC and error correction data. The ninth track holds mainly track and time information, similar to the compact disc. The PASC data is spread across the tape in a checkerboard pattern that increases the system's robustness against dropouts. This technique is similar to the interleaving of data used in the compact disc. Cross Interleaved Reed-Solomon Code (CIRC) protects the data against further errors. Like with the compact disc, a significant portion of data can be lost without affecting the data output

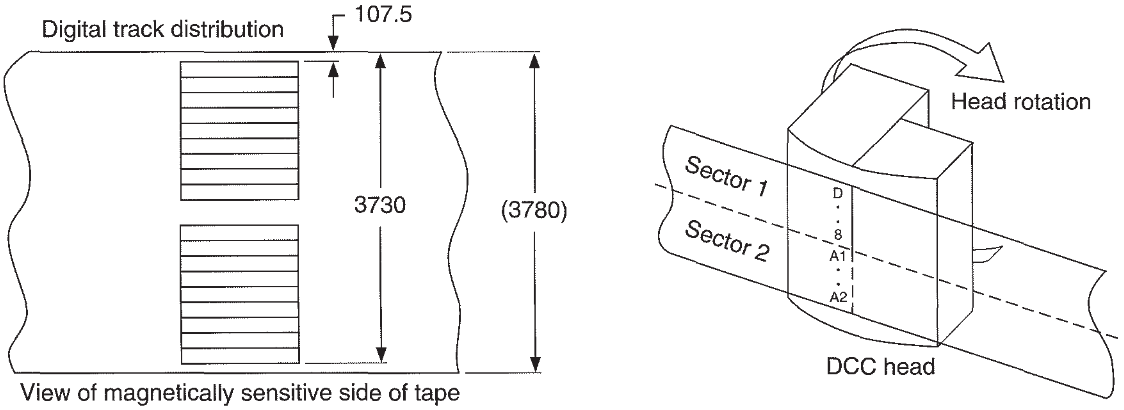
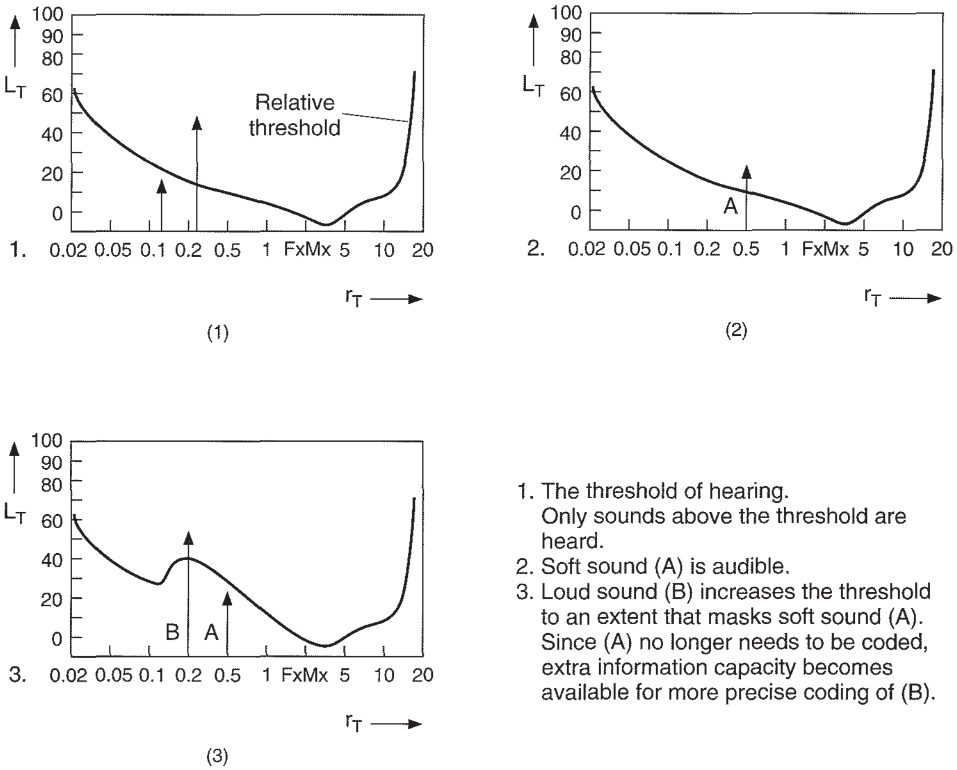


FIGURE 25.8 DCC head assembly.



1. The threshold of hearing. Only sounds above the threshold are heard.
2. Soft sound (A) is audible.
3. Loud sound (B) increases the threshold to an extent that masks soft sound (A). Since (A) no longer needs to be coded, extra information capacity becomes available for more precise coding of (B).

FIGURE 25.9 Hearing threshold.

(Fig. 25.10). The data storage requirements of the DCC require a digital bitstream of a maximum of 353 kilobits per second. This is as compared to the uncompressed bitstream which has a data rate of 1.4 megabits per second. This is a data reduction of four times. Notice that this requires a significant amount of computational power. Special-purpose, hardwired digital signal processors are used to accomplish the decoding task in real time (see Fig. 25.11 for a block diagram of and Table 25.2 for specifications of a complete DCC system).

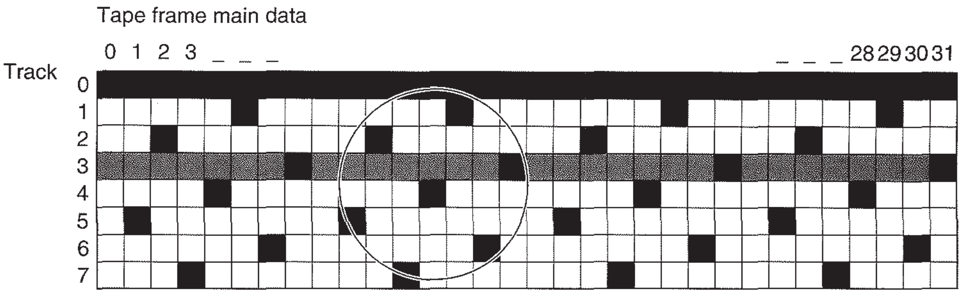
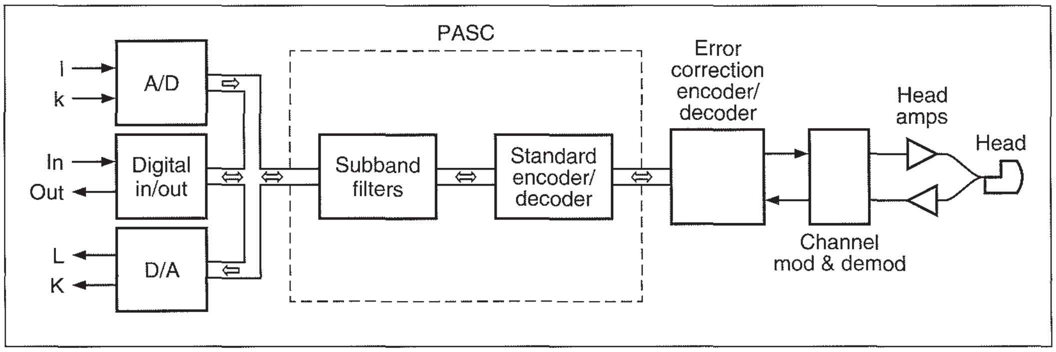


FIGURE 25.10 DCC data storage. A block as large as the circle, or an entire track can be lost without affecting data recovery.

TABLE 25.2 Digital Compact Cassette Technical Specifications

Achievable audio performance	
Number of channels	Stereo
Frequency range at fs = 48 kHz	5–20,000 Hz
at fs = 44.1 kHz	5–20,000 Hz
at fs = 32 kHz	5–14,500 Hz
Dynamic range	>105 dB
THD (including noise)	>95 dB
Wow and flutter	Quartz crystal precision
Signal format	
Sampling frequencies	48 kHz, 44.1 kHz, 32 kHz
Coding	PASC
Audio bit rate	384 bits/s (at 48 kHz)
Error correction systems	C1, C2 Reed-Solomon block code
Modulation system	8–10 (ETM)
Pre-emphasis	Optional
Cassette	
Recording time	Up to 2 × 45 min (D90) Provision for 2 × 60 min (D120)
Tape type	(Video) chrome or equivalent
Tape width	3.78 mm
Tape speed	4.76 cm/s
Number of tracks	8 digital audio, 1 subcode
Track width	>185 µm
Track pitch	195 µm

MiniDisc (MD). The MiniDisc[32] was developed by Sony as a combination of the compact disc and a recordable medium. The recording technology used is a magneto-optical (MO) system. The form factor of the medium is very similar to a common 3.5-in floppy disk. The encoding technique in principle is similar to the PASC system. The encoding technique employed here is known as ATRAC. The MD is certainly a rugged portable medium, although subjective testing indicates that the response of the MD ATRAC system is not quite capable of encoding compact disc quality sound.



Schematic diagram: DCC

FIGURE 25.11 Complete DCC playback system.

25.4.2 Signal Processing

Digital Signal Processing Technology. Currently, we are within a radical change in signal processing in all fields of consumer and industrial electronics. The proliferation of sound capabilities in the commonplace personal computer is a testament to this change, with further improvements and functionality coming yearly. Driven by the requirement of increased feature content and the rapidly reducing price of high-speed digital semiconductors, we will truly see a radical change in signal-processing architectures through the end of the century. Consider that the price of a state-of-the-art programmable digital signal processor (DSP) integrated circuit in 1980 was hundreds of dollars. This part could perform on the order of 1 to 5 million instructions per second (MIPS).[33] Today, a state-of-the-art programmable DSP is certainly less than one hundred dollars, with numerous units available for five dollars. Furthermore, today's DSPs will support advanced features such as shared memory, parallel processing, and high-level language support. We are beginning to see further levels of functional integration whereby formerly external functions such as analog-to-digital converters, digital-to-analog converters, and memory management units are integrated directly on the DSP. Simultaneously, these parts are exceeding 50 MIPS.[34] Additional advances are being made in integrating fixed function digital logic on the DSP. The result is that future DSPs will be tailored to specific tasks with further reduced cost. Current state-of-the-art integrated circuit manufacturing technologies are at the 1-micron feature width. Trends indicate that this will be approaching (and likely surpassing) the 0.5-micron level within the next decade. Thus, we can expect a quadrupling of functionality in the same silicon die area.

Automotive Audio Applications. Certainly the impact of the rapid advances in DSP technology on automotive entertainment systems will be far reaching. The initial impact has already happened in the area of simple audio signal processing. Numerous aftermarket and OEM offerings contain room emulation or Hall effect features. Similar to the effect available in home entertainment systems, this effect simply emulates the reflective and absorptive properties of various listening environments in the automobile. The user can then experience the ambience of a jazz room, stadium, or concert hall. Here, DSP is an enabling technology. While room emulation can be done in analog technology, it is most expensive and performs poorly. Given the sampling frequency of 44.1 kHz of most digital audio media, the DSP must be capable of processing the signal data within one sample period (about 27 μs). This is relatively easy with available DSP chips. The proliferation of digital media furthers this enabling. Analog media must first be converted to digital before the signal can be processed in the digital domain. Given that a medium (DCC, MD, or DAB) exists in a digital form, processing the signal immediately is straightforward; no conversion is necessary. A discussion of the relative merits of analog versus digital signal processing is beyond the scope of this discussion. The significant features of digital signal processing are accurate design results, ease of feature addition and extension, predictable quality, and ease of manufacturability. Consider the requirement for the addition of adding particular equalizer characteristics to an analog radio that does not have the equalizer feature installed. Considerable redesign is required to add the analog components required to achieve the equalizer function. In a DSP system, adding the equalizer function is simply a matter of adding the proper algorithms in the coding of the processor to achieve the desired signal response.

With the rapid increase in speed of DSP circuits, the amount of signal processing that can be done with the DSP is rapidly increasing. Speeds of commercially available DSP chips are such that they are capable of processing the multiplex signals of an FM receiver. Now the methods of combating multiplex distortion and other reception problems are transferred from the realm of analog circuits to the implementation of DSP algorithms. This movement of the processing towards the front end of the receiver will continue. Processing the radio frequency signal directly seems to be just over the horizon.

The signal-processing capabilities of automotive entertainment systems have seen a many-orders-of-magnitude increase and have been driven by the fact that, today, transistors are less costly than staples.[35] Consider that in 1930 we had radios with six vacuum tubes. Before the year 2000 we will have audio systems with multiple control microprocessors and multiple DSPs totaling the equivalent of hundreds of thousands of transistors performing calculations on the order of GIPS (10^9 instructions per second).

GLOSSARY

Acoustics The study of the generation, transmission, and reception of sound.

Analog signal A signal in which the information of interest is communicated in the form of a continuous signal. The magnitude of this signal is proportional (or *analog*ous) to the actual quantity of interest.

CIRC (Cross Interleaved Reed-Solomon Code) The encoding process used for playback error correction in the compact disc playback system.

Digital signal A signal in which the information of interest is communication in the form of a number. The magnitude of this number is proportional to (within the limitations of the resolution of the number) the actual quantity of interest.

Digital signal processing (often abbreviated DSP—as a verb). The processing of analog signals which have been converted to digital form. The processing usually involves repeated additions and multiplications.

Digital signal processor (often abbreviated DSP—as a noun). A monolithic integrated circuit optimized for digital signal-processing applications. Portions of the device are similar to a conventional microprocessor. The architecture is highly optimized for the rapid, repeated additions and multiplications required for digital signal processing. Digital signal processors may be implemented as programmable devices or may be realized as dedicated high-speed logic.

Dynamic loudness compensation A type of loudness compensation which is dependent on the actual signal level rather than the position of the volume control. This is greatly enabled through the use of digital signal processor integrated circuits.

Equalization The modification of the frequency response of a system. This may be done to compensate for deficiencies in the transmission/reception system or based on the subjective requirements of the user.

Loudness compensation The characteristic of an audio system to boost the low- and (sometimes) high-frequency components of an audio signal. This is to compensate for the human ear's decreasing sensitivities at these ends of the spectrum. The amount of boost is usually dependent on the position of the volume or level control. See also **Dynamic loudness compensation**.

Monaural Traditionally refers to an audio signal that has only one channel of information. Also may refer to the signal component in a stereo transmission which is common between the two (left and right) channels. Also commonly referred to as "mono."

Multiplex stereo The system standardized in the early 1960s to transmit and receive a stereo signal. This system was designed to be compatible with existing FM systems which were monaural. See text for details.

Psychoacoustics The study of the interaction of acoustics and the human hearing and brain systems.

Signal A fluctuating quantity that is proportional to some physical quantity. In the context here, the signal is usually in the form of a voltage or current. See also **Analog signal** and **Digital signal**.

Signal processing Modification of the time or frequency characteristics of a signal.

Stereo Classically, from the Greek *stereos,* meaning "solid." Traditionally refers to an audio system which has two or more independent channels of audio information. A typical stereo system consists of primarily a left and a right channel. Modern surround sound systems contain from four to as many as seven channels of information.

Sound A vibratory disturbance in the pressure and density of a fluid or in the elastic strain in a solid, with frequency in the range of about 20 to 20,000 cycles per second, capable of being detected by the organs of the human hearing system.

REFERENCES

1. F. Alton Everest, "Fundamentals of Sound," Chap. 1 in Ballou (ed.), *Handbook for Sound Engineers,* 2d ed., Howard Sams, Indianapolis, Ind. 1987.
2. F. Alton Everest, "Psychoacoustics," Chap. 2 in Ballou (ed.), *Handbook for Sound Engineers,* 2d ed., Howard Sams, Indianapolis, Ind. 1987.
3. *Websters Ninth New Collegiate Dictionary,* 1985.
4. J. Hebrank and D. Wright, "Are Two Ears Necessary for Localization of Sound Sources in the Median Plane?," *Journal of the ASA,* V56N3, 1974, pp. 935–938.
5. Donald W. Matteson, *The Auto Radio (A Romantic Genealogy),* Thornridge Publishing, Jackson, Mich., 1987.
6. Natalie Cole, *Unforgettable,* compact disc, Elektra Records, 1991.
7. Frank Zappa, *Jazz from Hell,* compact disc, Barking Pumpkin Records, 1988.
8. Jehl, *Menlo Park Reminiscences,* 1937 and 1990, Henry Ford Museum/Dover.
9. Audio Engineering Society Centennial Booklet, *The Evolution of Recordings . . . From Cylinder to Video Disk,* 1977.
10. William F. Boyce, *Hi-Fi Stereo Handbook,* Howard Sams, Indianapolis, Ind., 1967.
11. Motorola Publication, *Introduction to the Motorola C-Quam System,* 1985.
12. *Stereophonic FM Broadcasting,* BTR, vol. 7, no. 2, July 1961.
13. R. Summerville, "Advances in Tackling FM Reception Problems," SAE 850022, Society of Automotive Engineers, Warrendale, Pa., 1986.
14. Philips Semiconductors, *Car Radio Design-in Guide,* Doc # 9398-706-42011.
15. B. Goldes, *Improved Test Generator for Radio Noise Blanker,* personal communication.
16. R. W. Landee, D. C. Davis, A. P. Albrecht, and L. J. Giacoletto, *Electronic Designers Handbook,* 2d ed., McGraw-Hill, New York, 1977.
17. D. Manquen, "Magnetic Recording," Chap. 27, in Ballou (ed.), *Handbook for Sound Engineers,* 2d ed., Howard Sams, Indianapolis, Ind., 1987.
18. *Publication 958-Digital Audio Interface, First Edition, 1989–03,* International Electrotechnical Commission (IEC).
19. G. Alexandrovich, "Recording and Playback," Chap. 25, in Ballou (ed.), *Handbook for Sound Engineers,* 2d ed., Howard Sams, Indianapolis, Ind., 1987.
20. "Digital Audio Input/Output Circuit (DAIO) TDA1315," Philips Semiconductors Data Sheet, 1993.
21. "Audio Local Area Network Transceiver Circuit (A-Lan) CS8425," Crystal Semiconductor Data Sheet, 1992.

22. E. Geddes, H. Blind, *The Equalized Sound Power Measurement Technique,* Audio Engineering Society, New York, 1989.
23. David L. Clark, *Effective Listening (Seminar Notes),* DLC Designs Publication/Seminar, Dearborn, Mich., 1990.
24. David L. Clark, Test Car, 1990.
25. Philips Semiconductors Internal Application Notes 901026R2, *RDS (Radio Data System).*
26. CENELEC European Committee for Standardization, *RDS Specification (EN 50 067).*
27. *Electronic Industries Association Publication,* May 26, 1993.
28. "Policy snags delay DAB," *Electronic Engineering Times,* Sept. 20, 1993.
29. M. Halbe, "DAB Receiver Technology and Implementation Considerations," *1st International Symposium on DAB,* 1992.
30. Philips Marketing Brochure, "*DCC Fundamentals—The New Digital Sound System.*"
31. David Clark, "DA/B/Xing DCC," *Audio Magazine,* April 1992.
32. Leonard Feldman, "The Mechanics of Sony's Minidisc," *Audio Magazine,* Dec. 1992.
33. Texas Instruments, TMS32010 Data Sheet, 1982.
34. Motorola, DSP56004 Data Sheet, 1993.
35. Robert W. Keyes, "The Future of the Transistor," *Scientific American,* June 1993.

ABOUT THE AUTHOR

Thomas Chrapkiewicz has been an applications engineer for Philips Semiconductors since 1993. As an analog design engineer from 1977 to 1983 at both ADM Technology (formerly Audio Designs and Manufacturing) and Harris Broadcast Products, he designed audio circuits and systems for broadcast, recording, and film applications. He then spent 10 years designing analog and digital signal-processing circuits for Ford Motor Co.

CHAPTER 26
MULTIPLEX WIRING SYSTEMS

Fred Miesterfeld
Chrysler Corporation

26.1 VEHICLE MULTIPLEXING

Production and proposed passenger vehicle multiplexing and data communications network systems will be thoroughly examined in this chapter. The systems covered are those methods that are relevant to the electronic engineer who has the assignment of applying multiplexing techniques to high-volume production. In passenger vehicle design, cost is the universal method of determining whether or not a design will be put into production. If a multiplex network design can be applied while delivering functional improvements at a system-cost saving, then this design is the most likely network design to be accepted.

The SAE Vehicle Network for Multiplexing and Data Communications (Multiplex) Committee has defined[1] three classes of vehicle data communication networks:

> *Class A.* A potential multiplex system usage where vehicle wiring is reduced by the transmission and reception of multiple signals over the same signal bus between nodes that would have ordinarily been accomplished by individual wires in a conventionally wired vehicle. The nodes used to accomplish multiplexed body wiring typically did not exist in the same or similar form in a conventionally wired vehicle.
>
> *Class B.* A potential multiplex system usage where data (e.g., parametric data values) are transferred between nodes to eliminate redundant sensors and other system elements. The nodes in this form of a multiplex system typically already existed as stand-alone modules in a conventionally wired vehicle.
>
> *Class C.* A potential multiplex system usage where high data rate signals, typically associated with real-time control systems, such as engine controls and antilock brakes, are sent over the signal bus to facilitate distributed control and to further reduce vehicle wiring.

The Class B network is intended to be a functional superset of the Class A network; i.e., the Class B bus must be capable of communications that would perform all of the functions of a Class A bus. This feature protects the use of the same bus for all Class A and Class B functions or an alternate configuration of both buses with a *gateway* device. In a similar manner, the Class C bus is intended as a functional superset of the Class B bus.

Generally, this section will deal only with the requirements for the lowest three layers of the seven-layer ISO open system interconnects (OSI) model (Ref. ISO 7498). These layers in descending order are the network layer, data link layer, and the physical layer.

26.1.1 Background of Vehicle Network Architectures

A wide variety of network topologies[2] can be envisioned by network designers. The message structure described in this section is very flexible and useful in exchanging information between network nodes. The following discussion describes two network architectures which are likely configurations that can use this message definition set: a single-network architecture and a multiple-network architecture.

The selection would be application-specific and, thus, it is the system designer's choice as to which network architectures to use. It should be noted that the hardware that supports these two message structures is generally not interchangeable. It is recommended that care be taken in choosing which message definition to use, because the selection is generally irreversible because of hardware limitations.

Header Selection. The header field (header) is a one-, two-, or three-byte field within a frame and contains information about the message priority, message source, target address, message type, and in-frame response. The multiple network architecture is usually associated with the single-byte header protocol. Figure 26.1 (1) illustrates the header byte as the message identifier (ID), which is primarily used for functional "broadcast"-type messages and implicitly defines all the required information about the message. It is unnecessary to specify the source or destination of functional-type messages. Reception becomes the exclusive responsibility of the receiving node. Figure 26.1 (2) also illustrates header bytes, which are primarily used for physical-type messages, and has two bytes: the first is the ID and the second is the target address.

1. Header for Functional Message Frames

2. Header for Physical Message Frames

FIGURE 26.1 Single-byte header protocol.

The single-network architecture is usually associated with the multiple-byte header protocol, shown in Fig. 26.2. The first byte of the frame defines the priority and message types, functional or physical.

FIGURE 26.2 Multiple-byte header protocol.

Architecture Selection. Consideration must be given by the network designer as to whether a single-network architecture or a multiple-network architecture is preferable for an application. For example, a multiple-network architecture could be based on one network optimized around data communication (Class B) protocol requirements, and another network optimized around sensor type (Class A) multiplexing requirements. The Class B network may be characterized such that low latency is a significant requirement of the protocol and where the short functional type of messages can most effectively be used. A Class A network could handle the vehicle's event-driven multiplexing requirements. See the next section on Class A networking for more information on Class A multiplexing considerations.

Without regard to either the header or architecture selection, in Class B communications the network consists of the interconnection of intelligent nodes such as an engine controller, a body computer, a vehicle instrument cluster, and other modules. Such a network normally does not significantly reduce the base vehicle wiring but provides an intermodule data communications capability for distributed processing. The data shared between modules may be repetitive in nature and sometimes requires handshaking between modules or acknowledgment of data reception. As a result of handling the repetitive data and response-type data, a network can be optimized around functional addressing. Functional addressing sends data on the network, which can be received by one or more nodes without regard to the physical location of the module but only by their "interest" in those specific functions. In general, the transmitting node does not care which, if any, nodes receive the data it is sending. When physical addressing is required in a data communications network (Class B), it is usually for vehicle maintenance purposes and can be easily handled without reducing network bandwidth.

The nature of Class A multiplexing requires the interconnection of limited intelligence nodes, often simply sensors or actuators. These Class A networks can significantly reduce the base vehicle wiring as well as potentially remove redundant sensors from the vehicle. The data shared between nodes in this case are generally event-driven in nature. In most vehicles, the number of event-driven signals predominates, but they are only needed infrequently. The message to "turn headlamps on," for example, can be easily seen as event-driven. Because these messages are infrequent (only sent once when the signal changes) they generally require acknowledgment, either within the same message or a separate handshake/response message.

The single-network architecture carries both the Class A and Class B messages on one network and the multiple-byte header has the advantage of having more bits available for use in assigning message identifiers, priorities, message types, etc. The characteristics of both time-critical and event-driven messages must be accommodated on a message-by-message basis. In general, this level of complexity will need the flexibility of the multiple-byte header structure. It should be clear that both network architectures must be cost effective for the application and the specific nodes on each network.

The multiple network architecture tends to separate the Class A messages from the Class B messages and optimize each network and node interface for the specific characteristics of each network class. The time-critical messages could be exchanged on one network, while the event-driven messages are sent on another. For example, the data communication (Class B) repetitive messages can be handled on one network and the sensor and control (Class A) multiplexing requirements on another network. This architecture requires both networks to work together to achieve the total vehicle network requirements. If information is needed between the multiple networks, care must be exercised to meet the needs of each of the networks. This concept of multiple networks is not limited to two, but can be extended to several separate networks if desired.

Class A Network. Class A multiplexing[3] is most appropriate for low-speed body wiring and control functions. The example most often used to illustrate the benefits of Class A multiplexing is the base exterior lighting circuit. However, this example is the hardest function to cost-justify. The base exterior lighting system is extremely simple and very low cost. A multiplex network applied to this lighting system could result in increased wiring complexity and cost. Data integrity in the lighting system can be a stringent requirement for Class A multiplexing; e.g., a single-bit error that results in headlights "off" when they should be "on." Adequate data integrity in a Class A multiplex network is a constraint and bit-error checking may be required.

In the future, the results could change if new features, such as low-current switching or lamp-outage warning, became a requirement or new lamp technology, such as smart bulbs, became a reality. In general, the addition of new features will play a major role as to when and how multiplexing will become a cost-effective solution.

Other Driving Forces. The design of vehicles to minimize manufacturing complexity is a major force that will lead to architecture partitioning development. The properly developed multiplex architecture can be very effective in reducing the number of parts in the assembly plants and built-in testability can substantially reduce vehicle build test time.

Example Class A Systems. To illustrate how a Class A multiplex network could be used to simplify the vehicle wiring situation, consider the vehicle theft alarm system shown in Fig. 26.3. Although this example does not represent the epitome in theft alarm features, it does illustrate the nonmultiplexed condition. The horn actuator and the sensor switches are all wired directly to the theft alarm module. The module is then armed by activating the dash arm switch. The module can be disarmed by either the driver door key switch, passenger door key switch, or the trunk key switch. When the module is armed, the horn is sounded when the hood, door, or trunk is tampered with.

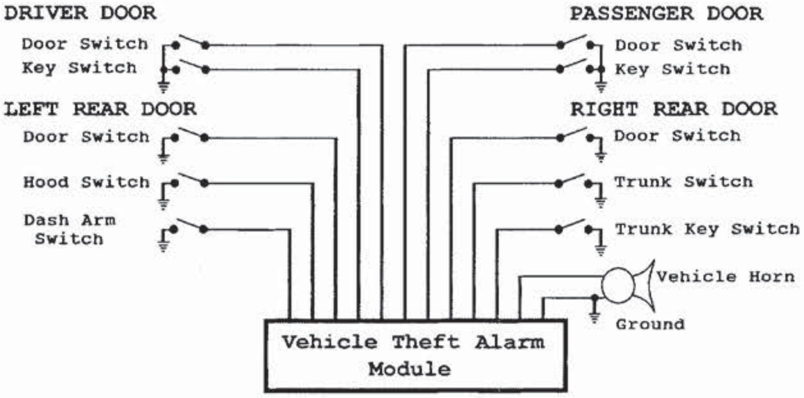


FIGURE 26.3 Vehicle theft alarm system.

The vehicle theft alarm system shown in Fig. 26.4 illustrates a near-optimal configuration of a Class A network. The sensors and actuators are integrated with the multiplexing electronics so that they can communicate over a single wire to the theft alarm module. The integration of electronics into the sensors and actuator improve sensor diagnostics because the sensor status and condition can be reported back to the controlling module. The integrity of the sensor status and condition can be linked to the mechanical operation of the sensor. This level of switch integrity cannot be achieved with normal switch-biasing methods. In a theft alarm system, there is an added benefit: the sensor condition can be used to set off the alarm and foil the tampering of a would-be thief.

The I/O requirements support T-tap connections, which can be highly automated in the production of wiring harnesses, reducing the wire bundle size and eliminating dual crimps. The configuration also supports the concept of adding sensors or actuators as the option requires without changing the theft alarm module configuration to support the optional features. This expandability feature allows the cost of the option to drive the system cost.

To show how this configuration is flexible and easily expandable, consider the example condition in which some versions of theft alarms are built as originally described, but an upscaled version is offered as an option in which the unit is armed by the driver locking the doors. To support this option, the dash arm switch would be eliminated and the driver door lock switch would be configured with the integrated switch multiplex at a different address. The same theft alarm module's software could then reconfigure itself without hardware modifications.

There are approximately seven sensors to every actuator in a real vehicle body system. This theft alarm system is typical with 10 sensors (switches) to one actuator (horn).

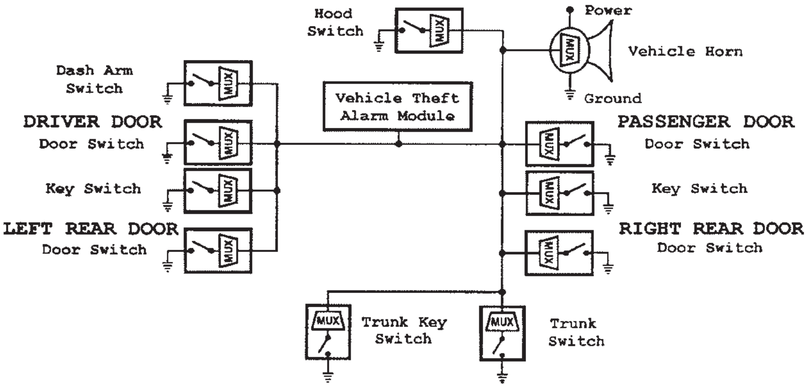


FIGURE 26.4 Multiplexed vehicle theft alarm system.

The sensors and multiplexing electronics can be integrated into the switch component. This configuration eliminates separate wiring and mounting of the multiplex module. Some component manufacturers have even been working on two wire (signal and ground) sensors in which the power to run the sensor has been supplied by the multiplex signal. For an example component, see Sec. 26.3.7. These sensors have been designed to include the multiplex circuit integrated with the Hall effect device in the same TO92 size package. The multiplexer portion is very small and requires approximately 300 logic gates.

The actuator driver and multiplexer can similarly be integrated into the horn or motor. This configuration also reduces wiring and mounting complexity. Actuators normally require more power then sensors and usually require three wires; signal, power, and ground. However, some manufacturers are developing a method to eliminate one of these wires by placing the signal on the power wire.

Class B Data Communications. The vehicle system designer now has many architecture partitioning options. A prime example is when to integrate many features into a module or when to employ a dedicated node. Care must be taken or the partitioning strategy may not achieve optimal results. The issue is much more complex when vehicle multiplexing is involved in this partitioning strategy. The most popular networking strategy is the Class B single-network architecture. However, this architectural strategy does not always result in an optimal solution.

A hypothetical vehicle will be described to illustrate this point. Figure 26.5 illustrates the part of a data communications network that contains a body computer, an instrument cluster, and a message center. In this example, all the sensors that feed the network enter through the body computer.

As illustrated in Fig. 26.5, all sensors are wired directly to the body computer. This example shows that for a base vehicle with only a small amount of electronic content, where all the sensors are directly wired to the body computer, the wire bundle size and number of connector pins is attainable. As additional features are made standard, either by consumer demands or government regulations, it becomes more and more difficult to implement the required system. This added complexity is due to the tremendous number of interconnecting wires from sensors to the modules. The build complexity and troubleshooting problems make this option a limited solution for this partitioning strategy.

The Class B single-network architecture strategy would solve this complex problem by adding a sensor node and reducing the number of interconnecting wires. Conventional sensors are connected directly to the node which serves as a gateway to the other modules over the Class B data network. Figure 26.6 illustrates the dramatic reduction in the number of cir-

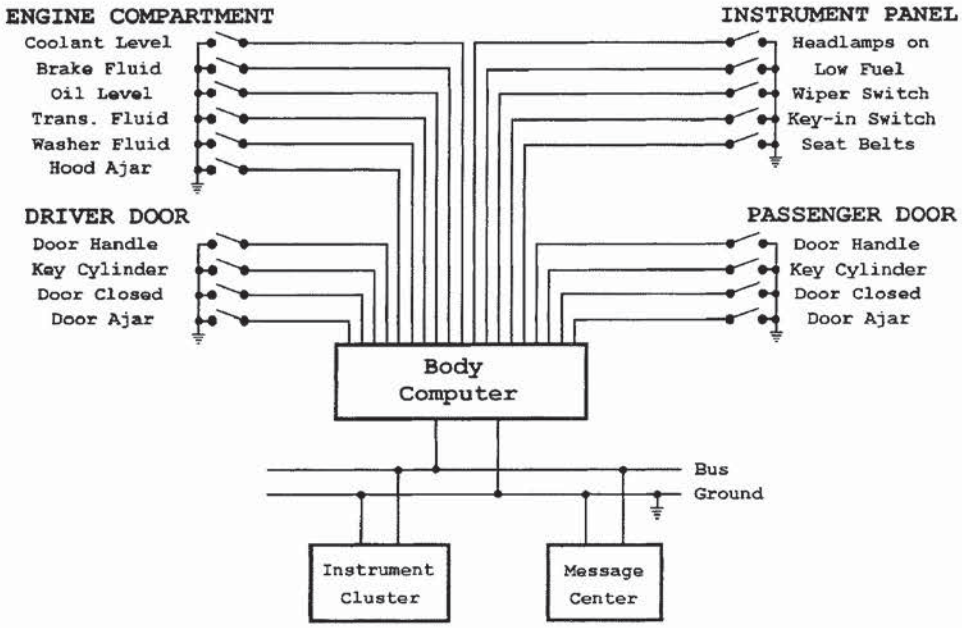


FIGURE 26.5 Data communications network with body computer, instrument cluster, and message center.

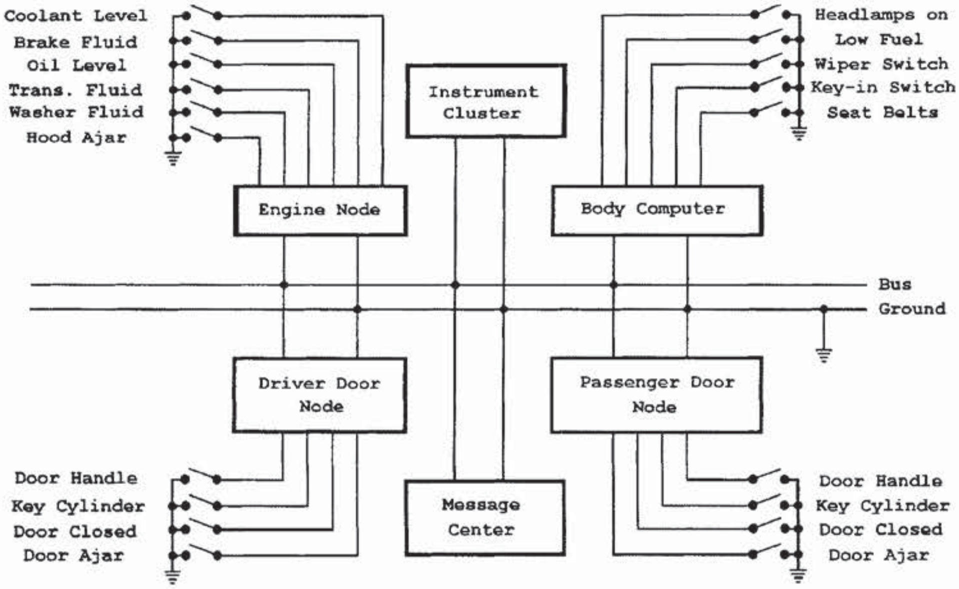


FIGURE 26.6 Data link with body computer, instrument cluster, message center, and three sensor nodes.

cuits required. This method is effective in reducing the number of sensor wires connected through "crunch points" such as the bulkhead or door hinge. This reduction in wiring, however, is obtained at the expense of three added sensor nodes.

Class B multiplexing is a very useful technique for reducing many of the problems encountered by the automotive system engineers. This section will demonstrate that in many situations the multiplex strategy, shown by Fig. 26.6, leads to a less than optimum system architecture. It is highly desirable to have a multiplexing architecture which would permit the use of smaller module connectors, reduce the number of wires crowding through the congested areas, and attain a solution without introducing more modules to mount, wire, and service.

Engine Compartment Node. In this hypothetical example, it may be desirable to integrate the node with the engine controller module, which would reduce module count and wiring circuits at the same time. The integration solution is not always possible because the engine controller already has an uncontrollably large module connector and would add a separate part just to cover an option.

Door Nodes. In this hypothetical example, the best location for the door node would be inside the door (See Fig. 26.7). By placement inside the door, the number of circuits through the door hinge is minimized, but without making further improvements the same wiring complexity inside the door still exists. These further improvements generally could integrate the electronics and mechanics into a single package.

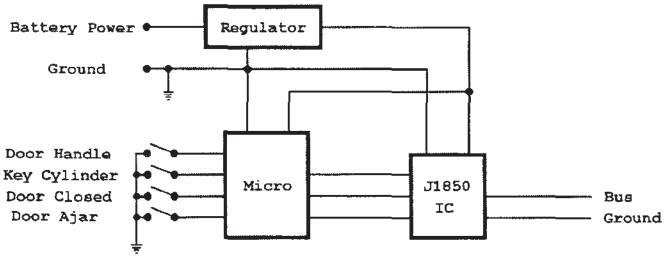


FIGURE 26.7 A driver door node.

General Node Concerns. General node concerns include the following:

- In order to achieve minimum cost, nodes tend to become application-specific and not generic because they usually can cover only one feature product.
- To cover more than one product, nodes tend to become intelligent and employ a microcomputer, and may negatively impact the system cost and complexity.
- Using conventional sensors remote from the node does not normally improve sensor diagnostics; e.g., the node cannot tell if the sensor switch is off or if the wire is disconnected. Refer to the previous section, "Example Class A Systems," for a discussion on switch integrity.
- The door node illustrated in Fig. 26.7, and nodes in general, can be effective in some wire bundle size and weight reduction but further improvements are possible with Class A sensor/actuator networking. The number of connector pins for the system can also be reduced with Class A networking.

Multiple Network Architecture. The multiple network architecture is the second strategy that solves many of these concerns. This architecture requires the development of many types of specialized network hardware components to efficiently handle each application. These components are connected together by a gateway on the Class B network for diagnostics purposes. Figure 26.8 illustrates this local area network (LAN) solution.

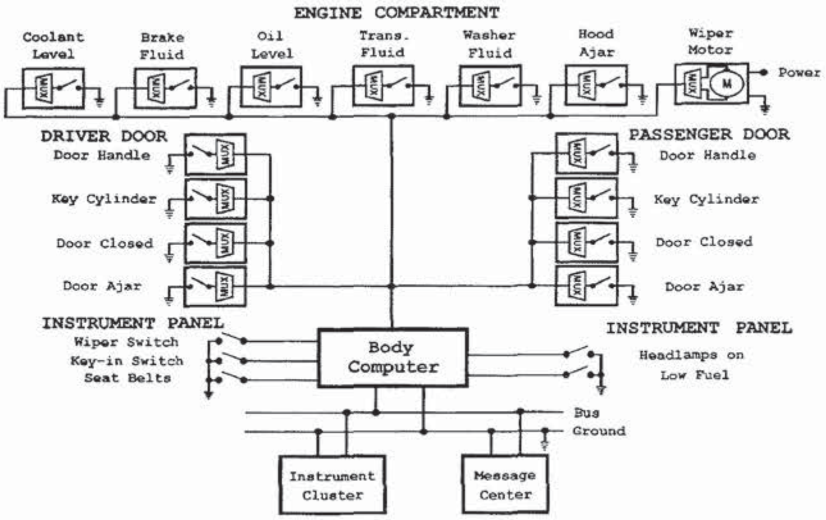


FIGURE 26.8 Data link with body computer, instrument cluster, message center, and Class A network for sensors.

Multiple network architecture strategy requires the integration of electronics into the sensors, actuators, and motors so that they can communicate over a single wire into the module that utilizes them. Since the sensor and actuator components contain the added multiplex electronics, the separate installation and wiring of the multiplex module is eliminated. Unlike the single-network architecture strategy, the integration of electronics into the sensors, actuators, and motors normally does improve sensor diagnostics because the sensor status and condition can be reported back to the controlling module. This method makes use of a Class A LAN without adding components to the vehicle system.

Figure 26.8 shows that the Class A LAN eliminates the need for the engine compartment node and two-door sensor module while still reducing wiring at the crunch points. The multiplex architecture shown in Fig. 26.8 significantly simplifies the same system shown in Fig. 26.6. This simplification is made possible by separating the Class B intermodule communications network from the Class A sensor-to-module communications.

The Class A LAN connects all the multiplexed components in parallel. The I/O requirements support T-tap connections, which can be highly automated in the production of wiring harnesses. This configuration reduces bundle size and eliminates dual crimps. The configuration also supports the concept of adding sensors or actuators, as the option requires, without changing the body computer configuration to support the option. This add-on feature allows the option to dictate the cost and not the cost of the added node dominating.

The two different Class A networks using multiplexed sensors are shown in Fig. 26.8 and Fig. 26.9 and illustrate a body computer flexibility trait that is not available in the other architectural approaches. The typical base vehicle using the Class A network is shown in Fig. 26.9. The body computer in the base, medium, and premium vehicle systems all have the wiper switch, seat belt switch, headlamps on, low fuel level, key-in ignition switch, Class A interface, and the Class B multiplex interface. This trait across option rates allows additional inputs to be connected without modifying the hardware in the body computer.

Unlike this hypothetical situation just used for illustration purposes, a real vehicle will have several actuators as well as sensors connected to the body computer. As it was previously discussed, there are approximately seven sensors to every actuator in a real vehicle body system. For illustrative purposes, 14 sensors and one actuator (wiper motor) were shown in this example. The principles shown, however, apply similarly to any number of actuators.