

Failure Modes of Electronics



Frances Metzger

First Edition, 2011

ISBN 978-93-81157-19-0



© All rights reserved.

Published by:

The English Press

4735/22 Prakashdeep Bldg,

Ansari Road, Darya Ganj,

Delhi - 110002

Email: info@wtbooks.com

Table of Contents

Chapter 1 - Failure Modes of Electronics

Chapter 2 - Electromigration

Chapter 3 - Electrostatic discharge, Overvoltage & Overcurrent

Chapter 4 - Semiconductor-Related Failures

Chapter 5 - Stress-Related Failures

Chapter 6 - Optoelectronics Failures

Chapter 7 - Microelectromechanical Systems Failures

Chapter 8 - Vacuum Tubes Failures

WWT

Failure Modes of Electronics

Electronic devices have a wide range of failure modes. These can be distinguished by their development in time (sudden failure or gradual degradation), by environmental effects (e.g. corrosion, ionizing radiation) or by the electrical parameter which was exceeded (e.g. electrostatic discharge, overvoltage, overcurrent, etc.).

Failures most commonly occur at the beginning and near the end of the lifetime of the parts. Burn-in procedures are used to detect early failures.

Presence of parasitic structures, irrelevant for normal operation, may become important in the context of failures; such structures can be both a source of failure and a protective device.

A sudden fail-open failure can cause multiple secondary failures, when the event is fast and the circuit contains an inductance. The suddenly interrupted current flow in combination with the inductance then causes large voltage spikes, which for very fast events may exceed 500 volts. A burned metallization on a chip may then cause secondary overvoltage damage.

Packaging-related failures

Electronic packaging, acting as the barrier between the materials of the electronic parts and the environment, is very susceptible to environmental factors. Thermal cycling may cause material fatigue due to mechanical stresses induced by thermal expansion, especially when the thermal expansion coefficients of the materials are different. Humidity or presence of aggressive chemicals can cause corrosion of the packaging materials, leads, or cause failure of encapsulation and following damage to the part inside, leading to electrical failure. Exceeding the allowed environmental temperature, whether too high or too low, can cause overstressing of the wire bonds inside the package, tearing the connections, cracking the semiconductor dies, or causing cracks to the packaging itself. Absorption of humidity into the packaging material and subsequent

heating to high temperature (e.g. during soldering) may also cause cracking. Mechanical damage could also fall here.

Majority of failure of electronics parts is typically packaging-related.

Bonding wires can be severed or shorted together during encapsulation, or touch the chip die, usually its edge. Dies can crack due to mechanical overstress or thermal shock; initial defects introduced during e.g. wafer sawing or scribing can develop to fractures later. Lead frame may contain excessive material or burrs, causing shorts. Ions of e.g. alkali metals or halogens can be released from the packaging materials and migrate to the semiconductor dies, causing corrosion or parameter deterioration.

Glass-metal seals commonly fail by forming radial cracks. The cracks originate at the pin-glass interface and continue outwards; other failure causes are weak oxide layer on the pin-glass interface and poor formation of glass meniscus around the pin.

Moisture and other gases may be present in the package cavity, either as impurities trapped during manufacture, or from outgassing of the materials used (monomers, curing agents, etc.), or even from chemical reactions (e.g. when the packaging material gets overheated; the released reaction products, often of ionic nature, then can facilitate corrosion and cause a delayed failure). Helium is often added into the inert atmosphere of the packagings, as a tracer gas to detect leaks during testing. Carbon dioxide may result from oxidation of organic materials with residual oxygen. Hydrogen may be released from some organic materials. Moisture can be outgassed by polymers. Amine-cured epoxies may outgas ammonia.

Formation of cracks and growth of intermetallics in die attachment materials may lead to formation of voids and delamination, impairing the heat transfer from the chip die to the substrate and heatsink, and cause a thermal-related failure.

As some semiconductors, notably silicon and gallium arsenide, are transparent in infrared, infrared microscopy can be used to check the integrity of die bonding and under-die structures.

Red phosphorus, used as a charring-promoter flame retardant, facilitates silver migration when present in device packaging materials. The phosphorus particles are normally coated with aluminium hydroxide. If this coating is incomplete, the phosphorus particles oxidize to phosphorus pentoxide, which is strongly hygroscopic and reacts with moisture to phosphoric acid. Phosphoric acid acts as a corrosive electrolyte, which together with electric fields facilitates dissolution and migration of silver, forming shorts between advanced packaging pins, lead frame leads, tie bars and chip mount structures, and/or chip pads. The silver bridge may be interrupted with thermal expansion of the package; disappearance of the failure when the chip is heated with a heat gun and its reappearance after several minutes later is an indication of this problem.

Delamination and thermal expansion may move the chip die relative to the packaging, deforming and possibly shorting and/or cracking the bonding wires.

Electrical contacts

- Failures of soldered joints
- Failures of electrical contacts – mechanical faults, corrosion
- Failures of cables – fraying, breaking of the conductors, corrosion, fire damage

Soldered joints, whether on boards, cables, or inside the electronic parts themselves, can fail in many ways; by electromigration, mechanical overstress, formation of brittle intermetallic layers, or material fatigue due to excessive thermal cycling. Such failures can be apparent only at high or low joint temperatures, hindering the debugging.

Thermal expansion mismatch between the part package and the printed circuit board material stresses the part-to-board bonds. Leaded parts are able to absorb the strain by bending. Leadless chip packages rely on the properties of the solder to absorb the stresses. Thermal cycling may lead to fatigue cracking of the solder joints, especially with less plastic solders. Various approaches are used to alleviate the temperature-induced strains.

Loose particles can form in the device cavity; a piece of bonding wire, a fragment of the chip die, flakes of plating, particles of die attachment material, fragments of the case, weld flash, and other materials may migrate inside the packaging cavity and cause shorts, often intermittent and sensitive to mechanical shocks.

Corrosion may cause buildup of oxides and other nonconductive products on the electrical contact surfaces. The contacts, when closed, then show unacceptably high resistance. Corrosion products may migrate and cause shorts.

Tin whiskers can form on the tin-coated metals, e.g. on the internal side of the packagings. Loose whiskers then can cause intermittent short circuits inside the packaging.

Printed circuit boards

Printed circuit boards are vulnerable to environmental influences. The traces are prone to corrosion, the vias can be insufficiently plated-through or insufficiently filled with solder. The traces may be improperly etched, either etched-through entirely, weakened, or insufficiently etched and leaving shorts between traces. The traces may crack under mechanical loads; the thin crack then often causes unreliable circuitboard operation, dependent on the physical warping on the board. Residues of solder flux may facilitate corrosion.

Residues of other materials on the surface of the boards can cause leaks. Polar nonionic compounds can attract water molecules from the atmosphere, forming a thin layer of conductive moist coating between the traces (some antistatic agents act the same way). Ionic compounds, especially chlorides, tend to facilitate corrosion. Alkali metal ions may migrate through plastic packaging and influence the operations of the semiconductors. Chlorinated hydrocarbon residues may hydrolyze, with release of corrosive chlorides; chlorinated solvent residues trapped in the packagings may cause problems years later. Polar molecules may dissipate high-frequency energy, causing dielectric losses.

Above the glass transition temperature of the boards, the resin matrix softens and becomes significantly susceptible to diffusion of contaminants. As an example, polyglycols from the flux can enter the board and increase its humidity intake, with corresponding deterioration of dielectric and corrosion properties.

Multilayer substrates, using ceramics instead of fiber-reinforced polymers, suffer from mostly the same problems.

Conductive anodic filaments (CAF) may grow within the boards, along the fibers of the composite material. The metal is introduced to the vulnerable surface typically from plating the vias, then migrates in presence of ions, moisture, and electrical potential. Drilling damage and poor glass-resin bonding promotes such failures. The formation of CAF usually begins by bonding failure between the glass fiber and the resin matrix, and a layer of adsorbed moisture then provides a channel through which ions and corrosion products migrate. In presence of chloride ions, the precipitated material is atacamite (copper chloride hydroxide); its semiconductive properties then may lead to increased leaks, deteriorated dielectric strength, and short circuits between the traces. Absorbed glycols from the flux residues aggravate the problem. The difference in thermal expansion of the fibers and the matrix also weakens the bond when the board is subjected to high temperature during soldering; the lead-free solders, which generally require higher soldering temperatures, are expected to increase the incidence of CAF. CAF incidence depends on absorbed humidity; below certain threshold it does not occur.

Delamination can occur, separating the board layers, cracking the vias and conductors and introducing pathways for corrosive contaminants and migration of conductive species.

Semiconductors

- Reliability (semiconductor)

Many failures result in generation of large amount of hot carriers in the chip structure, namely hot electrons. These are observable under an optical microscope, as they generate near-infrared photons detectable by e.g. a CCD camera. Latchups can be observed this way.

The location of the failure site, if visible, on the chip die may present clues to the nature of the overstress; whether the site is located at the place with highest current density, highest temperature, the highest electric field gradient, etc., size of the damage, secondary damage (fused leads, cracked die, reflowed die attachment...).

Liquid crystal coatings can be used for localization of faults. Cholesteric liquid crystals respond to temperature, are thermochromic; these are used for visualisation of locations of heat production on the chips. Nematic liquid crystals respond to voltage; these are used for visualising current leaks through oxide defects, and for visualising of charge states on the chip surface, allowing seeing the logical states on the individual structures and conductors.

During laser marking of plastic-encapsulated packages, the laser beam may reach and damage the chip die, if the glass spheres used as fillers in the epoxy resin packaging material line up in such way that they conduct the laser light to the chip.

Parameter-related

- GaAs MMICs:
 - Degradation of I_{DSS} : caused by gate sinking and hydrogen effects ("hydrogen poisoning"). Most common and easiest to detect. Affected by reduction of the active channel of the transistor (gate sinking) or depletion of the donor density in the active channel (hydrogen poisoning).
 - Degradation in gate leakage current: occurs at accelerated life tests or high operation temperatures; suspected to be caused by surface-state effects.
 - Degradation in pinch-off voltage": common failure mode for GaAs devices, operating at high temperature. Primarily results from semiconductor-metal interactions and degradation of gate metal structures. Can be hindered by suitable barrier metal inhibiting diffusion between gold and GaAs. Also can be caused by presence of hydrogen.
 - Increase in drain-to-source resistance: observed in devices operating at high temperature, caused by metal-semiconductor interactions. Caused by gate sinking and ohmic contact degradation.
 - Degradation in RF performance: caused by multiple factors. Surface-state density and material related effects play major roles.

In some cases, the normal presence of tolerances in the circuits can cause erratic behavior difficult to trace. For example a combination of a weak driver transistor with a higher series resistance together with the capacitance of the gate of the subsequent transistor, each within the normal "good" specifications, can significantly increase the propagation delay of the signal. Such faults can manifest only at very specific environmental conditions, high clock speeds in combination with low (but within specifications) power supply voltages, and/or specific circuit signal states. Significant variations can occur on a single die. Overstress induced damage, by e.g. creating ohmic shunts or lowering transistor output current, can potentially increase such delays, leading to erratic behavior of the circuit. As the propagation delays show significant dependence on power supply

voltage, normally allowed fluctuations of power supply voltage can trigger such erratic behavior.

Vias are a common source of unwanted serial resistance on chips. Defective vias show higher resistance than they should have and therefore increase propagation delays. As their resistivity drops with increasing temperature, degradation of maximum operating frequency of the chip with decreasing temperature is an indicator of such fault.

Mousebites are regions of partially missing metalization. The conductor is still present but its width is locally decreased. Such defects usually do not show during electrical testing, but present a major reliability risk. The increased current density in the damaged region may exacerbate electromigration problems. A very significant degree of voiding is needed to create a temperature-sensitive propagation delay.

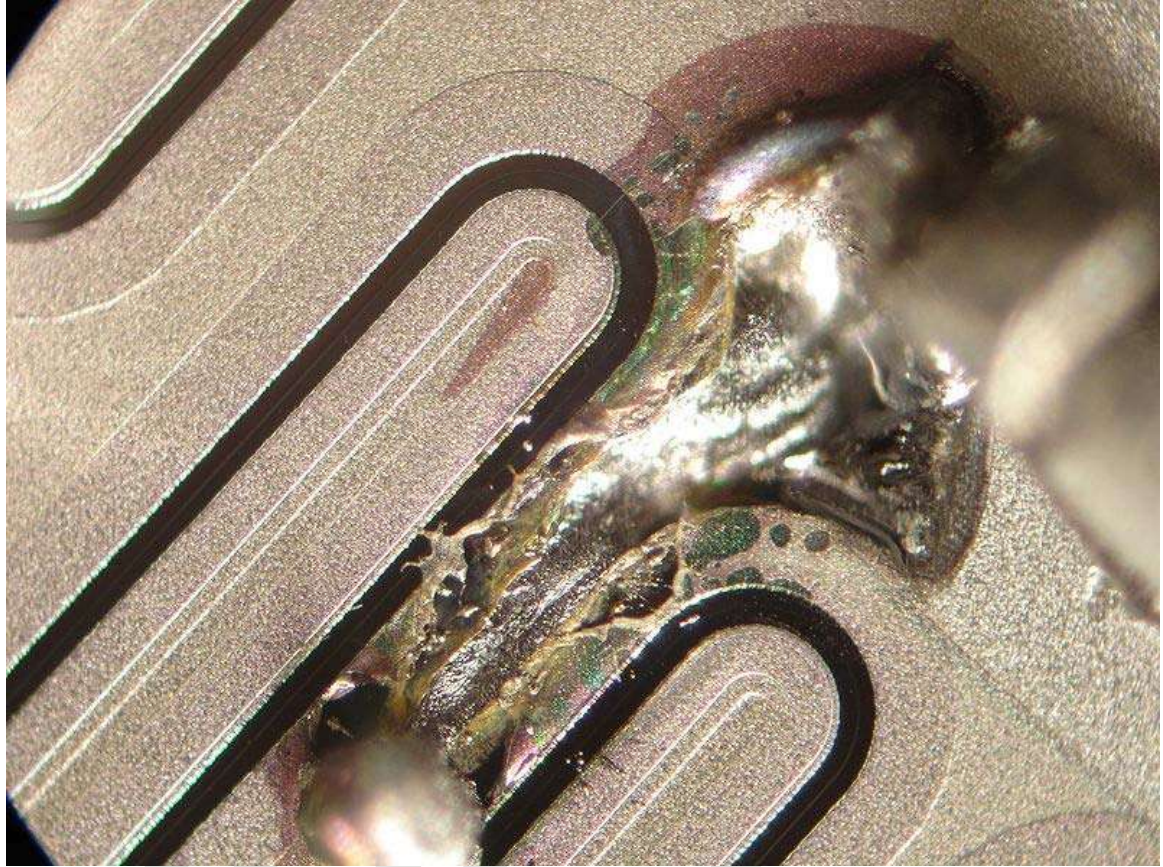
Metalization and wire bonding related

Metalization-related faults are more common and more serious cause of FET transistor degradation than bulk semiconductor material processes. Amorphous materials are promising; the lack of grain boundaries hinders interdiffusion and corrosion.

- Electromigration, caused by high current density can move atoms out of the active regions, leading to emergence of dislocations and point defects, acting as nonradiative recombination centers and producing heat instead of light.
 - Al-gate electromigration in power MESFETs can occur with large RF signals. The current densities in the "fingers" of the gate can be sufficient to cause electromigration, leading to voids and interruptions of the gate fingers and consequent loss of control of drain current. Gold metallization is less susceptible, the issue is therefore limited to aluminium.
 - Drain contacts in power FET transistors are depleted on their end while source contacts get material deposited on them.
 - In structures using aluminium metalization over a refractory metal barrier layer, the electromigration affects primarily the aluminium layer. The underlying refractory metal is highly resistant to electromigration, so the conductor does not fail entirely; its resistance just somewhat, often erratically, increases. The displaced aluminium can however cause shorts to neighbouring structures. Addition of 0.5-4% of copper to the aluminium metal significantly increases resistance to electromigration; copper accumulates on the alloy grain boundaries, increasing the energy needed to dislodge the metal atoms from them.
- Metal diffusion caused by high electrical currents or voltages at elevated temperatures can move metal atoms from the electrodes into the active regions. Some materials, notably indium tin oxide and silver, are subject to electromigration which causes leakage current and, in LEDs, nonradiative recombination along the chip edges. A barrier metal layer can be used to hinder the electromigration effects. Metal diffusion can cause changes in dimensions

(and therefore parameters) of the transistor gates and other semiconductor junctions. The migration of gate layer in MESFET transistors is known as **gate sinking**; it reduces the dimensions of the active channel and causes change in its effective level of doping, leading to deterioration of electrical parameters.

- Al/GaAs: Gallium arsenide in contact with aluminium, a common construction of MESFET gates, is susceptible to interdiffusion. Arsenic and especially gallium migrate into the aluminium layer, creating a zone with depleted stoichiometry, and buildup of a $\text{Al}_x\text{Ga}_{1-x}\text{As}$ region with gradient of concentration of aluminium. AlAs crystals are formed. This manifests e.g. as an increase of the Schottky barrier height. Forward current accelerates the process in comparison with thermal-only effect. Ti/GaAs shows similar effects. Aluminium-metalized GaAs devices two decades ago had lifetimes of thousands hours; contemporary TiPtAu metalization has lifetimes reaching millions of hours.
- Au/GaAs: Other interface layers are used with gold; examples are Au/TiW/GaAs, Au/TiPt/GaAs, and Au/Ta/GaAs. However TiW, while best for high temperatures, has different coefficient of thermal expansion than GaAs, which leads to growth of crystal defects under the metalization and reduced carrier mobility. The industry standard structures for gates on GaAs are based on Au/Pt/Ti or Au/Pd/Ti; Ti serves as a thin interlayer to facilitate adhesion, Pt or Pd is a barrier metal hindering diffusion of Au, and Au is the thick layer conductor. Grain boundaries in the barrier metal can however facilitate increased diffusion of gold into GaAs, leading to gate sinking.
- Ohmic contact degradation. The boundary between a metalization layer and the semiconductor can degrade. In case of GaAs, a layer of gold-germanium-nickel alloy (or gold-germanium alloy) is used to achieve low contact resistance. The ohmic contact is achieved by diffusion of germanium into GaAs, forming a highly n-doped region under the metal that facilitates the connection. A thick layer of gold is then deposited over the thin layer of AuGe. Gallium atoms can migrate through the thin layer and get scavenged by the gold above, creating a defect-rich Ga-depleted zone under the contact. Gold (and oxygen) migrate in the other direction, resulting in increased resistance of the ohmic contact and depletion of effective doping level. Formation of intermetallic compounds also plays a role.
- Short circuits; mechanical stresses, high currents, and corrosive environment can lead to formation of whiskers, causing short circuits. These effects can occur both within the packaging of individual devices and on the level of circuit boards.



Micro-photograph of a failed TO3 power transistor due to short circuit

- Formation of intermetallic compounds, e.g. the well-known gold-aluminium intermetallics (the dreaded white and purple plagues), leading to increased contact resistance and vastly decreased mechanical reliability. This is limited to older devices.
- Formation of silicon nodules. Aluminium interconnects may be doped with silicon to saturation during deposition, to prevent alloy spikes. During thermal cycling, the silicon atoms, originally homogeneously distributed, may migrate and clump together, forming nodules. The nodules act as voids in the metallization, increasing its local resistance and lowering the device lifetime.
- Corrosion of metallization. Aluminium is highly susceptible to humidity; especially negatively biased structures can be interrupted easily. Degradation of aluminium metallization was a common cause of failures of early plastic-encapsulated integrated circuits. Gold is susceptible to anodic corrosion in presence of humidity, forming voluminous conductive gold(III) hydroxide. Nickel can be extruded from metallization in presence of humidity and electric field, forming filaments along the electric field gradient that can short the electrodes. Arsenic can be leached from GaAs in presence of moisture. To cause **dry corrosion** of aluminium, only trace amounts of water and ionic contaminants are

required; all plastics are somewhat permeable. Phosphate ions can be leached from phosphosilicate glasses used for passivation of the chips; overcoat with oxide or oxynitride layer can be used in modern processes to create a moisture barrier on top of the phosphate glass layer, and/or a borophosphosilicate glass can be used to lower the phosphate content. Halogenides also rapidly corrode aluminium. Chlorides can be transported through the packaging from the outside. Bromides can be liberated from brominated flame retardants present in the packaging plastics when heated above 250 °C; overheating such package during storage, manufacture or use can increase the part's susceptibility to corrosion in the future.

- Sodium contamination, together with less common lithium and potassium. These ions are mobile in silicon dioxide layers even at normal temperature. (Their counterpart anions in contrast stay immobilized in the oxide structure.) The electric fields present during operation of the semiconductors cause migration of the mobile ions, leading to buildup of charged areas in the gate oxide. The NMOS gates are especially susceptible; a positive gate bias repels the cations towards the junction, depressing its threshold voltage. Even low concentrations of mobile ions can cause shifts by few millivolts, enough to cause trouble for analog circuits. The ion migration is a slow process, causing slow gradual drift of circuit parameters up to a possible failure. This effect was observed especially in early metal-gate CMOS logic circuits. Baking the affected chips at 200 °C for a few hours can temporarily reverse the effect by rediffusing the ions through the oxide. Doping the polysilicon of the base region with phosphorus is an effective way for immobilizing the ions. Alkali ions can migrate from the outside of the package; plastics can hinder their movement but can not slow them entirely. Nitride or phosphate glass layers are used as a chip die protection against externally originated contaminants.
- Very narrow interconnection cracks can maintain functionality by electron tunneling across the crack. Cold temperatures cause contraction of the metal, widening the crack and reducing the tunneling. Devices suffering this failure can run at higher frequencies at higher temperatures; this unusual behavior is symptomatic for an interconnection crack.
- Metalization step coverage, or microcracking, is an insidious unscreenable manufacturing failure. The metallization layer at some places, where their height differs, due to geometrical constraints and deposition technique forms locally weakened sites. The metal on such steps is thinner than required, or even develops a microcrack. The increased current density then leads to other effects leading to interruption of the layer and premature failure of the device.

Semiconductor-related

- Nucleation and growth of dislocations are known mechanisms for degradation of the semiconductor junctions. This requires a presence of an existing defect in the

crystal and is accelerated by heat, high current density, and emitted light. In case of LEDs, Gallium arsenide and aluminium gallium arsenide are more susceptible to this mechanism than gallium arsenide phosphide and indium phosphide. Because of different properties of the active regions, gallium nitride and indium gallium nitride are virtually insensitive to this kind of defect.

- Accumulation of charge carriers trapped in the gate oxide of MOSFETs. This introduces permanent gate biasing, influencing the transistor's threshold voltage. This effect may be caused by hot carriers injection, ionizing radiation (one of the total dose effects), or even nominal use; in case of EEPROM cells and related structures this is the major wear mechanism limiting the number of erase-write cycles.
- Ionizing radiation and neutron radiation have multiple effects, both transient and permanent. They can cause defects in the semiconductor, creating recombination centers and shortening the lifetime of minority carriers, degrading the performance of bipolar junction transistors; it also causes accumulation of charge carriers discussed above. More details are described in the problematics of radiation hardening.
- Migration of charge carriers from floating gates, limiting the lifetime of stored data in EEPROM and flash EPROM structures.
- Improper passivation; corrosion effects are a significant source of delayed failures. Semiconductor materials, metallic interconnects, and passivation glasses are all susceptible to corrosion. The surface of semiconductor, subjected to moisture, develops a layer of oxide; the hydrogen liberated from water then reacts with deeper layers of the material, yielding volatile hydrides.
- Stress Induced Leakage Current is an increase in the gate leakage current of a MOSFET, due to defects created in the gate oxide during electrical stressing.
- Hot carrier injection, occurring in MOSFET transistors. At high field gradients, the charge carriers (electrons in NMOS, holes in PMOS) are accelerated to high speeds. When the transistor operates at saturation with high drain-source voltage, the pinched-off region under the gate (created by the electric field) produces hot carriers near the drain end. Due to lower mobility of holes, the NMOS devices produce hot carriers at lower field gradients than PMOS; e.g. a 3 micrometer NMOS will generate hot electrons at 10 V, while a PMOS of the same size will not generate hot holes below 20 V. Some of the hot carriers, after collision with atoms in the drain region, are deflected into the gate oxide. Most return back, but some become trapped in the oxide imperfections. These then accumulate and cause a persistent gate bias, increasingly shifting the threshold voltage as more hot carriers accumulate. Like with alkali ions, the damage can be partially or fully annealed by heating the chip without power at 200-250 °C for a couple of hours. Avalanche breakdown is also a major source of hot carriers; in case of diffused

junction it usually occurs near the surface, where the dopant concentration is highest.

- Parasitic channels can form anywhere where a source-drain structure exists. The role of the gate can be played by a metallization trace above, or even by static charge built up or migrated into the overlaying insulator, protective overcoat, or passivation layer. As the migrating charge carriers are usually electrons, parasitic PMOS transistors tend to be formed. The bipolar chips are more sensitive, likely because of more stringent requirements regarding absence of mobile ionic species for CMOS and BiCMOS technologies.
- Burnt fuses, used for programming the integrated circuits, can under certain conditions reform. Polysilicon fuses may crack instead of vaporizing, which prematurely interrupts the current and leaves enough material to allow reforming the fuse later. Programming both polysilicon and metalization fuses after packaging the circuit prevents the metal from dispersing; the metal stays in the vicinity of the burnt fuse and can migrate back.
- Random access memory chips suffer from several types of failures:
 - Address fault, where the memory address decoder is faulty
 - Stuck-at fault, where a data line somewhere on the chip is shorted to H or L
 - Bridging fault, where two inputs or outputs are shorted together
 - Transition fault
 - n-cell coupling fault
 - Delay fault, where a propagation delay somewhere in the circuit is unacceptably high, causing faulty operation at high speeds
 - Retention fault, where a DRAM cell capacitor does not reliably hold charge for sufficient time

The failures are also classed as bit failures (single, double, triple, quadruple, multiple), row failures (where the entire row of bits fails), column failure (same for the memory array column), cross (where both a row and a column failure occurs), continuous block, peripheral logic, and systemic defect.

Stress-related

Most stress-related failures are electrothermal in nature. The locally increased temperature can lead to immediate failure by melting or vaporizing metalization layers, melting the semiconductor, or creating other structural changes. The diffusion and electromigration effects tend to be accelerated by high temperature, which shortens the lifetime of the device. Damages to junctions that do not lead to immediate failure manifest as altered current-voltage characteristics of the junctions.

Electrical overstress failures can be classified as thermally induced failures, electromigration related failures, and electric field related failures.

- Thermal runaway: Nonhomogenities in the substrate, causing localized loss of thermal conductivity, can cause thermal runaway where heat causes damage which causes more heat etc. Most common ones are voids caused by incomplete soldering, or by electromigration effects and Kirkendall voiding.
- Current crowding, non-homogenous distribution of the current density over the junction, formation of current filaments. This may lead to creation of localized hot spots, which poses risk of thermal runaway.
- Reverse bias: Although e.g. the LED is based on a diode junction and is nominally a rectifier, the reverse-breakdown mode for some types can occur at very low voltages and essentially any excess reverse bias causes immediate degradation, and may lead to vastly accelerated failure. 5 V is a typical, "maximum reverse bias voltage" figure for ordinary LEDs, some special types may have lower limits.
- Overcurrent can cause failures of the bonding wires. In some cases the semiconductor junctions can withstand high enough current to melt the bonding wires.
- Zener diodes in reverse bias, when severely overloaded, fail as a short circuit. A sufficiently high voltage causes an avalanche breakdown of the Zener junction; the voltage across the junction together with a significant current being forced through causes extreme localized heating; the junction and metallization melt, and an alloy of silicon and aluminium shorts the diode's terminals. This is sometimes intentionally used in semiconductors as a type of programming fuses.
- Latchups can occur when the device is subjected to an overvoltage or undervoltage pulse. The opened SCR parasitic structure then can cause an overcurrent-based permanent failure. In integrated circuits, latchups are divided by their cause to internal (transmission line reflections, ground bounces, power supply overshoots) and external (signals injected to the chip via its I/O pins from the outside). External latchups can be triggered by an electrostatic discharge. Effects of ionizing radiation and cosmic rays are also included in external effects. Susceptibility to latchup is tested according to the JEDEC78 test standard. Latchup can be triggered by charge carriers injected into the chip substrate by e.g. a current flowing through an ESD protection diode, or through another latchup.
- FET transistors are sensitive to dV/dt failures; excessively fast voltage transients can cause the transistor to open.
- Bipolar transistors are more thermally sensitive than FETs; the thermal runaway phenomena limit their operation margins at higher ambient temperatures. Bipolar transistors are also more sensitive to degradation of die cooling mechanisms (die bond defects, part-heatsink attachment degradation).

Electrostatic discharge

Electrostatic discharge (ESD), a subclass of electrical overstress (EOS), may cause immediate failure of the semiconductor device, a permanent shift of its parameters, or a latent damage causing increased rate of degradation. ESD failure has at least one of three components: localized heat generation, high current density, and high electric field gradient. Currents of several amperes can be present for several hundreds of nanoseconds; the energy deposited to the device structure then causes the damage.

ESD discharge in a real circuit containing capacitances and/or inductances causes a "ringing" waveform, a damped wave with rapidly alternating polarity. Affected junctions are rapidly stressed in alternating forward and reverse polarization.

ESD damage has four basic mechanisms:

- **Oxide rupture/breakdown**, occurring at field strengths above 6–10 MV/cm. Designing the circuit so the oxide layers are protected by junctions with lower avalanche breakdown thresholds than the oxide breakdown voltage can protect against this failure mode.
- Junction damage manifests as increased reverse-bias leakage, up to total shorting.
 - **Junction filamentation**, also known as second or thermal breakdown. Localized overheating causes melting of the silicon; molten silicon has 30 or more times lower resistance; the current through the junction therefore flows in few narrow filaments of high current, resulting in a thermal runaway. In MOSFET devices, the filamentation region is usually near the surface, in the gate-drain overlap region, where the insulator layer serves as a thermal insulation. Devices where the hot spots occur deeper in the structure are less susceptible and are often used as ESD protection structures. Dopants diffuse easily in the molten region; shorter events lead to localized thinning of the base region, resulting in increased leakage, longer events may provide enough time for formation of an ohmic channel across the base, resulting in emitter-collector (or source-drain) short.
 - **Junction spiking**, where metalization is involved; the aluminium metalized devices are more susceptible as the eutectic melting point of aluminium-silicon alloy is 577 °C instead of 1415 °C for silicon. Refractory barrier metals inhibit this effect. When the molten silicon region reaches the metalization, a violent exchange of materials occurs.
- **Metalization and polysilicon burn-out**, where the damage is focused to the conductive and resistive elements – metal and polysilicon interconnects, thin film and diffused resistors. Thin film resistors are the most susceptible. The thermal damage causes localized degradation to destruction of the conductive structure. The critical current densities can be lower in structures surrounded by thermal insulators where the Joule heat can not easily dissipate.
- **Charge injection**, where the hot carriers generated by an avalanche breakdown are injected into the oxide layer. This mechanism is the same as for hot carrier

injection. The result is an increased leakage current from the FETs prebiased by the injected charges in the gate oxide.

For **catastrophic failures**, there are three basic ESD-related mechanisms:

- **Junction burnout**, formation of a conductive path through the junction, shorting it
- **Metalization burnout**, melting or vaporizing part of the metal interconnect, interrupting it
- **Oxide punch-through**, formation of a conductive path through the insulating layer between two conductors or semiconductors; the gate oxides are thinnest and therefore most sensitive. The damaged transistor shows a low-ohmic junction between gate and drain terminals.

ESD can cause a **parametric performance failure**; the device still operates, but its parameters are shifted. The failure may manifest in stress testing. In some cases, the degree of damage can lower over time (so called *cold healing*).

ESD can also cause **latent failures**, which manifest themselves in a delayed fashion and are difficult to impossible to test for. They have these main reasons:

- Damage to insulators: weakening of the insulator structures, leading to accelerated breakdown and/or increased leakage
- Damage to junctions: lowering the lifetime of minority carriers with consequent bipolar transistor gain loss; increasing resistance in forward biased state; increasing leakage in reverse biased state
- Damage to metalization: weakening of the conductor, leading to increased resistance or increased rate of electromigration

Catastrophic failures require the highest discharge voltages; they are easiest to test for, and rarest to occur. Parametric performance failures occur at intermediate discharge voltages and occur more often. Latent failures occur at low voltages and are the most common; for each parametric performance failure there are 4–10 latent ones.

Modern high-integration circuits are more ESD sensitive. The features are smaller, their capacitance is lower and for the same amount of charge the deposited voltage is higher. The silicidation of the conductive layers makes them more conductive, reducing the ballast resistance that can have a degree of protective role.

The gate oxide of some MOSFET transistors can be damaged by as little as 50 volts potential. The gate is isolated from the junction; potential accumulated on it causes extreme stress on the thin dielectric layer. Stressed oxide can shatter and fail immediately (gate rupture – occurs in nanoseconds, does not require a sustained electric current, and is irreversible; usually the gate and backgate of the affected transistor end up connected together). The gate oxide does not have to fail immediately; the gate leakage can increase

by stress induced leakage current, the oxide damage can lead to a delayed failure after hundreds or thousands of operation hours. On-chip capacitors using oxide or nitride dielectric are vulnerable to the same kind of damage. Small structures are more vulnerable than large ones by the virtue of their lower capacitance; the same amount of charge carriers will charge the capacitor the structure forms to a higher voltage. All thin layers of dielectric, e.g. the protective oxide layers over emitter regions of transistors or the insulation between two interconnections, are vulnerable; not just the MOS gates. Chips made by processes employing thicker oxide layers are less vulnerable.

The degree of gate oxide damage depends on the size of the gate. Older transistors had larger gate regions with higher capacitance; a discharge of the accumulated potential frequently caused a gate rupture, causing a **hard breakdown**. Newer ultrathin oxide layers are more commonly damaged by a **soft breakdown**; while the damage is still irreversible, its most significant effect is an increase of the noise voltage of the gate, by up to 4 orders of magnitude. Gate oxide breakdown does not always have to lead to failure; gate-to-channel breakdowns usually do not lead to a hard failure, unlike the gate-to-source or gate-to-drain breakdowns.

Gate oxides can also be stressed during manufacture, by so called **antenna effect**; charges introduced during e.g. ion implantation or dry etching accumulate in conductive structures, causing a voltage buildup across the vulnerable dielectrics.

Breakdown of the gate oxide can form a number of different structures with varying voltage-current characteristics. The simplest case is a connection between two regions with the same doping; a resistive path is then formed. In case of opposite doping regions, the resulting structure is a diode. A metal-to-semiconductor connection may create an ohmic contact or a Schottky diode.

The formation of an ohmic path across an insulator or a junction may not always lead to a hard failure. If the resistance of the path is higher than a critical value, only a degradation of parameters occurs.

Interruption of a conductor may also not always lead to a hard failure. Capacitive coupling can still occur between the structures, maintaining partially degraded functionality, reducing signal strength or affecting gate output voltage range. Faults in drain or source connections of CMOS transistors may lead to formation of quasi-memory cells, making the defect manifestation dependent on previous states of the logic circuit. Loss of an output driver can lead to high-impedance output at some circuit states; the load capacitance then can maintain previous state for some time as a dynamic memory cell.

Semiconductor junctions are more robust than thin dielectrics. A sufficiently high voltage leads to an avalanche breakdown. The electric current is usually concentrated to a small area, due to current crowding; a semiconductor heated above certain limit starts losing resistance with increasing temperature and the region concentrates more current in itself, leading to more heat production and a thermal runaway. The extreme current density can

migrate the metalization and short the junction, the heat can melt and recrystallize or even shatter the junction. Such damage usually manifests as a shorted junction.

Current-induced failures are more common in bipolar junction devices, where Schottky or pn junctions are predominant. The high power of the discharge (often above 5 kilowatts for less than a microsecond) is capable of melting and vaporizing silicon and metal. Thin-film resistors can have their value altered by a discharge path forming a shunt across part of their length (value decrease) or get part of the layer vaporized (value increase); this can be problematic in precision analog applications where the values are critical.

A junction which suffered an avalanche breakdown typically has increased leakage. Avalanching the base-emitter junction of a NPN transistor permanently lowers its beta.

Device terminals connected only to MOS gates or capacitors are the most vulnerable to the effects of electrostatic discharge. In case both insulated gates and junctions are connected, avalanching usually takes precedence before gate oxide damage. Substrates of chips and large diffusion areas in the semiconductors, e.g. collectors of high-power transistors, are less vulnerable due to their large size and ability to dissipate more energy. Small diffusion areas, e.g. base and emitter regions of small NPN transistors, are more vulnerable.

Newer CMOS output buffers using a lightly doped drain and silicide source/drain are more ESD sensitive. The N-channel driver usually suffers damage, usually in the oxide layer or n+/p well junction. The damage is caused by current crowding during the snapback of the parasitic NPN transistor. In PMOS-NMOS totem-pole structures, the NMOS transistor is virtually always the one damaged.

Typical ESD-related failure distribution for bipolar/Schottky devices is 90% junction burnout and 10% metalization burnout. For MOSFET devices it is 63% metalization burnout and 27% oxide punchthrough. MIL-HDBK-263 offers more complete discussion.

The structure of the junction influences its ESD sensitivity. Corners and defects can lead to current crowding, reducing the damage threshold. Forward-biased junctions are less sensitive than reverse-biased; in the first case the Joule heat is dissipated through the thicker layer of the material while in the latter it is concentrated in the narrow depletion region.

LEDs and lasers grown on sapphire substrate are more susceptible to ESD damage.

In NMOS transistors, the ESD pulse can lead to formation of a metal filament between source and drain, by a phenomenon called *electrothermomigration*. The junction heat can also melt the polysilicon gate, forming polysilicon filaments between gate-source and gate-drain.

The damage can be observed on the I/V curve; undamaged devices show sharp knees, while damaged ones are significantly softened.

Optoelectronics

- List of LED failure modes
- Catastrophic optical damage of semiconductor lasers at high power, when the part is overstressed
- Hydrogen darkening of optical fibers in presence of hydrogen
- Phosphor degradation, in white LEDs, cathode ray tubes, plasma displays, fluorescent lights, etc.
- Corrosion of optical materials; silicate glasses are attacked by moisture. Lenses and optical fibers are prone to deterioration by condensed moisture. Phosphate glasses are more susceptible to corrosion than silica-based glasses, different formulations of glasses can have vastly different.

MEMS

Microelectromechanical systems suffer from specific types of failures.

- Stiction causes the moving parts stick to other surfaces on contact. An external impulse can sometimes release the adhesion and restore functionality. Non-stick coatings, reduction of surface contact area, and increased awareness virtually eliminated the problem in contemporary systems.
- Particle contamination can cause failures by particles lodged between the elements and blocking their movements. Conductive particles may short out circuits, namely electrostatic actuators.
- Wear damages the surfaces; debris, removed from the surfaces during their mutual movements, can be a source of particle contamination.
- Fractures may cause loss of mechanical parts
- Material fatigue may induce cracks in moving structures
- Electrostatic discharge
- Oxide charging, causing electrostatic attraction between parts, may lead to electrostatically mediated stiction
- Dielectric breakdown, causing a short circuit and irreversibly damaging the MEMS structure

Vacuum tubes

- Vacuum tubes and fluorescent lights are susceptible to degradation of hot cathodes, leading to gradual loss of emission of electrons. The vacuum inside the tubes can be compromised by outgassing of materials inside, diffusion of gases through the envelope, or envelope failure. Burnout of the hot cathode filaments leads to a sudden failure. Cold cathode devices tend to be more reliable.
- Multipactor effect

- Phosphor degradation

Passive elements

Resistors

Resistors can fail open (going to infinite resistance), fail short (going to close to zero resistance), or their value can increase or decrease under environmental conditions (e.g. corrosion, material aging) or because of exceeding of performance limits (e.g. overheating).

Thin film resistors are formed from a thin film of a suitable material, e.g. chromium or tantalum nitride.

- Resistor#Failure modes
- The value of certain thin film metal resistors can change when they overheat, due to annealing of their crystalline structure. Extreme overheating may lead to open circuit failure. In integrated circuits, diffused resistors are preferable for applications where high transient currents are to be encountered, e.g. ESD protection, as they are in close contact with the semiconductor substrate which serves as additional heatsinking.
- Mechanical defects from manufacturing can cause intermittent problems or failures. Improperly crimped caps on carbon or metal resistors can become loose and lose contact permanently or intermittently, or the resistor-to-cap interface resistance increase can shift the value of the resistor. Leads to the through-hole resistors are welded to the caps by spark or by butt weld; a faulty weld can become loose. Rough handling can cause such defects or lead to manifestation of latent defects. Thin cracks in the ceramic substrate may cause an open fault, often only annoyingly intermittent.
- Deformation of wire-wound resistors can lead to shorting of adjacent loops of resistive wire, causing partial loss of resistance.
- Operation in oxidizing atmosphere causes oxidation to the outer layer of the resistive wire, reducing its active diameter and increasing its resistance. Operation in reducing atmosphere, e.g. in presence of hydrogen, has the opposite effect.
- Ceramic and carbon resistor cores are prone to cracking under mechanical loads and shocks.
- Under overload, the power dissipated on the resistor can cause heating above the maximum rating, and melting or oxidizing of the resistive element. The protective lacquer or polymer is charred and pyrolyzed, with release of the characteristic smell and possible formation of a conductive path. The resistive element may be interrupted or weakened.
- At high voltages, arcing can occur between parts of the resistor surface, with possible formation of conductive paths or vaporization of the resistive material.
- Damage to insulation layer of the end caps can lead to short between the cap and an underlying circuit board trace.

- Carbon composition resistors, rods of carbon/ceramic composite with embedded leads, may absorb water in humid environments during operation or storage, with resistance changes of as much as 15% up or down.
- Wire-wound resistors are prone to corrosion of the resistive wire. Cap crimping and welds between the wire, caps, and leads are also weak points of this type.
- Metal and carbon film resistors, composed of a thin film of resistive material (typically cut into a spiral) on a (typically) ceramic core, are prone to corrosion and electrolysis of the resistive layer, when the protective layer is penetrated by moisture and contaminants. The electrolytic damage can erode the layer up to an open circuit failure; externally applied voltage is needed for this deterioration mode. Conductive contaminants may form bridges between the loops of the spiral and lead to resistance drop.
- Surface-mount resistors can suffer delamination of their structure where dissimilar materials join, e.g. between the ceramic substrate and the resistive layer, or between the resistive layer and the contact terminations.
- Nichrome thin-film resistors in integrated circuits can be attacked by phosphorus from the passivation glass, which corrodes them and increases their resistance.
- Laser-trimmed resistors may develop instabilities due to the heat damage and microcracks in the resistive layer adjoined to the site of the kerf.
- SMD resistors with silver metallization of termination contacts may suffer open-circuit failure in sulfur-rich environment (e.g. close to sulfur-vulcanized rubber), due to buildup of silver sulfide. Presence of either sulfur dioxide (emitted e.g. by heated rubber) and hydrogen sulfide can be the cause. Conformal coatings do not prevent the failure. The failure is however fairly rare.
- Materials may migrate through the resistor structure and alter resistance or cause shorts. Silver can migrate from the electrodes of thick film chip resistors. Silver atoms were detected as far as 100 micrometers from the electrodes. In addition to lowering the resistance, the silver atoms create nonhomogeneities in current distribution and lead to current crowding, instability of characteristics over time, and increase of noise.
- Copper dendrites may grow from copper oxide present in some materials (e.g. from the layer facilitating adhesion of metallization to a ceramic substrate), and bridge the trimming kerf.
- The resistive layer of thick film chip resistors may degrade when subjected to overvoltage or overcurrent.
- The electrodes may crack or corrode.

Potentiometers and trimmers

Potentiometers and trimmers are three-terminal electromechanical parts, containing a resistive path with a wiper contact with adjustable position. All the failure modes listed for resistors apply to these parts too. Additionally, mechanical wear on the wiper and the resistive layer, together with corrosion, surface contamination, and mechanical deformations, may lead to intermittent path-wiper resistance changes. These are especially annoying with audio amplifiers. Many types are not perfectly sealed, and contaminants and moisture may readily enter the parts. An especially common

contaminant is the solder flux. Mechanical deformations, with impaired wiper-path contact, can occur by warpage of the part housing during soldering or mechanical stress during mounting. Excess stress on leads can cause substrate cracking and open failure when the crack penetrates the resistive path.

Capacitors

Capacitors are characterized by their capacitance, parasitic resistance in series (equivalent series resistance) and parallel (leakage) – both often frequency-dependent and voltage-dependent, breakdown voltage, and dissipation factor. Structurally, capacitors consist of electrodes separated by the dielectric (sometimes soaked with liquid electrolyte), the leads, and the housing. Deterioration of any of these structures may cause shift of parameters or open or short failure. Short failures and increase of leakage are the most common failure modes of capacitors, followed by open failures.

- Dielectric breakdown occurs due to overvoltage or aging of the dielectric material leading to breakdown voltage falling below the operating voltage; some types of capacitors are "self-healing", the internal arcing vaporizes parts of the electrodes around the failed spot of the dielectric and breaks the contact, others form a conductive pathway through the dielectric, leading to a short or to partial loss of dielectric resistance.
- Electrode materials may migrate across the dielectric, forming conductive paths.
- Leads can be separated from the capacitor by rough handling during storage, assembly or operation; this leads to an open failure. The failure can occur inside the packaging, invisible from the outside but measurable.
- Dissipation factor may increase due to contamination of the capacitor materials, whether from manufacture or by penetration along the leads or through the packaging. Flux and solvent residues are a common source of problems.
- Leakage can result from contaminants forming conductive paths across the capacitor plates or from altering the dielectric parameters. Moisture or solvents can be absorbed by the dielectric if the coating cracks. Cracks in dielectric can also form leakage paths.
- Excessive charging or discharging current may cause partial fusing of the electrodes, leading to open failure or to capacitance shift. Excessive charging or discharging currents may also fuse the leads to the electrodes.
- Partial debonding of leads can cause loss of capacitance in multilayer capacitors.
- Delamination of multilayer capacitors may cause partial capacitance loss and deterioration of other electrical parameters.
- Voids, cracks, pores and other defects may impair the parameters of the dielectric, causing leakages, loss of capacitance, and loss of maximum breakdown voltage.
- Cracks in the dielectric may present pathways for metal migration and arcing.
- Chemical attack on the dielectric may degrade the insulation resistance.
- Thermal shocks, especially during soldering, may cause mechanical defects in the parts, and alter (temporarily or permanently) the properties of the dielectrics.

- Capacitors in low-voltage high-impedance applications generally fail due to lowered dielectric resistance. Capacitors in high voltage and/or low impedance applications tend to fail due to dielectric breakdown.
- Pinholes and voids cause increased leakage in film capacitors. When sufficient current is available, the defect site can be burned away and the capacitor "self-heals". When insufficient current is available, the current can cause migration of electrode material through the defect, gradually decreasing the leakage path resistance.
- Film capacitors are sensitive to excessive ripple voltage; the thin metallization on the dielectric may not sustain high currents, and partially fuse with a loss of capacitance.
- Polymer film dielectrics tend to become brittle with age, especially at increased temperature. Fractures may develop. Moisture may cause degradation of the polymer.
- Corrosion and cracking of the internal leads can cause open failure.
- Overvoltage in a ceramic capacitor may cause an avalanche breakdown in a weak spot of the dielectric, resulting in a thermal runaway. The damaged area may show localized melting and cracking. The energy deposited in the failure point can be high enough to cause the capacitor to explode, create a miniature fireball, and char the encapsulation and possibly the underlying circuitboard.
- Mica capacitors may show intermittent open failures from separation of the electrode plates from the termination at certain temperatures. Reconstituted mica capacitors may fail short due to weaknesses in the dielectric. As mica capacitors are typically used in high voltage applications, the results may be severe.

Electrolytic capacitors

- Aluminium electrolytic capacitors suffer from gradual increase of leakage and equivalent series resistance, and loss of capacitance due to drying out of the electrolyte. Power dissipation due to high ripple currents and high internal resistance cause rise of the internal temperature of the capacitor, which accelerates the deterioration rate, especially when the internal temperature exceeds maximum design temperature. Such capacitors usually fail short.
- Electrolyte contamination, especially with moisture, can lead to corrosion of the electrodes. The deterioration of their area can lead to capacitance loss, loose particles of conductive material can cause shorts.
- Electrolyte may under certain conditions evolve gas. This leads to increased pressure inside the capacitor housing, in extreme but frequent cases leading to an explosion.
 - Capacitor plague is an extreme kind of premature electrolytic capacitor degradation and failure.
- Tantalum capacitors are susceptible to electrical overstress. Exceeding of the maximum surge voltage may permanently degrade the dielectric and even cause open or short failure.

- The most common failure mode of solid tantalum capacitors is an electrical short. The most common manufactured-related cause of this failure is a defect in the dielectric oxide.
- The failure locations may be visible as discolored dielectric, or as locally melted anode.
- Defects in the tantalum chip anode material, e.g. impurities, are transferred to the dielectric oxide layer formed during the capacitor manufacture. These locations then present weak spots that may fail prematurely. Thermal shock e.g. during soldering may generate new defects.
- The thickness of the oxide layer is limited by the size of the powder from which the tantalum slug is made.
- On voltage transients, tantalum capacitors may show a momentary short or increase of leakage current, called "scintillation". If the current through the failure site is small, the site self-heals; if the current is sufficient to heat the site above 400 °C, thermal runaway occurs and the part shorts.

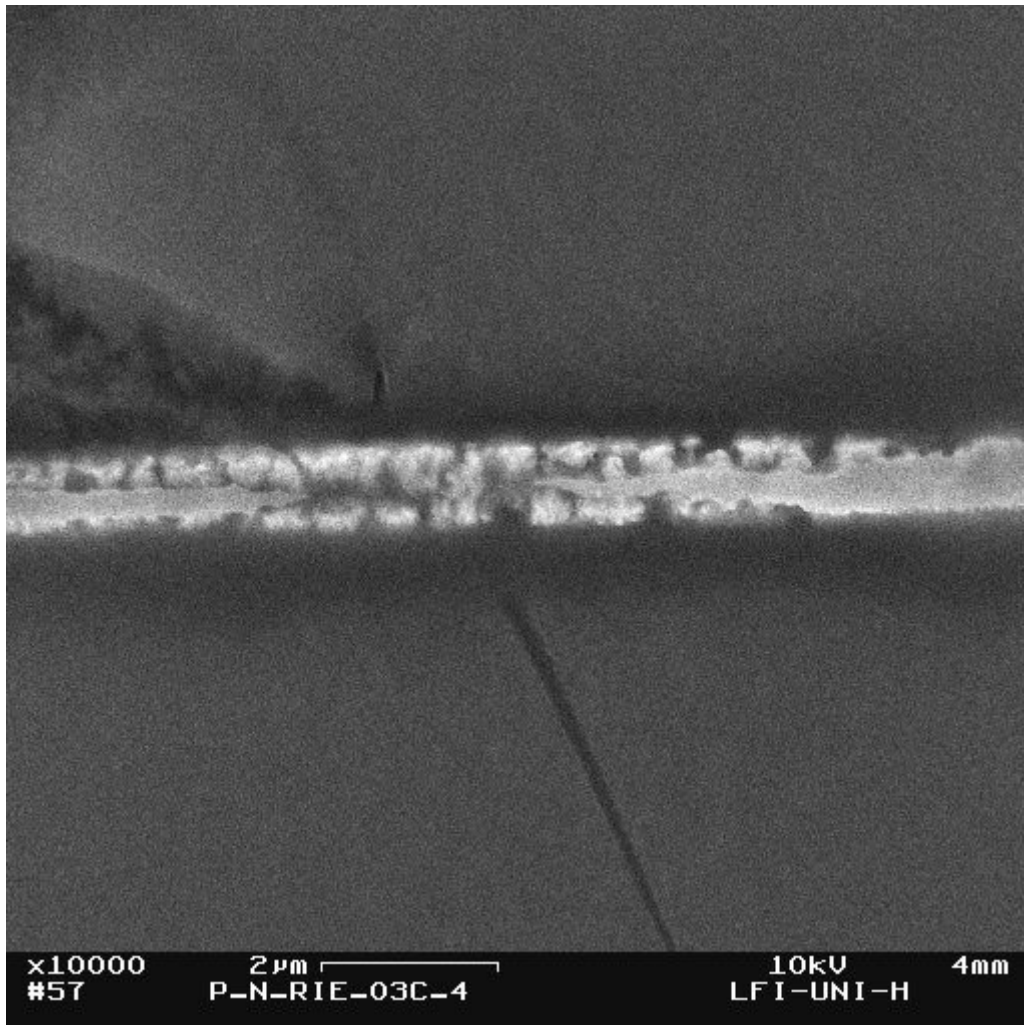
Crystals

Crystals for crystal oscillators are thin slabs of a piezoelectric material, typically quartz, with deposited electrodes on the slab surfaces, mounted in a hermetically sealed housing, providing the circuit with oscillations at a stable frequency.

- Cracks of crystal slab can occur at extremely high drive levels (especially with low-frequency crystals), or due to mechanical shocks strong enough to make the crystal impact the inside of the housing.
- Failure of lead connection can lead to an open circuit failure.
- Outgassing of materials inside the housing, most often improperly cured epoxy, can lead to frequency shifts of the crystal.
- Moisture present in the housing can condense on the crystal at low temperatures, causing significant frequency shifts.
- Ionizing radiation and neutron radiation can cause frequency shifts. Swept quartz, from which the mobile alkali metal ions were removed, is used for rad-hard crystals.
- Dehermetization of the housing, by rough handling or corrosion, may lead to penetration of contaminants and moisture into the housing, with consequent changes in resistance and frequency, and possible lead-to-lead and lead-to-packaging leakage paths.

Electromigration

Electromigration is the transport of material caused by the gradual movement of the ions in a conductor due to the momentum transfer between conducting electrons and diffusing metal atoms. The effect is important in applications where high direct current densities are used, such as in microelectronics and related structures. As the structure size in electronics such as integrated circuits (ICs) decreases, the practical significance of this effect increases.



SEM image of a failure caused by electromigration in a copper interconnect. The passivation has been removed by RIE and HF

History

The phenomenon of electromigration has been known for over 100 years, having been discovered by the French scientist Gerardin. The topic first became of practical interest in 1966 when the first integrated circuits became commercially available. Research in this field was pioneered by a number of investigators throughout the fledgling semiconductor industry. One of the most important engineering studies was performed by Jim Black of Motorola, after whom Black's equation is named. At the time, the metal interconnects in ICs were still about 10 micrometres wide. Currently interconnects are only hundreds to tens of nanometers in width, making research in electromigration increasingly important.

Practical implications of electromigration

Electromigration decreases the reliability of ICs. In the worst case it leads to the eventual loss of one or more connections and intermittent failure of the entire circuit. Since the reliability of interconnects is not only of great interest in the field of space travel and for military purposes but also with civilian applications like for example the anti-lock braking system of cars, high technological and economic values are attached to this effect.

Due to the relatively high life span of interconnects and the short product lifecycle of most consumer ICs, it is not practical to characterize a product's electromigration under real operating conditions. A mathematical equation, the Black's equation, is commonly used to predict the life span of interconnects in integrated circuits tested under "stress", that is external heating and increased current density, and the model's results can be extrapolated to the device's expected life span under real conditions. Such testing is known as high temperature operating life (HTOL) testing.

Although electromigration damage ultimately results in failure of the affected IC, the first symptoms are intermittent glitches, and are quite challenging to diagnose. As some interconnects fail before others, the circuit exhibits seemingly random errors, which may be indistinguishable from other failure mechanisms (such as electrostatic discharge damage). In a laboratory setting, electromigration failure is readily imaged with an electron microscope, as interconnect erosion leaves telltale visual markers on the metal layers of the IC.

With increasing miniaturization the probability of failure due to electromigration increases in VLSI and ULSI circuits because both the power density and the current density increase. In advanced semiconductor manufacturing processes, copper has replaced aluminium as the interconnect material of choice. Despite its greater fragility in the fabrication process, copper is preferred for its superior conductivity. It is also intrinsically less susceptible to electromigration. However, electromigration (EM) continues to be an ever present challenge to device fabrication, and therefore the EM research for copper interconnects is ongoing (though a relatively new field).

A reduction of the structure (scaling) by a factor k increases the power density proportional to k and the current density increases by k^2 whereby EM is clearly strengthened.

In modern consumer electronic devices, ICs rarely fail due to electromigration effects. This is because proper semiconductor design practices incorporate the effects of electromigration into the IC's layout. Nearly all IC design houses use automated EDA tools to check and correct electromigration problems at the transistor layout-level. When operated within the manufacturer's specified temperature and voltage range, a properly designed IC device is more likely to fail from other (environmental) causes, such as cumulative damage from gamma-ray bombardment.

Nevertheless, there have been documented cases of product failures due to electromigration. In the late 1980s, one line of Western Digital's desktop drives suffered

widespread, predictable failure 12–18 months after field usage. Using forensic analysis of the returned bad units, engineers identified improper design-rules in a third-party supplier's IC controller. By replacing the bad component with that of a different supplier, WD was able to correct the flaw, but not before significant damage to the company's reputation.

Electromigration can be a cause of degradation in some power semiconductor devices such as low voltage power MOSFETs, in which the lateral current through the source contact metallisation (often aluminium) can reach the critical current densities during overload conditions. The degradation of the aluminium layer causes an increase in on-state resistance, and can eventually lead to complete failure.

Fundamentals

The material properties of the metal interconnects have a strong influence on the life span. The characteristics are predominantly the composition of the metal alloy and the dimensions of the conductor. The shape of the conductor, the crystallographic orientation of the grains in the metal, procedures for the layer deposition, heat treatment or annealing, characteristics of the passivation and the interface to other materials also affect the durability of the interconnects. There are also grave differences with time dependent current: direct current or different alternating current forms cause different effects.

Forces on ions in an electrical field

Two forces affect ionized atoms in a conductor. The direct electrostatic force F_e as a result from the electric field therefore having the same direction. The force from the exchange of momentum with other charge carriers F_p showing toward the flow of charge carriers. In metallic conductors F_p is caused by a so-called "electron wind" or "Ion wind".

The resulting force F_{res} on an activated ion in the electrical field is

$$F_{res} = F_e - F_p = q \cdot Z^* \cdot E = q \cdot Z^* \cdot j \cdot \rho$$

Electromigration occurs when some of the momentum of a moving electron is transferred to a nearby activated ion. This causes the ion to move from its original position. Over time this force knocks a significant number of atoms far from their original positions. A break or gap can develop in the conducting material, preventing the flow of electricity. In narrow interconnect conductors, such as those linking transistors and other components in integrated circuits, this is known as a **void** or **internal failure** open circuit.

Electromigration can also cause the atoms of a conductor to pile up and drift toward other nearby conductors, creating an unintended electrical connection known as a **hillock failure** or **whisker failure** (short circuit). Both of these situations can lead to a malfunction of the circuit.

Failure mechanisms

Diffusion mechanisms

In a homogeneous crystalline structure, because of the uniform lattice structure of the metal ions, there is hardly any momentum transfer between the conduction electrons and the metal ions. However, this symmetry does not exist at the grain boundaries and material interfaces, and so here momentum is transferred much more vigorously. Since the metal ions in these regions are bonded more weakly than in a regular crystal lattice, once the electron wind has reached a certain strength, atoms become separated from the grain boundaries and are transported in the direction of the current. This direction is also influenced by the grain boundary itself, because atoms tend to move along grain boundaries.

Diffusion processes caused by electromigration can be divided into grain boundary diffusion, bulk diffusion and surface diffusion. In general, grain boundary diffusion is the major electromigration process in aluminum wires, whereas surface diffusion is dominant in copper interconnects.

Thermal effects

In an ideal conductor, where atoms are arranged in a perfect lattice structure, the electrons moving through it would experience no collisions and electromigration would not occur. In real conductors, defects in the lattice structure and the random thermal vibration of the atoms about their positions causes electrons to collide with the atoms and scatter, which is the source of electrical resistance. Normally, the amount of momentum imparted by the relatively low-mass electrons is not enough to permanently displace the atoms. However, in high-power situations (such as with the increasing current draw and decreasing wire sizes in modern VLSI microprocessors), if many electrons bombard the atoms with enough force to become significant, this will accelerate the process of electromigration by causing the atoms of the conductor to vibrate further from their ideal lattice positions, increasing the amount of electron scattering. High current density increases the number of electrons scattering against the atoms of the conductor, and hence the speed at which those atoms are displaced.

In integrated circuits, electromigration does not occur in semiconductors directly, but in the metal interconnects deposited onto them.

Electromigration is exacerbated by high current densities and the Joule heating of the conductor, and can lead to eventual failure of electrical components. Localized increase of current density is known as current crowding.

Balance of atom concentration

A governing equation which describes the atom concentration evolution throughout some interconnect segment, is the conventional mass balance (continuity) equation

$$\frac{\partial N}{\partial t} + \nabla \cdot \vec{J} = 0$$

where $N(\vec{x}, t)$ is the atom concentration at the point with a coordinates $\vec{x} = (x, y, z)$ at the moment of time t , and J is the total atomic flux at this location. The total atomic flux J is a combination of the fluxes caused by the different atom migration forces. The major forces are induced by the electric current, and by the gradients of temperature, mechanical stress and concentration. $\vec{J} = \vec{J}_c + \vec{J}_T + \vec{J}_\sigma + \vec{J}_N$. Define

$$\vec{J}_c = \frac{NeZD\rho}{kT} \vec{j}$$

the fluxes mentioned above. Here e is the electron charge, eZ is the effective charge of the migrating atom, ρ the resistivity of the conductor where atom migration takes place, $\vec{j}$ is the local current density, k is Boltzmann's constant, T is the absolute temperature. $D(\vec{x}, t)$ is the time and position dependent atom diffusivity.

$$\vec{J}_T = -\frac{NDQ}{kT^2} \nabla T$$

We use Q the heat of thermal diffusion.

$$\vec{J}_\sigma = \frac{ND\Omega}{kT} \nabla H$$

here $\Omega = 1 / N_0$ is the atomic volume and N_0 is initial atomic concentration, $H = (\sigma_{11} + \sigma_{22} + \sigma_{33}) / 3$ is the hydrostatic stress and $\sigma_{11}, \sigma_{22}, \sigma_{33}$ are the components of principal stress. $\vec{J}_N = -D\nabla N$.

Assuming a vacancy mechanism for atom diffusion we can express D as a function of the

$$D = D_0 \exp\left(\frac{\Omega H - E_A}{kT}\right)$$

hydrostatic stress where E_A is the effective activation energy of the thermal diffusion of metal atoms. The vacancy concentration represents availability of empty lattice sites, which might be occupied by a migrating atom.

Electromigration-aware design

Electromigration reliability of a wire (Black's equation)

At the end of the 1960s J. R. Black developed an empirical model to estimate the MTTF (mean time to failure) of a wire, taking electromigration into consideration:

$$\text{MTTF} = A(J^{-n})e^{\frac{E_a}{kT}}$$

Here A is a constant based on the cross-sectional area of the interconnect, J is the current density, E_a is the activation energy (e.g. 0.7 eV for grain boundary diffusion in

aluminum), k is the Boltzmann's constant, T is the temperature and n a scaling factor (usually set to 2 according to Black). It is clear that **current density J and (less so) the temperature T are deciding factors** in the design process that affect electromigration.

The temperature of the conductor appears in the exponent, i.e. it strongly affects the MTTF of the interconnect. For an interconnect to remain reliable in rising temperatures, the maximum tolerable current density of the conductor must necessarily decrease.

Wire material

The most common conductor used in integrated circuits is aluminium, due to its good adherence to substrate, good conductivity, and formation of ohmic contacts with silicon. However, it soon appeared that pure aluminium is susceptible to electromigration. Research shown adding 2-4% of copper to aluminium increases resistance to electromigration about 50 times. The effect is attributed to grain boundary segregation of copper, which greatly inhibits the diffusion of aluminium atoms across grain boundaries.

It is known that pure copper used for Cu-metallization is more electromigration-robust than aluminum. Copper wires can withstand approximately five times more current density than aluminum wires while assuming similar reliability requirements. This is mainly due to the higher electromigration activation energy levels of copper, caused by its superior electrical and thermal conductivity as well as its higher melting point. Further improvements can be achieved by alloying copper with about 1% palladium, which, similar to copper in aluminium, inhibits diffusion of copper atoms along grain boundaries.

Bamboo structure and metal slotting

It is obvious that a wider wire results in smaller current density and, hence, less likelihood of electromigration. Also, the metal grain size has influence; the smaller grains, the more grain boundaries and the higher likelihood of electromigration effects. However, if you reduce wire width to below the average grain size of the wire material, the resistance to electromigration increases, despite an increase in current density. This apparent contradiction is caused by the position of the grain boundaries, which in such narrow wires as in a bamboo structure lie perpendicular to the width of the whole wire. Because the grain boundaries in these so-called “bamboo structures” are at right angles to the current, the boundary diffusion factor is excluded, and material transport is correspondingly reduced.

However, the maximum wire width possible for a bamboo structure is usually too narrow for signal lines of large-magnitude currents in analog circuits or for power supply lines. In these circumstances, slotted wires are often used, whereby rectangular holes are carved in the wires. Here, the widths of the individual metal structures in between the slots lie within the area of a bamboo structure, while the resulting total width of all the metal structures meets power requirements.

Blech length

There is a lower limit for the length of the interconnect that will allow electromigration to occur. It is known as “Blech length”, and any wire that has a length below this limit will not fail by electromigration. Here, a mechanical stress buildup causes a reversed migration process which reduces or even compensates the effective material flow towards the anode. The Blech length must be considered when designing test structures for electromigration.

Via arrangements and corner bends

Particular attention must be paid to vias and contact holes, because generally the ampacity of a (tungsten) via is less than that of a metal wire of the same width. Hence multiple vias are often used, whereby the geometry of the via array is very significant: Multiple vias must be organized such that the resulting current is distributed as evenly as possible through all the vias.

Attention must also be paid to bends in interconnects. In particular, 90-degree corner bends must be avoided, since the current density in such bends is significantly higher than that in oblique angles (e.g., 135 degrees).

Electromigration in Solder Joints

The typical current density at which electromigration occurs in Cu or Al interconnects is 10^6 to 10^7 A/cm². For solder joints (SnPb or SnAgCu lead-free) used in IC chips, however, electromigration occurs at much lower current densities, e.g. 10^4 A/cm². It causes a net atom transport along the direction of electron flow. The atoms pile up at the anode, voids are generated at the cathode and back stress is induced during electromigration. The typical failure of a solder joint due to electromigration will occur at the cathode side. Due to the current crowding effect, voids form first at the corner of the solder joint. Then the voids extend and cause a failed circuit. Electromigration also influences formation of intermetallic compounds.

Electrostatic Discharge, Overvoltage & Overcurrent

Electrostatic discharge (ESD) is the sudden and momentary electric current that flows between two objects at different electrical potentials. The term is usually used in the electronics and other industries to describe momentary unwanted currents that may cause damage to electronic equipment.

ESD is a serious issue in solid state electronics, such as integrated circuits. Integrated circuits are made from semiconductor materials such as silicon and insulating materials such as silicon dioxide. Either of these materials can suffer permanent damage when subjected to high voltages; as a result, there are now a number of antistatic devices that help prevent static build up.

Causes of ESD

One of the causes of ESD events is static electricity. Static electricity is often generated through tribocharging, the separation of electric charges that occurs when two materials are brought into contact and then separated. Examples of tribocharging include walking on a rug, rubbing plastic comb against dry hair, ascending from a fabric car seat, or removing some types of plastic packaging. In all these cases, the friction between two materials results in tribocharging, thus creating a difference of electrical potential that can lead to an ESD event.

Another cause of **ESD** damage is through electrostatic induction. This occurs when an electrically charged object is placed near a conductive object isolated from ground. The presence of the charged object creates an electrostatic field that causes electrical charges on the surface of the other object to redistribute. Even though the net electrostatic charge of the object has not changed, it now has regions of excess positive and negative charges. An ESD event may occur when the object comes into contact with a conductive path. For example, charged regions on the surfaces of styrofoam cups or plastic bags can induce potential on nearby ESD sensitive components via electrostatic induction and an ESD event may occur if the component is touched with a metallic tool.

Types of ESD

The most spectacular form of ESD is the **spark**, which occurs when a strong electric field creates an ionized conductive channel in air. This can cause minor discomfort to people, severe damage to electronic equipment, and fires and explosions if the air contains combustible gases or particles.

However, many ESD events occur without a visible or audible spark. A person carrying a relatively small electric charge may not feel a discharge that is sufficient to damage sensitive electronic components. Some devices may be damaged by discharges as small as 10 V. These invisible forms of ESD can cause device outright failures, or less obvious forms of degradation that may affect the long term reliability and performance of electronic devices. The degradation in some devices may not become evident until well into their service life.

Sparks

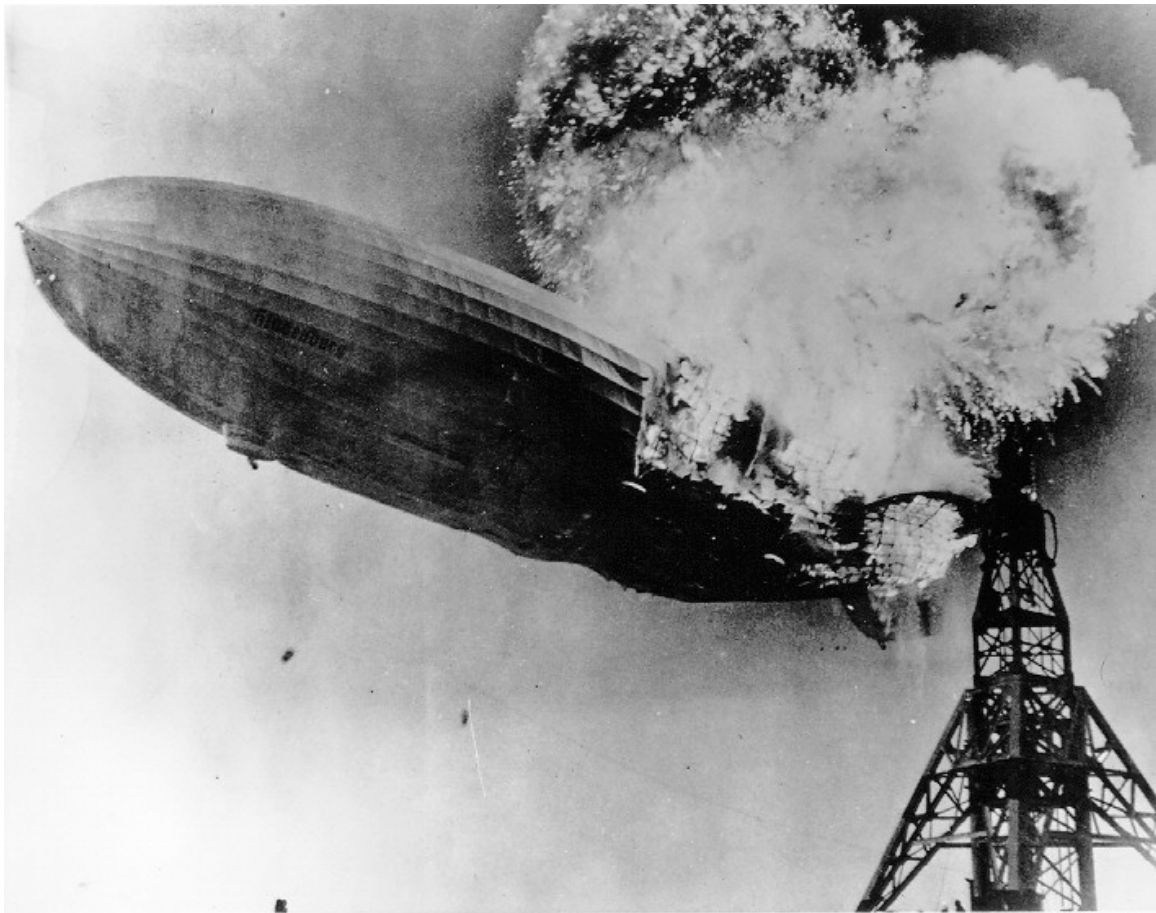
A spark is triggered when the electric field strength exceeds approximately 4–30 kV/cm — the dielectric field strength of air. This may cause a very rapid increase in the number of free electrons and ions in the air, temporarily causing the air to abruptly become an electrical conductor in a process called dielectric breakdown.



Lightning over Rymań. Northern Poland.

Perhaps the best known example of a natural spark is a lightning strike. In this case the potential difference between a cloud and ground, or between two clouds, is typically hundreds of millions of volts. The resulting current that flows through the ionized air causes an explosive release of energy. On a much smaller scale, sparks can form in air during electrostatic discharges from charged objects that are charged to as little as 380 V (Paschen's law).

Earth's atmosphere consists of 21% oxygen (O_2) and 78% nitrogen (N_2). During an electrostatic discharge, the intervening atmosphere becomes electrically overstressed. The diatomic oxygen molecules are split, and then recombine to form ozone (O_3), which is unstable, or reacts with metals and organic matter. If the electrical stress is high enough, nitrogen oxides (NO_x) can form. Both products are toxic to animals, and nitrogen oxides are essential for nitrogen fixation. Ozone attacks all organic matter by ozonolysis and is used in water purification.



The *Hindenburg*, moments after catching fire.

Sparks can cause serious explosions because of the high temperatures reached in a spark. Methane and coal dust explosions have been caused by electrostatic discharges. The Hindenburg disaster has been attributed to spark discharge igniting flammable panels tainted with thermite, which burned vigorously, violently, and extremely swiftly, which

ultimately led to the ignition of hydrogen gas held in or leaking from the airship at the time. The ship had just passed through a thunderstorm and therefore may have acquired a large charge. Discharge supposedly occurred when mooring ropes were dropped as it came in to land in New Jersey in 1937.

Damage prevention in electronics



A portion of a static discharger on an aircraft. Note the two sharp 3/8" metal micropoints and the protective yellow plastic.

Prevention of ESD bases on Electrostatic Protective Area (EPA). EPA can be a small working station or a large manufacturing area. The main principle of an EPA is that there are no highly charging materials in the vicinity of ESD sensitive electronics, all conductive materials are grounded, workers are grounded, and charge build-up on ESD sensitive electronics is prevented. International standards are used to define typical EPA and can be found for example from International Electrotechnical Commission (IEC) or American National Standards Institute (ANSI).

ESD prevention within an EPA may include using appropriate ESD-safe packing material, the use of conductive filaments on garments worn by assembly workers, conducting wrist straps and foot-straps to prevent high voltages from accumulating on

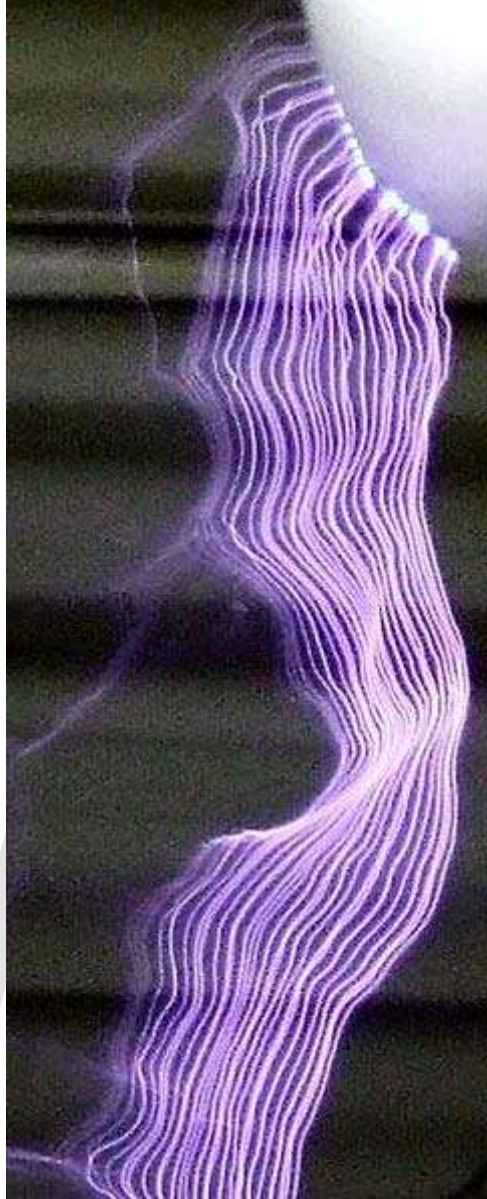
workers' bodies, anti-static mats or conductive flooring materials to conduct harmful electric charges away from the work area, and humidity control. Humid conditions prevent electrostatic charge generation because the thin layer of moisture that accumulates on most surfaces serves to dissipate electric charges.

Ion generators are sometimes used to inject ions into the ambient airstream. Ionization systems help to neutralize charged surface regions on insulative or dielectric materials. Insulating materials prone to triboelectric charging should be kept away from sensitive devices to prevent accidental charging of devices through induction. On aircraft, static dischargers are used on the trailing edges of wings and other surfaces.

Manufacturers and users of integrated circuits must take precautions to avoid ESD. ESD prevention can be part of the device itself and include special design techniques for device input and output pins. External protection components can also be used with circuit layout.

Due to dielectric nature of electronics component and assemblies, electrostatic charging can not be completely prevented during handling of devices. Most of ESD sensitive electronic assemblies and components are also so small that manufacturing and handling is made with automated equipment. ESD prevention activities are therefore important with those processes where component is touching on equipment surfaces. In addition, it is important to prevent ESD when electrostatic discharge sensitive component is connected with other conductive parts of the product itself. An efficient way to prevent ESD is to use materials that are not too conductive but will slowly conduct static charges away. These materials are called static dissipative and have resistivity values in the range of 10^5 to 10^{11} ohm-meters. Materials in automated manufacturing which will touch on conductive areas of ESD sensitive electronic should be made of dissipative material, and the dissipative material must be grounded.

Simulation and testing for electronic devices



Electric discharge showing the ribbon-like plasma filaments from multiple discharges from a Tesla coil.

For testing the susceptibility of electronic devices to ESD from human contact, an ESD Simulator with a special output circuit, called the human body model (HBM) is often used. This consists of a capacitor in series with a resistor. The capacitor is charged to a specified high voltage from an external source, and then suddenly discharged through the resistor into an electrical terminal of the device under test. One of the most widely used models is defined in the JEDEC 22-A114-B standard, which specifies a 100 picofarad capacitor and a 1500 ohm resistor. Other similar standards are MIL-STD-883 Method 3015, and the ESD Association's ESD STM5.1. For conformance to European Union standards for Information Technology Equipment, the IEC-61000-4-2 test specification is used. Guidelines and requirements are given for test cell geometries, generator

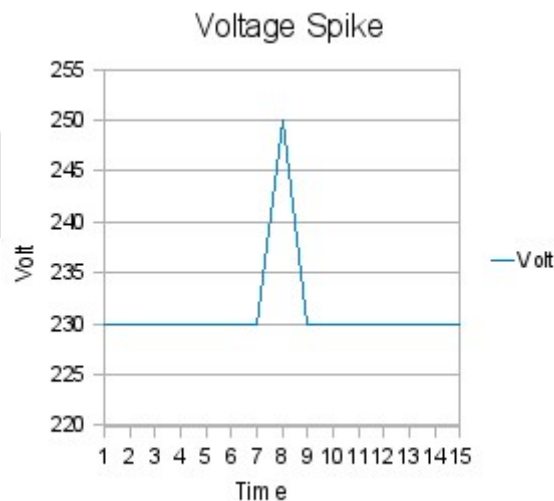
specifications, test levels, discharge rate and waveform, types and points of discharge on the "victim" product, and functional criteria for gauging product survivability.

A Charged Device Model (CDM) test is used to define the ESD a device can withstand when the device itself has an electrostatic charge and discharges due to metal contact. This discharge type is the most common type of ESD in electronic devices and causes most of the ESD damages in their manufacturing. CDM discharge depends mainly on parasitic parameters of the discharge and is strongly depending on size and type of component package. One of the most widely used CDM simulation test models is defined by the JEDEC.

Other standardized ESD test circuits include the following:

- Machine model (MM)
- Transmission line pulse (TLP)

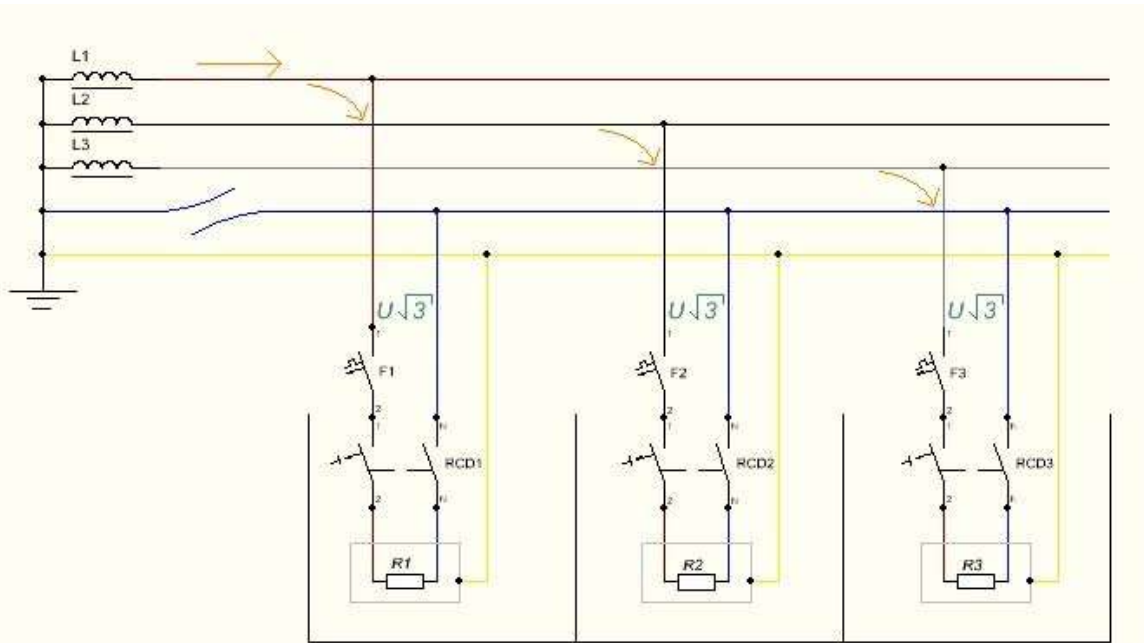
Overvoltage



Voltage spike.

When the voltage in a circuit or part of it is raised above its upper design limit, this is known as **overvoltage**. The conditions may be hazardous. Depending on its duration, the overvoltage event can be transient—a voltage spike—or permanent, leading to a power surge.

Explanation



Lack of 3-phase electric system connected by star. If zero breaks off, small-power appliances will be destroyed by overvoltage

Electronic and electrical devices are designed to operate at a certain maximum supply voltage, and considerable damage can be caused by voltage that is higher than that for which the devices are rated.

For example an electric light bulb has a wire in it that at the given rated voltage will carry a current just large enough for the wire to get very hot (giving off light and heat), but not hot enough for it to melt. The amount of current in a circuit depends on the voltage supplied: if the voltage is too high, then the wire may melt and the light bulb would have "burned out". Similarly other electrical devices may stop working, or may even burst into flames if an overvoltage is delivered to the circuit of which these devices are part.

Sources

Natural

A typical natural source of transient overvoltage events is lightning. Bursts of solar wind following solar flares are also known to cause overvoltage in electrical circuits, especially onboard space satellites.

Man made

Man-made sources of spikes are usually caused by electromagnetic induction when switching on or off inductive loads (such as electric motors or electromagnets), or by switching heavy resistive AC loads when zero-crossing circuitry is not used - anywhere a large change of current takes place. One of the purposes of electromagnetic compatibility compliance is to eliminate such sources.

An important potential source of dangerous overvoltage is electronic warfare. There is intensive military research in this field, whose goal is to produce various transient electromagnetic devices designed to generate electromagnetic pulses that will disable an enemy's electronic equipment. A recent military development is that of the exploding capacitor designed to radiate a high voltage electromagnetic pulse. Another intense source of an electromagnetic pulse is a nuclear explosion.

Conduction path

The transient pulses can get into the equipment either by power or data lines, or directly through space from a strong electromagnetic field change - an electromagnetic pulse (EMP). Filters are used to prevent spikes entering or leaving the equipment through wires, and the devices coupled electromagnetically to space (such as radio-frequency pick-up coils in MRI scanners) are protected by shielding.

Overvoltage protection devices

- Arcing horns
- Zener diode
- Avalanche diode
- Transil
- Trisil
- Spark gap
- Gas filled tube
- Metal Oxide Varistor

Overcurrent

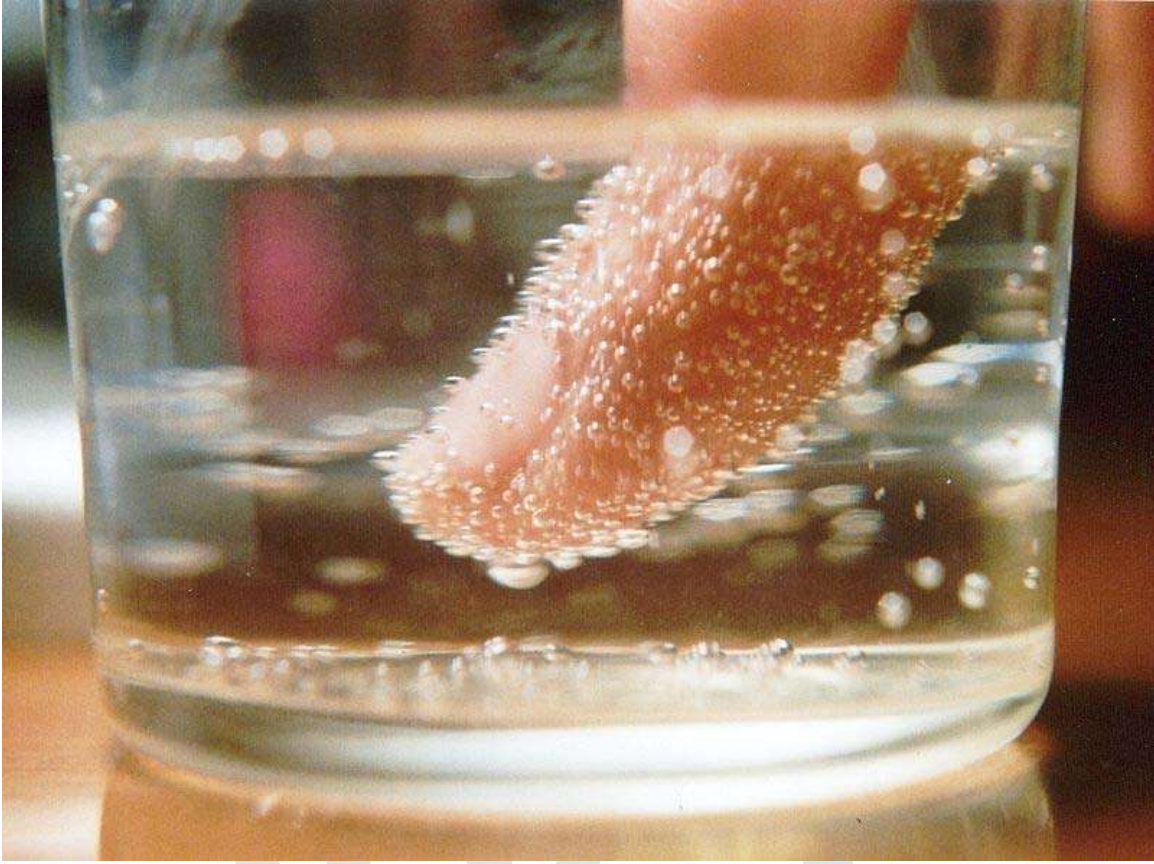
In electricity supply, **overcurrent** or **excess current** is a situation where a larger than intended electric current exists through a conductor, leading to excessive generation of heat, and the risk of fire or damage to equipment. Possible causes for overcurrent include short circuits, excessive load, and incorrect design. Fuses, circuit breakers, temperature sensors and current limiters are commonly used protection mechanisms to control the risks of overcurrent.

Semiconductor-Related Failures

Nucleation

Nucleation is the extremely localized budding of a distinct thermodynamic phase. Some examples of phases that may form via nucleation in liquids are gaseous bubbles, crystals or glassy regions. Creation of liquid droplets in saturated vapor is also characterized by nucleation. Nucleation of crystalline, amorphous and even vacancy clusters in solid materials is also important, for example to the semiconductor industry. Most nucleation processes are physical, rather than chemical, but a few exceptions do exist (e.g. electrochemical nucleation). A good example would be the famous Diet Coke and Mentos eruption. Nucleation normally occurs at *nucleation sites* on surfaces contacting the liquid or vapor. Suspended particles or minute bubbles also provide nucleation sites. This is called *heterogeneous nucleation*. Nucleation without preferential nucleation sites is *homogeneous nucleation*. Homogeneous nucleation occurs spontaneously and randomly, but it requires superheating or supercooling of the medium. Nucleation is involved in such processes as cloud seeding and in instruments such as the bubble chamber and the cloud chamber.

Examples of nucleation



Nucleation of carbon dioxide bubbles around a finger

- Pure water freezes at -42°C rather than at its freezing temperature of 0°C if no crystal nuclei, such as dust particles, are present to form an ice nucleus.
- Presence of cloud condensation nuclei is important in meteorology because they are often in short supply in the upper atmosphere.
- All natural and artificial crystallization process (of formation of solid crystals from a homogeneous solution) starts with a *nucleation* event.
- Bubbles of carbon dioxide *nucleate* shortly after the pressure is released from a container of carbonated liquid. Nucleation often occurs more easily at a pre-existing interface (*heterogeneous nucleation*), as happens on boiling chips and string used to make rock candy. So-called Diet Coke and Mentos eruptions are a dramatic example.
- Nucleation in boiling can occur in the bulk liquid if the pressure is reduced so that the liquid becomes superheated with respect to the pressure-dependent boiling point. More often nucleation occurs on the heating surface, at *nucleation sites*. Typically, nucleation sites are tiny crevices where free gas-liquid surface is maintained or spots on the heating surface with lower wetting properties. Substantial superheating of a liquid can be achieved after the liquid is de-gassed and if the heating surfaces are clean, smooth and made of materials well wetted by the liquid.

- Nucleation is relevant in the process of crystallization of nanometer sized materials, and plays an important role in atmospheric processes.
- Nucleation is a key concept in polymer, alloy and ceramic systems.
- In chemistry and biophysics, nucleation can also refer to the phaseless formation of multimers which are intermediates in polymerization processes. This sort of process is believed to be the best model for processes such as crystallization and amyloidogenesis.
- In molecular biology, nucleation is used to term the critical stage in the assembly of a polymeric structure, such as a microfilament, at which a small cluster of monomers aggregates in the correct arrangement to initiate rapid polymerization. For instance, two actin molecules bind weakly, but addition of a third stabilizes the complex. This trimer then adds additional molecules and forms a nucleation site. The nucleation site serves the slow, or lag phase of the polymerization process.
- Some champagne stirrers operate by providing many nucleation sites via high surface area and sharp corners, speeding the release of bubbles and removing carbonation from the wine.
- Sodium acetate heating pads use a metal disk as a nucleation centre for exothermic crystallization

Ionizing radiation



Radiation hazard symbol.



Ionizing radiation hazard symbol (recently introduced).

Ionizing radiation consists of particles or electromagnetic waves energetic enough to detach electrons from atoms or molecules, thus ionizing them. The degree and nature of such ionization depends on the energy of the individual particles or waves, and not on their number. An intense flood of particles or waves will not cause ionization if these particles or waves do not carry enough energy to be ionizing. Roughly speaking, particles or photons with energies above a few electron volts (eV) are ionizing. Ionization produces free radicals, which are atoms or molecules containing unpaired electrons, that tend to be especially chemically reactive due to their electronic structure.

Examples of ionizing particles are alpha particles, beta particles, neutrons, and cosmic rays. The ability of an electromagnetic wave (photons) to ionize an atom or molecule depends on its frequency. Radiation on the short-wavelength end of the electromagnetic spectrum—high frequency ultraviolet, x-rays, and gamma rays—is ionizing. Lower-energy radiation, such as visible light, infrared, microwaves, and radio waves, are not ionizing.

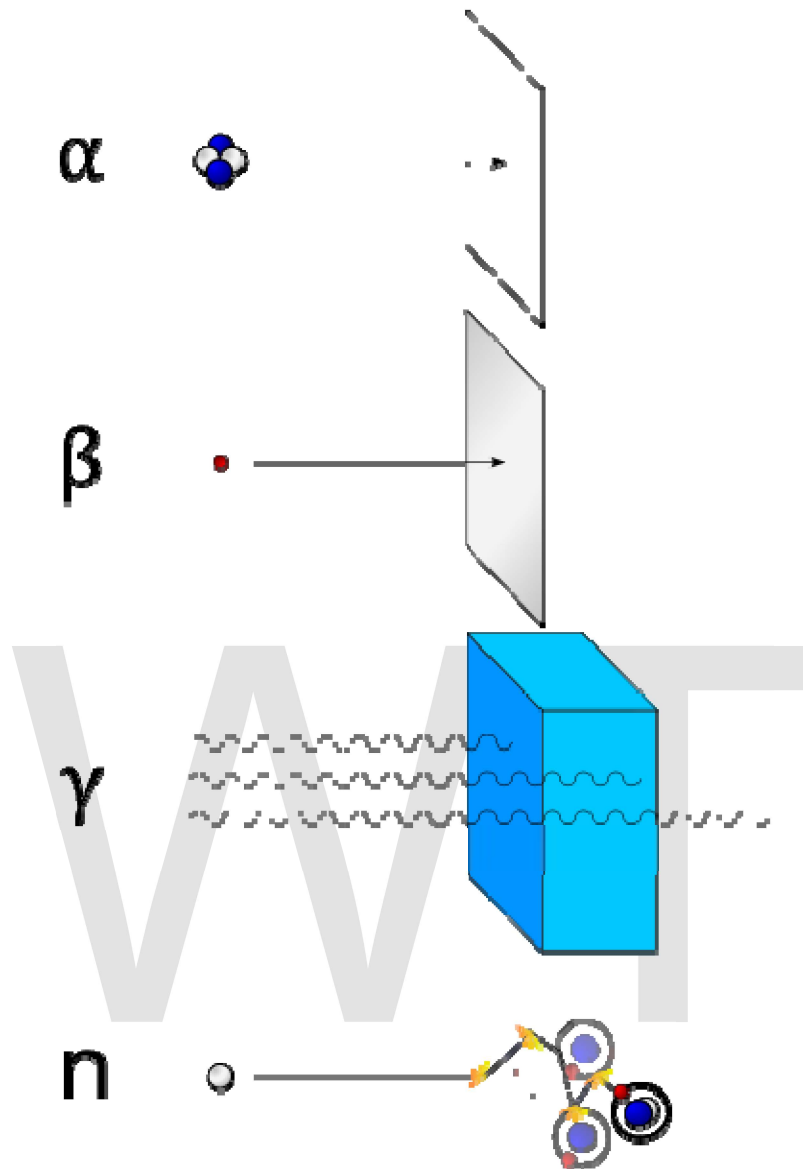
The latter types of lower energy electromagnetic radiation may damage molecules, but the effect is generally indistinguishable from the effects of simple heating. Such heating does not produce free radicals until higher temperatures (for example, flame temperatures or "browning" temperatures, and above) are attained. In contrast, damage done by ionizing radiation produces free radicals, even at room temperatures and below, and production of such free radicals is the reason these and other ionizing radiations produce quite different types of chemical effects from (low-temperature) heating. Free radical

production is also a primary basis for the particular danger to biological systems of relatively small amounts of ionizing radiation that are far smaller than needed to produce significant heating. Free radicals easily damage DNA, and ionizing radiation may also directly damage DNA by ionizing or breaking DNA molecules.

Ionizing radiation is ubiquitous in the environment, and also comes from radioactive materials, x-ray tubes, and particle accelerators. It is invisible and not directly detectable by human senses, so instruments such as Geiger counters are usually required to detect its presence. In some cases it may lead to secondary emission of visible light upon interaction with matter, as in Cherenkov radiation and radioluminescence. It has many practical uses in medicine, research, construction, and other areas, but presents a health hazard if used improperly. Exposure to radiation causes damage to living tissue, and high doses can result in mutation, radiation sickness, cancer, and death.

Types

WWT



Alpha (α) radiation consists of a fast moving helium-4 (^4He) nucleus and is stopped by a sheet of paper. Beta (β) radiation, consisting of electrons, is halted by an aluminium plate. Gamma (γ) radiation, consisting of energetic photons, is eventually absorbed as it penetrates a dense material. Neutron (n) radiation consists of free neutrons which are blocked using light elements, like hydrogen, which slow and/or capture them.

Various types of ionizing radiation may be produced by radioactive decay, nuclear fission and nuclear fusion, and by particle accelerators and naturally occurring cosmic rays. Muons and many types of mesons (in particular charged pions) are also ionizing.

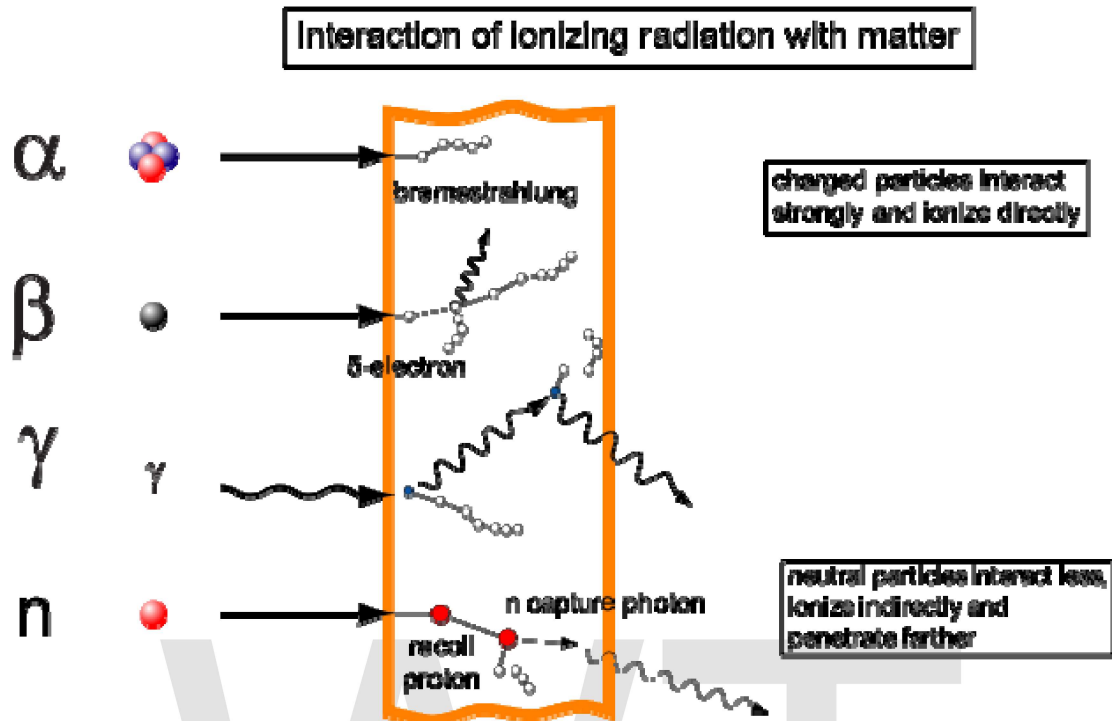
In order for a particle to be ionizing, it must both have a high enough energy and interact with the atoms of a target.

Photons interact electromagnetically with charged particles, so photons of sufficiently high energy also are ionizing. The energy at which this begins to happen with photons (light) is in the high frequency end of the ultraviolet region of the electromagnetic spectrum.

Charged particles such as electrons, positrons, muons, protons, alpha particles, and heavy atomic nuclei from accelerators or cosmic rays also interact electromagnetically with electrons of an atom or molecule. Muons contribute to background radiation due to cosmic rays, but by themselves are thought to be of little hazard importance due to their relatively low dose. Pions (another very short-lived sometimes-charged particle) may be produced in large amounts in the largest particle accelerators, but are not a theoretical biological hazard except near such machines, which are in practice are subject to heavy security while in use.

Neutrons, on the other hand, having zero electrical charge, do not interact electromagnetically with electrons, and so they cannot directly cause ionization by this mechanism. However, fast neutrons will interact with the protons in hydrogen (in the manner of a billiard ball hitting another, head on, sending it away with all of the first ball's energy of motion), and this mechanism produces proton radiation (fast protons). These protons are ionizing because they are charged, and interact with the electrons in matter.

A neutron can also interact with other atomic nuclei, depending on the nucleus and the neutron's velocity; these reactions happen with fast neutrons and slow neutrons, depending on the situation. Neutron interactions in this manner often produce radioactive nuclei, which produce ionizing radiation when they decay, then they can produce chain reactions in the mass that is decaying, sometimes causing a larger effect of ionization.



Types of radiation - gamma rays are represented by wavy lines, charged particles and neutrons by straight lines. The little circles show where ionization processes occur.

An ionization event normally produces a positive atomic ion and an electron. High-energy beta particles may produce bremsstrahlung when passing through matter, or secondary electrons (δ -electrons); both can ionize in turn. Energetic Beta-particles, like those emitted by ^{32}P , are quickly decelerated when passing through matter. The energy lost to deceleration is emitted in the form of X-rays called "Bremsstrahlung" which translates "Braking Radiation". Bremsstrahlung is of concern when shielding beta emitters. The intensity of bremsstrahlung increases with the increase in energy of the electrons or the atomic number of the absorbing medium.

Unlike alpha or beta particles, gamma rays do not ionize all along their path, but rather interact with matter in one of three ways: the photoelectric effect, the Compton effect, and pair production. By way of example, the **figure** shows Compton effect: two Compton scatterings that happen sequentially. In every scattering event, the gamma ray transfers energy to an electron, and it continues on its path in a different direction and with reduced energy.

In the same figure, the neutron collides with a proton of the target material, and then becomes a fast recoil proton that ionizes in turn. At the end of its path, the neutron is captured by a nucleus in an (n, γ)-reaction that leads to a neutron capture photon.

Neutron radiation

Neutron radiation is a kind of non-ionizing radiation which consists of free neutrons. A result of nuclear fission or nuclear fusion, it consists of the release of free neutrons from both stable molecules and isotopes, and these free neutrons react with nuclei of other stable molecules to form new isotopes of previously non-isotopic molecules, which in turn produce radiation. This will result in a chain reaction of nuclear radiation, which makes radiation dangerous and harmful over great areas of space.

Sources

Neutrons may be emitted from nuclear fusion or nuclear fission, or from any number of different nuclear reactions such as from radioactive decay or reactions from particle interactions (such as from cosmic rays or particle accelerators). Large neutron sources are rare, and are usually limited to large-sized devices like nuclear reactors or particle accelerators (such as the Spallation Neutron Source).

Neutron radiation was discovered as a result of observing a beryllium nucleus reacting with an alpha particle thus transforming into a carbon nucleus and emitting a neutron, $\text{Be}(\alpha, n)\text{C}$. The combination of an alpha particle emitter and an isotope with a large (α, n) nuclear reaction probability is still a common neutron source.

Uses

Cold, thermal and *hot* neutron radiation is most commonly used for scattering and diffraction experiments in order to access the properties and the structure of materials in crystallography, condensed matter physics, biology, solid state chemistry, materials science, geology, mineralogy and related sciences. Neutron radiation is also used in select facilities to treat cancerous tumors due to its highly penetrating and damaging nature to cellular structure. Neutrons can also be used for imaging of industrial parts termed neutron radiography when using film, neutron radioscopy when taking a digital image, such as through image plates, and neutron tomography for three dimensional images. Neutron imaging is commonly used in the nuclear industry, the space and aerospace industry, as well as the high reliability explosives industry.

Ionization mechanisms and properties

Neutron radiation is often called *indirectly ionizing radiation*. It does not ionize atoms in the same way that charged particles such as protons and electrons do (exciting an electron), because neutrons have no charge. However, neutron interactions are largely ionizing, for example when neutron absorption results in gamma emission and the gamma subsequently removes an electron from an atom, or a nucleus recoiling from a neutron interaction is ionized and causes more traditional subsequent ionization in other atoms. Because neutrons are uncharged, they are more penetrating than alpha radiation or

beta radiation. In some cases they are more penetrating than gamma radiation, which is impeded in materials of high atomic number. In hydrogen, a low energy neutron may not be as penetrating as a high energy gamma.

Health hazards and protection

In health physics neutron radiation is considered a fourth radiation hazard alongside the other types of radiation. Another, sometimes more severe hazard of neutron radiation, is neutron activation, the ability of neutron radiation to induce radioactivity in most substances it encounters, including the body tissues of the workers themselves. This occurs through the capture of neutrons by atomic nuclei, which are transformed to another nuclide, frequently a radionuclide. This process accounts for much of the radioactive material released by the detonation of a nuclear weapon. It is also a problem in nuclear fission and nuclear fusion installations, as it gradually renders the equipment radioactive; eventually the hardware must be replaced and disposed of as low-level radioactive waste.

Neutron radiation protection relies on radiation shielding. In comparison with conventional ionizing radiation based on photons or charged particles, neutrons are repeatedly bounced and slowed (absorbed) by light nuclei, so a large mass of hydrogen-rich material is needed. Neutrons readily pass through most material, but interact enough to cause biological damage. Due to the high kinetic energy of neutrons, this radiation is considered to be the most severe and dangerous radiation available. The most effective materials are e.g. water, polyethylene, paraffin wax, or concrete, where a considerable amount of water molecules are chemically bound to the cement. The light atoms serve to slow down the neutrons by elastic scattering, so they can then be absorbed by nuclear reactions. However, gamma radiation is often produced in such reactions, so additional shielding has to be provided to absorb it.

Because the neutrons that strike the hydrogen nucleus (proton, or deuteron) impart energy to that nucleus, they in turn will break from their chemical bonds and travel a short distance, before stopping. Those protons and deuterons are high linear energy transfer particles, and are in turn stopped by ionization of the material through which they travel. Consequently, in living tissue, neutrons have a relatively high relative biological effectiveness, and are roughly ten times more effective at causing cancers or LD-50s compared to photon or beta radiation of equivalent radiation exposure.

Effects on materials

Neutrons also degrade materials; bombardment of materials with neutrons creates collision cascades that can produce point defects and dislocations in the materials. At high neutron fluences this can lead to embrittlement of metals and other materials, and to swelling of some of them. This poses a problem for nuclear reactor vessels, and significantly limits their lifetime (which can be somewhat prolonged by controlled annealing of the vessel, reducing the number of the built-up dislocations). Graphite

moderator blocks are especially susceptible to this effect, known as Wigner effect, and have to be annealed periodically; the well-known Windscale fire was caused by a mishap during such an annealing operation.

Neutron radiation and nuclear fission

The neutrons in reactors are generally categorized as slow (thermal) neutrons or fast neutrons depending on their energy. Thermal neutrons are similar to a gas in thermodynamic equilibrium but are easily captured by atomic nuclei and are the primary means by which elements undergo atomic transmutation.

In order to achieve an effective fission chain reaction, the neutrons produced during fission must be captured by fissionable nuclei, which then split, releasing more neutrons. In most fission reactor designs, the nuclear fuel is not sufficiently refined to be able to absorb enough fast neutrons to carry on the fission chain reaction, due to the lower cross section for higher-energy neutrons, so a neutron moderator must be introduced to slow the fast neutrons down to thermal velocities to permit sufficient absorption. Common neutron moderators include graphite, light water and heavy water. A few reactors (fast neutron reactors) and all nuclear weapons rely on fast neutrons. This requires certain changes in the design and in the required nuclear fuel. The element beryllium is particularly useful due to its ability to act as a neutron reflector or lens. This allows smaller quantities of fissile material to be used and is a primary technical development that led to the creation of neutron bombs.

Cosmogenic neutrons

Cosmogenic neutrons, neutrons produced from cosmic radiation in the Earth's atmosphere or surface, and those produced in particle accelerators can be significantly higher energy than those encountered in reactors. Most of them activate a nucleus before reaching the ground; a few react with nuclei in the air. The reactions with nitrogen 14 lead to the formation of carbon 14, widely used in radiocarbon dating.

SILC (semiconductors)

Stress Induced Leakage Current (SILC) is an increase in the gate leakage current of a MOSFET, due to defects created in the gate oxide during electrical stressing. SILC is perhaps the largest factor inhibiting device miniturization. Increased leakage is a common failure mode of electronic devices.

Oxide Defects

The most well studied defects assisting in the leakage current are those produced by charge trapping in the oxide. This model provides a point of attack and has stimulated

researchers to develop methods to decrease the rate of charge trapping by mechanisms such as N_2O nitridation of the oxide.

Hot carriers injection

Hot carriers injection (HCI) is the phenomenon in solid-state or semiconductor electronic devices where either an electron or a “hole” gains sufficient kinetic energy to overcome a potential barrier necessary to break an interface state. The term hot electron comes from the effective temperature used to model carrier density (with fermi dirac function). One should not think that the term "hot" refers to the actual temperature of the MOS transistor. At higher temperatures, the mean free path (distance between two collisions with atoms in the substrate) is shorter, which decreases the energy gain by the carrier. As a result, Hot Carrier Degradation is more important at low temperature.

HCI and CMOS Semiconductor Technology

Semiconductor Physics of HCI

The term “hot carrier injection” usually refers to the effect in MOSFETs, where a carrier is injected from the conducting channel in the silicon substrate to the gate dielectric, which usually is made of silicon dioxide (SiO_2).

To become “hot” and enter the conduction band of SiO_2 , an electron must gain a kinetic energy of 3.3 eV. For holes, the valence band offset in this case dictates they must have a kinetic energy of 4.6 eV.

When electrons are accelerated in the channel, they gain energy along the mean free path. This energy is lost in two different ways:

1. The carrier hit an atom in the substrate. Then the collision create a cold carrier and an additional electron-hole pair. In the case of nMOS transistors, additional electrons are collected by the channel and additional holes are evacuated by the substrate.
2. The carrier hit a Si-H bond and break the bond. An interface state is created and the Hydrogen atom is released in the substrate.

The probability to hit either an atom or a Si-H bond is random. And the average energy involved in each process is the same in both case.

This is the reason why the substrate current is monitored during HCI stress. A high substrate current means a large number of created electron-hole pairs and thus an efficient Si-H bond breakage mechanism.

When interface states are created, the threshold voltage is modified and the subthreshold slope is degraded. This leads to lower current, and degrades the operating frequency of integrated circuit.

Scaling and HCI

Advances in semiconductor manufacturing techniques and ever increasing demand for faster and more complex integrated circuits (ICs) have driven the associated Metal–Oxide–Semiconductor field-effect transistor (MOSFET) to scale to smaller dimensions.

However, it has not been possible to scale the supply voltage used to operate these ICs proportionately due to factors such as compatibility with previous generation circuits, noise margin, power and delay requirements, and non-scaling of threshold voltage, subthreshold slope, and parasitic capacitance.

As a result internal electric fields increase in aggressively scaled MOSFETs, which comes with the additional benefit of increased carrier velocities (up to velocity saturation), and hence increased switching speed, but also presents a major reliability problem for the long term operation of these devices, as high fields induce hot carrier injection which affects device reliability.

Large electric fields in MOSFETs imply the presence of high-energy carriers, referred to as “**hot carriers**”. These hot carriers that have sufficiently high energies and momenta to allow them to be injected from the semiconductor into the surrounding dielectric films such as the gate and sidewall oxides as well as the buried oxide in the case of silicon on insulator (SOI) MOSFETs.

CMOS Reliability Impact of HCI

The presence of such mobile carriers in the oxides triggers numerous physical damage processes that can drastically change the device characteristics over prolonged periods. The accumulation of damage can eventually cause the circuit to fail as key parameters such as threshold voltage shift due to such damage. The accumulation of damage resulting degradation in device behavior due to hot carrier injection is called “**hot carrier degradation**”.

The useful life-time of circuits and integrated circuits based on such a MOS device are thus affected by the life-time of the MOS device itself. To assure that integrated circuits manufactured with minimal geometry devices will not have their useful life impaired, the life-time of the component MOS devices must have their HCI degradation well understood. Failure to accurately characterize HCI life-time effects can ultimately affect business costs such as warranty and support costs and impact marketing and sales promises for a foundry or IC manufacturer.

Relationship to Radiation Effects

Hot carrier degradation is fundamentally same as the ionization radiation effect known as the total dose damage to semiconductors, as experienced in space systems due to solar proton, electron, X-ray and gamma ray exposure.

HCI and NOR Flash Memory Cells

HCI is the basis of operation for a number of non-volatile memory technologies such as Electrically Erasable Programmable Read-Only Memory (EEPROM) cells. As soon as the potential detrimental influence of HC injection on the circuit reliability was recognized, several fabrication strategies were devised to reduce it without compromising the circuit performance.

NOR flash memory exploits the principle of hot carriers injection by deliberately injecting carriers across the gate oxide to charge the floating gate. This charge alters the MOS transistor threshold voltage to represent a logic '0' state. An uncharged floating gate represents a '1' state. Erasing the NOR Flash memory cell removes stored charge through the process of Fowler–Nordheim tunneling.

Because of the damage to the oxide caused by normal NOR Flash operation, HCI damage is one of the factors that cause the number of write-erase cycles to be limited. Because the ability to hold charge and the formation of damage traps in the oxide affects the ability to have distinct '1' and '0' charge states, HCI damage results in the closing of the non-volatile memory logic margin window over time. The number of write-erase cycles at which '1' and '0' can no longer be distinguished defines the endurance of a non-volatile memory.

Stress-Related Failures

Most stress-related failures are electrothermal in nature. The locally increased temperature can lead to immediate failure by melting or vaporizing metalization layers, melting the semiconductor, or creating other structural changes. The diffusion and electromigration effects tend to be accelerated by high temperature, which shortens the lifetime of the device. Damages to junctions that do not lead to immediate failure manifest as altered current-voltage characteristics of the junctions.

Electrical overstress failures can be classified as thermally induced failures, electromigration related failures, and electric field related failures.

Thermal runaway

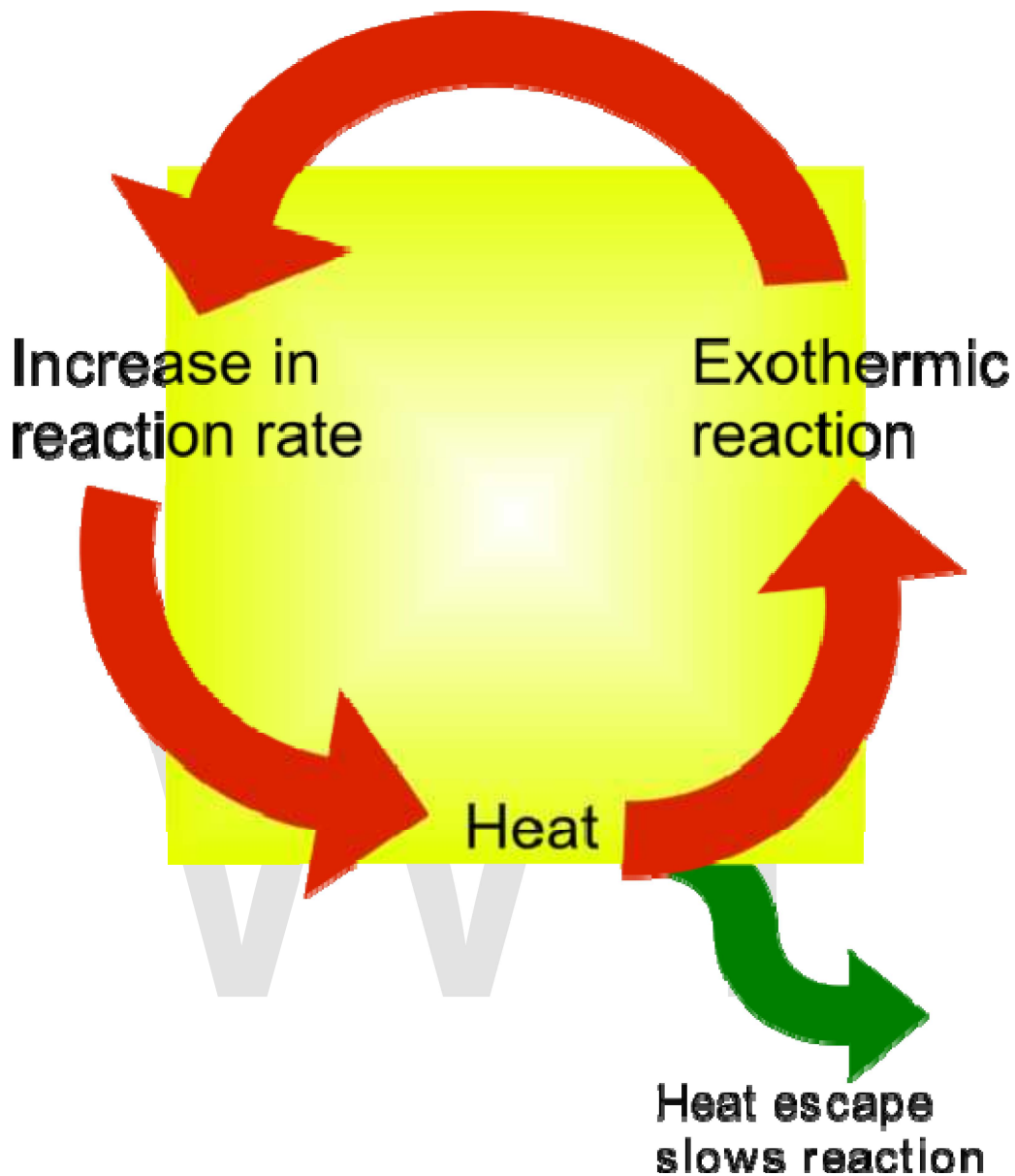


Diagram of thermal runaway.

Thermal runaway refers to a situation where an increase in temperature changes the conditions in a way that causes a further increase in temperature, leading (in the normal case of an exothermic reaction) to a destructive result. It is a kind of positive feedback.

Chemical engineering

In chemical engineering, thermal runaway is a process by which an exothermic reaction goes out of control, often resulting in an explosion. It is also known as a "runaway reaction" in organic chemistry.

Thermal runaway occurs when the reaction rate increases due to an increase in temperature, causing a further increase in temperature and hence a further increase in the reaction rate. It has contributed to industrial chemical accidents, most notably the 1947 Texas City disaster from overheated ammonium nitrate in a ship's hold, and the disastrous release of a large volume of methyl isocyanate gas from a Union Carbide plant in Bhopal, India in 1984. Thermal runaway is also a concern in hydrocracking, an oil refinery process. Thermal runaway may result from exothermic side reaction(s) that begin at higher temperatures, following an initial accidental overheating of the reaction mixture. This scenario was behind the Seveso disaster, where thermal runaway heated a reaction to temperatures such that in addition to the intended 2,4,5-trichlorophenol, poisonous 2,3,7,8-tetrachlorodibenzo-p-dioxin was also produced, and was vented into the environment after the reactor's rupture disk burst.

Thermal runaway is most often caused by failure of the reactor vessel's cooling system. Failure of the mixer can result in localized heating, which initiates thermal runaway. Similarly, in flow reactors, localized insufficient mixing causes hotspots to form, where thermal runaway conditions occur, which causes blowouts of reactor contents and catalysts. Incorrect component installation is also a common cause. Many chemical production facilities are designed with high-volume emergency venting to limit the extent of injury and property damage when such accidents occur.

Some laboratory reactions must be run under extreme cooling, because they are prone to hazardous thermal runaway. For example, in Swern oxidation, the formation of the sulfonium chloride must be performed in a cooled system ($-30\text{ }^{\circ}\text{C}$), because at room temperature the reaction undergoes thermal runaway explosively.

The UK Chemical Reaction Hazards Forum publishes previously unreported chemical accidents to assist the education of the scientific community, with the aim of preventing similar occurrences elsewhere. Almost 150 such reports are available to view at the present time (Jan 2009).

Microwave heating

Microwaves are used for heating of various materials in cooking and various industrial processes. The rate of heating of the material depends on the energy absorption, which depends on the dielectric constant of the material. The dependence of dielectric constant on temperature varies for different materials; some materials display significant increase with increasing temperature. This behavior, when the material gets exposed to microwaves, leads to selective local overheating, as the warmer areas are better able to accept further energy than the colder areas—potentially dangerous especially for thermal insulators, where the heat exchange between the hot spots and the rest of the material is slow. These materials are called *thermal runaway materials*. This phenomenon occurs in some ceramics.

Semiconductors

Silicon shows a peculiar instability. Its electrical resistance increases with temperature up to about 160 °C, then starts decreasing, and drops further when the melting point is reached. This leads to thermal runaway phenomena within the semiconductor junction areas; the resistance of the area which becomes hot above this threshold decreases, more of current flows through it causing yet more heating in comparison with the surrounding areas, which leads to further temperature increase and resistance decrease. This leads to the phenomenon of current crowding and formation of current filaments, and is one of the underlying principles behind many semiconductor junction failure mechanisms.

Bipolar transistors

Leakage current increases significantly in bipolar transistors (notably germanium-based bipolar transistors) as they increase in temperature. Depending on the design of the circuit, this increase in leakage current can increase the current flowing through the transistor and with it the power dissipation. This causes a further increase in C–E current. This is frequently seen in a push–pull stage of a class AB amplifier. If the transistors are biased to have minimal crossover distortion at room temperature, and the biasing is not made temperature dependent, as the temperature rises, both transistors will be increasingly turned on, causing current and power to further increase, eventually destroying one or both devices.

To avoid thermal runaway the operating point of BJT should be $V_{ce} \leq 1/2V_{cc}$

The proper practice is to mount a thermal feedback transistor or other device on the heat sink, which generates the crossover bias voltage, so that as the output transistors heat up, so does the thermal feedback transistor, causing the thermal feedback transistor to also start to turn on at a slightly lower voltage, and thus reduce the crossover bias, and reduce the heat released by the output transistors. It can be surprising what professional sound equipment does not have thermal feedback and would thus need modification.

If multiple bipolar transistors are connected in parallel (which is typical in high current applications) one device will enter thermal runaway first, taking the current which originally was distributed across all the devices and exacerbating the problem. This effect is called **current hogging**. Eventually one of two things will happen, either the circuit will stabilize or the transistor in thermal runaway will be destroyed by the heat. Hence current hogging term is related to thermal runaway.

Power MOSFETs

Power MOSFETs display increase of the on-resistance with temperature. Power dissipated in this resistance causes more heating of the junction, which further increases the junction temperature, in a positive feedback loop. (However, the increase of on-resistance with temperature helps balance current across multiple MOSFETs connected in parallel and current hogging does not occur). If the transistor produces more heat than the heatsink can dissipate, the thermal runaway happens and destroys the transistor. This

problem can be alleviated to a degree by lowering the thermal resistance between the transistor die and the heatsink.

- Java applet demo of MOSFET thermal runaway

Digital electronics

The leakage currents of transistors increase with temperature. In rare instances, this may lead to thermal runaway in digital electronics. This is not a common problem, since leakage currents make up a small portion of overall power consumption, so the increase in power is fairly modest - for an Athlon 64, the power dissipation increases by about 10% for every 30 degrees Celsius. For a device with a TDP of 100 W, for thermal runaway to occur, the heat sink would have to have a thermal resistivity of over 3 K/W (kelvins per watt), which is about 6 times worse than a stock Athlon 64 heat sink. (A stock Athlon 64 heat sink is rated at 0.34 K/W, although the actual thermal resistance to the environment is somewhat higher, due to the thermal boundary between processor and heatsink, rising temperatures in the case, and other thermal resistances..) A heat sink with a thermal resistance of over 0.5 to 1 K/W would result in the destruction of a 100 W device even without thermal runaway effects.

Electronics and tropical environments

Many electronic circuits contain provisions against thermal runaway. This is most often seen in transistor biasing arrangements. However when equipment is used above its designed ambient temperature, thermal runaway can in some cases still occur. This occasionally causes equipment failures in tropical countries, and when air cooling vents are blocked.

Tantalum capacitors

Tantalum capacitors are under some conditions prone to self-destruction by thermal runaway. The capacitor typically consists of a sintered tantalum sponge acting as the anode, a manganese dioxide cathode, and a dielectric layer of tantalum pentoxide created on the tantalum sponge surface by anodizing. The tantalum oxide layer may have weak spots that undergo dielectric breakdown during a voltage spike. The tantalum then comes to direct contact with manganese dioxide and the leakage current causes a local heating; a chemical reaction then produces manganese(III) oxide and regenerates (self-heals) the tantalum oxide layer.

If the energy dissipated at the failure point is high enough, a self-sustaining exothermic reaction may occur, similar to the thermite reaction, with tantalum as fuel and manganese dioxide as oxidizer, destroying the capacitor and occasionally producing smoke and possibly flame.

Batteries

When handled improperly, some rechargeable batteries can experience thermal runaway, resulting in overheating. Sealed cells will sometimes explode. Especially prone to thermal runaway are lithium-ion batteries. Reports of exploding cellphones occasionally appear in newspapers. In 2006, laptop batteries from Apple, HP, Toshiba, Lenovo, Dell and other notebook manufacturers were recalled because of fire and explosions. The Pipeline and Hazardous Materials Safety Administration (PHMSA) of the U.S. Department of Transportation recently established regulations regarding the carrying of certain types of batteries on airplanes because of their instability in certain situations. This action was partially inspired by a fire on a UPS airplane.

Current crowding

Current crowding (also **current crowding effect**, or **CCE**) is a nonhomogenous distribution of current density through a conductor or semiconductor, especially at the vicinity of the contacts and over the PN junctions.

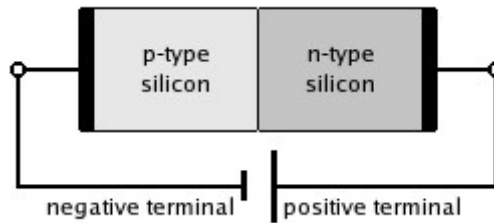
Current crowding is one of the limiting factors of efficiency of light emitting diodes. Materials with low mobility of charge carriers, eg. AlGaInP, are especially prone to current crowding phenomena. It is a dominant loss mechanisms in some LEDs, where the current densities especially around the P-side contacts reach the part of the emission characteristics with lower brightness/current efficiency.

Current crowding can lead to localized overheating and formation of thermal hot spots, in catastrophic cases leading to thermal runaway. Nonhomogenous distribution of current also aggravates electromigration effects and formation of voids. Formation of voids causes localized nonhomogeneity of current density, and the increased resistance around the void causes further localized temperature rise, which in turn accelerates the formation of the void. Conversely, localized lowering of current density may lead to deposition of the migrated atoms, leading to further lowering of current density and further deposition of material and formation of hillocks, which may cause short circuits.

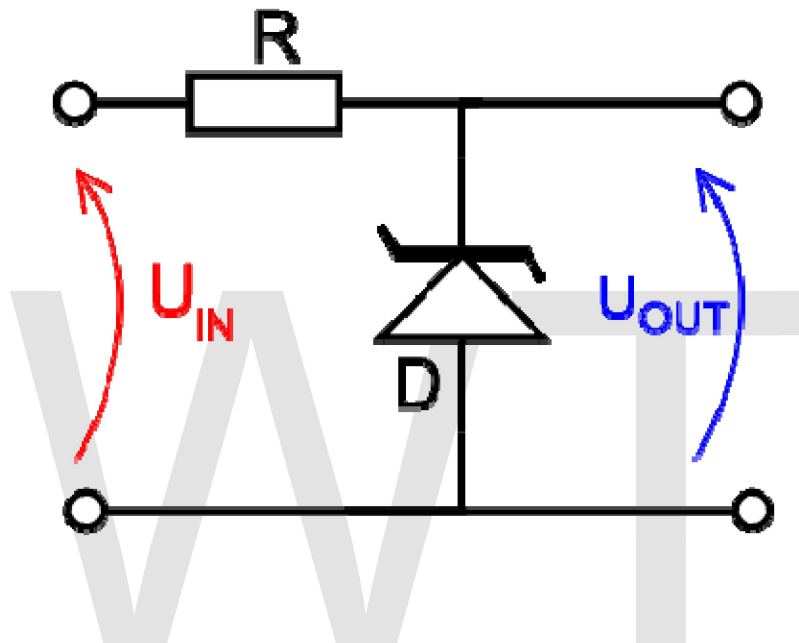
In large bipolar transistors, the resistance of the base layer influences the distribution of current density through the base region, especially at the emitter side.

Current crowding occurs especially at the areas of localized lowered resistance, or in areas where the field strength is concentrated (eg. at the edges of layers).

Reverse bias



A silicon p–n junction in reverse bias.



Reverse biased usually refers to how a diode is used in a circuit. If a diode is reverse biased, the voltage at the cathode is higher than that at the anode. Therefore, no current will flow until the diode breaks down. Connecting the *P-type* region to the *negative* terminal of the battery and the *N-type* region to the *positive* terminal, corresponds to reverse bias. The connections are illustrated in the following diagram:

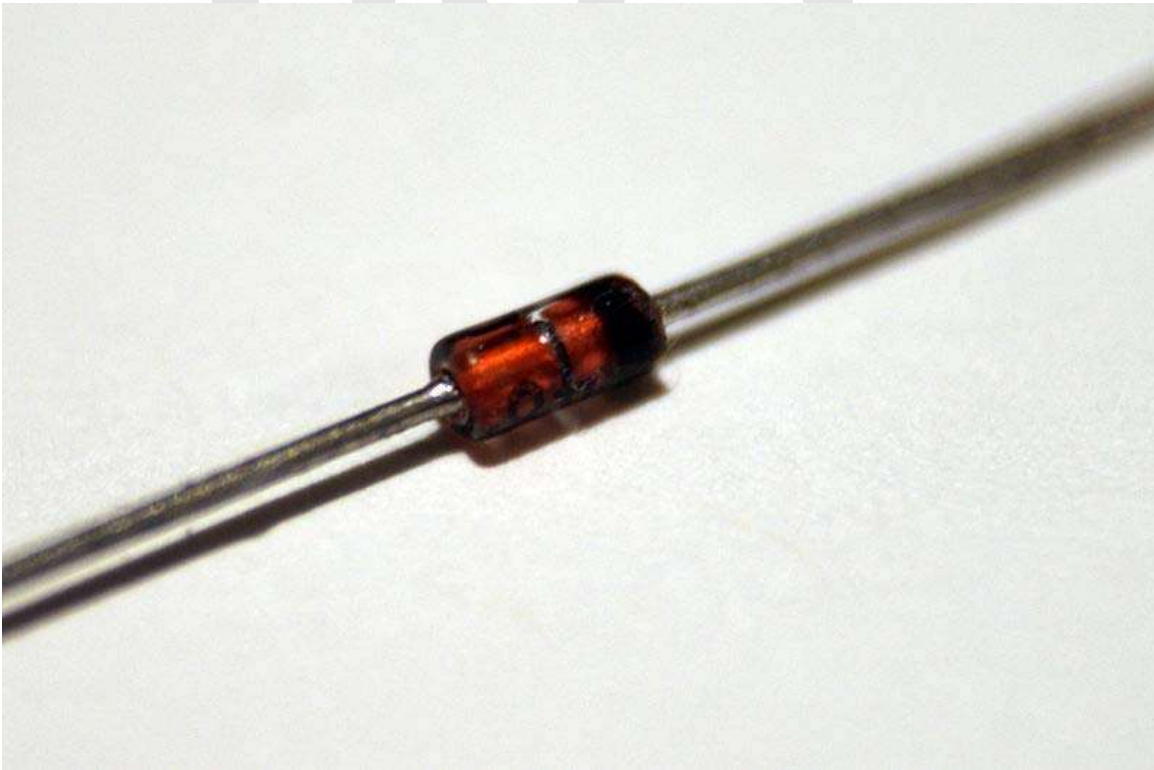
Because the p-type material is now connected to the negative terminal of the power supply, the 'holes' in the P-type material are pulled away from the junction, causing the width of the depletion zone to increase. Similarly, because the N-type region is connected to the positive terminal, the electrons will also be pulled away from the junction. Therefore the depletion region widens, and does so increasingly with increasing reverse-bias voltage. This increases the voltage barrier causing a high resistance to the flow of charge carriers thus allowing minimal electric current to cross the p–n junction. The increase in resistance of the p–n junction results in the junction to behave as an insulator. This is important for radiation detection because if current was able to flow, the charged particles would just dissipate into the material. The reverse bias ensures that charged particles are able to make it to the detector system.

The strength of the depletion zone electric field increases as the reverse-bias voltage increases. Once the electric field intensity increases beyond a critical level, the p-n junction depletion zone breaks-down and current begins to flow, usually by either the Zener or avalanche breakdown processes. Both of these breakdown processes are non-destructive and are reversible, so long as the amount of current flowing does not reach levels that cause the semiconductor material to overheat and cause thermal damage.

This effect is used to one's advantage in zener diode regulator circuits. Zener diodes have a certain - low - breakdown voltage. A standard value for breakdown voltage is for instance 5.6V. This means that the voltage at the cathode can never be more than 5.6V higher than the voltage at the anode, because the diode will break down - and therefore conduct - if the voltage gets any higher. This effectively regulates the voltage over the diode.

Another application where reverse biased diodes are used is in Varicap diodes. The width of the depletion zone of any diode changes with voltage applied. This varies the capacitance of the diode. For more information, refer to the Varicap article.

Zener diode



Zener diode



Current-voltage characteristic of a Zener diode with a breakdown voltage of 17 volt. Notice the change of voltage scale between the forward biased (positive) direction and the reverse biased (negative) direction.

A **Zener diode** is a type of diode that permits current not only in the forward direction like a normal diode, but also in the reverse direction if the voltage is larger than the breakdown voltage known as "Zener knee voltage" or "Zener voltage". The device was named after Clarence Zener, who discovered this electrical property.

A conventional solid-state diode will not allow significant current if it is reverse-biased below its reverse breakdown voltage. When the reverse bias breakdown voltage is exceeded, a conventional diode is subject to high current due to avalanche breakdown. Unless this current is limited by circuitry, the diode will be permanently damaged. In case of large forward bias (current in the direction of the arrow), the diode exhibits a voltage drop due to its junction built-in voltage and internal resistance. The amount of the voltage drop depends on the semiconductor material and the doping concentrations.

A Zener diode exhibits almost the same properties, except the device is specially designed so as to have a greatly reduced breakdown voltage, the so-called Zener voltage. By contrast with the conventional device, a reverse-biased Zener diode will exhibit a controlled breakdown and allow the current to keep the voltage across the Zener diode at the Zener voltage. For example, a diode with a Zener breakdown voltage of 3.2 V will exhibit a voltage drop of 3.2 V even if reverse bias voltage applied across it is more than its Zener voltage. The Zener diode is therefore ideal for applications such as the generation of a reference voltage (e.g. for an amplifier stage), or as a voltage stabilizer for low-current applications.

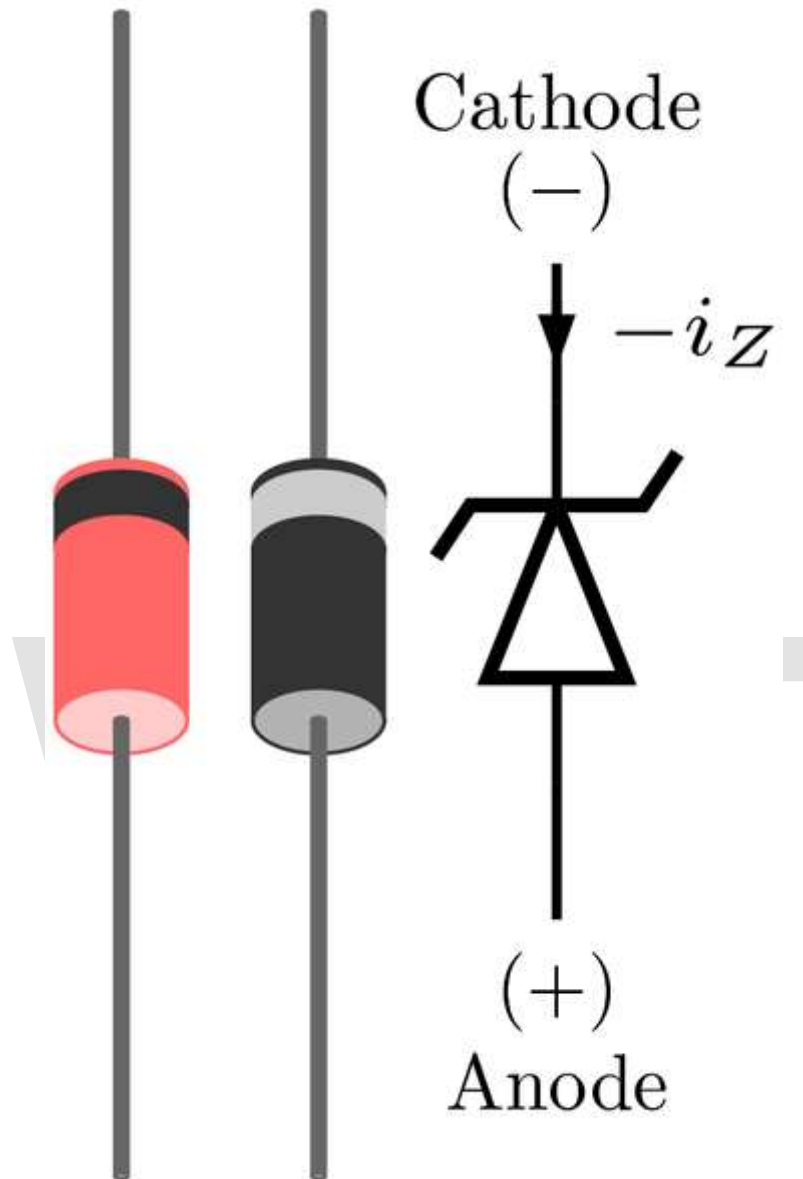
The Zener diode's operation depends on the heavy doping of its p-n junction allowing electrons to tunnel from the valence band of the p-type material to the conduction band of the n-type material. In the atomic scale, this tunneling corresponds to the transport of valence band electrons into the empty conduction band states; as a result of the reduced barrier between these bands and high electric fields that are induced due to the relatively high levels of dopings on both sides. The breakdown voltage can be controlled quite accurately in the doping process. While tolerances within 0.05% are available, the most widely used tolerances are 5% and 10%. Breakdown voltage for commonly available zener diodes can vary widely from 1.2 volts to 200 volts.

Another mechanism that produces a similar effect is the avalanche effect as in the avalanche diode. The two types of diode are in fact constructed the same way and both effects are present in diodes of this type. In silicon diodes up to about 5.6 volts, the Zener effect is the predominant effect and shows a marked negative temperature coefficient. Above 5.6 volts, the avalanche effect becomes predominant and exhibits a positive temperature coefficient. In a 5.6 V diode, the two effects occur together and their temperature coefficients neatly cancel each other out, thus the 5.6 V diode is the component of choice in temperature-critical applications. Modern manufacturing techniques have produced devices with voltages lower than 5.6 V with negligible temperature coefficients, but as higher voltage devices are encountered, the temperature coefficient rises dramatically. A 75 V diode has 10 times the coefficient of a 12 V diode.

All such diodes, regardless of breakdown voltage, are usually marketed under the umbrella term of "Zener diode".

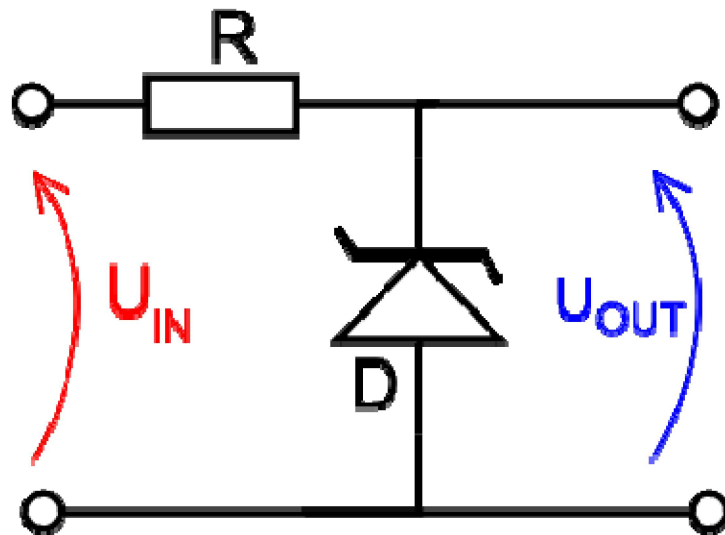
Uses





Zener diode shown with typical packages. *Reverse* current $-i_Z$ is shown.

Zener diodes are widely used as voltage references and as shunt regulators to regulate the voltage across small circuits. When connected in parallel with a variable voltage source so that it is reverse biased, a Zener diode conducts when the voltage reaches the diode's reverse breakdown voltage. From that point on, the relatively low impedance of the diode keeps the voltage across the diode at that value.



In this circuit, a typical voltage reference or regulator, an input voltage, U_{IN} , is regulated down to a stable output voltage U_{OUT} . The intrinsic voltage drop of diode D is stable over a wide current range and holds U_{OUT} relatively constant even though the input voltage may fluctuate over a fairly wide range. Because of the low impedance of the diode when operated like this, Resistor R is used to limit current through the circuit.

In the case of this simple reference, the current flowing in the diode is determined using Ohms law and the known voltage drop across the resistor R . $I_{Diode} = (U_{IN} - U_{OUT}) / R_{\Omega}$

The value of R must satisfy two conditions:

1. R must be small enough that the current through D keeps D in reverse breakdown. The value of this current is given in the data sheet for D . For example, the common BZX79C5V6 device, a 5.6 V 0.5 W Zener diode, has a recommended reverse current of 5 mA. If insufficient current exists through D , then U_{OUT} will be unregulated, and less than the nominal breakdown voltage (this differs to voltage regulator tubes where the output voltage will be higher than nominal and could rise as high as U_{IN}). When calculating R , allowance must be made for any current through the external load, not shown in this diagram, connected across U_{OUT} .
2. R must be large enough that the current through D does not destroy the device. If the current through D is I_D , its breakdown voltage V_B and its maximum power dissipation P_{MAX} , then $I_D V_B < P_{MAX}$.

A load may be placed across the diode in this reference circuit, and as long as the zener stays in reverse breakdown, the diode will provide a stable voltage source to the load.

Shunt regulators are simple, but the requirements that the ballast resistor be small enough to avoid excessive voltage drop during worst-case operation (low input voltage concurrent with high load current) tends to leave a lot of current flowing in the diode much of the time, making for a fairly wasteful regulator with high quiescent power dissipation, only suitable for smaller loads.

Zener diodes in this configuration are often used as stable references for more advanced voltage regulator circuits.

These devices are also encountered, typically in series with a base-emitter junction, in transistor stages where selective choice of a device centered around the avalanche/Zener point can be used to introduce compensating temperature co-efficient balancing of the transistor PN junction. An example of this kind of use would be a DC error amplifier used in a regulated power supply circuit feedback loop system.

Zener diodes are also used in surge protectors to limit transient voltage spikes.

Another notable application of the zener diode is the use of noise caused by its avalanche breakdown in a random number generator that never repeats.

Latchup

Latchup is a term used in the realm of integrated circuits (ICs) to describe a particular type of short circuit which can occur in an improperly designed circuit. More specifically it is the inadvertent creation of a low-impedance path between the power supply rails of a MOSFET circuit, triggering a parasitic structure which disrupts proper functioning of the part and possibly even leading to its destruction due to overcurrent. A power cycle is required to correct this situation.

The parasitic structure is usually equivalent to a thyristor (or SCR), a PNPN structure which acts as a PNP and an NPN transistor stacked next to each other. During a latchup when one of the transistors is conducting, the other one begins conducting too. They both keep each other in saturation for as long as the structure is forward-biased and some current flows through it - which usually means until a power-down. The SCR parasitic structure is formed as a part of the totem-pole PMOS and NMOS transistor pair on the output drivers of the gates.

The latchup does not have to happen between the power rails; it can happen at any place where the required parasitic structure exists. A spike of positive or negative voltage on an input or output pin of a digital chip, exceeding the rail voltage by more than a diode drop, is a common cause of latchup. Another cause is the supply voltage exceeding the absolute maximum rating, often from a transient spike in the power supply, leading to a breakdown of some internal junction. This frequently happens in circuits which use multiple supply voltages that do not come up in the proper order after a power-up,

leading to voltages on data lines exceeding the input rating of parts that have not yet reached a nominal supply voltage.

Yet another common cause of latchups is ionizing radiation which makes this a significant issue in electronic products designed for space (or very high-altitude) applications.

Latchup prevention

It is possible to design chips that are latchup-resistant, where a layer of insulating oxide (called a *trench*) surrounds both the NMOS and the PMOS transistors. This breaks the parasitic SCR structure between these transistors. Such parts are important in the cases where the proper sequencing of power and signals cannot be guaranteed (e.g., in hot swap devices).

Devices fabricated in lightly doped epitaxial layers grown on heavily doped substrates are also less susceptible to latchup. The heavily doped layer acts as a current sink where excess minority carriers can quickly recombine.

Another possibility for a latchup prevention is the *Latchup Protection Technology* circuit. When a latchup is detected, the LPT circuit shuts down the chip and holds it powered-down for a preset time.

Most silicon-on-insulator devices are inherently latchup-resistant. Latchup is the low resistance connection between tub and power supply rails.

Optoelectronics Failures

Catastrophic optical damage

Catastrophic optical damage (COD), or **catastrophic optical mirror damage (COMD)**, is a failure mode of high-power semiconductor lasers. It occurs when the semiconductor junction is overloaded by exceeding its power density and absorbs too much of the produced light energy, leading to melting and recrystallization of the semiconductor material at the facets of the laser. This is often colloquially referred to as "blowing the diode." The affected area contains a large number of lattice defects, negatively affecting its performance. If the affected area is sufficiently large, it can be observable under optical microscope as darkening of the laser facet, and/or as presence of cracks and grooves. The damage can occur within a single laser pulse, in less than a millisecond. The time to COD is inversely proportional to the power density.

Catastrophic optical damage is one of the limiting factors in increasing performance of semiconductor lasers. It is the primary failure mode for AlGaInP/AlGaAs red lasers.

Short-wavelength lasers are more susceptible to COD than long-wavelength ones.

The typical values for COD in industrial products range between 12 and 20 MW/cm².

Causes and mechanisms

At the edge of a diode laser, where light is emitted, a mirror is traditionally formed by cleaving the semiconductor wafer to form a specularly reflecting plane. This approach is facilitated by the weakness of the [110] crystallographic plane in III-V semiconductor crystals (such as GaAs, InP, GaSb, etc.) compared to other planes. A scratch made at the edge of the wafer and a slight bending force causes a nearly atomically perfect mirror-like cleavage plane to form and propagate in a straight line across the wafer.

But it so happens that the atomic states at the cleavage plane are altered (compared to their bulk properties within the crystal) by the termination of the perfectly periodic lattice

at that plane. Surface states at the cleaved plane have energy levels within the (otherwise forbidden) band gap of the semiconductor.

The absorbed light causes generation of electron-hole pairs. These can lead to breaking of chemical bonds on the crystal surface followed by oxidation, or to release of heat by nonradiative recombination. The oxidized surface then shows increased absorption of the laser light, which further accelerates its degradation. The oxidation is especially problematic for semiconductor layers containing aluminium.

Essentially, as a result when light propagates through the cleavage plane and transits to free space from within the semiconductor crystal, a fraction of the light energy is absorbed by the surface states whence it is converted to heat by phonon-electron interactions. This heats the cleaved mirror. In addition the mirror may heat simply because the edge of the diode laser—which is electrically pumped—is in less-than-perfect contact with the mount that provides a path for heat removal. The heating of the mirror causes the band gap of the semiconductor to shrink in the warmer areas. The band gap shrinkage brings more electronic band-to-band transitions into alignment with the photon energy causing yet more absorption. This is thermal runaway, a form of positive feedback, and the result can be melting of the facet, known as *catastrophic optical damage*, or COD.

Deterioration of the laser facets with aging and effects of the environment (erosion by water, oxygen, etc.) increases light absorption by the surface, and decreases the COD threshold. A sudden catastrophic failure of the laser due to COD then can occur after many thousands hours in service.

Improvements

One of the methods of increasing the COD threshold in AlGaInP laser structures is the sulfur treatment, which replaces the oxides at the laser facet with chalcogenide glasses. This decreases the recombination velocity of the surface states.

Reduction of recombination velocity of surface states can be also achieved by cleaving the crystals in ultrahigh vacuum and immediate deposition of a suitable passivation layer.

A thin layer of aluminium can be deposited over the surface, for gettering the oxygen.

Another approach is doping of the surface, increasing the band gap and decreasing absorption of the lasing wavelength, shifting the absorption maximum several nanometers up.

Current crowding near the mirror area can be avoided by prevention of injecting charge carriers near the mirror region. This is achieved by depositing the electrodes away from the mirror, at least several carrier diffusion distances.

Energy density on the surface can be reduced by employing a waveguide broadening the optical cavity, so the same amount of energy exits through a larger area. Energy density of 15–20 MW/cm² corresponding to 100 mW per micrometer of stripe width are now achievable. A wider laser stripe can be used for higher output power, for the cost of transverse mode oscillations and therefore worsening of spectral and spatial beam quality.

In the 1970s, this problem, which is particularly nettlesome for GaAs-based lasers emitting between 1 μm and 0.630 μm wavelengths (less so for InP based lasers used for long-haul telecommunications which emit between 1.3 μm and 2 μm), was identified. Michael Ettenberg, a researcher and later Vice President at RCA Laboratories' David Sarnoff Research Center in Princeton, New Jersey, devised a solution. A thin layer of aluminum oxide was deposited on the facet. If the aluminum oxide thickness is chosen correctly, it functions as an anti-reflective coating, reducing reflection at the surface. This alleviated the heating and COD at the facet.

Since then, various other refinements have been employed. One approach is to create a so-called non-absorbing mirror (NAM) such that the final 10 μm or so before the light emits from the cleaved facet are rendered non-absorbing at the wavelength of interest. Such lasers are called **window lasers**.

In the very early 1990s, SDL, Inc. began supplying high power diode lasers with good reliability characteristics. CEO Donald Scifres and CTO David Welch presented new reliability performance data at, e.g., SPIE Photonics West conferences of the era. The methods used by SDL to defeat COD were considered to be highly proprietary and have still not been disclosed publicly as of June, 2006.

In the mid-1990s IBM Research (Ruschlikon, Switzerland) announced that it had devised its so-called "E2 process" which conferred extraordinary resistance to COD in GaAs-based lasers. This process, too, has never been disclosed as of June, 2006.

Hydrogen darkening

Hydrogen darkening is a physical degradation of the optical properties of glass. Free hydrogen atoms are able to bind to the SiO₂ silica glass compound forming hydroxyl (OH) - a chemical compound that interferes with the passage of light through the glass.

The problem is particularly relevant to fiber optic cables — particularly in oil and gas wells where fiber optic cables are used for distributed temperature sensing. Hydrogen can be present due to the cracking of hydrocarbons in the well. The darkening of the fiber can distort the DTS reading and possibly render the DTS system inoperable due to the optical loss budget being exceeded.

To prevent this, coatings such as carbon are applied to the fiber, and hydrogen capturing gels are used to buffer the fiber, and other proprietary techniques may be used to prevent hydrogen atoms from reaching the glass fiber via the cable sheath.

Microelectromechanical Systems Failures

Microelectromechanical systems suffer from specific types of failures.

- Stiction causes the moving parts stick to other surfaces on contact. An external impulse can sometimes release the adhesion and restore functionality. Non-stick coatings, reduction of surface contact area, and increased awareness virtually eliminated the problem in contemporary systems.
- Particle contamination can cause failures by particles lodged between the elements and blocking their movements. Conductive particles may short out circuits, namely electrostatic actuators.
- Wear damages the surfaces; debris, removed from the surfaces during their mutual movements, can be a source of particle contamination.
- Fractures may cause loss of mechanical parts
- Material fatigue may induce cracks in moving structures
- Electrostatic discharge
- Oxide charging, causing electrostatic attraction between parts, may lead to electrostatically mediated stiction
- Dielectric breakdown, causing a short circuit and irreversibly damaging the MEMS structure

Fatigue (material)

In materials science, **fatigue** is the progressive and localized structural damage that occurs when a material is subjected to cyclic loading. The nominal maximum stress values are less than the ultimate tensile stress limit, and may be below the yield stress limit of the material.

Fatigue occurs when a material is subjected to repeated loading and unloading. If the loads are above a certain threshold, microscopic cracks will begin to form at the surface. Eventually a crack will reach a critical size, and the structure will suddenly fracture. The shape of the structure will significantly affect the fatigue life; square holes or sharp corners will lead to elevated local stresses where fatigue cracks can initiate. Round holes

and smooth transitions or fillets are therefore important to increase the fatigue strength of the structure.

Fatigue life

ASTM defines *fatigue life*, N_f , as the number of stress cycles of a specified character that a specimen sustains before failure of a specified nature occurs.

One method to predict fatigue life of materials is the Uniform Material Law (UML). UML was developed for fatigue life prediction of aluminum and titanium alloys by the end of 20th century and extended to high-strength steels.

Characteristics of fatigue



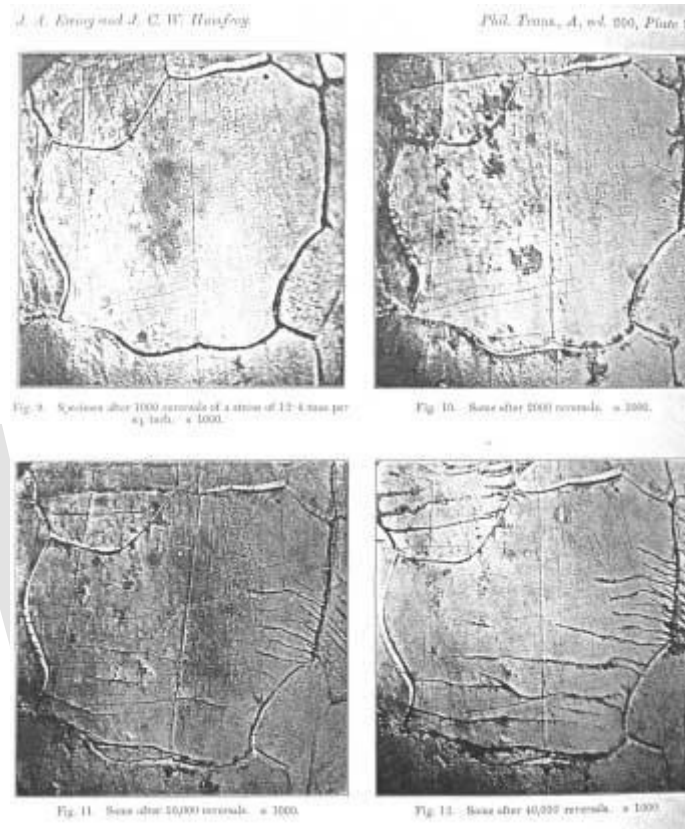
Fracture of an aluminium crank arm. Dark area of striations: slow crack growth. Bright granular area: sudden fracture.

- In metals and alloys, the process starts with dislocation movements, eventually forming persistent slip bands that nucleate short cracks.
- Fatigue is a stochastic process, often showing considerable scatter even in controlled environments.
- The greater the applied stress range, the shorter the life.
- Fatigue life scatter tends to increase for longer fatigue lives.
- Damage is cumulative. Materials do not recover when rested.
- Fatigue life is influenced by a variety of factors, such as temperature, surface finish, microstructure, presence of oxidizing or inert chemicals, residual stresses, contact (fretting), etc.
- Some materials (e.g., some steel and titanium alloys) exhibit a theoretical fatigue limit below which continued loading does not lead to structural failure.
- In recent years, researchers (see, for example, the work of Bathias, Murakami, and Stanzl-Tschegg) have found that failures occur below the theoretical fatigue limit at very high fatigue lives (10^9 to 10^{10} cycles). An ultrasonic resonance technique is used in these experiments with frequencies around 10–20 kHz.
- High cycle fatigue strength (about 10^3 to 10^8 cycles) can be described by stress-based parameters. A load-controlled servo-hydraulic test rig is commonly used in these tests, with frequencies of around 20–50 Hz. Other sorts of machines—like resonant magnetic machines—can also be used, achieving frequencies up to 250 Hz.
- Low cycle fatigue (typically less than 10^3 cycles) is associated with widespread plasticity in metals; thus, a strain-based parameter should be used for fatigue life prediction in metals and alloys. Testing is conducted with constant strain amplitudes typically at 0.01–5 Hz.

Timeline of early fatigue research history

- 1837: Wilhelm Albert publishes the first article on fatigue. He devised a test machine for conveyor chains used in the Clausthal mines.
- 1839: Jean-Victor Poncelet describes metals as being *tired* in his lectures at the military school at Metz.
- 1842: William John Macquorn Rankine recognises the importance of stress concentrations in his investigation of railroad axle failures. The Versailles train crash was caused by axle fatigue.
- 1843: Joseph Glynn reports on fatigue of axle on locomotive tender. He identifies the keyway as the crack origin.
- 1848: Railway Inspectorate report one of the first tyre failures, probably from a rivet hole in tread of railway carriage wheel. It was likely a fatigue failure.
- 1849: Eaton Hodgkinson is granted a *small sum of money* to report to the UK Parliament on his work in *ascertaining by direct experiment, the effects of continued changes of load upon iron structures and to what extent they could be loaded without danger to their ultimate security*.
- 1854: Braithwaite reports on common service fatigue failures and coins the term *fatigue*.

- 1860: Systematic fatigue testing undertaken by Sir William Fairbairn and August Wöhler.
- 1870: Wöhler summarises his work on railroad axles. He concludes that cyclic stress range is more important than peak stress and introduces the concept of *endurance limit*.



Micrographs showing how surface fatigue cracks grow as material is further cycled.
From Ewing & Humfrey (1903)

- 1903: Sir James Alfred Ewing demonstrates the origin of fatigue failure in microscopic cracks.
- 1910: O. H. Basquin proposes a log-log relationship for SN curves, using Wöhler's test data.
- 1945: A. M. Miner popularises A. Palmgren's (1924) linear damage hypothesis as a practical design tool.
- 1954: L. F. Coffin and S. S. Manson explain fatigue crack-growth in terms of plastic strain in the tip of cracks.
- 1961: P. C. Paris proposes methods for predicting the rate of growth of individual fatigue cracks in the face of initial scepticism and popular defence of Miner's phenomenological approach.
- 1968: Tatsuo Endo and M. Matsuishi devise the rainflow-counting algorithm and enable the reliable application of Miner's rule to random loadings.

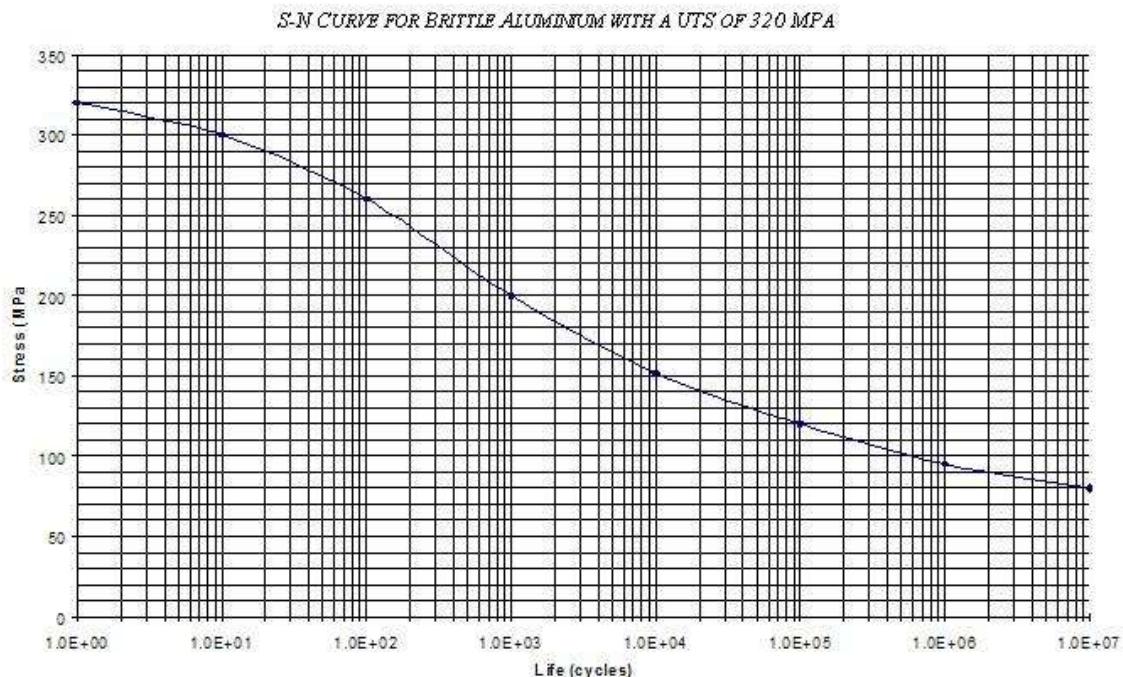
- 1970: W. Elber elucidates the mechanisms and importance of crack closure in slowing the growth of a fatigue crack due to the wedging effect of plastic deformation left behind the tip of the crack.

High-cycle fatigue

Historically, most attention has focused on situations that require more than 10^4 cycles to failure where stress is low and deformation primarily elastic.

The S-N curve

In high-cycle fatigue situations, materials performance is commonly characterised by an *S-N curve*, also known as a *Wöhler curve*. This is a graph of the magnitude of a cyclic stress (S) against the logarithmic scale of cycles to failure (N).



S-N curves are derived from tests on samples of the material to be characterised (often called *coupons*) where a regular sinusoidal stress is applied by a testing machine which also counts the number of cycles to failure. This process is sometimes known as *coupon testing*. Each coupon test generates a point on the plot though in some cases there is a *runout* where the time to failure exceeds that available for the test. Analysis of fatigue data requires techniques from statistics, especially survival analysis and linear regression.

Probabilistic nature of fatigue

As coupons sampled from a homogeneous frame will manifest variation in their number of cycles to failure, the S-N curve should more properly be an *S-N-P curve* capturing the probability of failure after a given number of cycles of a certain stress. Probability distributions that are common in data analysis and in design against fatigue include the lognormal distribution, extreme value distribution, Birnbaum–Saunders distribution, and Weibull distribution.

Complex loadings



Spectrum loading

In practice, a mechanical part is exposed to a complex, often random, sequence of loads, large and small. In order to assess the safe life of such a part:

1. Reduce the complex loading to a series of simple cyclic loadings using a technique such as rainflow analysis;
2. Create a histogram of cyclic stress from the rainflow analysis to form a fatigue damage spectrum;
3. For each stress level, calculate the degree of cumulative damage incurred from the S-N curve; and
4. Combine the individual contributions using an algorithm such as *Miner's rule*.

Miner's rule

In 1945, M. A. Miner popularised a rule that had first been proposed by A. Palmgren in 1924. The rule, variously called *Miner's rule* or the *Palmgren-Miner linear damage hypothesis*, states that where there are k different stress magnitudes in a spectrum, S_i ($1 \leq i \leq k$), each contributing $n_i(S_i)$ cycles, then if $N_i(S_i)$ is the number of cycles to failure of a constant stress reversal S_i , failure occurs when:

$$\sum_{i=1}^k \frac{n_i}{N_i} = C$$

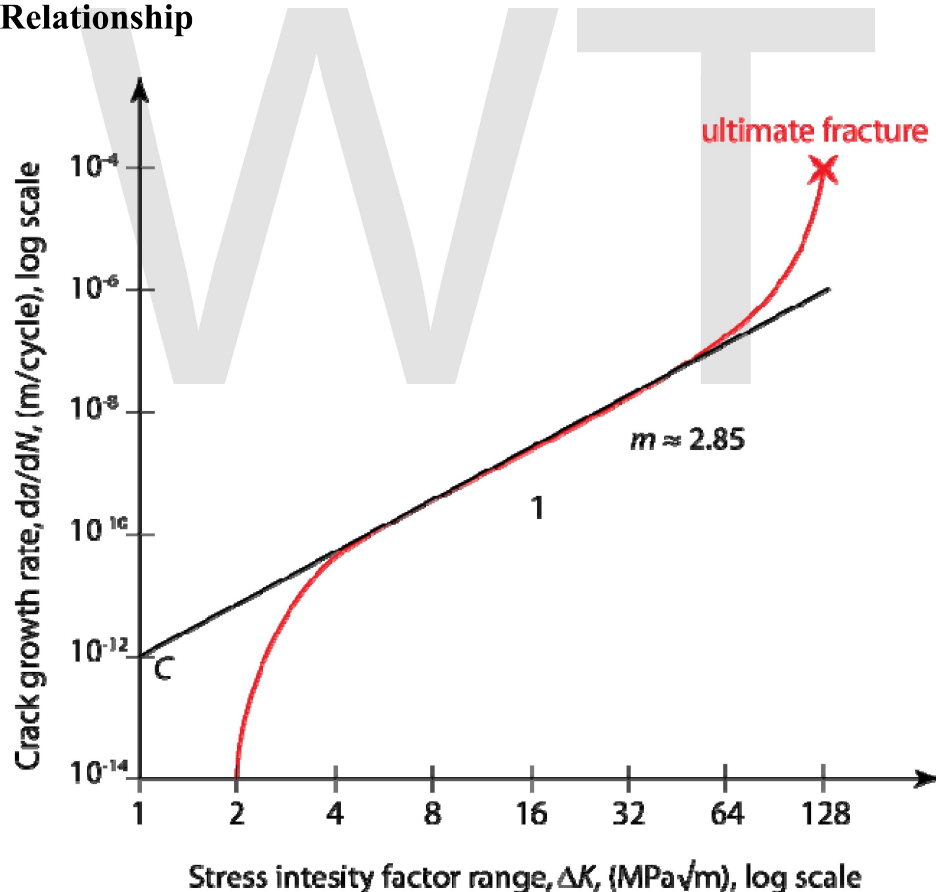
C is experimentally found to be between 0.7 and 2.2. Usually for design purposes, C is assumed to be 1.

This can be thought of as assessing what proportion of life is consumed by stress reversal at each magnitude then forming a linear combination of their aggregate.

Though Miner's rule is a useful approximation in many circumstances, it has several major limitations:

1. It fails to recognise the probabilistic nature of fatigue and there is no simple way to relate life predicted by the rule with the characteristics of a probability distribution. Industry analysts often use design curves, adjusted to account for scatter, to calculate $N_i(S_i)$.
2. There is sometimes an effect in the order in which the reversals occur. In some circumstances, cycles of low stress followed by high stress cause more damage than would be predicted by the rule. It does not consider the effect of overload or high stress which may result in a compressive residual stress. High stress followed by low stress may have less damage due to the presence of compressive residual stress.

Paris' Relationship



Typical fatigue crack growth rate graph.

In Fracture mechanics, Anderson, Gomez and Paris derived relationships for the stage II crack growth with cycles N , in terms of the cyclical component ΔK of the Stress Intensity Factor K

$$\frac{da}{dN} = C(\Delta K)^m$$

where a is the crack length and m is typically in the range 3 to 5 (for metals).

This relationship was later modified (by Forman, 1967) to make better allowance for the mean stress, by introducing a factor depending on $(1-R)$ where R = min stress/max stress, in the denominator.

Low-cycle fatigue

Where the stress is high enough for plastic deformation to occur, the account in terms of stress is less useful and the strain in the material offers a simpler description. Low-cycle fatigue is usually characterised by the *Coffin-Manson relation* (published independently by L. F. Coffin in 1954 and S. S. Manson 1953):

$$\frac{\Delta \epsilon_p}{2} = \epsilon'_f (2N)^c$$

-where:

- $\Delta \epsilon_p / 2$ is the plastic strain amplitude;
- ϵ'_f is an empirical constant known as the *fatigue ductility coefficient*, the failure strain for a single reversal;
- $2N$ is the number of reversals to failure (N cycles);
- c is an empirical constant known as the *fatigue ductility exponent*, commonly ranging from -0.5 to -0.7 for metals in time independent fatigue. Slopes can be considerably steeper in the presence of creep or environmental interactions.

A similar relationship for materials such as Zirconium, used in the nuclear industry.

Fatigue and fracture mechanics

The account above is purely empirical and, though it allows life prediction and design assurance, life improvement or design optimisation can be enhanced using fracture mechanics. It can be developed in four stages.

1. Crack nucleation;
2. Stage I crack-growth;
3. Stage II crack-growth; and
4. Ultimate ductile failure.

Factors that affect fatigue-life

- **Cyclic stress state:** Depending on the complexity of the geometry and the loading, one or more properties of the stress state need to be considered, such as stress amplitude, mean stress, biaxiality, in-phase or out-of-phase shear stress, and load sequence,
- **Geometry:** Notches and variation in cross section throughout a part lead to stress concentrations where fatigue cracks initiate.
- **Surface quality.** Surface roughness cause microscopic stress concentrations that lower the fatigue strength. Compressive residual stresses can be introduced in the surface by e.g. shot peening to increase fatigue life. Such techniques for producing surface stress are often referred to as *peening*, whatever the mechanism used to produce the stress. Low Plasticity Burnishing, Laser peening, and ultrasonic impact treatment can also produce this surface compressive stress and can increase the fatigue life of the component. This improvement is normally observed only for high-cycle fatigue.
- **Material Type:** Fatigue life, as well as the behavior during cyclic loading, varies widely for different materials, e.g. composites and polymers differ markedly from metals.
- **Residual stresses:** Welding, cutting, casting, and other manufacturing processes involving heat or deformation can produce high levels of tensile residual stress, which decreases the fatigue strength.
- **Size and distribution of internal defects:** Casting defects such as gas porosity, non-metallic inclusions and shrinkage voids can significantly reduce fatigue strength.
- **Direction of loading:** For non-isotropic materials, fatigue strength depends on the **direction of the principal stress**.
- **Grain size:** For most metals, smaller grains yield longer fatigue lives, however, the presence of surface defects or scratches will have a greater influence than in a coarse grained alloy.
- **Environment:** Environmental conditions can cause erosion, corrosion, or gas-phase embrittlement, which all affect fatigue life. Corrosion fatigue is a problem encountered in many aggressive environments.
- **Temperature:** Extreme high or low temperatures can decrease fatigue strength.

Design against fatigue

Dependable design against fatigue-failure requires thorough education and supervised experience in structural engineering, mechanical engineering, or materials science. There are three principal approaches to life assurance for mechanical parts that display increasing degrees of sophistication:

1. Design to keep stress below threshold of fatigue limit (infinite lifetime concept);
2. Design (conservatively) for a fixed life after which the user is instructed to replace the part with a new one (a so-called *lified* part, finite lifetime concept, or "safe-life" design practice);
3. Instruct the user to inspect the part periodically for cracks and to replace the part once a crack exceeds a critical length. This approach usually uses the technologies of nondestructive testing and requires an accurate prediction of the rate of crack-growth between inspections. This is often referred to as damage tolerant design or "retirement-for-cause".

Stopping fatigue

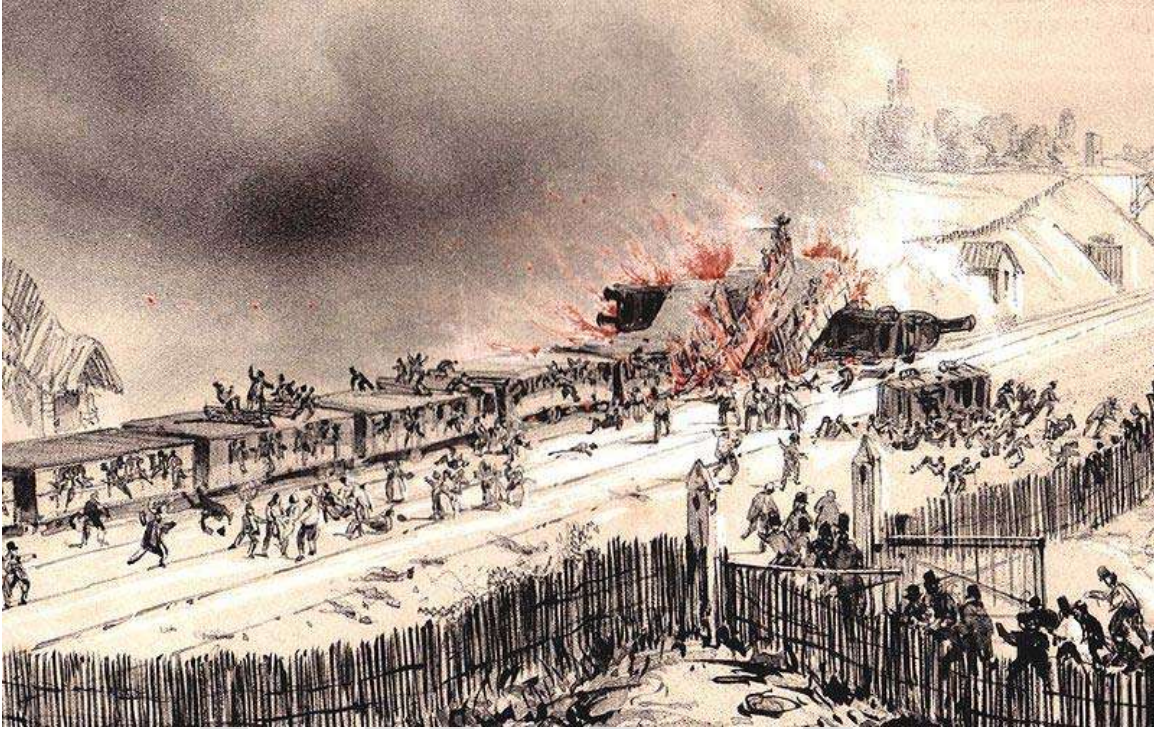
Fatigue cracks that have begun to propagate can sometimes be stopped by drilling holes, called *drill stops*, in the path of the fatigue crack. This is not recommended as a general practice because the hole represents a stress concentration factor which depends on the size of the hole and geometry. There is thus the possibility of a new crack starting in the side of the hole. It is always far better to replace the cracked part entirely.

Material change

Changes in the materials used in parts can also improve fatigue life. For example, parts can be made from better fatigue rated metals. Complete replacement and redesign of parts can also reduce if not eliminate fatigue problems. Thus helicopter rotor blades and propellers in metal are being replaced by composite equivalents. They are not only lighter, but also much more resistant to fatigue. They are more expensive, but the extra cost is amply repaid by their greater integrity, since loss of a rotor blade usually leads to total loss of the aircraft. A similar argument has been made for replacement of metal fuselages, wings and tails of aircraft.

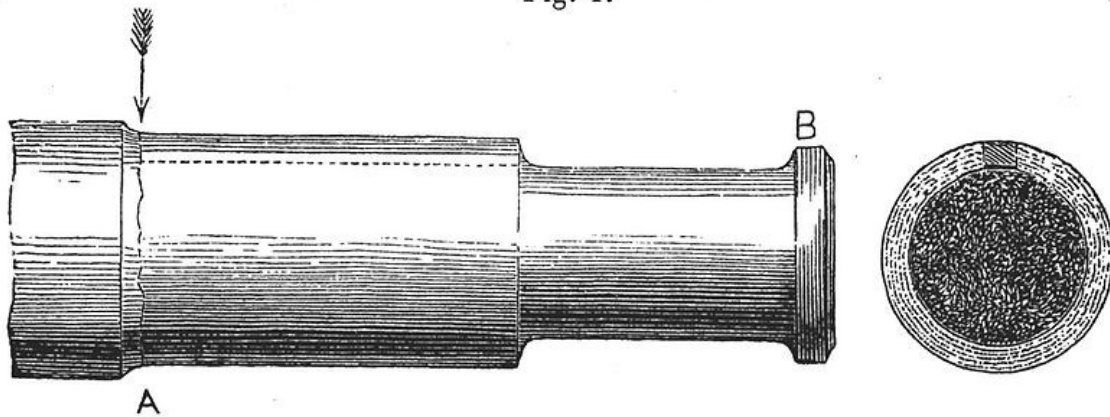
Infamous fatigue failures

Versailles train crash



Versailles train disaster

Fig. 1.

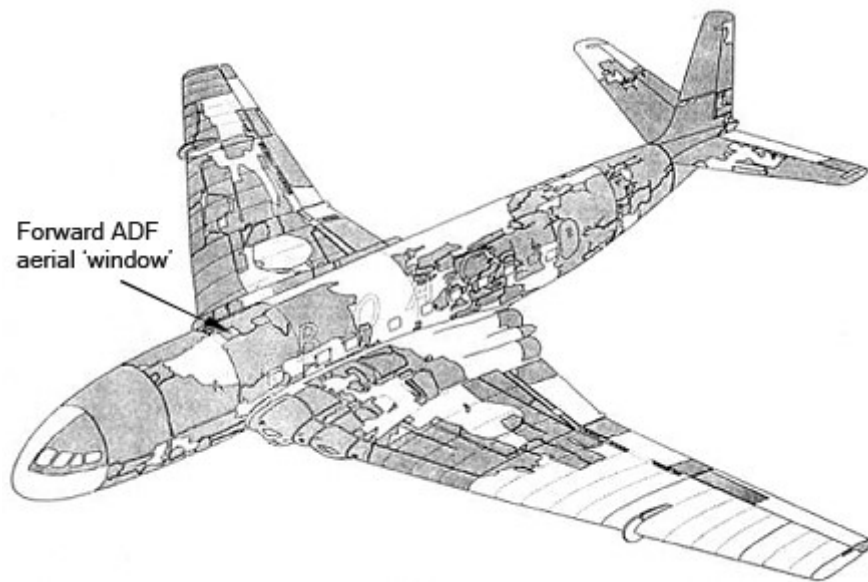


Drawing of a fatigue failure in an axle by Joseph Glynn, 1843.

Following the King's fete celebrations at the Palace of Versailles, a train returning to Paris crashed in May 1842 at Meudon after the leading locomotive broke an axle. The carriages behind piled into the wrecked engines and caught fire. At least 55 passengers were killed trapped in the carriages, including the explorer Jules Dumont d'Urville. This accident is known in France as the "Catastrophe ferroviaire de Meudon". The accident was witnessed by the British locomotive engineer Joseph Locke and widely reported in Britain. It was discussed extensively by engineers, who sought an explanation.

The derailment had been the result of a broken locomotive axle. Rankine's investigation of broken axles in Britain highlighted the importance of stress concentration, and the mechanism of crack growth with repeated loading. His and other papers suggesting a crack growth mechanism through repeated stressing, however, were ignored, and fatigue failures occurred at an ever increasing rate on the expanding railway system. Other spurious theories seemed to be more acceptable, such as the idea that the metal had somehow "crystallized". The notion was based on the crystalline appearance of the fast fracture region of the crack surface, but ignored the fact that the metal was already highly crystalline.

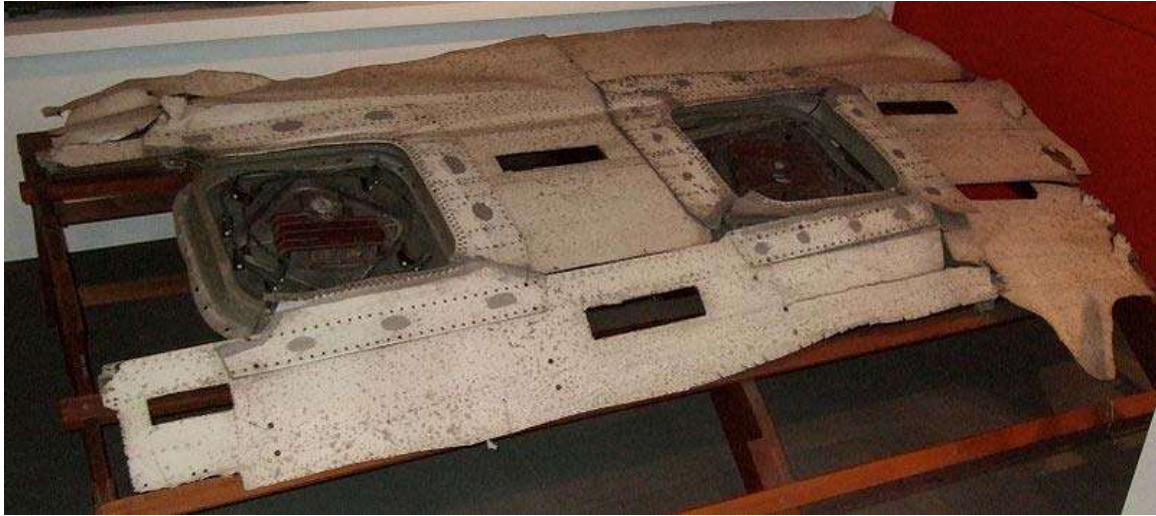
de Havilland Comet



The recovered (shaded) parts of the wreckage of *G-ALYP* and the site (arrowed) of the failure

Two de Havilland Comet passenger jets broke up in mid-air and crashed within a few months of each other in 1954. As a result systematic tests were conducted on a fuselage immersed and pressurised in a water tank. After the equivalent of 3,000 flights investigators at the Royal Aircraft Establishment (RAE) were able to conclude that the crash had been due to failure of the pressure cabin at the forward Automatic Direction Finder window in the roof. This 'window' was in fact one of two apertures for the aerials of an electronic navigation system in which opaque fibreglass panels took the place of the window 'glass'. The failure was a result of metal fatigue caused by the repeated pressurisation and de-pressurisation of the aircraft cabin. Another fact was that the supports around the windows were riveted, not bonded, as the original specifications for the aircraft had called for. The problem was exacerbated by the punch rivet construction technique employed. Unlike drill riveting, the imperfect nature of the hole created by

punch riveting caused manufacturing defect cracks which may have caused the start of fatigue cracks around the rivet.

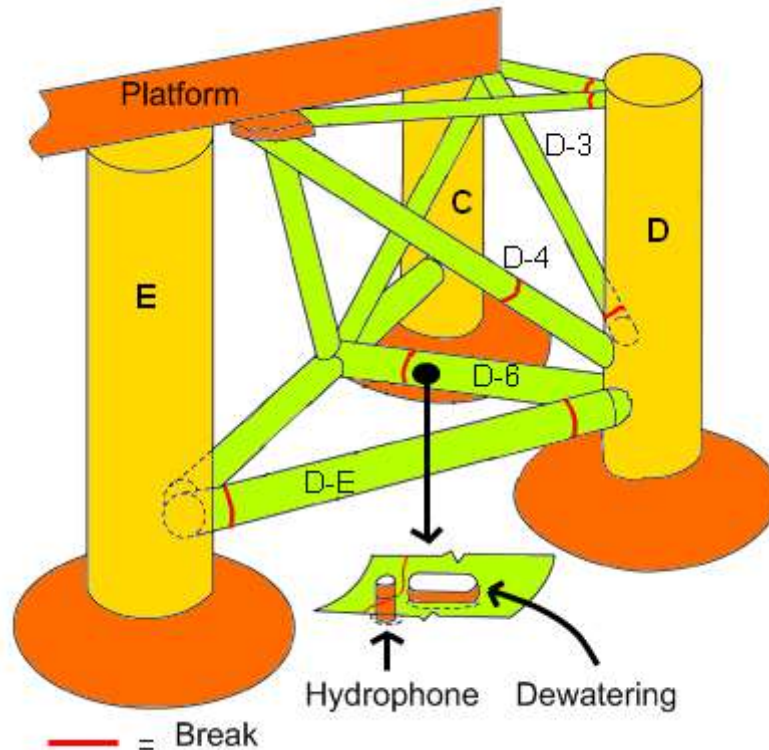


The fuselage fragment of *G-ALYP* on display in the Science Museum in London

The Comet's pressure cabin had been designed to a safety factor comfortably in excess of that required by British Civil Airworthiness Requirements (2.5 times the cabin proof pressure as opposed to the requirement of 1.33 times and an ultimate load of 2.0 times the cabin pressure) and the accident caused a revision in the estimates of the safe loading strength requirements of airliner pressure cabins.

In addition, it was discovered that the stresses around pressure cabin apertures were considerably higher than had been anticipated, especially around sharp-cornered cut-outs, such as windows. As a result, all future jet airliners would feature windows with rounded corners, the curve eliminating a stress concentration. This was a noticeable distinguishing feature of all later models of the Comet. Investigators from the RAE told a public inquiry that the sharp corners near the Comets' window openings acted as initiation sites for cracks. The skin of the aircraft was also too thin, and cracks from manufacturing stresses were present at the corners.

Alexander L. Kielland oil platform capsizes



Fractures on the right side of the Alexander L. Kielland rig

The *Alexander L. Kielland* was a Norwegian semi-submersible drilling rig that capsized whilst working in the Ekofisk oil field in March 1980 killing 123 people. The capsizing was the worst disaster in Norwegian waters since World War II. The rig, located approximately 320 km east from Dundee, Scotland, was owned by the Stavanger Drilling Company of Norway and was on hire to the U.S. company Phillips Petroleum at the time of the disaster. In driving rain and mist, early in the evening of 27 March 1980 more than 200 men were off duty in the accommodation on the *Alexander L. Kielland*. The wind was gusting to 40 knots with waves up to 12 m high. The rig had just been winched away from the *Edda* production platform. Minutes before 18:30 those on board felt a 'sharp crack' followed by 'some kind of trembling'. Suddenly the rig heeled over 30° and then stabilised. Five of the six anchor cables had broken, the one remaining cable preventing the rig from capsizing. The list continued to increase and at 18.53 the remaining anchor cable snapped and the rig turned upside down.

A year later in March 1981, the investigative report concluded that the rig collapsed owing to a fatigue crack in one of its six bracings (bracing D-6), which connected the collapsed D-leg to the rest of the rig. This was traced to a small 6 mm fillet weld which joined a non-load-bearing flange plate to this D-6 bracing. This flange plate held a sonar device used during drilling operations. The poor profile of the fillet weld contributed to a reduction in its fatigue strength. Further, the investigation found considerable amounts of lamellar tearing in the flange plate and cold cracks in the butt weld. Cold cracks in the welds, increased stress concentrations due to the weakened flange plate, the poor weld

profile, and cyclical stresses (which would be common in the North Sea), seemed to collectively play a role in the rig's collapse.

W T

Vacuum Tubes Failures

Multipactor effect

The **multipactor effect** is a phenomenon in radio frequency (RF) amplifier vacuum tubes and waveguides, where, under certain conditions, secondary electron emission in resonance with an alternating electric field leads to exponential electron multiplication, possibly damaging and even destroying the RF device.

Description

The multipactor effect occurs when electrons accelerated by radio-frequency (RF) fields are self-sustained in a vacuum (or near vacuum) via an electron avalanche caused by secondary electron emission. The impact of an electron to a surface can, depending on its energy and angle, release one or more secondary electrons into the vacuum. These electrons can then be accelerated by the RF fields and impact with the same or another surface. Should the impact energies, number of electrons released and timing of the impacts be such that a sustained multiplication of the number of electrons occurs, the phenomenon can grow exponentially and may lead to operational problems of the RF system such as damage of RF components or loss/distortion of the RF signal.

Mechanism

The mechanism of multipactor depends on the orientation of an RF electric field with respect to the surface. There are two types of multipactor: two-surface multipactor on metals and single-surface multipactor on dielectrics.

Two-surface multipactor on metals

This is a multipactor effect that occurs in the gap between metallic electrodes. Often, an RF electric field is normal to the surface. A resonance between electron flight time and rf field cycle is a mechanism for multipactor development.

The existence of multipactor is dependent on the following three conditions being met: The average number of electrons released is greater than or equal to one per incident electron (this is dependent on the secondary electron yield of the surface) **and** the time taken by the electron to travel from the surface from which it was released to the surface it impacts with is an integer multiple of one half of the RF period **and** the average secondary electron yield is greater than or equal to one.

Single-surface multipactor on dielectrics

This is a multipactor effect that occurs on a dielectric surface. Often, an RF electric field is parallel to the surface. The positive charge accumulated on the dielectric surface returns electrons back to the surface. A single surface multipactor event is also possible on a metallic surface in the presence of a crossed static magnetic field.

Frequency-Gap Product in Two Surface Multipactor

The conditions under which multipactor will occur in two surface multipactor can be described by a quantity called the frequency-gap product. Consider a two surface setup with the following definitions:

- d , distance or gap between the surfaces
- ω , angular frequency of the RF field
- V_0 , peak plate to plate RF voltage
- E_0 , peak electric field between the surfaces, equal to V_0/d .

The RF voltage varies sinusoidally. Consider the time at which the voltage at electrode A passes through 0 and starts to become negative. Assuming that there is at least 1 free electron near A, that electron will begin to accelerate to the right toward electrode B. It will continue to accelerate and reach a maximum velocity $\frac{1}{2}$ of a cycle later just as the voltage at electrode B begins to become negative. If the electron(s) from electrode A strike electrode B at this time and produce additional free electrons, these new free electrons will begin to accelerate toward electrode A. The process may then repeat causing multipactor. We now find the relationship between the plate spacing, RF frequency, and RF voltage that causes the strongest multipactor resonance.

Consider a point in time at which electrons have just collided with electrode A at position $d/2$. The electric field is at zero and is beginning to point to the left so that the newly freed electrons are accelerated toward the right. Newton's equation of motion of the free electrons is

$$a(t) = \frac{F(t)}{m}$$

$$\ddot{x}(t) = \frac{qE_0}{m} \sin(\omega t)$$

The solution to this differential equation is

$$x(t) = -\frac{qE_0}{m\omega^2} \sin(\omega t) + \frac{qE_0}{m\omega} t - \frac{d}{2}$$

where we've assumed that when the electrons initially leave the electrode they have zero velocity. We know that resonance happens if the electrons arrive at the rightmost

electrode after one half of the period of the RF field, $t_{\frac{1}{2}} = \frac{\pi}{\omega}$. Plugging this into our solution for $x(t)$ we get

$$\begin{aligned} x(t_{\frac{1}{2}}) &= -\frac{qE_0}{m\omega^2} \sin(\omega t_{\frac{1}{2}}) + \frac{qE_0}{m\omega} t_{\frac{1}{2}} - \frac{d}{2} \\ \frac{d}{2} &= -\frac{qE_0}{m\omega^2} \sin(\omega \frac{\pi}{\omega}) + \frac{qE_0}{m\omega} \frac{\pi}{\omega} - \frac{d}{2} \end{aligned}$$

Rearranging and using the frequency f instead of the angular frequency gives

$$fd = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{qV_0}{m}}$$

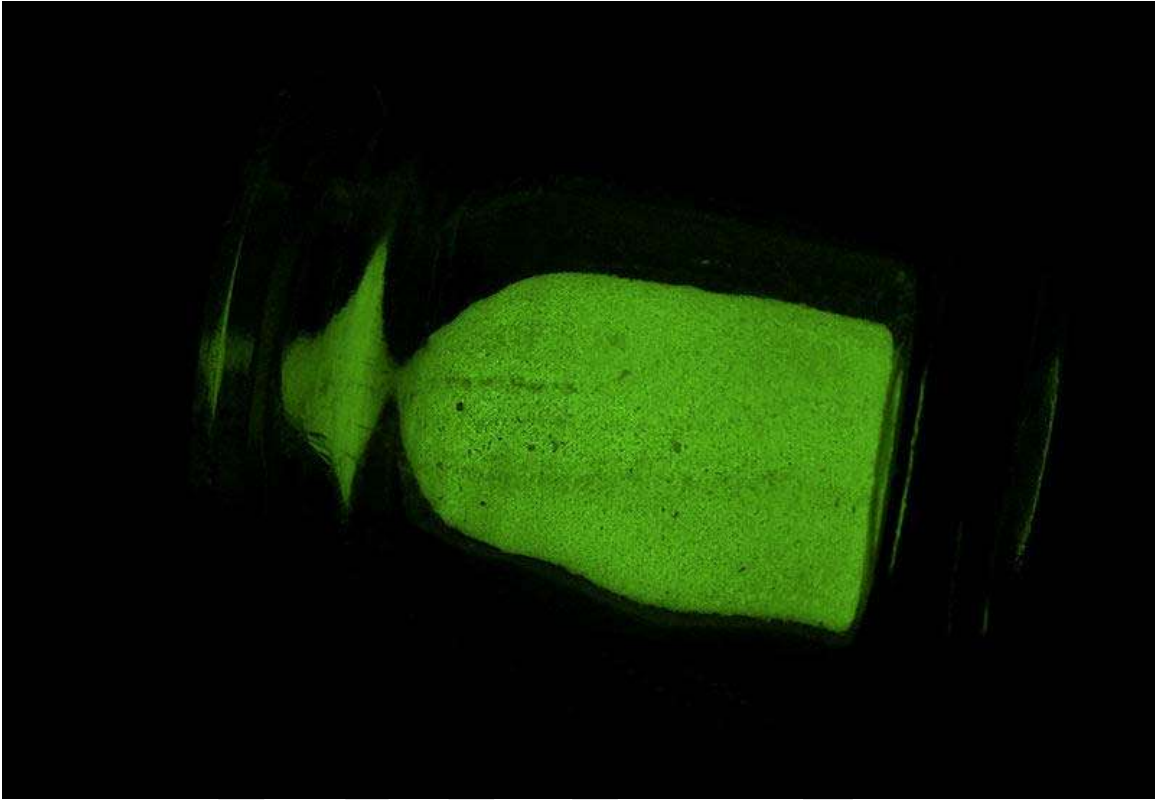
The product fd is called the frequency-gap product. Keep in mind that this equation is a criterion for greatest amount of resonance, but multipactor can still occur when this equation is not satisfied.

History

This phenomenon was first observed by the French physicist Camille Gutton, in 1924, at Nancy.

Multipactor was identified and studied in 1934 by Philo T. Farnsworth, the inventor of electronic television, who attempted to take advantage of it as an amplifier. More commonly nowadays, it has become an obstacle to be avoided for normal operation of particle accelerators, vacuum electronics, radars, satellite communication devices, and so forth. A novel form of multipactor has been proposed (Kishek, 1998), and subsequently experimentally observed, in which charging of a dielectric surface considerably changes the dynamics of the multipactor discharge.

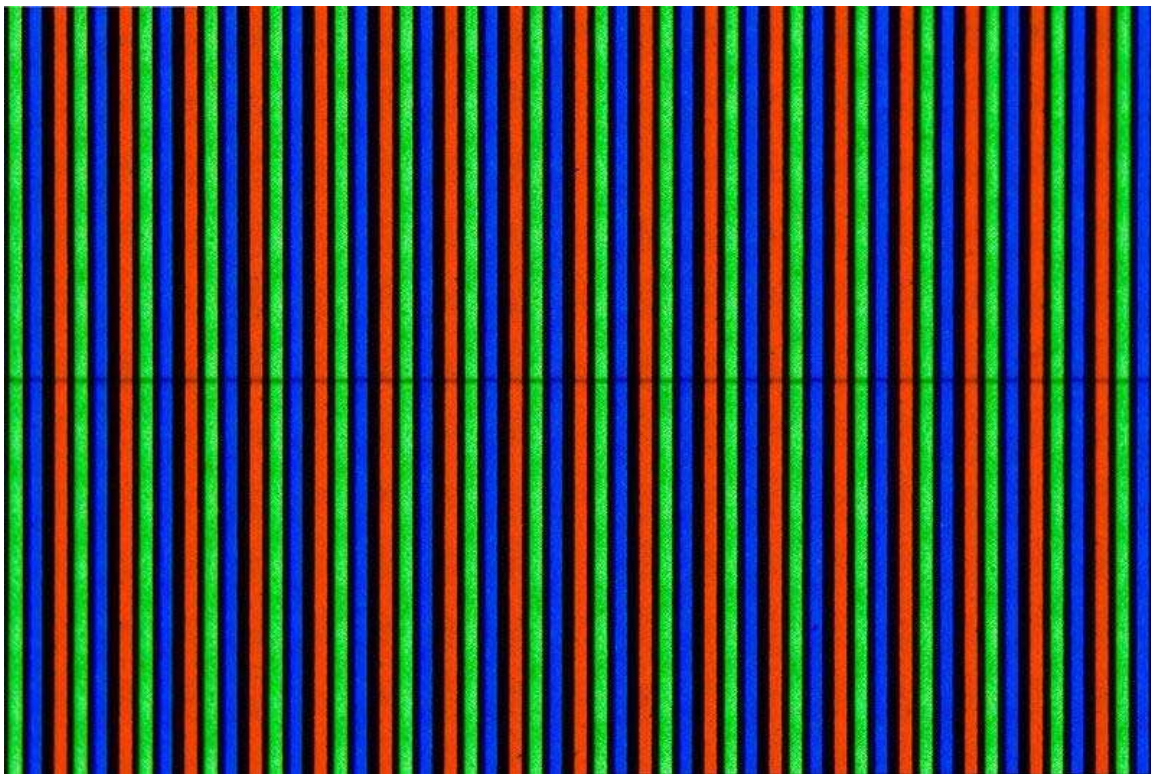
Phosphor



Example of phosphorescence



Monochrome monitor



Aperture grille CRT phosphors

A **phosphor**, most generally, is a substance that exhibits the phenomenon of **luminescence**. Somewhat confusingly, this includes both phosphorescent materials, which show a slow decay in brightness ($>1\text{ms}$), and fluorescent materials, where the emission decay takes place over tens of nanoseconds. Phosphorescent materials are known for their use in radar screens and glow-in-the-dark toys, whereas fluorescent materials are common in CRT screens, sensors, and white LEDs.

Phosphors are transition metal compounds or rare earth compounds of various types. The most common uses of phosphors are in CRT displays and fluorescent lights. CRT phosphors were standardized beginning around World War II and designated by the letter "P" followed by a number.

Note: Phosphorus, the chemical element from the phenomenon draws its name from, emits light under certain conditions, but this is due to chemiluminescence, not phosphorescence.

Principles

A material can emit light either through incandescence, where all atoms radiate, or by luminescence, where only a small fraction of atoms, called **emission centers** or **luminescence centers**, emit light. In inorganic phosphors, these inhomogeneities in the crystal structure are created usually by addition of a trace amount of dopants, impurities

called **activators**. (In rare cases dislocations or other crystal defects can play the role of the impurity.) The wavelength emitted by the emission center is dependent on the atom itself, and on the surrounding crystal structure.

The scintillation process in inorganic materials is due to the electronic band structure found in the crystals. An incoming particle can excite an electron from the valence band to either the conduction band or the exciton band (located just below the conduction band and separated from the valence band by an energy gap). This leaves an associated hole behind, in the valence band. Impurities create electronic levels in the forbidden gap. The excitons are loosely bound electron-hole pairs that wander through the crystal lattice until they are captured as a whole by impurity centers. The latter then rapidly de-excite by emitting scintillation light (fast component). In case of inorganic scintillators, the activator impurities are typically chosen so that the emitted light is in the visible range or near-UV where photomultipliers are effective. The holes associated with electrons in the conduction band are independent from the latter. Those holes and electrons are captured successively by impurity centers exciting certain metastable states not accessible to the excitons. The delayed de-excitation of those metastable impurity states, slowed down by reliance on the low-probability forbidden mechanism, again results in light emission (slow component).

Phosphor degradation

Many phosphors tend to gradually lose efficiency, by several mechanisms. The activators can undergo change of valence (usually oxidation), the crystal lattice degrades, atoms - often the activators - diffuse through the material, the surface undergoes chemical reactions with the environment with consequent loss of efficiency or buildup of a layer absorbing either the exciting or the radiated energy, etc.

Electroluminescent devices degrade in dependence on frequency of driving current and the luminance level, temperature; moisture impairs phosphor lifetime very significantly as well.

Examples:

- $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ (BAM), a plasma display phosphor, undergoes oxidation of the dopant during baking. Three mechanisms are involved; absorption of oxygen atoms into oxygen vacancies on the crystal surface, diffusion of Eu(II) along the conductive layer, and electron transfer from Eu(II) to adsorbed oxygen atoms, leading to formation of Eu(III) with corresponding loss of emissivity. Thin coating of aluminium phosphate or lanthanum(III) phosphate is effective in creation a barrier layer blocking access of oxygen to the BAM phosphor, for the cost of reduction of phosphor efficiency. Addition of hydrogen, acting as a reducing agent, to argon in the plasma displays significantly extends the lifetime of $\text{BAM}:\text{Eu}^{2+}$ phosphor, by reducing the Eu(III) atoms back to Eu(II).

- $\text{Y}_2\text{O}_3\text{:Eu}$ phosphors under electron bombardment in presence of oxygen form a non-phosphorescent layer on the surface, where electron-hole pairs recombine nonradiatively via surface states.
- ZnS:Mn , used in AC thin film electroluminescent (ACTFEL) devices degrades mainly due to formation of deep electron traps, by reaction of water molecules with the dopant; the traps act as centers for nonradiative recombination. The traps also damage the crystal lattice. Phosphor aging leads to decreased brightness and elevated threshold voltage.
- ZnS-based phosphors in CRTs and FEDs degrade by surface excitation, coulombic damage, build-up of electric charge, and thermal quenching. Electron-stimulated reactions of the surface are directly correlated to loss of brightness. The electrons dissociate impurities in the environment, the reactive oxygen species then attack the surface and form carbon monoxide and carbon dioxide with traces of carbon, and nonradiative zinc oxide and zinc sulfate on the surface; the reactive hydrogen removes sulfur from the surface as hydrogen sulfide, forming nonradiative layer of metallic zinc. Sulfur can be also removed as sulfur oxides.
- ZnS and CdS phosphors degrade by reduction of the metal ions by captured electrons. The Me^{2+} ions are reduced to Me^+ ; two Me^+ then exchange an electron and become one Me^{2+} and one neutral Me atom. The reduced metal can be observed as a visible darkening of the phosphor layer. The darkening (and the brightness loss) is proportional to the phosphor's exposition to electrons, and can be observed on some CRT screens that displayed the same image (e.g. a terminal login screen) for prolonged periods.

Materials

Phosphors are usually made from a suitable host material with an added activator. The best known type is a copper-activated zinc sulfide and the silver-activated zinc sulfide (*zinc sulfide silver*).

The host materials are typically oxides, nitrides and oxynitrides, sulfides, selenides, halides or silicates of zinc, cadmium, manganese, aluminium, silicon, or various rare earth metals. The activators prolong the emission time (afterglow). In turn, other materials (such as nickel) can be used to quench the afterglow and shorten the decay part of the phosphor emission characteristics.

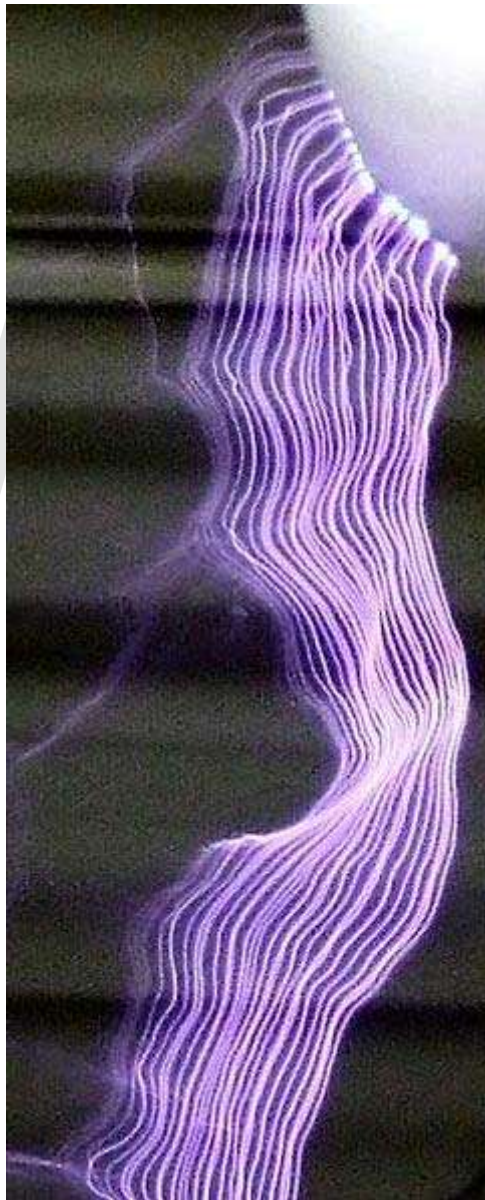
Many phosphor powders are produced in low-temperature processes, such as sol-gel and usually require post-annealing at temperatures of $\sim 1000^\circ\text{C}$, which is undesirable for many applications. However, proper optimization of the growth process allows to avoid the annealing.

Phosphors used for fluorescent lamps require a multi-step production process, with details that vary depending on the particular phosphor. Bulk material must be milled to obtain a desired particle size range, since large particles produce a poor quality lamp coating and small particles produce less light and degrade more quickly. During the firing

of the phosphor, process conditions must be controlled to prevent oxidation of the phosphor activators or contamination from the process vessels. After milling the phosphor may be washed to remove minor excess of activator elements. Volatile elements must not be allowed to escape during processing. Lamp manufacturers have changed composition of phosphors to eliminate some toxic elements, such as beryllium, cadmium, or thallium, formerly used.

The commonly quoted parameters for phosphors are the wavelength of emission maximum (in nanometers, or alternatively color temperature in kelvins for white blends), the peak width (in nanometers at 50% of intensity), and decay time (in seconds).

Electrical breakdown



Electrical breakdown in an electric discharge showing the ribbon-like plasma filaments from a Tesla coil.

The term **electrical breakdown** has several similar but distinctly different meanings. The term can apply to the failure of an electric circuit. Alternatively, it may refer to a rapid reduction in the resistance of an electrical insulator that can lead to a spark jumping around or through the insulator. This may be a momentary event (as in an electrostatic discharge), or may lead to a continuous arc discharge if protective devices fail to interrupt the current in a high power circuit.

Electrical system failure

The most common meaning is related to automobiles and is the failure of an electric circuit or associated device resulting in a loss of vehicle function (a breakdown). Common problems include battery discharge, alternator failure, broken wires, blown fuses, etc.

Failure of electrical insulation

The second meaning of the term is more specifically a reference to the breakdown of the insulation of an electrical wire or other electrical component. Such breakdown usually results in a short circuit or a blown fuse. This occurs at the breakdown voltage. Actual insulation breakdown is more generally found in high-voltage applications, where it sometimes causes the opening of a protective circuit breaker. Electrical breakdown is often associated with the failure of solid or liquid insulating materials used inside high voltage transformers or capacitors in the electricity distribution grid. Electrical breakdown can also occur across the strings of insulators that suspend overhead power lines, within underground power cables, or lines arcing to nearby branches of trees. Under sufficient electrical stress, electrical breakdown can occur within solids, liquids, gases or vacuum. However, the specific breakdown mechanisms are significantly different for each, particularly in different kinds of dielectric medium. All this leads to catastrophic failure of the instruments.

Disruptive devices

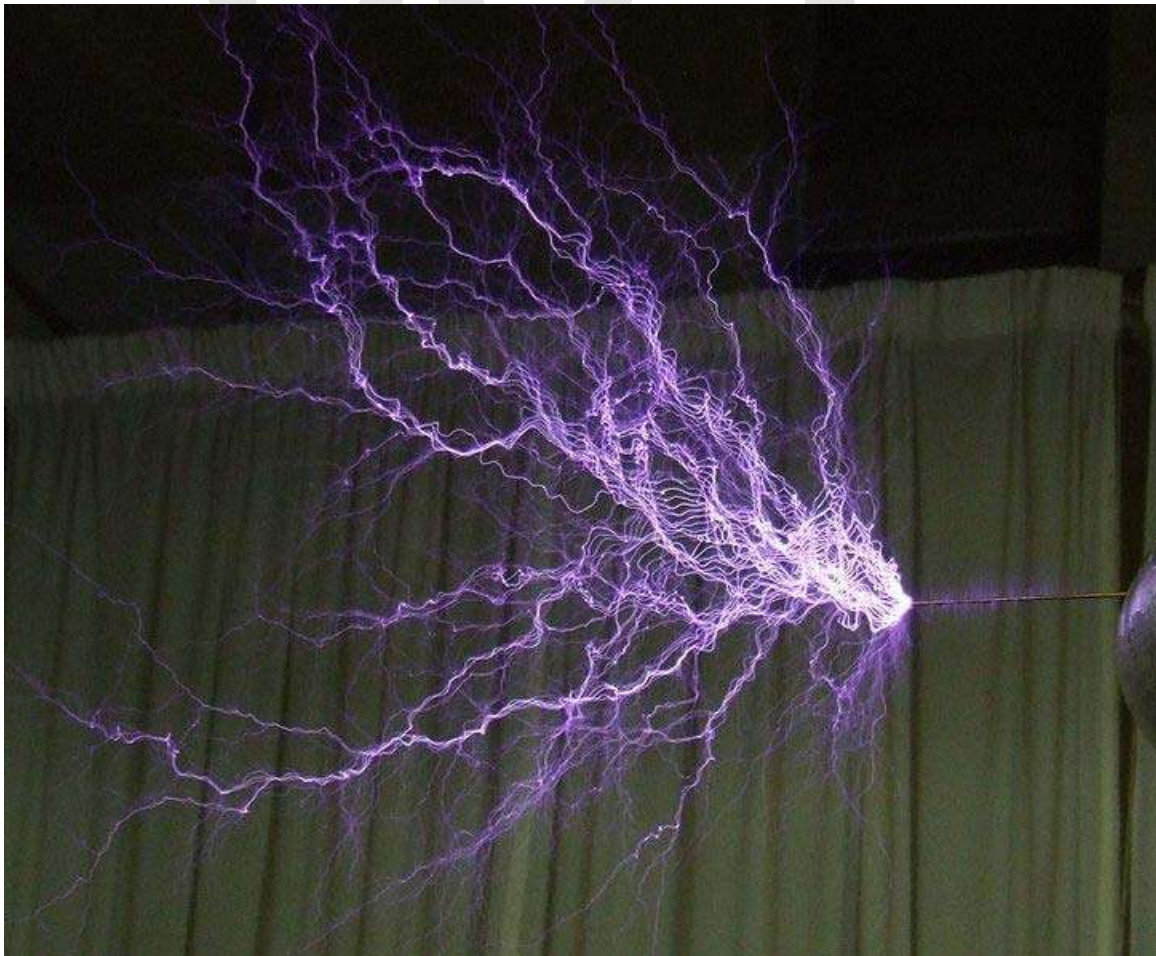
A **disruptive device** is a device that has a dielectric, whereupon being stressed beyond its dielectric strength, has an electrical breakdown. This results in the sudden transition of part of the dielectric material from an insulating state to a highly conductive state. This transition is characterized by the formation of an electric spark, and possibly an electric arc through the material. If this occurs within a solid dielectric, physical and chemical changes along the path of the discharge will cause permanent degradation and significant reduction in the material's dielectric strength. A spark gap is a type of disruptive device that uses a gas or fluid dielectric between spaced electrodes. Unlike solid dielectrics,

liquid or gaseous dielectrics can usually recover their full dielectric strength once current flow (through the plasma in the gap) has been externally interrupted.

Mechanism

Electrical breakdown occurs within a gas (or mixture of gases, such as air) when the dielectric strength of the gas(es) is exceeded. Regions of high electrical stress can cause nearby gas to partially ionize and begin conducting. This is done deliberately in low pressure discharges such as in fluorescent lights or in an electrostatic precipitator.

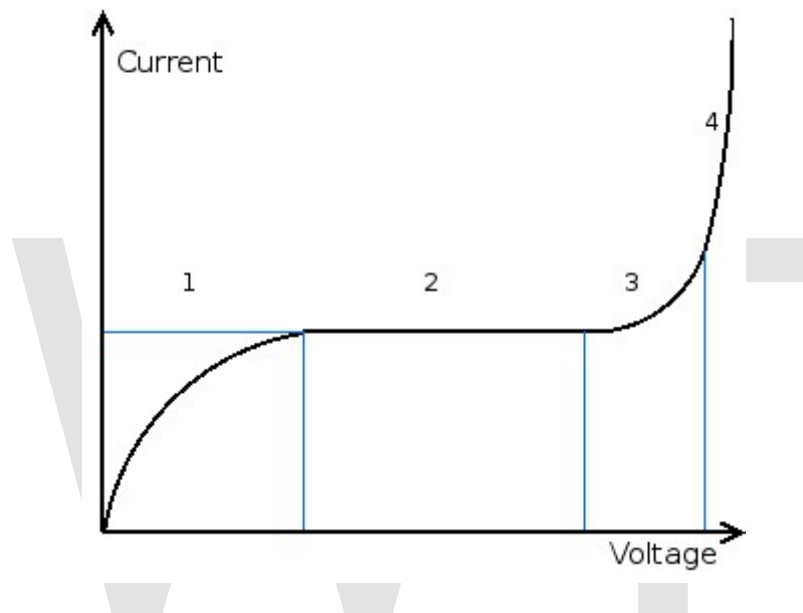
Partial electrical breakdown of the air causes the "fresh air" smell of ozone during thunderstorms or around high-voltage equipment. Although air is normally an excellent insulator, when stressed by a sufficiently high voltage (an electric field strength of about $3 \times 10^6 \text{V/m}$), air can begin to break down, becoming partially conductive. If the voltage is sufficiently high, complete electrical breakdown of the air will culminate in an electrical spark or arc that bridges the entire gap. While the small sparks generated by static electricity may barely be audible, larger sparks are often accompanied by a loud snap or bang. Lightning is an example of an immense spark that can be many miles long. The color of the spark depends upon the gases that make up the gaseous media.



Electric discharge showing the lightning-like plasma filaments from a Tesla coil.

If a fuse or circuit breaker fails to interrupt the current through a spark in a power circuit, current may continue, forming a very hot electric arc. The color of an arc depends primarily upon the conductor materials (as they are vaporized and mix within the hot plasma in the arc). Although sparks and arcs are usually undesirable, they can be useful in everyday applications such as spark plugs for gasoline engines, electrical welding of metals, or for metal melting in an electric arc furnace.

Voltage-current relation



Voltage-current relation before breakdown

Before breakdown, there is a non-linear relation between voltage and current as shown in the figure. In region 1, there are free ions that can be accelerated by the field and induce a current. These will be saturated after a certain voltage and give a constant current, region 2. Region 3 and 4 are caused by ion avalanche as explained by the Townsend discharge mechanism.

Corona breakdown

Partial breakdown of the air occurs as a corona discharge on high voltage conductors at points with the highest electrical stress. As the dielectric strength of the material surrounding the conductor determines the maximum strength of the electric field the surrounding material can tolerate before becoming conductive, conductors that consist of sharp points, or balls with small radii, are more prone to causing dielectric breakdown. Corona is sometimes seen as a bluish glow around high voltage wires and heard as a sizzling sound along high voltage power lines. Corona also generates radio frequency noise that can also be heard as 'static' or buzzing on radio receivers. Corona can also occur naturally at high points (such as church spires, treetops, or ship masts) during

thunderstorms as St. Elmo's Fire. Although corona discharge is usually undesirable, until recently it was essential in the operation of photocopiers (Xerography) and laser printers. Many modern copiers and laser printers now charge the photoconductor drum with an electrically conductive roller, reducing undesirable indoor ozone pollution. Additionally, lightning rods use corona discharge to create conductive paths in the air that point towards the rod, deflecting potentially-damaging lightning away from buildings and other structures.

Corona discharge ozone generators have been used for more than 30 years in the water purification process. Ozone is a toxic gas, even more potent than chlorine. In a typical drinking water treatment plant, the ozone gas is dissolved into the filtered water to kill bacteria and viruses. Ozone also removes the bad odours and taste from the water. The main advantage of ozone is that the overdose (residual) decomposes to gaseous oxygen well before the water reaches the consumer. This is in contrast with chlorine which stays in the water and can be tasted by the consumer.

Corona discharges are also used to modify the surface properties of many polymers. An example is the corona treatment of plastic materials which allows paint or ink to adhere properly.

