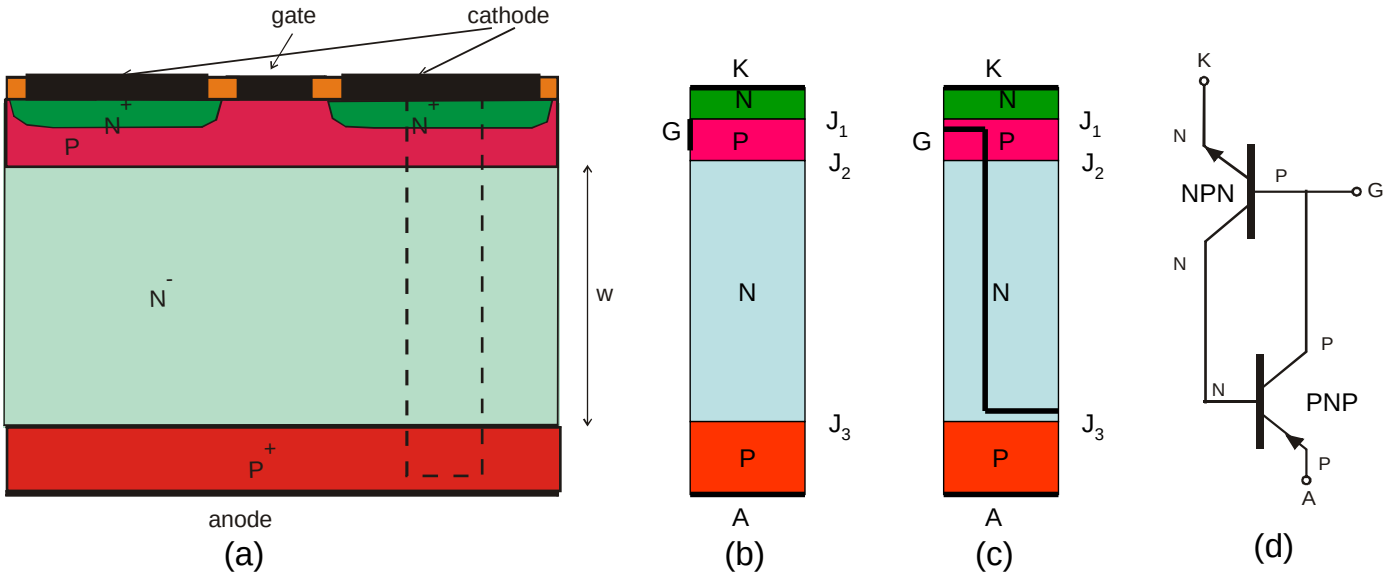
The Silicon Controlled Rectifier (SCR)

The Silicon Controlled Rectifier, also called Thyristor, is one of the oldest power devices, and it is actually employed as power switch for the largest currents (several kiloamperes) and voltages (several kilovolts) – as presented in the introductory lesson.
Its name indicates that the device behaves a a PIN diode (or rectifier) in his ON state, with a very low forward drop (if compared with the one of the other devices), but it is normally in OFF state, and to be brought in the ON state it requires a trigger on a separate control terminal (called gate). From this effect we can consider the SCR as a silicon device that behaves as a PIN diode – or rectifier - in forward bias if a command is done on the control terminal.
The main drawback of the SCR is that we can not control it in the opposite way – i.e. we can not act on the control terminal to bring it in the OFF state when it is already conducting (ON state). Then in order to switch off the SCR we can only rely on the natural current commutation, in other words we must revert the voltage across it, from forward to reverse bias (as for the PIN diode) to interrupt the current flowing into it.
The device, like the PIN diode and the BJT, is a bipolar device: its operation relies on both type of carriers, and the low voltage drop in conduction is actually due to the conductivity modulation (already discussed for the PIN diode) that takes place at high injection in the thick and low doped region (needed to sustain high voltages in OFF state).

The cross-section of the structure of an SCR is depicted in fig. (a): the two main terminals are the anode A (on the bottom of the chip) that is biased at positive voltages in forward state, and the cathode K(on top of the chip) that is grounded. The gate terminal is the control electrode that allows to switch on the SCR.
Considering the main portion indicated in dashed line, we see that there are four doping layers(N^+ P N^- P^+) and three P/N junctions from cathode to anode (fig. (b). A simple analysis of this structure can be done by considering it made by two complementary BJT (NPN and PNP) closely connected, as indicated in fig. (c,d):



D.C. analysis an trigger condition of the SCR

An SCR approximate analysis, to understand the behavior in forward bias (V_{AK} >0) , and to define the trigger condition that will change its state from OFF to ON, can be done by referring to the configuration with two BJT seen before.

a) forward blocking (OFF) state

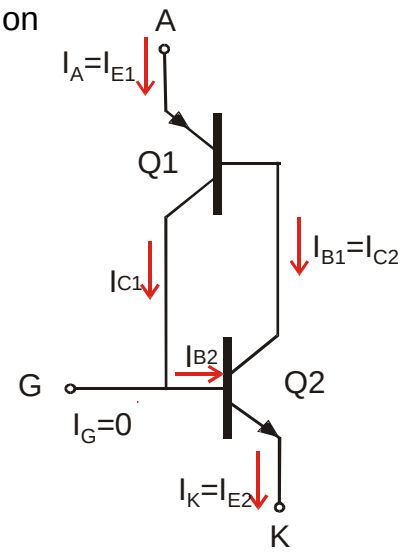
When I_B = 0, and assuming V_{AK} > 0, both BJT are in the active region and we have: $I_A = I_K = I = I_{C1} + I_{C2}$. At high voltage where multiplication takes place (as seen for BJT eq. (26)) we have:

$$I_{C1} = M\alpha_1 I_{E1} + MI_{C01} = M\alpha_1 I + MI_{C01}$$
$$I_{C2} = M\alpha_2 I_{E2} + MI_{C02} = M\alpha_2 I + MI_{C02} \qquad (1)$$

Summing the two currents:

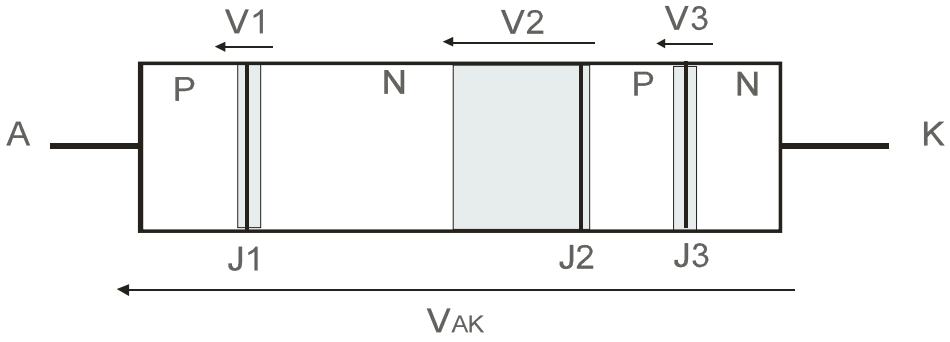
$$I = I_{C1} + I_{C2} = (M\alpha_1 + M\alpha_2)I + M(I_{C01} + I_{C02})$$

$$I = \frac{M(I_{C01} + I_{C02})}{1 - M(\alpha_1 + \alpha_2)} \qquad (2)$$



The current *I* is negligible at low voltages and increase largely only when $M(\alpha_1+\alpha_2) = 1$. The max voltage V_{MAX} is near to the breakdown voltage of J_2 because the α of the BJT is quite low.

In the forward blocking state the P/N junctions J_1 and J_3 are forward biased, while the N/P junction J_2 is reverse biased, as indicated in figure.
The central N region must be thick and low doped to sustain high reverse voltages with V_{AK} positive, needed to keep the SCR in the OFF condition even at high anode voltages if the SCR is not triggered.



The PNP transistor Q_1 has α_1 very low (< 0.5), because the width of the N base region is large; also the NPN transistor Q_2 has a relatively low α_2 because the P layer is not thin nor low doped. We must consider that both α_1 and α_2 depend on the V_{AK} voltage, because the effective width of the base for the two BJT (mainly Q1) decreases if V_{AK} is increased, due to the depletion width of region J_2 . The voltage dependence of alphas is important in defining the IV curves of the SCR in the off-state region.

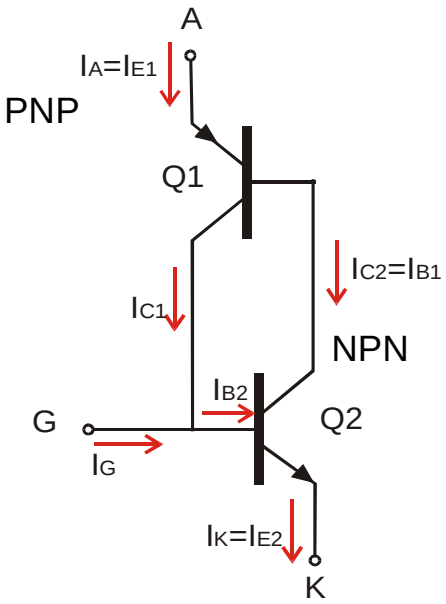
b) the triggering condition

If a gate current is injected from the gate electrode : $I_K = I_A + I_G$ and eq. (1) are changed as:

$$I_{C1} = M\alpha_1 I_A + MI_{C01}$$
$$I_{C2} = M\alpha_2 I_K + MI_{C02} = M\alpha_2 (I_A + I_G) + MI_{C02} \quad (3)$$

$$I_A = I_{C1} + I_{C2} = M\alpha_1 I_A + M\alpha_2 (I_A + I_G) + M(I_{C01} + I_{C02})$$

and we have: $$I_A = \frac{M\alpha_2 I_G + M(I_{C01} + I_{C02})}{1 - M(\alpha_1 + \alpha_2)} \quad (4)$$



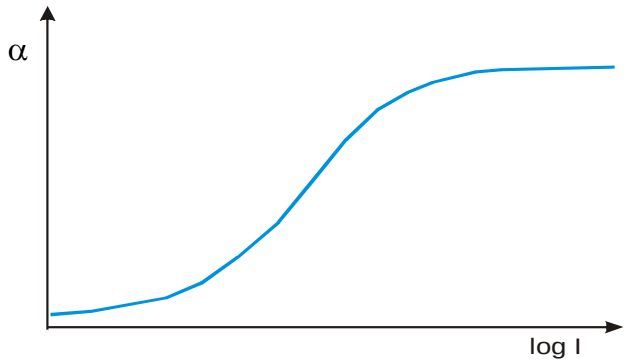
Eq. (4) is the general equation for the trigger condition of the SCR, valid also for low voltages with no multiplication (M=1):

$$I_A = \frac{\alpha_2 I_G + (I_{C01} + I_{C02})}{1 - (\alpha_1 + \alpha_2 (I_K))}$$

in that case we can still have a current increasing indefinitely (trigger condition) if the condition $\alpha_1+\alpha_2(I_K) = 1$ holds.

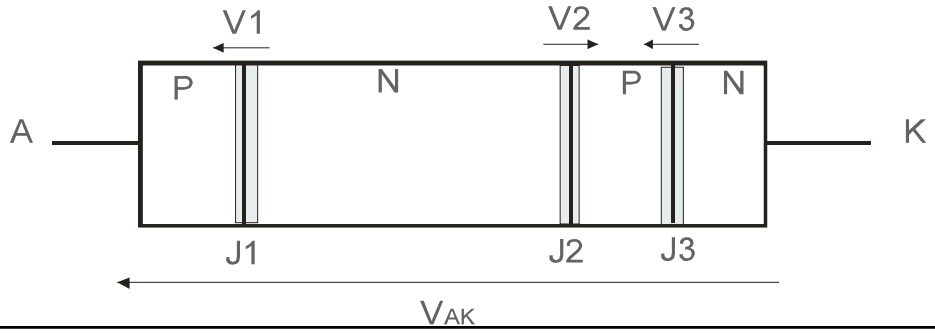
Why the emitter current gain α is a function the junction current? As seen in the current gain analysis of BJT (pag. 4), α is basically the injection factor γ of the emitter junction: $\gamma = I_{nE}/I_E$. At low current we must consider for the current I_E not only the hole and electron components but also the generation/recombination current in the depleted layer, that has a dependence $\exp(V_J/2V_T)$ from the junction voltage. In the case of SCR we have for the α_2 expression:

$$\alpha_2 \cong \frac{I_n}{I} = \frac{I_n}{I_n + I_p + I_G} \cong \frac{I_n}{I_n + I_G} = \frac{I_{n0} \exp\left(\frac{V_J}{V_T}\right)}{I_{n0} \exp\left(\frac{V_J}{V_T}\right) + I_{G0} \exp\left(\frac{V_J}{2V_T}\right)} \quad (5)$$

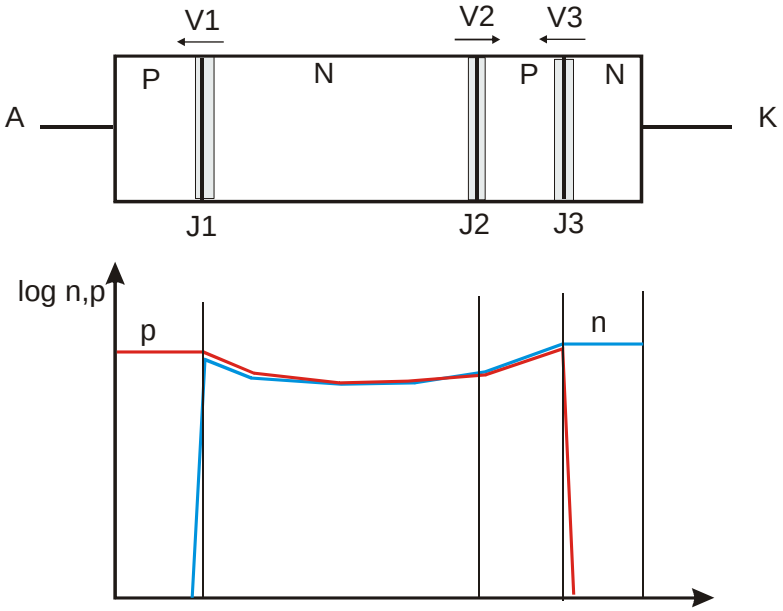


At low currents I_G is a substantial part of the total current I, and α is low; if the current I increases (because V_J is increased), the I_G component becomes negligible with respect to I_n and α increases near to unity. If we then increase the total J_3 current by injecting a gate current I_G, α_2 will be increased from <0.5 to about 0.8 - 0.9, and the trigger condition will be met.

When the condition $(\alpha_1+\alpha_2)=1$ is met, we reach the turn-over point in the IV curve of the SCR, because $dI_A/dV_{AK} = 0$. The total current I_A increases without limits, because both BJTs Q_1 and Q_2 enter into saturation mode. A positive feedback develops for this configuration, because the increase of I_{C2} current of Q_2 corresponds an increase of the drive base current I_{B1} of Q_1, and that in turn increases I_{C1} and hence the drive base current I_{B2} of Q_2, and so on.
This situation is self-sustaining (latching state), because it can be maintained even if the initial gate current I_G is removed, because now the base current I_{B2} of Q_2 is generated by the I_{C1} current of Q_1. The I_G current can be then a short pulse to trigger the switch from OFF to ON state, but it does not need to be maintained after the triggering (this is an important aspect for the drive circuitry).
If both Q_1 and Q_2 are in saturation, the N/P collector junction J_2 (for both BJT) must be forward biased, as indicated in figure, and the whole V_{AK} voltage drops to a very low value, being the (algebric) sum of three forward biased junctions, with J_1 and J_2 summing and J_3 subtracting: $V_{AK} = V_1 + V_2 – V_3$.

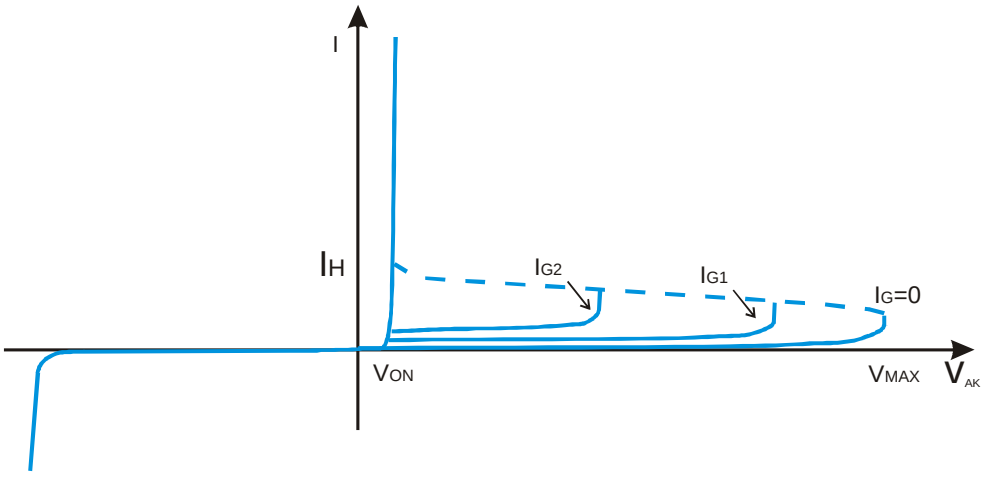


In the latching condition (ON state) the thick an low doped N^- region is in conductivity modulation, but in that case we have a much higher conductivity, because in this state both the PN junctions J_1 and J_2 are in forward bias and inject holes from both sides of the N^- region.
As a consequence the hole distribution is more flat than in BJT in saturation, as indicated in the figure, even with thicker layers, (more than 100 μm, as needed to sustain reverse voltage of several kilovolts), and the voltage drop ΔV across this layer is still low.



IV characteristics of the SCR

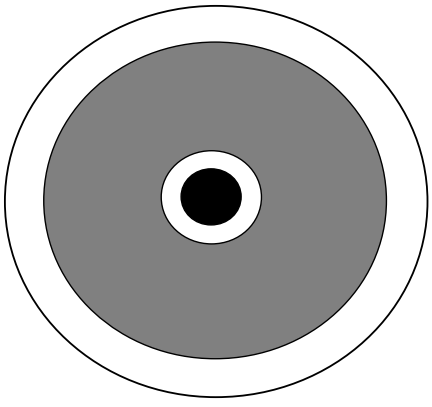
From the previous analysis, we can now understand the IV characteristics of the SCR.
In reverse bias ($V_{AK}<0$), both J_1 and J_3 are reverse biased, and J_1 will sustain the voltage up to his breakdown voltage.
In forward bias, if we inject an I_g pulse the turnover point is reduced and the SCR enters in a negative resistance region (dashed line) up to the stable ON state (latching state) where the voltage drop V_{ON} is of the order of that of a single P/N diode.
For larger values of I_G, the voltage V_{AK} of the turnover point gets lower, because a lower voltage is needed to hold the condition $\alpha_1+\alpha_2(V,I_G)=1$.



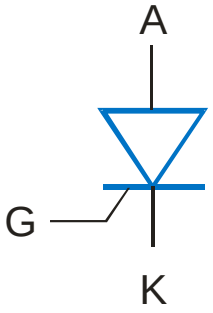
To bring the SCR from ON to OFF state, we must reduce the current I_A below the holding value I_H, (a value quite low respect to the operating current).
This is normally done by reverting the V_{AK} voltage, as done for the PIN diode.

Turn-on dynamics

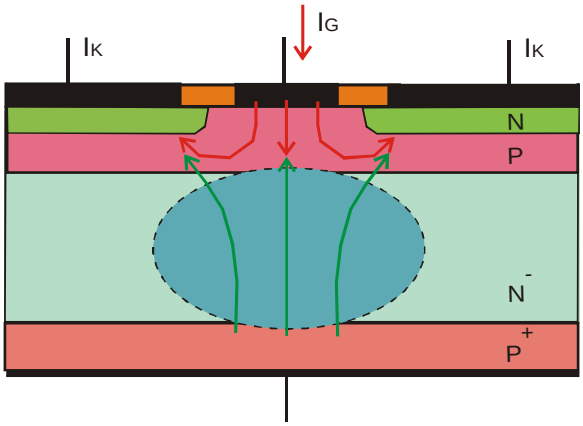
The lay-out of a typical SCR is reported in fig. (a), together with the electrical symbol for the SCR device (fig. (b)).
For large current ratings (up to several kiloamp), the device is usually made on a single wafer, as indicated in the figure for a wafer of 2 inches.
The small gate area with respect to the large cathode one (here the cathode lateral length is about 2 cm), will make an important issue in the turn-on dynamics, due to the lateral spreading of the carriers.



(a)

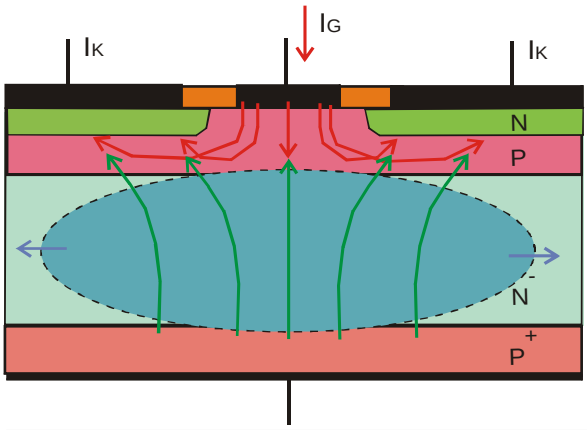


(b)



(a)

With reference to the cross-section picture during the turn-on phase, just after the reaching of the turnover voltage (fig. a), the high injection region ($p=n>>N_D$) in the N^- region initially starts to develop down the gate contact area, where the I_G current inject holes into the P base.



(b)

Then the conductivity modulation region spreads laterally in the large cathode area (fig. b), through the carrier diffusion (there is no large lateral electric field), filling eventually the whole cathode.
Only at that time the whole N^- region will present a low voltage drop, and the current will grow up to the final transient value, because it is equally distributed across the whole cathode area.

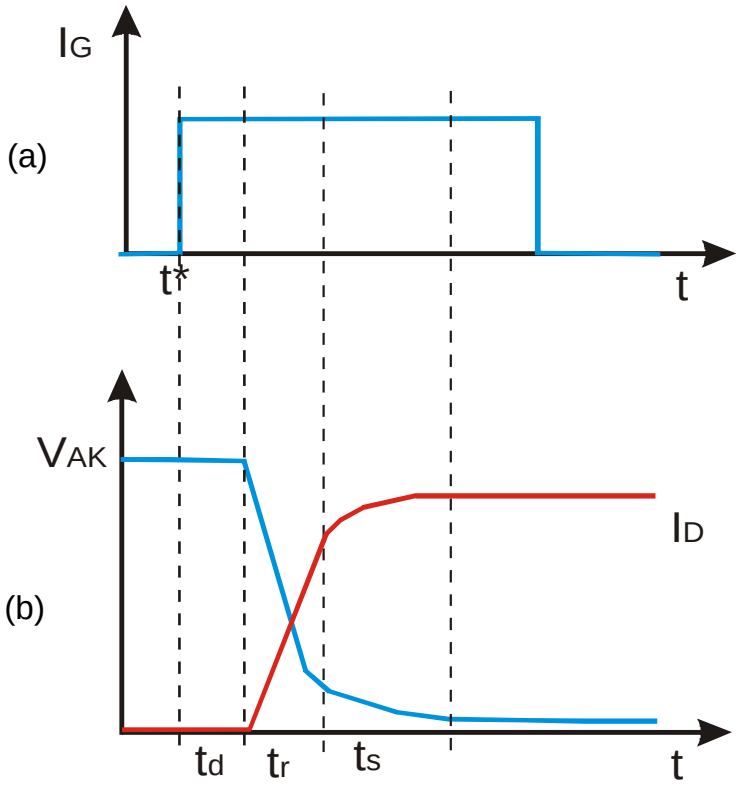
a) di/dt limitation

The I_A current constriction in the initial part of the turn-on transient poses some limitation on the current rise time, that must be not less than the time needed for the conductivity region to spread across the whole cathode area.
If the current rise time is lower than this $di/dt|_{MAX}$ value (usually given in the datasheet), the current forced into the SCR will flow only in the central part of the chip, giving rise to an excessive current density and to a strong power dissipation in that small area, with an unacceptable temperature rise (and device failure).
The speed of the lateral carrier propagation is about some hundreds of μm/μs, so the current rise time can be of the order of hundreds of μs if the lateral cathode dimention is of some cm.

b) dv/dt limitation

The limitation on a $dV_{AK}/dt|_{MAX}$ is due to the unwanted turn-on that could be caused by an excessively short rise time of the anode voltage, even if there is no I_G pulse applied (the device should stay in OFF state). Recalling the two BJT model introduced, if the V_{AK} has a rapid increase, the current I_{C1} in the off state (eq. 3) is the sum of the leakage current I_{C01} and the capacitive current in C_{CG1}: $I_{C1} = I_{C01} + C_{CG1}\, dV_{AK}/dt$. This induces an extra current I_{C1}' that has the same effect in the base of Q_2 as an added I_G .
The SCR then can be triggered in ON if that displacement current is of the order of the I_G needed for triggering. The dV/dt limitation indicated the max allowed voltage rise time that does not generate a critical I'_{C1}.

Turn-on waveforms

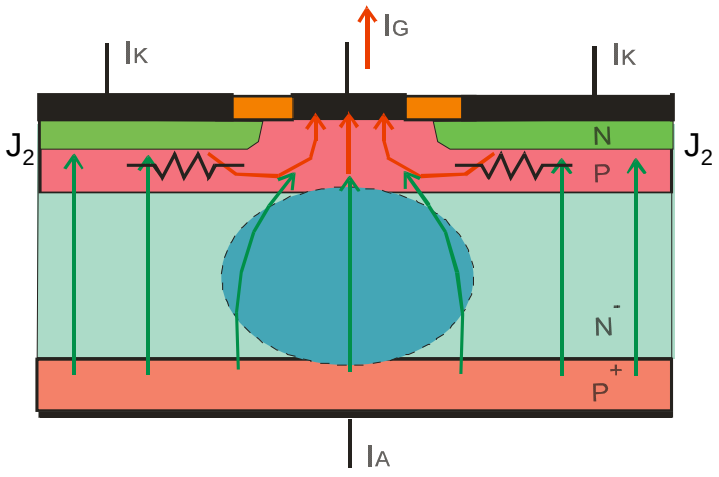


The turn-on dynamics of the SCR is sketched in these plots. Starting from the time t* when the I_G current is applied (fig. a), we mist consider (fig. b):

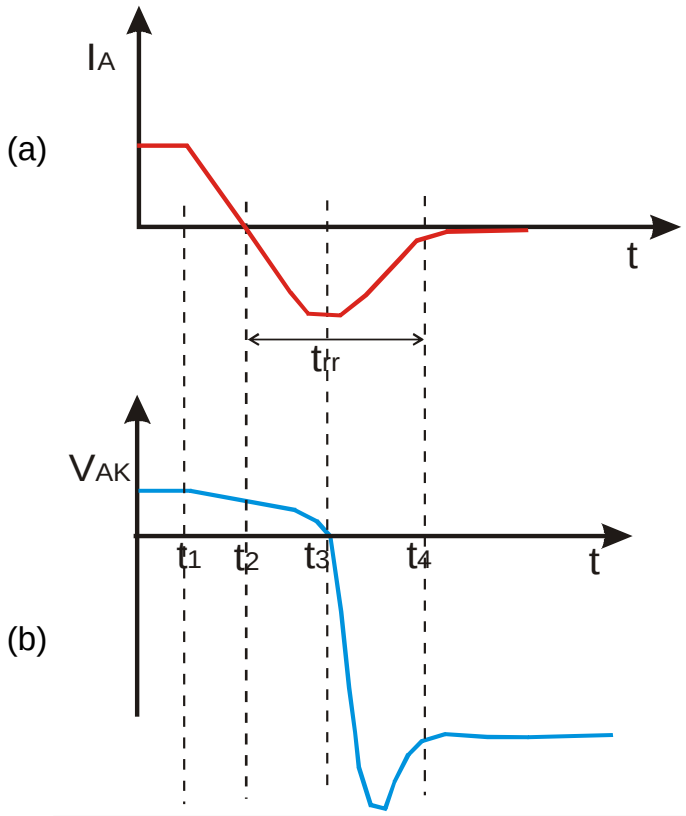
the delay time t_d, that is the time needed to reach the triggering condition $(\alpha_1+\alpha_2)=1$ and the turnover point.
the rise time t_r that is the time needed to create the high injection region and the conductivity modulation near the gate region
the spreading time t_s that is the time needed to the spreading of the high injection region across the cathode area.

The minimum gate pulse duration must be the sum of these three times to allow the full switching of the SCR.

Why the SCR can not be switched OFF by reverting the gate current?
When the gate current is reversed, it will remove only the stored charges underneath the gate contact, as indicated in figure. Then only the lateral part of junction J_2 will be reverse biased: the spreading resistance of the thin and long P region under the cathode area will give rise to a transverse voltage drop that will keep in forward bias most part of the cathode/gate J_2 junction.
Then the SCR will continue to stay in its latching condition, despite of the reverse gate current I_G applied.
The only way to switch off the SCR, as said before, is to reduce the anode current I_A below the holding value I_H.



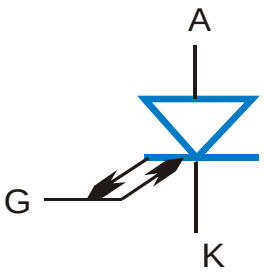
Turn-off waveforms



The turn-off dynamics of the SCR when the anode voltage (and current) is reverted, is similar to the one of the PIN diode, because both devices rely on the natural current commutation (the anode voltage must be reverted to switch off the device).

The I_A current then start to decrease below the I_H value and then reverse (fig. a). The stored charges in the wide N^- region must recombine and at time t_3 the I_A current reaches its minimum value.
The V_{AK} voltage first reduces below the V_{ON} (fig. b) and then start to increase in reverse bias when the current I_A reaches it minimum (the N^- region start to be depleted from mobile charge).

The Gate Turn-Off Thyristor (GTO)



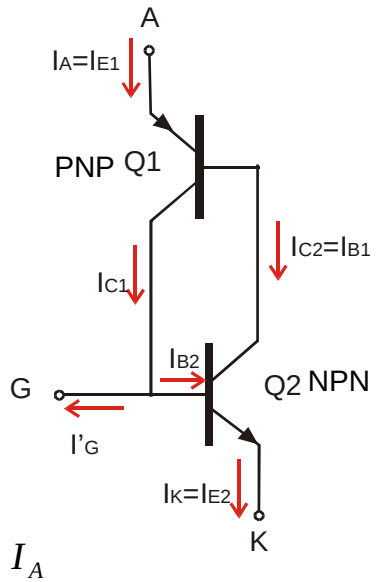
The GTO is basically an SCR with the capability to being switched off even with positive anode voltages, with a negative gate current pulse applied; it is switched on (analogously to the SCR) with a positive gate current pulse. The electrical symbol of the GTO is reported in figure, where the bidirectional gate current is indicated.

The ON state characteristic and the turn-on behavior is equal to the one of the SCR so we will concentrate on the turn-off capability of this device.

The turn-off can be initially identified in an approximate way by using the two complementary BJT configuration seen before for the SCR. Recalling the previous analysis done, the latching state is maintained by the regenerative feedback action of the two transistors Q_1 and Q_2 connected in current loop. To inhibit the regenerative action we must bring the transistor Q_2 out from saturation using a negative gate current I_G'; in the active region the collector junction (J_1) of Q_2 will become reverse biased and the egenerative feedback is blocked.
The condition for Q_2 to exit from saturation is:

$$\frac{I_{C2}}{\beta_2} > I_{B2} \quad \text{where} \quad \beta_2 = \frac{\alpha_2}{1-\alpha_2} \qquad (6)$$

Recalling the schematic for the two transistors equivalent, we have:



$$I_{B2} = I_{C1} - I'_G = \alpha_1 I_A - I'_G \qquad I_{C2} = (1-\alpha_1) I_A \qquad (7)$$

Substituting (7) in (6) we have:

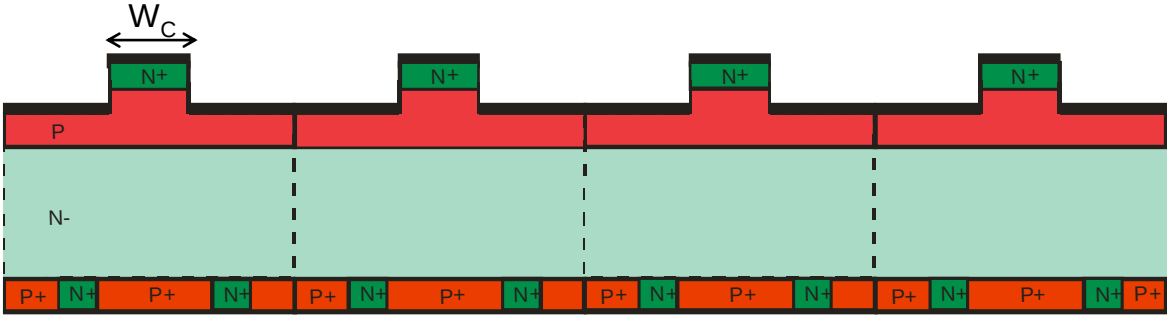
$$I_G' > \alpha_1 I_A - \frac{(1-\alpha_1) I_A}{\beta_2}$$

and from the value of β_2:

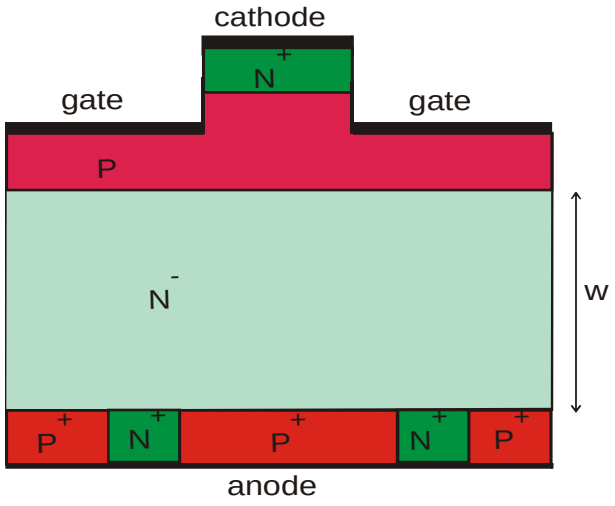
$$I_G' > I_A \left[\alpha_1 - \frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_2} \right] = I_A \left[\frac{\alpha_1 + \alpha_2 - 1}{\alpha_2} \right] \quad \Rightarrow \quad I_G' > \frac{I_A}{\beta_{OFF}}$$

where a turn-off current gain: $\beta_{OFF} = \dfrac{I_A}{I_G'} = \dfrac{\alpha_2}{\alpha_1 + \alpha_2 - 1}$ can be defined.

From the simplified one-dimensional analysis made with the two transistor equivalent, in principle it will be possible to turn-off the SCR, but with the usual values of α_1 and α_2 the turn-off β is about 1, and we need to switch off the device a gate current equal to the anode one. Moreover, this analysis neglect the transverse voltage drop due to the base spreading resistance that inibit the SCR turn-off.
We need to change the gate and cathode structure in order to a) increase the α_2 value of the NPN transistor, so to have a turn-off $\beta \geq 10$, and b) reduce the base spreading resistance.
The main change is to resort to an interdigitated structure for the gate and cathode contacts, as schematically indicated in the GTO cross section. The cathode is raised above the gate, and its lateral width W_C is contained, and the gate is contacted by an inner metallization plane. As a result the GTO is made of many elementary cells as indicated.



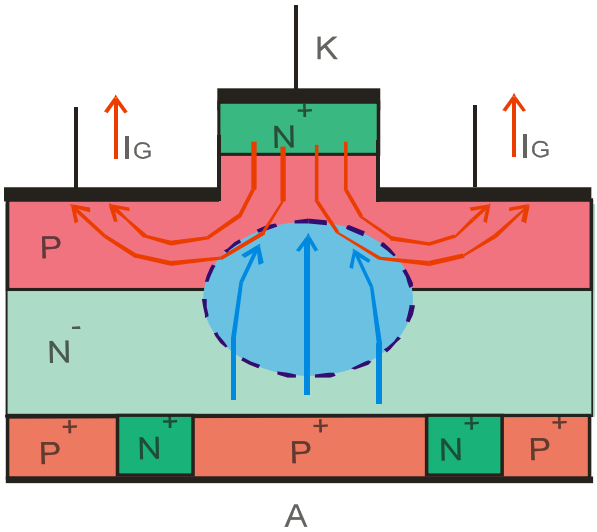
GTO cross-section



With reference to the elementary cell structure of the GTO, we can do the following comments:

a) the gate contact is obtained by etching the top surface after P and N^+ doping on the whole wafer; in this way the cathode has an high periphery/area ratio and the charge removal from the cathode region is more effective
b) the doping and thickness of the gate layer are chosen to obtain a quite high α_2 value and a turn-off β of about 10.
c) the anode region is made of alternate P^+ and N^+ regions (the N^+ with a lower area than the P^+ one): the N^+ layers are named "anode shorts" because they act as localized short circuits across the anode junction J_3. Their role is to allow a faster removal of the stored charge in the N^- region and to reduce the turn-off time.

GTO turn-off



The turn-of dynamics is sketched in this figure, with reference to the elementary cell of the GTO.

At the beginning of the turn-off transient, the negative gate current removes the stored charge from the gate area far from the cathode region. The high injection region with p and n carriers then concentrates below the cathode area, and subsequently it shrinks down due to the carrier removal through the gate lateral region, until the P/N^- junction (collector junction of the NPN Q_2) becomes reverse biased, and the regenerative feedback between Q_1 and Q_2 is blocked.

The anode shorts N^- in the anode layer help in removing the stored charge accumulated near the anode junction that would be otherwise blocked by the N/P junction. Then the recovery time is reduced and the switching speed is increased.
As for the PIN diode, the dynamics of the turn-off is dependent also from the lifetime in the low doped region: the combined effects of the lifetime control and of the anode shorts are beneficial in the recovery time reduction, but are negative on the on-state voltage V_{ON} in forward conduction: as a result a trade-off relation between the reverse recovery losses and forward steady-state losses can be defined for any GTO (and SCR) device, linking the best choice of the relevant parametersa to the specific application (high frequency operation or low frequency/high current operation).
It must be noted that the anode shorts will pose a significant limitation in the GTO ratings with respect to the SCR ones: the GTO can not withstand reverse voltages (contrarily to the SCR) because the J_1 junction is shorted by the N^+ shorts, and can not sustain reverse bias. Asd a consequence, the GTO can be used only as a controlled switch for positive anode voltages.