POWER ELECTRONICS

Luca Colombo | Politecnico di Milano | a. a. 2022/23

DISCLAIMER

These notes cover the arguments of the course 'Power Electronics' held by Professor M. Ghioni at Politecnico di Milano during the academic year 2022-2023.

Since they have been authored by a student, errors and imprecisions can be present.

These notes don't aim at being a substitute for the lectures of Professor Ghioni, but a simple useful tool for any student (life at PoliMi is already hard as it is, cooperating is nothing but the bare minimum).

Please remember that for a complete understanding of the subject there is no better way than directly attending the course (DIY), which is an approach that I personally suggest to anyone. Indeed, the course is really enjoyable and the professor very clear and helpful.

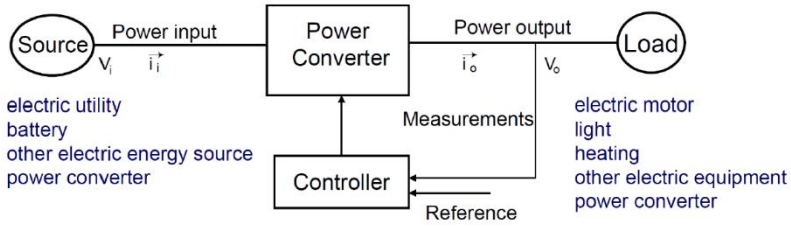
In any case, if you found these notes particularly helpful and want to buy me a coffee for the effort, you're more than welcome: https://paypal.me/LucaColomboxc

lucacolombo29@gmail.com

POWER ELECTRONICS

It's not the level of power that defines what power electronic is. Power electronics deals with the use of electronics for converting and controlling the flow of electrical energy from the source to the load in the most efficient way (power processing).

The basic diagram of any power electronic system is the following.

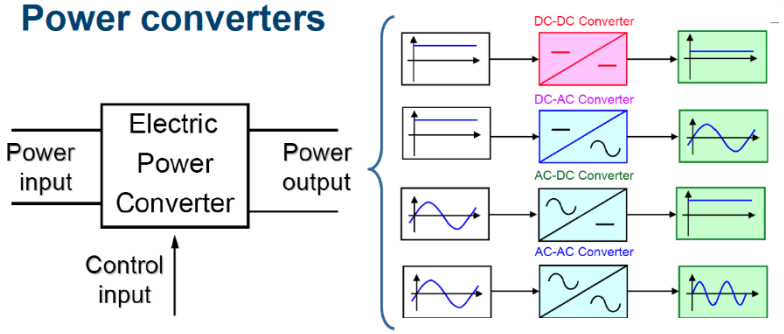


The power converter is used to adapt the source to the load. The other transfer of the power converter is the control of the energy flow from the source to the load.
The power electronic circuit is typically a feedback circuit. E.g., we need to regulate the voltage provided to the load from the source. We measure the output voltage, we compare it with a reference and the error is amplified and used to modulate the flow of energy to compensate for the error.

POWER CONVERTERS

There are 4 types:

- **DC/DC**: it takes a DC voltage in input and produces it at the output, and at the output it can be higher, lower or changed with respect to the input.
- **AC/DC**: rectifier.
- **DC/AC**: this operation is called inversion. These converters are used to drive motors.
- **AC/AC**: typically used in high power applications.

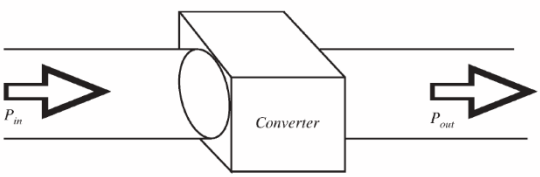


Dc-dc conversion:	Change and control voltage magnitude
Ac-dc rectification:	Possibly control dc voltage, ac current
Dc-ac inversion:	Produce sinusoid of controllable magnitude and frequency
Ac-ac cycloconversion:	Change and control voltage magnitude and frequency

DC-DC conversion plays a central role in the growing field of power management.

Efficiency

It's one of the main guidelines when designing DC/DC converters. High efficiency means that the conversion is provided with no or minimum power dissipation. Another figure of interest is the **dimension**, alongside with the **reliability** and the **cost**.



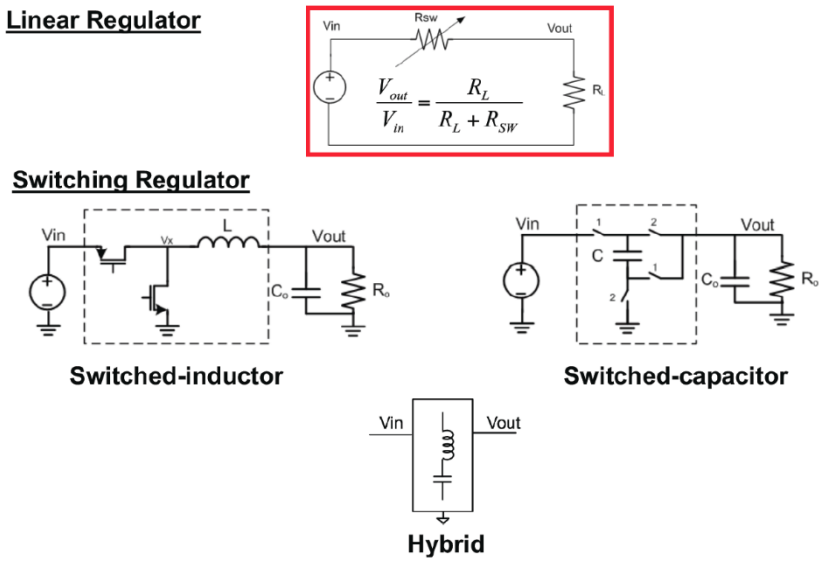
$$\eta = \frac{P_{out}}{P_{in}}$$

$$P_{loss} = P_{in} - P_{out} = P_{out}\left(\frac{1}{\eta} - 1\right)$$

- High efficiency leads to low power loss within converter
  - small size and reliable operation is then feasible
- Efficiency is a good measure of converter performance

The efficiency is defined as the ratio between the output power and the input power. The higher the efficiency, the lower the power lost in the conversion. Power lost is typically converted in heat and has to be dissipated with heatsinks, which are bulky. Moreover, if we push efficiency we also increase the battery life.

Types of DC/DC converters

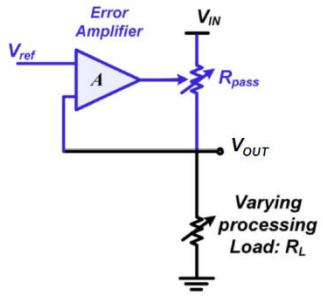


- **Linear regulator**: typically it is a sort of voltage divider. They have some problems, and in order to overcome these problems we can use different converters.
- **Switching converters**: the difference with respect to the regulators is the presence of a feedback loop. They use transistors as switches and we have 3 subcategories: switched inductors, switched capacitor and hybrid converters. Hybrid are switched capacitors with small inductances to have resonances and increase the efficiency of the system.

Linear voltage regulator

We have a feedback loop to set the output voltage equal to the reference voltage and the error is used to modulate the Rpass to regulate the flow of power and keep Vout equal to Vref.
Assuming the amplifier is absorbing a negligible current, so that all the current flows in the two resistances, the output power is simply Vout*Iout, as in the image.
Then the efficiency is the ratio between the output voltage and input one.

Resistor-divider-based DC-DC converter

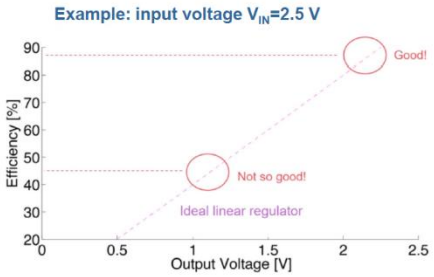


$$P_{OUT} = V_{OUT} \cdot I_{OUT}$$
$$P_{IN} = V_{IN} \cdot I_{IN}$$

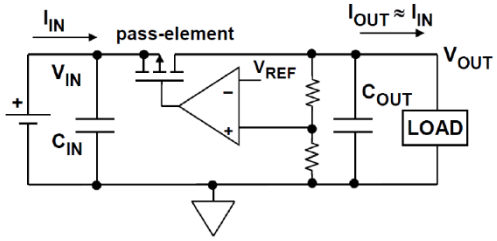
- Ignore quiescent current

$$I_{IN} \cong I_{OUT}$$

$$\Rightarrow \quad \eta = \frac{P_{OUT}}{P_{IN}} \cong \frac{V_{OUT}}{V_{IN}}$$



A real voltage regulator will be the following. Again, we are assuming the bias currents as negligible.



Assume:

- V_{IN} = 12 V
- V_{OUT} = 5 V
- R_L = 1 Ω

→ I_{OUT} = 5 A

- $I_{IN} \approx I_{OUT}$

$$P_{IN} \cong V_{IN} I_{OUT} = 12V \cdot 5A = 60W$$
$$P_{OUT} = V_{OUT} I_{OUT} = 5V \cdot 5A = 25W$$
$$P_{LOSS} \cong P_{IN} - P_{OUT} \cong (V_{IN} - V_{OUT}) I_{OUT} = 35W\ !$$

$$\Rightarrow \quad \eta = \frac{P_{OUT}}{P_{IN}} \cong \frac{V_{OUT}}{V_{IN}} = 40\%$$

This 40% efficiency is not good. 35W are dissipated by the passive element, transformed into heat and they must be dissipated → we need a heat sink, but it is bulky, so it is better not to use it.

Linear regulation: summary

- Very small implementations possible → high power density
- Simple, low-cost design; low output noise
- Good efficiency for low V_{IN}/V_{OUT} ratios
  - Low drop-out (LDO) regulators, a special class of linear regulators that have a low minimum input-to-output voltage ratio (i.e., "drop-out" voltage)
  - Used mainly for regulating an un-regulated power supply that has a slightly higher voltage than the required V_{OUT}
- Step-down only, i.e. $V_{OUT} < V_{IN}$
- Very inefficient at large conversion ratios

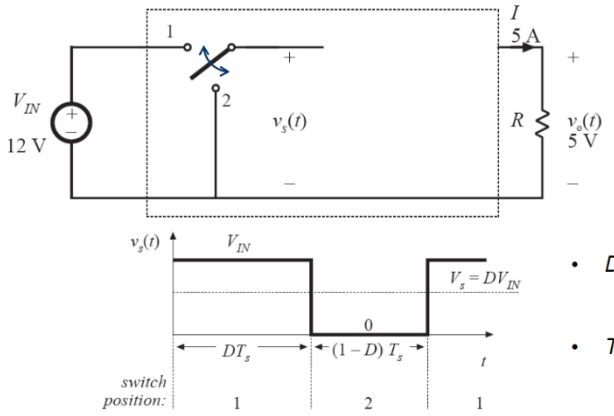
How can we get better efficiency?

Linear regulators are used only in the applications where the difference between the input and output voltages is minimal. They are called **low-dropout regulators** (LDO).

KEY DESIGN GUIDELINES

To seek for high efficiency, it is better to avoid magnetics (from signal processing theory), while from power processing it is better do avoid lossy elements (resistors and transistors in the linear region).

Example – Switching DC-DC converter (Buck Converter)



- D = duty cycle
  → $0 \le D \le 1$
- T_s = switching period
  = 1 / f_s

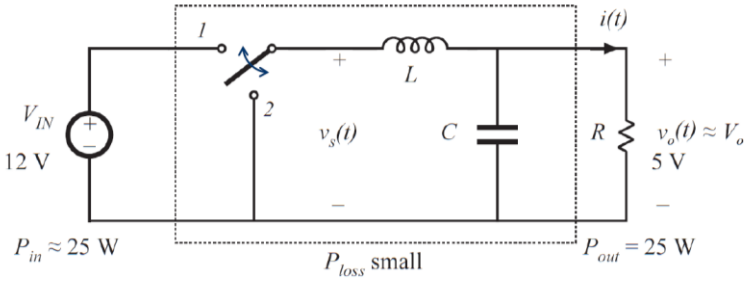
DC component of $v_s(t)$ = average value:

$$V_s = \frac{1}{T_s}\int_0^{T_s} v_s(t)\,dt = DV_{IN}$$

Input voltage is 12V and output one is 5V. The output current is 5 A because of the resistive load. We are replacing the linear regulator with a single pole ideal switch. The switch is switched periodically between position 1 and position 2. The voltage v_s is Vin if switch is in position 1, it is 0 if the switch is in position 2. In the end I get a square waveform with a fixed period that is the reciprocal of the switching frequency. The Duty Cycle is the ratio between the time in position 1 and the total switching period.

The square waveform has a non-zero average value that we can calculate. V_s (upper letter because constant) is the DC multiplied by the input voltage. But the load must be biased with a constant voltage, not with a square wave one. To extract the average voltage from the switch I add a low pass filter that gets rid of the switching harmonic and keeps only the zero frequency component.
It is a second order filter, a 1st order one cannot be used because a simple RC the current in the R will produce a large power dissipation, so we use a L instead of the R.

- Addition of (ideally lossless) *L*-*C* low-pass filter, for removal of switching harmonics:
  - Choose filter cutoff frequency f_c much smaller than switching frequency f_s



⇨ Main idea: switch the input waveform, then filter to take the average

At the output of the LP filter, if properly sized, we get a stable voltage depending on the DC.

If everything were ideal, the efficiency is 1.

With

- ideal switches, i.e.: V_{sw} (on) = 0
  I_{sw} (off) = 0
  T_{sw} = 0

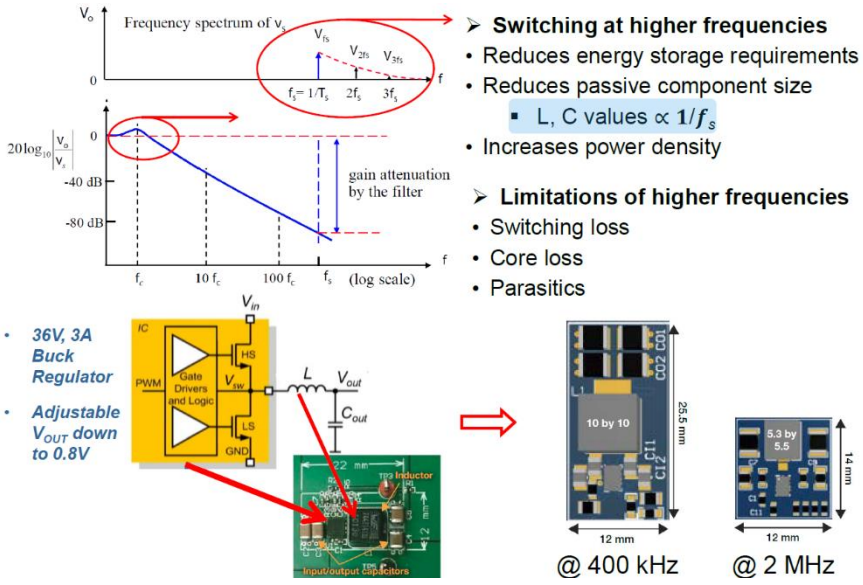
$\Longrightarrow$ $\boxed{\eta = 1}$

- ideal inductor, i.e. $R_L = 0$

- ideal capacitor, i.e. $R_C = 0$

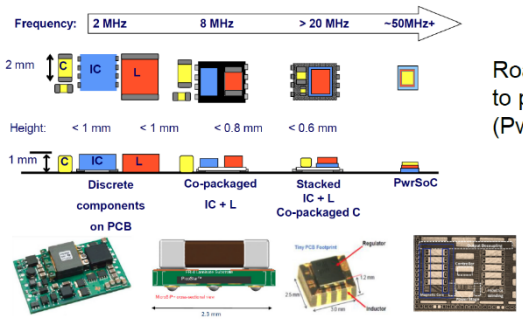
There is another interesting point. Not only the efficiency can be large, but the size of the converter, that is typically limited by the size of the passive components (L and C) can be reduced by increasing the switching frequency. The first plot is the spectrum of the switching (with the harmonics), while the second plot is the Bode plot of the filter transfer function.
If we increase the switching frequency all the harmonics are pushed to higher frequencies, and to get the same attenuation for the first harmonic we can move the cut-off frequency of the filter accordingly.
The corner frequency of a second order filter is fc = 1/(2*pi*sqrt(LC)) and increasing fc means we can decrease the value of L and C, so decrease the size of both the inductor and the capacitor.

This allows also to increase the power density, defined as the power transferred over the area. Increasing the power density is important to keep the electronics compact. So the smaller the switching frequency, the smaller the occupied area.



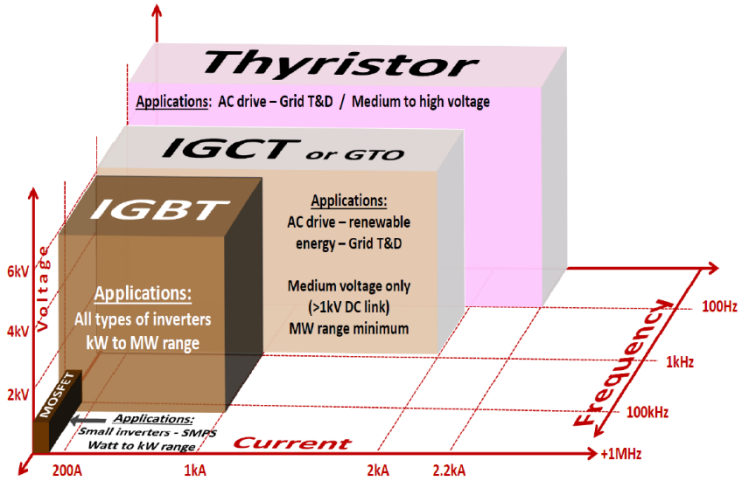
However, increasing the switching frequency we also increase the switching losses of the active components and also in the magnetic cores of the magnetic passive components.
So to get a large efficiency and a small volume we don't have just to increase the switching frequency, but we should also change the topology of the converter.

Switching DC-DC converter: roadmap to PwrSoC

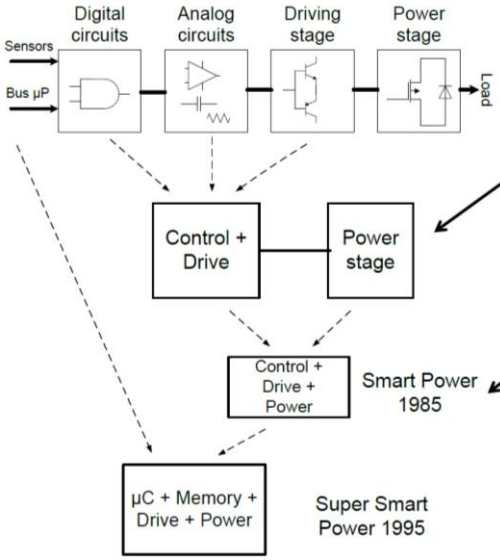


Roadmap for evolution to power supply on chip (PwrSoC) technology

Power semiconductor device capabilities



Evolution of power electronic circuits



For currents exceeding a few amps, it is more cost effective to use discrete power devices with appropriate control ICs.

For currents below ~1 A, it is feasible to integrate power devices and control electronics on a single monolithic chip to provide greater functionality.

IDEAL AND REAL SWITCHES

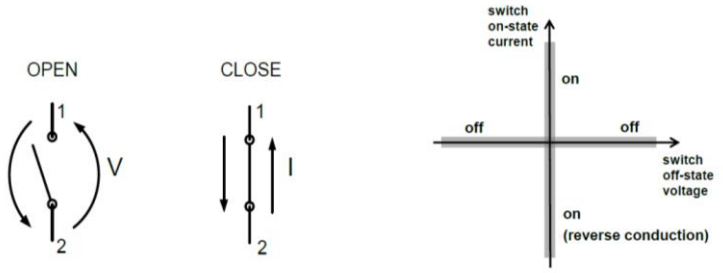
IDEAL SWITCH

SPST (single pole single throw)

Characteristics of an ideal switch

- Zero current leakage when off, while blocking arbitrarily high forward and reverse voltages.
- Zero voltage drop when on, while conducting arbitrarily high currents.
- Instantaneous switch on/off times. Switched at arbitrarily high frequencies.
- Require zero power to control (switch on/off).

In order for a switch to be considered ideal, the 4 requirements in the image must be fulfilled. Moreover, by definition, an ideal switch is a four-quadrant switch. It means that when it is open I can apply to it either a voltage between terminal 1 and terminal 2 or in the opposite direction, and the current must be zero despite the voltage or direction of the voltage. In addition, when the switch is closed, the current must be able to flow in both the directions.



Four-quadrant switch

Active switch: switch state is controlled exclusively by a third terminal (control terminal).

Passive switch: switch state is controlled by the applied current and/or voltage at terminals 1 and 2.

The switch can be real, so it is controlled exclusively by a third terminal (**control terminal**), or passive, where the switch is controlled by applied current or voltage between two terminals (e.g. diodes, while transistors are active switches).

REAL SWITCHING DEVICES

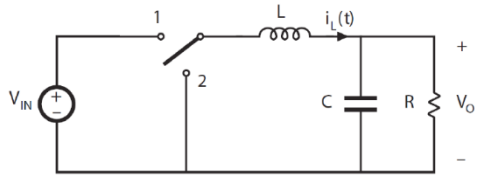
We can implement a switch by using a power semiconductor device. All power semiconductor devices work as a single pole single throw switch (SPST). Of course, there is no key device that is good for any application.

The first thing to do is to understand the application, and then to select the power semiconductor device that suits the application. The best device is the one that meets all the application requirements in terms of size, efficiency and cost-specific current capability.

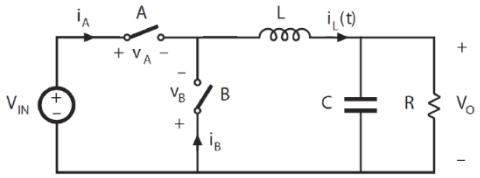
- All power semiconductor devices work as SPST switches.
- Real devices approach ideal characteristics with varying success.
- Each device has its pros and cons:
  - there's no "one-fits-all" device;
  - the best device is the one that meets application needs in terms of size, efficiency and Amps/$ capability.

Practical implementation of SPDT switches

SPDT switch…

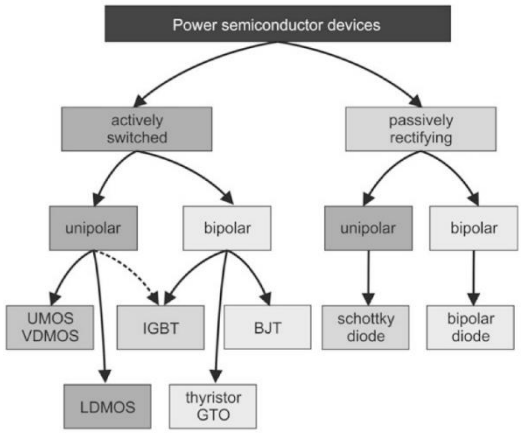


replaced with two SPST switches working in antiphase



SPDT: single pole double throw switch (deviatore). It cannot be implemented with a power semiconductor device, we can have only SPST, so to implement SPDT we need to use two real switches operated in anti-phase.

POWER SEMICONDUCTOR DEVICES CLASSIFICATION



- *"Unipolar"* or *"majority carrier"* devices (i.e., Schottky diodes, power MOSFETs) use only one type of charge carrier.
- *"Bipolar" or "minority carrier devices"* (i.e., thyristors, bipolar junction transistors, IGBTs, PiN diodes) use both majority and minority carriers.

We have a first split in two big categories: active and passive devices. Then each category can be further split in unipolar (majority carrier device: it uses only one type of carrier to sustain the current in the device) or bipolar (minority carrier device: uses both types of carriers to sustain the current, e.g. BJTs). The typical unipolar passive device is the Schottky diode, while the bipolar one is the pn junction.

DIODE

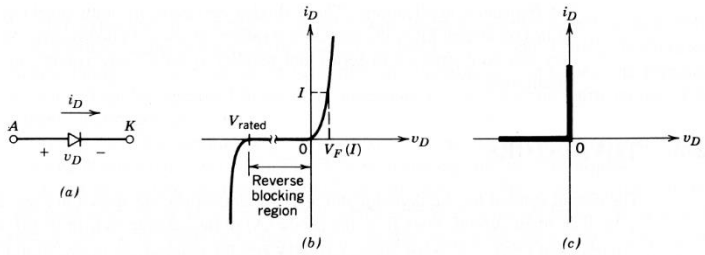


Figure 2-1 Diode: (*a*) symbol, (*b*) *i–v* characteristics, (*c*) idealized characteristics.

- A passive switch
- Single-quadrant switch:
  - can conduct positive on-state current
  - can block negative off-state voltage

Passive switch because the diode acts as a closed switch when the current goes from anode to cathode, while it acts as a closed switch if we apply a reverse voltage. It is a single-quadrant switch and the one in the center is the classical I-V characteristic.

If we apply a reverse bias, in principle there is no current flowing through the diode (except for the leakage current due to the thermal generation in the depletion region, but in power semiconductor devices it can be considered negligible). The diode can sustain the reverse voltage without making any current flow up to the breakdown voltage, from which it start to conduct. We cannot apply any reverse voltage higher than the breakdown voltage.

The last plot is the idealized characteristic of the diode. It conducts any current with zero voltage drop and it blocks any current when reverse biased.

MOSFET

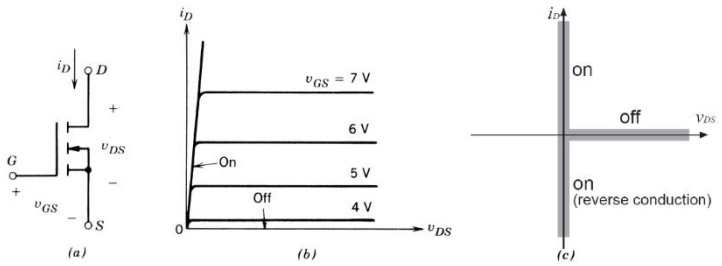


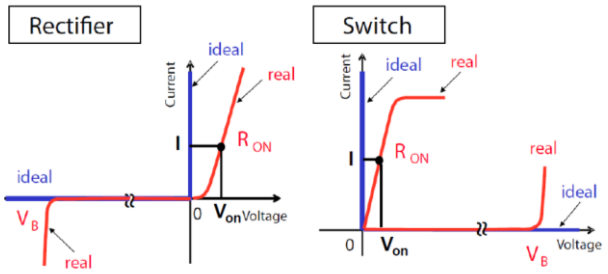
Figure 2-9 *N*-channel MOSFET: (*a*) symbol, (*b*) *i–v* characteristics, (*c*) idealized characteristics.

- An active switch, controlled by terminal G
- Normally operated as single quadrant switch:
  - can conduct positive on-state current (can also conduct negative current in some circumstances)
  - can block positive off-state voltage

The one in the image is an n-channel mosfet and it can be operated as a switch. When we apply a gate to source voltage below the threshold voltage, the mosfet is off (even if in reality we have a small reverse current). If we drive the mosfet into the ohmic region (Vgs and Vds > Vt), the mosfet is behaving as a small resistor (from the channel resistance), so it acts as a closed switch with a small series resistance. it is clearly not an ideal switch because of the resistance.

In principle the mosfet is bidirectional in current (two-quadrant device), because in the ohmic region the current can flow in both directions. However, in most application the mosfet is operated with just one current direction.

Difference between real and ideal power device



- Ideal devices show "zero" voltage drop in the on-state and "zero" leakage current (and "infinite" breakdown voltage) in the off-state.
- Real devices exhibit:
  - a finite semiconductor resistance in the on-state (→ on-state voltage drop or forward voltage drop);
  - a finite leakage current in the off-state;
  - a maximum blocking voltage in the off-state limited by breakdown.

We can see that the ideal device shows zero voltage drop in the on-state region and zero leakage current in the on-state, with infinite breakdown voltage. But the real one has a finite semiconductor series resistance when on, and this resistance, if we increase the current that flows in the switch, causes an ohmic drop producing a non-zero voltage drop between the drain and the source, so the mosfet is dissipating power (we are considering the switch case).

When the mosfet is off, in principle there is zero current, but also the leakage one is so small that it can be neglected. Moreover, a real device cannot block any voltage, there is always a maximum blocking voltage that is determined either by the Vbreak in a diode or the Vbreak of the drain to bulk junction in the mosfet.

We are concerned with the non-idealities because the main contribution to the power dissipation in a converter is coming from the active or passive switches.

Power dissipation (loss) in real switches

- Due to their non-ideal characteristics, all switches suffer from losses:
  - Off-state leakage: when a switch is off, an extremely small current will still 'leak' through the switch. The power loss due to off-state leakage can be assumed to be zero.
  - Forward on-state voltage drop causes **conduction loss**
  - Finite switching times lead to large transient power dissipation, called **switching loss**.
- Finite power required to control the switch, called **base or gate drive loss**.

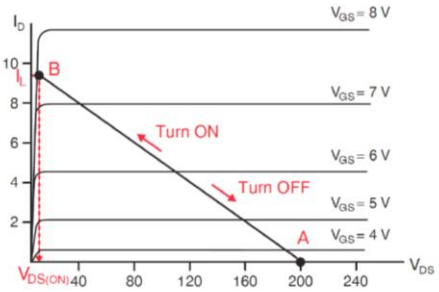
When the device is off, the current in principle is not zero so there is a small power dissipation, but nowadays the quality of the Silicon technology is so high that it can be neglected.

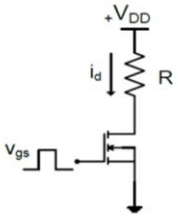
We must be concerned with the forward on-state voltage drop, which causes the conduction loss. If we assume that the switch is closed, there is a constant power dissipation given by I*Von, if I is constant.

Moreover, there is another source of loss coming from the fact that the switched is on and off periodically in a converter, so it has to perform a transition between states. During these transitions there is a lot of energy dissipation → switching loss.

There is also a third loss coming from the finite power required to control the switch → base or gate drive loss.

SWITCH LOSSES





- **Off-State Leakage (point A)**

  Leakage currents are generally so small that the power loss in the off-state can be assumed to be zero.

- **Conduction loss (point B)**

  When the switch is on, there is a slight voltage drop across the switch. $V_{DS(ON)}$ is usually large enough to cause significant power losses.

- **Switching loss (A ↔ B)**

  It takes a finite amount of time to turn on (off) the switch. Finite switching times lead to large transient power dissipation.

Let's consider the circuit on the left, which is used to switch on and off a resistive load. The diagonal black line is the load line; point A is obtained assuming there is no current, computing the voltage drop from source to drain (it is Vdd) and point B is calculated assuming Vds is 0 and so current in the mosfet is Vdd/R.

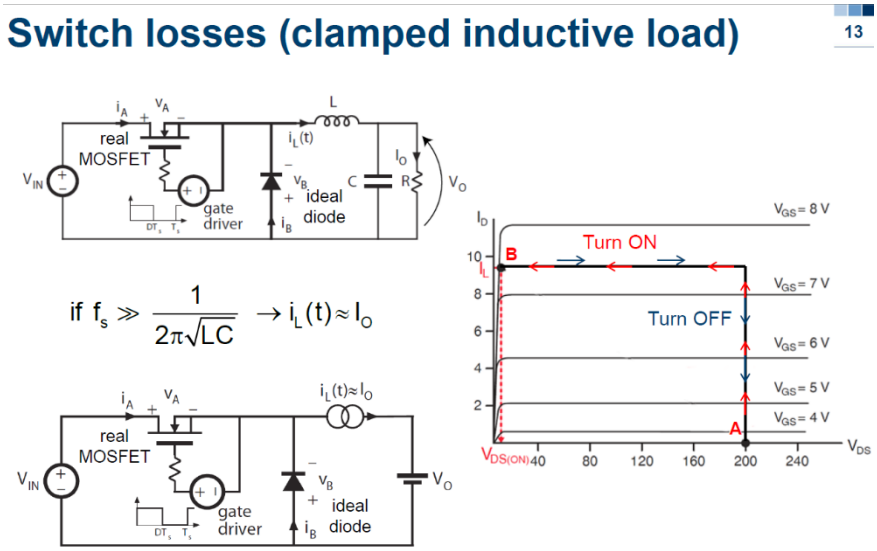
Let's start with the transistor off. There is no current in the load (neglecting the leakage current), so there is no power dissipation in point A. The situation is different in case B. If Vgs > Vt, the mosfet is in the ohmic region and we are in point B where the voltage between drain and source is not 0, but Vds_on; due to the current that is flowing (I_l), power dissipated is I_l * Vds_on and it is called conduction loss.

If the mosfet does a transition between on and off from point A to point B, I'm moving along the load line and the mosfet passes from a situation where we have a non-zero voltage Vds to a non-zero current in the mosfet. So when switching on, the transistor is dissipating energy. This dissipation is the source of the switching losses. If the switch was ideal, there would be no dissipation, because the passage from point A to point B would be instantaneous.

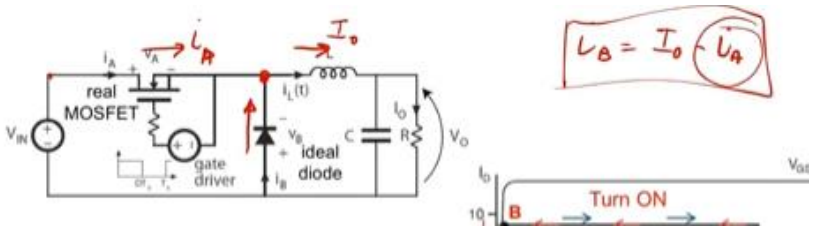
In power electronic, the typical load that we have is an inductive one (to be placed in series to the resistor). When we turn off the mosfet, the mosfet is going to blow, so we need to add a clamping diode (diodo di ricircolo). Something similar happens in converters.

The mosfet is turning on and off periodically, and let's assume that we have reached the steady state. The average inductor current is constant. If fs is much larger than the cut-off frequency we can assume that

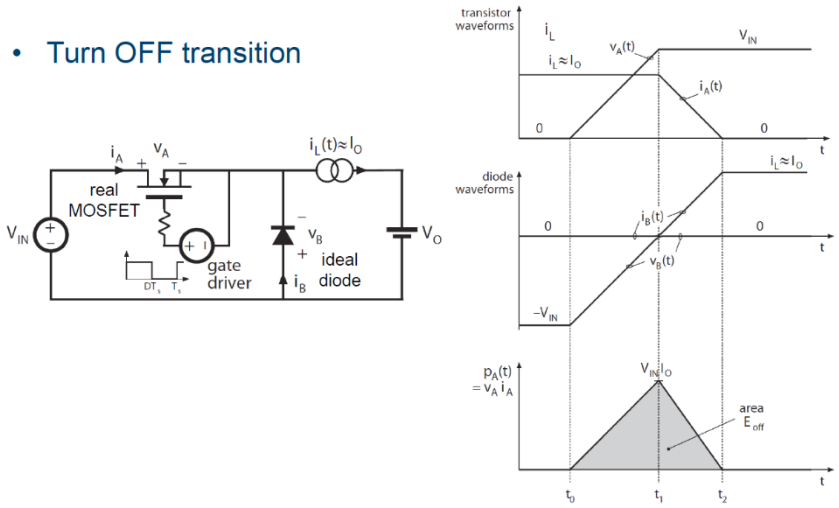
the instantaneous inductor current is almost equal to the current in the load (I0), so constant. Hence we can replace the inductor with a current generator and the capacitor with a battery.



If we assume that the mosfet is off, there is energy in the inductor and the current that was flowing in the inductor is redirected in the clamping diode (I'm in point A).
When we turn on the mosfet, the current that flows in it starts to increase, but the current in the diode is I0 – i_A.

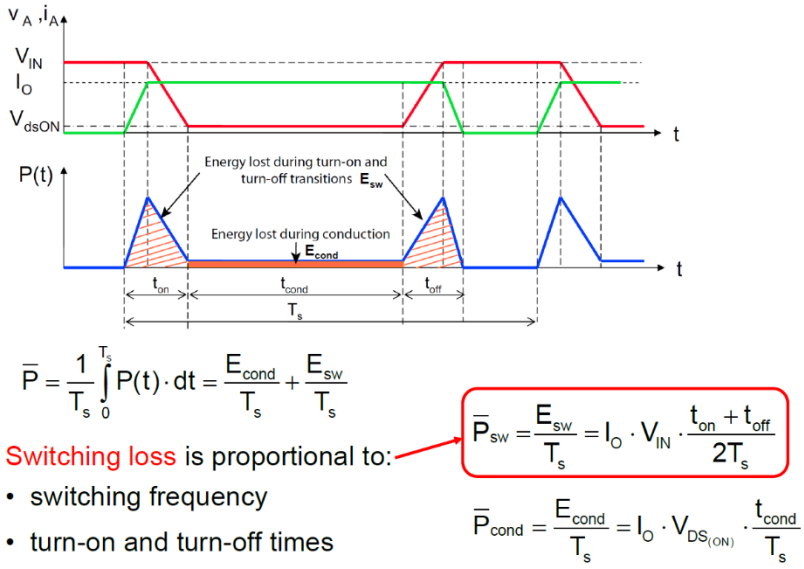


Hence until the current i_A reaches I0, there will be current in the diode. So the mosfet is operating in the plane as if the voltage is still clamped by the presence of the diode (vertical transition). When i_B = i_A, there is no more current in the diode, the diode is disengaged and so the voltage across the mosfet stops increasing.
So the load line for a clamping inductive load is not actually a line, but a square box trajectory.
Of course, the same is valid for the opposite transition. Everything is show in the following plots.



We can see that during the turn off transition, firstly the voltage increases linearly the current stays constant (diode not engaged) and when the diode is engaged the voltage is clamped (first plot). If we plot the instantaneous power dissipation (last plot), we get a triangular instantaneous power dissipation. The area is the power dissipated.

Switching loss



$$\bar{P} = \frac{1}{T_s}\int_0^{T_s} P(t)\cdot dt = \frac{E_{cond}}{T_s} + \frac{E_{sw}}{T_s}$$

Switching loss is proportional to:

- switching frequency
- turn-on and turn-off times

$$\bar{P}_{sw} = \frac{E_{sw}}{T_s} = I_O \cdot V_{IN} \cdot \frac{t_{on} + t_{off}}{2T_s}$$

$$\bar{P}_{cond} = \frac{E_{cond}}{T_s} = I_O \cdot V_{DS_{(ON)}} \cdot \frac{t_{cond}}{T_s}$$

This is the same reasoning of the previous image for the blue line, while the red line is the voltage across the mosfet device, while the green line is the current in the mosfet. The mosfet is initially off, then turned on and off again. This happens periodically and we want to compute the average power dissipated by the mosfet.
To calculate it, we can compute the energy dissipate in a cycle and divide it by the period to have the average power.

We can split the integral in two components: the conduction energy Econd, which is the area subtended by the blue line when the switch is conductive (on) and a second term Esw, energy dissipated in the turn on and off transients.
Esw is computed calculating the area of the triangle. The peak value of the triangle is Vin*Io and the base is t_on or t_off, and I have to put them together. The switching loss is the average power dissipated during the turning on or off transients. We can do the same thing with the conduction loss.

Two important observations:

1. Switching loss is inversly proportional to the switching period → the higher the switching frequency, the higher the switching loss. So to increase the switching frequency without increasing the switching loss I need to reduce t_on or t_off or both.
2. Also the conduction losses are apparently proportional to the switching period, but t_cond/Ts is the duty cycle, which is fixed once decided the application.

t_on and t_off of a mosfet device are typically in the tens of ns range, while the switching frequency is in the order of few hundred of kHz up to few MHz, so much larger than t_on and t_off.

SWITCH REQUIREMENTS

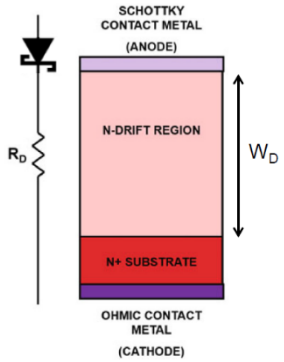
In order to approach ideal characteristics, the following requirements must be met:

- High blocking voltage while OFF
- Low voltage drop while ON ⟹ low conduction loss
- Fast turn-on and turn-off ⟹ low switching loss

These requirements are conflicting!

These three characteristics are conflicting, so we need to make trade-offs.

SCHOTTKY BARRIER DIODE (SBD)



- Unipolar device.
- The drift region is designed to support the blocking voltage.
- Assume a low, uniform doping concentration for the drift region.
- Neglect any junction curvature effects by assuming a parallel-plane configuration.
- Simple 1D analysis can be performed.

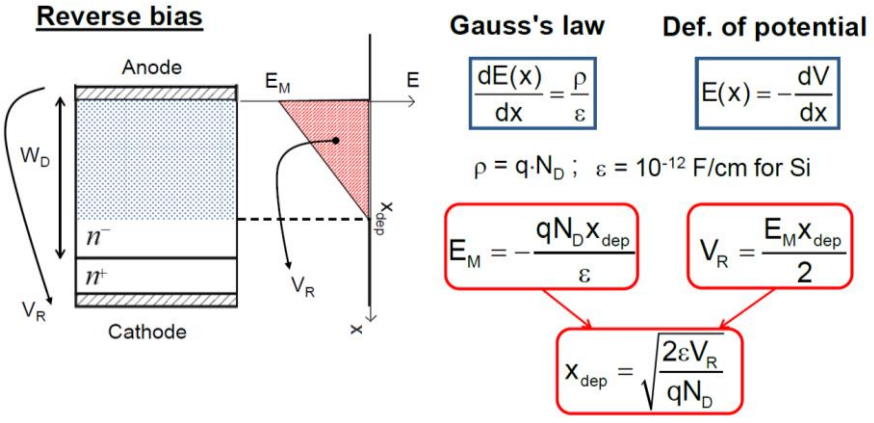
- W_D is the width of the low-doped n region (drift region)
- $N_D \sim 10^{13} \div 10^{16}$ typ.

It is similar to a pn junction, but we have a metal deposited on top of Silicon which is a stack of two regions: a low doped region over which we have the metal and a highly doped substrate. The junction is between the metal (anode) and the semiconductor low doped. We have also a series resistance due to the n-drift region.

It is a unipolar device because the metal cannot inject holes in the n-type semiconductor, so conduction is sustained by electrons. Moreover, the drift region is sized to allocate the depletion region when we reverse bias the diode. The n+ substrate is used to minimize the series resistance towards the metal contact.

We assume that the doping level of the drift region is constant and we will neglect any termination structure (assuming thus a 1D device).

REVERSE BIAS

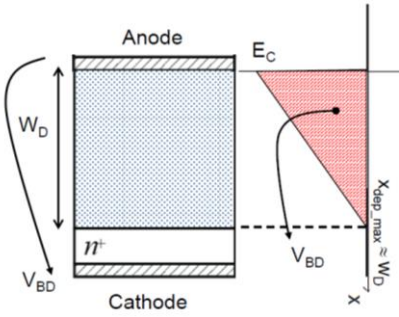


- The solution of Gauss's equation leads to a triangular electric field distribution , within a uniformly doped drift region.
- The slope of the field profile is determined by the doping concentration.

The device is not able to conduct current and it is operated as an open switch. The Gauss law and the potential definition can be simplified in the case of a 1D device.
We notice that the electric field decreases linearly in the depletion region starting from a maximum value that occurs at the boundary between the metal and the semiconductor.

x_dep is the thickness of the depletion layer. V_r is the area of the triangle.
The slope of the electric field is proportional to the doping concentration, the higher, the steeper.

BREAKDOWN



- Breakdown occurs when $|E_M| = E_C$
  - for silicon, $E_C \approx 2.5 \cdot 10^5$ V/cm (weakly dependent on N_D)

$$x_{dep\_max} = \sqrt{\frac{2\varepsilon V_{BD}}{qN_D}}$$

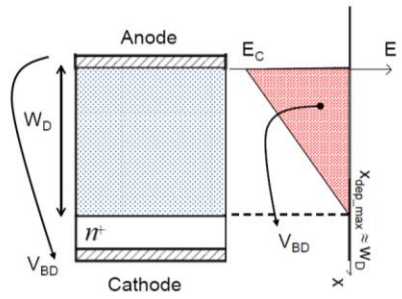
➢ W_D is chosen such that the space charge region barely reaches the n+ layer at breakdown i.e.,

$$x_{dep\_max} \approx W_D \text{ (non punch-through design)}$$

If we increase the reverse bias voltage, the electric field becomes higher and higher until we reach a value so that the junction breaks down. Breakdown electric field is almost constant.
We have a corresponding maximum thickness of the depletion layer with Ec. Vbd is the voltage that we are applying in order for the electric field to reach Ec.

The Shottky diode is designed in such a way that the depletion region almost reaches the n+ region at breakdown.
The rational of this design is to block the voltage.

Design for a target blocking voltage



At breakdown:

➢ $x_{dep\_max} = W_D$

Therefore:

$$V_{BD} = \frac{E_C W_D}{2}$$

$$W_D = \sqrt{\frac{2\varepsilon V_{BD}}{qN_D}}$$

Combining the two equations we get:

$$W_D = \frac{2 \cdot V_{BD}}{E_C}$$

$$N_D = \frac{\varepsilon E_C^2}{2qV_{BD}}$$

Design example

$V_{BD} = 100\text{V}$

$W_D = 8\ \mu\text{m}$

$N_D \approx 2 \cdot 10^{15}\ \text{cm}^{-3}$

From the system we can get the Wd and Nd. Thus we can tune the design of the Shottky diode to clap the voltages we want.

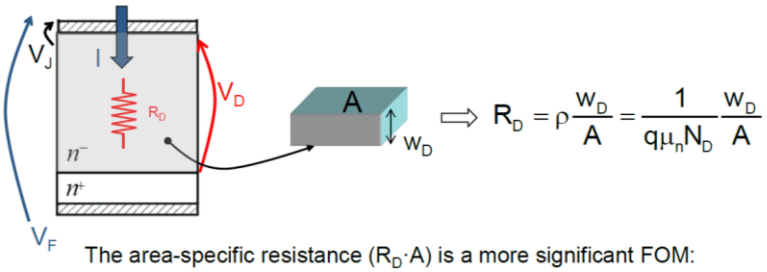
FORWARD BIAS

In this regime we inject a current from the anode to the cathode and the current flows. In forward bias there is a small potential barrier between metal and semiconductor (small means 0.3 or 0.4 V) and also

another contribution. The current that flows in the diode flows also in the drift layer, hence through a resistive path. This generates an additional voltage drop Vd.

Forward bias

- In forward bias, only the small potential barrier between metal and semiconductor material must be overcome (around 0.3 – 0.5 V for Silicon).
- An addidional voltage drop V_D develops across the low-doped base region due to the current flow in the ON condition (forward bias), resulting in $V_F=V_J+V_D$



$$\Rightarrow R_D = \rho\frac{w_D}{A} = \frac{1}{q\mu_n N_D}\frac{w_D}{A}$$

The area-specific resistance ($R_D \cdot A$) is a more significant FOM:

$$R_D \cdot A = R_{ON,sp} = \frac{w_D}{q\mu_n N_D} \quad [\Omega \cdot cm^2]$$

where $R_{ON,sp}$ is referred to as *ideal specific on-resistance of the drift region.*

We can calculate the series resistance Rd by definition, knowing that we are using a block of Silicon. Of course, I cannot increase indefinitely the area, so the figure of merit to use is the area specific resistance Rd*A = Ron,sp. It is an ideal computation because I'm considering a 1D device.

The goal of the designer is to obtain a small Ron,sp. In fact, the smaller the better.

Baliga's FOM

- Ideal specific on-resistance of the drift region

$$R_{ON,sp} = \frac{w_D}{q\mu_n N_D}$$

Replacing $V_{BD} = \frac{E_C w_D}{2}$ and $N_D = \frac{\varepsilon E_C^2}{2qV_{BD}}$ in the above equation we get:

$$R_{ON,sp} = \frac{4V_{BD}^2}{\varepsilon\mu_n E_C^3}$$

The denominator of this equation ($\varepsilon \cdot \mu_n \cdot E_C^3$) is commonly referred to as Baliga's figure of merit (FOM) for power devices.

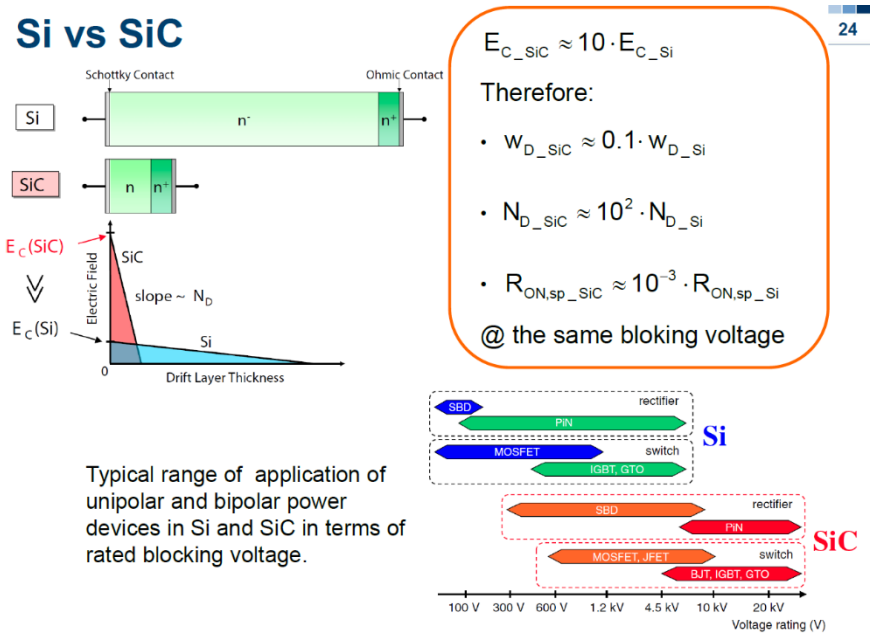
- Baliga's FOM is an indicator of the impact of the semiconductor material properties on the resistance of the drift region.
- The cubic dependence of $R_{ON,sp}$ on E_C favors wide-bandgap (WBG) semiconductors such as SiC an GaN .

	Si	GaAs	SiC	GaN
Breakdown Electric Field (MV/cm)	0.3	0.4	2.4	3.0
Electron mobility (cm^2/Vs) at 300K	1350	8500	370	900
Relative dielectric constant	11.8	13.1	10	9.5
BFOM = $\varepsilon \cdot \mu_n \cdot E_C^3$ normalized to Si	1	17	119	537

The one in the blue box is the Ron,sp computed before. Still using the same equations for Em and Vr seen previously, we can get another Ron,sp definition by substituting in the Ron,sp formula in the blue box.

We notice that the larger Vbd, the larger the specific Ron. Moreover, Ron is inversely proportional to the critical electric field Ec. The product epsilon*mu_n*Ec^3 is called Baliga's FOM.
So at the same blocking voltage I can get a smaller Ron,sp if I use some material having a larger Ec with respect to Silicon. These materials are called wide-bandgap materials, such as SiC and GaN.

Si vs SiC



It is a comparison where the top diode is made using Silicon, and the bottom one using SiC. The devices have the same blocking voltage, so the area of the red triangle is the same of the blue one. The difference is in the Ec value, so in order to have the same area we need to decrease the doping level in the drift region of the Si diode to have it large, which translates in a large diode and large Ron,sp. With SiC we can have a larger doping and thinner device.

To recap, the great advantage of unipolar devices are the fast switch on and switch off times. This means that we can increase the switching frequency without increasing too much the switching loss, because turn on and off occur very fast. This is why mosfet devices can be operated at hundreds of kHz.
The disadvantage is the trade-off that we cannot use a unipolar device if we want to have at the same time a large blocking voltage and a large current, because we would end up with a large Ron,sp, which, combined with a large current, generates a large power loss. So for large currents or high blocking voltages better not to use unipolar devices, unless we change Si with another material.

THE LIMIT OF SILICON

$$R_{ON,sp} = \frac{4V_{BD}^2}{\varepsilon \cdot \mu_n \cdot E_C^3} \quad [\Omega \cdot cm^2]$$

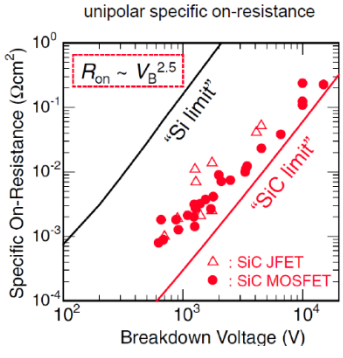
See B.J. Baliga, "Fundamentals of Power Semiconductor Devices" 2nd edition, Springer International Publishing, 2019

Replacing:

μ_n = 1200 cm²/Vs, $E_C = 7.7 \cdot 10^5 \cdot V_{BD}^{-\frac{1}{6}}$ we get:

$$R_{ON,sp} = 5.9 \cdot 10^{-9} \cdot V_{BD}^{2.5} \quad [\Omega \cdot cm^2]$$

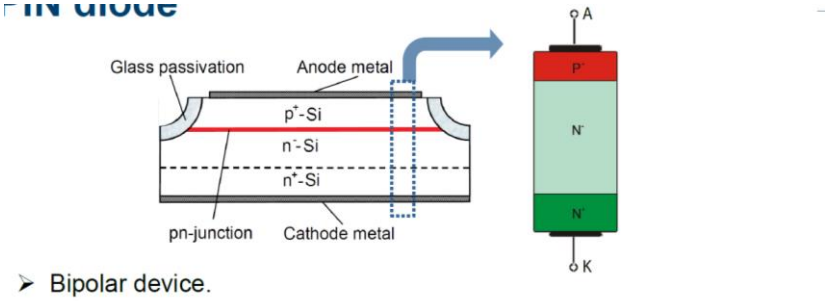
- Unipolar Si device has min $R_{ON,sp} \propto V_{BD}^{2.5}$
- Si typical unipolar devices up to 300 V - 600 V



To overcome the trade-off of Si in the unipolar devices we should use bipolar devices.
Conversely, in the previous image we have replaced the critical electric field Ec in the Ron,sp formula. We have also a formula that relates Ec with Vbd, and if we replace this Ec expression in the Ron,sp and consider u_n = 1200 cm^2/Vs, we get a proportionality between Ron,sp and Vbd only.

The relationship is plot in the right plot. The black line is a fundamental barrier we cannot overcome. The other barrier is the SiC one, and there is a 3 order of magnitude difference between the two limits.

PIN POWER DIODE

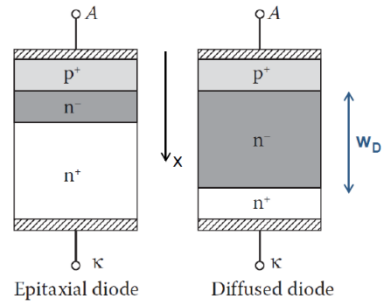


- Bipolar device.
- An n+ layer must be added between the low-doped n side of the p+n junction and the cathode metal layer.
- The middle region is not "i" (intrinsic), but n--type, with a much lower doping level than in the outer regions.
- Termination structures will be omitted from now on (1-D structure)

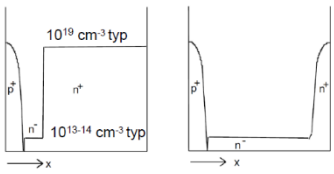
→ p-i-n structure

It is a bipolar device, and the advantage of a bipolar device with respect to a unipolar one is the possibility of exploiting the conductivity modulation, which allows to overcome the Si limit.

The PIN diode is a pn junction diode with an additional layer of Si which is highly doped with acceptors with respect to the Schottky diode. It is a bipolar device because the conduction of current is sustained by both electrons and holes. The n- low doped region is called intrinsic region, but there is no intrinsic Si.



- In epitaxial PIN diodes the *drift region width* w_D ranges from a few μm to tens of μm.
- In diffused PIN diodes w_D is of the order of hundred μm.



There are two different categories of PIN diodes, epitaxial and diffused. The difference is the thickness of the intrinsic layer. In the former it is up to few tens of um, in the latter hundreds of um.
The junction is the p+/n- one and it can be considered as a **step junction** (so across the junction the doping level abruptly steps) **highly asymmetric**. If so, we can use the same equations used for the Schottky diode.

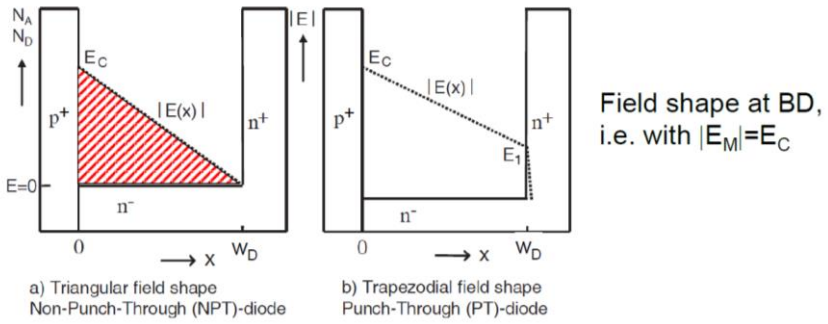
PIN DIODES: NPT AND PT DESIGNS

The plot on the left displays the doping level as a function of the internal position x inside the device. On the same figure there is also the plot of the electric field (red) inside the n- region when we reverse bias the junction. The base layer is the drift layer, and its width determines the blocking voltage.

There can be two possible designs for a PIN diode:

1. Non-Punch Through design: at Vbd, so when the peak E reaches Ec, the extension of the depletion layer barely reaches the n+ layer (as in the Schottky diode).
2. Punch Through design: the electric filed is allowed to penetrate in the n+ layer (right plot) and we have a trapezoidal electric field profile.

The advantage in the PT design is that we can get the same blocking voltage by shrinking the width of the intrinsic layer, having a small Ron,sp.



a) Triangular field shape Non-Punch-Through (NPT)-diode

b) Trapezodial field shape Punch-Through (PT)-diode

Field shape at BD, i.e. with $|E_M|=E_C$

The base width along with the base-doping concentration determines the blocking voltage. Two different cases of the field shape can be distinguished:

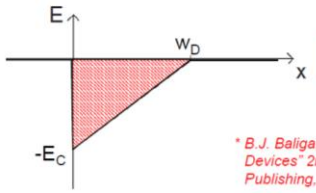
- w_D is chosen such that the space charge does not reach the n+-layer (triangular field shape) → non-punch-through (NPT) diode.
- w_D is chosen such that the space charge penetrates the n+-layer, then the field shape is trapezoidal → punch-through (PT) diode.

NPT Design: blocking voltage

For an ideal NPT diode, w_D is selected such that, at maximum reverse voltage, the end of the triangular field is located at this point.

$$W_D = \frac{2V_{BD}}{E_C} \quad (1)$$

$$N_D = \frac{\varepsilon E_C^2}{2qV_{BD}} \quad (2)$$



$$E_C = 7.7 \cdot 10^5 \cdot V_{BD}^{-\frac{1}{6}} \quad (3)^*$$

* B.J. Baliga, "Fundamentals of Power Semiconductor Devices" 2nd edition, Springer International Publishing, 2019

Plugging (3) in (1) and (2), we get:

$$W_{D(NPT)} = 2.6 \cdot 10^{-6} \cdot V_{BD}^{\frac{7}{6}} \quad (cm)$$

$$N_{D(NPT)} = 1.85 \cdot 10^{18} \cdot V_{BD}^{-\frac{4}{3}} \quad (cm^{-3})$$

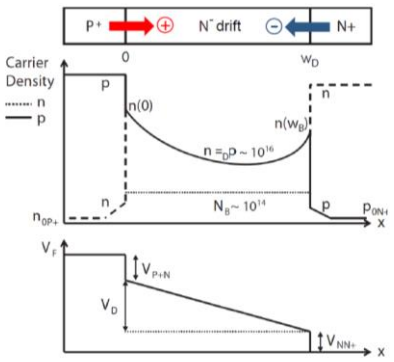
⇒ Large V_{BD}, e.g. > 10^3 V requires:

- w_D > 80 µm;
- N_D < $1.8 \cdot 10^{14}$ cm^{-3}.

The approach is the same as in the Schottky diode. The electric field profile is triangular and we want to design the thickness of the drift layer such that at Ec the depletion layer barely reaches the n+ layer to get the minimum possible Ron,sp. We use the exact same formulas.
In reality, Ec is not constant but it depends on the Vbd voltage as in the red box. Then, by combining all the formulas we get the ones in the yellow box.

With large blocking voltage we need Wd > 80 um with a corresponding doping level as in the image.

PIN DIODE – FORWARD BIAS

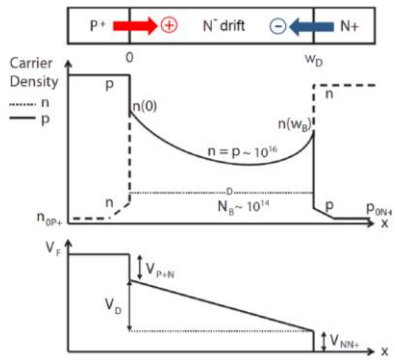


1) When the forward bias applied to the rectifier increases, the injected minority carrier concentration also increases in the drift region until it ultimately exceeds the doping concentration (N_D). This is defined as *high-level injection*.

2) Forward bias injects holes into the drift region from p^+ layer. Electrons are attracted into the drift region from n^+ layer to preserve charge neutrality.
→ *double injection*.

What happens if we forward bias the PIN junction? This is the big difference with respect to the unipolar device.
If we apply a forward bias, we start injecting majority charges from the p side of the junction (holes in this case). Since the doping level of the drift layer is very low, by increasing the forward voltage the concentration of the charges that are injected in the n- layer increases and soon overcomes the concentration of the majority carriers (electrons, 10^14). However, if we forward bias the junction we are injecting an enormous quantity of holes and the concentration soon overcomes the one of electrons in the n- region. This particular condition is called high injection: the concentration of minority carriers is higher than the one of majority carriers (e- in the n- region).

Since the concentration of holes is overcoming the one of electrons, we have a charge imbalance, which is not allowed in the drift layer, so the device is injecting electrons from the n+ side to reach a charge neutrality condition.
In the end we get a drift layer which is full of holes and electrons. The result is that the conductivity of the drift layer, due to the high concentration of e- and holes, is increasing. This mechanism is called conductivity modulation. Of course this happens if we are in a high injection condition (n_a is the concentration of electrons and holes, which is almost the same.



3) If $W_D \leq$ ambipolar diffusion length L_a, carrier distributions are quite flat with $p(x) \approx n(x) \approx n_a$.

4) For $n_a >> N_D$, the resistance of the drift region will be quite small.
→ *conductivity modulation*.

A high concentration of both electrons and holes in the drift region will strongly reduce the voltage drop across this region with respect to the ohmic case where only majority carriers were present.

Conductivity modulation allows us to have a very small resistance, so even if we push a large forward current we don't have a large power dissipation because we have decreased the resistance value.

In the second plot we have the voltage profile inside the device; we notice that the forward voltage has a first drop inside the device caused by the junction, then there is a voltage drop in the drift layer. The total voltage drop in the drift layer is Vd. Then there is also a small voltage drop in the n+/n- junction.

Let's see in detail what happens in the forward condition and which is the resistance of the intrinsic layer.

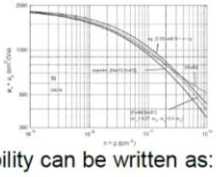
- $I_F = \frac{q(\mu_n + \mu_p) n_a A V_D}{w_D}$ (Ohm's law)

  - No diffusion current in the drift region (n_a=const)
  - V_D=voltage drop across the drift region

- The mobility is dependent on the carrier concentration:

$$\mu_n + \mu_p = \frac{\mu_o}{1 + \frac{n_a}{n_b}}$$

$n_b \approx 10^{17}\ cm^{-3}$ is a fitting parameter



- The forward diode current including density-dependent mobility can be written as:

$$I_F = \frac{q\, n_a A V_D}{w_D} \frac{\mu_o}{1 + \frac{n_a}{n_b}}$$

- The latter equation can be inverted to calculate V_B/I_F, assuming $n_a >> n_b$

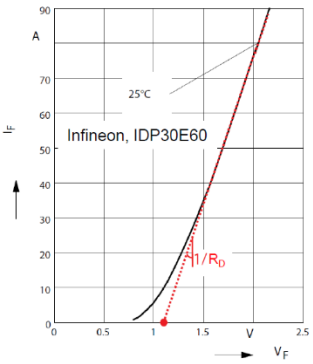
$$\frac{V_D}{I_F} = R_D = \frac{w_D}{q\, \mu_o n_b A}$$

To find the resistance value, we assume that:

- There are no diffusion currents in the drift layer. This is true if we assume a flat concentration profile. We are applying the Ohm's law in the drift layer: J = sigma*E, where sigma is the conductivity. Then E = Vd/Wd, and J = I/A.
- Mobility depends on carriers concentration: if we increase the charge concentration in the drift layer we have carrier to carrier scattering and hence a reduction of the mobility. It has a decreasing shape that is described by the second formula. u_0 is a constant term, n_a is the concentration of electrons and holes in the conductivity modulated layer, while n_b is a fitting parameter for the curves.
  Then we substitute in the first expression and we get the third one. The third one can be inverted to get Vd/If, which is the resistance of the conductivity modulated layer (drift layer).

  If we assume that n_a >> n_b, the simplified result is in the blue box; the resistance of the drift layer, in conditions of conductivity modulation, can be written as above.

In the next plot we have the I-V characteristic of a PIN diode. Initially it increases as an exponential, but then it is described by a simple resistive one, because at large currents the series resistance of the drift layer becomes dominant, so current increases linearly. The reciprocal of the slope of this characteristic is exactly Rd. **Hence at large forward currents the drift region behaves as an ohmic conductor having a conductivity of q*u_0*n_b**.



$$R_D = \frac{V_D}{I_F} = \frac{w_D}{q\, \mu_o n_b A}$$

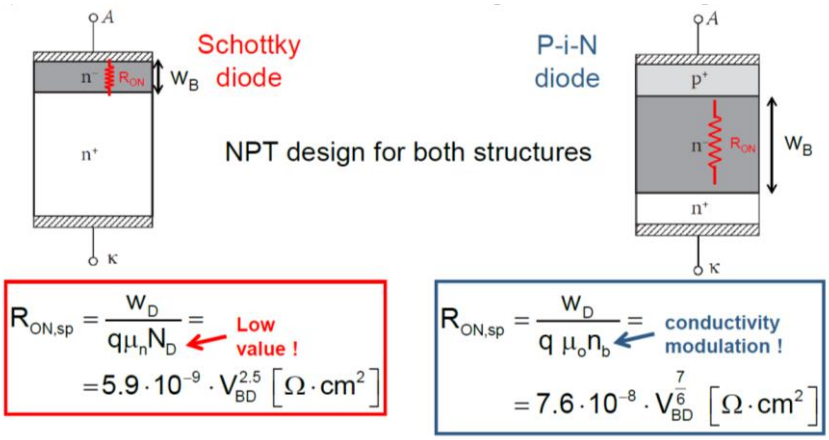
- At large forward currents, the drift region behaves like an ohmic conductor having a conductivity of $q\mu_o n_b$.
- The voltage drop across R_D is a significant contribution to total forward diode voltage:

$$V_F = V_{JP+N} + V_{JNN+} + I_F R_D.$$

The voltage drop across Rd is hence a significant contribution to the total forward diode voltage, and we want to minimize this contribution to minimize the conduction losses. To do so we need to increase n_a.

Our analysis will be done not on Rd, but on Ron,sp = Rd*A = Wd/q*u_0*n_b.

SPECIFIC ON RESISTANCE: UNIPOLAR VS BIPOLAR

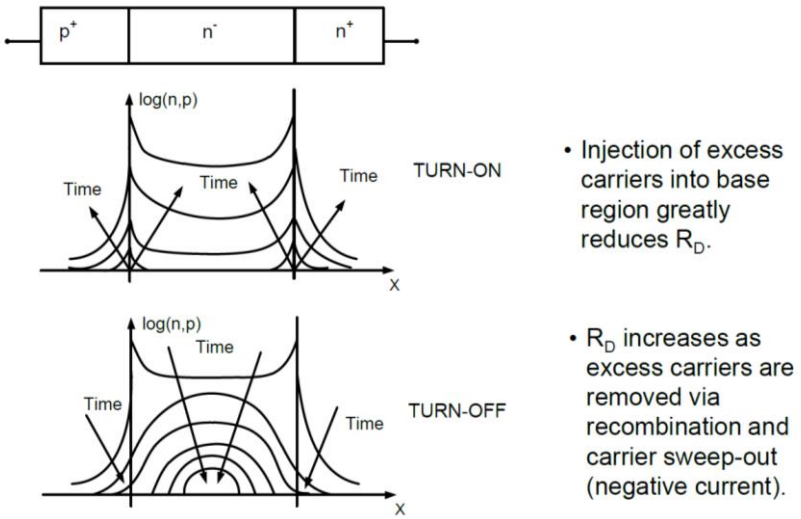


- These equations clearly show that as breakdown voltages increase, bipolar devices and majority-carrier devices suffer reductions in their current-carrying capabilities. However, the reduction in bipolar devices is less severe compared to majority-carrier devices.

We are comparing the figure of merit for a unipolar device with the one for a bipolar device.
The Ron,sp for the unipolar one is depending on Nd, which is the doping level. The PIN diode has a similar sxpression, but the difference is in the n_b term.
In a unipolar device Nd is in the order of 10^15, while n_b is at least 10^17. Hence if the sizes are the same, a bipolar device shows a Ron,sp which is a couple of order of magnitude smaller than the corresponding unipolar device. This difference comes from the conductivity modulation that is allowed by the presence of both n type and p type charge carriers.

This conductivity modulation happens in the same way in BJT.

PIN diode: turn on and turn off

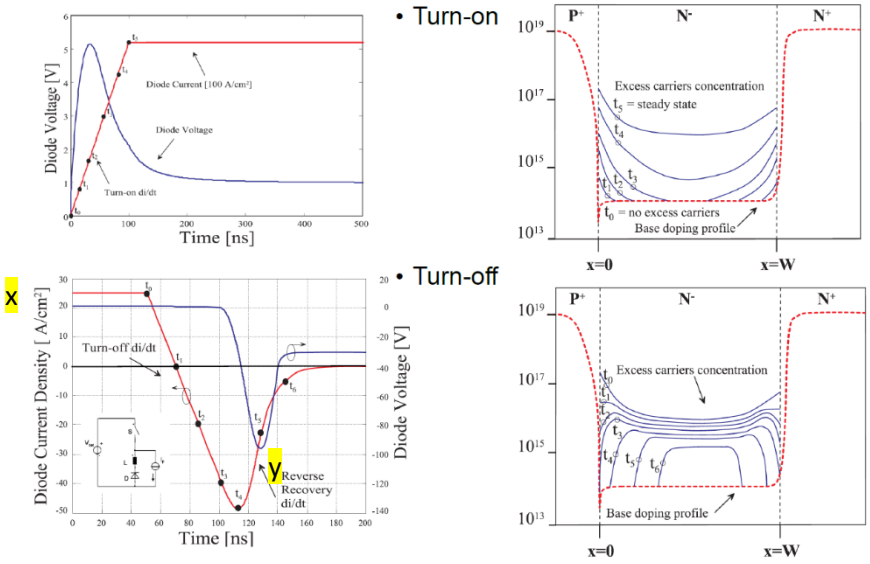


- Injection of excess carriers into base region greatly reduces R_D.
- R_D increases as excess carriers are removed via recombination and carrier sweep-out (negative current).

The price to pay for a small Ron,sp is that the switch on and switch off times are much longer for bipolar devices than for unipolar device, and this is a consequence of the charge accumulation in the drift layer. We are flooding the drift layer with charges, holes and electrons when we turn on the device, and when we turn it off we need to get rid of them, and this takes time.

The consequence is that <u>we cannot use a bipolar device at high switching frequencies</u>.

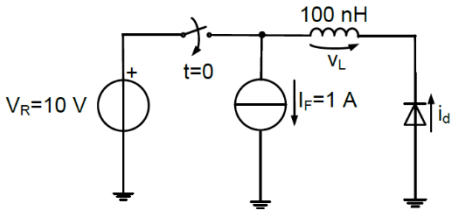
The situation is worse in the turn off transition because we start with a lot of holes and electrons in the drift layer and we cannot say the device is off until we get completely rid of these charge carriers.
The results of the simulations are in the following curves.



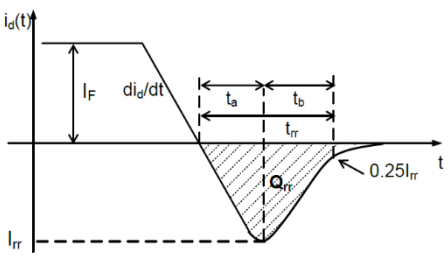
To justify these curves we must assume a small inductance in series with the diode. Initially the diode is driven with a constant current in the forward bias region (current red is flat, plot x) and at time t0 the switch is closed and I reverse bias the device.
When we turn off the device the current starts to decrease with a constant slope determined by the inductance up to t1 when it is 0. But the diode is not off, because there is still charge in the drift layer that must be extracted with a negative current. This is the so-called reverse recovery transient. The accumulated charge in the drift layer can sustain a negative current.
When the current returns back to zero we have removed all the charge carriers. The reverse recovery transient is a problem because we need for it to be over to have the device off and it is a transient also responsible for an additional power dissipation.

The current reaches the peak (valley) exactly when the voltage across the diode is equal to the value y.

Turn-off: di/dt



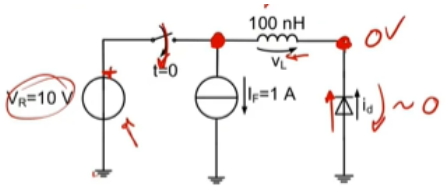
$$i_d = i_L$$
$$v_d \approx 0$$
$$v_L \approx -V_R = -10V$$
$$\frac{di_d(t)}{dt} = \frac{di_L(t)}{dt} = -\frac{V_R}{L} = -100\,A/\mu s$$
$$i_d(t) = I_F - \frac{V_R}{L}t$$



We have a PIN diode with a parasitic inductance in series.

Initially the switch is off and there is a current generator that pumps a current of 1A in the device. At t = 0 the switch is closed applying 10V, so we are reverse biasing the diode. However we have the inductor between the voltage generator and the diode.

We can assume that as far as there are accumulated charges in the drift layer, the voltage drop across the diode is negligible, more or less 1V, with respect to 10V. So Vd initially across the diode is 0.
Current in the diode is the same that flows in the inductor. So at t = 0 we have 10V on one side of the inductor and 0V on the other side because the diode keeps the voltage to more or less 0V.



We can use the relationship for the inductor. Since Vl is constant and equal to 10V we can easily calculate di/dt, which is negative and equal to -Vr/L (r = reverse). So the slope of the current is fully determined by the value of the inductor and the reverse bias we apply.

As for the voltage, in the previous image the blue line was the voltage and we noticed that it was around zero and it stays 0V until we get close to the peak of the reverse current. This because as far as there are holes and electrons in the drift layer, the diode can still be considered to be forward bias. The voltage drop becomes consistent after we have overcome the minimum of the current.

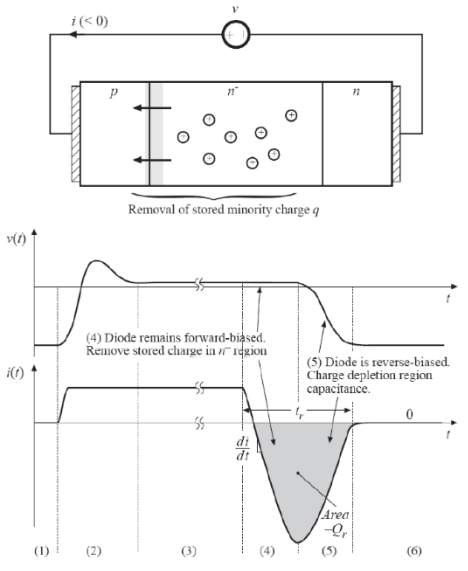
From a qualitative standpoint, the current reaches zero and becomes negative, reaching a valley (**peak reverse recovery current**) and then goes back to zero. The area subtended by the reverse recovery transient is called reverse recovery charge, and this charge is related to the charge accumulated inside the drift layer.

The duration of the recovery transient in principle is infinite because we have an exponential recovery, but to set a time we typically consider the time from the point where the current is 0 to when it reaches 25% of the peak value (trr). trr can be split in two parts:

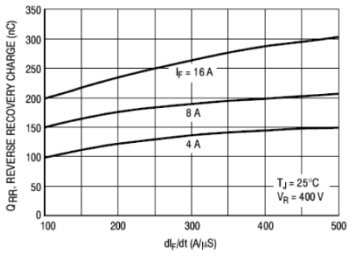
- t_a: time needed for the reverse transient to move from the zero crossing to the peak.
- t_b: time needed for the recovery to move from the peak to the conventional stopping point.

Reverse recovery charge

The reverse recovery charge Qr depends on several parameters.



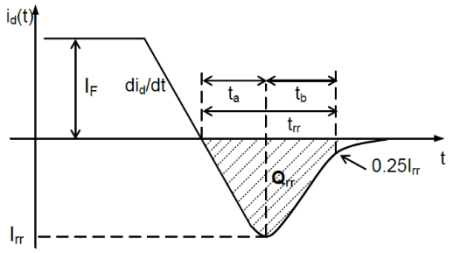
- Q_r depends on diode on-state forward current, and on the rate-of-change of diode current during diode turn-off transition



In the plot on the bottom right there is Qr as a function of the slope of the current that we are using to turn off the device, and the slope of the current strongly depends on how we designed the circuit. Moreover, there is a strong dependance also on the initial current. The larger the current, the larger the quantity of charge carriers in the drift layer and so the larger the Qr.

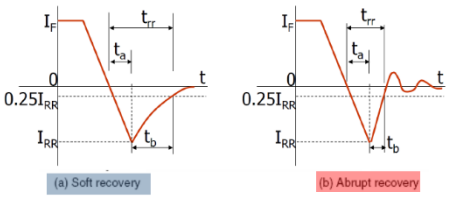
Qr increases if we increase the slope of the current. In fact, if we assume the initial current is 4A, if we increase the slope from 200 to 400, Qr increases.

As for the relationship between Qr, trr and di/dt, we have the following (we simply compute the area of the triangle).



$$t_{rr} = \sqrt{(1+RSF)\cdot\frac{2Q_r}{di/dt}}$$

$$I_{rr} = \sqrt{\frac{2Q_r}{(1+RSF)}\cdot\frac{di}{dt}}$$



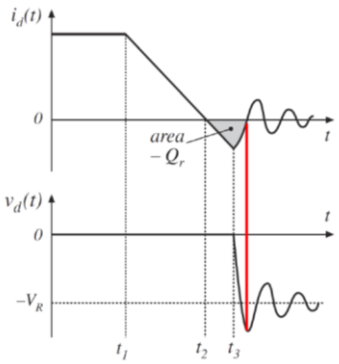
Recovery Softness Factor

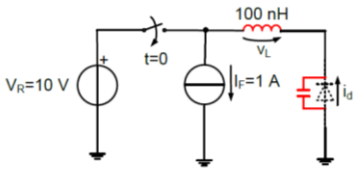
$$RSF = \frac{t_b}{t_a}$$

We can define the recovery softness factor as the ration between, t_b and t_a. if the RSF is large, the device has a soft recovery, so the t_b duration is larger than t_a and it is call soft because the current softly goes to ground in relatively a lot of time.
If RSF < 1, we have an abrupt recovery.

In principle a abrupt recovery seems better than a soft one, bit this is not always true, because one drawback of the abrupt is that we are exciting the parasitic inductances in series with the diode, producing ringings, and ringings ends up in a lot of setting time.

Abrupt recovery: ringing





- The inductor and capacitor then form a series resonant circuit, which rings with decaying sinusoidal waveforms.
- This ringing is eventually damped out by the parasitic loss elements of the circuit, such as the inductor winding resistance, inductor core loss, and capacitor equivalent series resistance.

The problem of ringing comes from a parasitic capacitance across the diode, which is the junction capacitance. The ringing comes from the interaction of the parasitic inductor and parasitic capacitance.

If we look at the voltage drop across the diode, the current reaches zero when the voltage across the diode, which is also the voltage across the parasitic capacitance, is not equal to the steady state value -Vr. So we are in an unbalanced situation, because in the capacitor we have an additional energy stored with respect to the steady state. This energy starts to bounce back and forth between the capacitor and the inductor producing ringings.

TYPES OF POWER DIODES

Standard recovery

- Reverse recovery time not specified, intended for 50/60Hz

Fast recovery and ultra-fast recovery

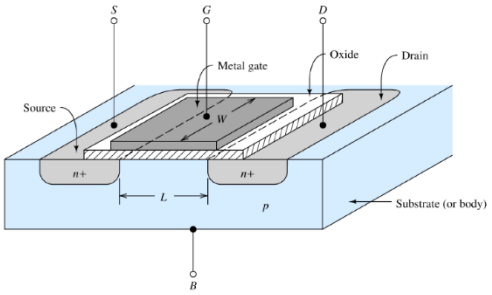
- Reverse recovery time and recovered charge specified
- Intended for converter applications

Schottky diode

- A majority carrier device
- Essentially no recovered charge
- Model with equilibrium i-v characteristic, in parallel with depletion region capacitance
- Restricted to low voltage (few devices can block 100V or more)

We are interested in fast and ultrafast recovery, not in the standard one. These devices are good for converters.

POWER MOSFET TRANSISTOR



Saturation region

$$I_D = \frac{1}{2}\mu_n C'_{ox}\frac{W}{L}(V_{GS}-V_T)^2$$

Ohmic region

$$I_D = \mu_n C'_{ox}\frac{W}{L}(V_{GS}-V_T)V_{DS} - \frac{V_{DS}^2}{2}$$

$$R_{DS(ON)} = \frac{V_{DS}}{I_D} = \frac{1}{\mu_n C'_{ox}\frac{W}{L}(V_{GS}-V_T)}$$

High current capability (i.e. low $R_{DS(ON)}$) requires:

- large W (→ area !)
- short L (→ technology; punchthrough)

Difficult to achieve high blocking voltage and low $R_{DS(ON)}$

In power electronics we don't want to use the mosfet in saturation because the power dissipation would be too high. We want to use it as a switch, either on or off. It is off if Vgs < Vt, otherwise it is in the ohmic region.
In the ohmic region the relationship between the drain current and the gate to source voltage is a linear one, provided that Vds^2/2 can be neglected.

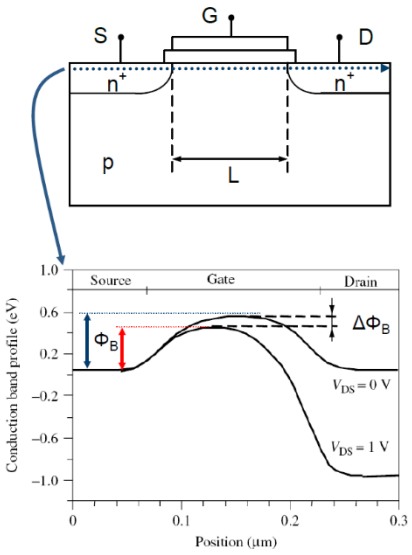
Rds,on is the resistance shown by the device in the ohmic region and it has to be as small as possible in order to dissipate the least possible power.
To have a small Rds,on = Vds/Id we should have large W and small L (in the formula u_n is the surface mobility, not the bulk mobility), and the oxide thickness cannot be decreased too much. Moreover, the Vgs cannot be increased too much because we apply a larger electric field and the oxide might break down.

Of course, W cannot be increased indefinitely because we would use too much Silicon. Furthermore, if we decrease L we cannot decrease it too much because we would reach the punchthrough.
In general, it is difficult with a classical mosfet structure to achieve high blocking voltage and low Rds,on at the same time. And this is a problem because we want both.

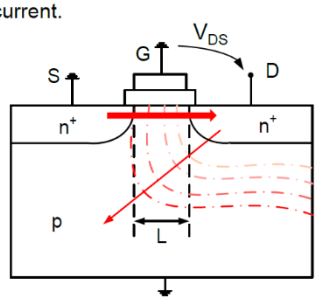
PUNCHTHROUGH

The one in the image is the cross-section of the standard mosfet and let's assume that we want to apply a given voltage. We put the source at ground and also the gate and we apply a positive voltage to the drain. We are in this way reverse biasing the drain to body junction. If so, we are expanding the depletion layer; the larger the Vds, the more expanded the depletion layer.

When the depletion layer reaches the source, the current starts to increase from the source to the drain even if we are working below threshold. Considering the band diagram of the mosfet structure (bottom left image), at thermal equilibrium this is the band diagram. We have an energy barrier that prevents electrons to overcome it and moving to the drain. But if we increase the drain voltage, we are shifting down the bands at the drain side and if we apply high voltages we are reducing the barrier up to the point where there will no more a barrier.
If we decrease the length to decrease Rds,on, this is the effect. If I decrease the length and increase the doping to avoid the deformation, we eventually reach the breakdown of the drain to body junction.

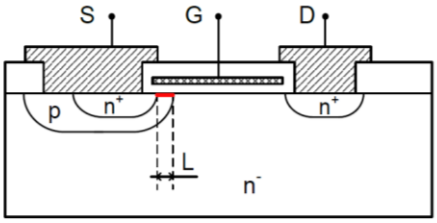


- Punchthrough is observed when the depletion region of the body-drain p-n junction reaches the source region at drain voltages below the rated avalanche voltage of the device.
- The potential barrier between source and drain will be lowered and carriers will start to move from source to drain, causing a rapidly increasing current.

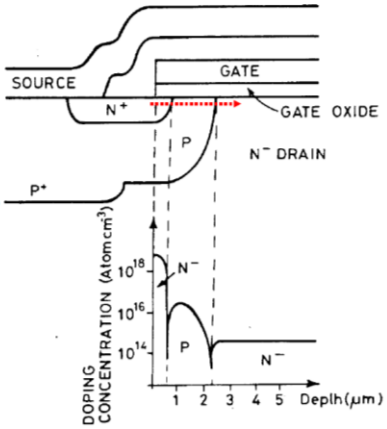


LATERAL DOUBLE-DIFFUSED MOS TRANSISTOR (DMOS)

The solution is, instead of using the standard structure we use this one.



- The device is named DMOS transistor due to the simultaneous double diffusion of source and channel dopants to fabricate a short channel device.
- L determined by the difference in lateral extension of the P (body)and N+ (source) regions produced by their diffusion cycles
- L_{typ} from < 1 μm to 2 μm
- body and source short circuited
- body-drain depletion region extends mainly into the drain → no punchthrough



From a functional point of view the equations are the same and also the behaviour, the only difference is the structure, we can apply high voltages without encountering the punchthrough condition.
We notice that we have a n channel mosfet and the body region is a p diffused region inside the n doped structure, and the source is a n region diffused inside the p region.
It is called double diffused because the body diffusion and the source diffusion are obtained in the following way. We have the substrate and the oxide over it. We open a window in the oxide and we diffuse firstly donor atoms (e.g. Arsenic) in the substrate and then, with the same window we diffuse and implant acceptors (e.g. Boron). After this, we apply a thermal treatment. What happens is that the dopants start to diffuse and we get the final structure.

What happens is that Boron atoms are very small, so the diffusion coefficient of Boron in Silicon at room temperature is very high, so it is diffusing faster. Conversely, Arsenic atoms are very large, so the diffusion coefficient is very small at room temperature. With the thermal treatment, the body region extends laterally more than the source region, because Boron moves laterally. The extension of the body region with respect to the source region defines the channel length.

The other important feature of this structure is the gate, that is a polysilicon gate and then laterally there is another n+ diffused well for the drain. It is a planar device, the three contacts are on the same side of the Silicon wafer.
Moreover, the source contact is contacting at the same time the source and the body, so there is an intrinsic body to source shortcircuit.

There are two reasons for this:

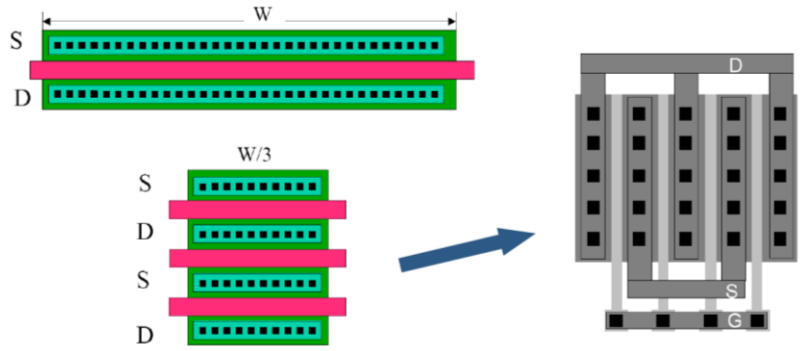
1. Reduce the body effect.
2. Avoiding the turn on of a parasitic diode.

The DMOS is more effective to implement a power device, and to understand this we can look at the plot on the right.
We can see the plot of the doping concentration as a function of the position moving along the x axis. We start from the source, where we have a very high doping, then we have the body region and then we have the drain, which is very low doped.

When we apply a voltage to turn on the mosfet, we apply a positive drain to source voltage, and the junction that sustains the voltage applied is the body to drain junction. However, since the body region is much more doped with respect to the drain region, most of the depletion layer extends in the drain region. Hence we prevent the punchthrough. To sustain even higher blocking voltages we simply have to move the drain contact further and further away from the gate.
The channel length still remains the same, we are just changing the distance between the drain contact and the channel.

So the channel length is determined by different lateral diffusion of the dopands.

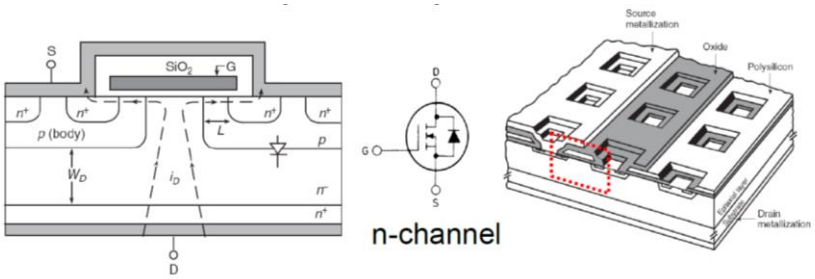
Lateral DMOS with large W/L ratio



- Split the transistor with large W/L ratio into the parallel of n elementary transistors with reduced W/L (stacked layout)
- Drain metallization on top: inefficient use of silicon

In a power mosfet power device the form factor is very large which is not obtained designing a long W on the wafer, what is typically done is splitting the device into sections and put them in parallel.
The metal is on top of the Silicon and this design is good if the form factor is not high, because for form factors greater than 1000 this design (on the right) doesn't work because we use too much Silicon. So what we can do is to move the drain contact from the surface down to the Silicon bottom.

VERTICAL DMOS TRANSISTOR



- Drain contact moved to the backside.
- Lightly doped drain drift region. W_D determines blocking voltage rating. Typically: 8 µm for 60V MOSFETs, 60-70 µm for 600V MOSFETs.
- Division of source into many small areas connected electrically in parallel. Maximizes gate width-to-channel length ratio.
- Source metallization covers the whole surface area. Parasitic BJT held in cutoff by body-source short

It is the so-called cellular structure, it includes many single identical cells which are working in parallel.
A cell, looking from the surface of the Silicon, is a second diffusion inside the body diffusion that goes all around the perimeter of the p body diffusion and with a hole in the middle.

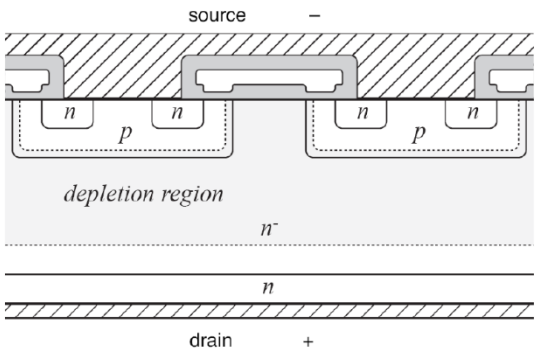
In the top left image we have the cross-section of two contiguous cells. The channel is all around the perimeter of the cell, and still it is defined by the different diffusion of dopands. The gate is a polysilicon gate.

When we turn on the mosfet with a Vgs > Vt we are forming a conductor channel, we are injecting electrons from the n+ layer into the drain and these electrons move from the top to the bottom of the Silicon wafer and are collected by the drain.
By doing this we are sparing some area because one of the contact is below. The other advantage is that cells are made with a p type diffusion inside an n- layer. The reason for this low n- doped layer is that the blocking voltage is determined by the doping of this layer and its thickness (epitaxial layer).
In fact, if we want to block the voltage and the device is in the off condition, we are applying a positive Vds, Vs is grounded and also Vg and so we are applying a reverse bias between the body and the drain region. The larger the voltage we want to block, the larger the layer.

This structure has a cellular shape with multiple cells because we want to increase the overall width of the mosfet at a given area.
The last important feature of the VDMOS is the source metallization. The source metal completely covers the surface of Silicon, it contacts both the source and the body.

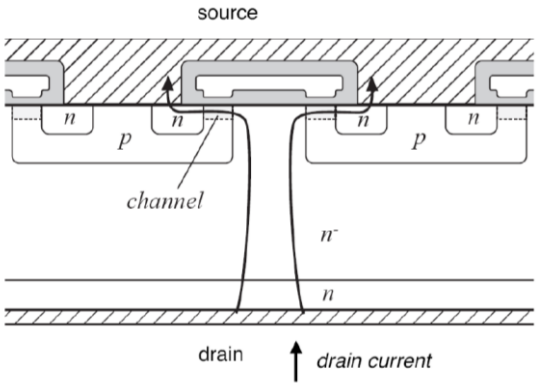
VDMOS: off state

Vgs < Vt. We are generating a depletion layer inside the epitaxial layer that is larger the larger the blocking voltage. The higher the blocking voltage, the lower the doping, the higher the thickness and so the larger the resistance, it is the same situation of the Schottky diode.



- p-n^- junction is reverse-biased
- off-state voltage appears across n^- region

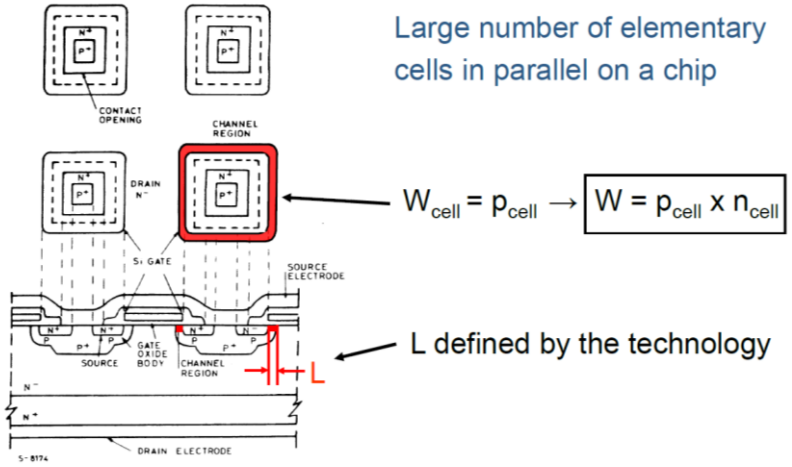
VDMOS: on state



- positive gate voltage induces conducting channel
- drain current flows through n^- region and conducting channel
- on resistance = total resistances of n^- region, conducting channel, source and drain contacts, etc.

We apply a positive Vgs > Vt, the mosfet enters the ohmic region, a channel is formed, electrons are injected from the source to the drain. Current is from bottom to top because of convention.

VDMOS: W/L



W/L is established by including a suitable number of cells

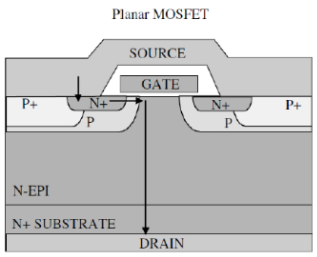
To decrease the Rds,on we want to increase the width of the channel without using too much area.
We have the cellular structure and considering 4 cells. The mosfet includes a large number of cells. What is the channel width of a single cell? It is all around the perimeter of the cell, but if we have a number n of cell, since the cells are operated in parallel, the total channel width is the perimeter of a single cell multiplied by the number of cell. Increasing or decreasing the number of cells we put in parallel we change the value of Rds,on.
We cannot do anything about the channel length.

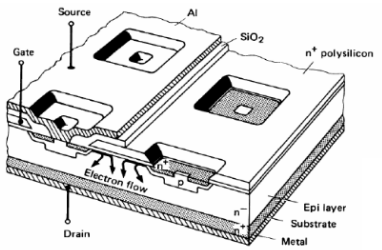
Of course, the larger the form factor the smaller Rds,on, and this can be done increasing the number of cells.
Cells are operated in parallel because the source, gate and drain contacts are the same for all. Furthermore, the cells can be square or hexagonal (to obtain the maximal channel width in a given Silicon layer).

Discrete and integrated VDMOS

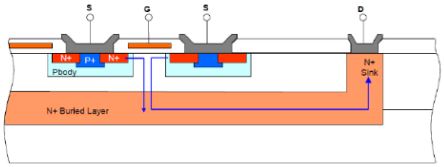
There are also VDMOS that are integrated in Silicon chips with other analog and digital circuits. This can be done with BCD technologies.
In this case, the top structure of the VDMOS is the same, but the difference is in the drain. Below the body cells there is a highly doped n+ buried layer (orange); so electrons are injected, they firstly go

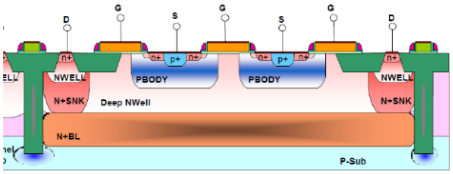
vertically down to the buried layer and then they move horizontally and are collected by the drain contact, which is on the top surface. We have to do this if we want to put on the same silicon chip some power mosfet device and analog and digital circuits.





Discrete





Integrated (BCD technologies)

Drain-source on-state resistance: Rds,on

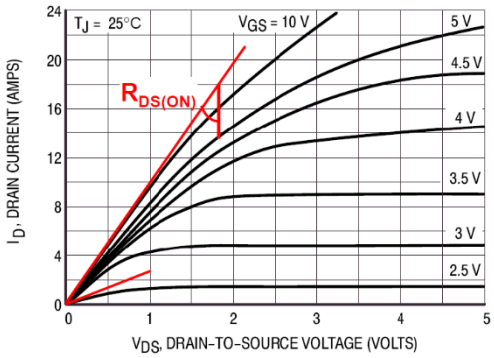
The most important thing is to get a low voltage drop between the terminals of the switch when the switch is on to minimize the conduction loss.

If I consider a mosfet device, this means that I have to reduce the Rds,on.

NB: in a traditional mosfet we have an intrinsic diode coming from the body to drain junction, which is a source to drain diode because the body is connected with the source.

To reduce the Rds,on we can have two possibilities:

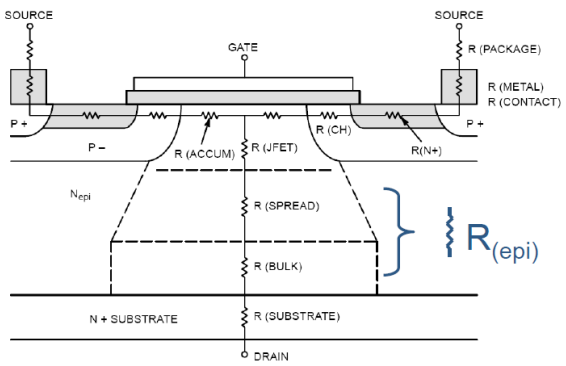
1. Increasing the Vgs, but we cannot do this indefinitely or we breakdown the silicon.
2. Increase the number of cells. If we take the Rds,on of a single cell, since the cells are in parallel, the overall Rds,on is the parallel between the Rds,on. Of course there is a drawback, in fact the larger the number of cells, the larger the area

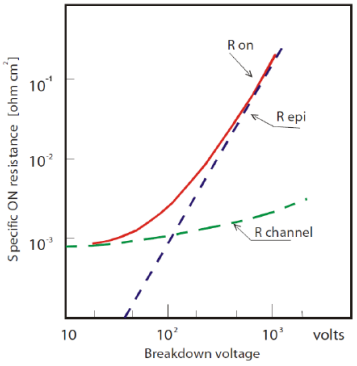


- Unipolar conduction
  - no conductivity modulation
- $V_{GS}\uparrow$; $R_{DS(ON)}\downarrow$
- Area $\uparrow$; $R_{DS(ON)}\downarrow$
  - $R_{DS(ON)} = \frac{R_{DS(ON)}[cell]}{n_{cell}}$

The product $R_{DS(ON)}$ x area ($\Omega \cdot mm^2$) is a key figure of merit

In order to see whether we can play with Rds,on we have to look at the structure of the VDMOS (left image) and identify all the contributions to the series resistance that an electron sees by going from the source, in the channel and then in the drain.
There are many contributions. The first one is the resistance of the metal contact, then the channel resistance, then electrons are injected into the drain and they firstly move horizontally seeing the contribution R(accum) and the R(jfet) and the final contribution given by the combination of the spread resistance and bulk resistance (grouped in a single contribution, R(epitaxial)).





On–Resistance Component	1000 V Standard MOSFET	250 V Standard MOSFET	50 V Standard MOSFET	50 V, High Cell Density MOSFET
Channel	0.4%	5.1%	40.5%	20.3%
Accumulation	0.1%	1.6%	12.6%	5.5%
JFET	8.3%	19.1%	12.7%	16.3%
EPI	91.1%	72.5%	20.1%	34.2%
Substrate	0.1%	1.7%	14.1%	23.7%

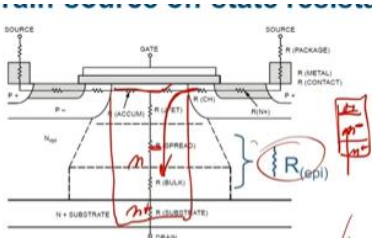
For BV_{DSS} > 200V $R_{(EPI)}$ is dominant ⇨ $R_{DS(ON)} \times area \propto (BV_{DSS})^{2.5 \div 2.7}$

"limit of silicon" **for unipolar conduction**

In the table we can see the impact of the different contributions depending on the blocking voltage of the mosfet.
If we consider a high blocking voltage, the channel resistance is negligible (0.4%), and the dominant contribution is the R(epi). The reason is simple. Since I want to block thousands of volts, I need to reduce the doping level of the epitaxial layer and increase the thickness to allocate the depletion layer. The obvious result is that the resistance increases because it is the resistivity times the length divided by the area, and we are increasing the length and the resistivity (resistivity increases because we decrease the doping).

To lower blocking voltage, 250V, still the R(epi) dominates. If we further decrease the blocking voltage, the dominant contribution comes from the channel resistance.
Hence if the mosfet has to provide a blocking voltage larger than 200V, the R(epi) is dominant with respect to the other contributions. If we are in this case, there is a trade-off between Rds,on and the area that we cannot avoid (red formula in the image).

This relationship is the one that we found for a Schottky barrier diode. The difference is that in a Schottky barrier diode we have metal, n- layer and n+ layer and electrons are moving from bottom to top. In the mosfet in this case we still have the same situation.

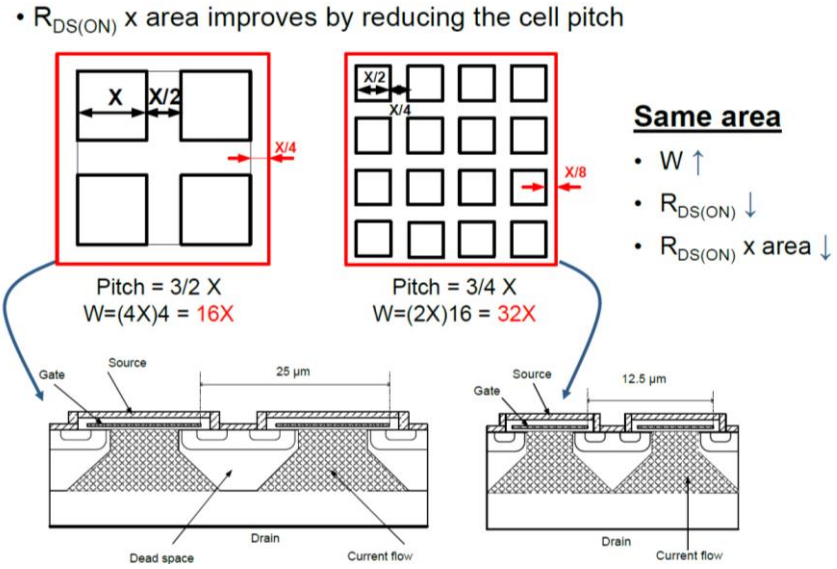


Hence we have the same limitations, so to compute the limiting Rds,on we can do the same analysis we did with the Schottky diode. In fact, the mosfet is a unipolar device, the current is sustained just by electrons.

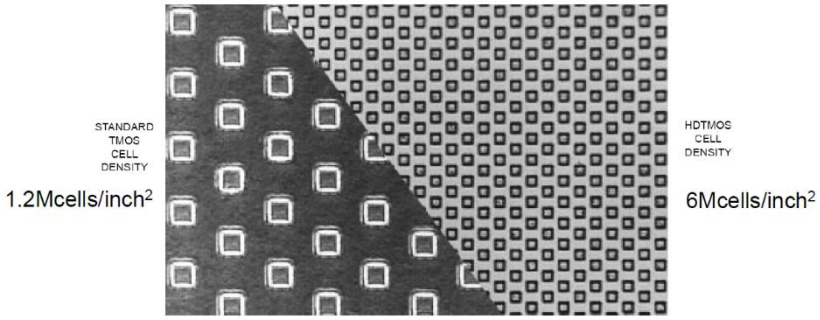
The right plot shows the R(epi) and R(ch) depending on the breakdown voltage (blocking voltage). The larger the blocking voltage, the larger the contribution coming from R(epi).

If we are considering lower blocking voltage devices, the situation is different because is R(ch) which is dominant. Can we minimize this resistance? Yes.

Blocking voltage smaller than 100V



We want a solution to reduce Rds,on for low voltage. A simple way is to **increase the cell pitch**. Inside the red area of silicon in the left case we have 4 cells and the pitch is the sum of the cell width and cell to cell distance. If we compute the total channel width it is the perimeter of a cell multiplied by the number of cell. If I keep the same area but half the cell pitch, the cell sizes are halvened. I get more cells, and the total channel width is increased. Hence we are decreasing the Rds,on with the same area.



On–Resistance Component	50 V Standard MOSFET	50 V, High Cell Density MOSFET
Channel	40.5%	20.3%
Accumulation	12.6%	5.5%
JFET	12.7%	16.3%
EPI	20.1%	34.2%
Substrate	14.1%	23.7%

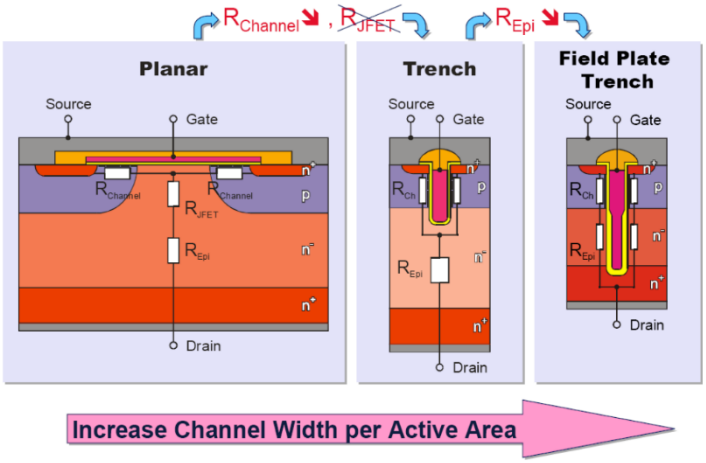
Progress in VDMOS structures

To overcome the R(jfet) resistance limitation we can introduce some more complex non-planar structure where we explore trenches.

In the classical planar structure, electrons move from top to bottom, and the narrower the region between the p wells, the higher the R(jfet).

We can create a trench where the yellow part is the oxide, and the trench is filled with polysilicon. Polysilicon is then contacted to the gate, and the trench is penetrated in the silicon between two adjacent

cells. The channel is no more horizontal, it is vertical (red part is the gate). By applying a positive voltage we are inducing an inversion and electrons are injected vertically, they immediately go vertically and no more horizontally before, so we eliminate the R(jfet) contribution.



Eventually, the trench can go all the way through the epitaxial layer. When we apply a positive voltage Vgs to turn on the device, we get an inversion layer in the purple region and an accumulation layer in the epitaxial layer. The accumulation layer is reach of electrons and so the R(epi) decreases.

A comparison between devices is in the following image.

75V MOSFET technology trends - Comparing MOSFETs of equal die size

	HUF75545P3 (Older Planar)	FDB045AN08A0 (Newer Trench)
$R_{DS(on)}$	10mΩ	4.5mΩ
Q_g	235nC	138nC
P_d	270W	310W
t_{rr} (@ 25°C)	100ns	53ns
Q_{rr} (@ 25°C)	300nC	54nC

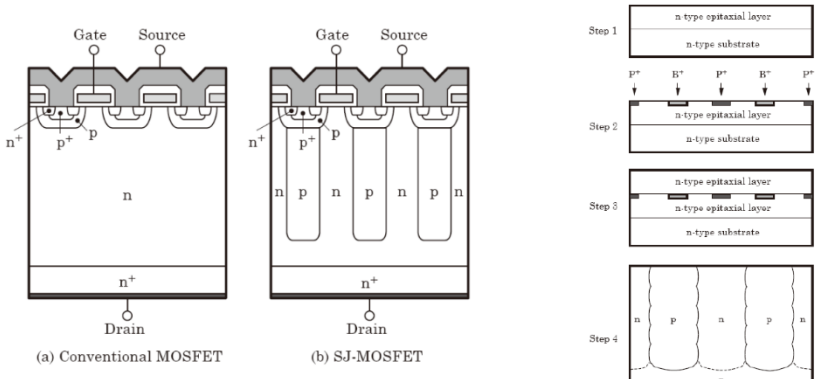
4.5 mOhm is a very small resistance, that is similar to the resistance of wires and metal lines. This is also the reason why we need to use advantage packages for these devices, not through hole.

Qrr is the reverse recovery charge of the intrinsic body-drain diode, and the trr is the reverse recovery time for the intrinsic body-drain diode (from source to drain). Moving from a standard technology to an advanced one we reduce both. This means that switching off the parasitic diode is much easier and faster.

SUPERJUNCTION MOSFET

It is a power mos device and the difference with respect to a standard DMOS device (on the left) is that we have pillars that we grow below the bodies of the cells. To build these pillars we start from an epitaxial layer on the top of which we implant the Boron and Phosphorus regions. Then we grow a second epitaxial layer, we perform a second implantation and so on.

By using the SJ mosfet we can break the Rds,on vs area trade-off. In fact if the mosfet device has a blocking voltage larger than 200V, the main component to Rds,on is the R(epi), and if this is true there is a trade-off with the area. To overcome this trade-off, either we change the material or we introduce the SJ mosfet.



- Superjunction MOSFETs reported at the end of '90s
- Break the $R_{DS(ON)} \times area \propto BV_{DSS}^{2.5 \div 2.7}$ limit of conventional VDMOS
- Complex fabrication

Standard MOS and SJMOS: Rds,on vs Area

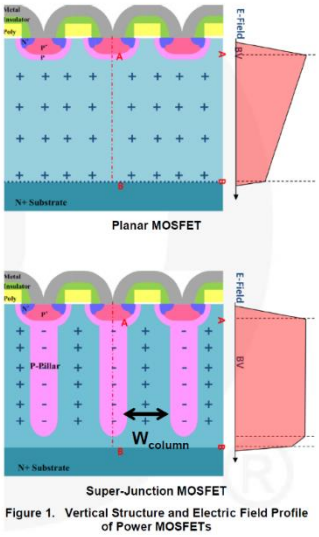
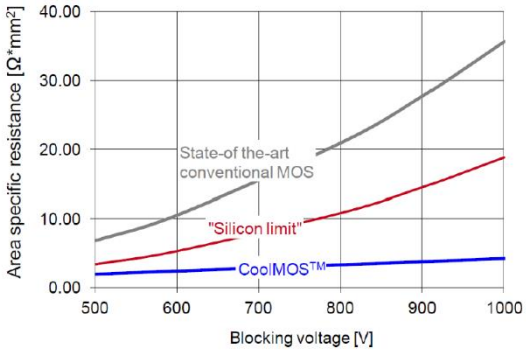


Figure 1. Vertical Structure and Electric Field Profile of Power MOSFETs



$$R_{ON,sp} = \frac{2V_{BD} \cdot W_{column}}{\varepsilon \mu_n E_C^2} \left[\Omega \cdot cm^2\right]$$

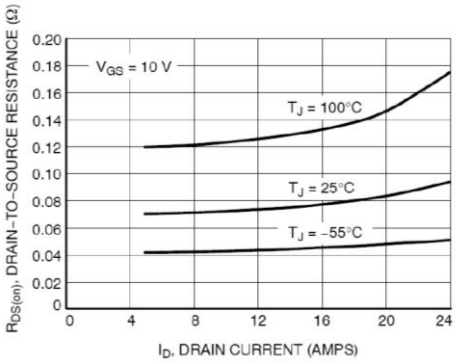
Assumption: N-columns and P-columns have the same width

CoolMOS is the name of the SJMOS commercially. It is important for medium power applications. In a standard device the field profile in the drift layer when we are reverse biasing and turning off the mosfet is trapezoidal or triangular, while in a SJMOS we get a square profile. The advantage is that we can get, at the same conditions, a larger blocking voltage.

In the plot we have the area specific on resistance as a function of the blocking voltage. The gray line represents the performance of a conventional pMOS, while the red line is the Silicon limit of a planar structure. By using the SJMOS we can overcome the Silicon limit. It makes sense to use a CoolMOS if we use at several hundreds of V, but for few tens of volts there is no point in using it, it's just more expensive.

In the green box we have the specific on resistance for the CoolMOS and we see that the dependance on the breakdown (blocking) voltage is linear.

Rds,on VS TEMPERATURE



Typical temperature dependence of mobility

	Si	Ge	GaAs
Electrons	$\propto T^{-2.4}$	$\propto T^{-1.7}$	$\propto T^{-1.0}$
Holes	$\propto T^{-2.2}$	$\propto T^{-2.3}$	$\propto T^{-2.1}$

- $R_{DS(ON)}$ increases with temperature (positive temperature coefficient - PTC).
- Parallel operation of MOSFETs is facilitated. Any imbalance between MOSFETs does not result in *current hogging*.
- Note: PTC does not however ensure even current sharing.

Rds,on is temperature dependent and the resistance depends on the conductivity of the channel and the conductivity of the channel depends on the temperature, because the mobility of electrons and holes in the channel depend on the temperature.
In the plot the parameter changing is the temperature. Rds,on loosely depends on the current but strongly on the temperature. Typically, the maximum temperature at which the mosfet is operated is 100°C. When we turn on the device we don't have to consider the Rds,on at room temperature, because the device dissipates power, so we have to consider the maximum Rds,on, that is the one measured at the maximum temperature at which we operate the mosfet.

The Rds,on of the power mosfet shows a positive temperature coefficient (PTC). This is good because it makes easy to work with several power mosfet in parallel.

If for example we want to drive a load that is absorbing 50A and we have devices capable of handling 12A, we need 5 of them in parallel to drive the load.
Having power mosfets in parallel doesn't arise any issue, because the PTC of the Rds,on prevents the mechanism of **current hogging**.
Current hogging is a problem that we find when we work with bipolar transistors in parallel, because they have a NTC, and the hotter the device, the larger the current that flows in the device (with the mosfet the hotter the device, the smaller the current). The consequence is that if we have a bad thermal design of the system and one transistor gets hotter than the others, this one is absorbing more current than the others because it is becoming more conductive, and it gets hotter and hotter up to the point where it blows. In mosfet we don't have this problem.

Logic-level MOSEFTs

Discrete power mosfets are operated usually with a gate to source voltage around 10V. However, there is a family of power mosfets, called logic level mosfets, where the Vgs is typically 5V.
The difference with respect to a classical mosfet is in the thickness of the oxide, in the logic level mosfets it is thinner. This results in a lower threshold voltage and so we can drive the mosfet in the ohmic region providing a reasonably low Rds,on with a logic level Vgs of 5V.

Of course, the maximum Vgs is limited due to the limited thickness of the oxide. In integrated CMOS technology, the Vgs of power devices depends on the technology.

- Logic level MOSFETs are specifically designed for operation from 5V logic and have guaranteed on-resistance at 5 or 4.5 V gate voltage.
- The logic-level version uses a thinner gate oxide and different doping concentrations. This has the following effects on the input characteristics:
  - gate threshold voltage is lower
  - transconductance is higher
  - input capacitance is higher
  - gate-source breakdown voltage is lower

Datasheet

In a power mosfet datasheet, the most important plot is the one reporting the Rds,on value vs temperature (bottom left). x is instead important to compute the switching properties of the mosfet. y is the value of the parasitic capacitances as a function of the applied voltages between gate, drain and source.

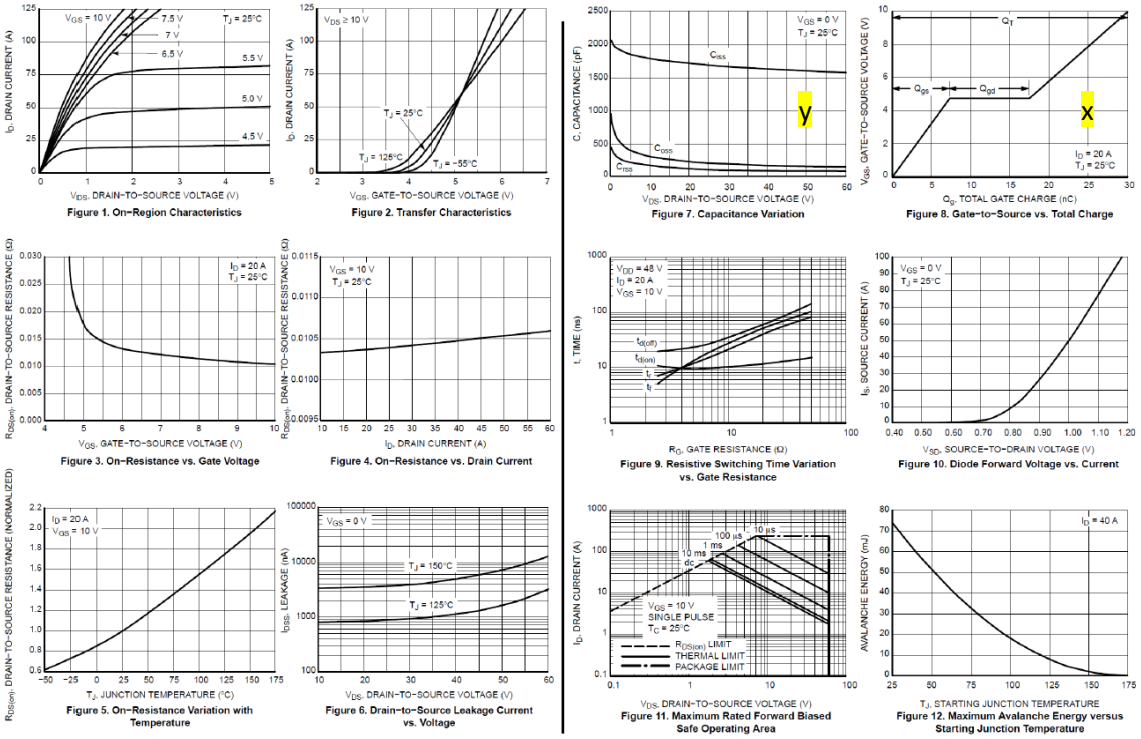


Figure 1. On−Region Characteristics
Figure 2. Transfer Characteristics
Figure 7. Capacitance Variation
Figure 8. Gate−to−Source vs. Total Charge
Figure 3. On−Resistance vs. Gate Voltage
Figure 4. On−Resistance vs. Drain Current
Figure 9. Resistive Switching Time Variation vs. Gate Resistance
Figure 10. Diode Forward Voltage vs. Current
Figure 5. On−Resistance Variation with Temperature
Figure 6. Drain−to−Source Leakage Current vs. Voltage
Figure 11. Maximum Rated Forward Biased Safe Operating Area
Figure 12. Maximum Avalanche Energy versus Starting Junction Temperature

ABSOLUTE MAXIMUM RATINGS OF POWER MOSFETS

Absolute maximum ratings is a set of variables like voltages, temperature, power, current that must not be exceeded. If we overcome the maximum ratings we loose the functionality of the mosfet (e.g. the mosfet enters the breakdown region if we go above Vbd and it becomes like a voltage generator). If for instance we overcome the maximal temperature usually nothing happens, we get a reduction of the mosfet lifetime, we don't have an immediate blow up of the mosfet.

Damages may happen when we overcome the maximum current limit. In this case we might blow the bonding wire. Anyway, these failure mechanisms will be seen later on.

There are some circumstances that we cannot foresee, even if we design the circuit such that the maximum ratings are never met. For instance, we are driving the motor and the motor gets stuck mechanically so we get an overcurrent. In this case the designer cannot foresee the mechanical block, so to avoid any damage in the mosfet we have to implement protection mechanisms for these unexpected events (overcurrent, overvoltage, overtemperature protection to prevent the overcoming of the maximum ratings).

In the table, the most important absolute maximum rating is the breakdown voltage between drain and source (in the example 60V). The same maximum rating is also expressed for Vgs to prevent the breakdown of the oxide.
Another figure is the continuous drain current, where we have written the maximum value of 63A, but it varies depending on temperature.

There is also a peak current that we can afford to flow only if it is very quick, and in this device is 252A. the difference between the pulsed drain current and continuous one is in the time duration of the current.

MAXIMUM RATINGS (T_J = 25°C unless otherwise stated)

Parameter			Symbol	Value	Units
Drain−to−Source Voltage			V_{DSS}	60	V
Gate−to−Source Voltage − Continuous			V_{GS}	±20	V
Gate−to−Source Voltage − Non−Repetitive (t_p = 10 μs)			V_{GS}	±30	V
Continuous Drain Current − $R_{\theta JC}$ (Note 1)	Steady State	T_C = 25°C	I_D	63	A
		T_C = 100°C		45	
Power Dissipation − $R_{\theta JC}$ (Note 1)	Steady State	T_C = 25°C	P_D	107	W
		T_C = 100°C		54	
Pulsed Drain Current	t_p = 10 μs		I_{DM}	252	A
Operating Junction and Storage Temperature			T_J, T_{STG}	−55 to 175	°C

NTP5864N

- The absolute maximum ratings table contains maximum limits for voltage, temperature, allowed current, power dissipation and other characteristics.
- Maximum ratings are limiting values of operation and should not be exceeded under the worst conditions.
- If any of these limits are exceeded, device functionality should not be assumed, damage may occur and reliability may be affected.

MAXIMUM RATINGS: Vgs

The physical limitation is the breakdown of the oxide. Typically it is 5-10 MV/cm for SiO2.
Of course, the breakdown of the oxide is an irreversible kind of damage.

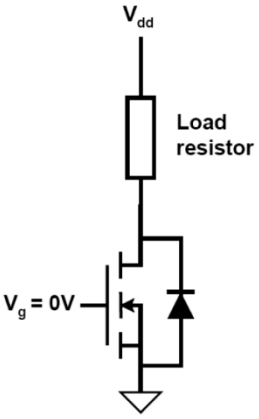
The most common way to destroy a mosfet device is by touching the gate with the finger, because the impedance of the gate is very high (we see a floating polysilicon surrounded by oxide) and if we have

some electrostatic charge on the finger, even few nC of charge break the oxide. Hence the mosfet must be handled always with a wrist arm connected to ground and with ESD protections placed in the mosfet, e.g. two Zener diodes. However, the drawback of placing Zener diodes is that we increase the parasitic capacitance between gate and source, degrading the switching times.

- $V_{GS(max)}$ = maximum gate-source voltage
- If $V_{GS} > V_{GS(max)}$ damage of gate oxide by large electric fields is possible
- E_{BD}(oxide) ≈ 5 -10 MV/cm
  - gate oxide typically 1000 Angstroms thick
  - $V_{GS(max)} < (5 \cdot 10^{6}$ V/cm) · (10^{-5} cm)= 50 V
  - typical $V_{GS(max)}$ = 20 -30 V
- Static charge on gate conductor can destroy gate oxide
- Handle MOSFETs with care (ground yourself before handling device)
- Place anti-parallel connected Zener diodes between gate and source as a protective measure

MAXIMUM RATINGS: Vds

- When a MOSFET is in the OFF state, it prevents current flowing through the load, except for a small leakage current, I_{DSS}
- Here, V_{dd} must not exceed the maximum specified drain-source voltage of the MOSFET, BV_{DSS}
- The ability of a MOSFET to block voltage
  - Decreases with temperature
  - Decreases with negative V_{GS}



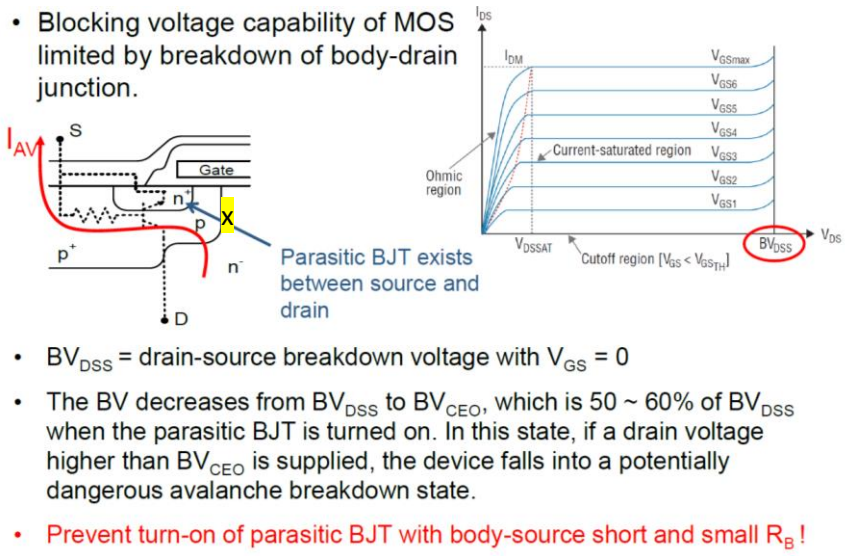
If the mosfet is in the off state, no current flows in the load (assumed resistive for simplicity), so the output node goes to Vdd.
We need to select a mosfet that is able to block this Vdd voltage. So the blocking voltage of the mosfet must be always larger than the bias supply of the application.
The maximum blocking voltage is determined by the breakdown of the drain to body junction. In fact, the breakdown typically occurs in position x, where the junction has a curvature because the electric field concentrates and peaks.

NB: source and body are shortcircuited, so the drain to body voltage is the drain to source.

There is a maximum voltage that we can apply that is the BVdss (conventional name for the breakdown voltage for a mosfet device in common source configuration), and we have to be careful not to overcome this limit. If we pass it, the device becomes a voltage generator, so it start conducting current and it is no more able to block the voltage.

There is a problem with power mosfet transistor, and it comes from the parasitic n+/p/n- parasitic in the source region. So we get a parasitic bipolar transistor that can be very annoying.

In fact, if for any reason this parasitic transistor is turned on, the breakdown voltage of the junction collapses down to a value called BVceo, which is typically 50% lower than BVdss.



- BV_{DSS} = drain-source breakdown voltage with $V_{GS} = 0$
- The BV decreases from BV_{DSS} to BV_{CEO}, which is 50 ~ 60% of BV_{DSS} when the parasitic BJT is turned on. In this state, if a drain voltage higher than BV_{CEO} is supplied, the device falls into a potentially dangerous avalanche breakdown state.
- Prevent turn-on of parasitic BJT with body-source short and small R_B !

To keep the parasitic transistor off we can apply a shortcircuit between source and body. Nevertheless, sometimes this is not sufficient because the spreading resistance of the body region might be high enough so that a current flowing through it produces a voltage drop turning on the parasitic bipolar transistor. This current can be a simple displacement current when we turn on the mosfet. If so, the BVdss collapses to smaller values, and if the Vdd is in the middle of the range, we might end up in a situation where the current is flowing and we cannot turn it off and the mosfet is going to blow.

Choosing BVdss of the MOSFET

- The voltage rating of the MOSFET must be larger than the bus voltage V_{dd}
- Allow safety margin for:
  - Bus voltage variations
  - Input voltage spikes
  - Ringing in synchronous buck
  - Motor drive spikes
  - Flyback transformer spikes
  - Change in V_{DS} at low temperature
  - Change in V_{DS} with negative V_{GS} drive

Typical MOSFET voltage ratings	
Portable equipment	*20V*
FPGA, VLSI supply	*20V,30V*
24V motor drive	*60V*
48V system	*80-100V*
80V system	*150-200V*
85VAC – 220VAC	*450-600V*
Three phase PSU	*800-1000V*

We must never select a mosfet with a breakdown voltage equal to the power supply voltage of my application, we need it sufficiently high with respect to PS. So we need some margins to take into account possible fluctuations of the PS voltages, spikes and ringings.
In the image we have a table with suggested margins when selecting a mosfet. In general, for automotive applications the mosfet rating is 60V. The next generation car battery is moving from 24V to 48V because we will have more and more electrical loads that are sinking more and more current. So to avoid Joule losses in the cables due to the flow of high currents we increase the voltage so that for the same load power less current is flowing.

Load damp event

Let's assume the alternator of the car is charging the battery and for any reason we disconnect the battery. We have an overvoltage produced on the alternator line that can reach up to 60V. The same is true when we are damping a load, e.g. there is an inductive load that is sinking current and we shut it off.

Due to the inductive nature of the load, if we suddenly cut the current, we have a sudden increase of the voltage (V = L * di/dt).

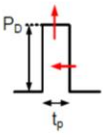
MAXIMUM RATINGS: Power Dissipation

- P_D is the maximum allowed continuous power dissipation for a device mounted on an infinite heat sink at 25°C.
- The power dissipation is calculated as that which would take the device to the maximum allowed junction temperature, T_{JMAX}:
  - T_{Jmax} < 150 – 175 °C with molded packages
  - T_{Jmax} < 200°C with metal packages

P_D= maximum power dissipation
$R_{\theta JC}$= juntion-to-case thermal resistance
$Z_{\theta JC}$= juntion-to-case thermal impedance

x $(T_{jmax} - T_C) = R_{\theta JC} \cdot P_D$ DC operation

$(T_{jmax} - T_C) = Z_{\theta JC}(t_p) \cdot P_D$ Pulsed operation



we'll see all this in part 10….

The maximum current rating in a mosfet is determined by the maximum power rating, which is internally determined by the maximum temperature that can be reached by the mosfet. This is why developing a correct thermal design for the application is important.

The maximum power dissipation Pd is in DC condition, so constant current and constant power dissipation (transients are over) it is defined in a well specified condition (it is a standard). Pd is measured with the device mounted on an infinite heat sink at 25°C.

The Pd is calculated in such a way that the temperature of the junction reaches the maximum allowed temperature as indicated in the image. The temperature is measured close to the channel because it is where the highest temperature is met. To relate power dissipation and temperature we need to introduce the concept of thermal resistance (R-thetaJC). In DC, the maximum Pd is the one such that relationship x is verified.

There is also a pulsed power dissipation. E.g. we are turning on the mosfet for a short period. In this case we can increase the power dissipation over the DC limitation. The shorter the duration of the pulse, the higher the pulse power that we can inject in the system.

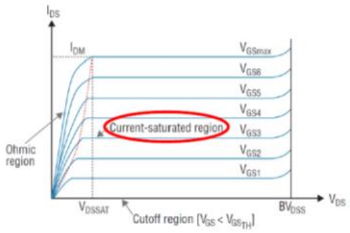
So the origin of the Pd in DC condition is the maximum junction temperature, because we don't want the temperature to overcome 150°C, which is the glass transition temperature of the plastic package. Moreover, the higher the temperature of the silicon, the higher the probability of failure of the device due to oxide breakdown or electromagnetic effect.

MAXIMUM RATINGS: Drain Current

- All MOSFETs have a specified maximum current rating
  - the maximum continuous (DC) current that the power MOSFET can pass in the forward direction is specified as I_D.
    - In general, the I_D value is specified so that the channel temperature will not exceed the maximum junction temperature T_{JMAX}.

$$P_D = I_D^2 \cdot R_{DS(ON)}(T_{jmax}) \rightarrow I_D = \sqrt{\frac{P_D}{R_{DS(ON)}(T_{jmax})}} = \sqrt{\frac{T_{jmax} - T_C}{R_{\theta JC} \cdot R_{DS(ON)}(T_{jmax})}}$$

  - the pulsed current that the power MOSFET can pass in the forward direction is specified as I_{DM}.
    - I_{DM} is determined by the saturation of the MOSFET current.
      - The maximum allowable pulsed current in turn depends on the pulse width, duty cycle and thermal conditions of the device.



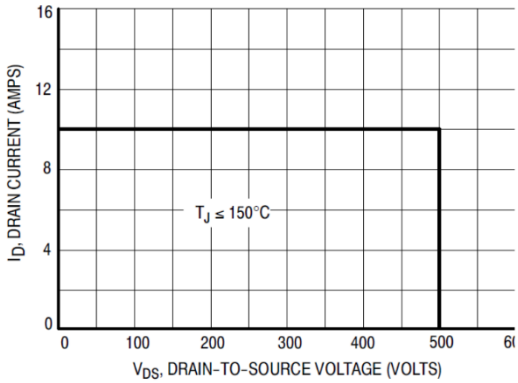
We want the maximum DC drain current Id. This maximum current is related to the maximum power that can be dissipated by the mosfet, which is related to the maximum junction temperature.
The maximum current is the current that produces a power dissipation equal to the maximum allowable power dissipation (always in DC).
The Rds,on to be used is the maximum one, in the worst case scenario (standard Tc is 25°C).

If for any reason I increase the case temperature Tc (working temperature) above 25°C, the current reduces.

Another specification we can look at is the maximum pulse current Idm. Instead of applying a constant DC current we can turn on the mosfet for a short period with a large current, dissipating a large power but in short period. If the pulse time is so short that the junction temperature is not changing too much, the maximum Idm is determined by the saturation of the mosfet, so the mosfet enters in the saturation regime. In general, the mosfet entering the saturation regime is the mechanism that limits the maximum current.

SWITCHING SAFE OPERATION AREA (SSOA)

- The *SSOA* is the boundary that the load line may traverse without incurring damage to the MOSFET. The fundamental limits are the peak current, I_{DM} and the breakdown voltage, BV_{DSS}.
- Obsolete !



MTP3N50E

In the case we use mosfet as a switch, turning just on and off, the only limitations we have to respect are 3:

1. Not overcoming the maximum breakdown voltage (vertical line in the plot)
2. Not overcoming the maximum current rating (horizontal line in the plot)
3. Not overcoming the maximum junction temperature

If we want to use a mosfet for linear applications, e.g. class A or class AB amplifiers, we need to consider the forward bias safe operating area, that is the one inside the rectangle.
Since we are not interested in linear applications, we can forget about it.

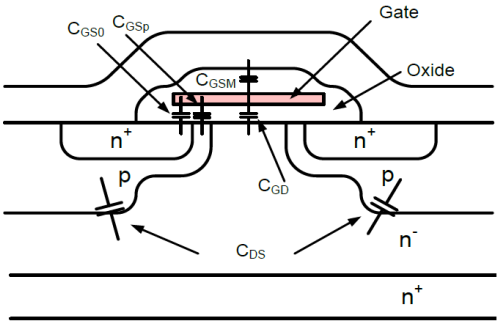
SWITCHING BEHAVIOUR OF POWER MOSFET

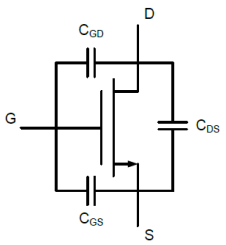
- Being majority carrier devices, power MOSFETs are free from the charge storage effects which impair the switching performance of bipolar devices.
- How fast a power MOSFET will switch is determined by the speed at which its internal capacitances can be charged and discharged by the drive circuit.
- MOSFET switching times are often quoted as part of the device data. However, these figures might be largely irrelevant to the true switching capability of the device.
- The quoted values are only a snapshot showing what will be achieved under the stated conditions.

When considering the switching behaviour of a mosfet we are interested in how much fast we can switch it on and off, and it is important because there are losses, and the faster the transient, the smaller the switch losses.
Moreover, the mosfet is a unipolar device, so we don't have any storage of minority carrier, when we turn it off we don't have to remove the excess of carriers stored in the drift layer (this has to be done in bipolar transistors). The limitation in the time with which we can turn it off and on is in how much fast we can charge or discharge the parasitic capacitance that we have between the gate and the drain, gate and source and source and drain.

In the datasheet we can find a table, called '**switching characteristics**', where there are some typical turn on and turn off delay times. These values are not really meaningful because they are just a snapshot of the transistor behaviour if we are using it in a specific condition. The meaning of this is just to compare mosfets from different manufacturers.
But to analyze how much the mosfet is fast in turning on or off, we have to rely on another characteristic of the mosfet, the gate charge curve.

PARASITIC CAPACITANCES OF THE POWER MOSFET





$C_{GS}=C_{GSM}+C_{GS0}+C_{GSp}$

$\approx$ independent of V_{GS}

- Data sheet capacitances

$C_{iss} = C_{GD} + C_{GS}$ input capacitance

$C_{rss} = C_{GD}$ reverse transfer capacitance

$C_{oss} = C_{GD} + C_{DS}$ output capacitance

Input Capacitance	(V_{DS} = 25 Vdc, V_{GS} = 0 Vdc, f = 1.0 MHz)	C_{iss}	–	323	450	pF
Output Capacitance		C_{oss}	–	107	150	
Transfer Capacitance		C_{rss}	–	34	70	

There are at least three parasitic capacitances: Cgs, Cgd and Cds. We are mostly interested in the Cgd and Cgs, because those two components are the main contributors to the transition time for the turn on

and turn off transients. Cds is not actually contributing to the transient time. However it has to be considered because the Cds is affecting adversely the power dissipation of the mosfet.

Cgd and Cgs

Cgs is including three different contributions. Firstly we have the Cgsp (capacitance between the gate and the underlying body region), Cgso (due to the overlap of the gate contact and the source diffusion area) and Cgsm (capacitor between the gate contact and the overlying source metal).

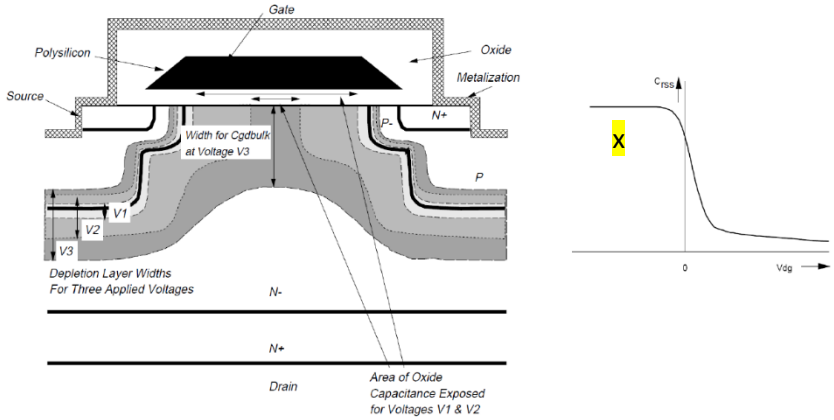
Is the Cgs dependent on the voltage that I'm applying between gate and source? Almost not, it is independent. In fact, the Cgsm contribution is independent from the voltage because it is like the capacitance of a parallel plate capacitor, whose capacitance is determined just by the thickness of the oxide that separates the gate and the top source metal. Also Cgso is independent, while Cgsp is dependent on the voltage because it depends on the fact that the channel is inverted or depleted. However, the contribution from the Cgsp is negligible with respect to the other two → in a power mosfet device the Cgs is independent on the applied voltage between gate and source.

As for Cgd, it is highly non-linear, highly dependent on the voltage applied between gate and drain.

If we look at the datasheet, we don't find the values of Cgd and Cgs, but the values of **Ciss** (input capacitance), **Crss** (reverse transfer capacitance) and **Coss** (output capacitance).

MOSFET – REVERSE TRANSFER CAPACITANCE

- C_{rss} is strongly and non-linearly dependent on applied voltage, V_{dg}.



- C_{rss} is is this capacitance which plays a dominant role during switching and which is also the most voltage dependent.

We consider just two operation points, which are switch on (mosfet in the ohmic region) or switch is off (mosfet in the cut off region).

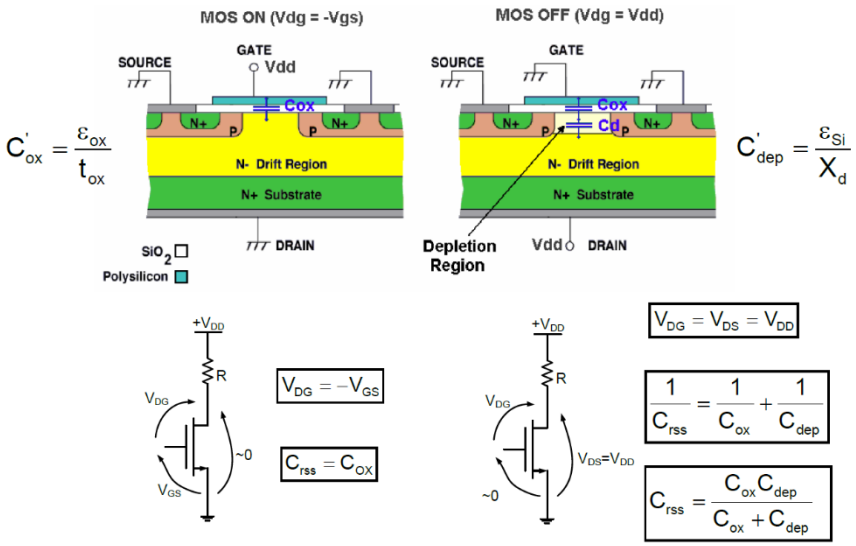
Switch on

Vg = 10V, the drain voltage is more or less equal to 0V. Vdg in this case is positive, 10V. Hence we are recalling electrons from the epitaxial layer and these electrons are accumulated at the interface between silicon and SiO2. The capacitor becomes a simple parallel plate capacitor, where the bottom plate is the accumulation layer. The specific capacitance Cgd' (' means specific) is the ratio between the dielectric constant of the oxide and its thickness. This is the reason why the capacitance for the mosfet in the ohmic region is large (x).

Switch off

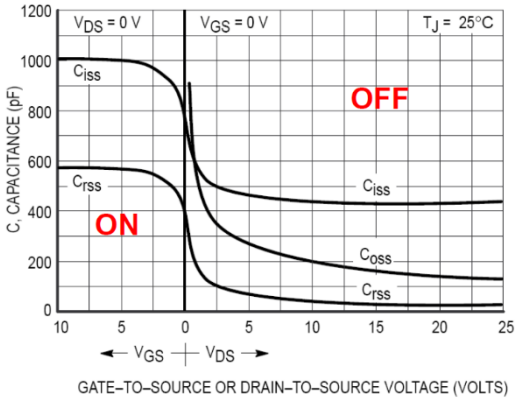
Turning off a mosfet means that we are grounding the gate, hence the drain voltage rises up up to more or less Vdd. The Vdg is Vdd in this case and the capacitance is much smaller because we are depleting the silicon region below the oxide, because the gate is grounded and we are applying a positive voltage to the drain. The conclusion is that now we have two series capacitances. The first one is the oxide capacitance, and the second one the depletion layer capacitance. The capacitance of the depletion layer is the ratio between the silicon dielectric constant and the thickness of the depletion layer. The depletion layer width might be very large because the doping concentration is not high, so the capacitance associated to the depletion layer is huge and we end up with this contribution dominating in the series.

This is true in a condition of deep depletion.



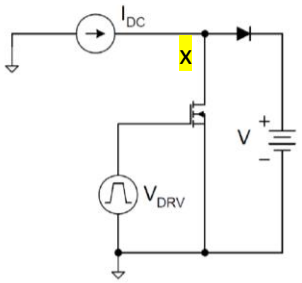
In the image above we are summarizing what said so far. In conclusion, the mosfet in the off region shows a Crss as a series of oxide capacitance and depletion layer capacitance dominated by the latter.

MOSFET CAPACITANCE VARIATION



The y axis reports the capacitance, and the x axis is split in two parts defining two half planes. In the left plane we see the mosfet operating in the on (ohmic region), so the gate to drain voltage is -Vgs and Ciss (Ciss = Crss + Cgs) is pretty large. In the off region the Vds is Vdd.
If we look at the distance between the Ciss and Crss in the off region, it is constant. It means that Cgs is independent from the voltage.

CLAMPED INDUCTIVE SWITCHING

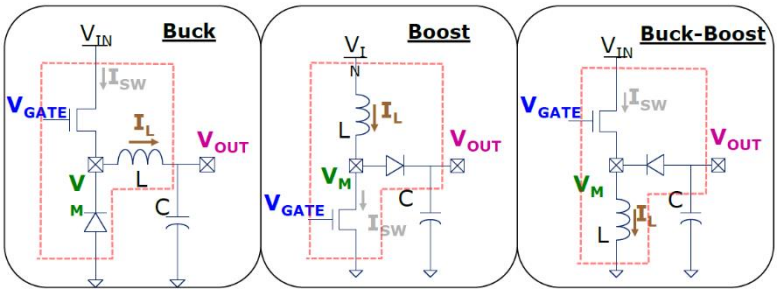


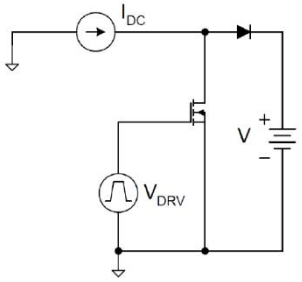
- Simplified clamped inductive switching model

- The DC current source represents the inductor. Its current can be considered constant during the short switching interval.
- The diode provides a path for the current during the off time of the MOSFET and clamps the drain terminal of the device to the output voltage symbolized by the battery

Turn on and turn off calculations will be done considering that the load of the mosfet (in the image we have a common source configuration) is a current generator and that node x, when the mosfet is off, is clamped to V thanks to the diode. This is a general situations that happens in 99% if the power electronics calculations.

The basic non-isolated DC/DC converters can be traced back to this simple structure.





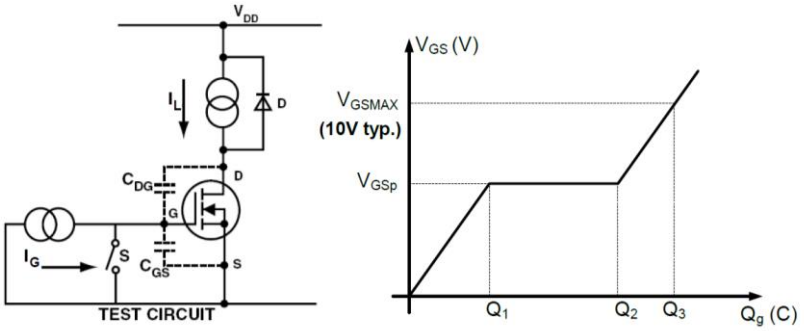
- This sub-network, and consequently this type of switching, is ubiquitous in power converters.

GATE CHARGE CURVE

- The nonlinearity of the parasitic capacitances make gate circuit design by conventional methods exceedingly difficult.
- To overcome this problem it has become standard practice to provide the *gate charge curve*
  - such a curve indicates the amount of charge Q_G that must be supplied to (removed from) the gate to effect the various stages of turn-on (turn-off).
- Gate charge curve permits a very simple design methodology for obtaining any desired switching time, and it enables the total charge and the total energy required to be estimated easily.

The gate charge curve it is an experimental curve, measured by the manufacturer and provided in the datasheet. It represents the amount of charge that we are supplying to the gate during the various phases of the turn on and turn off.

The gate charge curve is in the next image.



- The gate charge test circuit employs a constant current source to charge the MOSFET's input capacitance.
- A constant I_G ensures that C_{iss} is charged at a fixed rate (i = q/t).
- The V_{GS} waveform then, is a representation of V_{GS} versus gate charge as well as V_{GS} versus time.

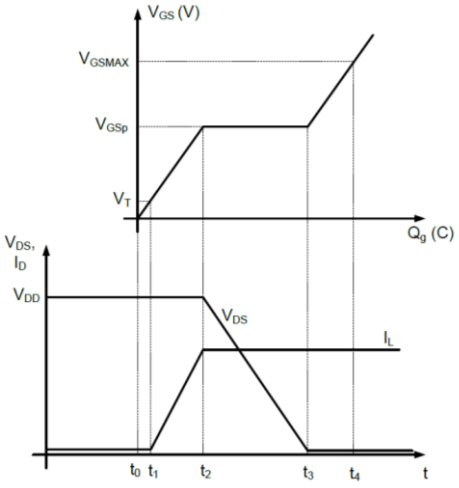
We start with the switch close and the mosfet is off. The inductor current is circulating in the clamping diode. At t = 0 we open the switch and if we are driving the mosfet gate with a constant current we are injecting a charge that is related to the current by i = q/t.
So we measure the gate charge Qg and the Vgs and we get the curve that is piecewise linear.

TURN ON TRANSIENT

In the image we have the gate charge curve and, in the same time frame in the bottom graph, the Vds and current Id transients. Each inflection point of the gate charge curve defines the beginning or the end of a distinct interval of the turn on transient.

At the beginning Vds = Vdd. Then at t0 I start injecting the current; firstly nothing happens until the Vgs reaches the mosfet threshold voltage. As soon as the threshold is overcame, the mosfet conducts current, but Vds stays constant because if the drain current is smaller than the load current there is some current in the clamping diode, that is on, so the drain node is clamped. At the first inflection point, the drain current becomes equal to the load current. From this point on, the drain current will stay constant, because forced by the current generator (load), and the Vgs remains constant.

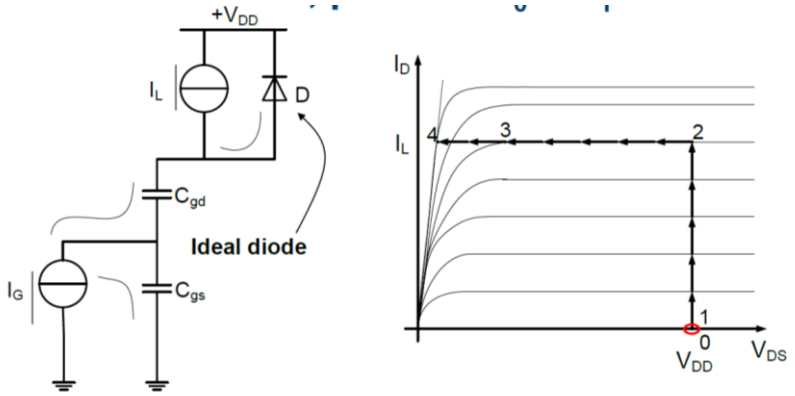
In the plateau region the Vds voltage reduces. When Vds approaches zero, the mosfet enters the ohmic region and the Vgs starts increasing again until the end of the transient. The end of the transient occurs when we inject a total gate charge so to obtain the desired Vgs,max.



- Each inflection point on the gate charge waveform defines the beginning or end of a distinct interval during the turn-on process.

Phase 1

It refers to the time interval between t0 and t1.



- At time t_0 the gate drive is activated.
- Current flows into the gate charging both C_{gs} and C_{gd}. After a short period the threshold voltage is reached and current begins to rise in the MOSFET.
- The drain-source voltage remains at the supply level (D is conducting).
- For $t_0 \le t < t_1$, $v_{GS} < V_{Th}$, i.e. the MOSFET remains in the cut-off region with $i_D = 0$.

At t0 the gate is activated and we start injecting current and we charge Cgs and Cgd.

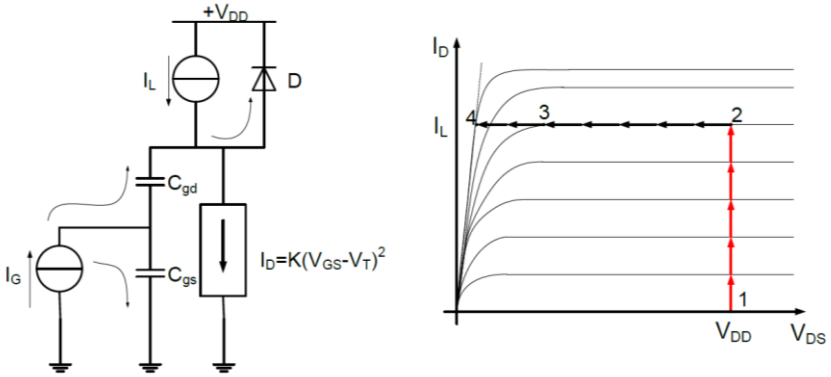
Phase 2

Between t1 and t2.

We are over threshold and the mosfet starts to be conductive. The mosfet is working in the saturation region because the drain is clamped to Vdd by the diode. In the saturation region the mosfet can be modelled as a constant current generator.

In this phase the mosfet current is increasing because we are still injecting charge both in Cgd and Cgs and the part of the charge that goes in the Cgs is increasing Vgs.

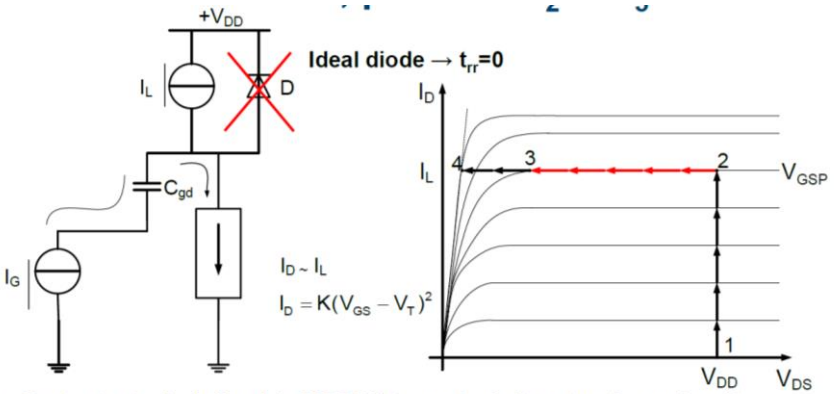
As said, voltage between drain and source is stuck at Vdd.



- At $t = t_1$ the device starts conducting.
- The drain-source voltage remains at the supply level as long as $i_d < I_L$ and the free-wheeling diode D is conducting.
- MOSFET operates in the saturation region.
- Current increases "*almost*" linearly until $i_D = I_L$.

Phase 3

From t2 to t3. We are using the assumption that the **diode is ideal**. The diode is turned off because the load current becomes equal to the drain current and initially the mosfet is still working in the saturation region. If the mosfet is working in the saturation region and I'm forcing the drain current to be equal to a constant value, the load current, it means that the Vgs has to stay constant → no more charge injected, Vgs = Vgs_plateau.



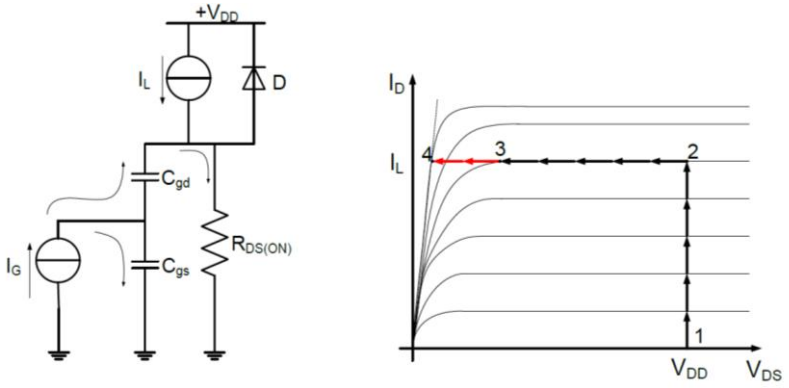
- For $t_2 < t < t_3$, $i_D = I_L$ ($I_G << I_L$). MOSFET operates in its saturation region. V_{gs} clamped at V_{GSP} → *"plateau"* region in the gate charge curve.
- The entire gate current now flows through C_{gd} causing the drain-source voltage to drop as C_{gd} is discharged.
- The rate at which v_{ds} falls is given by: $\frac{dV_{DG}}{dt} = \frac{dV_{DS}}{dt} = \frac{I_G}{C_{gd}}$

So Vgs is constant and we are injecting charges in Cgd, so we are increasing Vgd. If so, the drain voltage is dropping from Vdd to 0.

The voltage transient between Vdd and 0 would be linear assuming that Cgd was a linear or constant capacitor, but this is not true, so we expect that the voltage transient will be nonlinear. Initially the gate to drain capacitance is very small, so we are expecting a very steep variation of the Vd voltage. However, as soon as we approach the ohmic region (point 3), Cgd becomes larger and we have a less fast transient and nonlinear (because Cgd is nonlinear).

Phase 4

from t3 to t4, we enter the ohmic region at t3.

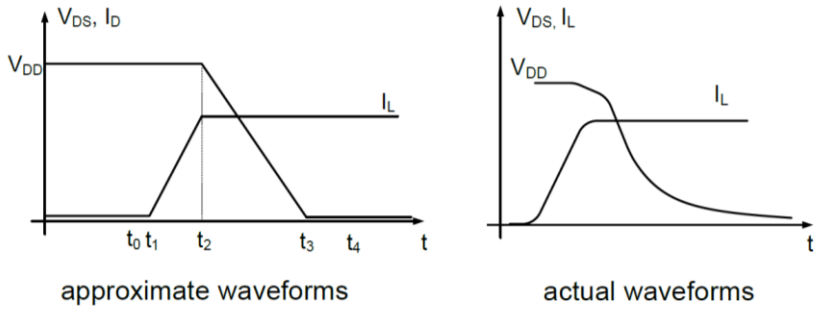


- For $t > t_3$,the device enters the ohmic region.
- The gate-source voltage becomes unclamped and continues to rise, causing $R_{DS(ON)}$ to drop.
- Turn-on transient stops when $V_{GS}=V_{GSMAX}$.

The mosfet becomes a resistor, so the Vgs is no more clamped because we are not operating in saturation. So the charge we are injecting with the current generator starts flowing also in the Cgs, making the Vgs to increase again. If the Vgs is increasing, we are reducing the Rds_on, meaning that the Vds is reducing from point 3 to point 4.

The transient is over when we reach Vgs,max.

Real transients



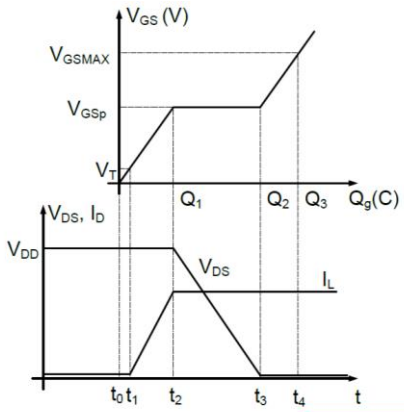
approximate waveforms actual waveforms

- Approximation acceptable in most practical applications

SWITCHING TIME CALCULATIONS

Let's get back to the turn on transient. We can identify 3 distinct time intervals: from t0 to t2, called **turn-on delay** and referred as **td(on)** in the datasheet, and it is the time needed by the current to reach the load level; from t2 to t3, called **rise time tr**, and it is the time interval where the voltage between the drain and source falls from Vdd to 0. The mosfet is operating in the plateau region; from t3 to t4, called **excess charge time** and the mosfet has entered the ohmic region.

Most of the power dissipation in the mosfet occurs during the first and second time intervals. Associated to each time interval there is a specific gate charge.



- Three distinct time intervals
  - $t_0 \rightarrow t_2$ (turn-on delay, $t_{d(on)}$)
  - $t_2 \rightarrow t_3$ (rise time, t_r)
  - $t_3 \rightarrow t_4$ (excess charge time)

- Associated gate-charge
  - Q_1
  - $(Q_2 - Q_1)$
  - $(Q_3 - Q_2)$

Assuming we are injecting a constant current, we can compute as below. Q1 is the charge associated to the turn on delay and it is found on the datasheet.

$$t_{d(on)} = \frac{Q_1}{I_G}$$

$$t_r = \frac{Q_2 - Q_1}{I_G}$$

Most relevant for estimating the switching losses

$$t_{3-4} = \frac{Q_3 - Q_2}{I_G}$$

Device operating in ohmic region: dissipation negligible

Note: the rise time is defined as the time between when v_{ds} is at 90% of its peak and it's fallen all the way down to 10% of its peak.

→ v_{ds} is falling, so why do we call this the rise time? Rise time actually refers to the rising current during the device turn on.

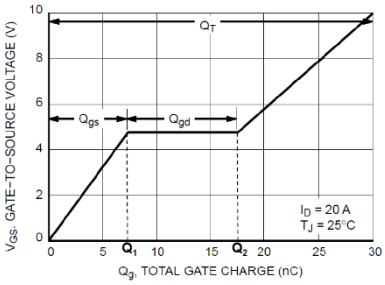
We are not really interested in t3→4 because the mosfet has already entered the ohmic region, so the Vds is small and power dissipation is negligible.

Example

Let's assume to turn on and off the mosfet with a constant current of 0.5A.

An interesting point is that <u>if we want to be fast in turning on and off the mosfet we just need to increase the current we are pushing into the gate</u>.

How much fast the mosfet is not only depends on the parasitic capacitances, but also on how much large the current that we are pumping or extracting in the gate is.

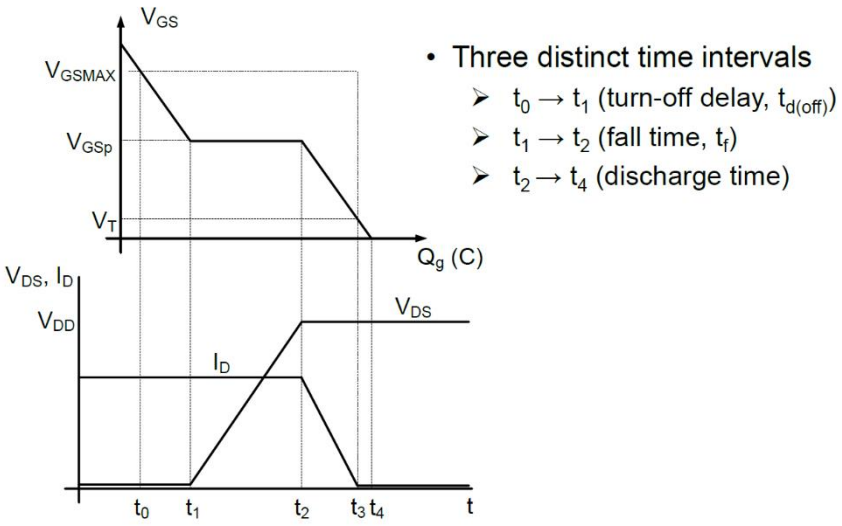


NTP5864N

- Assuming I_G=0.5 A:
  - $t_{d(on)} = \frac{Q_1}{I_G} = \frac{Q_{gs}}{I_G} = \frac{7.3nC}{0.5A} = 14.6ns$
  - $t_r = \frac{Q_2 - Q_1}{I_G} = \frac{Q_{gd}}{I_G} = \frac{10nC}{0.5A} = 20ns$

We cannot do this in a bipolar device, where turn on and off is completely determined by the parasitic capacitances, because we have to remove the excess charge accumulated.

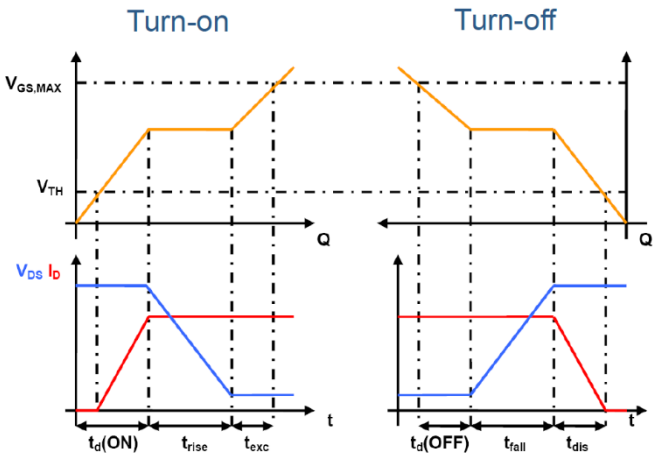
Turn off transient



Specular with respect to the turn on. We start from the mosfet operated in the ohmic region with Vgs = Vgs,max and the current in the mosfet that is i_L and we are in the ohmic region. In this transient we are extracting charges. Nothing happens until we exit the ohmic region. During the plateau the voltage increases from 0 to Vds, where the free wheeling diode is engaged clamping the Vds at Vdd.

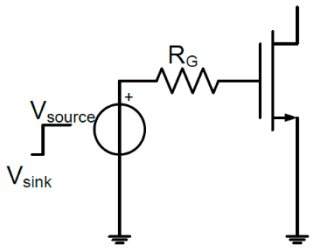
The turn off delay is the time needed to get out of the ohmic region the mosfet. Instead, the time interval t2→4 is not reported in the datasheets, but it is the most important, where the power dissipation occurs.

Summary

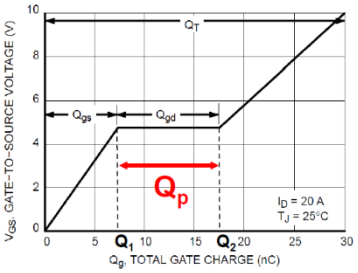


SWITCHING TIMES CALCULATIONS – VOLTAGE DRIVE

We are in the case where the gate driver is not an ideal current source. We can consider the Thevenin equivalent of the gate drive circuit and replace it with a voltage generator and a series resistance.
The voltage generator generates two levels of voltage. A low level called Vsink, that is extracting charges form the gate when applied (it might also be ground or a negative value), and a second level Vsource, which is pushing charges in the gate.



- Calculation of t_r and t_f is straightforward:
  - $v_{GS}=V_{GSplateau}$
  - I_G=const.



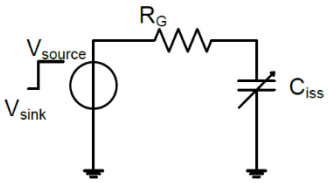
$$t_{fall} = \frac{Q_p}{I_G} = \frac{Q_p \cdot R_G}{V_{GSp} - V_{sink}}$$

$$t_{rise} = \frac{Q_p}{I_G} = \frac{Q_p \cdot R_G}{V_{source} - V_{GSp}}$$

where $Q_p = (Q_2-Q_1) = Q_{gd}$

The calculations of the switching time are elementary even if the gate driver is a voltage generator with a series resistance for t_fall and t_rise. This because the mosfet is operating in the plateau region for these two times. The plateau means that the Vgs voltage is clamped to Vgs_plateau.
The gate current Ig is constant, if Vsource is constant, and it is: Ig = (Vsource – Vgs)/Rg. Of course, Vsource > Vgs.

Instead, the calculations of td(on) and td(off) is less straight forward because Ciss is not constant. But this is not true in the sense that during td(on) and td(off), the Ciss is either constant equal to the maximum value or constant and equal to the minimum value.

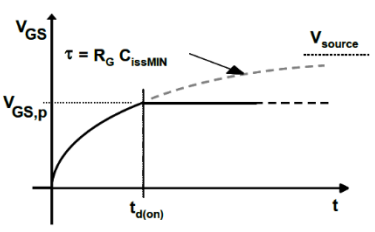


- Calculation of $t_{d(on)}$ and $t_{d(off)}$ is less straightforward, since C_{iss} is not constant
- However:
  - during $t_{d(on)}$, C_{iss}~$C_{iss(MIN)}$
  - during $t_{d(off)}$, C_{iss}~$C_{iss(MAX)}$

➔ Simple exponential transients !

$$t_{d(on)} = R_G \cdot C_{iss(MIN)} \cdot \ln\left[\frac{V_{source}}{V_{source} - V_{GSp}}\right]$$

$$t_{d(off)} = R_G \cdot C_{iss(MAX)} \cdot \ln\left[\frac{V_{GSMAX} - V_{sink}}{V_{GSp} - V_{sink}}\right]$$



td(on) is the time needed for the Vgs to reach the Vgs_plateau, so we have to invert the exponential function to calculate the time needed by the gate to source voltage to step up from 0 to Vgs_plateau.

Why is Ciss,max not available?

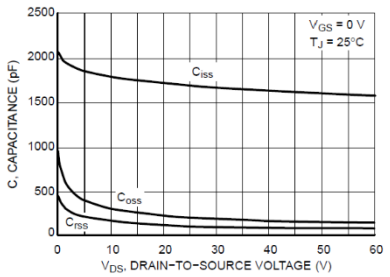


Figure 7. Capacitance Variation

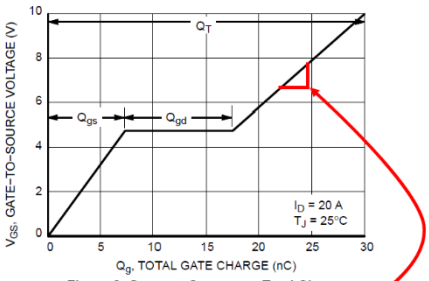
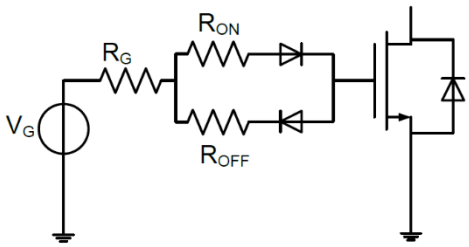


Figure 8. Gate−to−Source vs. Total Charge

$$C_{iss(MAX)} = \left.\frac{dQ}{dV}\right|_{ohmic}$$

Often Ciss,max is not available in the datasheet, reporting only Ciss for positive Vds. In this case we have to look at the gate charge curve, which is always present in the datasheet, and the inverse of slope of the curve in the last region is Ciss,max.

Separate adjustments of turn-on and turn-off times



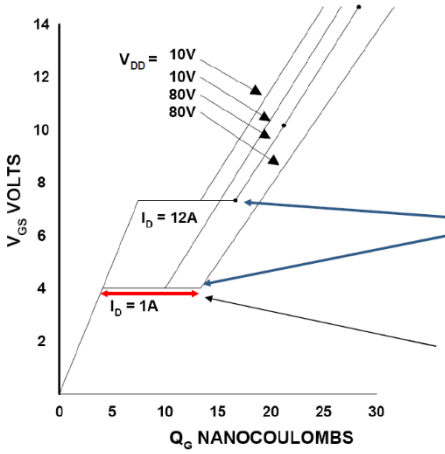
- R_{ON} and R_{OFF} properly chosen to adjust $(t_{d(on)} + t_r)$ and $(t_{d(off)} + t_f)$

If we use just a simple gate resistance Rg to turn the mosfet on and off we end up with a dependance of the turn on and off time transient one with the other, in fact both t_fall and t_rise depend on Rg.
If we want to decouple the t_fall and t_rise to be regulated separately, a simple way to do this is by using diode to allow or impede the flow of current.

If we work in a condition different from the one in the datasheet for Id and Vgs, can we still use the same gate-charge curve? Yes.

In the next image we have several gate charge curves. We can notice that the plateau region happens in different points, but this is not really a big issue, in general. So the difference in the current is translated into a difference in the plateau region, but it is not an issue because in the plateau region the mosfet is working in the saturation regime, so I_d = I_L.
In conclusion, the **Vgs_plateau depends on the square root of the load current**, which is in the end a small dependance.

The result is that we can use the standard gate charge curve reported in the datasheet even if we are working at a different current.



- Plateau region:
  - $I_D = K(V_{GS} - V_T)^2$
  - $I_D = I_L$
- therefore:
  - $V_{GS_P} = V_T + \sqrt{\frac{I_L}{K}}$

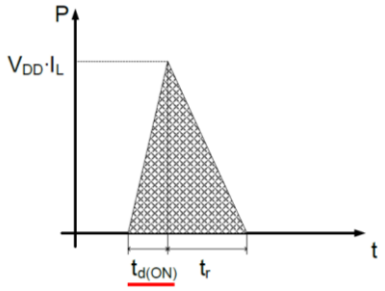
$$Q_{PLATEAU} = \int_0^{V_{DD}} C_{gd} dV_{DS}$$

- Gate charge curve at rated I_D is usually provided

Moreover, if also the Vds at which I'm working is different with respect to the one reported in the datasheet I don't care. The effect of a different final Vds, so bias polarity, is in the curve after the second inflection point → the plateau region is either increased or decreased. If the plateau is larger, also Q_plateau is increased.

SWITCHIN LOSSES

- Turn-on transient



$$E_{off/on} = \frac{P_p \cdot (t_{d(on)} + t_r)}{2} = \frac{V_{DD} \cdot I_L \cdot t_{on}}{2}$$

where

$$t_{on} = t_{d(on)} + t_r$$

$$P_{off/on} = f_{sw} \cdot E_{off/on} = f_{sw} \cdot \frac{V_{DD} \cdot I_L \cdot t_{on}}{2}$$

For example: f_{sw}=100kHz, I_D=6A, t_{on}=100$_{ns}$, V_{DD}=12V

$$P_{off/on} = 360mW$$

- Same approach for turn-off loss calculation

We still use the gate charge curve. Turn on and turn off transient have instantaneous power dissipation. In the turn on transient it happens during td(on), because the voltage is high and the current is ramping up → power dissipation increases linearly. Peak power dissipation is the highest when the current is I_L and the voltage is Vdd. Then the current stays constant and the voltage drops; again we have an instantaneous power dissipation that reduces linearly to 0. In the last section the power dissipation is negligible.

Power dissipation during the turn off transient is not relevant during td(off), but it is during t_fall and t_dis.

Turn on transient

As we can see in the image, the power dissipation has a triangular shape. The base of the triangle includes two time intervals, td(on) and t_rise. This instantaneous power dissipation profile is typical of the so-called **hard switches**. On the other hand we have soft switches, where we use resonances to have oscillations and we turn on or off the mosfet when the voltage Vds is 0 or when the current is 0.

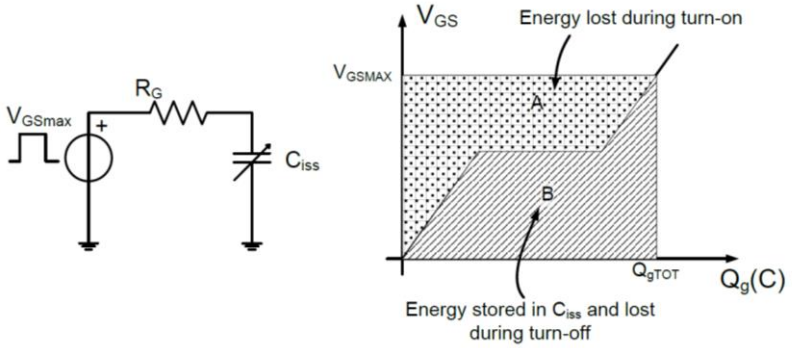
We can calculate the energy dissipated during the turn on, and it is the area of the triangle, and it is the energy spent in the transition between the off and on stages.

As for the average power dissipation (called switches loss), we have to assume that the switch is turned on and off periodically. To compute it we multiply the energy for the switching frequency, and we get the average power dissipation during the turn on, if we use just the energy for the turn on transient.

Gate driver loss

Another advantage of the gate charge curves is that they allow to calculate very easily the gate driver loss. The gate terminal of the power mosfet has to be driven by a gate driver, and this driver has to turn on the mosfet and off, so it has to spent and dissipate some energy.

There is an energy that enters into play during the turn on and during the turn off that is provided by the gate driver. How much is this energy?
The energy provided by the gate driver is simply the product between the total gate charge and Vgs,max. This is the energy to be provided to the mosfet during the turn on and extracted from the mosfet during the turn off.



$$E_{driver} = Q_{gTOT} \cdot V_{GSMAX} \Rightarrow P_{driver} = f_{sw} \cdot E_{driver} = f_{sw} \cdot Q_{gTOT} \cdot V_{GSMAX}$$

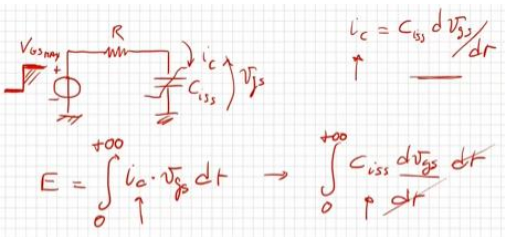
For example: $f_{sw} = 100kHz$, $Q_{gTOT} = 15nC$, $V_{GSMAX} = 12V$

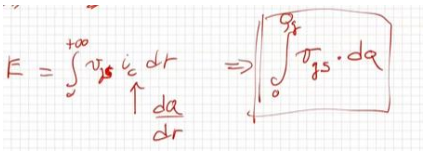
$$P_{driver} = 15mW$$

We can consider the equivalent circuit on the left, representing the gate driver with a Thevenin equivalent, and the gate is replaced with a nonlinear Ciss.
The area B subtended by the gate charge curve, in the charge interval between 0 and Qg,tot, is the charge stored in the Ciss once we have turned on the mosfet that has to be dissipated during the turn off. Whereas the area A is the energy lost across Rg when the mosfet is turned on.

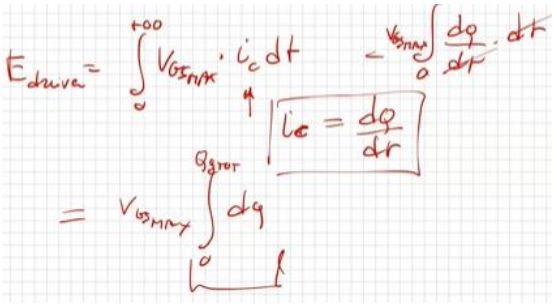
Edriver is the sum of the aera of the rectangle and the two triangles, once decided Vgs,max.

We can demonstrate the areas A and B. Let's assume we are turning on the mosfet; the Ciss is initially discharged and the gate driver voltage steps up between 0 and Vgs,max. We want to understand the energy stored in the Ciss, which is the integral between 0 and +inf of the product between ic and Vgs. However, having Ciss in the integral is not a good approach because Ciss is not constant.





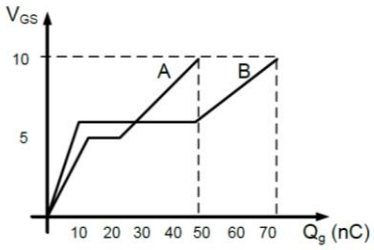
The one in the rectangle is the area subtended by the gate charge curve.
The total area A + B is the energy put into play by the gate driver.



The last integral is nothing else than the area of the rectangle.

SWITCHING PERFORMANCE

The typical mistake when selecting a power mosfet to get the fastest possible mosfet is looking at the Ciss and the larger the Ciss, the slower the turn on. But this is wrong, **we have to look at the total gate charge to compare different switching speeds of different mosfets**.



	A STVHD90	B SGSP386
Cell density/in^2	2.3 M	550 K
$R_{DS(on)}$ mΩ	23	40
C_{iss} pF	3000	1800
C_{oss} pF	1000	1100
C_{rss} pF	180	550
Gate charge nC	47	74

- Check the gate charge curve before deciding which MOSFET really has the edge in switching performance !

An example is in the image above. Mosfet B reaches the same Vgs with a higher Qg, but the Ciss is theoretically smaller.
The real difference is in the Crss, where it is small it means that the plateau region is smaller.

DRAIN TO SOURCE CAPACITANCE – Cds

- $C_{DS} = C_{oss} - C_{GD}$
- C_{DS} contributes to losses
- C_{DS} is highly nonlinear → $C_{DS} \approx \frac{C_O}{\sqrt{V_{DS}}}$ for $v_{DS} > V_{built-in}$

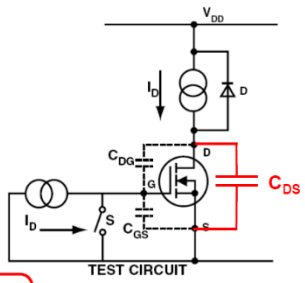
The energy stored in $C_{DS}(v_{DS})$ during the turn-off transition is:

$$W_{C_{DS}} \approx \int_{\text{turn-off transition}} v_{DS} \cdot i_C \, dt \approx \int_0^{V_{DD}} v_{DS} \cdot \frac{C_O}{\sqrt{V_{DS}}} dv_{DS}$$

$$= \frac{2}{3} C_O \cdot V_{DD}^{3/2} = \frac{1}{2}\left(\frac{4}{3} C_{DS}(V_{DD})\right) \cdot V_{DD}^2$$

$$= \frac{1}{2} C_{DS\_eq} \cdot V_{DD}^2 \quad \text{with} \quad C_{DS\_eq} = \frac{4}{3} C_{DS}(V_{DD})$$

$$P_{SW_{C_{DS}}} = \frac{1}{2} C_{DS\_eq} \cdot V_{DD}^2 \cdot f_{SW} \Rightarrow \text{MOSFET turn-on switching loss due to } C_{DS} \text{ discharge}$$



See part 3_additional reading materials: 3.4_MOSFET disipation due to Cds

The role of Cds is negligible as far as we deal with the turn on and turn off transients.
However, the Cds is contributing to the average switching losses of the mosfet, so we have to consider it when calculating the mosfet losses.

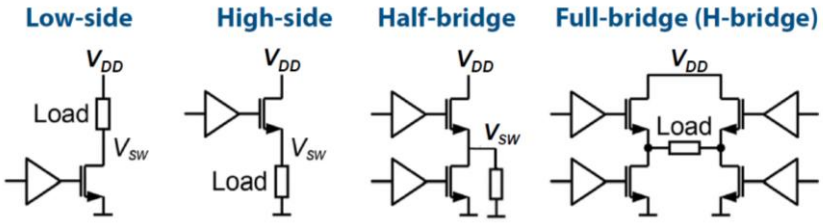
In fact, if we look at the figure and we assume the mosfet is turned off, the drain voltage ramps up to Vdd and the capacitor Cds is charged up to Vdd. In the opposite transition, when the mosfet is turned off, the energy stored in the Cds is discharged through the mosfet that is turning off, producing an energy dissipation. Of course, if we do this periodically, we translate the energy dissipation into a power dissipation.
So we have to calculate the energy stored in the Cds during the off time period that is released to the mosfet during the on time period. The energy is not ½*Cds*Vdd^2 because the Cds is not a constant capacitor.

The Cds is not reported in the datasheet, we find Coss and Cgd, and Cds = Coss – Cgd (or –Crss).
In the slide, W indicates the energy related to Cds (Co is a constant). It is the energy stored in the capacitor during the off transition and lost during the on transition.

Especially at high frequencies, the contribution of Cds to the overall switching loss becomes very important, and it must be considered also the parasitic capacitances more than just Cds (mainly in the buck converter).

POWER SWITCH CONFIGURATIONS

There are at least 4 power switch configurations depending on the position of the power semiconductor switch with respect to the load.



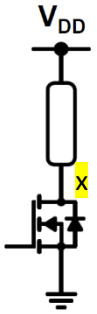
- Transistors can be on-chip (integrated) or discrete (external)
- On-chip: n- and p-channel transistor can be used as a power device if voltage ratings are suitable.
  - trade-off on-resistance R_{DSon} vs. C_{gate} (losses)
- High-power and high-voltage applications prefer n-channel transistors due to lower on-resistance R_{DSon} but gate supply > V_{DD} is required

- **Low side switch**: switch located between ground and load, and load connected to PS.
- **High side switch**: switch connected between the bias supply and load, and load connected to ground.
- **Half bridge**: combination of the previous two. The load is typically connected between the central node and ground.
- **Full bridge**: by connecting two half bridge configurations. It is typically used in class D amplifiers.

In general, the transistor can be on chip, so integrated with the rest of the circuit, or discrete, outside the control IC. The choice depends on the power level. A discrete power mosfet shows an Rds(on) of few mV, so it can handle large currents and block high voltages. In general, current handling capability of an on chip power mosfet is smaller.

In general, when we use a discrete mosfet, the choice is almost always an n channel power transistor because, at the same area, the mobility of electrons is higher than the mobility of holes, so the Rds(on) of the nMOS is smaller than the one of the pMOS.

LOW-SIDE SWITCHES



Low-side switch

Pros

- Easy to Drive
- Lower price driver

Cons

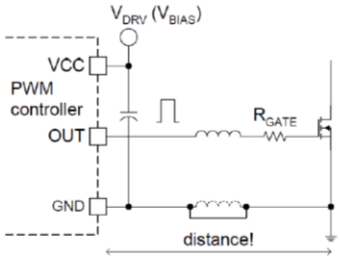
- No protection from shorts to ground
- Possible load corrosion

Low side switches refer to ground, so they are very easy to be driven because the gate driver refers to ground. This easiness in driving translates into a lower price. The drawbacks are that they don't offer any protections from shorts to ground. It means that if we accidentally get a short circuit between node x and

ground, the load is activated. This is a serious safety problem in automotive applications. Moreover, we might also have electrochemical corrosion if we leave the load attached to the positive PS.

Direct driver

Can I directly use the pwm output of a microprocessor to drive a power mosfet? Yes, but with some limitations.
The voltage generated by the uC is typically 5V, so we cannot use it to drive a vertical nMOS transistor with a maximum Vgs of 10V, but we can directly drive a logic-level power mosfet.
However, there is another problem. The maximum current that can be sourced by Arduino is 40mA, and if we have a Ciss huge it takes a long time to be charged → we cannot turn on and off the power mosfet at high switching speed.

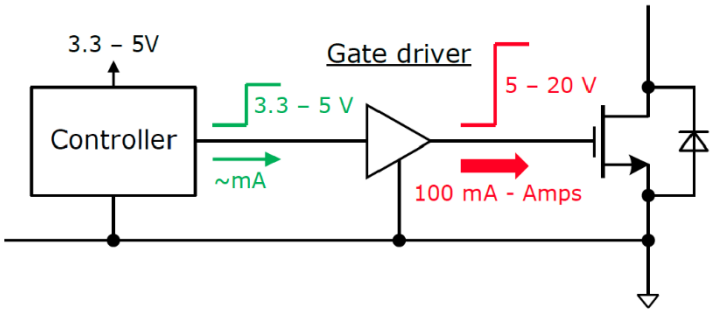


Yes, but..

- Logic-level power MOSFET (e.g. PWM_{HIGH} = 5V in Arduino)
- Limited current capability of the PWM driver (e.g. 40mA max from Arduino)
- PWM controller power dissipation to be evaluated
- Circuit layout is critical
- Not suitable for high-speed switching applications

Moreover, we must be careful in not connecting the power mosfet too far from the uC, otherwise the parasitic inductances play a role and affect the operations of the circuit.

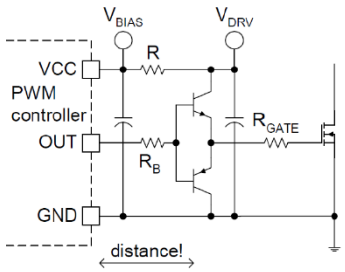
Gate drivers



Controller:	Switch:
› Power supply: 3.3 – 5 V	› On voltage: 5 – 20 V
› Drive current: ~mA range	› $i_g = 10nF\frac{15V}{100ns} = 1.5\ A$

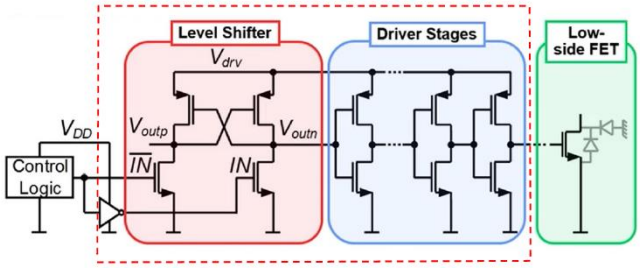
In general, the output level of the controller is too low to drive a power mosfet, so we need gate drivers for level shifting. Moreover, gate drivers are needed because the current that can be provided by the uC is too small, and we need more.

- Bipolar, noninverting totem-pole driver
- Commonly used and cost effective drive for *logic-level power MOSFETs*



The couple pnp-npn transistors implement a current booster.

Low-side driver ICs



- Level shifter followed by driver stage(s) (CMOS inverter)
- Single driver stage if small power MOSFET (low C_{gate})
- Cascaded driver stages if large power MOSFET (large C_{gate})

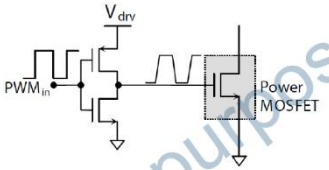
The gate driver typically integrates two stages:

- A level shifter needed to match the low digital voltage provided by the control logic to the relative large voltage to be applied between the gate and the source.
- Driver stage, which is an inverter chain.

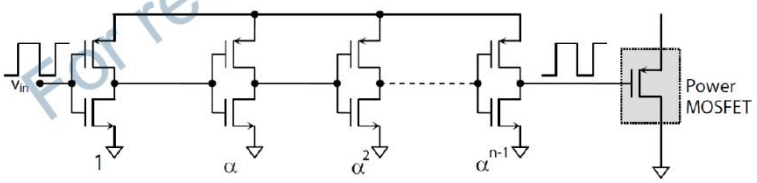
The level shifter is driven with logic levels.

Review of stages

- Single CMOS inverter not suitable to drive large power devices



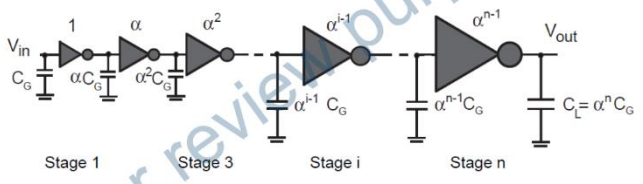
- Common solution: tapered buffer (inverter chain)



Tapered buffer

11

- n stages, each increasing in driver strength (W/L) by factor α.
- The first-stage inverter is built with minimum-sized transistors.
- Assume that the output load capacitance of each stage is primarily determined by the input (gate) capacitance of the next stage.



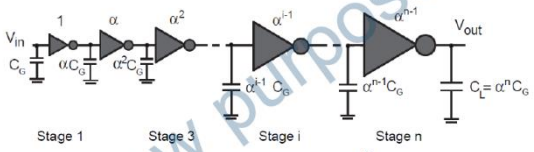
$$\alpha = \sqrt[n]{\frac{C_L}{C_G}}$$

- Different design goals depending on application needs:
  - design for minimum delay
  - design for minimum power (correlates with die area)

Tapered buffer: optimization for speed

12

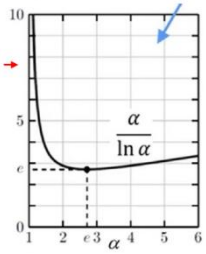
- Assume that the delay of the first stage driving an identical one is t_{apd}.
- The the delay of the i^{th} stage is: $t_{di} = \alpha \cdot t_{apd}$



➢ The global delay of the n stages is: $t_d = \sum_{i=1}^{n} t_{di} = n \cdot \alpha \cdot t_{apd}$

which, combined with $\alpha = \sqrt[n]{\frac{C_L}{C_G}}$, gives:

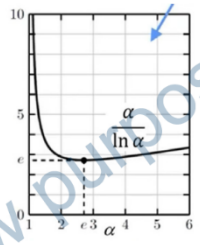
$$t_d = t_{apd} \frac{\alpha}{\ln \alpha} \ln\left(\frac{C_L}{C_G}\right)$$



Example

1

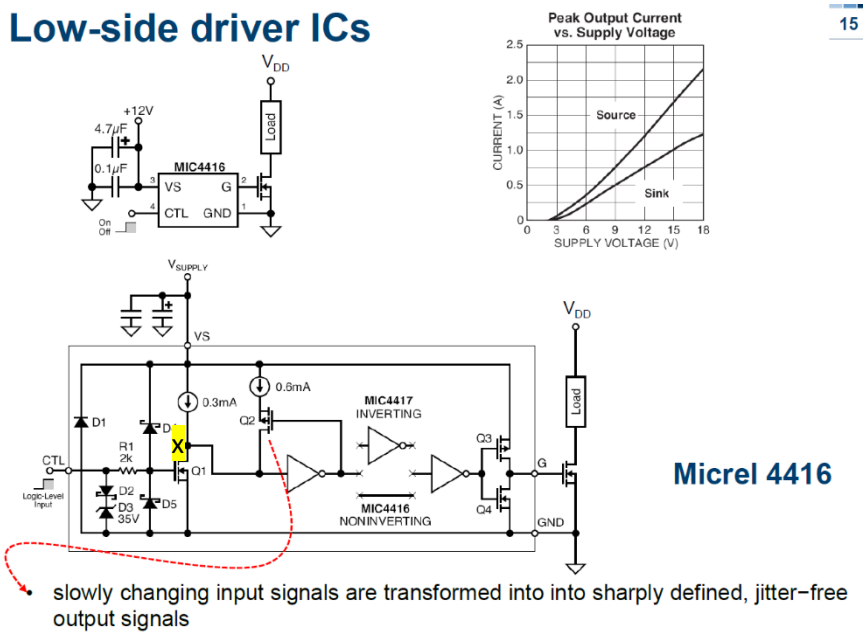
- Speed optimization



- Example : C_L = 300 pF, C_G = 30 fF, t_{apd} = 1 ns
  - α_{opt} = e, n_{opt}=ln(C_L/C_G) = 9 → t_{d_opt} = 25 ns
- However, by choosing a larger α, e.g.
  - α = 6.3, n = 5 → t_d = 32 ns

  we can reduce the number of stages in the buffer with little degradation of the overall delay.

In the next image we have an example of a low-side inverter. The digital level is applied to the CTL termina, we bias the IC with 12V and the output of the IC is connected to the gate of the mosfet, and that's all. We need however to evaluate the turn on and turn off transients.



In the top right image we have the current source and sink capabilities of the IC, and they vary depending on the supply voltage.

If we open the schematic, x is a level shifter, then we have a first inverter which has a positive feedback to sharpen the waveform, then we have the chain of inverters, another inverter and finally the a CMOS inverter as a last stage. So we are charging and discharging the gate of the power mosfet using constant currents from the CMOS inverter.

HIGH-SIDE SWITCHES

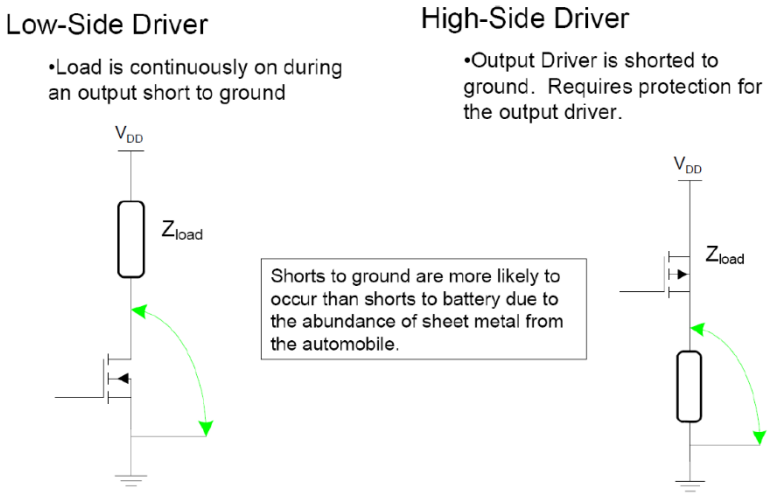
Pros

- Protection from shorts to ground: if the mosfet is on and we have an accidental node between the middle node and ground nothing happens.

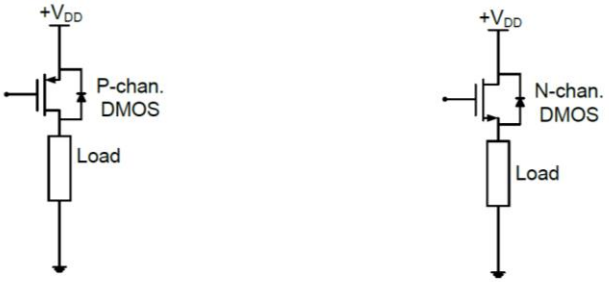
Cons

- Driving a high side switch is more difficult because the source of the mosfet is not grounded, not linked to a constant voltage. So we have to consider a sort of floating gate driver which is able to follow the source, which spans from 0V to Vdd.

The reason why high side switches are preferred in automotive applications is very simple. In fact, if we consider a car it is extremely easy to get in touch with pieces of metal that are conductive and we create a short to ground. With low side switches the load is activated and it is dangerous.



P vs N channel high-side switch



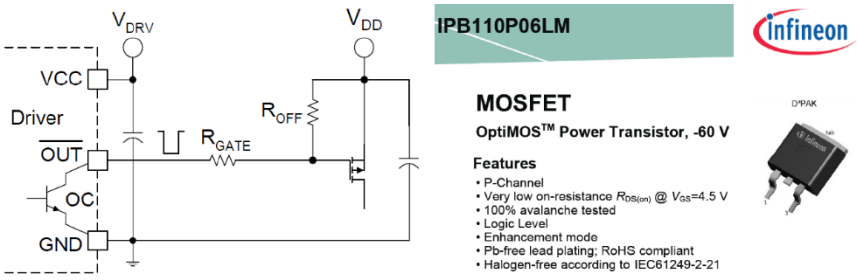
- P-channel MOSFET easier to drive but..
- R_{on} of p-channel is ~ a factor 2 higher than that of a n-channel having the same chip area.

⇨ **n-channel MOSFETs are mostly used as power switches**

We can implement a high side switch either with a pMOS, and the driver circuit is easy to be implemented, or a nMOS, and the gate driver is more complex.
The usage of the nMOS is preferred because the Rds(on) for the same area is smaller with respect to the pMOS.

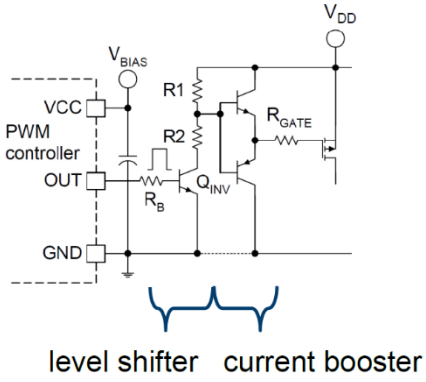
However, there are some pMOS available on market, like the following one, by Infineon®.
To drive a pMOS in a high side position in principle we can use an open collector driver, we basically connect its output with a voltage divider and in the end we have a Vgs given by Vdd divided according to the voltage divider. Of course we can regulate Roff and Rgate to get the desired Vgs.

Once again, the current capability of the driver is limited so we cannot switch fastly the power mosfet. moreover, another problem comes from the potential variability of the Vdd PS. In fact, if we have fluctuations in the Vdd we have also fluctuations in the Vgs.



- Use an open collector driver.
- Gate drive voltage is a function of the power line voltage, V_{DD} due to the R1, R2 divider.
- Limited current capability (high drive impedance).
- Not suitable for high-speed switching applications.

Level-shifted driver for high side pMOS switches



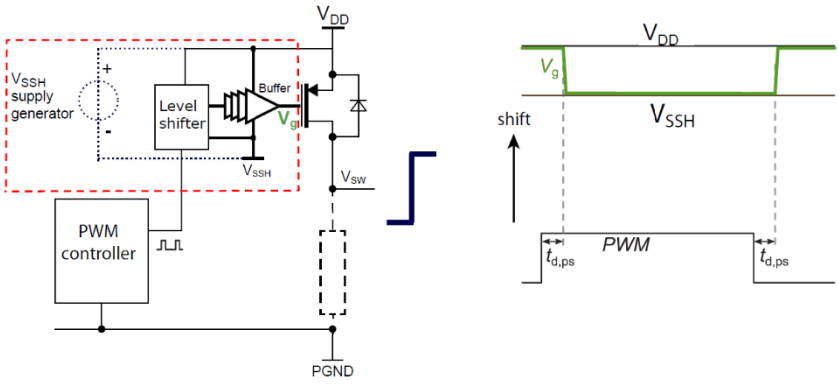
Disadvantage

- Gate drive voltage is still a function of the power line voltage, V_{DD} due to the R_1, R_2 divider.

To improve the previous solution we can use a current booster combined with a level shifter to drive quickly the p power mosfet.
However, we still have the Vgs dependance on the power voltage.

Hig-side pMOS switch – driver IC

The gate driver nowadays is not implemented with discrete components, we use IC acting as high side gate driver.



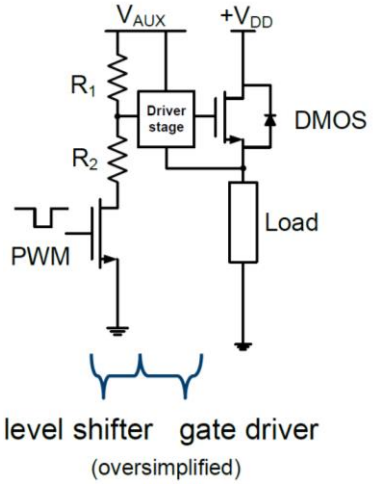
⇨ **Level shifting of a signal from one static level to another**

In the case of the pMOS, the typical IC driver is a combination of a level shifter and the tapered inverter chain. The voltage applied to the gate of the power mosfet swings between Vdd and a second reference Vssh.

The Vssh voltage is generated with an internal voltage generator integrated inside the circuit done either with a linear regulator or a charge pump.

Driving a nMOS high-side switch

First approach



- For efficient chip area, a n-channel DMOS is used as power switch.
- V_{AUX} must be higher than V_{DD} to drive the DMOS into the ohmic region with low R_{dson}.
- $V_{AUX} = V_G(on) = V_{DD} + V_{GSon}$
- V_{AUX} strongly related to V_{DD} (application).
- High power dissipation when the power NMOS is off.

The reason for using a nMOS for high-side switch is the lower Rds(on) with respect to a pMOS device. Let's use a simple gate driver which includes a level shifter that is a resistive divider driven by a control mosfet and a driver stage, which is usually a tapered inverter chain.

The problem with nMOS is that when we want to turn them on we have to bring the channel in the ohmic region, and this means that the final value of the source voltage will be closed to Vdd. To keep the mosfet in the ohmic region, Vg > Vdd, because Vgs > Vt. Hence we end up in needing a Vg larger than Vdd. This is the rational for using an auxiliary voltage generator Vaux which provides a voltage larger than Vdd.

But this is something we would like to avoid.

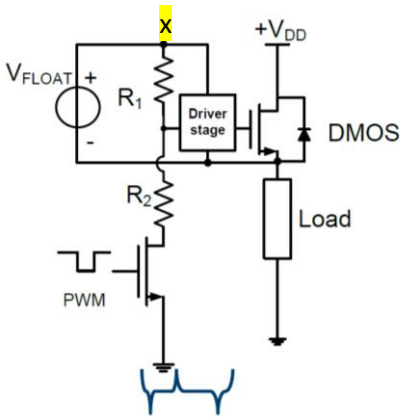
With this implementation there is also another problem. In fact, we want to keep the mosfet off, and to do so we are turning on the control input mosfet and we have a flow of current in the voltage divider, so **we are absorbing current from the Vaux and hence dissipating power**. The larger Vaux, the larger the power dissipation, for the same current.

Second approach

Instead of using a Vaux voltage generator which is referenced to ground, we can replace the Vaux with a floating voltage generator Vfloat. The lower terminal of the generator is connected to the source, so if the source is moving up and down depending on the fact that we are turning on and off the mosfet, the floating voltage generator follows.

We need the Vfloat to be able to generate efficiently high voltages to bring the mosfet in ohmic region, so Vfloat must be at least equal to the Vgs,on that we want to apply to the nMOS.

When the mosfet is off we are bringing the source of the power mosfet to 0 and so the voltage at node x with respect to ground is equal to Vfloat, and so the power dissipated by the floating voltage generator is Vfloat multiplied by the current in the voltage divider.

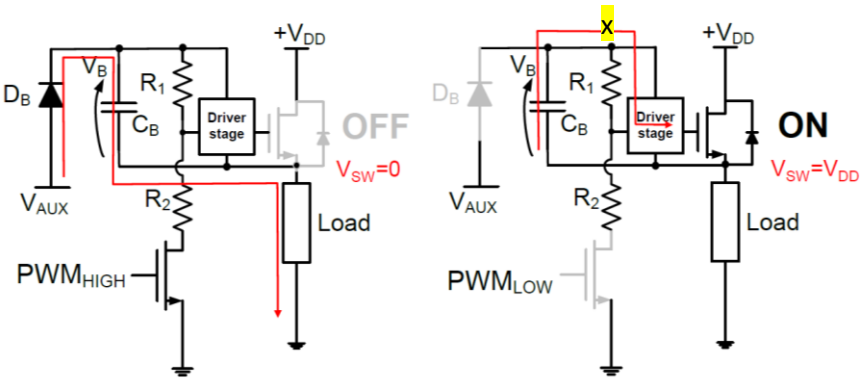


level shifter gate driver
(oversimplified)

- A floating voltage generator referenced to the source of the MOSFET is used instead of a ground-referred V_{AUX}.
- V_{FLOAT} is independent of V_{DD}.
- $V_{FLOAT} = V_{GSon}$.
- *Bootstrap* is a possible way to implement the V_{FLOAT} generator .

One typical way to implement a floating voltage generator is to use a bootstrap. We replace the voltage generator with a big capacitor which is able to keep the voltage across it constant and also able to provide all the charge needed to turn on the power mosfet. But the bootstrap capacitor needs to be charged.

BOOTSTRAP



- V_{AUX} = 10 – 15 V typ., independent of V_{DD}.
- $V_B = V_{AUX} – V_{DB} \sim V_{AUX} – 1V$
- The bootstrap technique is satisfactory for short ON times of a few milliseconds.

The bootstrap network is the combination of a capacitor, which is connected between the source and the upper node, a diode and an auxiliary voltage generator, but the Vaux generates a voltage in the order of 10-15V, that is the voltage needed on the gate of the mosfet to turn it on properly.

The diode is used to charge the bootstrap capacitor when the power mosfet is off (in fact the source is ground and the diode is on), and it disengages when the mosfet is on such that the bootstrap capacitor is able to provide charge to the gate of the power mosfet to turn it on.

During the off time period the bootstrap capacitor is charged up to Vb. During the on time period the diode is off, the current flows in the driver stage and the source increases. Since the bootstrap capacitor

is connected between source and node x and it is keeping the voltage almost constant, as soon as we increase the source the diode is disengages.

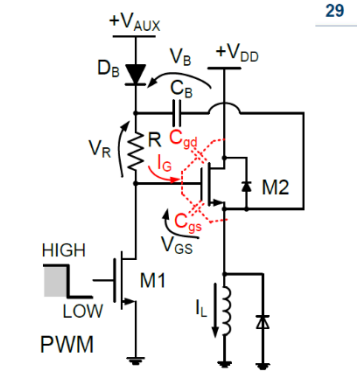
The bootstrap technique is a very good solution if we want to turn on the mosfet for short time periods (few ms). If we want the mosfet on for seconds, the bootstrap is not a good solution.

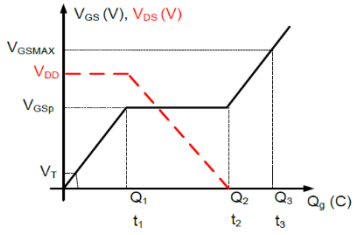
A simple view

Bootstrap: a simple view

29

- PWM = HIGH ; $V_G \sim 0$, $V_S \sim 0 \rightarrow V_{GS} \sim 0$
- PWM = HIGH → LOW; M1 OFF, $I_G = V_R/R$ (assume V_B = const.)
  1) $0 < t < t_1$
     $I_D < I_L$, $V_S = 0$, D_B ON, V_{GS} increases up to V_{GSp}
  2) $t_1 < t < t_2$
     $V_{GS} = V_{GSp} \rightarrow V_R = (V_B - V_{GSp})$ =const,
     I_G=const charges $C_{gd} \rightarrow V_S$ increases
     making D_B to turn off
  3) $t_2 < t < t_3$
     M2 into ohmic region → $V_S \sim V_{DD}$
     $I_G = (V_B - V_{GS})/R$ charges both C_{gs} and C_{gd}
     until $V_{GSmax} \sim V_B$, i.e., **$V_{Gmax} = (V_{DD} + V_B)$**



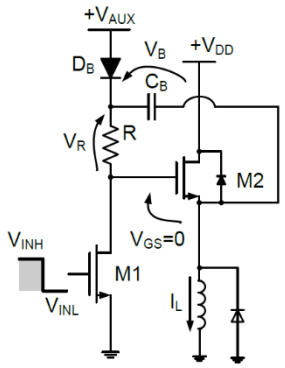


For the sake of simplicity the second resistor of the voltage divider has been removed. M1 is the driving mosfet, M2 the power mosfet. We assume that the load is an inductive load with a free-wheeling diode.

Let's start from a situation where the PWM is high. The M1 is on and it works in the ohmic region, keeping the gate of the power mosfet at ground → Vg, Vs and Vgs = 0.
The network starts commutating when the input signal moves from the high level to the low level. For the sake of simplicity we will assume that in the turn on transient the bootstrap capacitor keeps its voltage constant (like in a huge capacitor).

Turn-on transient (0)

- $V_{IN} = V_{INH} \rightarrow V_G \sim 0$, $V_S \sim 0 \rightarrow V_{GS} \sim 0$
  - power MOSFET M2 is OFF
  - $V_R \approx (V_{AUX} - V_{DB}) \approx V_B$
  - $I_R = V_R/R$



Input voltage is high level, so gate and source are ground. The current that flows in the network, assuming that M1 is in the ohmic region with 0 voltage drop across drain and source is I = (Vaux – Vd)/R, where Vd is the voltage drop across the diode. Vaux – Vd is called bootstrap voltage (Vb) because it is exactly the voltage on the bootstrap capacitor Cb.

Turn-on transient (1)

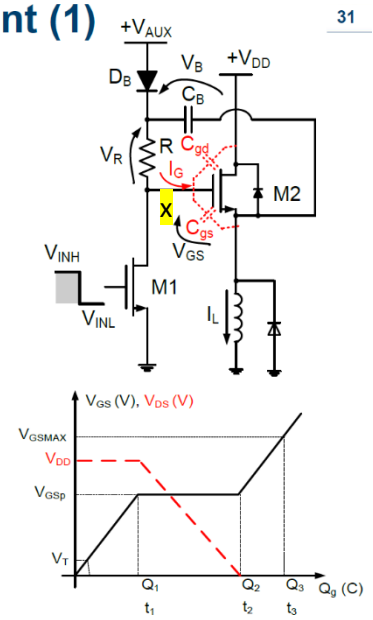
Bootstrap: turn-on transient (1)

31

- $V_{IN} = V_{INH} \rightarrow V_{INL}$, M1 OFF;
  (assume V_B = const.)

1) $0 < t < t_1$

- $i_G(t) = (V_B - v_{GS}(t))/R = v_R(t)/R$
- i_G charges C_{gs} and C_{gd}
- $i_D < I_L$, $V_S = 0$, D1 ON
- v_{GS} increases up to V_{GSp}



If we look at the gate charge curve, initially the mosfet is off and hence Vds = Vdd. At t = 0 I turn off M1 moving to the low voltage level.
The current that was initially flowing in M1 is redirected in the gate of M2 because there is a huge input gate capacitor that keeps the node x, at the very beginning of the transient, unchanged. In fact, we cannot change the voltage across a capacitor instantaneously. This current charges Cgs and Cgd.
Nothing happens to the source until we reach the threshold, so the inductive load is clamping the source at 0V. Hence in the initial part of the transient I'm simply charging Cgs and Cgd to make the Vg to increase.

At a given point I reach the threshold. Some current i_d will start to flow in M2, but as long as i_d < I_L, the voltage at the source is still clamped at ground because there is current flowing in the free-wheeling diode.
The real breakthrough happens at time t1 in the gate charge curve.

The current that charges Cgs and Cgd is the current that flows in the resistor R, which is (Vb – Vgs(t))/R.

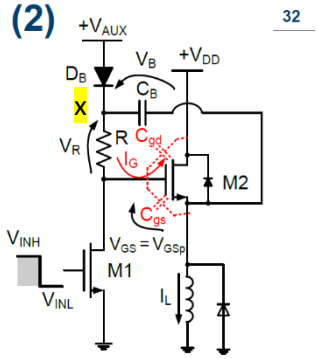
Turn-on transient (2)

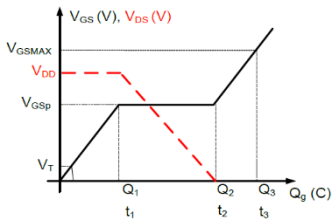
Bootstrap: turn-on transient (2)

32

2) $t_1 < t < t_2$

- $V_{GS} = V_{GSp}$
- $\rightarrow V_R = (V_B - V_{GSp})$ =const,
- $I_G = (V_B - V_{GSp})/R$ = const
- I_G charges C_{gd}
- $\rightarrow v_S$ increases making D_B to turn off
- $\rightarrow v_{DS}$ decreases





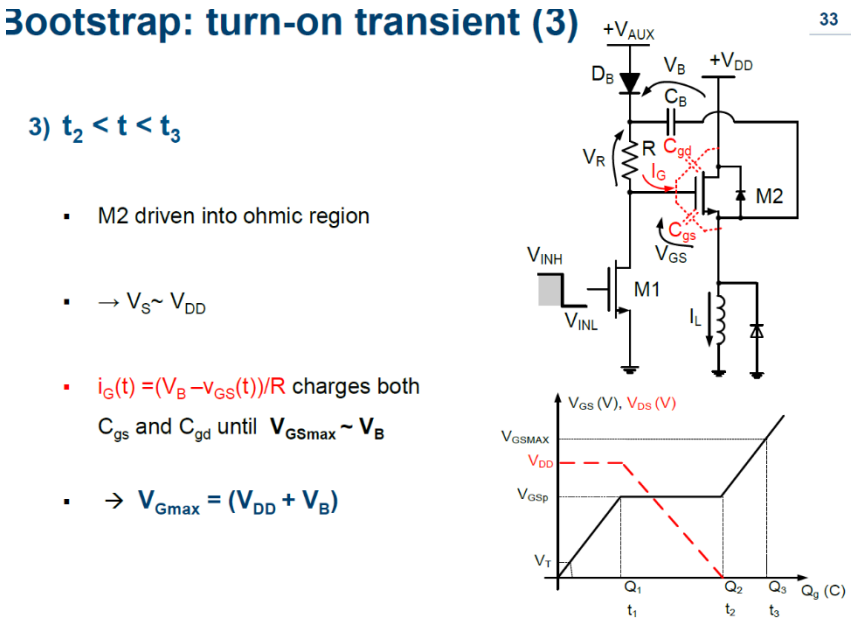
In phase 2 the mosfet enters the plateau region, and this happens when i_d = I_L. At this point the free-wheeling diode is disengaged and the mosfet current is forced to be equal to the load current, which is constant.

Vgs is now clamped because the mosfet is operating in the saturation region and so if we are setting the current that flows in the mosfet we are also setting the Vgs, which is equal to Vgs_plateau.
Once again, the current that enters in the gate of M2 is the current flowing in the resistor, which is (Vb – Vgs_plateau)/R → gate charged with a constant current.

This current doesn't go through the Cgs, because the voltage Vgs is clamped, so all the current injected in the gate is flowing through Cgd. As a consequence, the gate voltage is increasing, but also the source voltage is increasing, because Vgs is clamped. If the source voltage increases, also node x increases, and the diode Db is turned off.

From now on, all the charge needed to charge the gate of the mosfet will be provided by the bootstrap capacitor. Of course, during the plateau the Vds decreases because the source is increasing and the drain is constantly biased at Vdd.

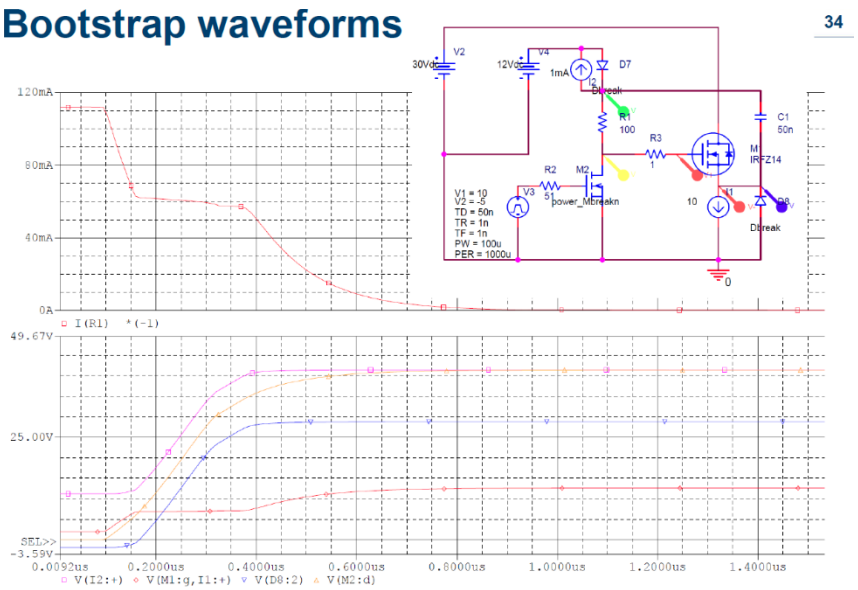
Turn-on transient (3)



I've passed the plateau region, meaning that I'm entering the ohmic region. We can assume that in the ohmic region Vs is almost Vdd and Vds = 0.
As for the current that flows into the gate, there is still current in it, set by (Vb – Vgs(t))/R. Now Vgs is free to increase because we are no more in the plateau region.
This current still charges the Cgs and Cgd until the Vgs reaches Vb. At this point the current is nihil and the transistor transient is over.

The final value of the gate voltage will be Vdd + Vgs,final = **Vdd + Vb = Vg,max**.
Hence the gate is brought above Vdd by the desired Vgs.

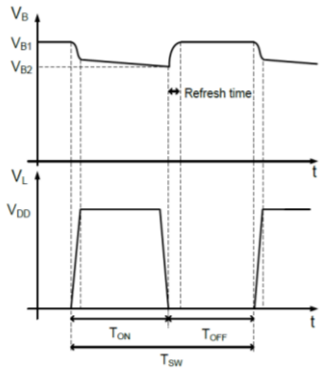
Simulations



Bootstrap capacitor

The boostrap capacitor, C_B, discharges due to:

- charge transfer from C_B to the power MOSFET input capacitance;
- reverse recovery current of the bootstrap diode;
- quiescent current of the driving stage and reverse leakage current of the bootstrap diode.



- T_{ON} connot be indefinitely long
- Refresh time must be sufficiently long to fully restore the charge lost by the bootstrap capacitor

With the bootstrap capacitor there is a problem, and it is the fact that it is discharged due to several causes. In fact, we cannot afford a big capacitor with a very high value.

On the top graph we see the bootstrap voltage as a function of time, and in the bottom one the voltage across the load. When the mosfet is turned on we notice that the voltage across the load is increased (second plot). Then the voltage is kept constant across the load (Ton).

Let's look at the turn on transient. The bootstrap voltage is initially equal to Vb1, then there is a sudden drop corresponding to the turn on of the mosfet and then a slow discharge of the bootstrap capacitor during Ton. When we turn off the mosfet the voltage across the bootstrap is restored to the original value. When we turn off the mosfet the source goes back to zero allowing the recharge current to flow in the load and restore the bootstrap voltage → restore happens when we turn off the mosfet.

Now we are more interested in the voltage drop when turning on the mosfet.

The reason for the fast drop in the bootstrap capacitor voltage is very simple. When we turn on the mosfet, we have to provide a charge to the gate, and most of the charge is provided by the bootstrap capacitor. Then the mosfet is turned on, but there is still a small reduction of the bootstrap voltage, and the bootstrap capacitor is discharging with a small current.

When the mosfet is on, the Db is reverse bias and, if so, we still have a small reverse bias current flowing into it. It is the leakage current of the diode (but also of the drain to body junction) that slowly discharges the Cb. **This small discharge is the reason why Ton cannot be indefinitely long**.
Hence refresh to the bootstrap capacitor must be applied, and this happens when we turn off the mosfet.

Vb1 to Vb2 voltage drop and Cb value

- The bootstrap capacitance must keep a gate to source voltage high enough for the time period between two successive turn on transitions.

From $\Delta Q_B = C_B \cdot \Delta V_B$ we get:

$$x\ (V_{B1} - V_{B2}) \cdot C_B = Q_{g_{TOT}} + Q_d + I_{leak} \cdot T_{ON}$$

Note 1:
- Q_{gTOT} is higher than the effective charge taken from C_B

C_B is designed by setting the allowable voltage drop ($V_{B1} - V_{B2}$), that is

$$C_B = \frac{Q_{g_{TOT}} + Q_d + I_{leak} \cdot \frac{1}{f_{SW}}}{V_{B1} - V_{B2}}$$

Note 2:
- T_{ON} is replaced with $1/f_{SW}$, overestimating the leakage current

For example:

$\Delta V_B = V_{B1} - V_{B2} = 2V$ (typ.); $Q_{g_{TOT}} = 12nC$; $Q_d = 5nC$; $I_{leak} = 1\mu A$; $f_{SW} = 100KHz$

$$\Rightarrow \quad C_{B(min)} = 8.5nF$$

We have to perform the charge balance x. The delta_Vb depends on how much charge we are extracting from the bootstrap capacitor during the Ton.
Delta_Vb is something that we would like to set as designers. We know the delta_Qb value because it depends on how much charge we are delivering to the gate and extracted due to the leakage current.

The total charge we are extracting is the sum of three different contributions; the Qg,tot that we have to deliver to the gate, the Qd, reverse recovery charge of the Db (to turn off the diode we need to provide Qd with the bootstrap capacitor) and then we have the charge extracted by the leakage current.

There are some approximations in this calculation. In fact, the Cb doesn't have to deliver the whole Qg,tot, because the charge to the gate in the first time interval up to t1 of the gate charge curve is provided by the auxiliary generator, because the diode is not yet disengaged. So we are making an overestimation.

As a second approximation, the leakage current is discharging the Cb only during the on time period, not during the off one. Nevertheless, the Ton is replaced with the switching period (sum of Ton and Toff).

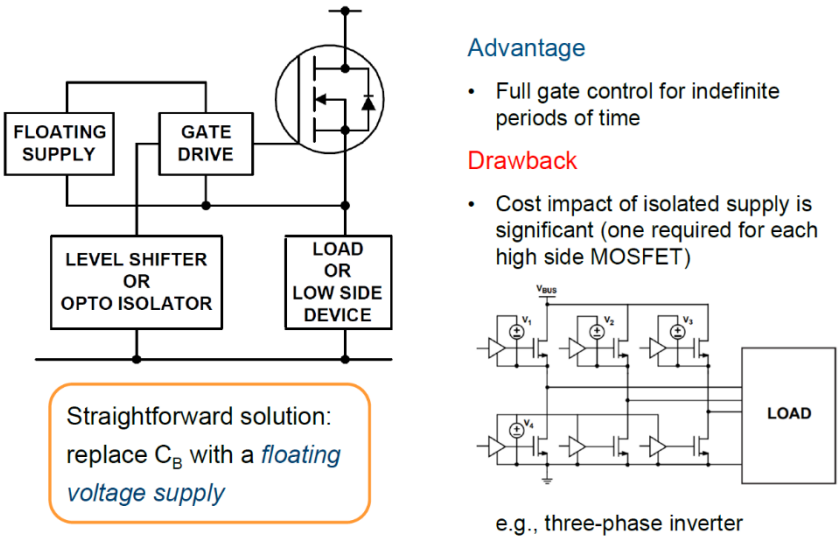
What happens if the power mosfet is kept on for a too long time?
The Cb discharges, as shown in the simulation below. The blue line is the power dissipated by the mosfet, the pink curve represents the voltage across the load. When we turn the mosfet on, initially the voltage across the load is the expected one (PS).
The red curve represents the Vgs of the power mosfet. Initially it is more or less 12V as expected, but due to the discharged of Cb it is reducing.

At a certain point the voltage across the load collapses from 30V to 5V and the Vgs reduces down to 7V. The Vds is 25V, so I'm dissipating a lot → the mosfet has collapsed in an operation region, the saturation, where the current is 10A and Vds = 25V and P = 250W and the mosfet is going to blow.

The mosfet is initially working in the ohmic region, because Vgs = 12 and Vl = 30V, which means that Vds = 0V. However, the Vgs is reducing because we are discharging the bootstrap capacitor. For some time the mosfet still work in the ohmic region up to the point x where the Vgs is not sufficient to keep it into the ohmic region and it enters the saturation region.
7V is the Vgs to be applied to get 10A in the saturation region, and this result can be obtained by looking at the Vds – Ids curve of the mosfet in the datasheet.

Is there any way to increase the on time?

Increasing the on time

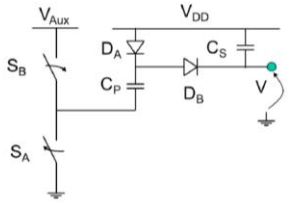


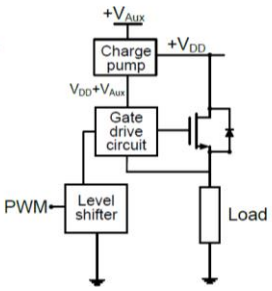
We cannot use the bootstrap, we need to use a floating voltage generator. However, this kind of solution might be not very cost effective.

How can we implement the floating voltage generator?
We have two possible solutions: an isolated DC/DC converter or a charge pump. The former is based on the usage of a high frequency transformer.

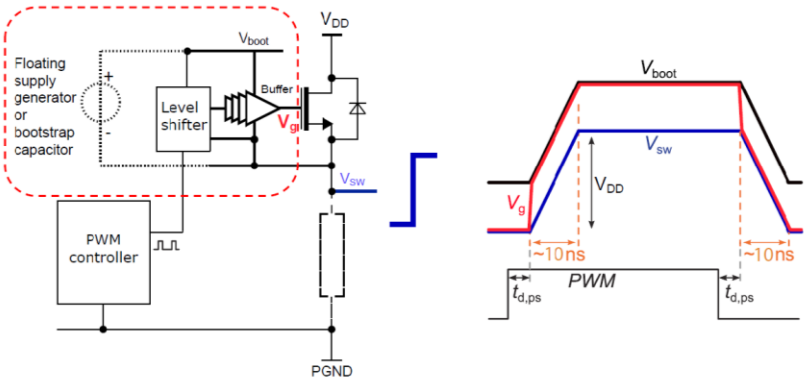
- Isolated DC/DC converter (see part 8)
  - exploits a high-frequency transformer that forms the isolation medium between the common auxiliary power supply and the floating secondary circuit;
  - can be external or fully integrated into the gate driver.

- Charge-pump
  - Used to generate an *over-rail* voltage V≈ V_{DD}+V_{Aux}





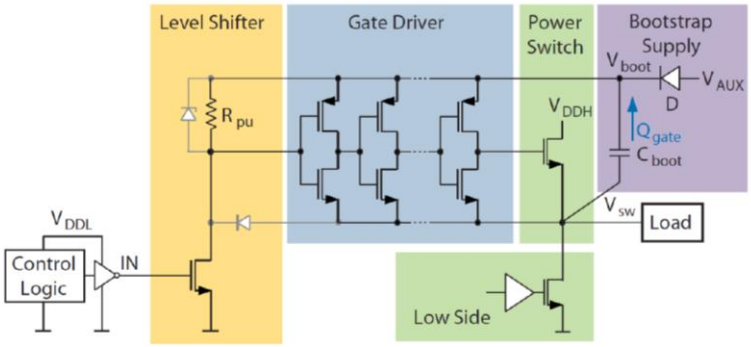
IC DRIVER FOR HIGH-SIDE NMOS SWITCH



- Level shifting of a signal from one static level to a "floating" level
- The driver must tolerate the violent voltage swings occurring during the switching transitions and drive the gate of the MOSFET above the positive supply rail V_{DD}.
  - this makes the high side driver design a challenging task.

The high-side driver is typically an integrated circuit that we can buy on the market and we just need to apply the logic signal to the IC.
Inside the IC there is a combination of a level shifter and a gate driver (tapered inverters cascade). There is, however, a big issue with the level shifter in a high side driver. The big issue comes from the fact that the level shifter is not a static one, but it is floating.
Floating level shifter means that the it is referred to the source voltage of the power mosfet. Hence the red signal is the one we have to apply to the power mosfet to turn it on, and the blue signal is the source voltage. The source voltage moves following the gate.

So the IC driver has to shift the level of the applied digital signal and also it has to float on the source.
In the case of the pMOS the IC driver was a static one.



- Gate driver stage (inverter chain): low voltage transistors even for large V_{DD}
- R_{pu}: tradeoff between speed and DC current (typ. 100 kΩ)
- Diode(s) to clamp voltage across R_{pu} and limit V_{GS} for driver

LEVEL SHIFTERS – KEY PERFORMANCES

1) High speed (N- and P-channel MOS switches)

- propagation delay in the subnanosecond range
  This affects the dead time and high switching frequency operations.

2) dV_{SW}/dt immunity (N-channel MOS switches)

- The switching node toggles between the ground and the input supply. In order to ensure reliable operation, the level shifter must tolerate the fast positive and negative dV_{SW}/dt slewing of several or tens of voltage per nanoseconds.

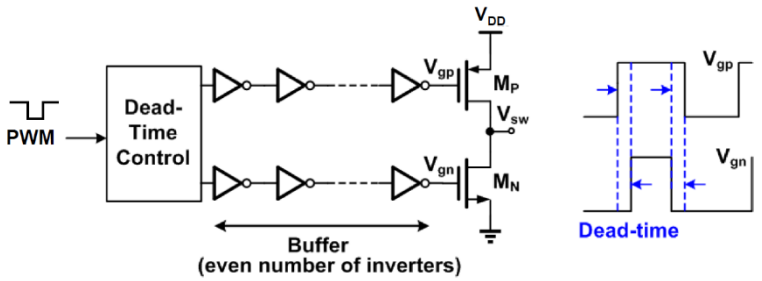
3) Low power dissipation(N- and P-channel MOS switches)

- The current consumption of the level shifter should be in the range of μA. Power dissipation is always a tradeoff with the propagation delay in the practical design.

1. We want high speed because the switching frequency is increasing.

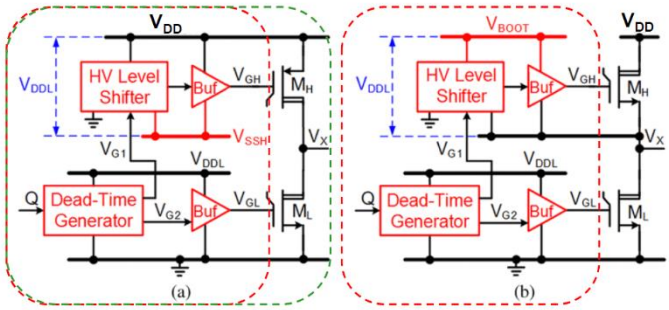
HALF-BRIDGE DRIVERS

Combination of high-side switch and low-side switch. In the image the high side and low side are driven synchronously. It is a good implementation if the power supply is a logic level power supply.



The dead-time control is used to introduce a dead time between the wavefronts of the digital signal controlling the switches. We need it because we might have, in principle, two identical switches for the high and low sides, but in the real world they will never be identical, one might turn on or off faster than the other, so they may be conductive at the same time.

If the Vdd is larger than the logic levels we cannot use the implementation above, we need a level shifter for the high side as below.



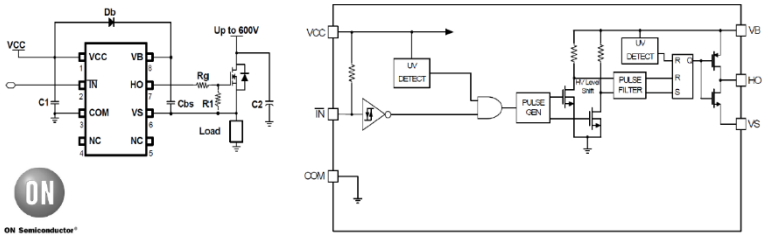
- Structures of generic HV gate drivers for (a) CMOS and (b) dual-nMOS power trains.
- Suitable for $V_{DD} > V_{gsMAX}$

If the application is a high current application, larger than few amps, the solution is the circuit on the right, two nMOS that are discrete components.

Examples

High-side driver IC (FAN7081-GF085)

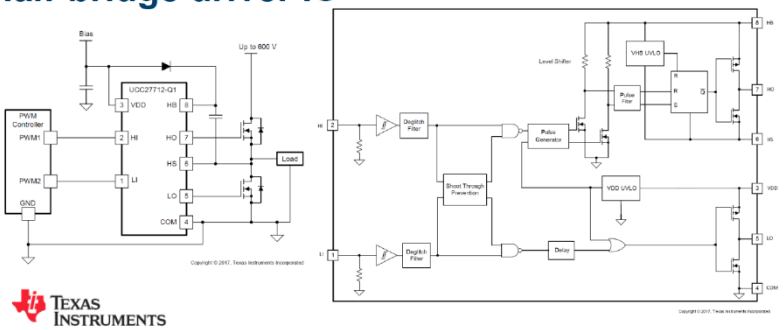
50



- High-side gate drive IC designed to operate up to 600V.
- High-side gate driver operation up to V_S=-5V (typical) at V_{BS}=15V.
- The UVLO circuits prevent from malfunction when V_{CC} and V_{BS} are lower than the specified threshold voltage.
- Minimum source and sink current capability of output driver is 250mA and 500mA respectively

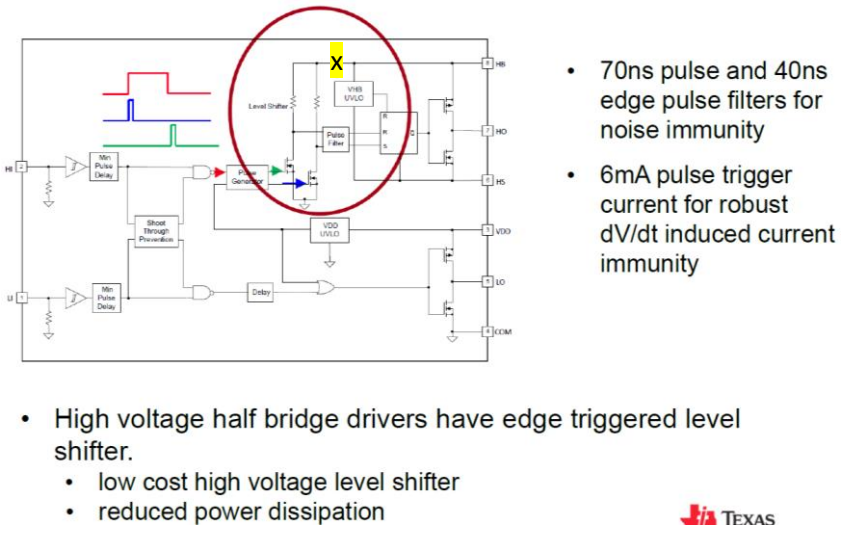
UCC27712-Q1 Automotive, 620-V, half-bridge driver IC

51



- Fully Operational up to 620-V, 700-V absolute maximum on HB Pin
- Peak output current 2.8-A Sink, 1.8-A Source
- dv/dt immunity of 50 V/ns
- Small propagation delay (100-ns Typical)

Level shifter implementation

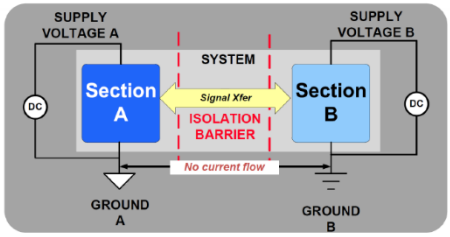


It is an example of how a level shifter can be implemented in a high voltage half bridge driver. It is a simple combination of a mosfet and a resistive load. But what if I have 600V at node x? If the mosfet is always on, I have a lot of current flowing in the resistance and a lot of power dissipation.
The solution to this problem is: the red one is the digital signal I'm applying to the high side mosfet to turn it on or off. This signal is split by the pulse generator in two very short pulses (green and blue). The blue leading edge turns on the mosfet for a very short time and the filp flop is set.
A second fast pulse (green) resets the flip flop and the nMOS is on.

So we are not consuming constant or DC power in the network, we are just using the level shifter in a dynamic way.

GALVANIC ISOLATION

Galvanic isolation is a mean of preventing continuous current or unwanted AC current to flow between two different parts of a circuit. However, it allows the power and signal transfer.



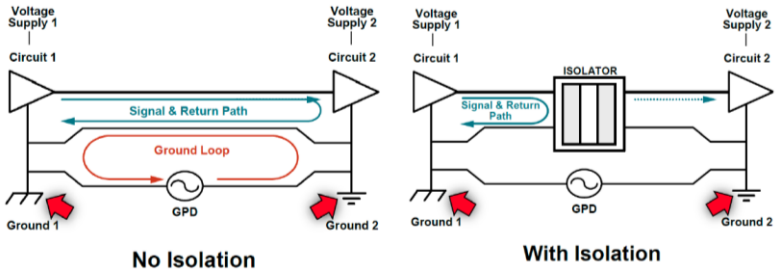
Reasons for Galvanic Isolation

- ✓ Safety of End User
- ✓ Protecting LV circuits from HV Circuits
- ✓ Filtering of Common-Mode Noise
- ✓ Eliminating Ground Loop Noise
- ✓ Level-Shift between Power Domains

- **Galvanic isolation** is a means of preventing DC and unwanted AC currents between two parts of a system, *while allowing signal and power transfer* between those two parts.

Example – breaking the ground loop

- Avoids disruptive ground loops where different ground potentials are involved



- Electrical Installation can cause large GPDs (ground potential difference) between two remote nodes.
- An isolator breaks the ground loop, thus removing signal path noise.
- The GPD yet still exists and the isolator must be robust enough to withstand the large voltage differences.

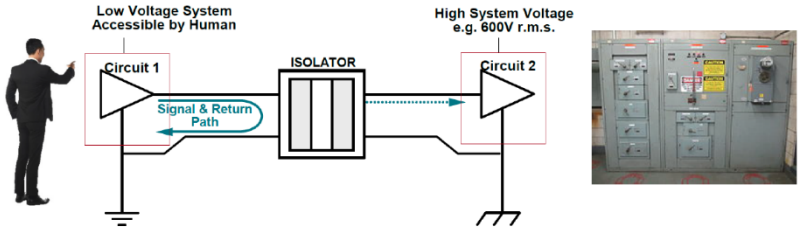
The most common application of galvanic isolation is to break a ground loop. Ground loop is generated when we have multiple connections between different ground levels. Due to the different ground levels we may have current circulating in the loop and we may end up with a differential signal building up in the path and interfering with the signal produced by the instrument.

To avoid this we introduce an isolator, that typically is of a galvanic type.

Example – safety protection against HV

Another reason to use galvanic isolation is for safety purposes.

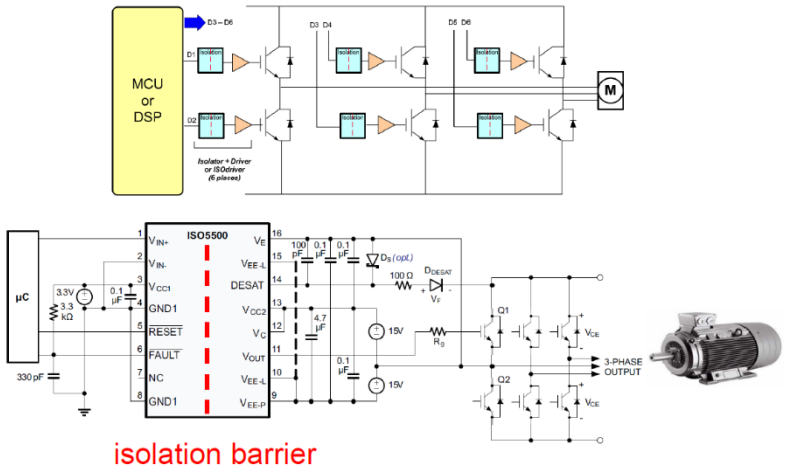
- Protects humans from mains voltage mains voltage



- Industrial equipment running of 100s of volts, temporary overvoltage of 1000s of volts, and 10000v surges.
- Isolation barrier is required to protect low voltage circuitry and human operators

Example – motor control

- Protect microcontroller from high voltage
- Many isolation paths



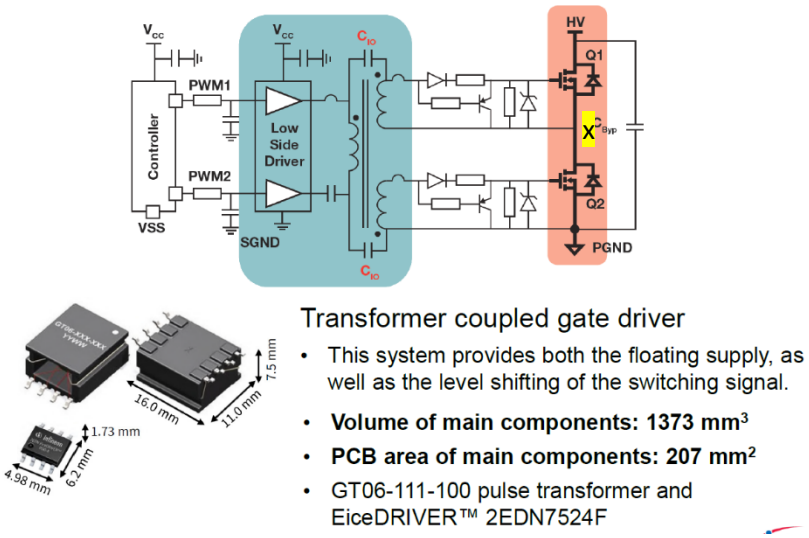
We consider an inductive motor driven with the combination of 3 half bridges inverters (triphase inverter) and each half bridge include a high side and a low side switch and these 6 switches are controlled by a uC.
We don't want that, for any reason, the high voltage we are applying to the motor gets back to the control circuit, destroying it.

ISOLATED GATE DRIVER

Pulse transformer

The isolation is based on the usage of a transformer. In this case the transformer is used double ended with two different PWM signals. This structure is used to drive the half bridge configuration, typically in the case of full bridge and half bridge converters.
PWM2 is typically equal to PWM1 but shifted by 180°.

The direction of the voltage depends on the position of the dots, where we have the dot we have the positive side of the voltage signal.
Moreover, the terminal of the secondary coil is connected to the source of the high side switch x. The secondary coil is able to float, so when the source voltage is increasing because we are turning on the mosfet, the secondary side of the transformer is able to float and bring the node above with the dot to float.



When we apply a positive voltage on the primary coil we are turning off the bottom mosfet (PWM1). In the second time period when we are applying a positive PWM2 pulse, the bottom mosfet is on and the top one is off.

There is a problem with transformers, and it is related to the **interwinding capacitance**, that is the capacitance between the primary and secondary side.
This problem comes from the fast voltage transient that occurs on the source terminal when we turn on the high side switch. If we have an interwinding capacitance Cio we have a current across it because the current is i_c = C*dV/dt, and this current may disturb the operation of the low side part of the circuit.

The common mode transient immunity (CMTI) is a specification that tells us the maximum dV/dt we can afford with our transformer. There is a maximum value above which the system is not working properly.

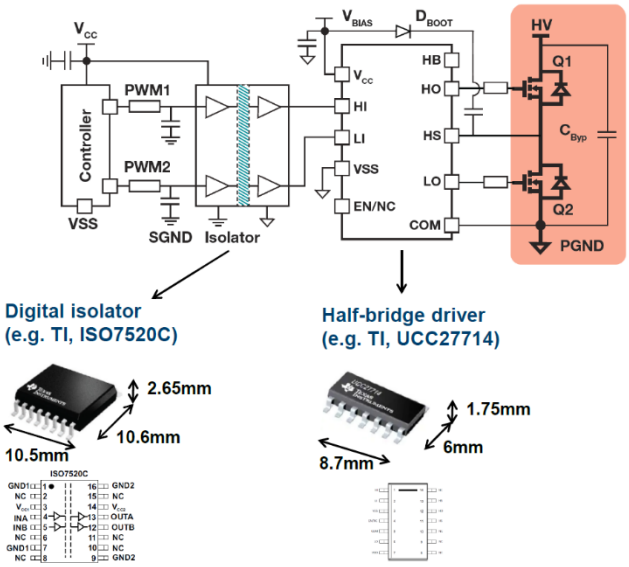
One of the advantages of this solution is that we don't need to provide any bias to the high side of the circuit, the voltage that is needed to turn on the mosfet is directly provided by the secondary side of the transformer.

However, there are two problems.

1. The interlining capacitance Cio coupling the primary and secondary side.
2. Overall size of the system, because we need to consider the size of the transformer and of the low side driver.

To reduce the size of the system we can replace the transformer with a digital isolator.

Digital isolator + driver



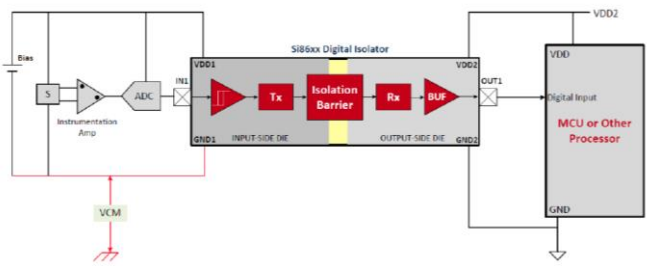
We have still two different PWM signals applied to the input of a digital isolator. The digital isolator is a simple circuit able to transfer a digital signal from the input to the output through an isolation barrier.
At the output of the digital isolator we still have a digital signal, so to drive the half bridge configuration we need a half bridge driver.

An interesting observation is that, in order to properly use the half bridge driver, we need a bias supply Vbias.
This bias supply is typically 10V or 15V, the one we need to provide to the gate and source of high side and low side switches.
The Vbias has to be isolated with respect to the low voltage side of the circuit, so we need to provide it by using an isolated DC/DC converter or something other. This is a potential disadvantage of this approach, because the transformer doesn't need any bias voltage.

The other advantage is the overall reduction of the size.

Digital isolator

- A digital isolator (also known as on-chip isolator) is used to get a digital signal across a galvanic isolation boundary .
- The basic operation of the CMOS digital isolator is analogous to that of the optocoupler.
- An RF carrier is used instead of light.
- The RF transmitter and receiver are separated by an isolation barrier.
- A standard CMOS integrated-circuit process with metal layers used to form transformers or capacitors to magnetically or capacitively couple data across an isolation barrier.



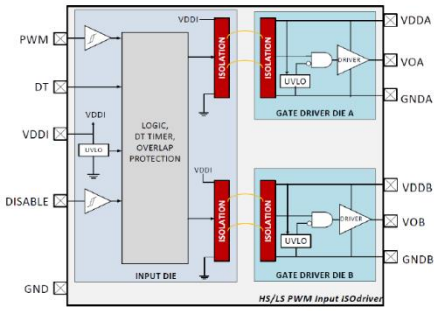
An example of digital isolator is the optocoupler, but opto-isolator are no more used in power electronics because their propagation delay is too high for practical applications. They are replaced by the digital isolators, that instead of using light as propagation mean use a radiofrequency carrier.

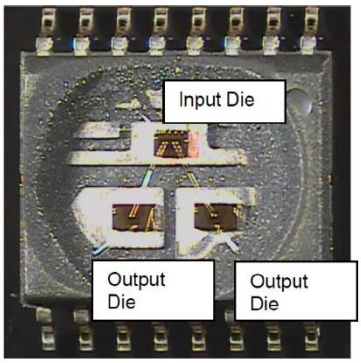
The advantage of digital isolators is that the modulated signal is transmitted over the isolation barrier using either coreless transformer that can be integrated in a CMOS process or capacitors → inductive coupling or capacitive coupling.

The second advantage is that propagations delay are pretty fast.

Integrated isolator + driver

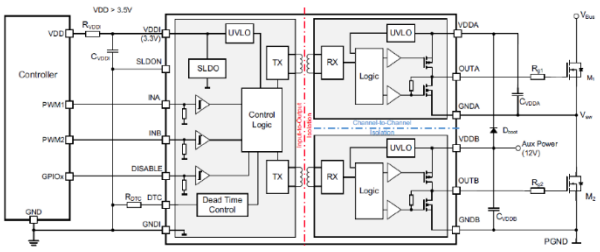
This is the best solution. It is based on the digital isolator and, in a single package (not in a single chip), we integrate 3 chips: a low voltage chip that handles all the low voltage parts of the high side and low side drivers and two isolators that couple the low voltage side to the high voltage side.

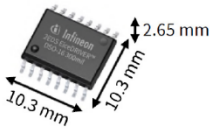




- Usually a three-chip SiP (half-bridge driver)
- Digital isolator technology is implemented

So we have a single package with 3 chips. An example of what we can find in the market is in the image below.



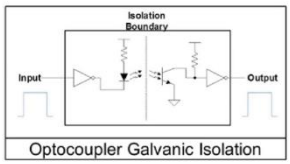


- EiceDRIVER™ 2EDS8265H reinforced isolated gate driver IC
- **Volume of main components: 281 mm³**
- **PCB area of main components: 106 mm²**

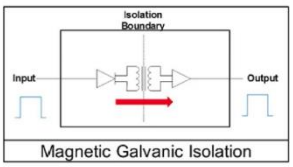
- 4 A / 8 A source / sink output current
- Up to 10 MHz PWM switching frequency
- PWM signal propagation delay typ. 37 ns
- 3 ns channel-to-channel mismatch
- Common Mode Transient Immunity CMTI >150 V/ns

POPULAR ISOLATION METHOD

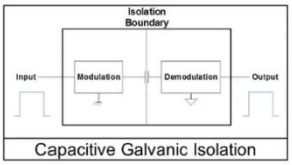
- **Optocoupler**
  - Signal transfer between two isolated circuits using light, LED + phototransistor, 1970s ~ (Avago, Fairchild, Toshiba and others)
- **Transformer**
  - Integrated micro-transformer and electronic circuitry, 2001~ (ADI, Infineon, Rohm and others)
- **Capacitor**
  - Signal transmission through capacitive isolation with On-Off-Keying (OOK) modulation, 2004~(Silabs, TI and others)



Optocoupler Galvanic Isolation



Magnetic Galvanic Isolation



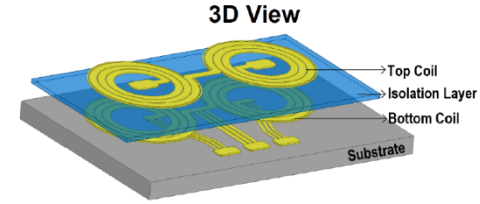
Capacitive Galvanic Isolation

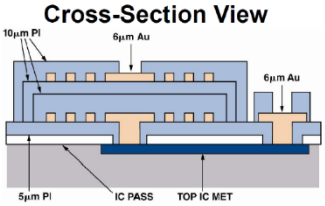
As far as digital isolation is concerned, we have two possibilities:

- Transformer: inductive coupling
- Capacitor: capacitive coupling

Inductive isolation

- ***Coreless planar transformers (CPT)*** were first developed as a solution for insulating the high voltage power circuit from the low voltage control circuit allowing integration on-chip.

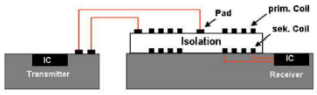




■ Two co-axial planar spiral coils
- Material = Gold
- Thickness = 6µm

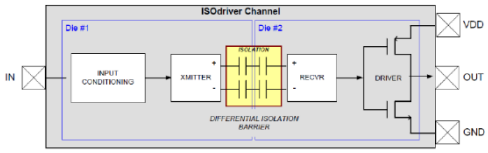
■ Isolation material
- Material = Polyimide } >5kV isolation
- Thickness > 20µm

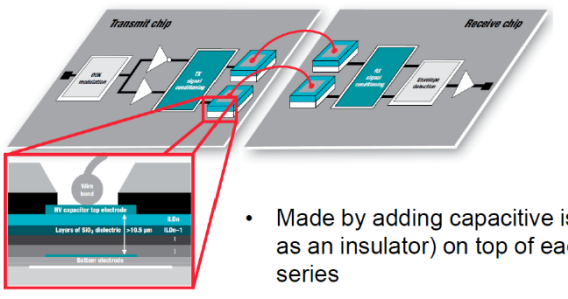
(Polyimide's dielectric breakdown strength >400V/µm)

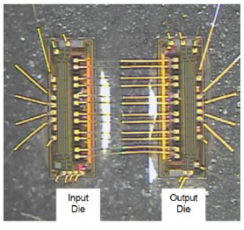


ANALOG DEVICES

Capacitive isolation

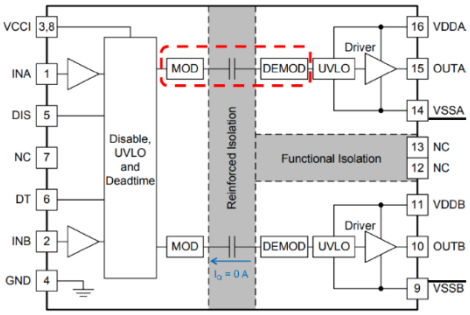


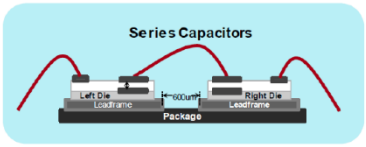


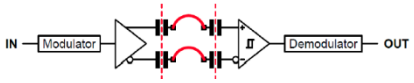


- Made by adding capacitive isolation barrier (SiO_2 as an insulator) on top of each die to connect in series
- 8 to10 times faster than optocouplers

Typical implementation is with two capacitors in series, one on the low voltage side of the circuit, and this capacitor is bonded to a second capacitor on the high voltage side of the circuit. Capacitor are standard capacitors made out of SiO2. The total isolation voltage can be from 12 to 8 kV peak.







- Reinforced isolation is realized by thick SiO_2 capacitors combined in series
  - Each channel uses high voltage isolation capacitors on both die
- Combined isolation capacitor thickness is > 21 µm
- 12.8 kV_{PK} surge voltage, 8 kV_{PK} transient over-voltage, 1.5 kV_{RMS} working voltage

Isolated gate driver comparison

Attribute		Opto-Coupler	On-chip Magnetic	On-chip Capacitive
Isolation Materials		Epoxy/Silicone gel	Polyimide	SiO_2 or equivalent
Signal Coupling		Optical (LED +diode)	Magnetic field	Electric field
Performance Across Temp & Time		Varies	Consistent	Consistent
Life Expectancy		~10 Yrs	~ 20 Yrs	~ 20 Years
Speed		Slow	Fast	Fast
Distance Through Insulation (DTI)		> 400 µm	~20 µm	~20 µm
Meets EN60950 >0.4mm DTI		Yes	No	No
Common Mode Transient Immunity (CMTI)		~25 kV/µs	> 100 kV/µs	> 100 kV/µs
EMI EMC	Susceptibility	Non-issue – too slow	Design techniques	Signal level dependent
	Radiation	Non-issue (light transmission)	Design techniques	Design techniques
Junction Temperature		Up to 125°C	Wide range (150 °C)	Wide range (150 °C)
Standards		UL1577 IEC60747-5-5	UL1577 VDE0884-11	UL1577 VDE0884-11
Modulation Method for Internal Signal Xfer		No modulation required	On-Off Keying	On-Off Keying
AEC Qualified Portfolio		Limited	Yes	Yes

PROTECTION CIRCUITS

Power mosfet can fail because of:

- Excessive Vds: we overcome the breakdown voltage.
- Excessive Id
- Excessive Tj (junction temperature)
- Excessive Vgs

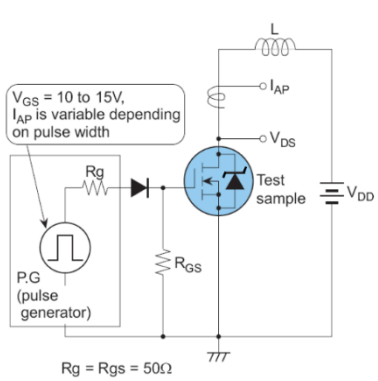
To avoid damages of power devices and destruction of them, we can implement some protection circuits, that come into play when there is an accident we cannot foreseen.

EXCESSIVE Vds – UNCLAMPED INDUCTIVE SWITCHING (UIS)

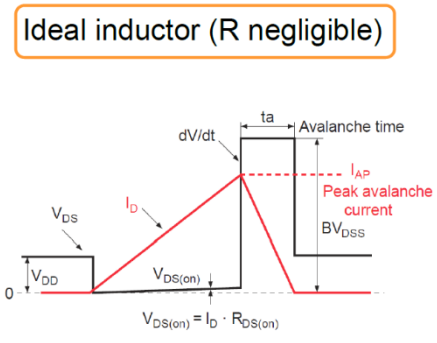
- The vast number of loads driven today are inductive in nature such as solenoids, transformers, inductors, etc.
- Power MOSFET failure due to Unclamped Inductive Switching conditions is one of the most prevalent failure modes encountered
- Recent advances in power MOS processing technology now enables power MOS transistors to dissipate energy while operating in the avalanche mode.
- The term *'Ruggedness'* when applied to a power MOS transistor, describes the ability of that device to dissipate energy while operating in the avalanche condition.

Usually the excessive Vds happens in power electronics because power electronics deals with inductive loads and if we are not operating it properly we might end up with an overvoltage between drain and source every time we turn off the mosfet.

The typical error is that we are not clamping the load with a free-wheeling diode, this is the case of the UIS.
However, there is a category of mosfet devices, the **Ruggedness mosfets** that are able to sustain a breakdown operation provided that some specifications are fulfilled.
A rugged mosfet can be distinguished because in the datasheet, instead of a simple diode, we have a Zener diode symbol between drain and source.



Test circuit



Voltage V_{DS} and current I_D waveforms

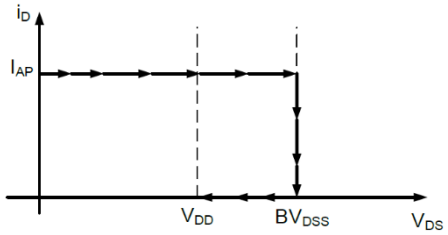
In the image above we have an ideal inductive load without a clamping diode. Initially the mosfet is off and there is no energy stored in the inductor. The drain current is 0 and the drain to source voltage is Vdd, because there is no current in the inductor.

At time t = 0, the pulse generator is producing a positive pulse, turning on the mosfet (we will neglect the turn on transient because its time constant is faster than the one related to the load).
The mosfet turns on, it enters the ohmic region and so the voltage between drain and source drops to 0. The current, instead, increases linearly because we are applying a voltage V_L = Vdd across the load and so di_L/dt = Vdd/L.
Actually, the voltage drop Vds is not actually zero, because as the current increases there is some voltage drop across the Rds(on) that produces a small drop.

At t = t1 the mosfet turns off again very quickly. The voltage between drain and source increases a lot until we reach the breakdown voltage BVdss. At this point V_L (always from left to right) is Vdd – BVdss. But BVdss must be larger than Vdd, otherwise the transistor would break down. So the voltage across the inductor changes its sign. Now, di/dt will be (Vdd – BVdss)/L, and it's negative. Hence the current goes down.

After a time t_avalanche, the current reaches 0, which means that there are no more carriers in the mosfet that are able to sustain the breakdown, which is over. So the Vds collapses back from BVdss to Vdd.
The current decreases linearly during the avalanche time because the voltage across the inductor is kept constant.

If we want to plot the turn off transient on the i_d vs Vds plot, the result is the one below. Initially we have Vds = 0 and a peak current Iap. When the mosfet turns off, the current stays constant until the Vds reaches BVdss. When the mosfet enters the breakdown the current is able to reduce. When the current reaches 0, the voltage collapses back to Vdd.



Turn-off switching trajectory during UIS

If we look at the datasheet of the power mosfet, there are, in general, three specifications related to the avalanche (the third one is the least important).

1. Maximum avalanche current rating: maximum current that can flow in the mosfet operating in the avalanche regime. It doesn't have to be calculated, because the maximum current in the avalanche regime is the one reached by the load before turning off the mosfet. It must be smaller than the rated avalanche current.
2. Avalanche energy rating, the maximum energy that can be dissipated inside the silicon chip when we turn off the load. The energy is calculated starting from the equation for the drain current in the avalanche regime. By putting the current equal to 0 we can then compute the avalanche time t_av. This value can be used to compute the energy as an upper limit for the integral.
   Also this energy must be smaller than the rated energy specified by the datasheet. The first term of the dissipated energy, ½ * L*I^2 is the energy stored in the inductor at the maximum level current in the inductor.

   The other factor is greater than 1, so the energy dissipated when turning off the load is bigger than the energy stored in the load. The smaller the difference the difference BVdss – Vdd, the larger the dissipated energy.

The energy dissipated by the mosfet can be larger than the energy stored in the inductor because when the mosfet is operated in the avalanche regime there is current flowing through it, and this current is also flowing through the voltage generator Vdd, so the voltage generator is providing an energy which is the product Vdd*Id. Only in a situation where the transient is very fast, which happens when BVdss >> Vdd the mosfet is dissipating the energy stored in the inductor. In all the other cases the dissipated energy is larger.

- There are three specifications on the data sheet relating to avalanche:
  1. the maximum avalanche current rating, which gives the maximum allowable current the which device can be subjected to in the avalanche mode;
  2. the non repetitive avalanche energy rating, which gives the maximum avalanche energy the device can handle in a single pulse with a starting junction temperature of 25°C.
  3. the repetitive avalanche energy rating, which gives the maximum allowable energy per pulse which the device can be subjected to on a repetitive basis;

From:

$$i_d(t) = I_{AP} - \frac{(BV_{DSS} - V_{DD})}{L} \cdot t \quad ; \quad E_{diss} = \int_0^{t_a} i_d(t) \cdot BV_{DSS} \cdot dt$$

it follows:

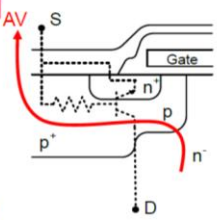
$$t_{av} = \frac{L \cdot I_{AP}}{(BV_{DSS} - V_{DD})} \quad ; \quad E_{diss} = \frac{1}{2} L I_{AP}^2 \cdot \frac{BV_{DSS}}{BV_{DSS} - V_{DD}}$$

To summarize the most important points about UIS, this can be handled only with a ruggedness mosfet, provided that we are respecting the specifications in the datasheet, e.g. the ones in the table below.
The physical reason for having a maximum current is the turn on of the parasitic bipolar transistor; the maximum current that can flow during the avalanche regime must be smaller than the current needed to turn on the parasitic bipolar transistor. If it turns on, BVdss collapses back.

I_{AR}	Avalanche Current, Repetitive or Not-Repetitive (pulse width limited by T_j max, $\delta < 1\%$)	50	A
E_{AS}	Single Pulse Avalanche Energy (starting T_j = 25 °C, $I_D = I_{AR}$, V_{DD} = 25 V)	400	mJ
E_{AR}	Repetitive Avalanche Energy (pulse width limited by T_j max, $\delta < 1\%$)	100	mJ

Avalanche current/ avalanche energy specification

- **Avalanche operation maximum current (I_{AR})**
  Maximum current that can flow through the device during avalanche operation (limited by the triggering of parasitic BJT). This limit must be considered as an absolute maximum rating.

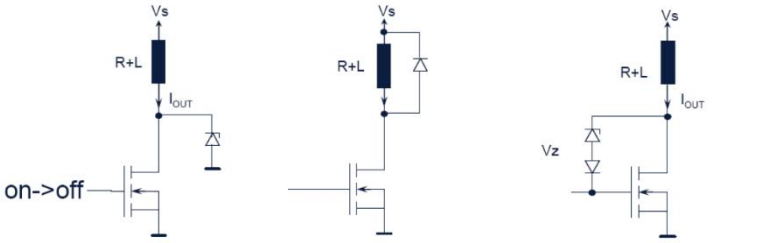


- **Energy during avalanche for single pulse (E_{AS})**
  Maximum energy that can be dissipated by the device during a single pulse avalanche operation (limited by T_{jmax} with starting T_J = 25°C). The E_{AS} value decreases as the starting junction temperature increases.

Conversely, there is also a maximum energy that can be dissipated inside the silicon because this energy is the energy needed to bring the temperature junction from the ambient temperature up to the maximum junction temperature.

OVERVOLTAGE PROTECTION CIRCUITS

If we don't have a ruggedness mosfet and we want to drive an inductive load we need protection circuits. The standard protection is the free-wheeling diode; otherwise we can use a Zener diode between ground and intermediate node, but it is not usually done, whereas the third alternative is the Zener clamp. It is a Zener diode connected between the drain and the gate with an additional standard diode in series with it, and this is done on the same silicon chip. It comes directly with the mosfet in the chip.



External (power) Z-diode from drain to GND

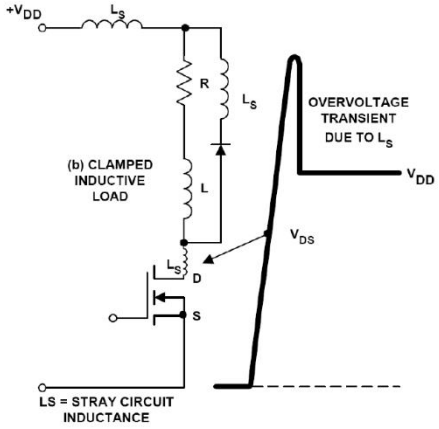
External or internal free-wheeling diode

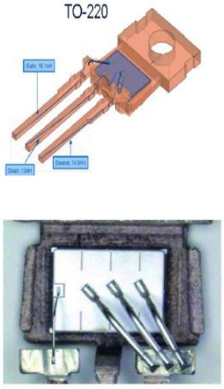
Switch off delay could be a disadvantage

Internal z-diode from drain to gate switches on the DMOS in case of overvoltage

High power dissipation at the DMOS

FREE-WHEELING DIODE PROTECTION





- Even if the main inductive component of the load is "clamped", stray circuit inductance still exists, however, and overvoltage transients will still be produced as a result

This type of protection might not be good. Protecting a mosfet with a free-wheeling diode means that we have to connect a free-wheeling diode in parallel to the load, and this connection is made outside the package of the mosfet. So **there is a small inductance that comes from the parasitic inductance of the bonding wires and pins of the package that is not protected**, because it is internal to the package (Ls).

So even if we use a free-wheeling diode there is a small but non-negligible inductance, and the mosfet may undergo a breakdown regime for a small amount of time due to the discharge of this unprotected parasitic inductance (tens of nH). This is why sometimes a ruggedness mosfet is used even if a free-wheeling diode is put in parallel to the load.

ZENER-CLAMP PROTECTION

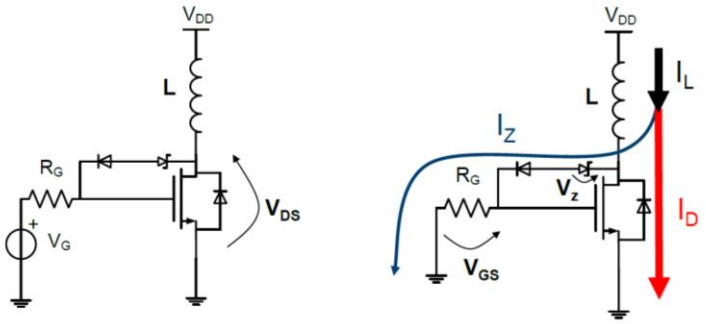
It consists of a Zener diode and a standard diode connected between the gate and the drain of a mosfet transistor.

Let's assume that the gate voltage applied by the driver (represented with a Thevenin equivalent) is high, and the mosfet is turned on, making some current flowing through the load. At t0 the mosfet is then turned off, the gate voltage is brought back to 0.

Assume that the turn on transient of the mosfet is quick with respect to the inductive load time constant, so for the sake of simplicity we can say that the mosfet turns off almost instantaneously. The current cannot be changed instantaneously, so initially we have a situation where the mosfet is turned off but there is still a current that flows in the inductor. Where does the current go?

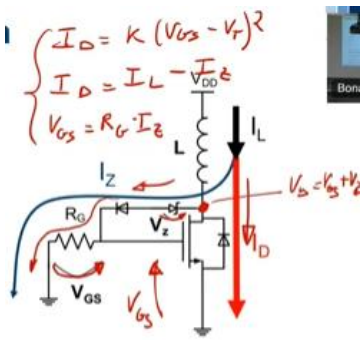
In a simplified analysis, we can assume that the circuit is operated as a sequence of quasi-static situations, even if everything happens simultaneously.

Hence initially the mosfet is off and so the current can only go through the Zener diode, making a Zener voltage Vz to develop across it, and across the gate resistance Rg. If we have a current flow in Rg, a voltage Vgs develops, which makes the transistor to turn on again.
The mosfet turns on again so that most of the current flows through the mosfet and the mosfet is operated in saturation regime.



- At turn-off, V_{DS} increases until it reaches the clamp level. The zener current I_Z flows through R_G turning on the power MOSFET.
- The condition: $V_{DD} < V_Z < BV_{DSS}$ must be satisfied.
- The energy is mostly dissipated by the power MOSFET working in saturation at $V_{DS} = (V_Z + V_{GS}) < BV_{DSS}$.

If there is current flowing through the Zener diode, the voltage that develops between gate and source, Vgs, is the product between the Zener current and Rg. The drain voltage Vd is Vgs + Vz. Vz is such to make the mosfet to work in the saturation regime.
The current Id must satisfy the following equation.



The new operation point satisfies the three equations above, for the mosfet.
The Zener voltage must be larger than Vdd, otherwise the Zener would always be on, but at the same time smaller than BVdss.

So the Zener-clamp protection avoids the mosfet to enter in the breakdown region, but the mosfet turns on and there is a current flowing through the mosfet, and this current discharges the inductance. Of course, to discharge the inductance, the Zener voltage must be larger than Vdd because we need to reverse the voltage across the inductance.

In conclusion, the energy stored by the inductor and the additional energy that comes from the voltage generator is mostly dissipated by the power mosfet working in saturation.

The standard diode is needed in series with the Zener diode because when the mosfet is turned on in the ohmic region normally, Vds drops to zero and we don't want a current in the forward bias diode.

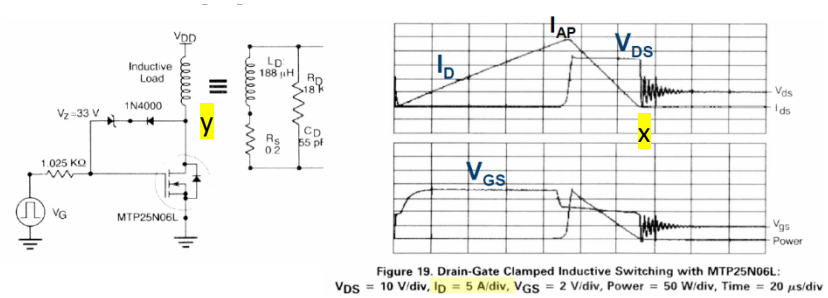
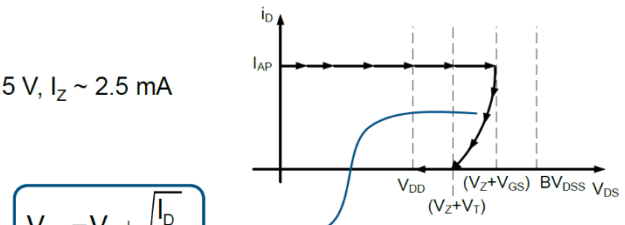


Figure 19. Drain-Gate Clamped Inductive Switching with MTP25N06L: V_{DS} = 10 V/div, I_D = 5 A/div, V_{GS} = 2 V/div, Power = 50 W/div, Time = 20 µs/div

- At turn-off:
  - V_{GS} ~ 2.5 V, I_Z ~ 2.5 mA
  - I_D ~ I_L



$$V_{GS} = V_T + \sqrt{\frac{I_D}{K}}$$

When the mosfet is on we have a linear increase of the current until the value Iap, then the mosfet is turned off. Vds goes almost to 35V, and Vz is 33V so we actually expect Vds above it.
If we look at the Vgs, once the mosfet is turned off, in reality is not really turned off, firstly it is operated in the ohmic region and then immediately enters the saturation region. When I turn off the input generator, Vgs drops down to 2.5V (from 5V) and reduces smoothly. This small reduction is due to the fact that the current is reducing because the voltage we are applying to the inductor is negative, so current decreases, the mosfet is in saturation regime and the Vgs must decrease correspondingly.

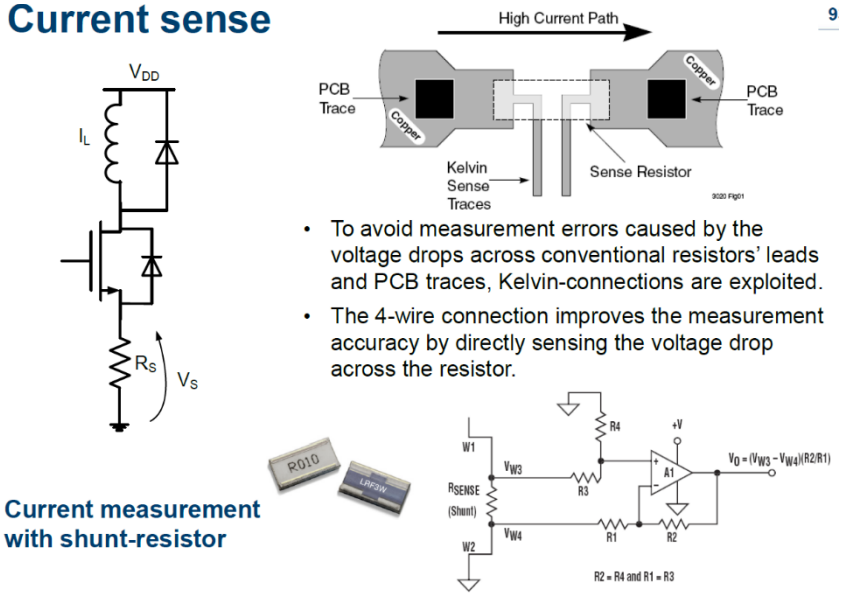
At point x the current reaches 0 and so no more current flows in the mosfet or in the Zener diode. If so, the node y collapses down to Vdd, because no more carriers can sustain the breakdown in the Zener diode.

The ringing is present because there are parasitic capacitances connected between node y and ground (Cds) and immediately before the collapse, the voltage across it is Vgs + Vz, so almost Vz. However, when the Zener stops and the voltage collapses, we have a capacitor charged at Vz. When the current reaches zero, the Zener is turning off, the mosfet is turning off but the voltage at node y is still Vz. The final voltage that the drain asks to reach is Vdd. So we have energy in excess in Cds and therefore this excess energy starts bouncing back and forth between the parasitic capacitance and the inductance until we reach the steady state. We reach the steady state because we have parasitic resistances in series which damp the oscillations.

What about the current in the Zener when on? It is almost 2.5mA according to KCL and KVL, and this value is negligible with respect to the current of several amps flowing in the mosfet.

Moreover, the turn off trajectory initially sees a Vds = 0 and a current Iap. When the gate driver goes to 0 the Vds increases up to Vz + Vgs, but still less than BVdss. At this point we are applying a reverse bias to the inductor, making the current to decrease. The current is decreasing and at the same time the voltage Vds slightly decreases. This decrease is because Vds = Vz + Vgs, and Vz is constant but Vgs reduces as we reduce the current.
When there is no more current the voltage collapses down to Vdd. It is a turn off trajectory different from the one of the unclamped inductive switching.

OVERCURRENT PROTECTION CIRCUITS



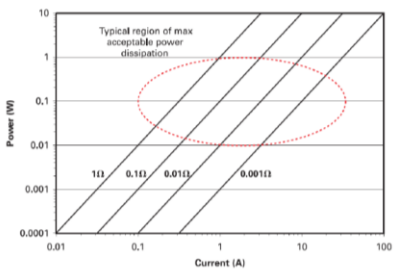
Circuits that switches off the mosfet when the current overcomes a certain level.
The simpler circuit uses a shunt resistance connected between source and ground; it is a very small resistance, few mOhm.

We have to measure the voltage drop across the resistance Rs and this voltage drop is proportional ot the current that flows in the mosfet. The proportionality factor is the sense resistance.

The Rs is typically a metal resistance and SMD. In order to avoid current measurements due to the voltage drop across the resistor leads and copper traces, we typically we need to use a 4 wire Kelvin configuration, that is the one in the bottom of the image.

Issue of Rs

Basically it is power dissipation, because the current that flows in the mosfet also flows in the shunt resistor, producing a power dissipation that is Rs*Id^2, assuming the mosfet is operated in DC.
I could reduce the power by reducing Rs, but if so we also reduce the voltage drop across it and this is a problem because it has to be read by a differential amplifier, which has its own offset voltage and noise, and when the voltage across Rs becomes comparable with the offset we cannot measure anything.



- As R_s increases, power dissipation increases.
- As R_s decreases, circuit offset voltages become a larger fraction of the measured sense voltage and therefore accuracy is reduced.
- A balance between required accuracy and power dissipated for a given current must be chosen.

- Ex.: 20A current must be sensed with maximum 1 W of dissipated power and with a best 2% accuracy.

$$P = I^2 \cdot R_S$$

$$R_S = \frac{P}{I^2} = \frac{1}{400}\Omega$$

$$R_S = 2.5\text{m}\Omega$$

$$V_{SENSE} = I \cdot R_S = 50\text{mV}$$

Therefore, for a best accuracy of 2% the input offset voltage can not be more than 2% of 50 mV, that is, 1 mV.

So the sensitivity of our measurement will depend on how much the voltage drop on the resistor will be with respect to the offset of the opamp used in the current sensing circuit.

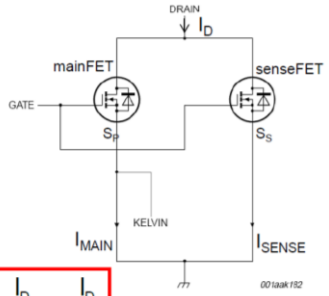
We can say that we have a **trade-off between power dissipation and sensitivity**.

The plot in the image is the power dissipation as a function of the current assuming different values of the shunt resistances.
For instance, let's assume that the mosfet is handling 20A and this current must be sensed with a maximum of 1W dissipation and a 2% accuracy. The calculations are the one in the previous image. With a Rs of 2.5 mOhm, the Vsense on Rs is 50mV, which means that the input offset voltage of the current sense amplifier must be 1mV, and it is a very low value.

So with a shunt resistor we have this trade-off that cannot be overcame. Another possibility is to use a current sensing mosfet.

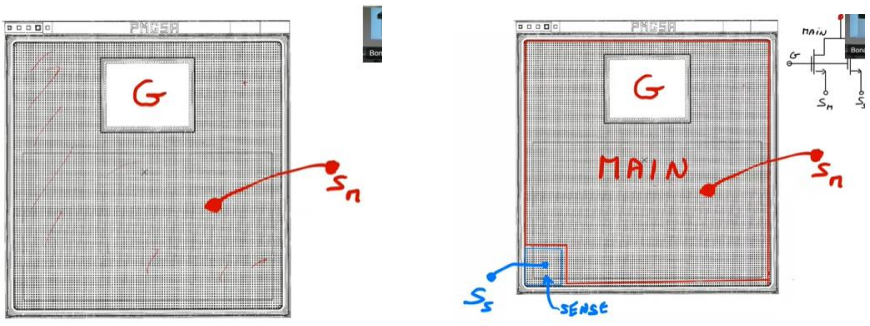
CURRENT SENSING MOSFET - SenseFET

- It is possible to isolate the SOURCE connections of several cells from those of the majority and bring them out onto a separate SENSE pin.
- The PowerMOS can now be thought of as two transistors in parallel with a common GATE and DRAIN but separate SOURCE pins. The ratio of cells from the main to sense FET determines the ratio of R_{DSon} of the two devices.
- When the devices are turned on, the load current will be shared as a ratio of their on-state resistances.
- The ratio of current through the mainFET to the current through the senseFET is known as the sense ratio (r).
- The sense ratio is typically 500 – 1000:1



$$I_{MAIN} = r \cdot I_{SENSE};\ I_{MAIN} + I_{SENSE} = I_D \Rightarrow I_{SENSE} = \frac{I_D}{(r+1)} \approx \frac{I_D}{r}$$

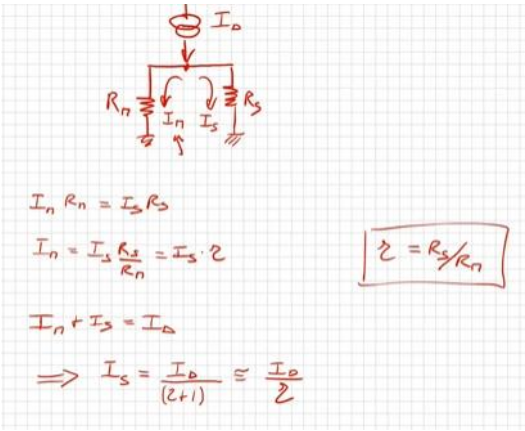
The idea is that we take we have a vertical DMOS with the drain connector on the back side of the Silicon, while on the top side we have metal that covers everything. The black dots are the cells of a vertical DMOS. G is the gate, the other metal that covers is the source.



Let's imagine now to take an angle of the DMOS and cut the metal in that portion, along the blue line. If we do so, we can isolate two different source connections, one including the cells in the red square, the other including the cells in the blue square. The sense source includes just a limited amount of cells, while the main source includes almost all the cells. Typically in the sense region we have 1/1000 of the cells in the main source.

However, the drain of the two mosfet is still the same, and also the gate.

Let's assume that we ground the two sources and we inject a current in the drain. How is it divided in the two paths?



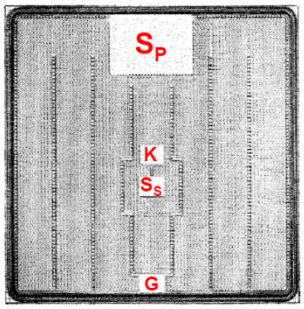
Im is the current through the main mosfet. Main resistance and source resistances are present because when on each mosfet is operated in the ohmic region. The ohmic resistance Rds(on) = Rcell/n, so the Rm of the main transistor is smaller than the one of the sense transistor → Rm << Rs.

r is also called current sense ratio. In general it spans from 200 to few thousands.

The conclusion is that when the mosfet is operated in the ohmic region, the current that flows in the sense mosfet is the total current Id divided by r.

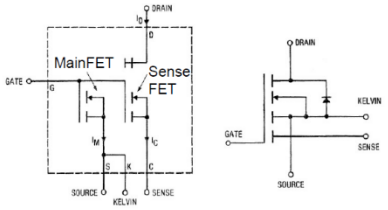
Details

The one in the left image is a real current sensing mosfet. It is indeed a five terminals device. The fifth contact is the **Kelvin contact**, which provides a path for the return of the sense current which doesn't include the parasitic resistance coming from the metal track and other sources. It is fundamental to get an accurate measurement of the current.

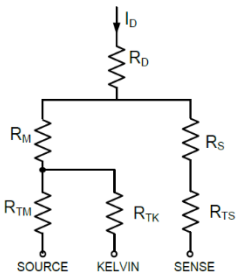


Current-sensing MOSFET die

- **The Kelvin connection provides a path for returning the sensing current to the source without including the source wire bond, the source pin, and much of the source metallization.**



schematic representation **symbol**

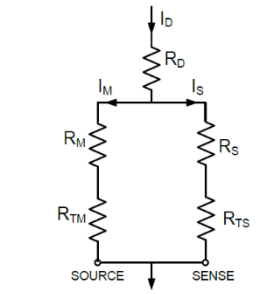


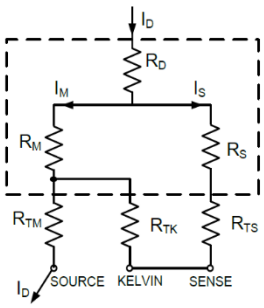
IRC530
R_D: 0.06 Ω
R_M: 0.08 Ω
R_S: 133 Ω

R_{TS}: 0.02 Ω

Rtm, Rtk and Rts are parasitic resistances in series with the main, kelvin and sense source contacts. These parasitic resistances are in the same order of magnitude of Rm, and Rd, and they produce an error in the sensing of the current.

Let's put the source and sense contacts to ground and let's have some current Id flowing. We can compute the current divider in the red box.





$$r=\frac{I_M}{I_S}=\frac{R_S+R_{TS}}{R_M+R_{TM}} \approx \frac{R_S}{R_M+R_{TM}}=\frac{R_S}{R_M}\cdot\frac{1}{1+R_{TM}/R_M} = 0.8n \text{ (for the IRC530)}$$

sense ratio ≠ cell ratio

$$n=\frac{n_M}{n_S}=\frac{R_S}{R_M}$$

cell ratio

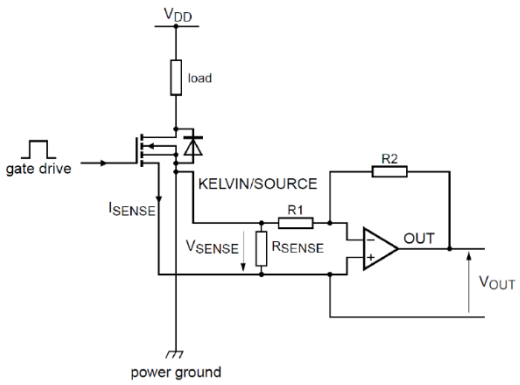
$$r=\frac{I_M}{I_S}=\frac{R_S+R_{TS}+R_{TK}}{R_M} \approx \frac{R_S}{R_M}=n$$

sense ratio = cell ratio

We can get rid of Rts because small with respect to Rs, but not Rtm at the denominator.
We notice that Rs/Rm is the cell ratio, between the number of cells in the main region and the sense region.
The cell ratio is known a priori, and it is not equal to the current ratio in this case, because of the parasitic resistances.

If instead we connect the sense and kelvin terminals, Rts and Rtk are negligible with respect to Rs, and now the sense ratio is actually the cell ration → the sense current that we measure is related to the drain current via the cell ratio.

How to measure the sense current



With $n = n_M/n_S = 1000$, $I_D = 20$ A,

$R_{SENSE} = 2.5\ \Omega$

⇩

$I_{SENSE} = 20$ mA

$V_{SENSE} = 50$ mV

$P_{SENSE} = 1$ mW

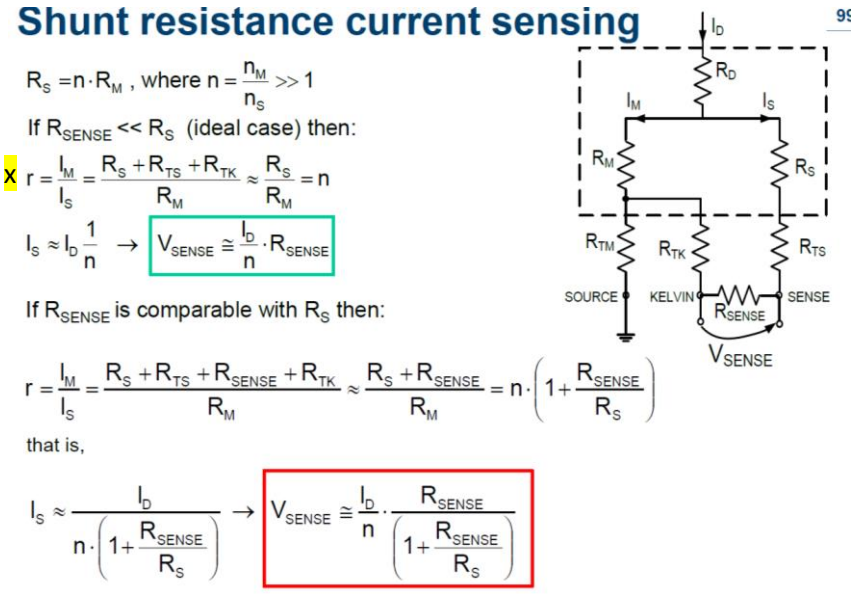
- The inclusion of R_{SENSE} increases the resistance of the sense leg

We insert a resistance between the sense and the kelvin pin, the Rsense. It is an external resistance and we have a voltage drop across it, Vsense.
The main source terminal is connected to ground, the sense source terminal is injecting current in the Rsense and the sense current is returned through the Kelvin pin. The voltage Vsense is then measured with an inverting amplifier.

Isense = Id/n.

In this situation I can increase the sense resistance because we can afford higher power dissipation, e.g. 10mW. The current is the same, so the resistance is multiplied by a factor 10 → 25 mOhm. If so, also Vsense becomes 500mV, so better than the offset of the amplifier.
Hence the current sensing mosfet allows us to break the trade off between accuracy and power dissipation.

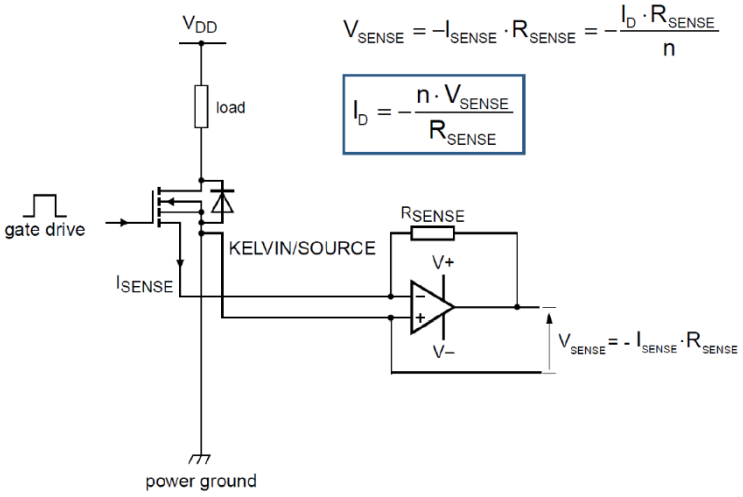
However, there is a drawback with this sense external resistor. The inclusion of sense resistance in the sense path increases the resistance of the sense path. We need to check if it is a problem. Let's include the external Rsense and check if the ratio between sense current and drain current is still equal to the cell ratio.



We have to recalculate the sense ratio. If Rsense is negligible with respect to Rs, we can drop the term in the numerator of x. If so, we have the formula in the green rectangle: the voltage we are measuring is strictly proportional to Id.

If instead Rs and Rsense are comparable, the Vsense is not strictly proportional to Id, because we have the term 1 + Rsense/Rs. Conceptually there is still a proportionality, but practically no, because Rs strongly depends on temperature, so it is an unpredictable factor.

So when we want to use a Rsense comparable with Rs, the solution is to use a virtual earth. Instead of putting the external Rsense in series with the sense leg path, we put the sense resistor in the feedback path of an inverting amplifier.



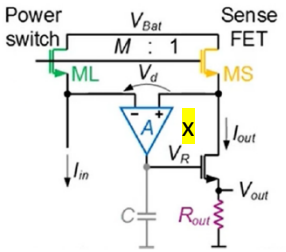
Because of epsilon = 0 between the terminals of the opamp, the sense and Kelvin terminals are at the same potential. This means that we are like shorting the Kelvin and sense terminals, even if we are actually not shorting them.
In this way the drain current is related in a predictable way to Rsense.

High-side current sensing

If the mosfet of which we want to measure the current is not a low side mosfet but it is a high side mosfet, in principle we could place a shunt resistance between the drain and the bias supply, but this is not practical.
It is easier to use a SenseFET; however, the mosfet in this case is not in the low side position, but in the high side one → not easy to extract the sense current.

The green and yellow are the mosfet, split in main and sense mosfet. We are using a feedback circuit that keeps the main source and the sense source at the same potential (x).

- Sense FET + sense amplifier



$$y \quad I_{IN} \cdot R_{POWER} \cong I_{OUT} \cdot R_{SENSE}$$

$$\frac{I_{OUT}}{I_{IN}} \cong \frac{R_{POWER}}{R_{SENSE}} = M$$

+ Lossless
+ Good accuracy
- Complexity

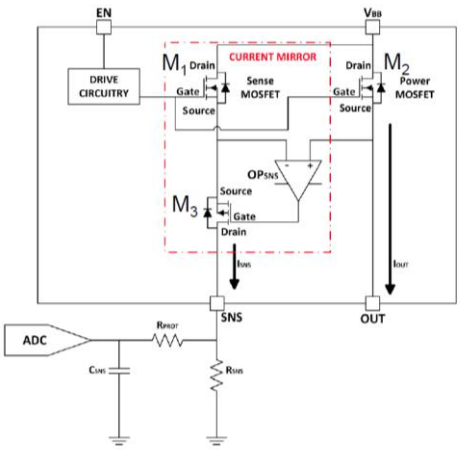
Having the same potential is important to have a sense current that is related in a predictable way to the sense current.

The sense current flows in the feedback nMOS transistor and goes in an external output resistance Rout. In the image, Rpower is Rmain, Rsense is Rs. Equation y holds because Vd = 0.

The issue with this implementation is that the current sensing circuit is pretty slow, so it is a good solution if the switching frequency is not that high. This can be understood computing the loop gain of the circuit and assuming the amplifier is a single pole amplifier.

There are several commercial product for high current sensing integrated inside a single silicon chip, integrating also the high side driver. The topology is the one seen so far.

The power mosfet is internal and integrated in the silicon chip?
Yes, because of course the sense and main mosfet are a part of the same mosfet.



- SENSE and POWER terminals are held at the same potential by M_3 and OP_{SNS}.
- The external resistor R_{SNS} transforms the current I_{SNS} into an analog voltage that is proportional to the load current.
- Common mode voltage is swinging over the whole range from ground to battery.
- Demanding combination of very high common-mode rejection and high-voltage handling capability, plus high gain, high accuracy, and low offset.

NB: the opamp has its input terminals connected to the source of the high side mosfet, which is going up and down between ground and Vdd. So the operational amplifier has to have a very large CMRR and it has to be able to sustain a large input common mode voltage, which reaches even the PS.

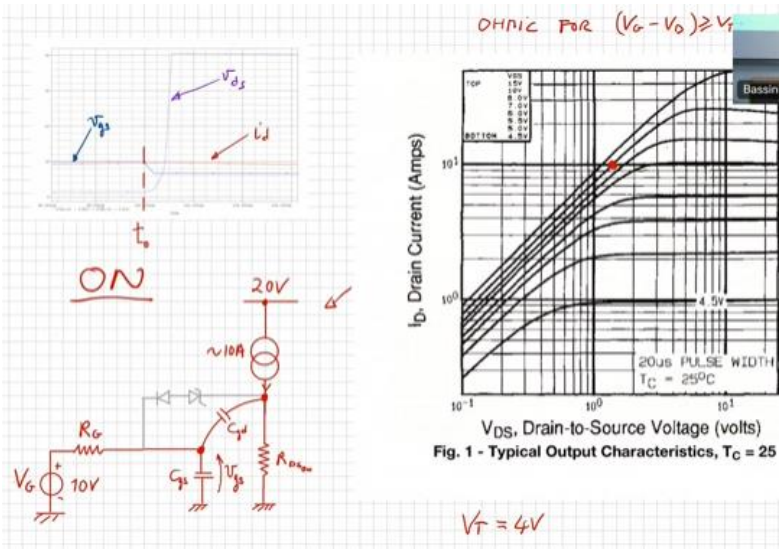
The advantage of this circuit is that the sense current is pushed on the Rsns resistor that the designer has to select properly and that is referred to ground. So we can easily measure the Vsense across this resistor referred to ground.

The difference with the configuration of two images before is that instead of a nMOS we have a pMOS.
What is the difference in using a pMOS instead of a nMOS in the loop?

Recap – Zener clamp protection explanation

During the turn off transient the mosfet doesn't turn off actually, but it moves from an ohmic region to a saturation region.
Let's better analyze the situation before the immediate 'turn off' of the mosfet. The current that flows in the inductor is almost constant, so we can assume the inductor is behaving like a constant current generator.

In the image below the mosfet is on and the current is 10A. The Vgs voltage is set by the voltage generator and it is 10V. Looking at the datasheet, a 10A current with a 10V Vgs leads to a small Vds (red dot). This is the starting point, the Zener is not active yet.

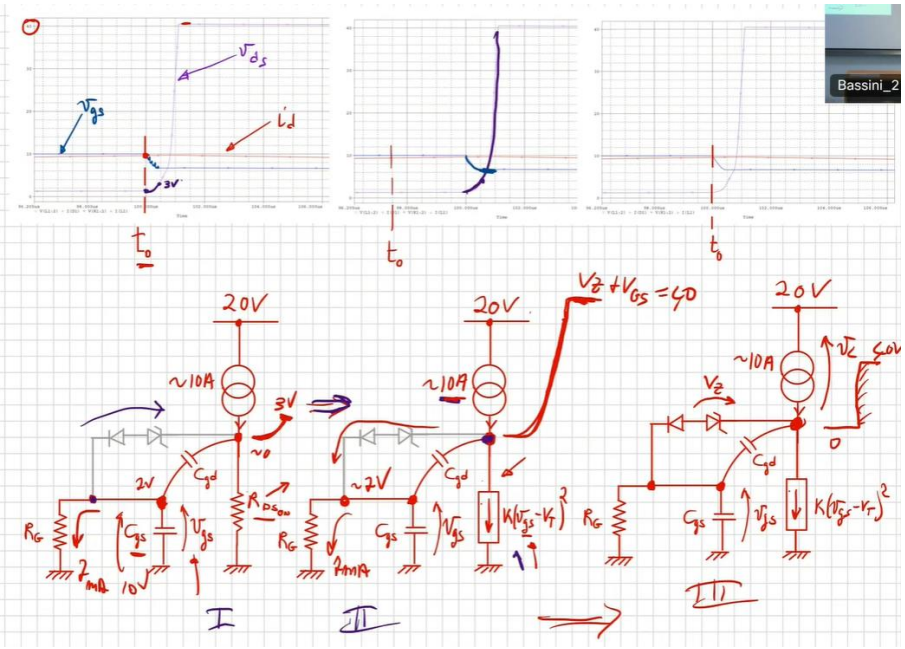


Immediately after the voltage applied to the gate is reduced down to 0, there is no more gate voltage Vgs, but the Cgs is still charged, so we still have 10V, producing a current of 10mA in the Rg. This 10mA current is discharging the Cgs, so the Vgs is reducing. We are moving horizontally in the Id vs Vds characteristic because the current is forced by the 10A current generator. So the Vds is increasing because if we reduce the Vgs, the Rds(on) is increasing. The Vds ramps up to 3V.
At this point, Vds = 3V, so since Vs = 0, Vd = 3V. The Vgs corresponding to this is 7V, so we are at the boundary between ohmic and saturation region.

At this point, the mosfet is entering the saturation region (Zener is still not active) and it acts as a voltage controlled current generator whose current is equal to the load current. The gate voltage is at 7V.
But the current of 7mA in the Rg is still trying to discharge Cgs. It is sufficient to move the Cgs by a small amount to change substantially this current.
If we try to reduce the mosfet current below the load current there is an unbalance at the drain node and the current from the load that is not entering the drain is charging Cgd producing an increase of Vds.

When the voltage Vds reaches Vz + Vgs = 40V (Vgs = 7mA*1k = 7V), the Zener starts to conduct current. Vds = 40V now and the drain node is clamped by the Zener diode, which is entering the breakdown region.

As for the Vgs, it is still around 7V.

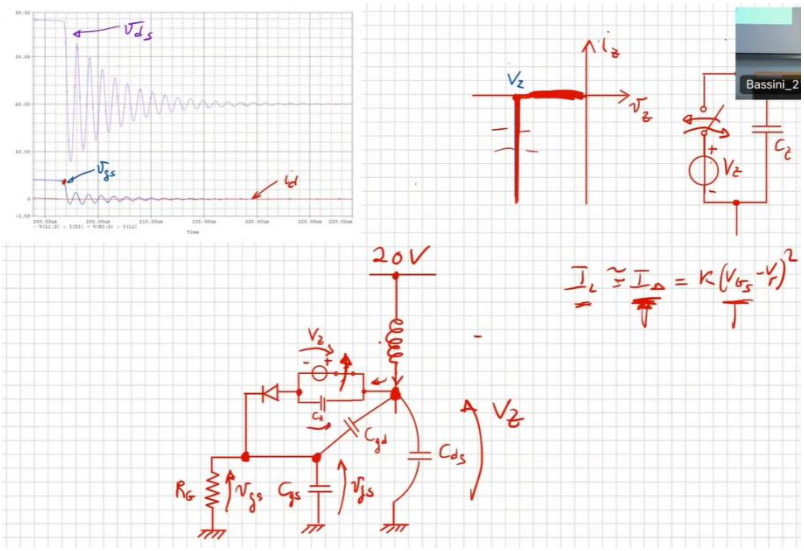


At this point, the current into the load is reducing. This because if we look at the voltage drop across the inductor, V_L = 20V – (Vgs + Vz) = -20V, so di_L/dt = V_L/L < 0 and the current starts decrease linearly.

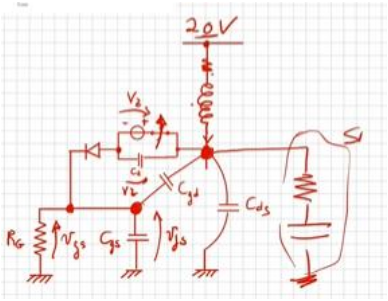
So **it is fundamental, for the current to reduce, that Vgs + Vz is larger than the bias voltage**.
Hence the current decreases, so also the Vgs decreases (Id = k(Vgs – Vt)^2). Furthermore, the current that flows into the Zener is determined by the amount of current flowing in the Rg resistance, which is 7mA.

When the load current is reduced and the Vgs goes below the threshold voltage, there is no more current generator in the mosfet, but still few mA flow in the Zener diode. The Zener diode can be modelled with an equivalent circuit made of a voltage generator, a switch and a capacitor. The switch is closed if there is current flowing through the Zener. If there is no more current in it, the switch is open.

When we are approaching the zero, the switch is opening and we are left with the Cds charged at Vz.
Vz != from the steady state value of the drain, which is 20V. Hence we have a ringing until we reach 20V starting from 40V.

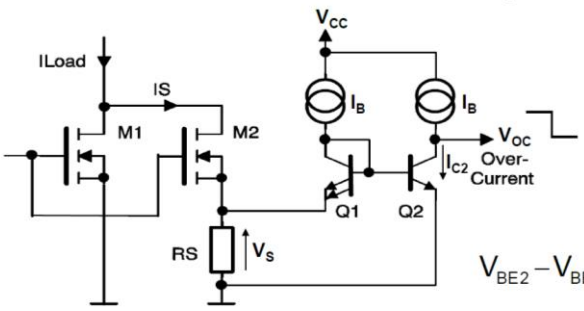


In order to damp efficiently these oscillations it is much better not to have a series parasitic resistance to the inductor, but a parallel one.
If we want to artificially damp the oscillations, the common way is to add a snubber network, which includes a resistor and a capacitor in series connected to the drain.



To design the R of the snubber we need to know the parasitic inductances and capacitances involved in the oscillations, and this can be done with an experimental measurement.

Example



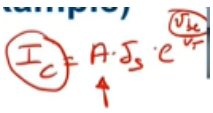
- Simple solution uses a shunt resistor implemented in the metal layer for the detection of overcurrent.
- Assuming the same collector currents for Q1 and Q2, the current limit threshold is well defined by the emitter area ratio of Q1 to Q2.
- Below threshold $I_{C2} < I_B$, then $V_{OC} \sim V_{CC}$
- Above threshold $I_{C2} > I_B$, then $V_{OC} = V_{CEsat2} \sim 0$

In some cases we want both the sense circuit and the current protection one to be integrated in the same chip, like in the image above.

We have two npn transistors and two current generators, M1 is the main mosfet, M2 is the sense mosfet and Rs is the internal sense resistor (implemented on the same silicon chip).
Q1 and Q2 are like a current mirror with an emitter degeneration due to Rs. Q1 has two arrows, meaning that the area of Q1 is bigger than Q2.

The two npn have the transistor current forced by the two Ib generators, and the Ib are identical. But is the Ib disturbing the Is in the Rs? No, because the Ib current in Rs is much smaller than Is, so Vs = Rs*Is.

Stated all this, let's analyze the operations of the circuit.
Let's assume the current flowing in the mosfet M2 is very small → Vs is almost 0, so we don't have to trigger the protection circuit. In this situation, Vbe1 and Vbe2 are not the same even if Ib is the same, because the area of Q1 is larger than the area of Q2. The larger the area, the smaller the Vbe at a given collector current.



So initially I expect that the Vbe1 < Vbe2. As a matter of fact, Vbe1 and Vbe2 are not independent, there is the relationship x linking them.
If Vs = 0, from x we can however assume that Vbe1 = Vbe2. This means that the current in the collector of Q2 is smaller than Ib, because Ib is collector current of Q1 and the area of the emitter of Q1 is bigger than the area of the emitter of Q2. This is the initial condition when the mosfet current is very low.

As for the output of the over-current protection circuit, at this point the output is Vcc, high, because we are sinking a current from the transistor Q2 that is smaller than the current Ib we are pushing.

Now, if I increase the current flowing in the mosfet (I_L), Vs increases but Vbe1 is not changing because it is a transdiode and it is set just by Ib. Since Vbe1 is not changing and Vs is increasing, Vbe2 = Vbe1 + Vs is increasing.
If Vbe2 is increasing, Ic2 is increasing. The critical point is when Ic2 = Ib. If we further increase Vs, Ic2 > Ib and the transistor 1 enters the saturation regime and the output voltage drops to 0.
Now we want to tune the parameters such that this situation happens when we want to limit the current.

Threshold sizing

$$V_{BE2} = V_{BE1} + I_S R_S$$

by replacing v_{be1} and v_{be2} with

$$V_{BE1} = V_T \ln \frac{I_B}{A_1 \cdot J_S}$$

$$V_{BE2} = V_T \ln \frac{I_{C2}}{A_2 \cdot J_S}$$

note : $V_T = \frac{k \cdot T}{q}$

we get

$$I_S = \frac{V_T}{R_S} \ln\left(\frac{I_{C2} \cdot A_1}{I_B \cdot A_2}\right)$$

Q2 transition threshold for $I_{C2} = I_B$, giving:

$$I_S\big|_{threshold} = \frac{V_T}{R_S} \ln\left(\frac{A_1}{A_2}\right)$$

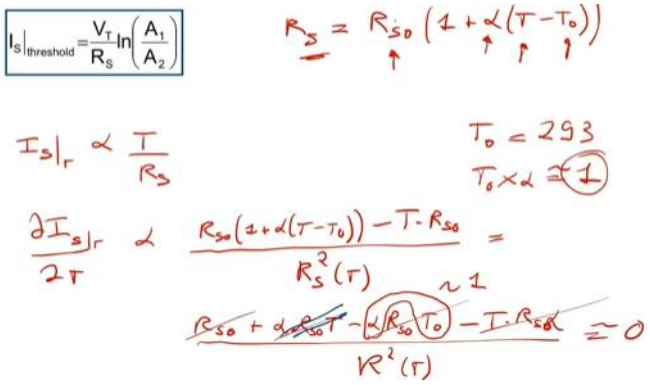
- Note: $V_T = kT/q$ depends on the absolute temperature. This temperature coefficient is first order compensated if the resistor R_S is made from aluminium interconnect metal.

With Vt we are indicating the thermal voltage, not the threshold voltage.
The reason to use to BJT is to eliminate the saturation current density term Js, and this is an advantage because Js strongly depends on temperature and we don't want the clipping point of our protection circuit to be dependent on temperature.

Then, the transition from the off region of Q2 to the saturation one when Ic2 = Ib.
In the end the threshold sense current is given by the blue box. We have a ratio between areas, which is a geometrical parameter. Since there is Vt, **we still are getting a dependance on temperature linearly**.

However, there is a trick we can use to avoid the dependance of the threshold current on the temperature, and it is to implement the Rs using a track of Aluminum. So it is a metal resistor, and Aluminum is chosen because the temperature coefficient of Al is 3.8*10^-3 per degree. So we can do the following reasoning.

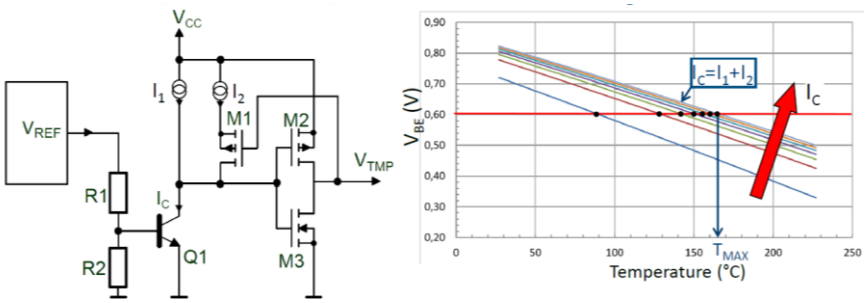
Rso is the resistance at room temperature, alpha is the temperature coefficient, and Rs is the resistance of Al.



So the Is|t will be proportional to T/Rs. We can compute the partial derivative.

We notice that the final result is almost equal to 0 at room temperature, at higher temperature it is more difficult.

OVERTEMPERATURE PROTECTION



- For $T < T_{MAX}$, $I_C < (I_1+I_2)$: then V_{TMP} ~ 0 and M1 is ON.
- For $T > T_{MAX}$, $I_C > (I_1+I_2)$: then V_{TMP} ~ V_{CC} and M1 is OFF

- Thermal protection of power devices requires a temperature sensor.
- The characteristic of the base emitter voltage is compared with a constant reference voltage, which is generated e.g. from a bandgap reference.
- The required temperature threshold (T_{MAX}) depends on the position of the temperature sensor in relation to the hot spot of the power device.

A possible way to implement an overtemperature protection is by using hysteresis.
If we want to implement an overtemperature protection we need to implement a temperature sensors first. Q1 is a temperature sensor, using a transistor. Of course a single transistor is not sufficient to implement a temperature sensor, we need to bias the base of the transistor with a fixed voltage which is independent on the temperature and on the bias supply we are applying to the IC. This is done by using a band-gap voltage reference.
Any IC we can imagine has a band-gap voltage reference inside if we need to generate a stable voltage.

So Q1 is biased with a fixed Vbe. The plot in the upper right depicts the Vbe dependance of a generic npn transistor as a function of the temperature. What changes between a curve and the other is the collector current Ic.

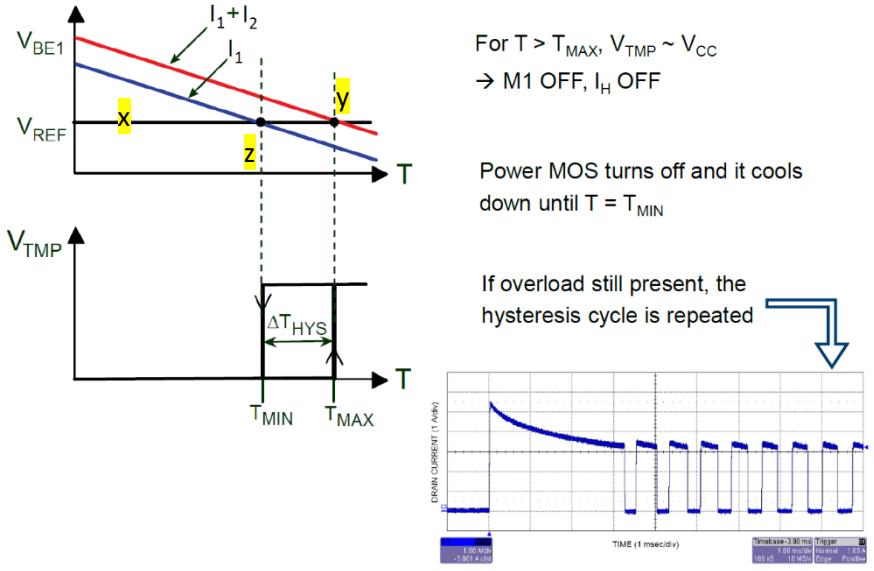
We are biasing Q1 with a fixed 0.6V. If we increase the temperature and we keep Vbe stable, the Ic increases, so we can use Q1 as a temperature sensor simply by looking at the Ic generated by it, the larger the temperature, the larger the Ic.

If the T of the IC is smaller than a certain Tmax (fixed by the designer), the Ic of Q1 is smaller than the sum of I1 and I2, viceversa it T > Tmax, Ic > I1 + I2.

Let's assume that Ic < I1 + I2. In this case we are pushing I1 and I2 in the collector of Q1, but the collector current is smaller, so it cannot sink all the current → collector voltage is high, to +Vcc. Then we have an inverter (M2 and M3), which means that low level at the input corresponds to high level at the output and viceversa, so in this case the output of the overtemperature protection is 0V. But 0V makes M1 to turn on, because it is a pMOS M1 with the source at Vcc, and this is also the reason why we can say that the total current entering in the collector of Q1 is I1 + I2. So the initial assumption is correct.

If we increase the temperature, the collector current increases up to when Ic = I1 + I2. If we overcome, even by a small amount, this equality, the transistor Q1 enters the saturation regime and the collector goes to 0V. As a consequence the output of the protection circuit swings to the high digital level and the pMOS is off. This is the origin of the hysteresis.

HYSTERESIS



Let's assume that the power mosfet is operated initially in normal conditions, meaning T < Tmax. The two curves in the upper left plot are the collector currents.
Initially the temperature sensor transistor is in point x. If we apply an overload, trying to make a current flow in the power mosfet that is larger than the maximum current, pMOS starts to become hotter, and so also the temperature sensor, so the operation point is moving horizontally towards the right.

As soon as we reach point y, the protection is triggered, and the protection signal is high. If so, the power mosfet gets immediately turned off, so it starts to cool down.
Hence we are moving in the backward direction, but the output of the protection is still high because of the hysteresis. In fact, once the temperature protection is triggered, M1 becomes cutoff, so there is no more I2 in the collector, we are left just with I1, so **we are moving the trigger point at a lower temperature**.

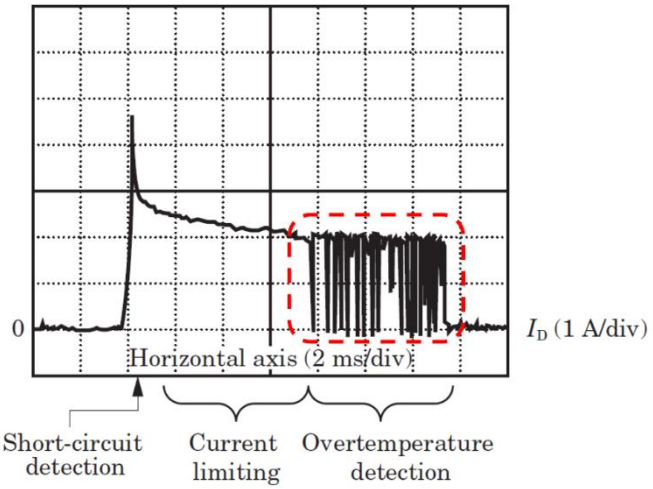
The output of the protection gets low when we reach the retriggering point z. The difference between Tmax and Tmin is the so-called hysteresis window, and it is defined by the user.

At point z the current Ic becomes smaller than I1, making the collector of Q1 to go up again, so the protection output voltage goes back to 0V.

When we are at Tmin, we have two possibilities:

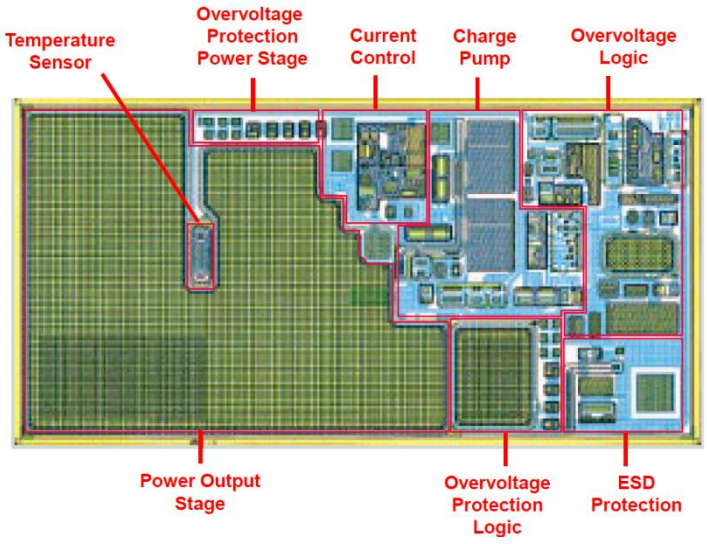
1. The reasons that caused the overload have been removed and the mosfet is cooling down getting back to the original working point x.
2. The reason for the overload are still there, so as soon as we turn on the power mosfet again, we it's becoming hotter again, and so we are cycling along the hysteresis. This is the situation represented in the bottom right plot. **The turn on and off occurs at a fixed frequency**, so we can filter out the disturbances with a notch filter at the frequency of the oscillations.

WITHOUT HYSTERESIS



If we don't use the hysteresis, we have an unpredictable switch on and off of the mosfet operated in the overload condition, it is erratic and we cannot predict when it turns on and off. This is a problem because this turn on and off produces a lot of electromagnetic interference in the ambient and it might disturb the electronics around.

Example – high side switch



The green is the power mosfet with the cellular structure, and the temperature sensor is placed inside the mosfet in the position where the maximum temperature is.

POWER ARCHITECTURE

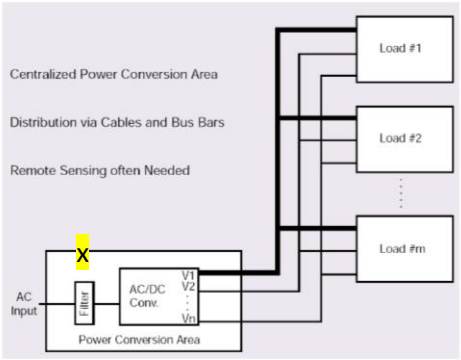
- All electronic equipment operates from a power source.
- These sources are, in order of prevalence, the AC powerline, portable batteries, or a fixed-location battery bank.
- Many electronic subsystems to be supplied by the single power source.
- Each of them has specific requirements in term of volts, amps, regulation, efficiency, etc.
- There are many ways to supply each subsystem matching the specific requirements: each way defines a *power architecture*.

DC/DC converters are very important because of the concept of **power architecture**. Any electronic equipment operate from a power source, and in order of prevalence, the power sources are AC powerline, portable battery and fixed location battery banks.
However, inside any electronic equipment there are a lot of substystems operated with a specific current and voltage. How can we separately bias each of the single subsystems starting from a single source of energy? Each configuration that we come up with is a power architecture.

How to select the power architecture

- The key parameters to be considered are:
  - Overall efficiency
  - Cost
  - Power density
  - Safety
  - Availability (Up-time)
  - Flexibility and ease of design .....
- When all the variations are taken into account, the number of possible architectures can be overwhelming and confusing.
- The best way to begin exploring these architectures is to first define the most basic and widely used ones.

CENTRALIZED POWER ARCHITECTURE



- A centralized power architecture has all of the power processing functions packaged at a single physical location within the equipment.
- The load electronics are packaged separately and connected to the power processing function by DC distribution hardware.

Drawbacks:

- Remote sensing is required to guarantee tight regulation
- High distribution losses
- Output voltages always turned on
- Complex routing

In a centralized power architecture, all the power processing functions are packaged in a single physical location within the system x. The loads are connected to the power supply units through wiring.

Drawbacks

- If we want to tightly regulate the voltage across each load we need a remote sensing, we need to measure the voltage at each single load and use it as a feedback value. Remote sensing is necessary because wirings have their own parasitic resistances, so if we have a large current absorbed by the load, a huge voltage drop can develop and the voltage across the load might be different with respect to the ideal one.
- High distribution losses, for the same reason as before. If the load absorbs a large current, a small parasitic resistance ends up in a huge power dissipation due to Joule effect.
- Output voltages are always turned on. To remove or change a load we need to shut down the entire system.
- Complex routing, because we need to distribute the single voltage, consider the remote sensing and so on.

These are the main reasons why this approach has been abandoned.

ISSUES

Wide range of power supply voltages

We have different voltage levels depending on the application, and generating them with a centralized power architecture is almost impossible.

Application	Voltage Rails
Set-Top Box	7
Personal Computer	10
Plasma TV	15
Low-End Server	20
Low-End Ethernet Switch	40
Mid-Range Server	60
Mid-Range Ethernet Switch	150
Carrier-Class Ethernet Switch	200+
Class 1 Telecom Switch	200+

IC technology trends

- All electronic systems now contain integrated high performance processors, ASICs and FPGAs.
- The supply voltage level of integrated circuit keeps reducing whereas the current increasing.
- Tighter voltage regulation is required to meet IC voltage requirement.
- High load step transient produce voltage transients that the supplies usually must correct. Fast loop response is required.
- Transient regulation requirements now often necessitate much more closely controlled dynamic behavior than that of the suppljes system designers considered more than adequate just a few years ago.
- IR drops in the distribution networks (that is, in the power and ground planes) must be reduced.

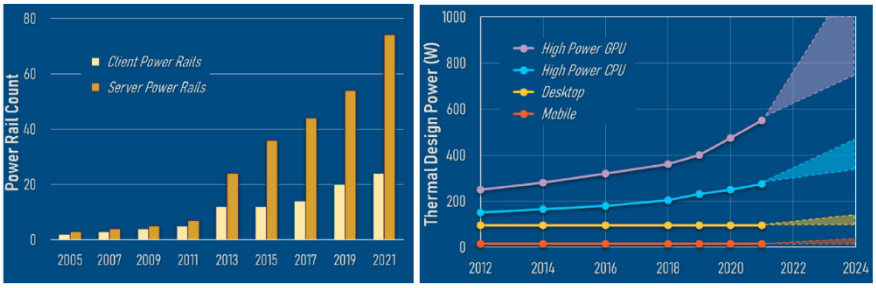
Some numbers

- High levels of current demand at multiple low supply voltages, tight voltage regulation, large and fast dynamic currents are commonly met in electronic systems .

Processor	Nominal Voltage	Output Tolerance ($\Delta V_{OUT(AC)}$)	Max Current	Max Load Step	Load Slew Rate
Communications ASIC	1.0 V	±50 mV (±5%)	120 A	60 A	12 A/µs
Automotive CPU	0.90 V	±22 mV (±2.5%)	90 A	60 A	100 A/µs
FPGA	0.88 V	±17 mV (±2%)	200 A	100 A	200 A/µs
Server processor (current gen)	1.8 V	±22 mV**	228 A	200 A	780 A/µs
Server processor (next gen)	1.8 V	±22 mV**	400 A	300 A	670 A/µs

Modern processor spec examples

Power supply systems for computing applications



- The number of power rails has been scaling up with the core count
- Overall power levels have been going up in the high-power segments

Power demand in future computing

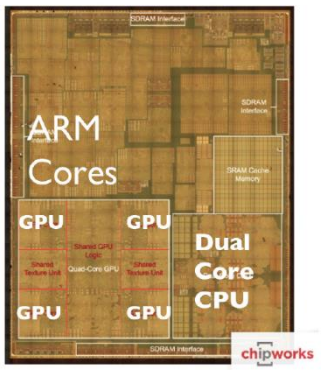
- **Processor-Level**
  - High current (>1,000 A/core)
  - High conversion ratio (48 V-1 V)
  - Fast control bandwidth (>1 A/ns)
  - High density (>100 A/cm^2)
  - High efficiency (>95%)
- **Server-Level**
  - Modular loads (up to 16 GPUs together)
  - Core-to-core communications

- Extreme prerformance 48V-1V PoL

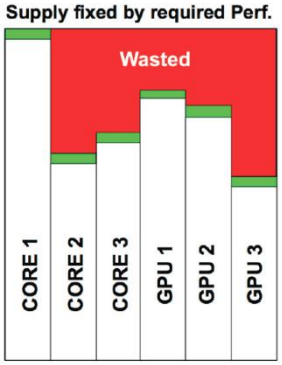
Power waste with single power supply

The need of multiple supply voltages is not only present in a simple board, but also inside a single chip component. An example is a microprocessor.
If we apply the same bias supply to each core inside the microprocessor, it is not a good idea because each core is operated at different voltages. So a part of the applied voltage is bringing a waste of power.

Hence the reason for multiple power supply is that each subsystem must be operated at the minimum voltage level that allows the subsystem to perform its task, in order to have a system operated with high efficiency → this is solved integrating DC/DC converters inside the chip.



Apple A8



- **Clear need for multiple independent supplies**
  - Per-block power management, dynamic supplies, …

The problem with this is that we are not speaking of a static situation, the cores might require different supply voltage depending on the task → voltage supply must be dynamic, and changes must be done in ns.

POWER DISTRIBUTION PROBLEM

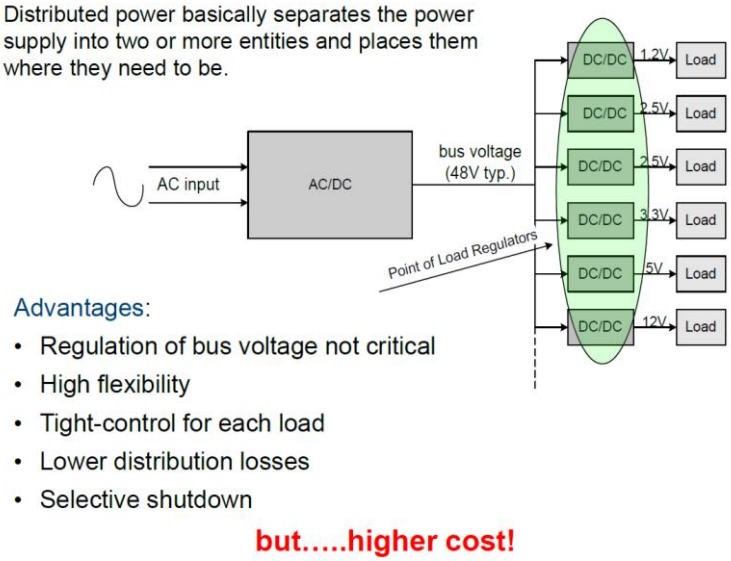
The bottom line is the presence of very complex power management needs in computing and networking systems. We don't have just to provide power, but also to manage it.

- Very complex power management architectures in computing & networking systems
  - Many low voltage, high current and tightly regulated voltage sources
  - Accurate management for sequencing, tracking and fault management
  - Low cost mandatory in high volume applications
  - High power density needed for volume reduction → thermal management issues

→ Distributed Power Architectures

If we think at all this problem, one solution might be the distributed power architecture.

DISTRIBUTED POWER ARCHITECTURE



Distributed power architecture separates the power supply typically in two entities. There is a first block where the AC input is converted into DC. This AC/DC converter is producing a loosely regulated voltage called bus voltage. Nowadays typical bus voltages are 12V or 48V.
The bus voltage is distributed using a voltage bus, which is the same concept of a data bus. Then, close to each subsystem the bus voltage is converted down to the needed voltage with a DC/DC converter. In particular, in this case it is a closed-loop DC/DC converter, so it is a **DC/DC regulator** (also called VRM).

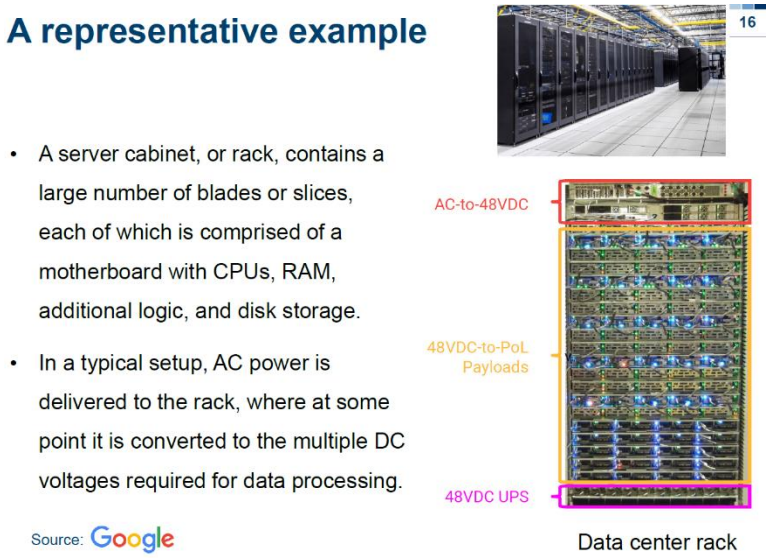
Advantages

- Regulation of the bus voltage is not critical. It is a great advantage because we can add or remove a load simply plugging it or removing it, because each load has its own PS.
- High flexibility, because we have DC/DC converters.
- Tight control for each load.
- Lower distribution losses. If we assume that one load is e.g. 100W and PS is 1V, we would need 100A to be supplied to the load. Without a DC/DC converter we would end up with 100A flowing in the wire without the DC/DC converter, and it's a too big value.
- Selective shutdown. We might have a failure in a subsystem, and we can remove it without turning off all the system.

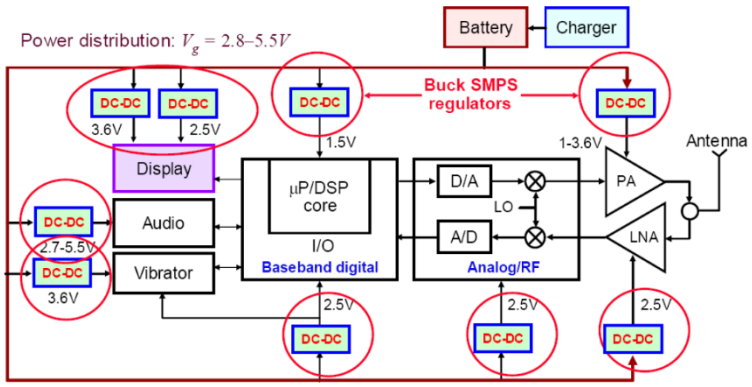
Features

- Centralized AC/DC converter to interface with the AC powerline.
- The AC/DC converter provides the functions of safety isolation, DC conversion, noise suppression and power factor correction.
- An isolated intermediate SELV (safety extra-load voltage) DC voltage (24 to 60 V) is distributed within the product.
- Individual load converters (DC/DC) are used for each load.
- The DC/DC converters are physically located at, or very close to, the load.
- The DC/DC converters are standardized modules or components.

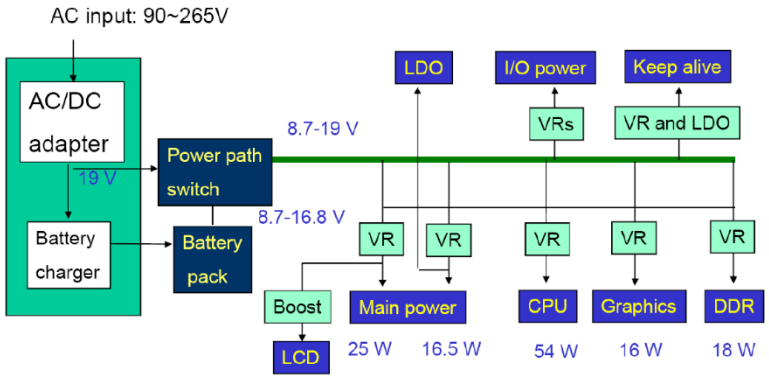
Example of a distributed power architecture



Power architecture for a smartphone

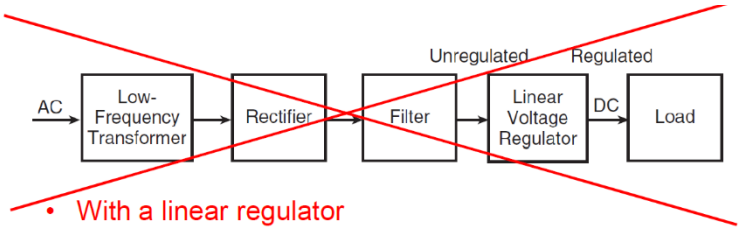


Power architecture for a computer

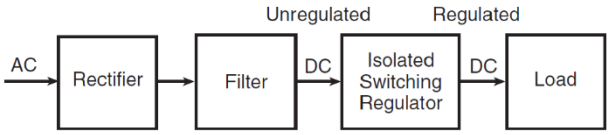


Also, if we consider the AC/DC converter, it is basically a DC/DC converter because the AC input is immediately rectified and filtered to obtain an almost flat unregulated voltage around 400V, which is then reduced down to the bus level by using a DC/DC converter (typically it is an isolated DC/DC converter working at high frequencies).

The difference between the old (top) approach and the new (bottom) one is that in the former the transformer is operated at 50Hz, in the latter at 100kHz. And, **at the same output power, the larger the switching frequency at which the transformer is operated, the smaller the size**.



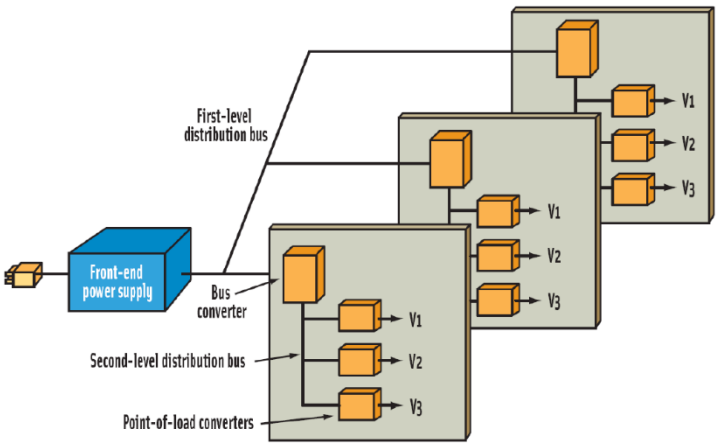
- With a linear regulator



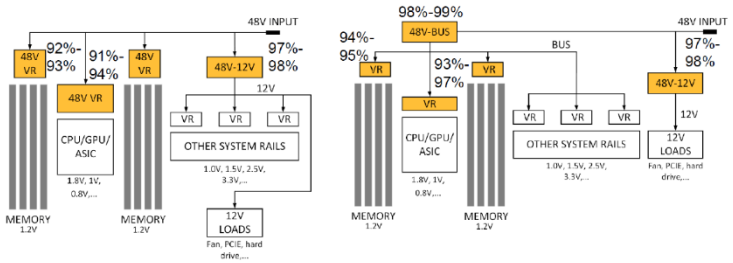
- With a switch-mode voltage regulator.

INTERMEDIATE BUS ARCHITECTURE

The idea is that we still we have an AC/DC converter to generate a first level bus voltage, this voltage is distributed and on each board there is a second DC/DC converter that generates a second level distribution bus voltage. Then to each single load we have a DC/DC converter.



Example architecture



Direct conversion approach

2-stage conversion approach

Key advantage of an IBA System

- System cost is reduced because only one isolated converter is needed and because low-cost, standardized non-isolated POL regulators are available in the market.

Basic challenge for an IBA System

- The intermediate-bus converter must have the highest efficiency and power density to provide a competitive edge for IBA versus DPA.

Source: Google

The advantage is that the first level converter is typically operated in open loop, whereas the second level converter can be operated with a larger efficiency because the difference between input and output voltage is smaller.

Bus converters

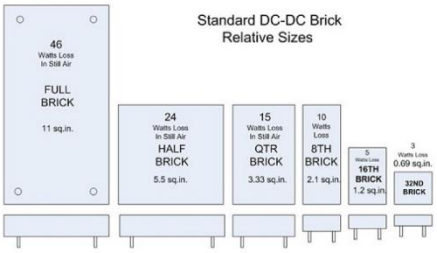
Unregulated bus converters
Input: 42~53V
Output: 9.7V~13.5V

Input: 42~53V
Output: 12V ±3%, 25A

Input: 36V~75V
Output: 12V ±1.5%, 14A

Input: 36V~75V
Output: 12V ±10%, 14A

Regulated bus converters



The six standard brick sizes

- **Full brick** – 4.6 x 2.4 x 0.5 inches or 116.8 x 61.0 x 12.7mm
- **Half brick** – 2.3 x 2.4 x 0.35 inches
- **Quarter brick** – 2.3 x 1.45 x 0.35 inches
- **8th Brick** – 1/8
- **16th Brick** – 1/16
- **32nd Brick** – 1/32

DC/DC POWER CONVERTERS

TYPES OF CONVERTERS

Linear regulators have the problem of very low efficiency, even if the cost is low and the power density very large. Also output noise is negligible. However, we use them only when the voltage drop between the input and the output is limited around 1V or less.

We will focus on **switching regulators**, and there are two families:

- Inductive: based on the use of inductors as filters. We can reach very high efficiency but at a very high price and large PCB area occupied. Once again, a regulator is a converter with feedback.
- Charge pump (also called switch cap): there are no inductors, just capacitors. The basic principle is the one of a charge pump and they can be very compact because we don't use inductor. Furthermore, efficiencies also in this case can be high. In general we can get just a fixed conversion ratio, whereas in inductive ones we can regulate the output voltage in the range we want.
- Hybrid converters: hybrid solution between the previous two. It is basically a switch cap converter using small value inductors to improve the performance of the converter.

- The choice of converter type depends on the power design priorities.

	Linear regulator	Switching regulator	
		Inductive	Charge Pump
Efficiency	20-60%	90-95%	75-90%
Ripple	Very low	Low	Moderate
EMI Noise	Very low	Moderate	Low
PCB Area	Very small	Largest	Medium
Cost	Lowest	Highest	Medium

Switching conversion techniques

Charge Pumps

- Simple & moderate cost
- Can be step up/down or invert
- Require only switches and capacitors
- Noise performance is architecture dependant

- Issues include:
  - Limited choices for V_{out} vs. V_{in}
  - Can be noisy
  - Limitations on output power

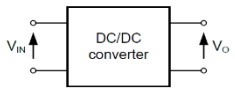
Inductor Based

- Can be step up/down or invert
- Can handle a wide range of input and output voltages
- Generally higher cost
- Can obtain high current levels

- Issues include:
  - Need for magnetics
  - Board layout can be critical.
  - Require more components than charge pumps

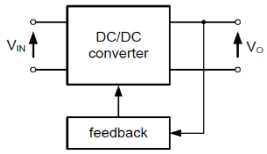
SM DC/DC converters

Regulators add a feedback loop to the DC/DC converter.



Converter

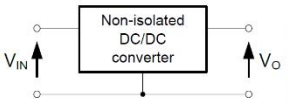
- Switch-mode dc-dc converters utilize one or more switches to transform dc from one level to another



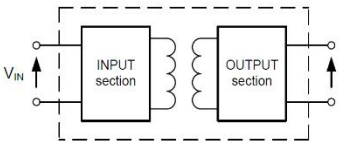
Regulator

- Feedback is included to convert the unregulated DC input to a controlled DC output with a desired voltage level

Isolated vs non-isolated



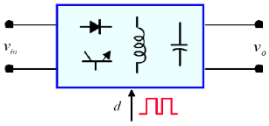
Non-isolated dc–dc converters may have *common negative* or *common positive* input and output terminals.



Isolated dc-dc converters use transformers to obtain dc isolation between the input and output and between the different outputs.

The isolated topology defends better against failures in one of the two sides of the circuit.

Converter topology

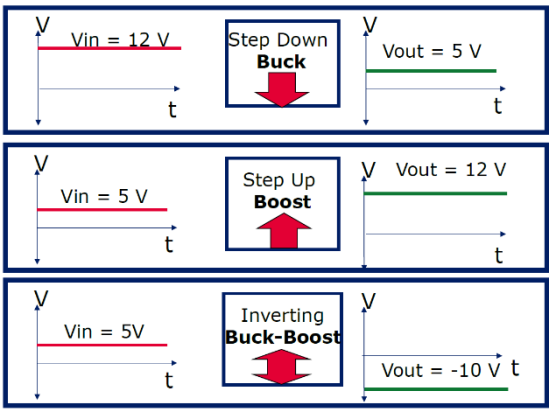


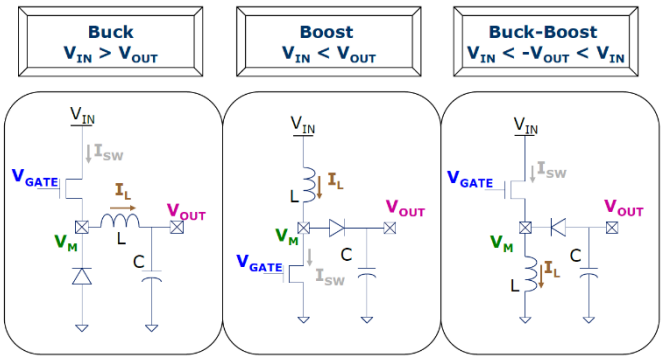
- A topology is the arrangement of the power devices and their magnetic elements
- 8 “mainstream” topologies:
  - non-isolated
    - buck, boost, buck-boost
  - isolated
    - flyback, forward, push-pull, half-bridge, full-bridge

A topology is the arrangement of the power devices and filter elements inside the DC/DC converter.

Non-isolated topologies

In the **Buck** converter the output voltage is always smaller than the input voltage. In the **Boost** converter the output voltage is higher than the input one. The **Buck-Boost** is an inverting converter, the sign of the output voltage is opposite with respect to the input one and the amplitude can be smaller or higher.

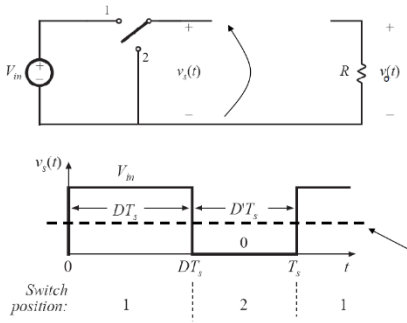




All topologies consists of the same basic components but are arranged differently

BUCK CONVERTER

We have a single pole double throw switch which is periodically switching between position 1 and 2 generating a square wave voltage v_s(t). It is the PWM modulation.
The impact of regulating the duty cycle is on the average voltage generated v_s(t), which is the input voltage multiplied by the duty cycle.



$$T_s = t_{on} + t_{off} = \text{const}$$

$$V_S = \frac{1}{T_s}\left[\int_0^{t_{on}} V_{IN}\,dt + \int_{t_{on}}^{T_s} 0\,dt\right] = \frac{t_{on}}{T_s} V_{IN} = D\,V_{IN}$$

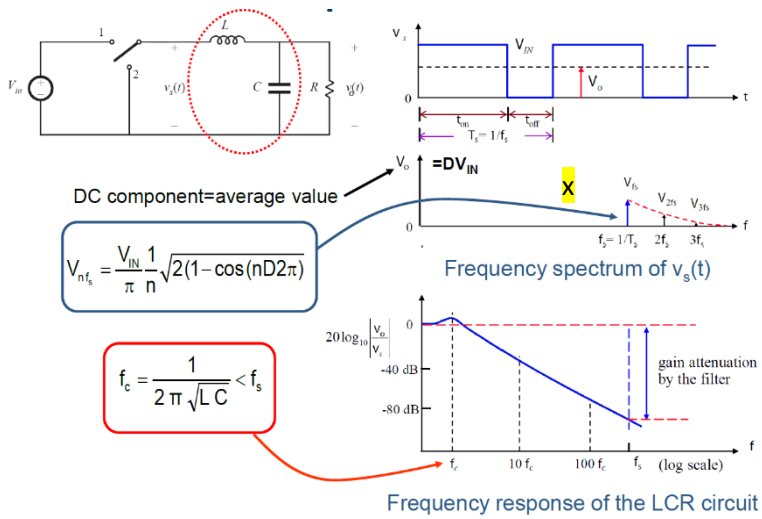
- switching at constant frequency:

  → pulse-width modulation (PWM) switching

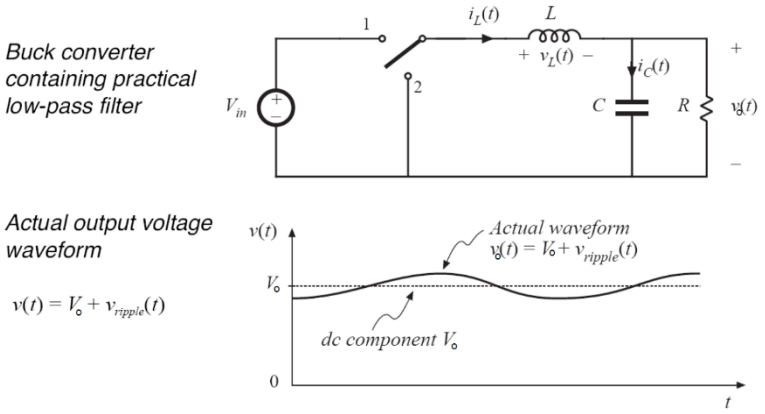
Insertion of a LP filter

Of course we cannot apply a square wave voltage to the load, we need to filter it to extract the average voltage. This is done with a second order filter, not a first order one because if we use a resistor we would have a very large power dissipation.

Plot x is the frequency spectrum of a square wave voltage waveform.

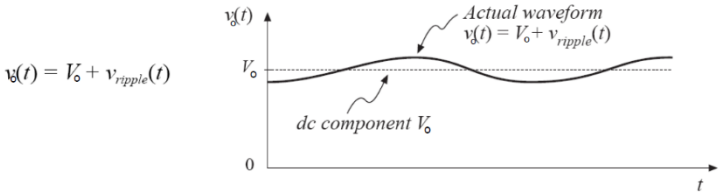


The DC component we are interested in is at f = 0, and we want to remove the other components at multiple of the switching frequency. This is done designing correctly the corner frequency of the LP filter, by choosing it much smaller than the switching frequency.
We can attenuate the other harmonics but not get rid of them, and the result is that we don't have a perfectly flat voltage across the load, but a flat one with some ripples on it.



In our analysis we will always apply the small ripple approximation.

Small ripple approximation



In a well-designed converter, the output voltage ripple is small. Hence, the waveforms can be easily determined by ignoring the ripple:

$$\left| v_{ripple} \right| \ll V_o$$

$$v_o(t) \approx V_o$$

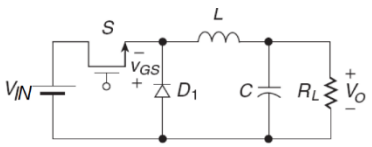
We assume that the peak to peak amplitude of the ripple is negligible with respect to the DC value of the output voltage, which is always true if the DC/DC converter has been properly designed.
If so, the instantaneous voltage at the output of the DC/DC converter can be replaced with the average voltage, which is the ideal voltage if we assume that the filter is ideal.

IMPLEMENTATION

We cannot implement a single pole double throw switch by using only semiconductor power device, we need to use several of them, and the one below is done with an active device (mosfet) and a passive one (diode). The diode D1 can eventually be replaced with an active component operated in antiphase with respect to the other switch.

The one in the image is a synchronous buck converter.
The switch is controlled by a pulse width modulator with a fixed frequency. The switch is on during t_on and off during t_off. By definition, the duty cycle is D = t_on/Ts, where Ts is the switching period.

Once we have filtered it, the output voltage is D*Vin.



- The SPDT switch is implemented by means of a DMOS transistor , S, and a diode, D1.
- The switch S is controlled by a *pulse-width modulator* and is turned on and off at the switching frequency $f_s = 1/T_s$.
- The switch S is ON during the time interval t_{on} and OFF during the time interval t_{off} . The duty cycle D defined as:

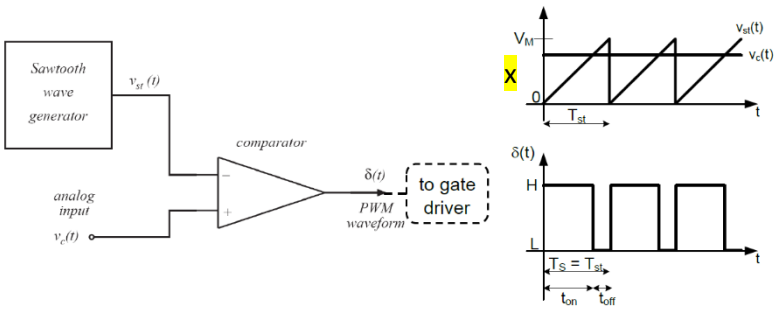
$$D \equiv \frac{t_{on}}{t_{on}+t_{off}} = \frac{t_{on}}{T_s}$$

- The output voltage V_O of the buck converter is always lower than the input voltage V_{IN} , therefore, it is a *step-down* converter.

PULSE WIDTH MODULATOR

We need to generate a PWM signal. The pulse width modulator is the combination of a Sawtooth wave generator, that is a voltage generator that generates a voltage waveform like x, and a comparator. The analog voltage used for comparison is called control voltage v_c(t). It is a DC voltage but it might not be.

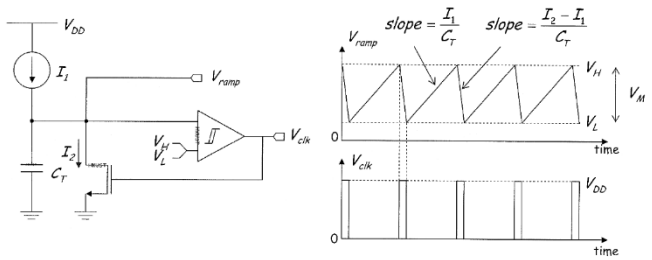
The output of the comparator depends of course on the relative value of the signal x and the control voltage.



Note that: $T_S = T_{st}$ ⇨ $D \equiv \frac{t_{on}}{T_s} = \frac{V_c}{V_M}$

By using the principle of similar triangles we can say that D = v_c/Vm. The user regulates the duty cycles regulating the control voltage.

Sawtooth voltage generator



- The ramp and clock signals are synchronous automatically
- The switching frequency is controlled by V_H, V_L, C_T, I_1 and I_2

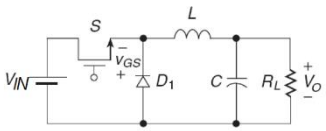
  for $I_2 \gg I_1$, $f_{sw} \approx \frac{I_1}{(V_H - V_L)C_T}$
- The amplitude of the ramp signal defined by V_H and V_L

It is a fixed current generator, a capacitor in which a constant current is injected and the voltage in the C is increased linearly until we reach the high threshold of a Schmitt trigger. When so, the output of the trigger goes up, the switch is closed and the voltage across the capacitor is reset.

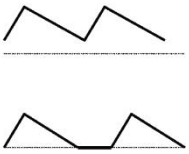
TYPES OF BUCK CONVERTERS

The buck converter can operate in, depending on the current that flows in the inductor:

- Continuous current mode (CCM): the current in the inductor never reaches 0 during the switching time intervals. However, due to the presence of the unidirectional device that the diode is, the converter might operate in the next type of conduction mode.
- Discontinuous current mode (DCM). In this case the current in the inductor reaches 0 within the switching time interval and stays 0 for a fraction of the switching time period.



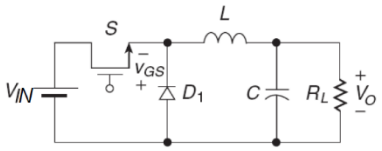
Two types of Conduction Modes
– Continuous Conduction Mode (CCM)
where Inductor current remains positive throughout the switching period
– Discontinuous Conduction Mode (DCM)
where Inductor current remains zero for some time in the switching period



Assumptions in the analysis of a Buck converter

- The converter is analyzed in steady state, which means that we have been switching the switch for a long time, so all the transients are over.
- The switches are ideal.
- The losses in the inductive and the capacitive elements are neglected.
- The dc input voltage has zero internal impedance.
- The small ripple approximation is applied.
- The converter is operating in CCM regime.

CCM ANALYSIS



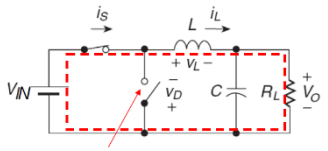
- DC voltage transfer function may be derived from volt-second balance:
  - average voltage across inductor is zero in steady state

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{v_{Lon}t_{on} + v_{Loff}t_{off}}{T_S} = 0$$

The current is never 0.
Which is the DC voltage transfer function (ratio between the output voltage and input one at steady state)?
To get it, we can apply the volt-second balance to the inductor. Vl,on*t_on + Vl,off*t_off = 0.

Ton

The active switch is closed and the passive switch (diode) is open because reverse biased. A voltage Vl,on = Vin – Vout is applied to L.
Then we write the formula in the red box.



Equivalent circuit when the switch is ON and the diode is OFF

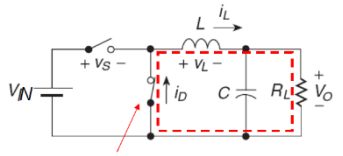
Reverse biased, thus the diode is open

- Switch closed for $t_{on}=DT_s$ seconds
- $v_{Lon} = V_{IN} - V_O$,

$$v_{Lon} t_{on} = (V_{IN} - V_O) DT_S$$

Toff

The active switch is open but there is current in the inductor because we are in CCM, so the current goes through the diode, which is turned on.



Equivalent circuit when the switch is OFF and the diode is ON

i_L continues to flow during t_{off}, thus the diode is closed. This is the assumption of "continuous conduction" in the inductor (CCM).

- Diode conducting for $t_{off}=T_s - t_{on} = T_s(1-D)$ seconds
- $v_{Loff} = -V_O$,

$$v_{Loff} t_{off} = -V_O (1-D) T_S$$

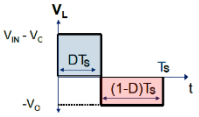
The voltage drop across the inductor is negative and -Vout.

Transfer characteristic

We apply the volt-second balance.

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{DT_S \cdot (V_{IN} - V_O) + (1-D)T_S \cdot (-V_O)}{T_S} = 0$$

$$D \cdot V_{IN} = D \cdot V_O + V_O - D \cdot V_O$$



The input/output equation becomes $\frac{V_O}{V_{IN}} = D$

From power balance, $V_{IN} I_{IN} = V_O I_O$, so

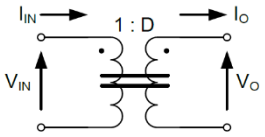
$$I_O = \frac{I_{IN}}{D}$$

The DC voltage transfer function is **D = Vout/Vin**.
We have assumed that the components are ideal, so there is no power lost in the circuit, so we can write a power balance Pin = Pout, with Pin that is the product between the input voltage and the average input current. In the end, **Iout/Iin = 1/D**.

The ideal buck converter working at steady state can hence be easily modelled with an equivalent circuit like below, which is an ideal transformer. It is not an ideal transformer because it also transform DC values.

- input power equals output power:

$$P_{IN} = P_O$$
$$V_{IN} I_{IN} = V_O I_O$$
$$\frac{I_{IN}}{I_O} = \frac{V_O}{V_{IN}} = D$$



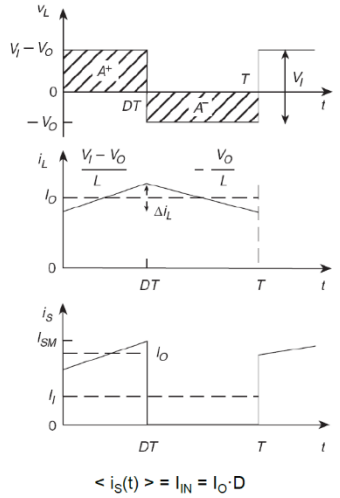
- step-down converter is equivalent to a dc transformer where the turns ratio is in the range 0-1

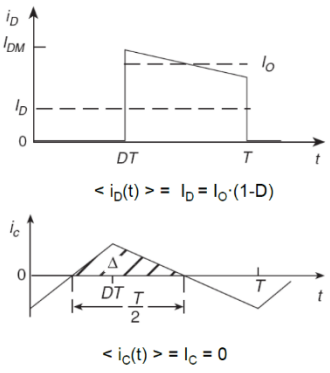
Of course, this model is good if the buck converter is ideal, if not we need some workarounds.

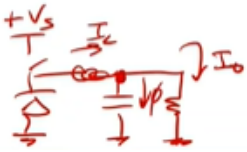
CCM waveforms

The ones below are the typical waveforms we can measure in a buck converter working in CCM. The first upper left is the voltage across the inductor. The areas A+ and A- are equal, because of the volt-second principle. The second left plot is the current in the inductor, whose average value is equal to the output current. At steady state, the average current in the capacitor is 0, so the average inductor current is equal to the output current (red circuit).
Besides the average value, there is also a current ripple due to the harmonics component of the square wave voltage on the switch.

i_s is the current in the switch, which is equal to the current in the inductor when the switch is closed.
i_d is the current in the diode, which is of course 0 when the switch is on.
The last plot is the instantaneous current in the capacitor.

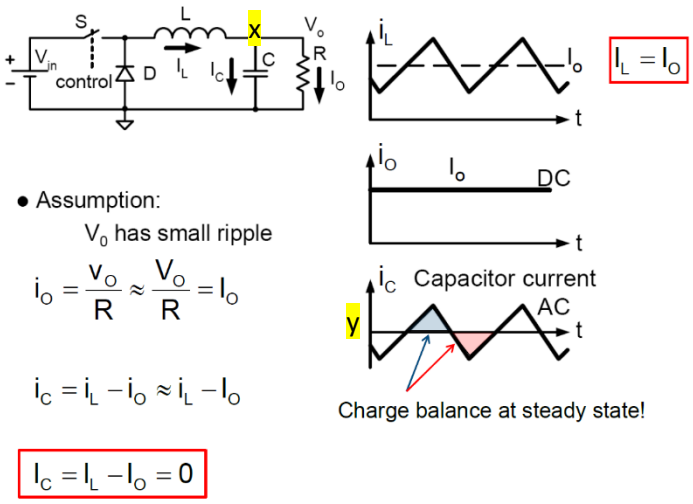






Capacitor current

It is important to understand the shape of the capacitor current to be able to calculate the output voltage ripple across the capacitor and therefore at the output of the DC/DC converter.
We start by applying the KCL at node x assuming the instantaneous value of the current. i_L coming from the inductor is the first plot, having a DC component Io plus a ripple. As for the current that flows in the load, we start from the **assumption that the output voltage ripple is negligible**.
In reality, the output voltage has a DC component plus a ripple, and so also the current, but we can assume that the amplitude of the ripple is negligible, and this is reasonable because the filter of the buck converter is design to satisfy this assumption.



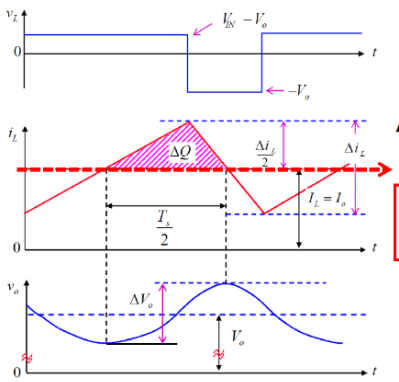
Given the assumption, we can say that the instantaneous output current in R, i_o, which is the instantaneous output voltage v_o divided by R, can be approximated as Vo, so the DC voltage value at the output of the buck converter, divided by R, so Io.

Now we can easily apply the KCL at node x. In the end we get that the average current flowing through the capacitor Ic = 0. This is of course true because we are operating at steady state.

Now we want to plot the capacitor current i_c using the small ripple approximation. We start from the inductor current, we subtract from it the average inductor current, so we are shifting the i_L waveform by a factor Io. The area subtended by the positive part of the curve of i_c must be equal to the negative one (charge balance at steady state).

Let's now remove the assumption that the ripple is negligible and let's calculate the ripple starting from plot y.

Output voltage ripple



$$\Delta Q = \frac{T_S}{2}\frac{\Delta i_L}{2}\frac{1}{2} = C \cdot \Delta V_O$$

$$\Delta V_O = \frac{\Delta Q}{C} = \frac{1}{C}\frac{1}{2}\frac{\Delta i_L}{2}\frac{T_s}{2}$$ x

$$\Delta V_O = \frac{\Delta i_L \cdot T_S}{8C}$$

$$\Delta V_O = \frac{(1-D)V_O}{8C \cdot L \cdot f_S^{\,2}}$$

$$\frac{\Delta V_O}{V_O} = \frac{(1-D)T_s^2}{8LC} = (1-D)\frac{\pi^2}{2}\left(\frac{f_c}{f_s}\right)^2$$

The first plot is the inductor voltage, which is Vin – Vout during the on time period D*Ts and -Vout during Ts, off time period. The second plot is the current flowing through the inductor.
We are not interested in i_L but in i_c. We can plot i_c on the same i_L graph shifting the origin of the x axis at a higher position (red dashed line).

Now we can calculate the voltage ripple across the capacitor. In fact, we have a capacitor which is charged during the pink period and discharged in the other subsequent part of the period.
At steady state, the charge injected must be equal to the one extracted.

When we are injecting charge in the capacitor, the voltage across the capacitor is going to increase, and this is what happens in the last plot. When we extract the charge, the voltage is going to decrease. We are at steady state, so the amount of increase must be equal to the among of decrease. We are interested in the peak-to-peak voltage ripple, which can be computed by dividing the charge injected in the capacitor by the capacitance value itself (x).

To get delta_Q we need to compute the area of the pink triangle. Its amplitude is half of the current ripple, so delta(i_L)/2, and it is matched by the negative delta_Q. The area of the two triangles is equal due to the charge balance. Then, the sum of the two bases is Ts, so each triangle has a base Ts/2.
In the end we get the formula in the red box. **It holds only for buck converters and forward converters, not for all the other converters**, because in a buck converter the current in the capacitor is a triangular one, and in the other converters is instead rectangular.

Previously, to compute the shape of the current in the capacitor we assumed the voltage ripple negligible. Then we are using the waveform used under this assumption to compute the voltage ripple. It is not correct, it is an approximation but it is a very good first order approximation of the voltage ripple.
So we calculate the current assuming that the voltage ripple is negligible and then we use the current to estimate the first order approximation of the voltage ripple.

Coming back to the expression in the red box, |delta(i_L)| = Vo(1-D)*Ts/L is the current across the inductor, and we can substitute in the formula.
Then we calculate the relative amplitude of the voltage ripple. The corner frequency of the filter is fc = 1/(2*pi*sqrt(LC)), we plug it into and we get the final expression.

The relative voltage ripple is proportional to the square of the ratio between the corner frequency and the switching frequency. The square is present because we are filtering the square wave voltage waveform with a second order filter. Moreover, the smaller the ratio fc/fs, the smaller the amplitude of the relative voltage ripple, because it means that the harmonics are more attenuated.

There is a strong assumption that we are making here. In fact, we are considering the ESR of the capacitor negligible, so the capacitor is ideal.
In the real case, this is true or not depending on the choice of the capacitor. A multilayer ceramic capacitor is almost ideal, so negligible series resistance, but if we use an electrolytic capacitor the equivalent series resistance has to be considered.

How to reduce the voltage ripple

Reduce voltage ripple

- Decrease the corner frequency → $L\uparrow$ or $C\uparrow$ → Larger size of the converter
- Increase the switching frequency → switching loss ↑, efficiency ↓
- How to reduce switching loss?
  - Use low switching loss device, such as MOSFET
  - Use soft switching topology, such as ZVS, ZCS
  - $f_s \approx (10\sim100)\, f_c \rightarrow \Delta V_o/V_o$ usually is less than 1%

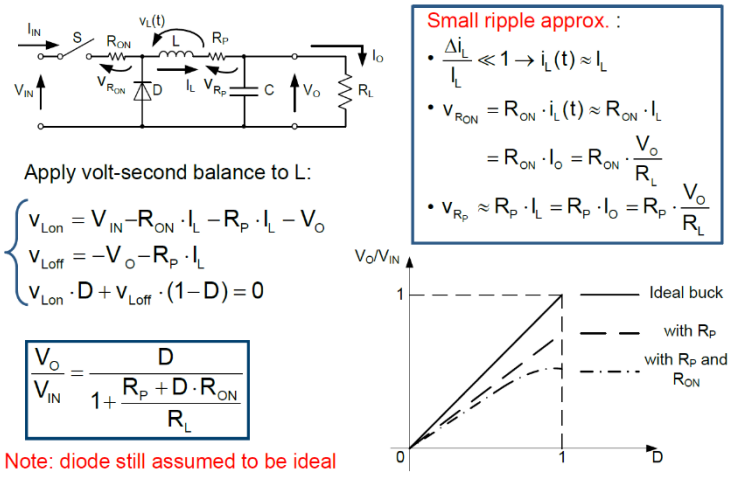
Looking at the expression, we might decrease the corner frequency with respect to the fs, but this means larger capacitor and inductor, and this impacts the size of the converter.

In general, DC/DC converters are designed in such a way that the relative amplitude of the voltage ripple is in the order of 1%.

NON IDEAL BUCK CONVERTER – Small ripple

In this circuit we are introducing a resistive component Ron which is used to take the switch resistance into account (we are considering Rds(on)) and also the inductor is no more ideal with its parasitic series resistance Rp.
The calculations are based on the **small ripple approximation**, but in this case we **are referring to the current ripple in the inductor**, not the voltage across the capacitor. It the condition is satisfied, we can replace the instantaneous current i_L(t) with the average current I_L.

Once we have introduced this approximation, we need to compute the DC voltage transfer function, considering the parasitism.
We apply the volt-second balance to the ideal inductor. To properly do so, we need to take the voltage drop on the Ron and Rp resistances into account.
V_Ron is, by definition, the product between Ron and the instantaneous current flowing through it, which is i_L(t) (I_L for the small ripple approximation), because when the switch is closed the diode is open. Same reasoning for V_Rp.

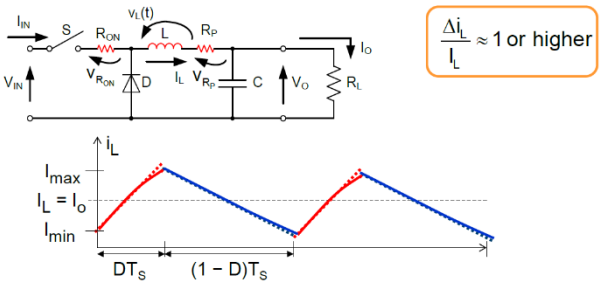


Then we apply the volt-second balance: Vl,onD + Vl,off(1-D) = 0.
To write Vl,on and Vl,off we use the KVL, always **under the small ripple approximation on the output voltage (v_o = Vo)**. For Vl_off the current recirculates in the diode, because we are under CCM.

We end up with a non-linear dependance of the DC voltage transfer function on the duty cycle D. The plots for the t.f. are in the bottom right plot. If only Rp is the non-ideal component, the slope of the t.f. is reduced but it is still linear.

Non negligible ripple

If the ripple is not negligible but linear, the approach can still be used. If neither negligible and nor linear we cannot.
If the ripple is not negligible and we draw the current in the inductor, we end up with pieces of exponential (red portions of the plot, and blue ones), so it is even not linear.
We still can apply the volt-second law, but we cannot replace the instantaneous current with the average current, because they are different.



Apply volt-second balance to L (no assumptions):

$$\begin{cases} v_{Lon} = (V_{IN} - V_O) - R_{ON} \cdot i_L(t) - R_P \cdot i_L(t) \\ v_{Loff} = -V_O - R_P \cdot i_L(t) \\ \int_0^{DT_S} \left[(V_{IN} - V_O) - R_{ON} \cdot i_L(t) - R_P \cdot i_L(t)\right] dt - \int_{DT_S}^{T_S} \left[V_O + R_P \cdot i_L(t)\right] dt = 0 \end{cases}$$

Hence we write the volt-second balance considering the instantaneous current flowing in the parasitic resistors.

So I need to compute the integral using the effective shape of the current i_L(t), which might not be very easy. However, there is a possible simplification, which occurs when the ripple is non negligible but linear.

Non negligible LINEAR ripple

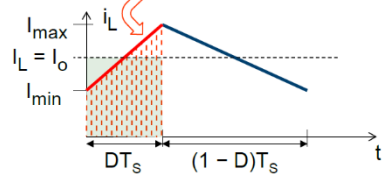
$$\int_0^{DT_S} \left[(V_{IN} - V_O) - R_{ON} \cdot i_L(t) - R_P \cdot i_L(t)\right] dt - \int_{DT_S}^{T_S} \left[V_O + R_P \cdot i_L(t)\right] dt = 0$$

$$x \quad V_{IN} \cdot D - V_O - \frac{R_{ON}}{T_S} \int_0^{DT_S} i_L(t)\,dt - \frac{R_P}{T_S} \int_0^{T_S} i_L(t)\,dt = 0$$

Area of trapezoid

$\approx D \cdot I_L$ linear ripple

$\equiv I_L = I_O$ always true



Linear ripple approx. :

- $\frac{L}{(R_P + R_{ON})} \gg DT_S$; $\frac{L}{R_P} \gg (1-D)T_S$

$\Rightarrow$ area red trapezoid $\approx$ area green rectangle

i.e. $\frac{1}{T_S} \int_0^{DT_S} i_L(t)dT\,dt \simeq D \cdot I_L = D \cdot I_O$

Replacing i_L (t) with I_L is still OK in the assumption linear ripple!

Firstly we can brought the constant term outside the integral and rewrite the integral and we get x. Still no approximation is applied.
The second integral in x is the integral from 0 to Ts, so over the full period of the current, divided by Ts. This is the definition of the average inductor current, which is equal to the output current in a buck converter, this is always true, no matter if the ripple is linear or not.

The approximation is in the first integral in x, it is extended over just Ton, and not the full period Ts, and the integral can be geometrically interpreted.
We are assuming that the ripple is linear, so we have a triangle and the integral is the dashed area of the trapezoid. The area of the trapezoid, if the ripple is linear, is identical to the area of the green rectangle.

So we can replace the integral with the simple product of the area of the rectangle.

To have a linear ripple, the exponential time constant during the on time period, L/(Rp +Ron) much be much larger than the duration of the on time period, DTs.
And also the exponential time constant during the off time period, L/Rp much be much larger than the duration of the off time period, (1-D)Ts.

The final result is that, if the ripple is linear, we get the formula calculated assuming a negligible ripple.

$$V_{IN}\cdot D - V_O - \frac{R_{ON}}{T_S}\int_0^{DT_S} i_L(t)dT\,dt - \frac{R_P}{T_S}\int_0^{T_S} i_L(t)dt = 0$$

Replace $i_L(t)$ with $I_L = I_O$ (ok if ripple is linear)

$$V_{IN}\cdot D - V_O - R_{ON}\cdot D\cdot I_O - R_P\cdot I_L = 0$$

Replace I_O with V_O/R:

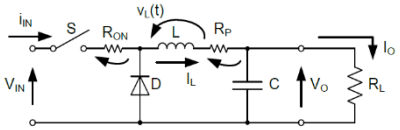
$$\frac{V_O}{V_{IN}} = \frac{D}{1+\frac{R_P + D\cdot R_{ON}}{R_L}}$$

Same result previously obtained with *small* ripple approximation

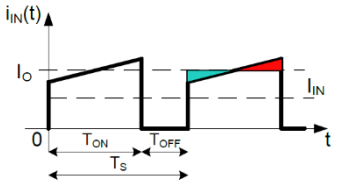
The conclusion is that we can replace the instantaneous current flowing in the inductor with the average current even if the ripple is non-negligible but linear.

BUCK CONVERTER'S EFFICIENCY

Let's compute the efficiency in a situation where there are parasitic components that are dissipating power. In this calculations, switching losses are not included.



$$\frac{V_O}{V_{IN}} = \frac{D}{1+\frac{R_P + D\cdot R_{ON}}{R_L}}$$ x



$$\frac{I_O}{I_{IN}} = \frac{1}{D}$$ Valid for small or linear ripple y

$< i_{IN}(t) > = I_{IN} = I_O\cdot D$

$$\eta = \frac{P_O}{P_{IN}} = \frac{V_O I_O}{V_{IN} I_{IN}} = \frac{1}{\cancel{D}}\,\frac{\cancel{D}}{1+\frac{R_P + D\cdot R_{ON}}{R_L}} = \frac{1}{1+\frac{R_P + D\cdot R_{ON}}{R_L}}$$

Let's start from the DC voltage transfer function computed including the effect of parasitics, x.
We need a second equation that relates the average input current and the average output current (if ripple negligible or non-negligible but linear), and it is the red box relationship. It holds even if we have parasitisms.

We start from considering the current at the input of the converter, i_in(t).
When the switch is on, the current that flows through the input is equal to the one in the inductor. This because the input is in series with the inductor. During the off time period, the input current is 0.
So we have a periodic trapezoidal waveform and we can compute the average input current, which is by definition the area of the trapezoid divided by the switching period.
Once again, if the current ripple is negligible or the current ripple is linear (red and green area are compensating), we can say that Iin = Io*D.

We multiply equation x with equation y and the result is the output power divided by the input power, so the efficiency.

In the ideal case the efficiency should be 1. If Rp != 0 but Ron = 0, the efficiency is constant but not 1.
If we plot the efficiency of a buck converter as a function of the output current Io, we get something that decreases like a slope. If Io is large, it means that R_L is small, and this makes the efficiency to be small. In principle, at 0 current we would expect an efficiency of 1, but this is not true. In fact, at a given current the efficiency starts decreasing.

This decrease is due to the switching losses, that we have neglected in our analysis so far.

Alternative approach for efficiency

This calculation can be applied **in the case of small ripple**.
By definition, the efficiency is the ratio between the output power and the input power, and the input power is nothing else the output power plus the losses. So if we can write the losses in some ways we can find the expression for the efficiency.

The only losses we are considering are conduction losses, no core losses or switching losses.

$$\eta = \frac{P_O}{P_O + P_{loss}} = \frac{P_O}{P_O + P_L + P_S}$$

Small ripple approx. : $\frac{\Delta i_L}{I_L} \ll 1 \rightarrow i_L(t) \approx I_L = I_O$

where

$$\langle P_S(t) \rangle = P_S = \frac{1}{T_S}\int_0^{T_S} R_{ON} \cdot i_S(t)^2 dt = \frac{1}{T_S} R_{ON} \int_0^{D \cdot T_S} I_O^2 dt = R_{ON} \cdot I_O^2 \cdot D$$

$$\langle P_L(t) \rangle = P_L = \frac{1}{T_S}\int_0^{T_S} R_P \cdot i_L(t)^2 dt \approx \frac{1}{T_S} R_P \int_0^{T_S} I_O^2 dt = R_P \cdot I_O^2$$

and $P_O = \frac{V_O^2}{R_L} = I_O^2 \cdot R_L$

$$\Rightarrow \quad \eta = \frac{1}{1 + \frac{R_P + D \cdot R_{ON}}{R_L}}$$

We have just two elements that are dissipating power during conduction, which are the switch and the inductor, so we have the power dissipated by these two components.
The average power dissipated by conduction is the product of the resistance and the square of the rms value of the current that flows in a resistor. We can compute this for the switch and inductor losses.

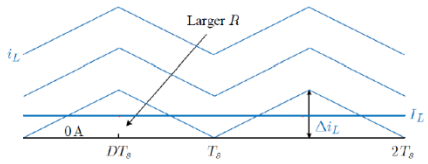
If we apply the small ripple approximation we replace the instantaneous current in the switch with the average current, and since the switch is in series with the inductor, the current that flows in the switch during D*Ts is actually I_L, which is equal to Io. Same reasoning then for P_L.

CCM AND DCM OPERATION

In real converters and especially if the buck converter includes unidirectional switches, the operation of the converter is in DCM. Unidirectional components are components where the current cannot change sign, e.g. the diode.

It is important to understand if the converter is operating in CCM or DCM because the DC voltage transfer function is different in the two cases. Also the dynamic behaviours are different.

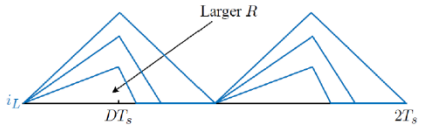
- Use of **unidirectional switch** forces the buck converter to exhibit two different operational modes: the continuous conduction mode and the discontinuous conduction mode.
- Continuous conduction mode (CCM) operation



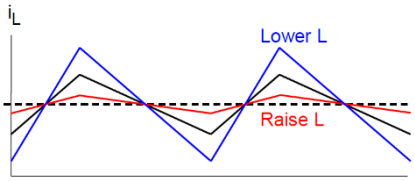
$$I_L = I_O = \frac{V_O}{R}$$

Note: by increasing R (i.e. by lowering I_O) the converter is pushed toward DCM

- Discontinuous conduction mode (DCM) operation. Typically occurs at light load (small load current)



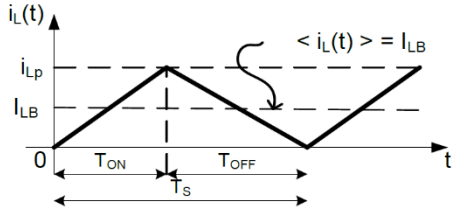
Effect of raising and lowering L while holding Vin, Vo, Io and f constant



- Lowering L increases ΔI and moves the converter toward DCM

This is the explains why a buck converter enters the DCM. In this case, the converter is working with a given average current and I decrease the value of the filtering inductor. The consequence is that the slope of the current during Toff and Ton is increasing, and we might end up in DCM.

BOUNDARY BETWEEN CCM AND DCM



- I_{LB} = critical current below which inductor current becomes discontinuous:

- If $I_L > I_{LB} \rightarrow$ CCM
- If $I_L < I_{LB} \rightarrow$ DCM

$$I_{LB} = \frac{1}{2} i_{Lp} = \frac{1}{2} \frac{V_O t_{off}}{L} = \frac{1}{2} \frac{V_O (1-D) T_s}{L} = \frac{1}{2} \frac{V_O (1-D)}{L f_s}$$

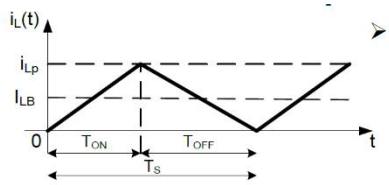
$$\rightarrow \quad I_{LB_{max}} = \frac{V_o}{2 L f_s}$$

The boundary condition is with the current that starts exactly at zero at the beginning of the switching cycle, reaches a peak value and then goes back to 0 exactly at the end of the switching period. The average inductor current in this case is called critical current.

We need an expression for I_LB. I_LB is the average current in the boundary condition, that is the area of the triangle divided by Ts, but the base of the triangle is exactly Ts, so I_LB = ½ * i_Lp.
Now we replace i_Lp with the corresponding value.

Condition for DCM operation

We are working in DCM if condition x is satisfied.



- Boundary if $I_L = I_{LB} = \frac{1}{2}\Delta i_L = \frac{1}{2}\frac{V_O(1-D)T_S}{L}$

Note: $I_L = I_O = \frac{V_O}{R}$

x

- DCM if $I_L < I_{LB} \Rightarrow \frac{V_O}{R} < \frac{1}{2}\frac{V_O(1-D)T_S}{L}$

i.e. $\frac{2L}{RT_S} < (1-D)$

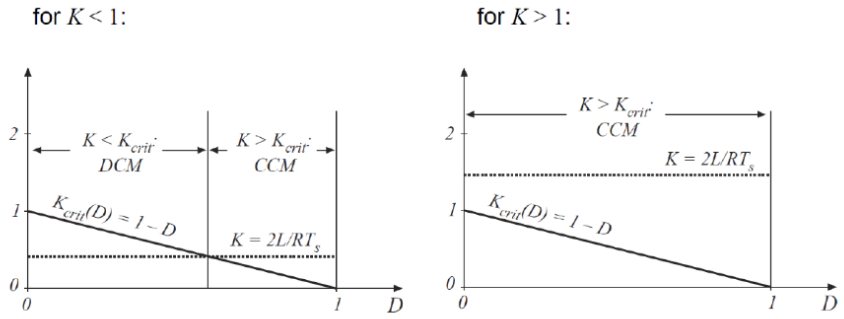
- This can also be expressed as: $K < K_{crit}(D)$ for DCM

where $K = \frac{2L}{RT_S}$ and $K_{crit}(D) = (1-D)$

Remembering that the average inductor current in a buck converter is equal to the average output current, we can rewrite the inequality, replacing I_L with Vo/R.
We can express the inequality using also K, where K is a dimensionless parameter, at a given duty cycle D.

K and Kcrit

K is a measure of the tendency of the converter to work in DCM. The left plot reports the Kcrit in continuous line vs the duty cycle. It is 1 at D = 0 and zero at D = 1. Dashed line is the K. The crossing between the two curves defined the critical duty cycles. If D < Dcrit, the converter works in DCM (K < Kcrit).



- The dimensionless parameter K is a measure of the tendency of a converter to operate in the discontinuous conduction mode.
- Large values of K lead to continuous mode operation, while small values lead to discontinuous mode for some values of duty cycle.
- If K is greater than one, then the converter operates in the continuous conduction mode for all duty cycles.

There is a special case. If the buck converter is designed in a way that K > 1, the converter is always operated in CCM, regardless the operation duty cycle.

Other conditions for DCM operation

The basic starting point is always I_L < I_LB, and then we do all the replacement. Simply by rearranging the same equation we can get two different conditions for the operation of the Buck converter in DCM.

1. **R > Rcrit**, where R is the load resistance. This condition might be useful when the converter is working at a given D, so we can compute the Rcrit and if R > Rcrit we are in DCM.
2. **L < Lcrit**, where L is the filter inductor. This might be useful for all the situations where the load is fixed and also the D. If the load is not resistive, R is the ratio between the output voltage and the absorbed current.

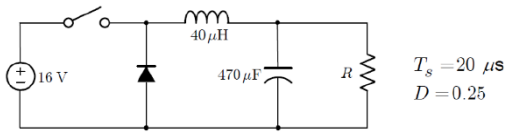
- It is natural to express the mode boundary in terms of the load resistance R or of the filter inductance L , rather than the dimensionless parameter K.

➢ DCM if $I_L < I_{LB} \Rightarrow \frac{V_o}{R} < \frac{1}{2}\frac{V_o (1-D)T_S}{L}$

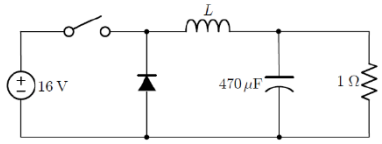
i.e. $\begin{cases} R > R_{crit} = \dfrac{2 \cdot L}{(1-D)T_S} \\ L < L_{crit} = \dfrac{R \cdot (1-D)T_S}{2} \end{cases}$

- These results can be applied to loads that are not pure linear resistors. An effective load resistance R is defined as the ratio of the dc output voltage to the dc load current: R = V/I.

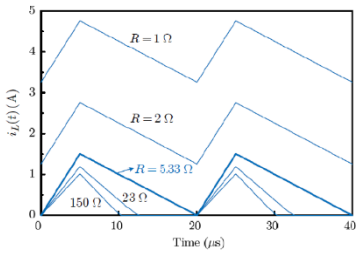
Example



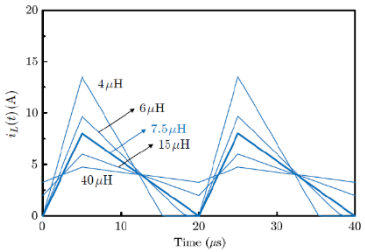
$T_s = 20\ \mu s$
$D = 0.25$



$T_s = 20\ \mu s$
$D = 0.25$



$$R_{crit} = \frac{2L}{(1-D)T_s} = \frac{2 \cdot 40 \times 10^{-6}}{(1-0.25)20 \times 10^{-6}} = 5.33\ \Omega$$



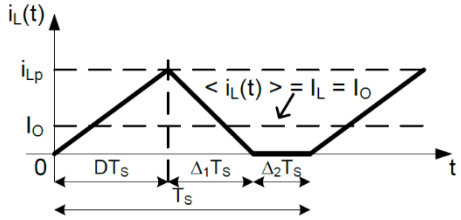
$$L_{crit} = \frac{(1-D)RT_s}{2} = \frac{(1-0.25)1 \cdot 20 \times 10^{-6}}{2} = 7.5\ \mu H$$

We want to see what happens changing the load resistance R. Firstly we need to compute Rcrit, which is 5.33 Ohm. We can also compute Lcrit.

Is there any advantage in working in DCM?

STEADY STATE ANALYSIS IN DCM

The starting point is the current flowing in the inductor in DCM. It starts from 0, reaches a maximum value i_Lp after DTs, so at the end of Ton, then the switch is turned off, the current goes down and reaches zero before the period is over, so for a fraction of the period the current is 0.
We can fractionate the period Ts in three components, and in the last one the current is 0.



Energy stored in L is fully released after $\Delta_1 T_S$

$$(V_{IN} - V_O)DT_S = V_O \Delta_1 T_S \Rightarrow V_O = \frac{D}{D+\Delta_1} V_{IN} \Rightarrow \boxed{\Delta_1 = D\left(\frac{V_{IN}}{V_O} - 1\right)} \quad (1)$$

V·s balance on L

$$I_o = i_{Lp} \frac{(D+\Delta_1)}{2} = \frac{V_O \Delta_1 T_S}{L} \frac{(D+\Delta_1)}{2} \quad (2)$$

A·s balance on C

substitution of (1) into (2) yields

$$\boxed{D = \frac{V_O}{V_{IN}} \sqrt{\frac{K}{1 - \frac{V_O}{V_{IN}}}}}$$

We need to calculate the DC voltage transfer function (also called **conversion ratio**). To do so, we apply two balances: the volt-second balance in L applied between 0 and the end of delta1Ts.
During the Ton, the voltage across the inductor is Vin – Vo and, multiplied by D*Ts is equal to the voltage across the inductor when the inductor is discharging (switch open) multiplied by delta1Ts.

To get rid of the delta1Ts we need a second equation, which comes from the ampere-second balance on the capacitor → average current in the capacitor is 0 and so the average inductor current must be equal to the average output current, this is the balance we are applying.
The average inductor current is, by definition, the area of the triangle divided by Ts. Then we replace i_Lp with: i_Lp = |delta(i_L-)| = Vo*delta1Ts/L.

We can rewrite the expression introducing Mvdc, which is the conversion ratio.

$$D = \frac{V_O}{V_{IN}} \sqrt{\frac{K}{1 - \frac{V_O}{V_{IN}}}} = M_{VDC} \sqrt{\frac{K}{1 - M_{VDC}}} \quad \text{where} \begin{cases} M_{VDC} = \frac{V_O}{V_{IN}} \\ K = \frac{2L}{RT_S} \end{cases}$$

rearranging:

$$\boxed{K M_{VDC}^2 + D^2 M_{VDC} - D^2 = 0}$$ → quadratic equation in M_{VDC}

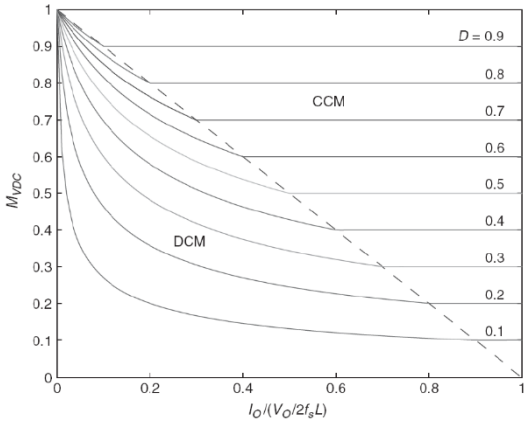
solving for M_{VDC}:

$$\boxed{M_{VDC} = \frac{V_O}{V_{IN}} = \frac{2}{1 + \sqrt{1 + \frac{4K}{D^2}}}}$$

The final result is the one in the red box. It is a much more complex dependance of Vo/Vin on the D, with respect to the CCM operation. Furthermore, the DC voltage t.f. depends on the factor K, so we get

also a dependance on the output current, which internally depends on the output resistance → i/o transfer function depends also on the load resistance.

Mvdc vs K

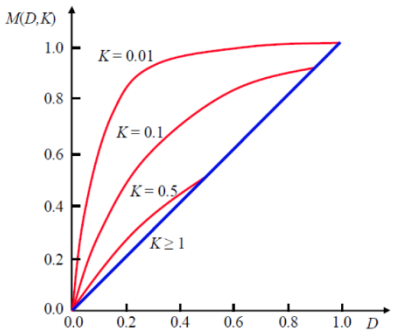


$$M_{VDC} = \frac{V_O}{V_{IN}}$$

$$\frac{I_O}{V_O/2f_sL} = K$$

There is a separation line. Below the line the converter works in DCM and we see the nonlinear dependance of the t.f. on the K and D, while in CCM the curves are flat, so no dependance on K, only on D.

However, it is better to plot the M as a function of the duty cycle. The blue line corresponds to a buck converter working in CCM, where M = D, which is true for K > 1.
If we assume K = 0.1, the device works in CCM only for D > 90%, because for K > Kcrit we are in CCM, and K = 0.1, Kcrit = 1 – D = 0.9.
For smaller D, the converter works in DCM → **the smaller the K, the larger Dcrit**.



$K < K_{crit}(D) = (1-D)$ for DCM

$$\frac{2}{1+\sqrt{1+\frac{4K}{D^2}}} > D$$

M_{VDC} : DCM M_{VDC} : CCM

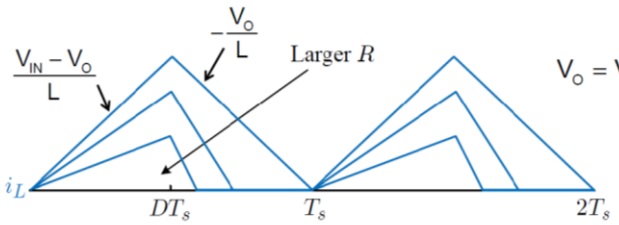
- M_{VDC} in larger in DCM than in CCM for the same duty ratio

If we consider the converter working at a give D, if it works in CCM, we have a certain value for M. However, if we are working in DCM, the conversion ratio M at the same D is larger.

Observation

When we work in CCM and we change the value of R, the characteristic of the current i_L, if R increases, shifts rigidly down. But when we enter in DCM, the slope of the current is reducing both in the on and off time periods.

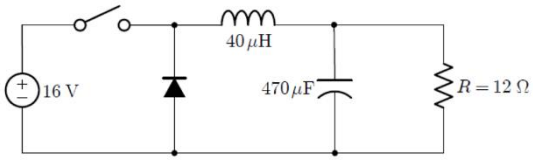
The slope of the current in the on time period, when current is increasing, is (Vin-Vo)/L. When we enter the DCM, the Vo is given by the formula x.
If we are increasing R we are reducing the denominator, so at the same input voltage we are increasing the output voltage, and at the same time we are reducing the slope in the on time period.



$$V_O = V_{IN}\frac{2}{1+\sqrt{1+\frac{8f_SL}{D^2R}}}$$ x

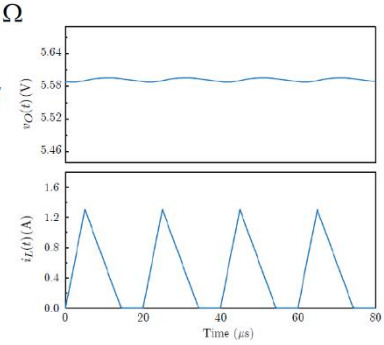
- In DCM operation, the output voltage increases with the load resistance even if the duty ratio remains unchanged.
- A larger load resistance decreases the slope of the on-time inductor current.
- A larger load resistance increases the slope of the off-time inductor current.

Example



D = 0.25, R=12 Ω > R_{crit} = 5.33 Ω

$$V_O|_{DCM} = V_{IN}\frac{2}{1+\sqrt{1+\frac{8f_SL}{D^2R}}} = 5.59V$$

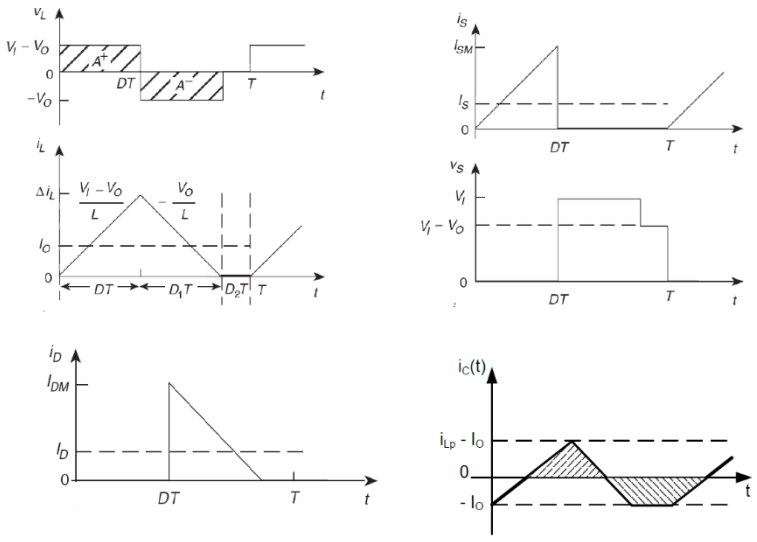


higher than

$$V_O|_{CCM} = V_{IN}\cdot D = 4\,V$$

Ts = 20us. We want to understand if the converter is in CCM or DCM; we compute the critical resistance, and since R > Rcrit, the converter is in DCM.
If the converter was in CCM, Vo = 4V with the same D.

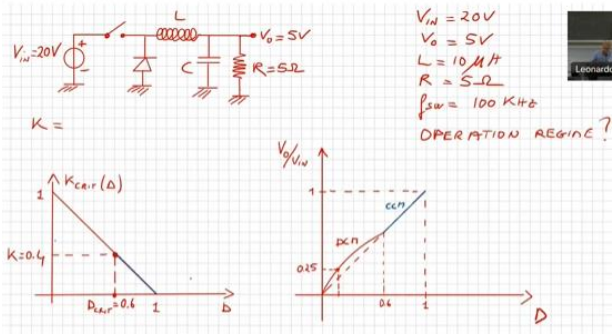
DCM waveforms



The switch voltage v_s, when the switch is on is 0, if the switch is ideal, but when the switch goes off we have two different time periods. v_s = Vin when the diode is on, but when the current reaches 0 and there is no more current inside the inductor, the voltage across the switch collapses down to Vin – Vo.

The last bottom right plot is the current through the capacitor, which is used to calculate the output ripple. To get this current we take the inductor current and we shift it rigidly down by a quantity Io. In the negative phase there is a trapezoid because for a certain period the current is 0. To calculate the capacitor contribution to the ripple we need to compute one of the two grey area.

Example



With this data, without knowing the D, which is the operating point? We just know that the D is operated by some circuit to get 5V at the output starting from 25V at the input.

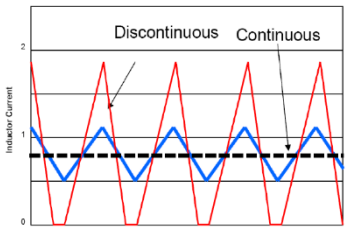
The first thing to compute is Dcrit. Then we compute the K for the device, which is K = 2L/(R*Ts) = 0.4. This done, we take the first plot, which is the line Kcrit = 1- D and we see the Dcrit in correspondence of K = 0.4. For D > Dcrit = 0.6, the converter will work in CCM.

Starting from this information, we use the plot Vo/Vin vs D (on the right). We still don't know D, but we don't care because we know the ratio Vo/Vin, so we know Mvdc = 0.25. We plot it in the plot and see where we are working. We see that we are in DCM. The corresponding D should be 0.18, and we can compute it from the DCM formula for the converter.

$$D = \frac{V_o}{V_{in}}\sqrt{\frac{K}{1-V_o/V_{in}}} = \sqrt{0.182}$$

This duty cycle is set by the feedback circuit, which senses the output voltage, compares it with the desired one, amplifies the error and produces, through the PWM, the desired D.

DISCONTINUOUS (DCM) VS CONTINUOUS (CCM)



Continuous – lower peak currents

- Lower conduction losses
- Smaller input filter required to reduce EMI
- Bigger inductance
- Complex dynamics

Discontinuous – higher peak currents

- Peak and RMS currents are larger in DCM
- Higher conduction losses, bigger core/skin effect losses
- Bigger input filter to reduce EMI
- Smaller inductance
- Simple dynamics

To perform a fair comparison let's start from a simple situation where the output current of the converter is fixed. The black dashed line is the average inductor current, which is equal to the output current.
The blue line represent a situation where the converter works in CCM, while the red in DCM, with the same output current and same converter.
If we are in DCM, to get the same average current we need to push a large current in the inductor, so the peak value of the current in the inductor is bigger than the peak value for the converter working in CCM. The more we push in DCM, the larger the peak current to get the same output current.

So in CCM the current has a lower peak value. Furthermore, the root mean square of the current, Irms of the inductor current, in CCM, is smaller than the rms value in DCM.

$$I_{rms}\Big|_{DCM} = I_{dc}\cdot\sqrt{\frac{4}{3}\cdot\frac{1}{(D+\Delta_1)}}$$

NOTE: $(D+\Delta_1)\leq 1$

Having a lower peak current and lower rms value leads to two big advantages:

1. Less conduction losses, which are due to the current flowing through resistive component, either the Rds(on) of the mosfet or the parasitic resistance of the inductor.
2. Smaller input filter to reduce EMI. We are forced to use them to reduce EMI. The closer we get to the DCM regime, the higher the contribution of the harmonics and in DCM we expect the harmonics to become even stronger. It is a huge disadvantage. The input filter is used before the buck to prevent the injection of the harmonics and of course, the stronger the weights of the harmonics, the larger the filter components.

The disadvantages of the CCM are that if we want to operate in this mode we have to use a larger inductor with respect to the same device working in DCM. Moreover, we have also complex dynamics.

In DCM, peak and rms values of the current are larger, so larger losses and larger input filter. Not only the losses are larger because the rms value is larger, but also because of proximity effect and magnetic core losses (see part 7).
Also ringing is a potential problem of a converter in DCM.

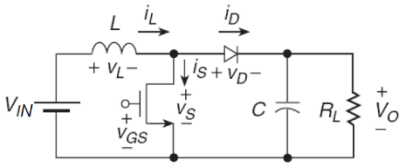
As for the advantages, with the same output current we can afford a smaller inductance, and also the dynamic is simpler, so it is easier to put it into a loop and have good phase margins with large BW (no RHP zero).
Another advantage is the '**no reverse recovery**'. In CCM a current in the off time period is always present in the diode, and when we turn on the switch again we have to turn off the diode reversing bias it. But since there was current flowing in the diode we have to remove the excess recovery charge, and this might be a problem in terms of efficiency. Conversely, in DCM the current in the inductor becomes 0 before the turn off period is over, so there is no current in the diode, hence when the switch is closed again we don't have to remove the reverse recovery charge.

If the output current is not larger than a few amps we can keep the peak current in the inductor to a reasonable level and we can work in DCM, but if we have tens of amps we have to operate in CCM, otherwise we end up with too large current peaks that are not easily manageable.

BOOST CONVERTER

The difference with respect to the buck converter is the position of the components.

CCM ANALYSIS



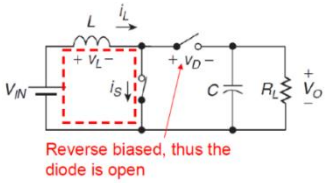
- DC voltage transfer function may be derived from volt-second balance:
  - average voltage across inductor is zero in steady state

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{v_{Lon}t_{on} + v_{Loff}t_{off}}{T_S} = 0$$

All components assumed ideal, we want the conversion ratio in CCM. As usual we apply the volt-second balance on the inductor. In a boost converter the inductor is in series with the input, so the input current is a triangular current in CCM, so there is no need in CCM to provide an input filter, the current is already filtered by the inductor.

Ton

In the on time period the current is flowing in the indicated red loop so V_Lon = Vin.

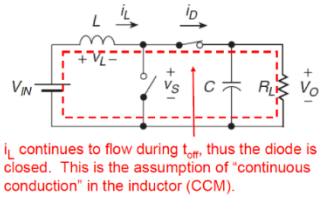


Equivalent circuit when the switch is ON and the diode is OFF

- Switch closed for $t_{on}=DT_s$ seconds
- $v_{Lon} = V_{IN}$

$$v_{Lon}t_{on} = V_{IN}DT_S$$

Toff



Equivalent circuit when the switch is OFF and the diode is ON

- Diode conducting for $t_{off}=T_s- t_{on} =T_s(1-D)$ seconds
- $v_{Loff} = (V_{IN} - V_O)$

$$v_{L_{off}}t_{off} = (V_{IN} - V_O)(1-D)T_S$$

Since we are in CCM there is current flowing in the inductor, the diode is on and the voltage drop across the inductor is Vin – Vo, assuming the diode still ideal with no voltage drop across it.

Transfer characteristic

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{DT_S \cdot V_{in} + (1-D)T_S \cdot (V_{IN} - V_O)}{T_S} = 0$$

$$V_O \cdot (1-D) = V_{IN} + D \cdot V_{IN} - D \cdot V_{IN}$$

The input/output equation becomes $\boxed{\frac{V_O}{V_{IN}} = \frac{1}{1-D}}$

From power balance, $V_{IN} I_{IN} = V_O I_O$, so $\boxed{\frac{I_O}{I_{IN}} = 1-D}$

- The output voltage V_O of the boost converter is always higher than the input voltage V_I .
- Therefore, it is a *step-up* converter.

If there are no losses, we can apply the power balance in the image and derive the relationship between output and input currents.

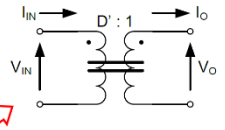
The output voltage in a boost converter is always larger than the input voltage. If D = 0, Vout = Vin.

No-losses case

- input power equals output power:

$$P_{IN} = P_O$$

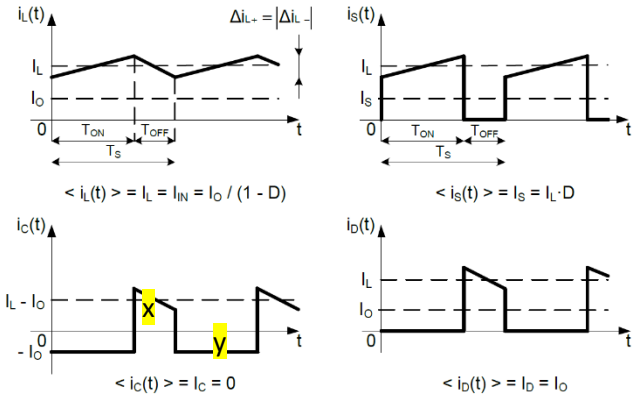
$$V_{IN} I_{IN} = V_O I_O$$



$$\boxed{\frac{I_{IN}}{I_O} = \frac{V_O}{V_{IN}} = \frac{1}{1-D} = \frac{1}{D'}}$$

At steady state the converter can be modelled with the ideal transformer in the image. D' = 1 – D.
It is better not to use this model.

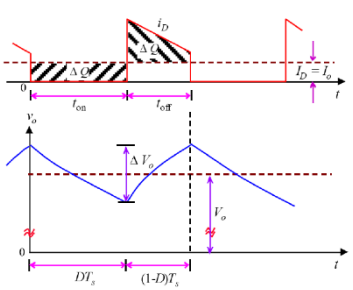
CCM waveforms

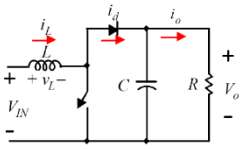


The average current of the diode is Io, so we can easily compute the instantaneous current in the capacitor by taking the diode current and shifting it down by Io. Then area x must be matched by area y.

Moreover, with respect to the buck converter, in a boost converter the inductor current is different with respect to the output current. This is important when calculating what happens in DCM.

Output voltage ripple





$$\Delta V_o = \frac{\Delta Q}{C} = \frac{I_o D T_s}{C} = \frac{V_o}{R}\frac{D T_s}{C}$$

$$\Downarrow$$

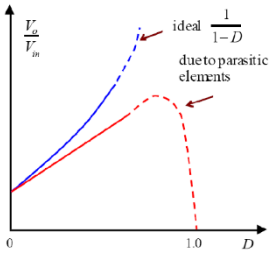
$$\frac{\Delta V_o}{V_o} = \frac{D T_s}{RC}$$

- small ripple assumption
- ESR is assumed to be zero

To compute it we have to consider the current that flows through the capacitor in area x of the previous image (where we are charging the capacitor) and in area y. The charge we are injecting is equal to the charge we are extracting in phase y. We get that the only contribution to the voltage ripple is due to the capacitive contribution.
In the formula, delta_Q is either the discharge or charge one, but obviously the discharge one is easier to be computed (rectangle y area).

With respect to the buck converter, in this case the relative ripple depends on the switching frequency (Ts) in the first order, while in the buck converter the dependance was of the second order (square of the fs).
In general, if we want to keep the output voltage ripple small in a boost converter we need to use a large capacitor with respect to a similar situation in a buck converter.

Parasitic components

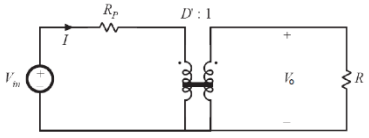


- The duty-ratio is generally limited before the parasitic effects become significant

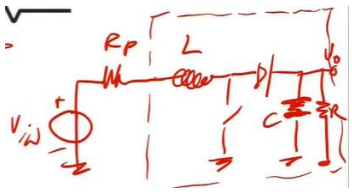
If the D is increased up to 1 I can apparently get an infinite output voltage, but it is not the real case. The output voltage in the real case drops after 80-85% due to parasitic components.

Example of the effect of parasitics: inductor resistance

- Current is continuous on R_P
- DC equivalent circuit can be used



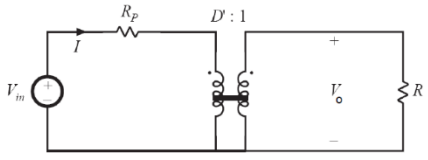
There are also other parasitics like the Rds(on) of the mosfet, the voltage drop across the diode that we can include. Here we are considering a boost converter and the parasitic resistance of an inductor, and we are assuming that the parasitic resistance of the inductor is in series with it.



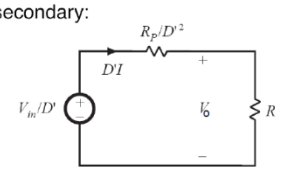
We can isolate the ideal boost converter and keep the parasitic resistance outside it. If we do so, we can easily take advantage of the transformer model seen for the boost converter. However, this model is valid only for ideal boost converters working at steady state (DC). However, since we are considering the parasitic resistance Rp outside the boost converter, we can do so.
R is the load resistance, and we can analyze the i/o transfer function at steady state by using this equivalent circuit.

We can move the resistance from the primary side of the transformer to the secondary side. If the turn ratio is n1:n2 for the transformer, to move the resistance we have to multiply the resistance by the (n2/n1)^2. Now n2 = 1, n1 = D' and so we get that to move the resistance it will be Rp/D'^2.
Also the input voltage generator has to be moved from the primary side to the secondary side, and this is just the multiplication by the turn ratio.

Converter equivalent circuit



Refer all elements to transformer secondary:

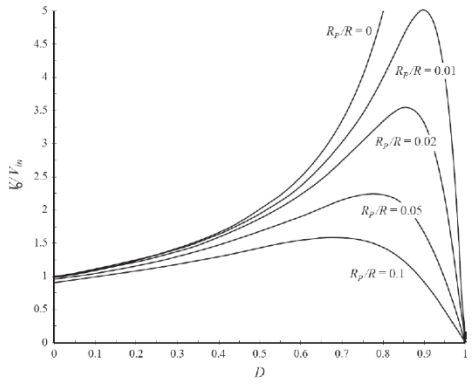


Solution for output voltage using voltage divider formula:

$$V_o = \frac{V_{in}}{D'} \frac{R}{R + \frac{R_P}{D'^2}} = \frac{V_{in}}{D'} \frac{1}{1 + \frac{R_P}{D'^2 R}}$$

$$\frac{V_O}{V_{IN}} = \frac{1}{D'} \frac{1}{1 + \frac{R_P}{D'^2 R}}$$

In the end we get a simple voltage divider.
In the result in the blue box we can find 1/D' that is the ideal DC voltage transfer function multiplied by a correction term due to the parasitic component. We can plot this function as a function of the duty cycle as below.

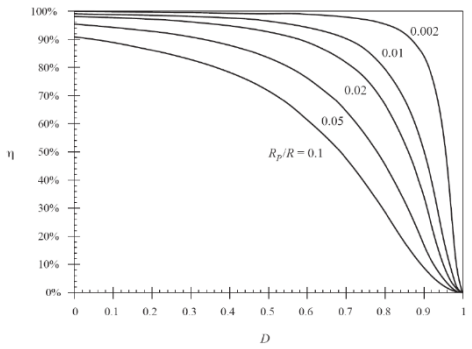


As expected, the DC voltage transfer function is peaking somewhere depending on the ration between Rp and R. For Rp/R = 0.1, the peak occurs around 0.85. The larger the value of the Rp, the smaller the value of the peak and the lower the duty cycle at which the peak is occurring. **It is impossible with a boost converter to get a DC voltage transfer function larger than 4 or 5**.
So to get an output voltage that is larger than the input voltage by a factor greater than 5 this is not a good conversion topology, we need quadratic boost converters.

Use of the transformer model

We can use this model because we are taking the Rp outside the ideal boost converter, but if we want to consider the Rds(on) and the voltage drop across the diode, we cannot use the transformer model.
If we cannot use it, the workaround is to apply the volt-second balance across the ideal inductor. Of course we may check if the small current ripple approximation (or the linear ripple one) can be applied.

Efficiency with parasitic components

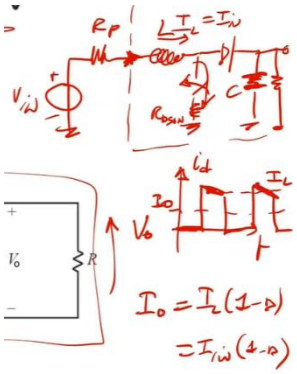


$$I_O = I_{IN}(1-D) = I_{IN}D' \quad \Rightarrow \quad \frac{V_O I_O}{V_{IN} I_{IN}} = \eta = \frac{1}{1+\frac{R_P}{D'^2 R}}$$ x

The parasitic component is dissipating power because the current is continuous over the parasitic resistance of the inductor, and the power dissipation by conduction is the product between Rp and the root mean square value of the current. We expect a reduction of the efficiency with respect to the ideal value.

By definition, the efficiency is the ratio between the input power and the output power VoIo/VinIin. The ratio Vo/Vin is known also with the parasitic resistance, while for Io/In we can apply the reasoning used with the buck converter; the presence of parasitic components doesn't affect the ratio between currents, which is Io/In = D' = 1 – D.
This comes from the fact that I_L = Iin because the inductor is in series with the input. When the switch is closed, the current flows through the switch, so there is no current in the diode. When the switch is

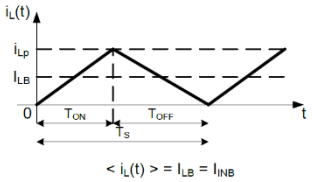
open, I_L flows through the diode in the boost converter. Hence the instantaneous diode current i_d is 0 when the switch is closed, equal to the inductor current when the switch is open.



In the end, the larger Rp with respect to the load resistance, the smaller the efficiency at a given D.

Moreover, in a real converter the efficiency reaches a peak close to Io = 0 and then it drops. This because with previous formula x we are just considering the conduction losses, not the switching losses, which reduce the efficiency at low output current.

BOOST CONVERTER – BOUNDARY BETWEEN CCM AND DCM



Note:
- $I_O \neq I_L$

$$I_{LB} = \frac{1}{2}\Delta i_L = \frac{1}{2} i_{Lp} = \frac{1}{2}\frac{V_{IN}}{L} DT_S \quad \Rightarrow \begin{cases} \text{If } I_L > I_{LB} & \rightarrow \text{CCM} \\ \text{If } I_L < I_{LB} & \rightarrow \text{DCM} \end{cases}$$

using

$$\frac{V_O}{V_{IN}} = \frac{1}{1-D}$$

$$\boxed{I_{LB} = \frac{T_s V_O}{2L} D(1-D)} \qquad \text{note: } I_{LB_{max}} = \frac{V_o}{8Lf_s}$$

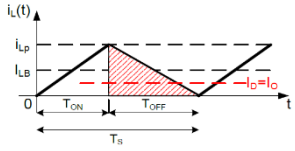
At the beginning of the switching cycle we notice the current is 0, it reaches a peak value and after Toff it returns to 0, so we are at the boundary.

The main difference with respect to the buck converter is that for boost converter the output current is different from the inductor current, and it must be considered when computing the DC voltage transfer function in DCM.
The boundary happens when the current starts from 0 and reaches 0 at the end of the switching period. The corresponding average current is called boundary current I_LB.

The average current is the area of the triangle, si i_Lp multiplied by the base of the triangle divided by 2. i_Lp is delta(i_L), so peak to peak current ripple, which can be written as Vin*D*Ts/L.
It is better to express the boundary current as a function of the output voltage, not a function of the Vin, because in general the converter is a regulator and so the output voltage is fixed. To do so, we use the relationship Vo/Vin = 1/(1-D), which is still valid at the boundary CCM – DCM.

The final boundary current is the one in the red box. We can notice that **the maximum boundary current happens for D = 0.5**.
If the inductor current is larger than the boundary current, the converter is operated in CCM. Viceversa, it is operated in DCM.

Condition for DCM operation



x

- Boundary if $I_L = I_{LB} = \frac{1}{2}\Delta i_L = \frac{V_O}{2L}D(1-D)T_S$

Note: $I_L = \frac{I_O}{1-D}$

- DCM if $I_L < I_{LB} \Rightarrow \frac{V_O}{R(1-D)} < \frac{V_O}{2L}D(1-D)T_S$

i.e. $\frac{2 \cdot L}{RT_S} < D(1-D)^2$

- This can also be expressed as: $K < K_{crit}(D)$ for DCM

where $K = \frac{2L}{RT_S}$ and $K_{crit}(D) = D(1-D)^2$

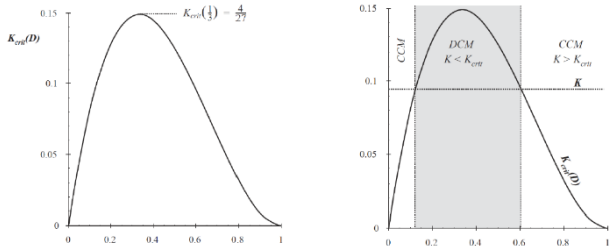
We start from the expression of the boundary current x. Then the converter is working in DCM if I_L < I_LB. We can replace I_L with something that depends on the output current. In fact, I_L = Io/(1-D). Then Io = Vo/R. These relationships can be plugged in the inequality and we get the formula in the blue box.

It is a condition very similar to the one obtained for the buck converter. The difference is that the Kcrit for the buck is simply 1 – D, while here it is different. K is a dimensionless parameter.

K and Kcrit vs D

- The conditions for operation in the continuous or discontinuous conduction modes are of similar form to those for the buck converter; however, the critical value is a different function of the duty cycle D.

$K_{crit}(D) = D(1-D)^2$ → maximum K_{crit} = 4/27 at D=1/3



The maximum occurs for D = 1/3, and the maximum value for Kcrit is 4/27.
To understand if the converter works in CCM or DCM we need to draw an horizontal line corresponding to the K factor of the converter, the line crosses the Kcrit curve in 2 points defining 3 regions. A first window has K > Kcrit, so we are in CCM, a second one where still K > Kcrit and an intermediate one where Kcrit > K, and we are in DCM. So we have two values of Kcrit defining a range of duty cycles, while in the buck converter we had just one value.

Condition for DCM operation – alternative definitions

The starting equation is always the same. Then, instead of solving the inequality for the K factor we can solve it for the resistance R or for the inductor.

The inequality for the resistance can be useful when D and L are fixed and we want to understand whether the boost converter is converter is working in DCM or CCM depending on the output resistance which,

in turn, depends on the output current (Io = Vo/R). Even if the load is not resistive, we can replace R with the voltage divided by current absorbed by the load-

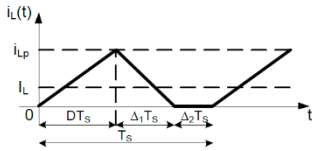
- It is natural to express the mode boundary in terms of the load resistance R or of the filter inductance L , rather than the dimensionless parameter K.

➢ DCM if $I_L < I_{LB} \Rightarrow \frac{V_O}{R(1-D)} < \frac{V_O}{2L} D(1-D)T_S$

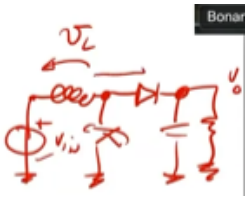
i.e. $\begin{cases} R > R_{crit} = \dfrac{2 \cdot L}{D(1-D)^2 T_S} \\ L < L_{crit} = \dfrac{RD(1-D)^2 T_S}{2} \end{cases}$

- These results can be applied to loads that are not pure linear resistors. An effective load resistance R is defined as the ratio of the dc output voltage to the dc load current: R = V/I.

STEADY STATE ANALYSIS IN DCM OF THE BOOST CONVERTER



Note:
- $I_O \neq I_L$



$V_{IN} D T_S + (V_{IN} - V_O)\Delta_1 T_S = 0 \rightarrow \frac{V_O}{V_{IN}} = \frac{\Delta_1 + D}{\Delta_1}$

V·s balance on L

Solving for Δ_1:

$$\Delta_1 = \frac{D}{\frac{V_O}{V_{IN}} - 1}$$

We focus on the inductor current, which has a time period where it is 0 (DCM is possible because we have a unidirectional device in our converter, the diode).
To compute the DC voltage transfer function we apply the volt second balance on the inductor only when the current is different from 0, and the ampere second balance on the capacitor.
The result of the volt second balance is delta 1, which we don't know, and we need to use the other balance to find it.

$i_{Lp} = \frac{V_{IN}}{L} DT_S$

and

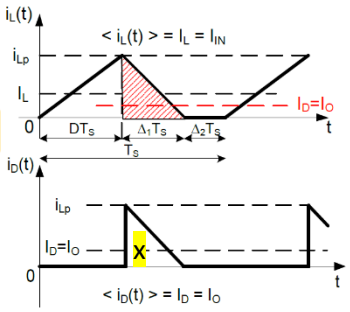
$I_D = I_O = i_{Lp} \frac{\Delta_1}{2}$

A·s balance on C

Substituting i_{Lp} and Δ_1

$I_D = I_O = \frac{V_{IN} T_S}{2L} \frac{D^2}{\frac{V_O}{V_{IN}} - 1}$

Replacing I_O with V_O/R and $2L/RT_S$ with K we obtain



$$D = \sqrt{K \frac{V_O}{V_{IN}} \left(\frac{V_O}{V_{IN}} - 1 \right)}$$

The ampere second balance implies that the average diode current is equal to the average output current. The diode current is not equal to the inductor current always, because it flows only in the discharge period in the diode. The average diode current is the area of triangle x divided by the switching period.

The final equation in the red box provides the D, but it is not very useful because the want the DC voltage t.f. as a function of D. So we replace Vo/Vin = Mvdc, we square the equation and we rearrange it solving the equation for Mvdc.

$$D = \sqrt{K\frac{V_O}{V_{IN}}\left(\frac{V_O}{V_{IN}} - 1\right)} = \sqrt{K M_{VDC}(M_{VDC} - 1)} \qquad \text{where} \begin{cases} M_{VDC} = \dfrac{V_O}{V_{IN}} \\ K = \dfrac{2L}{RT_S} < K_{crit} \end{cases}$$

rearranging:

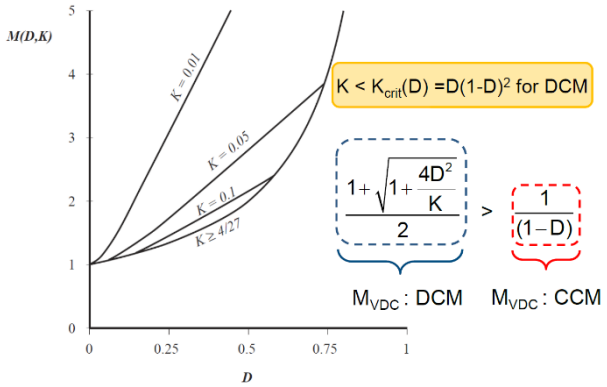
$$K M_{VDC}^2 - K M_{VDC} - D^2 = 0$$ → quadratic equation in M_{VDC}

solving for M_{VDC}:

$$M_{VDC} = \frac{1 + \sqrt{1 + \dfrac{2D^2 R}{f_S L}}}{2} = \frac{1 + \sqrt{1 + \dfrac{4D^2}{K}}}{2}$$

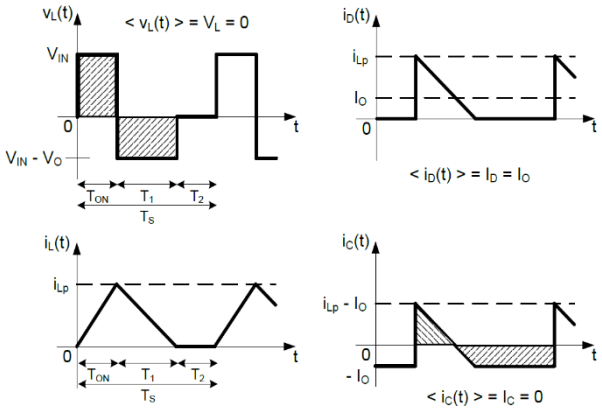
The Mvdc in DCM has a complex dependance on the D and also depends on the factor K (i.e. on the output resistance and in turn also on the current provided to the load), even if the converter is perfectly ideal.

Mvdc vs D



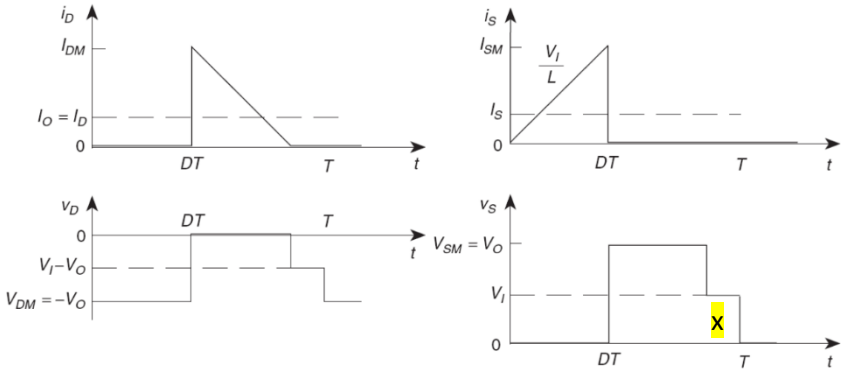
For a given D, Mvdc in DCM is larger than Mvdc in CCM.

DCM waveforms



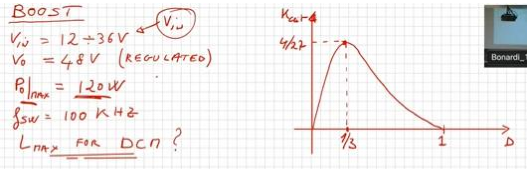
To compute the instantaneous current flowing in the capacitor, ic(t), we simply move the current id(t) down by Io and there must be a balance of areas, because <ic(t)> = 0 at steady state.

The v_s curve indicates the voltage across the switch; when the switch is closed, the voltage is 0, when the switch is open the voltage is equal to Vo in the first time period, because there is still current in the inductor making the diode becoming on and a short circuit. Then when the current in the inductor reaches 0, there is no more current in the diode and the voltage v_s collapses to Vin, because if there is no current V_L = 0. In reality, in the second phase x we should have some ringing due to parasitic capacitances and the resonance between the inductance and the parasitic capacitances.



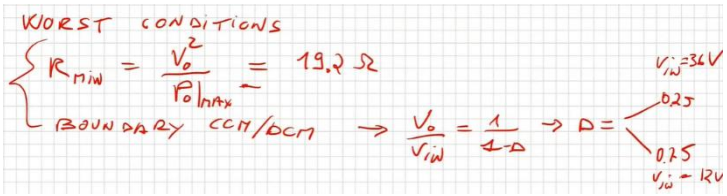
Example on the use of the Kcrit curve – 1

Regulated means that there is a feedback circuit that keeps the output voltage to this value. We need to



understand which are the worst conditions. As far as the output power is concerned, the worst concerned, the worst condition is the maximum output power, which means minimum output resistance and maximum output current. If the converter is working in DCM for the maximum power it will be working in DCM for any power smaller than the maximum power.

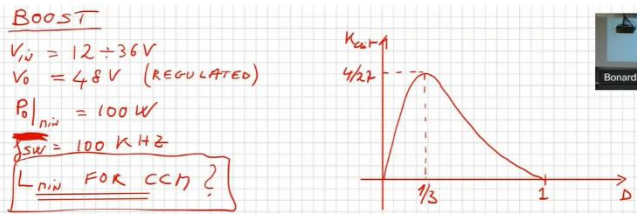
As for the resistance, the worst condition is the minimum one, as said. We want that at the worst condition the converter is operating at the boundary between DCM and CCM, so that it will be working in DCM for any other condition. At the boundary, Vo/Vin = 1/(1-D) is still valid.



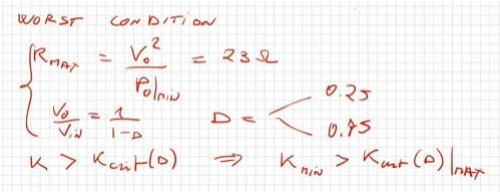
So the duty cycle ranges from 0.25 to 0.75, and I put these values in the curve. When K < Kcrit, the converter is in DCM. This inequality must be considered at the worst case, i.e. Kmax < Kcrit(D),min. If this inequality is satisfied, the device is working in DCM in any condition.
Kmax = 2L/(Rmin*Ts).
Kcrit,min = 0.75*(1 – 0.75)^2, it happens at the maximum D.

The result is L < 4.5uH, for this values the converter is operating in DCM.

Example on the use of the Kcrit curve – 2



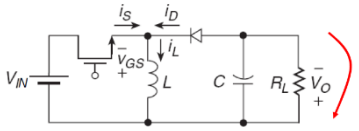
Again, we have to understand the worst condition for CCM, which occurs for the minimum output power, because if we increase it the resistance is going to reduce and therefore the converter is going to work in CCM. Again, we want to put ourself at the boundary between CCM and DCM at the worst condition.
The condition for CCM is K > Kcrit, but K is not fixed, it varies, so we rewrite it in the worst condition. So Kmin > Kcrit(D),max.



The maximum for Kcrit in this case, between 0.25 < D < 0.75, occurs for D = 1/3 and Kcrit = 4.27

BUCK-BOOST CONVERTER

CCM ANALYSIS



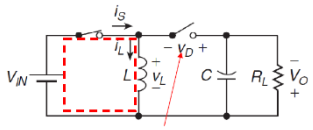
- DC voltage transfer function may be derived from volt-second balance:
  - average voltage across inductor is zero in steady state

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{v_{Lon}t_{on} + v_{Loff}t_{off}}{T_S} = 0$$

The BB converters are able to produce an output voltage that is reversed in sign with respect to the input voltage. In order not to bring the minus sign in all the calculations, the output voltage is taken positive in the opposite direction with respect to the buck and boost converters.

As in a boost converter, in a BB converter the input current is pulsed, because when the switch is off there is no current in input, while when the switch is on the current is I_L. This in general it requires an input filter.

Ton

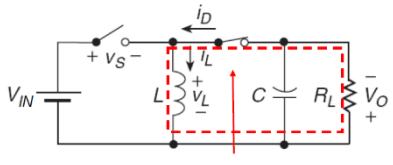


Equivalent circuit when the switch is ON and the diode is OFF

- Switch closed for $t_{on}=DT_s$ seconds
- $v_{Lon} = V_{IN}$

$$v_{Lon}t_{on} = V_{IN}DT_S$$

Toff



Equivalent circuit when the switch is OFF and the diode is ON

- Diode conducting for $t_{off}=T_s- t_{on} =T_s(1-D)$ seconds
- $v_{Loff} = -V_O$

$$v_{L_{off}}t_{off} = -V_O(1-D)T_S$$

Transfer characteristic

$$V_L = \frac{1}{T_S}\int_0^{T_s} v(t)dt = \frac{DT_S \cdot V_{in} - (1-D)T_S \cdot V_O}{T_S} = 0$$

$$V_O \cdot (1-D) = V_{IN} D$$

The input/output equation becomes $\frac{V_O}{V_{IN}} = \frac{D}{1-D} = \frac{D}{D'}$

From power balance, $V_{IN} I_{IN} = V_O I_O$, so $\frac{I_O}{I_{IN}} = \frac{1-D}{D}$

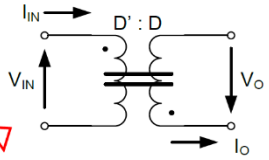
- The magnitude of V_O can be higher or lower than that of V_{IN} depending on the value of D
- The sign of V_O is reversed with respect to the sign of V_{IN}

No losses

- input power equals output power:

$$P_{IN} = P_O$$

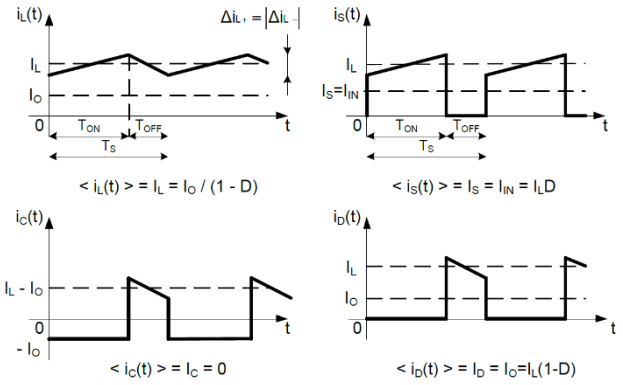
$$V_{IN} I_{IN} = V_O I_O$$



$$\frac{I_{IN}}{I_O} = \frac{V_O}{V_{IN}} = \frac{D}{1-D} = \frac{D}{D'}$$

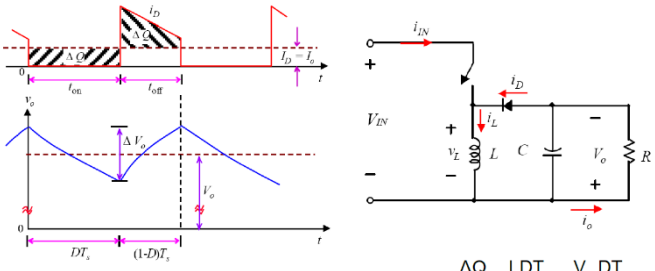
The dot convention highlights that the two are in the opposite positions, meaning that the output voltage is reversed.

CCM waveforms



Output voltage ripple

The ripple is quite similar to the one in the boost converter.



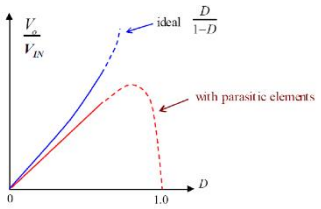
$$\Delta V_o = \frac{\Delta Q}{C} = \frac{I_o DT_s}{C} = \frac{V_o}{R}\frac{DT_s}{C}$$

$$\Downarrow$$

$$\frac{\Delta V_o}{V_o} = \frac{DT_s}{RC}$$

- small ripple assumption
- ESR is assumed to be zero

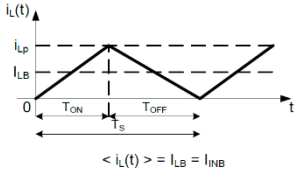
Parasitic components



- The duty-ratio is limited to avoid these parasitic effects from becoming significant

If we consider the ideal characteristic, and we consider the point D = 0.5, if the converter is working with D < 0.5, Vo/Vin is smaller than 1. While if D > 0.5, Vo/Vin > 1. So not only it can reverse the output voltage, but in absolute value it can be larger or smaller than the input voltage depending on the value od the duty cycle. It is similar to the SEPIC converter, but the SEPIC is not inverting.

BOUNDARY BETWEEN CCM AND DCM



$< i_L(t) > = I_{LB} = I_{INB}$

Note:
- $I_O \neq I_L$

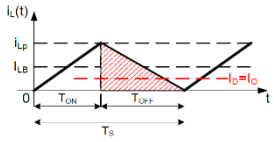
$$I_{LB} = \frac{1}{2}\Delta i_L = \frac{1}{2} i_{Lp} = \frac{1}{2}\frac{V_{IN}}{L} DT_S \Rightarrow \begin{cases} \text{If } I_L > I_{LB} \rightarrow \text{CCM} \\ \text{If } I_L < I_{LB} \rightarrow \text{DCM} \end{cases}$$

using

$$\frac{V_O}{V_{IN}} = \frac{D}{1-D}$$

$$\boxed{I_{LB} = \frac{T_s V_O}{2L}(1-D)} \qquad \text{note: } I_{LB_{max}} = \frac{V_o}{2Lf_s}$$

Condition for DCM operation



➢ Boundary if $I_L = I_{LB} = \frac{1}{2}\Delta i_L$

Note: $I_L = \frac{I_O}{1-D}$

➢ DCM if $I_L < I_{LB} \Rightarrow \frac{V_O}{R(1-D)} < \frac{V_O}{2L}(1-D)T_S$

i.e. $\frac{2\cdot L}{RT_S} < (1-D)^2$

➢ This can also be expressed as: $K < K_{crit}(D)$ for DCM

where $K = \frac{2L}{RT_S}$ and $K_{crit}(D) = (1-D)^2$

Other definitions for DCM

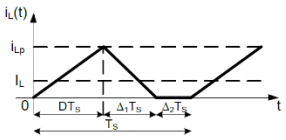
- It is natural to express the mode boundary in terms of the load resistance R or of the filter inductance L , rather than the dimensionless parameter K.

➢ DCM if $I_L < I_{LB} \Rightarrow \frac{V_O}{R(1-D)} < \frac{1}{2}I_{LB} = \frac{V_O}{2L}(1-D)T_S$

i.e. $\begin{cases} R > R_{crit} = \dfrac{2\cdot L}{(1-D)^2 T_S} \\ L < L_{crit} = \dfrac{R\,(1-D)^2 T_S}{2} \end{cases}$

- These results can be applied to loads that are not pure linear resistors. An effective load resistance R is defined as the ratio of the dc output voltage to the dc load current: R = V/I.

STEADY STATE ANALYSIS IN DCM



Note:
- $I_{OB} \neq I_{LB}$

$V_{IN} D T_S - V_O \Delta_1 T_S = 0 \rightarrow \frac{V_O}{V_{IN}} = \frac{D}{\Delta_1}$

V·s balance on L

Solving for Δ_1:

$\Delta_1 = D\frac{V_{IN}}{V_O}$

$i_{Lp} = \frac{V_{IN}}{L} DT_S$

and

$I_D = I_O = i_{Lp}\frac{\Delta_1}{2}$

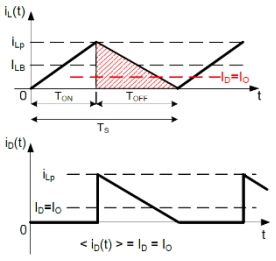
A·s balance on C

Substituting i_{Lp} and Δ_1

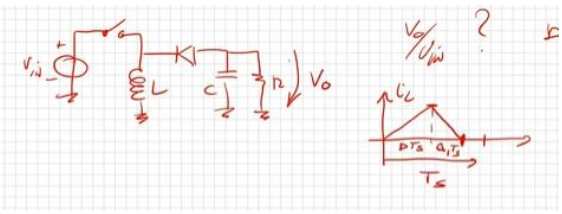
$I_D = I_O = \frac{V_{IN}^2 T_S}{2L}\frac{D^2}{V_O}$

Replacing I_O with V_O/R and $2L/RT_S$ with K we obtain

$\frac{V_O}{V_{IN}} = \frac{D}{\sqrt{K}}$



Let's compute the DC voltage transfer function in DCM for a BB converter. We can apply an energy balance or power balance; in a converter in DCM, the current in the inductor increases when the switch is on until we reach a peak. When the switch is off the current decreases and reaches zero before the period is over.



If we look at the schematic, initially there is no energy stored in the inductor, I close the switch and the energy starts to build up in the inductor because the current is flowing. The final energy stored in the inductor is ½ * L * i_Lp^2.
When the switch is open, the current flows in the opposite direction in the inductor following the direction of the diode until there is no more current in the inductor.

From an energy point of view, the energy that initially was stored in the inductor is fully released to the load. So in a BB converter in DCM we are alternatively charging the inductor and discharging it.
This occurs once per cycle, so if we multiply the energy by the switching frequency and we assume that there are no losses we get the output power.

$$\frac{1}{2} L i_{Lp}^2 f_{sw} = P_o$$

The energy that is stored in the inductor is transferred to the output.
Since i_Lp = Vin*D*Ts/L, we substitute in the Po formula and we get the Vo/Vin.

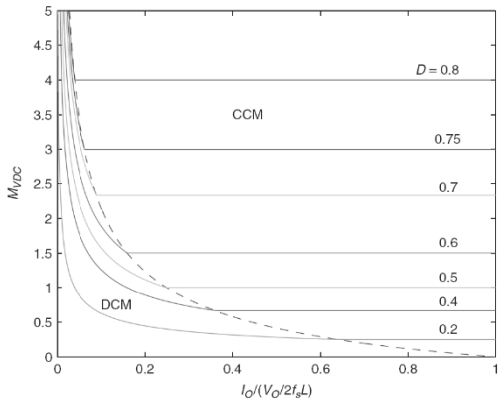
$$\begin{cases} \frac{1}{2} L i_{Lp}^2 f_{sw} = P_o = \frac{V_o^2}{R} \\ i_{Lp} = \frac{V_{in} D T_s}{L} \end{cases}$$

$$\frac{1}{2} L \frac{V_{in}^2 D^2 T_s^2}{L^2} f_{sw} = V_o^2/R$$

$$V_o/V_{in} = D\sqrt{\frac{R T_s}{2L}} = \frac{D}{\sqrt{K}}$$

The same result will be obtained in the case of the Flyback converter, which is nothing else than the isolated version of the BB converter.

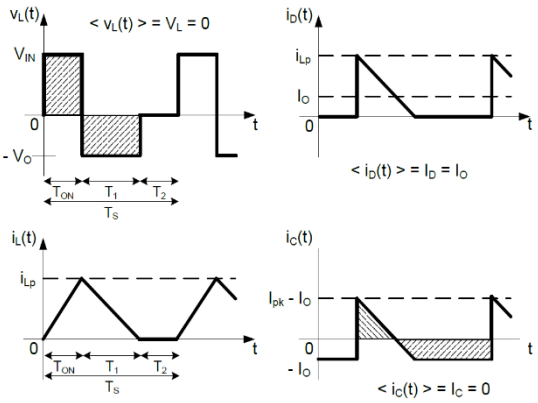
Mvdc vs K



$$M_{VDC} = \frac{V_O}{V_{IN}}$$

$$\frac{I_O}{V_O/2f_sL} = K$$

Waveforms



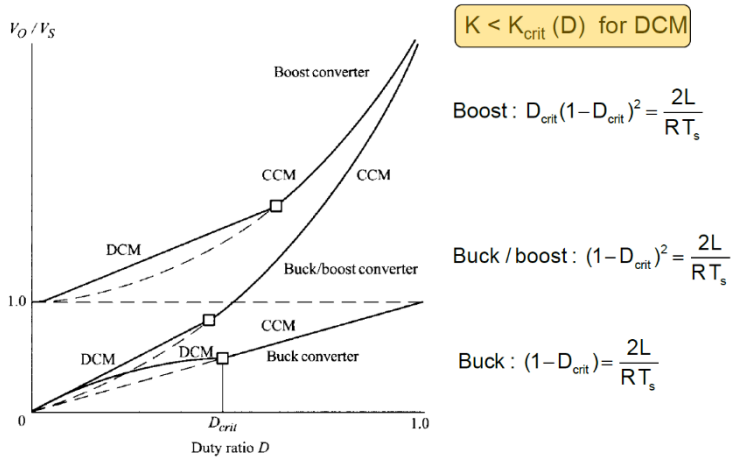
Summary of CCM – DCM characteristics

Table 5.2. Summary of CCM-DCM characteristics for the buck, boost, and buck-boost converters

Converter	$K_{crit}(D)$	*DCM* $M(D,K)$	*DCM* $D_2(D,K)$	*CCM* $M(D)$
Buck	$(1-D)$	$\frac{2}{1+\sqrt{1+4K/D^2}}$	$\frac{K}{D}M(D,K)$	D
Boost	$D(1-D)^2$	$\frac{1+\sqrt{1+4D^2/K}}{2}$	$\frac{K}{D}M(D,K)$	$\frac{1}{1-D}$
Buck-boost	$(1-D)^2$	$\frac{D}{\sqrt{K}}$	$\sqrt{K}$	$\frac{D}{1-D}$

with $K = 2L/RT_s$. *DCM occurs for* $K < K_{crit}$.

In the next plot, the characteristics of the three converters are displayed. Above the Dcrit the converter is in DCM. The characteristics are not linear.
The only peculiarity of the boost converter is that we have two values of Dcrit that identify where the converter is working in DCM.
We also notice that for the BB converter the characteristic is linear because in DCM it the t.f. is directly proportional to D.



See 'optional reading material' for further optional notes on the buck-boost converter.

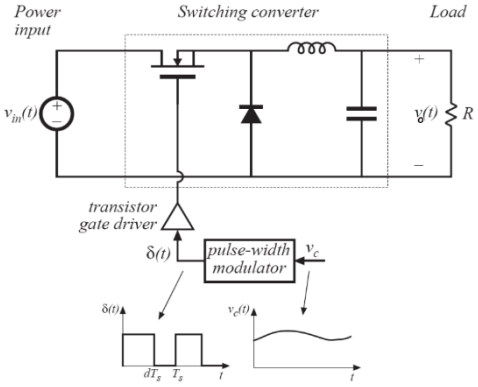
VOLTAGE REGULATORS

So far we have considered converters, which are operated in OL, without a feedback. The problem is that if the input voltage changes, also the output voltage changes because we don't have any feedback. Moreover, if the load current changes, the output voltage changes, so feedback is required.

A DC/DC regulator is nothing else than a DC/DC converter with a negative feedback.

There are disturbances:
- in $v_{in}(t)$
- in R

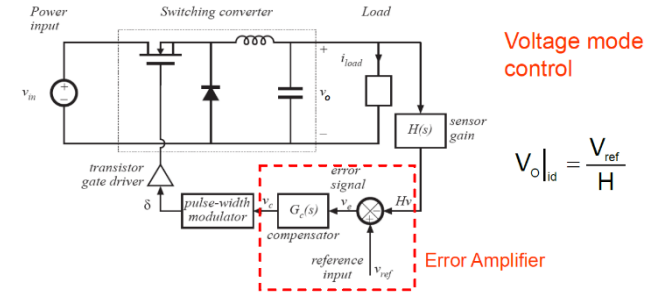
There are uncertainties:
- in component values
- in V_{IN}
- in R



In open loop, the gate controller is controlled with a PWM. If the converter works at steady state in CCM, Vo = Vin*D, where D = Vc/Vst (Vst is the sawtooth voltage).

The problem of the converter operated in open loop is that if we have a perturbation or disturbance on the input voltage, it is presented also on the output voltage, which cannot remain stable. The same happens if the load is not constant and we consider the parasitic components. In fact, the DC voltage transfer function depends on the load current and load resistance, so if we change the load we have a variation or fluctuation of the output voltage.
In general, we want a stable regulated output voltage.

ADDITION OF A FEEDBACK LOOP



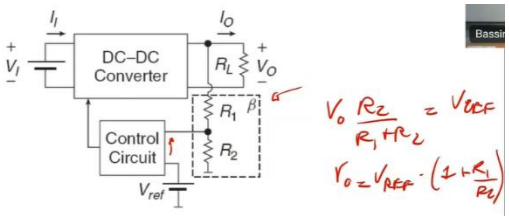
Negative feedback: build a circuit that automatically adjusts the duty cycle to obtain the specified output voltage, regardless of disturbances or component tolerances.

We measure the output voltage, that is the variable we want to control, this variable is multiplied by sensor gain, this output is compared with a reference voltage, the error is then amplified by an error amplifier compensator, and the output of the error amplifier will be connected at the input of the PWM to close the loop.

This simple control system is called voltage mode controller, because we are controlling the output voltage of the DC/DC converter. There are also more complex control system such as the current control, where we are controlling at the same time the current in the inductor and the output voltage.

When we use this simple controller, if we assume the loop t.f. is stable and the LG at 0Hz is infinite. In this case, the error signal tends to 0 making Vref = H*Vo, where Vref is a constant and H is the DC sensor gain.

Negative feedback



- The control circuit compares the feedback voltage with the reference voltage, generates an error voltage, amplifies it, and adjusts the duty cycle to keep the output voltage V_O constant.
- The load current I_O may vary over a very wide range: $I_{Omin} \le I_O \le I_{Omax}$. Consequently, the load resistance $R_L = V_O/I_O$ also varies over a wide range: $R_{Lmin} \le R_L \le R_{lmax}$.
- The input voltage of a voltage regulator is usually unregulated and can vary over a wide range: $V_{Imin} \le V_I \le V_{Imax}$. For example, the dc input voltage in telecommunications power supplies is $36 \le V_I \le 72V$ with a nominal input voltage $V_{Inom} = 48$ V.

We use a voltage divider because in general the Vref is generated inside the control chip by using a bandgap voltage generator, which produces a reference voltage typically around 1.2V, and this voltage is independent on voltage and power supply. So if we want higher voltage we need R1 and R2 to increase it.

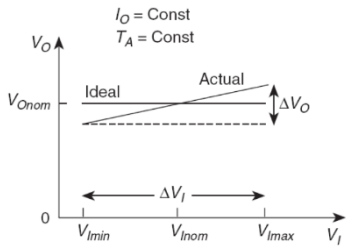
STATIC CHARACTERISTICS OF VOLTAGE REGULATORS

There are static parameters and dynamic ones to understand how good a regulator is. The static parameters are two: line regulation and load regulation.

Line regulation

It is a measure of the regulator's ability to maintain the prescribed voltage despite the variations of the input voltage.

- **Line regulation**, which is a measure of the regulator's ability to maintain the predescribed nominal output voltage V_{Onom} under slowly varying input voltage conditions.



$$\%LNR = \left.\frac{V_{O(max)} - V_{O(min)}}{V_{Onom}}\right|_{I_O=const} \cdot 100\%$$

An ideal regulator would show an output voltage Vo perfectly constant, independent on Vin. In real life, in general when we increase the Vin, even if the system has a feedback, the output voltage slightly increases. For the line regulation, we have the input voltage in a certain range, and in this range we measure the minimum and maximum output voltages. The percentage line regulation is as in the formula.

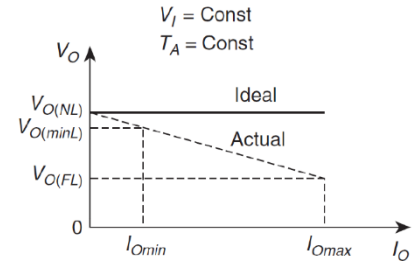
This measurement is performed keeping the output current constant. The smaller the percentage of line regulation, the better the performance of the regulator.

To get a good line regulation we need a stable feedback loop and high loop gain in DC, in principle infinite.

Load regulation

It is a measure of the regulator's ability to maintain a constant output voltage (around the nominal one) in case of slowly varying load conditions. In reality, there is a drop of the output voltage with the increasing output current.
To define %LOR we change the output current provided by the regulator from a minimum value up to a maximum one and we measure the corresponding output voltage. Once again, the previous consideration on the feedback loop holds true, and of course **the load must be stable**.

- **Load regulation**, which is a measure of the regulator's ability to maintain a constant output voltage V_{Onom} under slowly varying load conditions over a certain range of load current.



$$\%LOR = \left.\frac{V_{O(minL)} - V_{O(FL)}}{V_{O(FL)}}\right|_{V_{IN}=const} \bullet 100\%$$

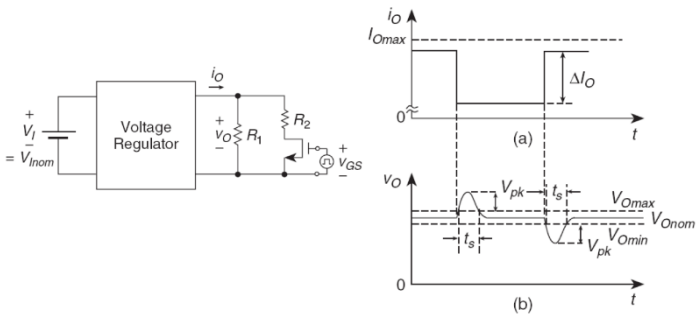
DYNAMIC CHARACTERISTICS OF VOLTAGE REGULATORS

We have two waveforms to define them, the line transient response and the load transient response.

Load transient response

It is the time response of the output voltage upon the application of a step variation in the load current. The switch is open or closed, and if open R2 isn't doing anything and the load consists of R1. Viceversa if the switch is closed, the load is R1||R2, getting a larger load current because the overall resistance is decreased. So we get a step variation of the load current between the two situations.

- **Load transient response**, i.e. the response to a step variation of the load.



Then we have to measure the time transient of the output voltage (b). Upon the increase of the output current the voltage in output reduces and the, due to the presence of the feedback it goes back to the original value.

We want to have a small value for the undershoot and a small settling time. The faster and the smaller the peak, the more ideal the regulator.

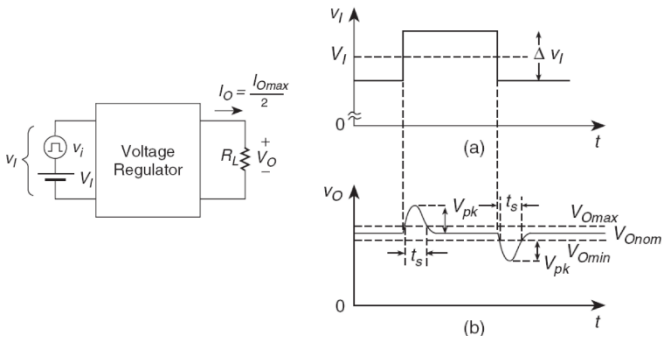
Which are the parameters of the feedback loop that enters into play to have a small setting time and a small amplitude for the undershoot?
If we assume the system is linear, v_o(t) output response of the system is the inverse Laplace transform of the output impedance multiplied by the variation of current delta_I over s. In general, the undershoot is determined by the ESR of the output capacitor, so there is not much we can do.

However, the settling time strongly depends on the frequency response of the Zo(s), output impedance and its BW. BW and amplitude of the closed loop output impedance are related to the closed loop gain, so to get a fast settling time it is better to have a large phase margin (to prevent ringing).

Line transient response

We apply a step variation of the input voltage and we measure the corresponding time transient on the output voltage. We still get an overshoot and an undershoot in the opposite direction. Also in this case the loop gain has an impact.

- **Line transient response**, i.e. the response to a step variation of the line voltage.



How to design the feedback loop?

- The first requirement is the development of a dynamic model of the regulator.
- A model is needed to perform the the dynamic analysis of the feedback system and eventually to design a suitable controller (compensator) for optimal dynamic performance.
- The dynamic performance includes the loop gain, frequency-domain transfer functions (output impedance, audiosusceptibility), and time-domain transient responses.

Firstly we need to develop a dynamic model for the regulator, we cannot replace the regulator with a t.f. because it is not a linear time invariant system.
Once we have the model, we can design the LG.

MODELING THE DC/DC CONVERTER

CHALLENGES

- Dynamic analysis is the main issue with the dc-dc converter.
- In general, dc-dc converters are ***nonlinear, time-varying*** systems.
- Conventional circuit analysis techniques are mainly intended for linear time-invariant (LTI) systems and therefore cannot be directly applied to dc-dc converters.
- While certain analytical methods are available for nonlinear systems, the analysis becomes overly complicated when adapted to dc-dc converters.

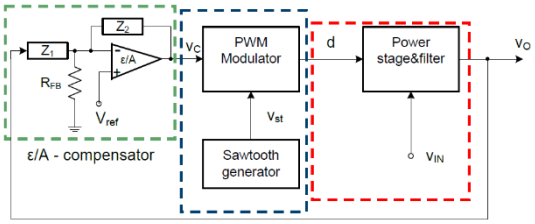
The goal is trying to convert a nonlinear time-varying system in a LTI.

The approach

- We introduce a modeling technique that eventually provides a ***small-signal model*** for the switching regulator.
  - The small-signal model is a ***linear time-invariant (LTI)*** circuit model to which all the standard circuit analysis techniques can directly be applied.
- The small-signal model is used to design a compensator so that closed-loop static and dynamic voltage regulation meet the specs.

Note: we will stick to CCM operation mode!

SMALL SIGNAL MODELING OF A SWITCHING REGULATOR

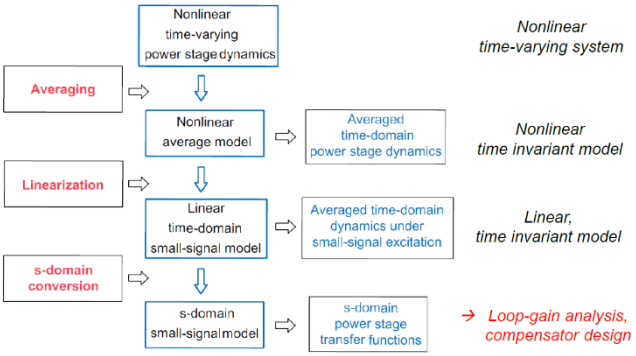


- The regulator is divided into three functional blocks: power stage, PWM block and ε/A-compensator (controller).
- First, each functional block is transformed into the respective small-signal model using a suitable modeling technique.
- The small-signal models of the three functional blocks are later merged to yield a complete small-signal model for the closed-loop converter.

The system is a generic DC/DC converter with a feedback loop and we will split it in three different blocks as in the image. The red block includes the DC/DC converter, the **power stage**. The blue block includes the **PWM** and the green block the **error amplifier**, which is a simple inverting amplifier. The green box is per se a LTI, the blue box too (under certain limitations), the red box not.

POWER STAGE MODELING

- Exploit the *averaging* technique to develop an AC equivalent circuit model that describes converter responses from the duty cycle d and the line input v_{in} to the output voltage v_o



We are considering the converter, because we are focusing on the power block, so diode, transistor, inductor and capacitor. The one in the image is the methodology we can use to develop a small signal dynamic model. We will exploit the **averaging technique**.

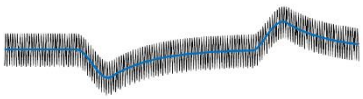
We start from a nonlinear time varying system. After the application of the averaging we get still a non linear mode, but time invariant. To transform it into a linear system we perform a linearization around the point of operation of the converter. The last step is to move from the time domain to the Laplace domain.

Once we have a Laplace domain small-signal model we have a t.f. that describes the operations of the DC/DC converter. Then we have to design a controller in feedback to obtain the desired CL characteristic for our system.
If we have the desired CL characteristics for the regulator, the thing we have to do is to, once the model for the DC/DC converter is found, convert the CL specifications into OL specifications for the loop t.f.. Then we have to shape the t.f. to get those specifications. Since the loop t.f. is the product of the feedforward t.f. and the feedback t.f., we can get the corresponding t.f. for the feedback.

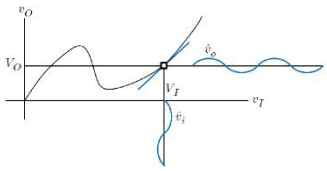
- The method of averaging is first applied to the time-varying power stage dynamics, providing a time-invariant average model.
  - In general, the average model is a nonlinear.
- As the second step, linearization is invoked to deal with the nonlinear relationships incurred during the averaging process.
  - The linearization produces a linear time-invariant small-signal model.
- As the last step, the time-domain small-signal model is converted into a frequency-domain, or s-domain, small-signal model, which provides transfer functions of power stage dynamics.

Essential steps

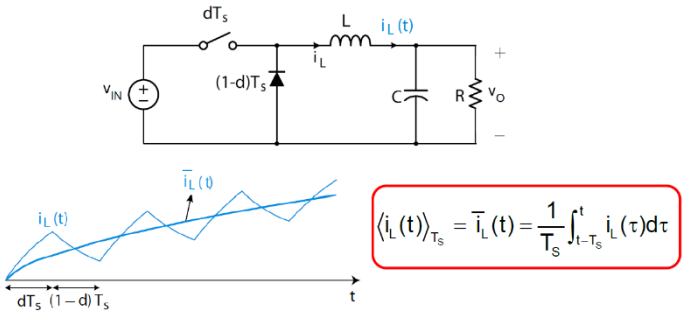
- **Averaging** to remove the time-variance from power stage dynamics



- **Linearization** to approximate a nonlinear function into a linear relationship under small-signal assumption



Averaging power stage dynamics



- Method of averaging produces smooth waveform $\bar{i}_L(t)$ that follows the time-averaged trajectory of the actual inductor current waveform $i_L(t)$.
- Two well-known averaging techniques are state-space averaging and circuit averaging.

The average is not from -inf to +inf. We want to get rid of the ripple because it isn't bringing any useful information as far as the dynamics of the system are concerned. What we do is to apply a moving average over a switching period.

There are two averaging techniques:

- State-space averaging
- Circuit averaging

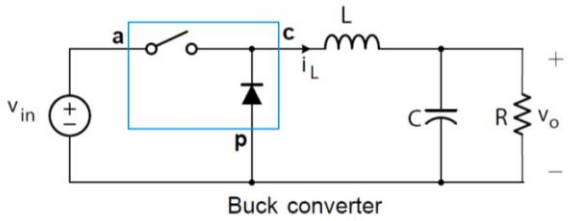
The final result of the two approaches is the same, but in the former we need to write the state space equations and apply the averaging to them in order to get a LTI, while in the latter the averaging is performed directly on the circuit.

CIRCUIT AVERAGING

Circuit averaging: rather than averaging and linearizing the converter state equations, the averaging and linearization operations are performed directly on the converter circuit i.e., all manipulations are performed on the circuit diagram, instead of on its equations.

The resulting average power stage model produces the continuous circuit waveforms that follow the original waveforms.

- Consider a buck converter operating in CCM :



Buck converter

- Switches combined in a three-terminal device:
  - a: active terminal
  - p: passive terminal
  - c: common terminal

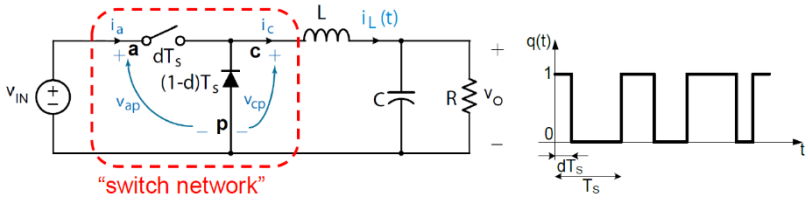
Let's consider for instance a buck converter in CCM and we want to apply the circuit averaging. The first thing to do is to identify the switch network, which is the network in the blue box. For simplicity, the switch is considered ideal and so the diode.

The switching network is a two ports network, that is fully characterized by 4 variables, v1, i1, v2, i2. Two of these variables can be considered independent variables, the other two are dependent variables on the previous two.

Introducing the switching function

Subscript a identifies the active terminal, p the passive terminal connected to the diode and c is the common terminal to the active branch and passive branch.

i_a is the same current i_c, which is instead forced by the inductor, and v_ap is forced by the input voltage generator. So it is easier to consider the i_c and v_ap as independent variables.



- PWM switch equations:

$$\begin{cases} v_{cp} = v_{ap} \cdot q(t) \\ i_a = i_c \cdot q(t) \end{cases} \quad \text{with} \quad q(t) = \begin{cases} 1 & \text{for } 0 < t < dT_S \\ 0 & \text{for } dT_S < t < T_S \end{cases}$$

and $v_{ap} = v_{IN}$, $i_c = i_L$

q(t) is known as "*switching function*"

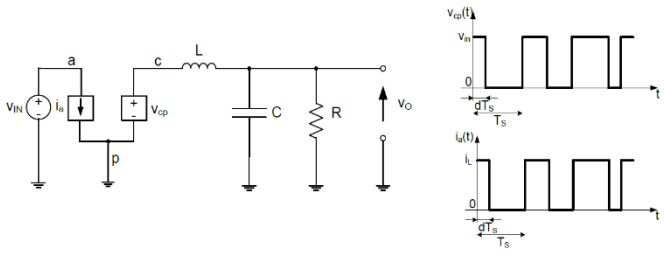
The dependent variables will be i_a and v_cp. Now we need to find the relationship between the independent and dependent variables, which is the result in the red box. q(t) is the switching function, which identifies whether the switch is open or closed. When the switch is open, the current that flows in the inductor in CCM forces the diode to close, and the diode is ideal.

As for the second dependent variable i_a, when the switch is closed, the diode is open and so i_a = i_c, but when the switch is off, i_a = 0.
So we have identified a switching network and defined 4 variables.

Obtaining a time invariant circuit topology

Once we have identified the switching network and the dependent variables we can redraw the circuit so that it is described by a time invariant topology, meaning that the topology of the circuit is not depending on the state of conduction of the switch. In fact, previously the circuit topology was changing depending on the state of conduction of the switch.

Now we replace the switch network with a current and a voltage generator which are reproducing the dependent variables.



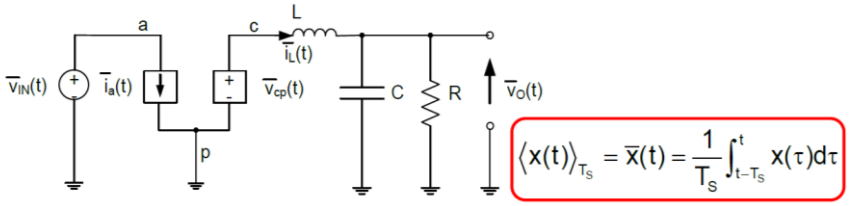
- Key step in circuit averaging: replace the PWM switch with voltage and current sources, to obtain a time-invariant circuit topology.
- The waveforms of the voltage and current generators are defined to be identical to the switch waveforms of the original converter.
- No approximation made so far!

So at the input we have replaced the switch network with a current generator generating a current i_a, which depends on i_c, and at the output we have replaced the switch network with a voltage generator generating a voltage v_cp.
Now the circuit topology is a time invariant one, there are no more switches.

The two voltage and current generators are generating exactly the current and voltages we can measure on the circuit with the switch, there are no approximations.

Average all the waveforms over Ts

Once we have written the time invariant topology, on that topology we apply the average. **Any variable that we can see in the circuit is averaged using a running average**.

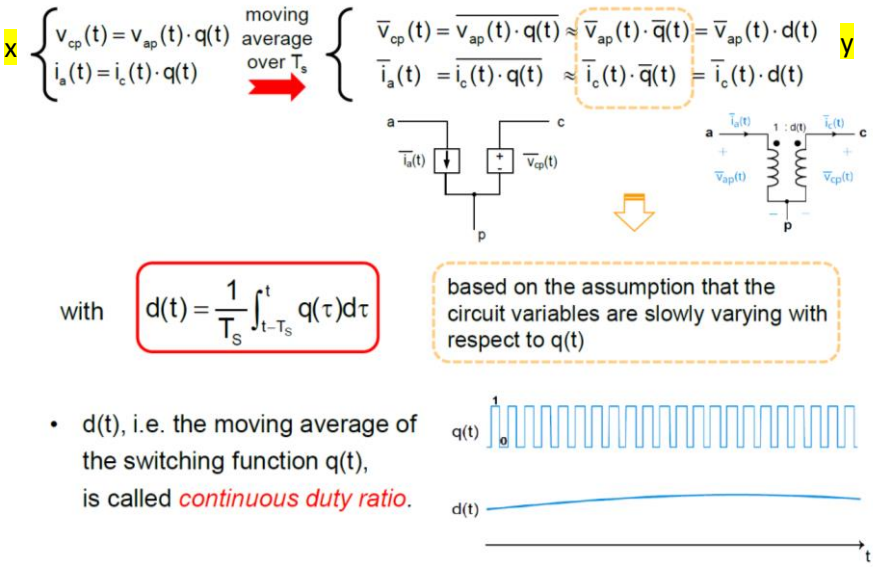


- Basic assumption: the natural time constants of the converter are much longer than the switching period. In this case, the averaging operation does not significantly alter the system response.
- Averaging over the switching period T_s removes the switching harmonics, while preserving the low-frequency components of the waveforms.
- In general, averaged variables are slowly-varying functions of time.
- The linear circuit components, including voltage and current sources, inductors, and capacitors remain invariant during the averaging process.

Running average is defined as in the red box.
The basic assumption that we are making applying this average is that the natural time constants of the converter are much longer than the switching period. The meaning is that applying the averaging doesn't change the dynamic behaviour of the system, it only removes the high frequency ripple.

This assumption is always met in the real world because when we design L and C to have a small ripple, having a small ripple means that this assumption is automatically verified.

Averaging the switch waveforms



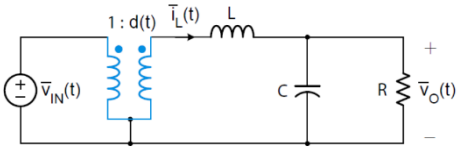
We start from the PWM switch equations x ad we apply a moving average over a period Ts. What we get are average values. The important step is that, **assuming that the circuit variables are slowly varying with respect to q(t), the average of the product can be replaced with the product of the averages**.

Again, this is verified if the natural time constants of the circuit are much larger than the switching period (always true in a well-designed system).

The running average of q(t) is called continuous duty ratio.
After the averaging we can write the average v_cp(t) as the average of v_ap(t) and the continuous duty ratio, and the same for the average i_a(t).
In the final two equations y we can replace the average current and voltage generators with a simple ideal transformer, able to transform both AC and DC signal, whose turn ration is 1:d(t).

As a final result, after the averaging we have replaced the switch network with a model, that is the ideal transformer, and this model provides a NLTI. The problem is that the model is still nonlinear, because there are still the products of functions depending on time.

Averaged model of the buck converter



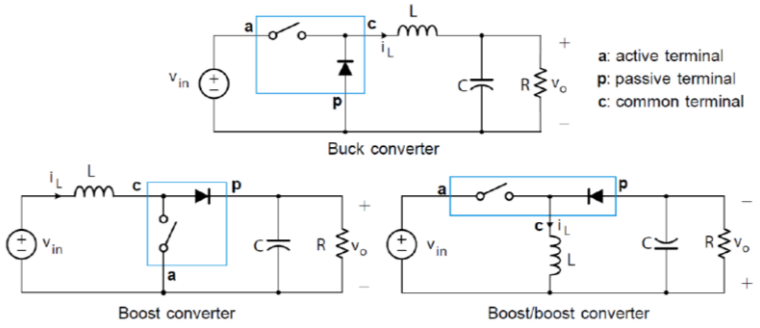
- *Note*: the averaged model of the buck converter could be derived considering the PWM switch as a standalone individual device.
- This idea is the basis of the so-called "*averaged switch modeling*"

There is a second possible approach. In fact, we identified a switch network, performed the averaging steps on the switching network and in the end we plugged the model in the buck converter.
Instead of focusing on the converter and do all these passages applying the average everywhere, we can try to extract from the converter (any converter) the switch network, average the switch network and once we have averaged the switch network and obtained a model, we put the model back into the original converter. This approach is called averaged switch modelling.

AVERAGED SWITCH MODELLING

- The central idea of the *averaged switch modeling* approach is to find an averaged circuit model for the *switch network*. The resulting averaged switch model can then be inserted into the converter circuit to obtain a complete averaged circuit model of the converter.
- An important advantage of the averaged switch modeling approach is that the same model can be used in many different converter

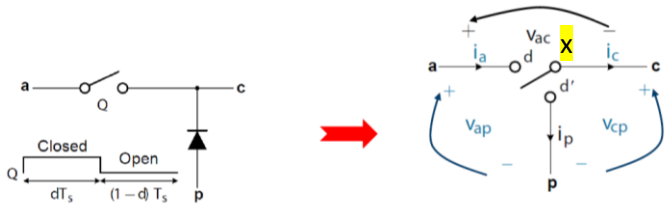
- Three basic converters



We identify the switch network in all the three converters, the one that includes the switch and the diode.

We extract the switch network, perform all the averaging operations on the extracted network and then we plug back the final model.

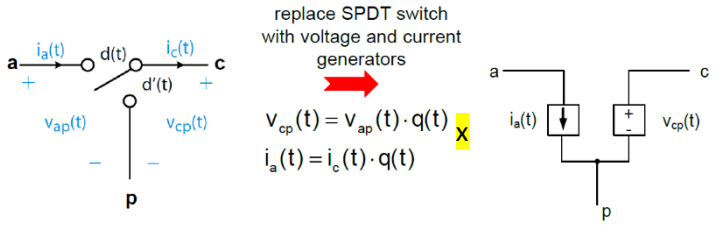
- The active and passive switches in the three basic converters can be lumped together in a single-pole, double-throw switch called the ***PWM switch***
- All the elements outside the PWM switch are linear passive elements which provide filtering, whereas the PWM switch is the only nonlinear element which performs the dc-to-dc conversion process.



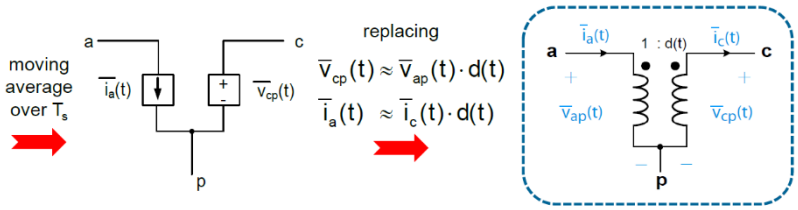
PWM switch

We end up with a single pole double throw switch (x) that represents the switch network including an active switch (mosfet) and a passive switch (diode).

Averaging steps



Note: this set of equations for the terminal currents and port voltages of the PWM switch is *independent* of the particular converter in which the PWM switch is implemented in.



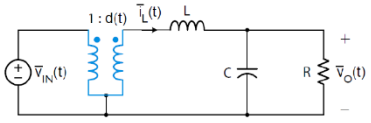
The PWM switch can be replaced by a time invariant topology which is defined by a current generator and a voltage generator. Set of equations x is independent on the type of converter we are using.
Then the moving average is applied to the time invariant topology and the PWM switch can be modelled with an ideal transformer with turn ratio 1:d(t), where d(t) is the running average of the switching function.

The model is still nonlinear, but it can be plugged in all the converters paying attention to the correct positioning of the active, passive and common terminals (next image). For instance, the boost converter has the active switch grounded, while the common terminal is always connected to the inductor. Instead, the buck-boost has the switch in series with the input.

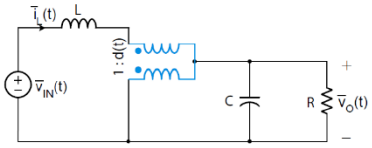
If the duty cycle was fixed and not a function of time d(t), the averaged model is already a LTI.

- The average models are simply obtained with a pin-to-pin replacement of the active-passive switch pair with the average model of the PWM switch.

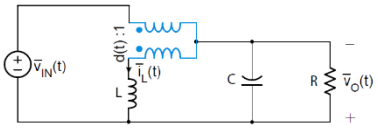
- Buck



- Boost

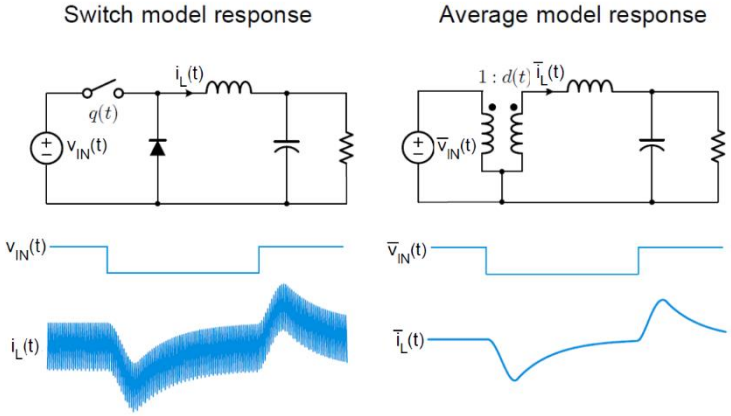


- Buck-boost



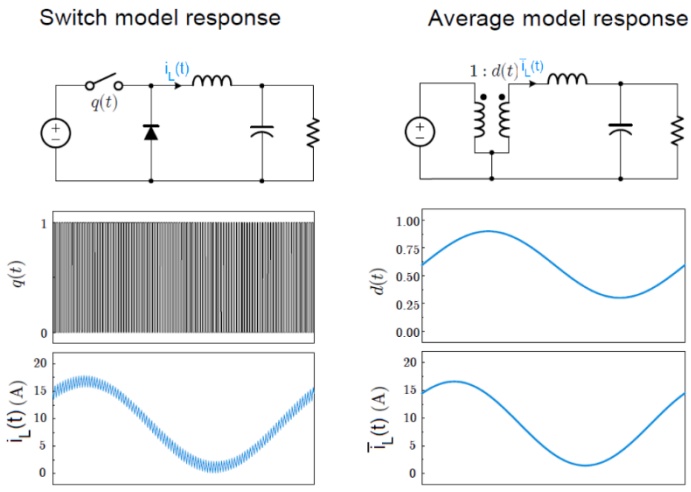
This network is still nonlinear. The last step will be linearization.

RESPONSE OF THE AVERAGE MODEL



We are investigating the response of the model to a step variation of the input voltage. On the left we have the original converter with the ripples on the inductor's current.

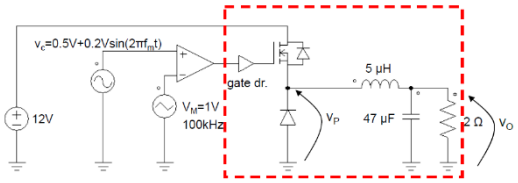
Practical example – Buck converter



In this case the duty cycle D is modulated in a sinusoidal way.

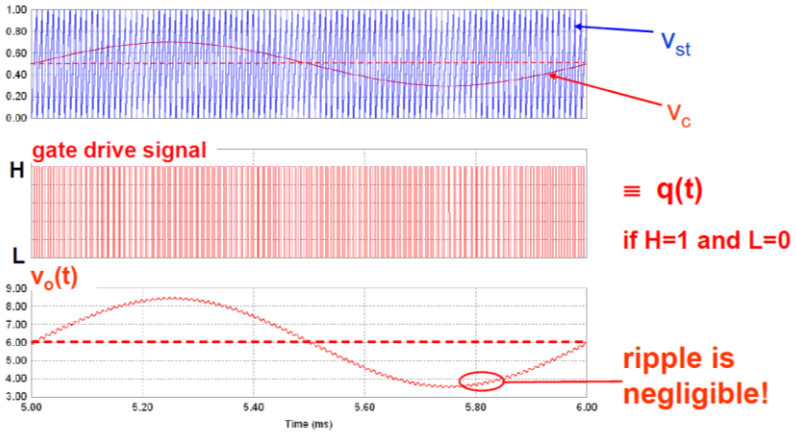
Practical example

Once again it is a buck converter.



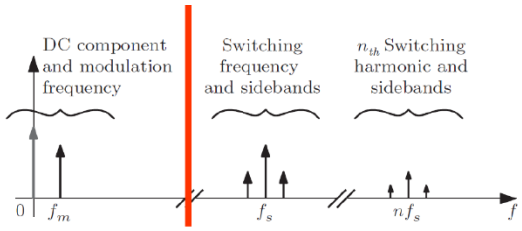
- v_c is modulated sinusoidally: $v_c(t) = V_c + V_{cm} \sin(2\pi f_m t)$
- the modulation frequency f_m is much smaller than the converter switching frequency f_s

The switch is drive with a sinusoidally modulated control signal v_c. In general, the control signal is not a sinusoidal voltage, but it comes from the feedback.
The frequency of the modulation is much smaller than the switching frequency. The converter waveforms are the following ones.



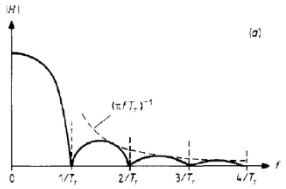
The blue line is the sawtooth voltage waveform and the red line in the same plot is the control signal.
The duty cycle is modulated according to the sinusoidal signal. Then the output voltage v_o(t) is the superposition of a DC value and the modulation. Of course we also have a ripple on the top of it. The amplitude of the modulated output signal is 2.4 with respect to the DC value, and it is not a surprise because it is 12/5, and 5 comes from the ratio between amplitude of the sawtooth voltage and the control voltage AC amplitude (1V/0.2V).

OUTPUT VOLTAGE SPECTRUM



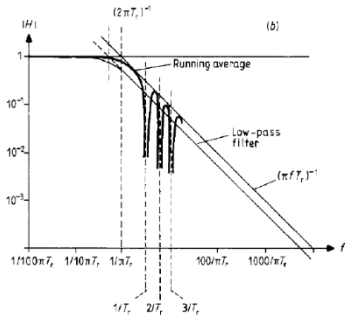
- Assuming that the converter duty cycle is sinusoidally modulated at a frequency $f_m \ll f_s$, the output voltage consists of a low-frequency component $\langle v_o(t) \rangle$, plus a high-frequency *switching ripple*.
- The low-frequency component of $\langle v_o(t) \rangle$ contains a dc term V_o and a spectral component at the modulation frequency f_m.
- The high-frequency content contains the switching frequency f_s and its harmonics, as well as all the modulation sidebands originating from nonlinear interactions between f_m and f_s components.

There is a DC component, a modulation component at fm plus a series of harmonics. At the switching frequency fs there is the switching harmonics, and also at the multiples of the switching frequency. There harmonics are at a frequency far away from the DC region, and they are producing the ripple, but we are interested in the underlying low frequency behaviour of the system.



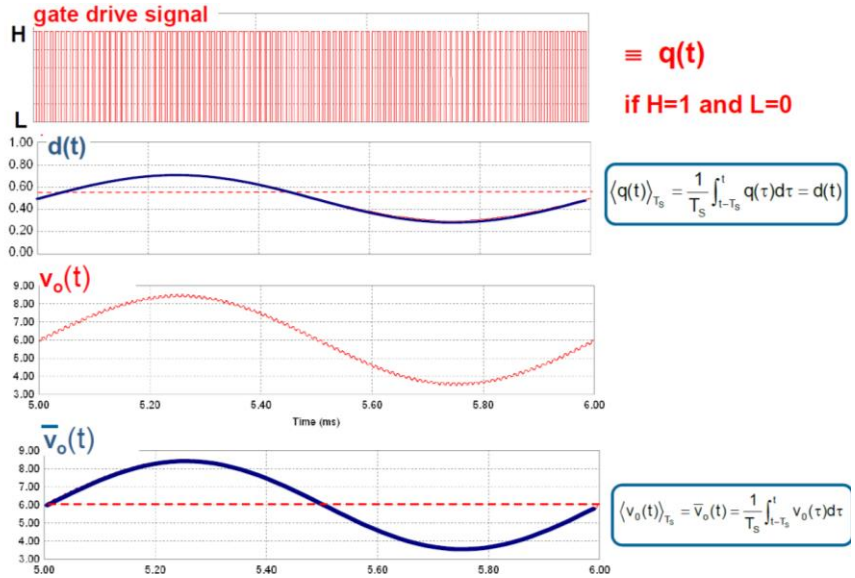
Frequency response of the moving average filter

To remove the switching harmonics we can use a simple moving average filter, which has notches at fs and multiple of fs to kill the harmonics.
We also preserve the low frequency component.



This moving average is just a conceptual step in the development of the switch model, because this operation is intrinsically done when we replace the switching function with the continuous duty ratio function d(t).

Averaged waveforms



However, we still have the problem of nonlinearity.

LINEARIZATION OF NONLINEAR FUNCTION

We linearize the model around a linearization point and assuming small signals → we will develop a small signal linear model around an operating point.

We start by writing any average variable, either a current or a voltage that we have in the circuit, as the sum of a DC value X and an AC signal x_tilde.
For simple algebraic nonlinear equations, the small signal model can be obtained by plugging the variables rewritten in this way into the nonlinear equations, separating the DC terms and AC terms, and then by dropping the high order terms.

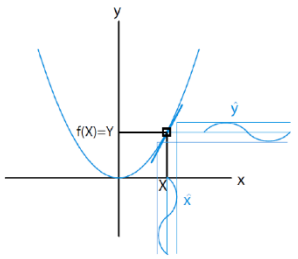
- Linearization is invoked to deal with the nonlinear relationships incurred during the averaging process. Linearization produces a linear time-invariant small-signal model describing the averaged time-domain dynamics upon the application of a small-signal excitation.
- An averaged variable (voltage or current) includes a DC component and a small-signal AC component

$$\langle x(t)\rangle_{T_s} = \bar{x}(t) = X + \tilde{x}$$

- For simple algebraic nonlinear equations, the small-signal model can be found by:
  - evaluating the nonlinear equation with the variables consisting of DC (steady-state) and AC (small-signal perturbation) components
  - equating only ac components of the input and output variables.Linearization is accomplished by perturbing and linearizing the averaged model about a quiescent operating point
  - retaining only the first-order, linear terms.

Graphical interpretation

We want to linearize the parabola around a generic point X.



$$Y + \tilde{y} = f(X) + \left.\frac{df}{dx}\right|_{x=X} \cdot \tilde{x}$$

$$\boxed{\tilde{y} = K \cdot \tilde{x}} \quad \text{with } K = \left.\frac{df}{dx}\right|_{x=X}$$

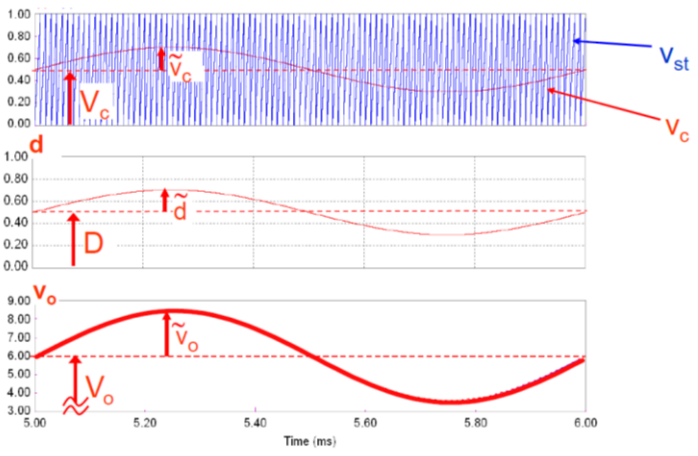
- Example: linearizing $y = x^2$ around a generic X

$$\underbrace{Y}_{\text{DC}} + \underbrace{\tilde{y}}_{\text{ac}} = (X+\tilde{x})^2 = \underbrace{X^2}_{\text{DC}} + \underbrace{2X\tilde{x}}_{\text{ac}} + \underbrace{\tilde{x}^2}_{2^{nd}\text{ order term}}$$

$\boxed{\tilde{y} = 2X\tilde{x}}$ → linearized ac equation

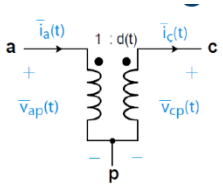
The result obtained using the derivative could have been obtained in a simpler way by plugging the values y and x written as the superposition of an AC value and a DC component. The simplification of the second order term can be done if the amplitude of the perturbation is negligible.

We can apply this approach to our model (still using the previous example). For the control signal the DC level is the dashed one, and the AC component is the amplitude peak to DC dashed value. The same for the output voltage.



LINEARIZATION OF THE AVERAGED MODEL

$$\bar{v}_{cp}(t) = d(t)\cdot \bar{v}_{ap}(t)$$
$$\bar{i}_a(t) = d(t)\cdot \bar{i}_c(t)$$



- Linearization

$$\bar{v}_{cp} = d\cdot \bar{v}_{ap} \Rightarrow (V_{cp} + \tilde{v}_{cp}) = (D+\tilde{d})\cdot(V_{ap}+\tilde{v}_{ap})$$
$$= DV_{ap} + V_{ap}\tilde{d} + D\tilde{v}_{ap} + \tilde{d}\tilde{v}_{ap}$$
DC ac DC ac ac 2nd order

$\tilde{v}_{cp} = V_{ap}\tilde{d} + D\tilde{v}_{ap}$ **ac equation** $V_{cp} = DV_{ap}$ **DC equation**

$$\bar{i}_a = d\cdot \bar{i}_c \Rightarrow (I_a + \tilde{i}_a) = (D+\tilde{d})\cdot(I_c+\tilde{i}_c)$$
$$= DI_c + I_c\tilde{d} + D\tilde{i}_c + \tilde{d}\tilde{i}_c$$

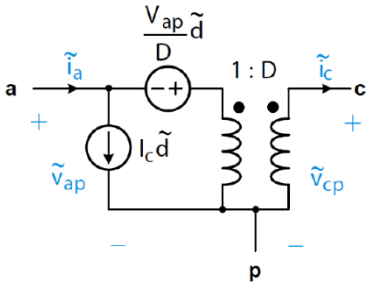
$\tilde{i}_a = I_c\tilde{d} + D\tilde{i}_c$ **ac equation** $I_a = DI_c$ **DC equation**

This is the final linearization procedure.
We start from the averaged model where the nonlinearity comes from the fact that v_cp is d(t)*v_ap(t). The second order terms can be dropped. The final step is then plugging the linearized equations inside the model, as below.

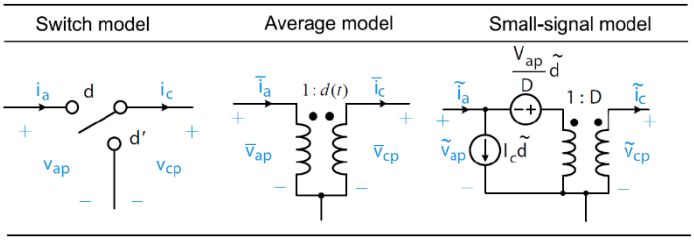
- ac equations of the PWM switch

$$\begin{cases} \tilde{v}_{cp} = V_{ap}\tilde{d} + D\tilde{v}_{ap} \\ \tilde{i}_a = I_c\tilde{d} + D\tilde{i}_c \end{cases}$$

- Circuit model for ac equation: PWM switch model



In the transformer there is no more d(t), but the turn ratio is 1:D and the AC component of d is included in the two generators.

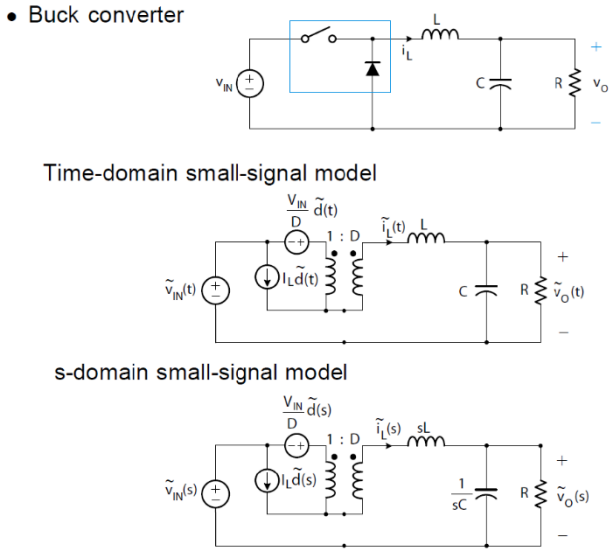


Switch model	Average model	Small-signal model
$v_{cp} = v_{ap}\cdot q(t)$ $i_a = i_c\cdot q(t)$	$\bar{v}_{cp}(t) = \bar{v}_{ap}(t)\cdot d(t)$ $\bar{i}_a(t) = \bar{i}_c(t)\cdot d(t)$	$\tilde{v}_{cp} = V_{ap}\tilde{d} + D\tilde{v}_{ap}$ $\tilde{i}_a = I_c\tilde{d} + D\tilde{i}_c$

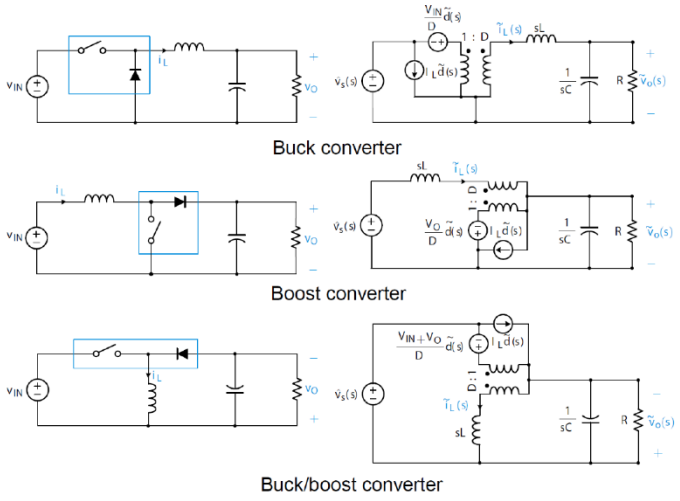
$$q(t) = \begin{cases} 1 \text{ for } dT_S \\ 0 \text{ for } d'T_S \end{cases} \qquad d(t) = \frac{1}{T_S}\int_{t-T_S}^{t} q(\tau)d\tau$$

TIME DOMAIN AND S-DOMAIN SMALL SIGNAL MODELS

The last step is getting the model in the Laplace domain, which can be done by replacing the voltages as a function of s and the capacitor and inductor with their impedances.



Small signal models in the Laplace domain

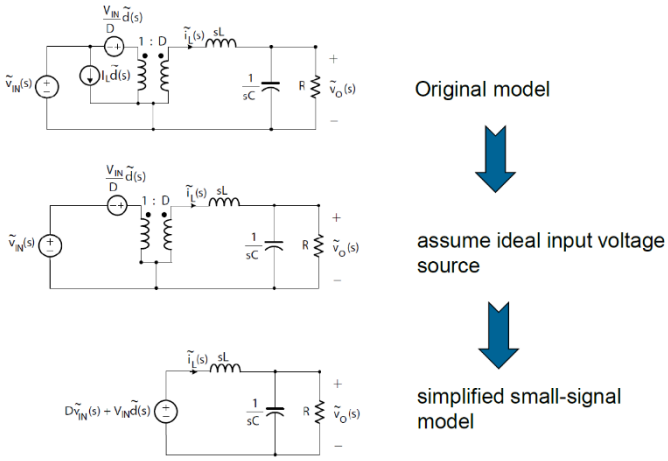


At this point, the final result is that the converter has became a single LTI network that we can analyze with the standard methodologies.

Buck converter small signal model

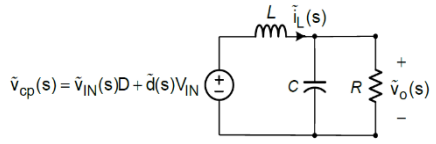
Let's suppose the buck converter is supplied by an ideal voltage generator without a series resistance. In this case we can simplify the model because the current generator is in parallel with the input generator so we can get rid of it (if the voltage generator is ideal) and the model reduces to a transformer and a small signal voltage generator in series with the input voltage generator.

The two generators in series can be brought on the secondary side of the transformer by multiplying them by the turn ratio D. We can hence compute v_o_tilde(s).

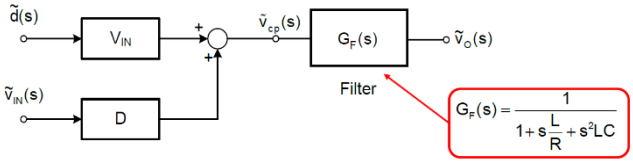


Block diagram

The output voltage is the voltage at the input multiplied by the filter t.f.. We are assuming that the capacitor has no ESR, otherwise we would have a negative zero at the numerator.

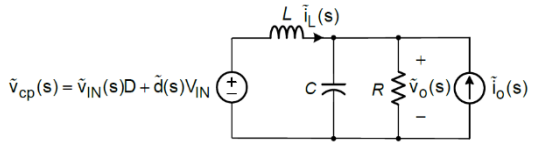


$$\tilde{v}_O(s) = \frac{\tilde{v}_{CP}(s)}{1 + s\frac{L}{R} + s^2 LC} = \frac{\tilde{v}_{IN}(s)\cdot D + \tilde{d}(s)\cdot V_{IN}}{1 + s\frac{L}{R} + s^2 LC}$$



$$G_F(s) = \frac{1}{1 + s\frac{L}{R} + s^2 LC}$$

Transfer functions



$$G_{od}(s) = \left.\frac{\tilde{v}_o(s)}{\tilde{d}(s)}\right|_{\substack{\tilde{v}_{IN}=0 \\ \tilde{i}_o=0}} = \frac{V_{IN}}{1 + s/Q\omega_o + s^2/\omega_o^2}$$

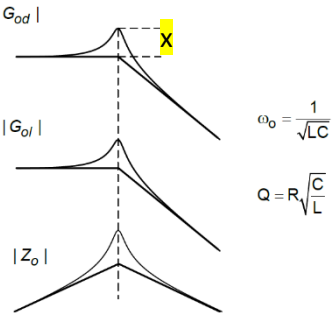
duty-cycle to output

$$G_{ol}(s) = \left.\frac{\tilde{v}_o(s)}{\tilde{v}_{IN}(s)}\right|_{\substack{\tilde{d}=0 \\ \tilde{i}_o=0}} = \frac{D}{1 + s/Q\omega_o + s^2/\omega_o^2}$$

line to output

$$Z_o(s) = \left.\frac{\tilde{v}_o(s)}{\tilde{i}_o(s)}\right|_{\substack{\tilde{v}_{IN}=0 \\ \tilde{d}=0}} = \frac{sL}{1 + s/Q\omega_o + s^2/\omega_o^2}$$

output impedance



$$\omega_o = \frac{1}{\sqrt{LC}}$$

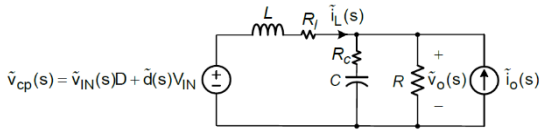
$$Q = R\sqrt{\frac{C}{L}}$$

The first t.f. considers how a perturbation of the duty cycle affects the output. It is computed putting v_in(s) to ground. Omega0 is the natural resonant frequency, 1/sqrt(LC). Q is the quality factor and it is the height of the overshoot in the frequency response (x).

Instead, the line to output transfer function is the t.f. between the perturbation of the output voltage and a perturbation on the input line voltage, keeping the duty cycle perturbation to 0.

The interesting t.f. is the output impedance, which is the ratio between the output voltage and the output current. So we inject a current in output and we measure the perturbation on the output voltage. Assuming that there is no ESR, the output impedance has a zero in the origin. Again, depending on Q we have more or less peaking in the transfer function.

Parasitics inclusion



$$G_{ol}(s) \equiv \frac{\tilde{v}_o}{\tilde{v}_{IN}} = D\frac{\left(1+\frac{s}{\omega_{esr}}\right)}{1+\frac{s}{Q\omega_o}+\frac{s^2}{\omega_o^2}}$$

$$G_{od}(s) \equiv \frac{\tilde{v}_o}{\tilde{d}} = V_{IN}\frac{\left(1+\frac{s}{\omega_{esr}}\right)}{1+\frac{s}{Q\omega_o}+\frac{s^2}{\omega_o^2}}$$

$$Z_o(s) \equiv \frac{\tilde{v}_o}{\tilde{i}_o} = R\|R_l\frac{\left(1+\frac{s}{\omega_z}\right)\left(1+\frac{s}{\omega_{esr}}\right)}{1+\frac{s}{Q\omega_o}+\frac{s^2}{\omega_o^2}}$$

$$Q = \frac{1}{\omega_o}\cdot\frac{R+R_l}{L+C(R_lR_C+R_lR+R_CR)} \cong R\sqrt{\frac{C}{L}}$$

$$\omega_o = \sqrt{\frac{1}{LC}\frac{R+R_l}{R+R_C}} \cong \sqrt{\frac{1}{LC}}$$

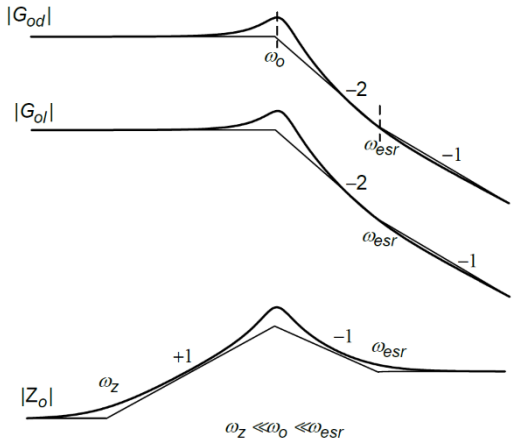
$R \gg R_C$
$R \gg R_l$

$$\omega_{esr} = \frac{1}{CR_C}$$

$$\omega_z = \frac{R_l}{L}$$

The most important effect is the appearance of a negative zero in the line to output t.f. and in the duty cycle to output t.f.. The denominator remains the same. In general, the impact of the ESR of the inductor and of the capacitor in Q and omega0 is almost negligible, given that R >> Rc and R >> Rl.

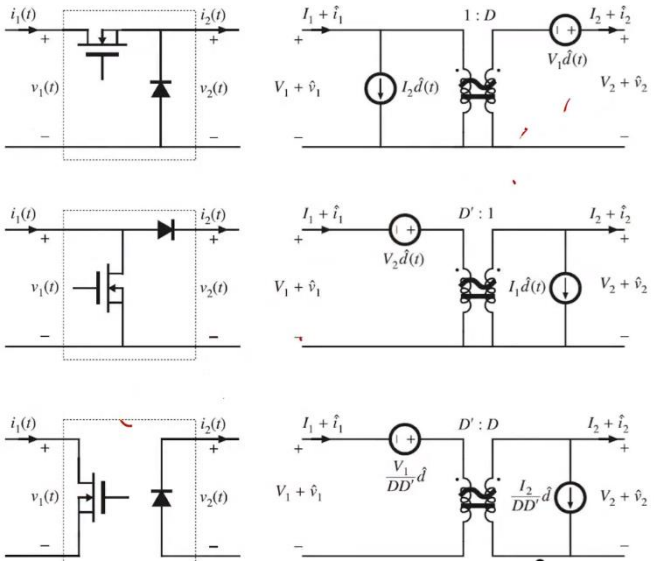
Another important observation is that the Rl is producing a second negative zero in the output impedance. The small signals transfer functions are as below.



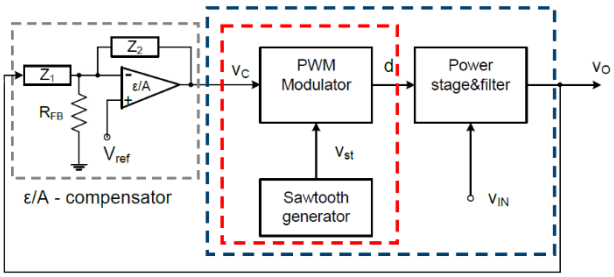
In general, the poles in the transfer functions are complex and conjugate. Zo is constant at 0 and infinite frequency. In fact, at HF the inductor is an open circuit and the capacitor is a short, so the output impedance is the parallel of Rc and R, so basically Rc. At zero frequency, the inductor is a short and the capacitor is an open, and Zo = R || Rl, so basically Rl.

MODELLING THE PWM

There are different ways in which we can model the switching network, as indicated in the image below.



We want to model the PWM.



Task: develop an ac equivalent circuit model of the PWM block

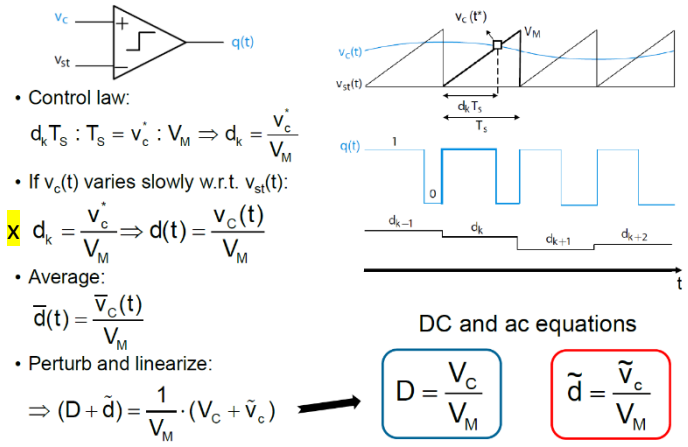
MODELLING THE PWM BLOCK

The PWM is a comparator that typically on the – terminal receives the sawtooth waveform and a control voltage on the + terminal. The control voltage is a slowly varying function of time with respect to the switching period.

The PWM produces at the output a square wave; assuming the low level is 0 and the high level is 1, the output voltage produced by the PWM is the switching function q(t).
Let's focus on the generic k switching period. Within it, the control voltage is compared to the sawtooth voltage waveform and the output of the comparator, if the v_c is higher than v_st, is high. Then in the next point, the control voltage crosses the sawtooth voltage waveform and if so the q(t) goes to 0.

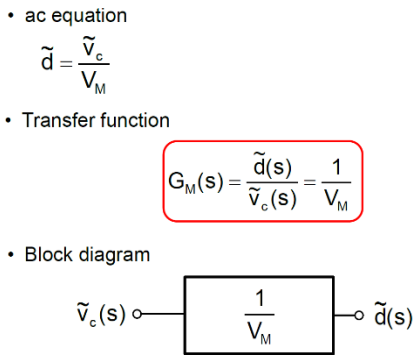
Can we compute the duty cycle at the k switching period?
It is the d_k and it is computed using the principle of similar triangles on the sawtooth waveform in a period. v_c* is the control voltage computed at t* where the two waveforms cross each other.
An interesting thing is that the values of the duty cycles are discrete values, because the PWM can produce a single value of duty cycle per switching period, and it is the last black plot.

However, if the control voltage varies very slowly with respect to the switching period, we can assume that the difference between d_k-1, d_k and so on is small, and so we can replace the discrete function with a continuous one. This is what is done at x.

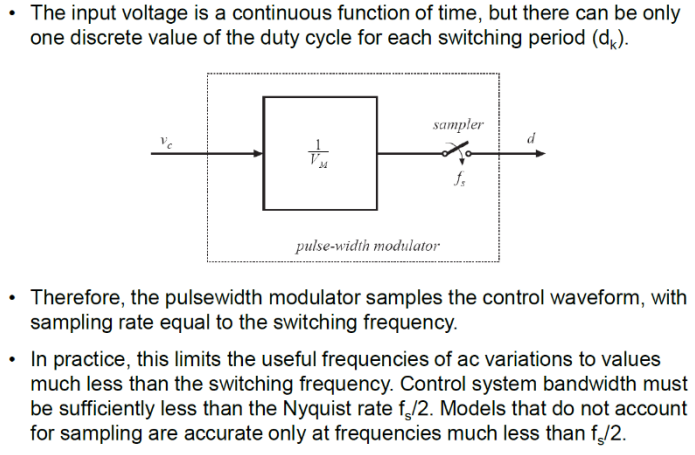


On the continuous function of time we apply the running average over a time base equal to the switching period and then we linearize.

Small signal AC model

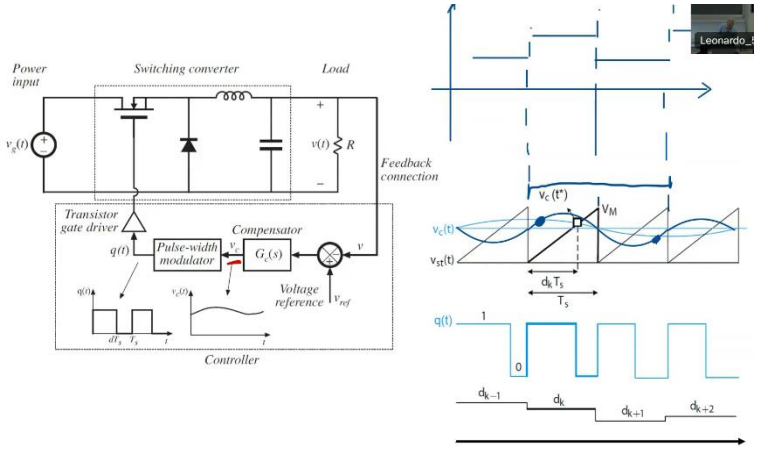


So the PWM can be replaced with a simple block having a constant gain 1/Vm.
In reality, it is not this simple. The input voltage to the PWM is a continuous function of time v_c, but there can be just a single discrete value of duty cycle at the output per switching period. So as a matter of fact the PWM is sampling the control voltage at the input of the PWM, with a sampling rate that is equal to the switching frequency. So the real model should have a sampler after the 1/Vm block. Thus discrete values of duty cycles are produced.



If the BW of v_c is larger than half of the switching frequency we have aliasing, and this has to be avoided. To avoid this, the sampling rate (i.e. switching frequency) has to be larger than twice the BW of the input signal.

More details

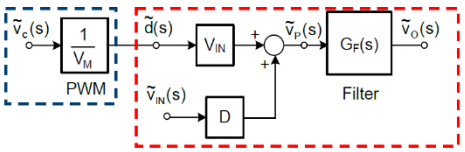


If we assume that the control signal is a sinusoidal signal with a frequency exactly half of the switching frequency, we have two samples of the sinusoidal voltage in a period (one sawtooth triangle).
If we increase the frequency of the control signal, we are generating a single value of the duty cycle, so it is an indistinguishable situation with respect to when we apply a constant voltage v_c.

Control to output transfer function

We have a block diagram including the DC/DC converter (filter and power part) and the PWM. The input of the PWM is the control signal, the output of the system is the output voltage.
The blue and red systems are in series, so I can compute the t.f. between the output voltage and the control voltage, i.e. the control to output t.f., by multiplying the t.f. of the PWM and the t.f. of the DC/DC converter.

- The *control-to-output* characteristic is simply the behavior of buck regulator when the error amplifier is removed from the system.



$$G_{oc}(s)=\left.\frac{\tilde{v}_O(s)}{\tilde{v}_C(s)}\right|_{\tilde{v}_{IN}=0}=G_M(s)\,G_{od}(s)=\frac{V_{IN}}{V_M}\frac{1}{1+sL/R+s^2LC}$$

where

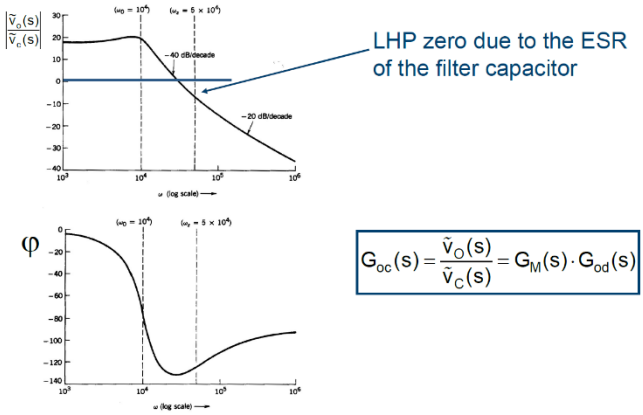
$$G_M(s)=\frac{\tilde{d}(s)}{\tilde{v}_C(s)}=\frac{1}{V_M}\quad;\quad G_{od}(s)=\frac{\tilde{v}_O(s)}{\tilde{d}_P(s)}=V_{IN}\frac{1}{1+sL/R+s^2LC}$$

note $$G_{oc}(s)=\frac{V_{IN}}{V_M}\frac{(1+sR_CC)}{1+s\left(\frac{L}{R}+R_CC\right)+s^2LC}$$

If an electrolytic capacitor is used with ESR=R_C << R

Goc is the feedforward t.f., also called plant transfer function. This t.f. also shows a negative zero if the ESR of the output capacitor is considered.

Typical t.f. of a buck converter



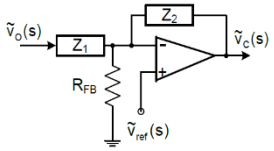
$$G_{oc}(s) = \frac{\tilde{v}_O(s)}{\tilde{v}_C(s)} = G_M(s) \cdot G_{od}(s)$$

We can identify a pair of complex conjugate poles and a real negative zero due to the ESR. The c.c. poles introduce a very quick reduction of the phase, which reaches a minimum and then there is a recover due to the negative zero contribution that adds 90°.

ERROR AMPLIFIER AND COMPENSATOR

This is the last block to be modelled in the regulator.

SMALL SIGNAL AC MODEL



- Typical implementation of a ε/A and compensator
- R_{FB} is used to set the DC value of the output voltage (see Part 7)

Assuming that the OA is ideal:

x $$\tilde{v}_C(s) = \tilde{v}_{ref}(s)\left(1+\frac{Z_2(s)}{Z_1(s)//R_{FB}}\right) - \tilde{v}_O(s)\frac{Z_2(s)}{Z_1(s)}$$

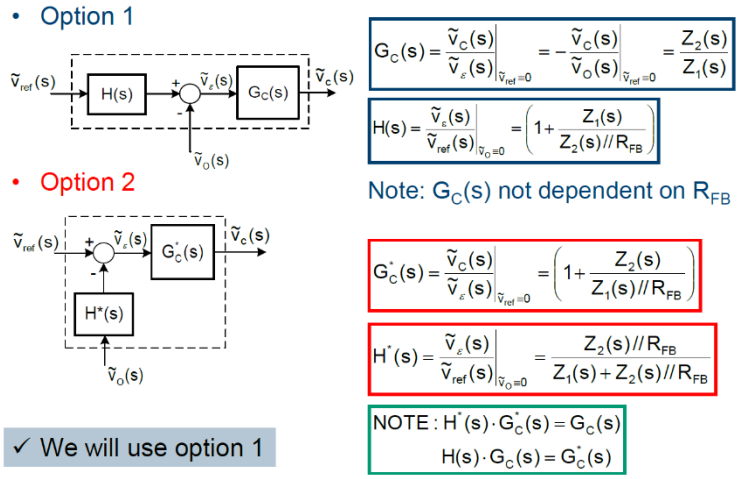
$$\tilde{v}_C(s) = \begin{cases} 1 \rightarrow \left[\tilde{v}_{ref}(s)\left(1+\frac{Z_1(s)}{Z_2(s)//R_{FB}}\right) - \tilde{v}_O(s)\right]\left(\frac{Z_2(s)}{Z_1(s)}\right) & G_C(s) \\ 2 \rightarrow \left[\tilde{v}_{ref}(s) - \tilde{v}_O(s)\left(\frac{Z_2(s)//R_{FB}}{Z_1(s)+Z_2(s)//R_{FB}}\right)\right]\left(1+\frac{Z_2(s)}{Z_1(s)//R_{FB}}\right) & G_C^*(s) \end{cases}$$

The error amplifier and compensator is just an inverting amplifier. We will assume by now the opamp as ideal. Z1 and Z2 are designed by the user to shape the loop properly.
Rfb is a resistor that is put between the inverting input of the opamp and ground and it plays a role only on DC level, it is used to determine the steady state value of the output voltage of the regulator.

The input signal to the error amplifier is the output voltage of the converter v_o, and the output of the compensator is the control voltage, which is then fed to the PWM. In general, v_ref is a constant DC value generated internally with a bandgap voltage generator, so in the case of small signal analysis it should be grounded. However, there are some applications such as the motion control where the v_ref is not a DC value, so we cannot ground it. Since we want to be general, we assume v_ref as a second input to the opamp.

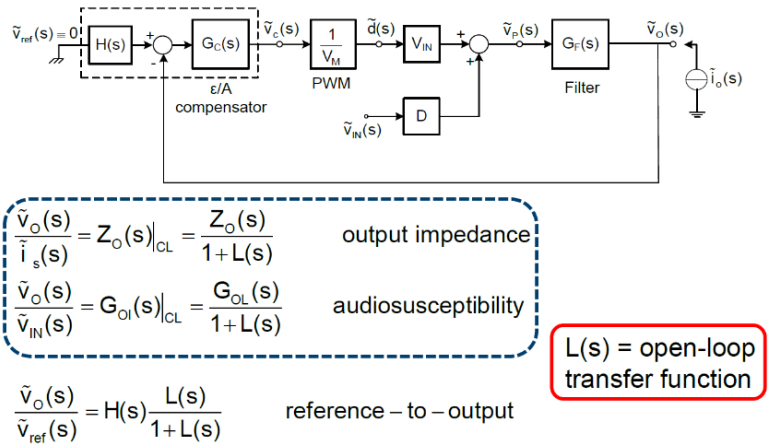
The system is LTI, so we can apply the superposition of effects. Then we can rearrange equation x in two possible ways. In the first case, Gc(s) is the t.f. of the compensation block.

The following are the block diagrams representing the two possibilities.



Gc(s) is not exactly the inverting amplifier t.f., there is a – missing which has been moved to the summing node in the model. We are not interested then in H(s) in the case of a regulator, because the v_ref is constant and so the node to which is attached is grounded.

BUCK REGULATOR SMALL SIGNAL AC MODEL



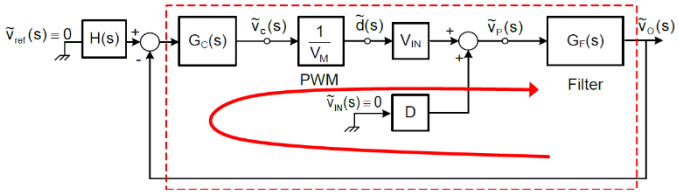
We have the model for the DC/DC converter, the model for the PWM and the model for the error amplifier using option 1. Using option 1 we end up in a unity feedback loop.
To get L(s) we put v_in to ground and perform the multiplication of all the blocks in the feedforward path. Moreover, the OL t.f. is different from the Gloop, the Gloop has a negative sign in this case.

The other important t.f. is the output impedance, ratio between the output votlage and current injected in the output. The CL Zo is the OL Zo divided by 1 + L(s), so the effect of the feedback is to reduce the OL Zo. The same is true for the line to output t.f., also called audiosusceptibility.

For the reference to output t.f., if we assume the reference is not constant we can define a t.f.. We are interested in the DC t.f. from v_ref to v_o because it is the one that determines the DC value of the output voltage.

Loop transfer function

The loop t.f. has the summing node outside it, so L(s) is simply the product of the forward block, it is the controller plant. The minus sign is not in the equation because it is given for granted and extracted from the loop, acting on the summing node.



$$L(s) = G_C(s) \cdot G_M(s) \cdot V_{IN} \cdot G_F(s)$$

Loop function $L(s)$ = products of the gains around the negative feedback loop.

$$\rightarrow \boxed{L(s) = G_C(s) \cdot G_{OC}(s)}$$

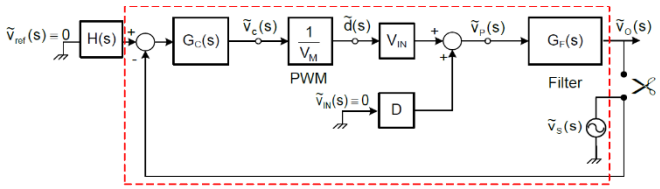
- In the literature, *L(s)* is considered *without* the phase reversal brought by the negative feedback:

$\angle L(j\omega_c) = -180°$ brings instability..

If we consider L(s), the system becomes unstable when the phase of the L(jwc), so calculated at the crossover frequency, is -180°.

L(s) vs Gloop(s)

Now also the summing node is considered. In this case Gloop(s) is negative and the system is unstable when the phase of Gloop at the crossover frequency is -360°.



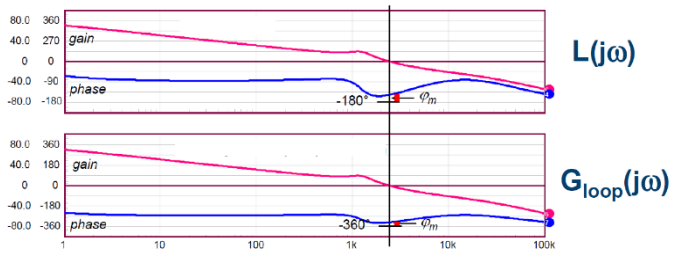
$$G_{loop}(s) = \frac{\tilde{v}_O(s)}{\tilde{v}_S(s)} = -G_C(s) \cdot G_M(s) \cdot V_{IN} \cdot G_F(s)$$ loop gain

$$\rightarrow \boxed{G_{loop}(s) = -G_C(s) \cdot G_{OC}(s) = -L(s)}$$

- In the real life, $G_{loop}(s)$ *includes* the phase reversal brought by negative feedback:

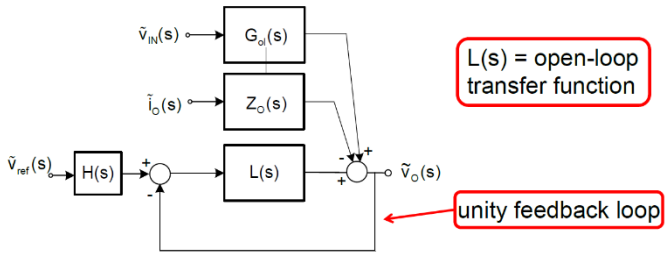
$\angle G_{loop}(j\omega_c) = -360°$ brings instability..

Phase margins for L(jw) and Gloop(jw)



If we use the L(s) the phase margin is 180° minus the value of the phase of L(s). We will use L(s).

COMPACT BLOCK DIAGRAM REPRESENTATION



$$\tilde{v}_O(s) = \tilde{v}_{ref}(s)\cdot H(s)\cdot\frac{L(s)}{1+L(s)} + \tilde{v}_{IN}(s)\cdot\frac{G_{OL}(s)}{1+L(s)} - \tilde{i}_O(s)\cdot\frac{Z_O(s)}{1+L(s)}$$

This is a block diagram of the system which can be used to compute all the t.f.. In general, Zo,OL is summed with a minus sign because the current we use to test the output impedance is extracted from the output, not injected in it.

EFFECT OF L(jw) ON CLOSED LOOP TFs

- Below the crossover frequency:

  $f < f_c$
  - ➢ both output impedance and audiosusceptibility are reduced in magnitude by $1/|1+L(j\omega)|$
- Above the crossover frequency:

  $f > f_c$
  - ➢ feedback loop has essentially no effect on output impedance and audiosusceptibility
- Asymptotic approximation becomes inaccurate in the neighborhood of the loop gain crossover frequency where $|L(j\omega)|\approx 1$

$$\left|\frac{1}{1+L(j\omega)}\right| = \begin{cases} \left|\frac{1}{L(j\omega)}\right| & \text{for } |L(j\omega)| \gg 1 \\ 1 & \text{for } |L(j\omega)| \ll 1 \end{cases}$$

The magnitude of the generic OL t.f. is reduced by a factor 1 + L(jw), with the corresponding approximations in the blue dashed box. If |L(jw)| << 1 the loop has no effect on the OL transfer functions.

Unfortunately, the dynamic behaviour of the system (stability, phase margin, setting time and so on) depends on the value of L(s) or Gloop(s) around the crossover frequency, so we cannot use the asymptotic behaviours.

If we are considering f << fc, 1 + L(s) is more or less 1, so |v_o/v_ref| = |H(jw)|. Since in regulators v_ref is a DC signal, we can consider this t.f. at DC (0 frequency).

- Feedback causes the transfer function from the reference input to the output to be insensitive to variations in the gains in the forward path of the loop
- At frequencies sufficiently less than f_c, i.e., for $|L(j\omega)| >> 1$:

$$\left|\frac{\tilde{v}_O(j\omega)}{\tilde{v}_{ref}(j\omega)}\right| = |H(j\omega)|\left|\frac{L(j\omega)}{1+L(j\omega)}\right| \approx |H(j\omega)|$$

- This result applies equally well to dc values:

$$\frac{V_O}{V_{ref}} = H(0) = \left(1+\frac{Z_1(0)}{Z_2(0)//R_{FB}}\right)$$

Equation x is the equation that will be used to set the value of v_o once v_ref is set.

Transfer functions of some CCM converters

Table 8.2. Salient features of the small-signal CCM transfer functions of some basic dc-dc converters

Converter	G_{g0}	G_{d0}	ω_0	Q	ω_z
buck	D	$\frac{V}{D}$	$\frac{1}{\sqrt{LC}}$	$R\sqrt{\frac{C}{L}}$	∞
boost	$\frac{1}{D'}$	$\frac{V}{D'}$	$\frac{D'}{\sqrt{LC}}$	$D'R\sqrt{\frac{C}{L}}$	$\frac{D'^2R}{L}$
buck-boost	$-\frac{D}{D'}$	$\frac{V}{D\,D'}$	$\frac{D'}{\sqrt{LC}}$	$D'R\sqrt{\frac{C}{L}}$	$\frac{D'^2R}{D\,L}$

where the transfer functions are written in the standard forms

Note: output capacitor with negligible ESR

$$G_{oc}(s) = G_{do}\frac{\left(1-\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{Q\omega_0}+\left(\frac{s}{\omega_0}\right)^2\right)}\cdot G_M(s) \qquad G_{ol}(s) = G_{go}\cdot\frac{1}{1+\frac{s}{Q\omega_0}+\left(\frac{s}{\omega_0}\right)^2}$$

Goc(s), the control to output tranfer function, is written as the product of the PWM t.f. Gm(s) and the t.f. between duty cycle and output. This t.f. can be written in a form that is general for the buck, the boost and the buck-boost, we have to change Gdo (zero frequency gain of the t.f.), w_z, w0 and Q.
The table reports the values for these parameters to have the t.f. rewritten.

If we consider a buck converter, w_z is infinite, so there is no zero at the numerator. However, if we consider a boost or buck-boost, the w_z is finite, meaning that those converters working in CCM show a real positive zero. And this real positive zero (RHPZ) is a pain in the back, because it gives a contribution to the phase of -90°.

NB: we are also assuming that the output capacitor is perfectly ideal, so no ESR. In case of ESR there would have been a negative zero.

As for Gol, it can be written for any converter as Ggo, zero frequency gain, divided by the second order polynomial. Once again, Ggo depends on the kind of converter. In the table, V indicates v_o, the output voltage.

Effect of a RHP zero on the step response

Consider a second order transfer function with a zero:

$$H_z(s) = \frac{\left(1-\frac{1}{\omega_z}s\right)\cdot\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2}$$

Note: DC gain = 1

H(s) is the transfer function of the original system (without the zero)

We can split the above transfer function into the sum of two terms:

$$H_z(s) = \frac{\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2} - \frac{1}{\omega_z}s\cdot\frac{\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2} = H(s) - \frac{1}{\omega_z}sH(s)$$

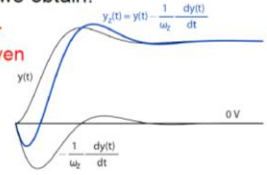
Denote the Laplace transform of the step response for this system by $Y_z(s) = H_z(s)1/s$.
Using the above decomposition of $H_z(s)$, we obtain:

$$Y_z(s) = \left(H(s)-\frac{1}{\omega_z}sH(s)\right)\frac{1}{s} = H(s)\frac{1}{s} - \frac{1}{\omega_z}sH(s)\frac{1}{s} = Y(s) - \frac{1}{\omega_z}sY(s)$$

Noting that the inverse Laplace Transform of sY(s) is $\dot{y}(t)$, we obtain:

$$y_z(t) = y(t) - \frac{1}{\omega_z}\dot{y}(t)$$

The step response of the second order system with a RHP zero at $s = \omega_z$ is given by the step response of the original system minus a scaled version of the derivative of the step response of the original system.

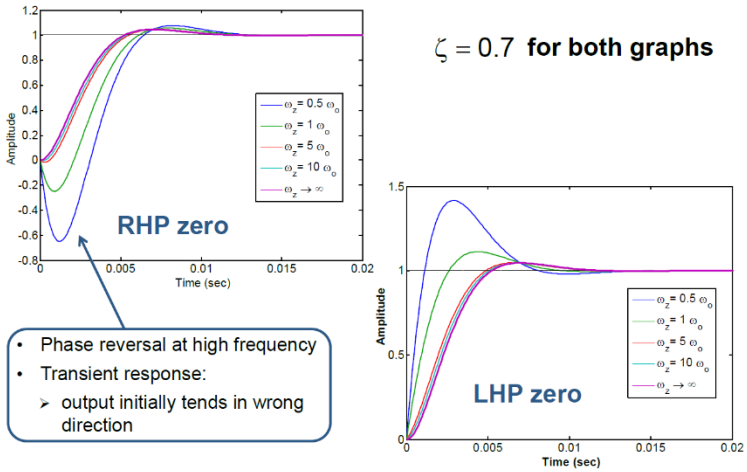


If we have a RHP zero and we apply a step at the input, the output firstly goes in a direction, and then in the opposite one, so we get a bipolar response.
So let's consider a second order system with a DC gain of 1 and a RHP zero. We can split the t.f. in two contributions. H is the transfer function of the original system but without the zero.

The Laplace transform of the step response of the original system is simply Yz(s) = Hz(s)/s, and again it can be written as the sum of two contributions. Y(s) is the Laplace transform of the step response of the original system without the zero.

If we apply the reverse Laplace transform we get that, in the time domain, the step response of the system with the RHP zero is equal to the superposition of the step response of the system without the zero and the negative scaled version of the time derivative of the step response of the system without the zero.

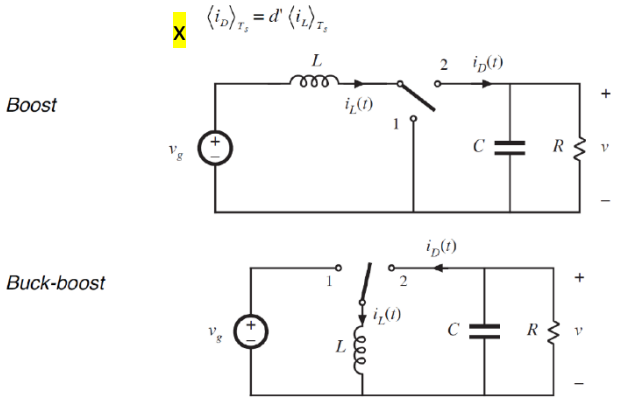
The responses are similar to the ones below.



If instead the zero is not negative but positive (LHP zero), the response is unipolar.

ZERO INSPECTION

We want to look at the circuit and understand if a zero is present. For instance, let's consider a boost converter (but the same considerations can be done for a buck-boost converter).



In the boost converter the current flows in the diode just when the switch is off, so we can write the equation x, where d' = 1 – d.

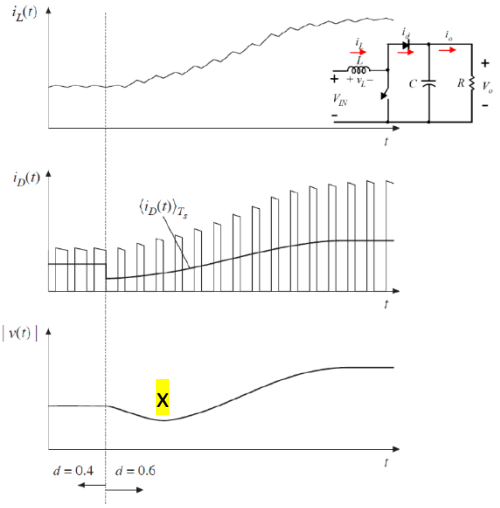
If we assume to increase the duty cycle as a step, e.g. from 0.4 to 0.6, what happens to the diode and inductor currents?
The diode current, upon the application of the step change of the duty cycle, displays a reduction of the width of the trapezoid, because the width is 1 – d. However, the top level of the trapezoid is the inductor current. But the inductor current initially will stay unchanged, it cannot change instantaneously, so for

few cycles the height of the trapezoid will remain the same. After few cycles it will increase. The result is that the running average diode current (black line) will initially decrease, because the base is shrinking and the moving average is the area of the rectangle divided by the switching period.

So initially the average diode current is decreasing, but the average diode current is also the current that we are pushing to the output node, so we have an initial transient situation where there will be some current drawn from the capacitor, producing a reduction of the output voltage x.

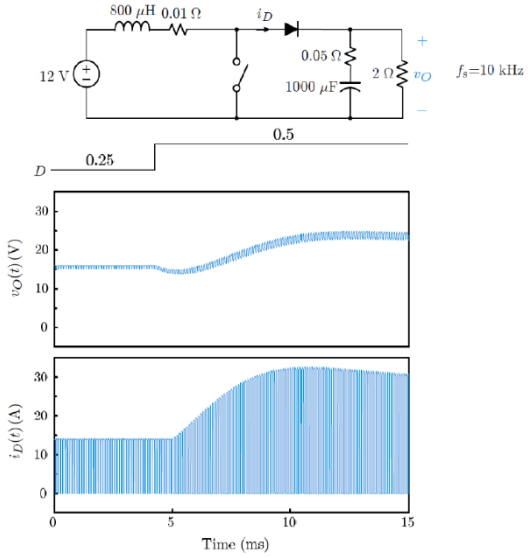
$$\langle i_D \rangle_{T_s} = d' \langle i_L \rangle_{T_s}$$

- Increasing $d(t)$ causes the average diode current to initially decrease
- As inductor current increases to its new equilibrium value, average diode current eventually increases



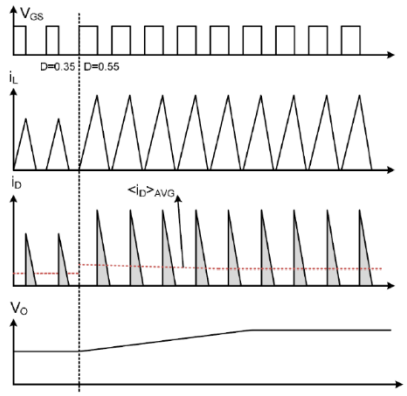
Then, step after step, the inductor current increases going to the steady state value for the new rectangle and so the output voltage will increase to the new final state. So we have a bipolar response in the output voltage, which is the signature of the presence of a RHP zero.

Simulation example



VOLTAGE MODE CONTROL IN DCM

- Single pole transfer functions
- Right half plane zero does not exist for boost and buck-boost



CONTROL-TO-OUTPUT TRANSFER FUNCTIONS OF BASIC CONVERTERS (DCM, VOLTAGE-MODE CONTROL)

Topology	Control-to-Output Transfer Function
Buck	$\frac{\hat{v}_o}{\hat{v}_{EA}} = \frac{2V_O}{V_P \cdot D} \cdot \frac{1-M}{2-M} \cdot \frac{1}{1+s/\omega_p}$ $\omega_p = \frac{2-M}{1-M} \cdot \frac{1}{R_L C_O}$, $M = \frac{V_O}{V_{IN}}$
Boost	$\frac{\hat{v}_o}{\hat{v}_{EA}} = \frac{2V_O}{V_P \cdot D} \cdot \frac{M-1}{2M-1} \cdot \frac{1}{1+s/\omega_p}$ $\omega_p = \frac{2M-1}{M-1} \cdot \frac{1}{R_L C_O}$, $M = \frac{V_O}{V_{IN}}$
Buck-boost	$\frac{\hat{v}_o}{\hat{v}_{EA}} = \frac{V_O}{V_P \cdot D} \cdot \frac{1}{1+s/\omega_p}$ $\omega_p = \frac{2}{R_L C_O}$, $M = \frac{V_O}{V_{IN}}$

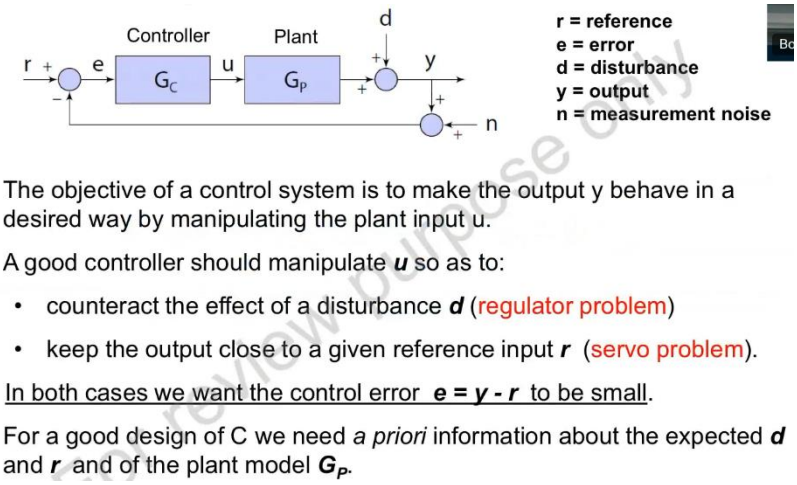
We won't model the converter working in DCM, but the approach is the same used for CCM, separating the switching network and modelling it.

In DCM, as a conclusion, and we have a voltage mode control system, the buck converter, the boost and buck-boost converters show a control to output t.f. having a single pole, not a c.c. pair of poles. So we don't have a second order system but a first order one. This because the two c.c. poles are split and one of them is sent to very high frequency where it has no impact on the dynamic behaviour of the converter.

Another advantage is that, if we look at the boost and buck-boost and at their t.f., we don't see any RHP zero. Also the zero doesn't disappear, but it is pushed to very high frequency.

COMPENSATOR DESIGN

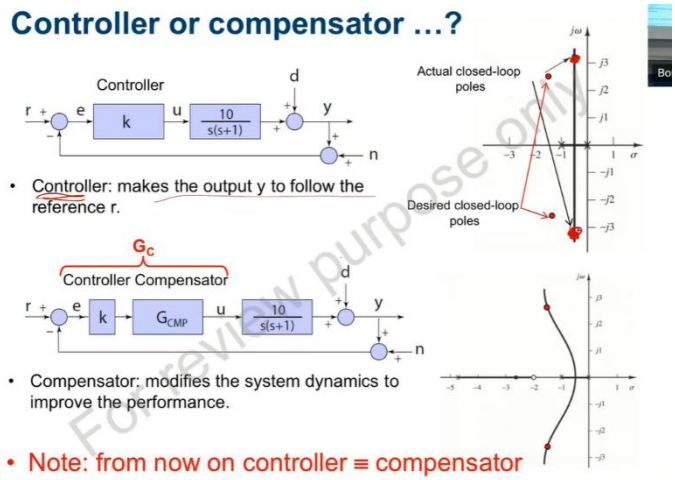
Recap – the control problem



- The objective of a control system is to make the output y behave in a desired way by manipulating the plant input u.
- A good controller should manipulate ***u*** so as to:
  - counteract the effect of a disturbance ***d*** (regulator problem)
  - keep the output close to a given reference input ***r*** (servo problem).
- In both cases we want the control error ***e = y - r*** to be small.
- For a good design of C we need *a priori* information about the expected ***d*** and ***r*** and of the plant model G_P.

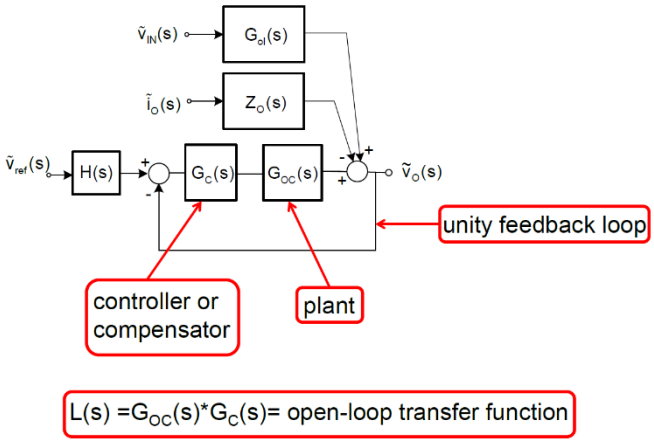
The objective of the control is to make the output to behave in a desired way by manipulating the plant input.
The difference between a regulator and a servo system is that in a regulator the controller has to be designed to counteract the effect of the disturbance and the reference input is a constant voltage, the only input is the disturbance. In the servo problem the reference is changing and we want the output to follow the reference. In both cases we need to reduce the error.

Controller or compensator?



COMPACT BLOCK DIAGRAM REPRESENTATION

Gc is the compensator and Goc the plant.



TYPICAL SPECIFICATIONS FOR A CLOSED LOOP SYSTEM

Static (steady-state) performance:

- **Desired behavior w.r.t. order k inputs in terms of a maximum absolute error**. This means that if we want a zero steady state error upon the application of a step reference signal (type 0 reference signal, we need a type 1 loop transfer function (one pole in the origin). This because pole in the origin means infinite amplitude at 0Hz and so zero amplitude of the static error.
- **Desired attenuation level w.r.t. constant disturbances acting on the forward loop**. In general if we have disturbances in a frequency that is inside the CL BW we want to attenuate them.
- **Tracking of a sinusoidal reference**. This is typical of servo systems.
- **Attenuation of sinusoidal disturbances and measurement noise**. In our case the noise comes from the output ripple that is injected back into the loop. We want to attenuate this because if we inject the ripple, the ripple is a signal that is changing at the switching frequency and this signal goes to the PWM, which is sampling at the switching frequency, so we might end up in a folding of the noise. To attenuate the HF noise we need to be sure that the magnitude of the loop gain at the switching frequency will be the smallest possible one → loop gain as high as possible in LF and as low as possible in LF.

Dynamic performance:

- Time domain specifications on the step response (mainly on the reference to output behavior).
- Frequency domain specifications through the resonance peak and the bandwidth (mainly on the reference to output behavior).

The dynamic performances deal with the loop t.f. around the crossover frequency.

Stability

- **Nominal stability**: the system is stable with no model uncertainty (closed loop poles have negative real part).
- **Robust stability**: the system is stable for all perturbed plants about the nominal model up to the worstcase model uncertainty (e.g. gain and phase margin are satisfactory).

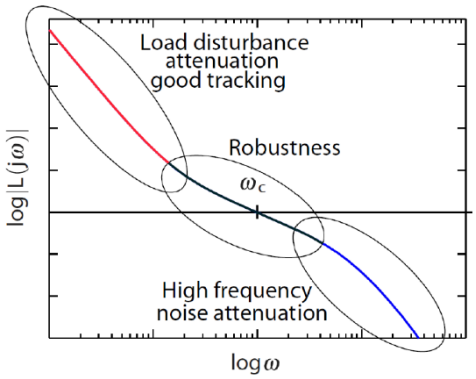
APPROACH: LOOP SHAPING

The basic idea is to convert the closed loop specifications into open loop specifications. Since the loop t.f. is the product of the plant t.f., that we know, and of the compensator t.f. we need to design, we can easily design the compensator t.f..

- Basic idea: convert specs on closed loop to specs on open loop
  → desired shape of L(s)

Closed-loop requirement →	Constraint on loop transfer function
Closed-loop stability	L(jω) verifies the Nyquist criterion
Steady state error	Gain and/or type
Disturbance attenuation on output	$\|L(j\omega)\|$ >>1 before crossover frequency
Noise attenuation on output	$\|L(j\omega)\|$ <<1 after crossover frequency
Settling time of step response	Crossover frequency (and phase margin)
Damping of step response	Phase margin
T(jω) has "sufficiently high" bandwidth	Crossover frequency (and phase margin)
T(jω) doesn't have high resonant peak	Phase margin

The table shows how to translate the closed-loop requirements into open loop ones.
In the following image there is the typical shape of the loop we want to get in order to satisfy the previous requirements. We notice the loop t.f. magnitude is high at 0 frequency and low at high frequencies. Instead, the behaviour of the loop t.f. around the crossover frequency determines the robustness of the system and its dynamic performances.

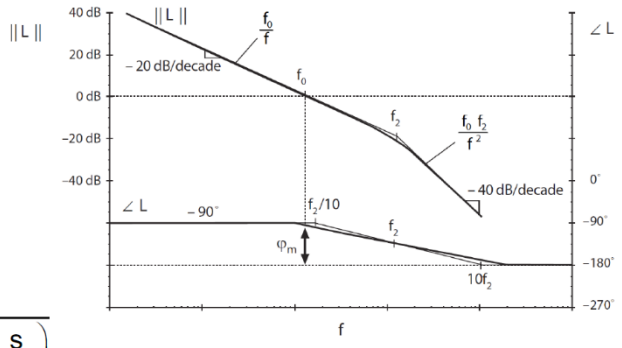


- Stability, transient performance, sensitivity to modeling inaccuracies heavily depend upon the L(jω) shape in the crossover region.

Typical loop shape for CCM voltage-mode regulators

L(s) can be well approximated in the vicinity of the crossover frequency as:

$$L(s) = \frac{1}{\left(\frac{s}{\omega_o}\right)\cdot\left(1+\frac{s}{\omega_2}\right)}$$



This is the loop shape we want to target in a CCM converter.

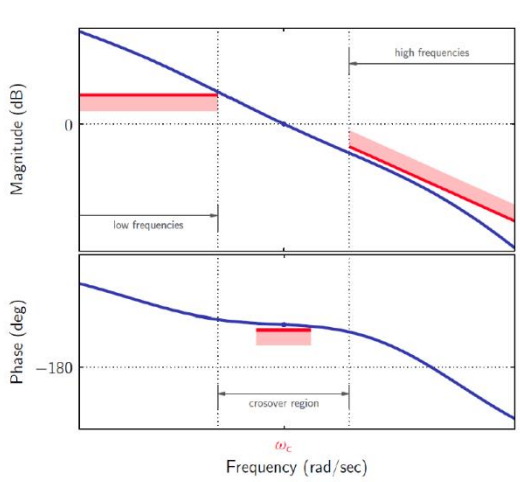
The ideal shape of the loop t.f. has a pole in the origin, so a roll off with -20dB/dec until we cross the 0dB axis and a second pole to further decrease the magnitude of the loop t.f. to attenuate the switching ripple.

The phase margin of this system is strictly related to the difference between f2 and fc.

Loop shaping – example

This is the typical loop shape that we want to target. The crossover frequency is typically located at frequencies between $1/10^{th}$ of the switching frequency and $1/5^{th}$ of the switching frequency.
The other thing we have to do is to increase the low frequency magnitude of the loop t.f. and this is done by including a pole in the origin (hence the loop t.f. is a type 1 t.f. having an integrator). The third thing we are interested in is increasing the phase margin, which has to be typically around 60° to 70°.

The last important thing is to decrease the HF gain for better noise suppression, and this is typically done by including an additional pole in the loop t.f. which is located at half the switching frequency usually.



We might want to:

- set required crossover freq. ω_c
  - typ. $\omega_s/10 < \omega_c < \omega_s/5$.
- increase low-frequency gain
  - e.g. pole at ω=0 for zero steady-state error with constant reference signal
- increase phase around ω_c
  - nominal φ_m of 60 to 70°
  - worst case φ_m of 20° to 30°
- decrease high-frequency gain for better noise suppression
  - additional pole at $\omega_s/2$ typ.

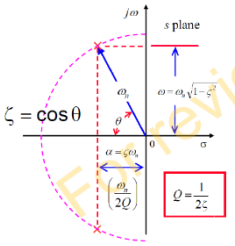
❑ Additional constraint: $\omega_c < \omega_{RHPZ}/2$ id the plant has a RHP zero at ω_{RHPZ}

If the system we are controlling is not a minimum phase system, so we have a RHP zero, things are more complex. In this case we have an additional constraint, that is that the crossover frequency must be occurring at a frequency below the frequency of the RHP zero. Typically, $1/5^{th}$ of the RHP zero frequency below.

REVIEW – CLOSED LOOP RESPONSE

If $L(s)=\dfrac{1}{\left(\dfrac{s}{\omega_o}\right)\cdot\left(1+\dfrac{s}{\omega_2}\right)}$, then $T(s)=\dfrac{L(s)}{1+L(s)}=\dfrac{1}{1+\dfrac{s}{\omega_o}+\dfrac{s^2}{\omega_o\cdot\omega_2}}$

or $T(s)=\dfrac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$, where $\omega_n=\sqrt{\omega_o\cdot\omega_2}$ and $\zeta=\dfrac{\omega_n}{2\omega_o}$



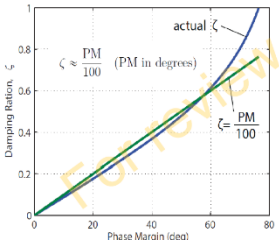
- When the phase margin is small, the closed-loop system is close to instability, so that there will be closed-loop poles near the jω-axis. That is, low PM → low damping ratio.

Starting from the loop t.f. L(s), the closed loop t.f. is L(s)/(1 + L(s)), which is a second order t.f.. This t.f. is not the t.f. we have in the regulator, because in the regulator there is a H(s) block before the L(s) and the unity feedback.
The two CL poles depends on the damping factor and on the natural frequency omega_n.

Phase margin vs. damping ratio

- Calculate $|L(j\omega)|$ at crossover frequency and express ω_c/ω_n as a function of ζ. Evalutate φ_m as a function ω_c/ω_n and replace. Result is:

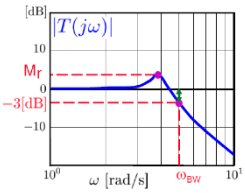
$$\varphi_m = a\tan\left(\frac{2\zeta}{\sqrt{\sqrt{1+4\zeta^4}-2\zeta^2}}\right) \quad \text{or} \quad \zeta=\frac{1}{2}\frac{\sin\varphi_m}{\sqrt{\cos\varphi_m}}$$

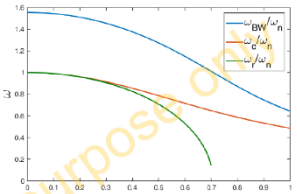


$$\zeta \approx \frac{\varphi_m}{100} \ (\varphi_m \text{ in degrees})$$

approximated formula for φ_m<60°

Impact of φ_m on closed-loop frequency response

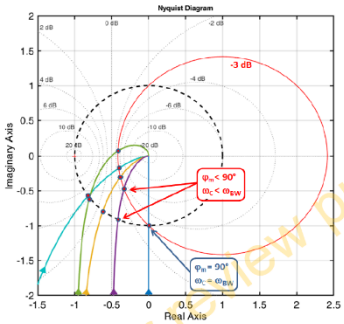




- Closed-loop resonant peak $M_r=\dfrac{1}{2\zeta\sqrt{1-\zeta^2}}\approx\dfrac{1}{2\sin(\varphi_m/2)}$
- Closed-loop resonance frequency $\omega_r=\omega_n\sqrt{1-2\zeta^2}$
- Closed-loop bandwidth $\omega_{BW}=\omega_n\sqrt{1-2\zeta^2+\sqrt{2-4\zeta^2+4\zeta^4}}$
- Crossover frequency (loop function) $\omega_C=\omega_n\sqrt{\sqrt{1+4\zeta^4}-2\zeta^2}$

NB: the BW is not the crossover frequency of the OL t.f.. They are coincident only if the phase margin is 90°.

Impact of φ_m on closed-loop bandwidth



$$\zeta \approx \frac{\varphi_m}{100} \ (\varphi_m \text{ in degrees})$$

$$\omega_{BW}=\omega_n\sqrt{1-2\zeta^2+\sqrt{2-4\zeta^2+4\zeta^4}}$$

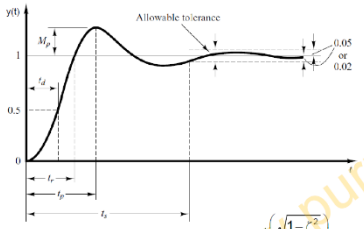
$$\omega_C=\omega_n\sqrt{\sqrt{1+4\zeta^4}-2\zeta^2}$$

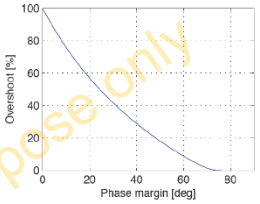
$$\omega_{BW}=\omega_C\frac{\sqrt{1-2\zeta^2+\sqrt{2-4\zeta^2+4\zeta^4}}}{\sqrt{\sqrt{1+4\zeta^4}-2\zeta^2}}$$

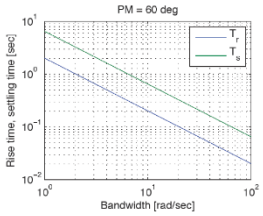
- In general, the closed-loop bandwidth ω_{BW} is close to the crossover frequency ω_C. A rule of thumb is that :

$1.2\ \omega_C < \omega_{BW} < 1.5\ \omega_C$

Closed-loop transient performance



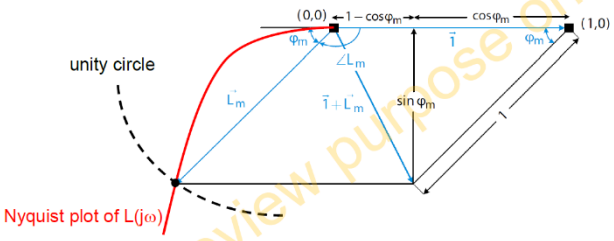




- 0-100% rise time $t_r=\dfrac{\pi-\tan^{-1}\left(\dfrac{\sqrt{1-\zeta^2}}{\zeta}\right)}{\omega_n\sqrt{1-\zeta^2}}$
- Settling time (5%) $t_s=\dfrac{3}{\zeta\omega_n}$
- Time to peak $t_p=\dfrac{\pi}{\omega_n\sqrt{1-\zeta^2}}$
- Peak overshoot $M_p=e^{\frac{-\pi}{\sqrt{1-\zeta^2}}}$

Exact analysis at crossover frequency

- Evaluation of $|1+L(j\omega)|$ at crossover frequency:

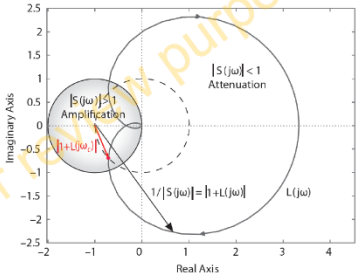


$$|1+L(j\omega_C)| = \sqrt{\sin^2\varphi_m + (1-\cos\varphi_m)^2} = \sqrt{2-2\cos\varphi_m} = 2\sin\left(\frac{\varphi_m}{2}\right)$$

Impact of φ_m on closed loop TFs

- Generic closed loop TF: $G_{CL}(s) = \frac{G_{OL}(s)}{1+L(s)} = G_{OL}(s)\cdot S(s)$
- At $\omega = \omega_C$:

$$|G_{CL}(j\omega_C)| = |G_{OL}(j\omega_C)|\cdot|S(j\omega_C)| = \frac{|G_{OL}(j\omega_C)|}{|1+L(j\omega_C)|} = \frac{|G_{OL}(j\omega_C)|}{2\sin(\varphi_m/2)}$$



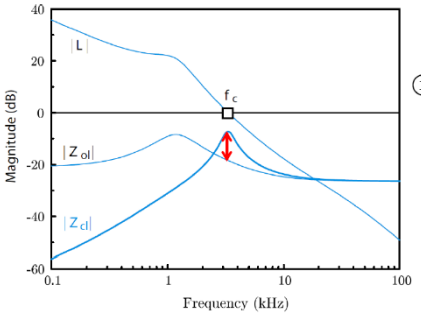
We want the magnitude of 1 + L(s) at the crossover frequency, because it is strongly determining the dynamic performances of the system. We can understand the magnitude by looking at the blue plot. The red line is the Nyquist plot of the loop t.f. and Lm is the vector L.

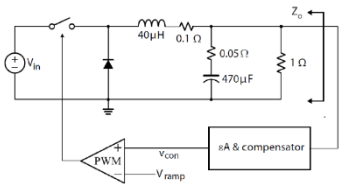
The impact is relevant on the CL t.f. we are interested in, e.g. the line t.f. or the output impedance. The generic OL t.f. is divided by 1 + L(s) when we close the loop. 1/(1 + L(s)) is called sensitivity function. Around the crossover region, the magnitude of the CL t.f. is the product of the magnitude of the OL t.f. and of the sensitivity function. The impact of this can be seen considering a graphical interpretation, the one in the right image is the Nyquist diagram. The dashed circle is the unity circle centered in the origin, while the gray one is centered in the critical point.
If the Nyquist plot crosses the unity circle within the gray region, the value of the vector 1 + L(jw_c), that is the vector joining the critical point and the crossover point, is smaller than 1, so we are amplifying the OL t.f.. If instead the unity circle is crossed outside the gray region, the OL t.f. is attenuated.

It is very important not to amplify the OL t.f., otherwise we get an oscillating response. The limiting situation is when the Nyquist plot of the loop t.f. crosses the unity circle exactly in the border of the gray circle. In this situation we don't have neither an amplification nor an attenuation. The phase margin corresponding to this critical situation is 60°.

EXAMPLE: IMPACT OF PHASE MARGIN ON CL Zo

$$|Z_{o_{CL}}(j\omega_C)| = \frac{|Z_{o_{OL}}(j\omega_C)|}{|1+L(j\omega_C)|} = \frac{|Z_{o_{OL}}(j\omega_C)|}{2\sin(\varphi_m/2)}$$



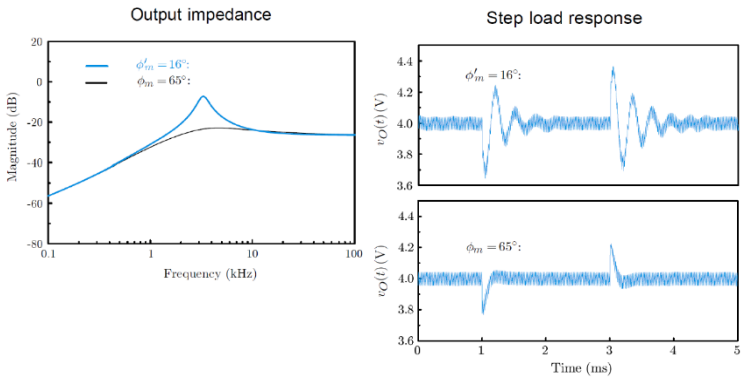


Example: buck regulator

$$|Z_{peaking}(j\omega_C)| = \frac{|Z_{o_{CL}}(j\omega_C)|}{|Z_{o_{OL}}(j\omega_C)|} = \frac{1}{2\sin(\varphi_m/2)}$$

The output impedance at the crossover is closed loop. If we consider for instance a buck converter, we notice that close to the crossover frequency we have an overpeaking in the CL output impedance. The

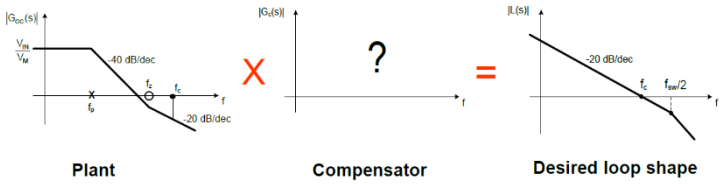
presence of this peak is the reason for an oscillating transient response as in the images below. The blue line is an example of 60° phase margin. With a 65° phase margin the step response is almost ideal.



These considerations are extremely important, because if we assume for instance that we are biasing the core of a microprocessor, a step load response with ringing leads to a 'blue screen' on the computer or a blowing of the microprocessor.

FROM LOOP SHAPING TO COMPENSATOR DESIGN

- The *converter* we want to compensate has an open-loop transfer function $G_{oc}(s)$
- On this plot, a crossover frequency is identified, f_c
- The designer reads the gain and the phase of $G_{oc}(j\omega)$ at f_c
- A *compensator* transfer function $G_c(s)$ is inserted so that it:
  - provides gain/attenuation at the crossover frequency: $|G_{OC}(j\omega_c)\cdot G_C(j\omega_c)|=1$
  - *boosts* the phase at the crossover: $\angle G_{OC}(j\omega_c)+\angle G_C(j\omega_c)+180°=\varphi_m$
  - provides pole at f=0
  - provides an additional pole at $f_c < f < f_{sw}$



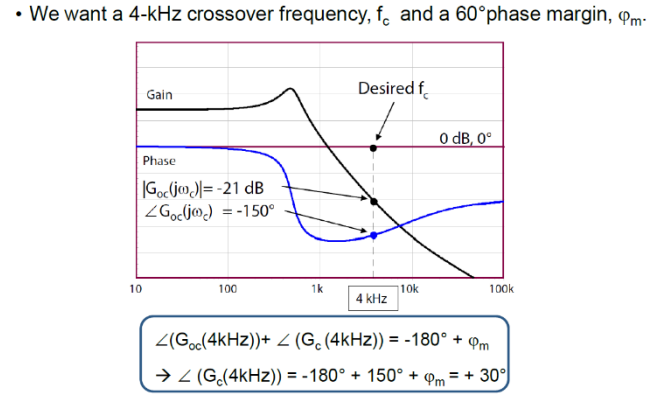
Let's assume a buck converter for the sake of simplicity, since the control to output t.f. is flat until the appearance of a c.c. pair of poles, and then we have a negative zero coming from the ESR of the capacitor. The starting point is the plant t.f., we want to compensate this t.f. with a compensator t.f., which is still unknown at this point, to get the desired loop shape.

We start by identifying the crossover frequency in the plant t.f., for instance $1/10^{th}$ of the switching frequency. The magnitude of the plant t.f. at the switching frequency is negative and so the first thing to do is to add a gain to the system, the compensator has to provide a gain to push fc up so that the final loop crosses the 0dB axis at fc.

The second thing we have to read is the phase of the plant t.f. at fc (crossover frequency). Once this is done, the compensator must provide a gain (or attenuation) at fc such that the product of the gain of the plant and the gain of the compensator is 1 at fc.
The other thing the compensator has to provide is a boost in the phase at the crossover, to get the desired phase margin.

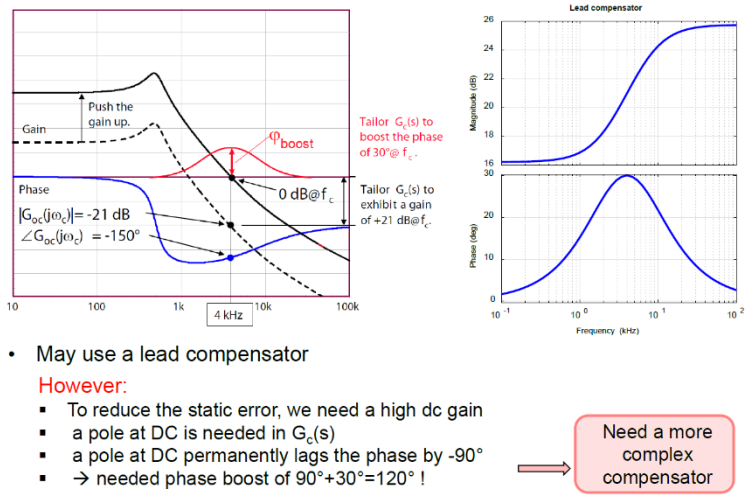
Then there are two additional requirements; we have to provide a pole at f = 0, to increase the magnitude of the loop gain at LF, and an additional pole between fc and the switching frequency fsw, usually at half fsw.

Example



It's a trial and error approach. Let's assume we have a buck converter and the one in the image above is the plant t.f..
As a first move we need to read the gain of the plant t.f. at fc, which is -21dB, so we have to provide a positive gain to compensate for this negative gain, and the second important thing is the phase, which is -150°. If we simply push up the t.f. without adding a phase boost we would end up with a phase margin of 30°, which is too small, so we need to add a phase boost around the fc in a way that the target phase margin is 60°.
The equation we have to consider is the one in the blue box, which says that the phase of the plant at the fc plus the phase of the compensator at fc must be the desired phase margin phi_m -180°.
In the end we get a phase boost that has to be 30°.

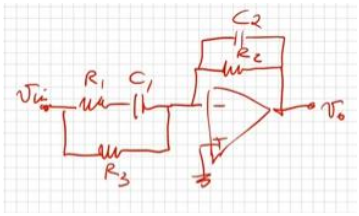
How can we practically boost the phase?
We need to put somewhere a negative zero to increase the phase. In principle we can think of using a lead compensator, which is a circuit providing a t.f. where the magnitude is the one on the right, with a phase that has a bell shape.



Then we have to properly design the position of the zero and of the pole in the lead compensator to get the desired phase boost (30° in the previous example), and also the value of the gain to provide the desired gain.

Lead compensator implementation

It is a proportional-derivative (PD) compensator (if we look at the gain behaviour). A possible implementation is the following.



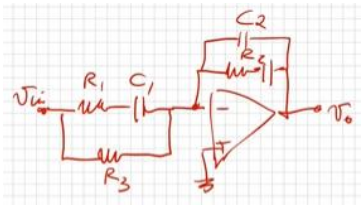
We can compute the t.f.

$$\frac{V_o}{V_{in}} = -\frac{R_2}{R_3}\,\frac{1+sC_1(R_1+R_3)}{(1+sR_2C_2)(1+sC_1R_1)}$$

We have a DC gain, which is needed because we need to push the plant gain, and then there is a zero and a pole, because the second pole can be pushed to HF and we can get rid of it. We can tune R1, R2, C1 and C2 to place the pole and the zero where we need.

Coming back to the lead compensator, using it we are boosting the phase making it equal to the target one, and we can boost the gain. However, we are not providing a pole in the origin to push the gain at LF towards infinite. If we are adding a pole in the origin we are also adding a -90° phase contribution, so we cannot add simply an integrator to the PD.
To the PD we add a second block that might be a PI, something that has a t.f. with a pole in the origin and a zero somewhere. The zero has to be inserted not to reduce the phase margin at the crossover frequency.

To have a PI we can add a simple capacitor in the R2 branch.



Because of the additional pole in the origin that adds -90° to the phase margin, the phase boost needs to be higher, in our case example 90° + 30° = 120°.

However, the standard approach is another one, based on standard compensators.

GENERIC COMPENSATORS

They are of 3 different types: I, II and III.

Type I compensator

It is a pure integrator, and because of this it adds also a phase lag of -90°.

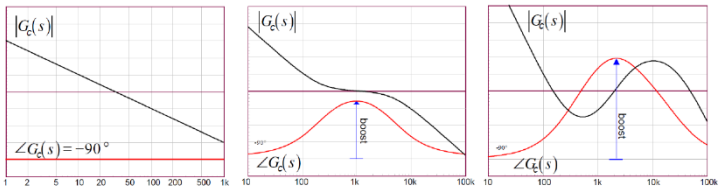
Type II compensator

It has one pole in DC, one zero and one pole after the zero. It can boost the phase but up to 90°. The phase is boosted with respect to the starting negative -90°, so we can recover the 90° lag introduced by the pole in the origin. It can be applied to plants and converter which have an already satisfactory phase margin in open loop.

Type III compensator

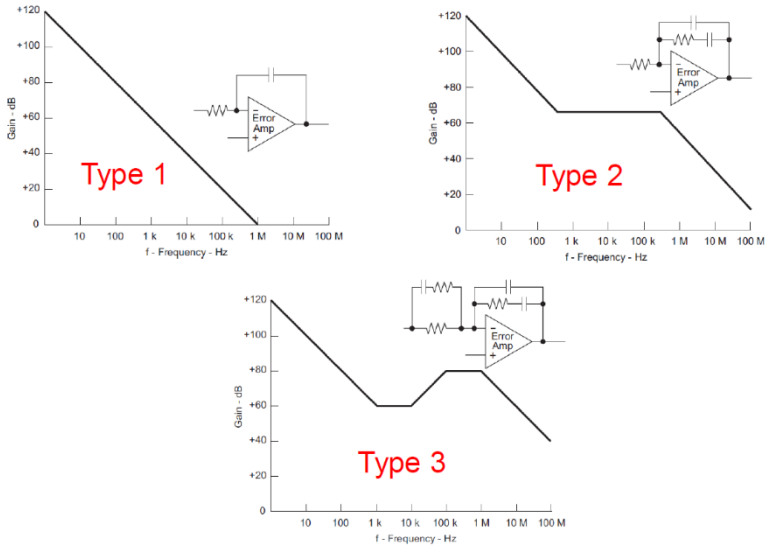
It has a pole in DC, a pair of real zeros and then also a pole pair. The phase boost we can get with this compensator is up to 180°, so not only we recover the 90° lag due to the pole in the origin, but we can also add further 90°. This is useful if the phase margin of the original system with no compensator is below the target phase margin.

- In most cases, Type I, Type II or Type III compensation network will properly compensate the system. The compensator type is determined by the amount of phase boost needed at the crossover frequency.
- All the compensators have apole at the origin which introduces a 90° phase lag.



- **Type 1**: 1 pole at DC, no phase boost
- **Type 2**: 1 pole at DC, 1 zero, 1 pole. Phase boost up to 90°
- **Type 3**: 1 pole at DC, 1 zero pair, 1 pole pair. Phase boost up to 180°

Networks



Generic type I compensator

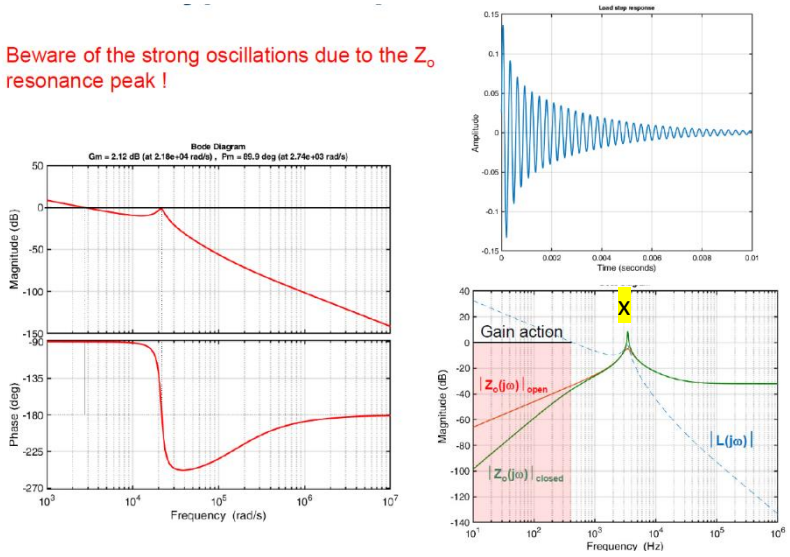
- It does not offer any phase boost.
- Type 1 compensator can be used in converters where the power stage phase shift is small, e.g., in an application where you would like to roll off the gain far away from the resonant frequency of a second-order filter.
- As in any integral type compensation, it brings the largest overshoot in the presence of a sudden load change.
- This type is widely used in PFC. In PFC we use very small BW, few Hz.

In the following image we have a buck converter which is compensated with a type I compensation. The red line is the loop t.f., which crosses the 0dB axis more or less at 2.5*10^3 rad/s.
Below we have the phase plot, and the phase margin at fc is quite large, 90°. However, the circuit is not ok, because the step load response of the system is shit, it has an absurd ringing, even though we have a phase margin of 90° and the system is stable.

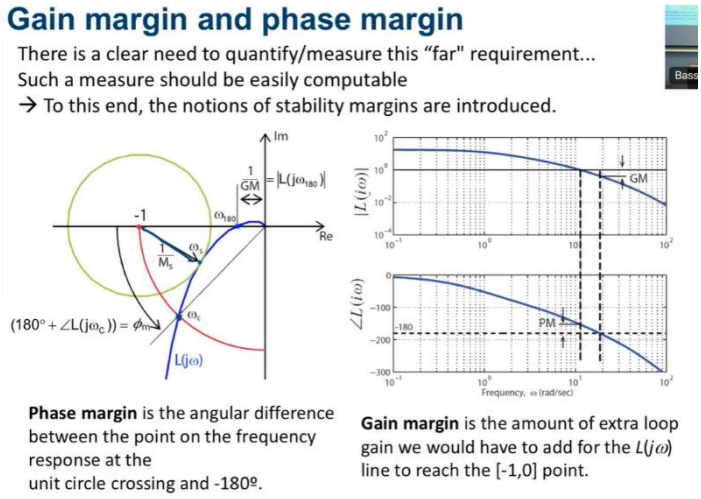
This is due to two reasons:

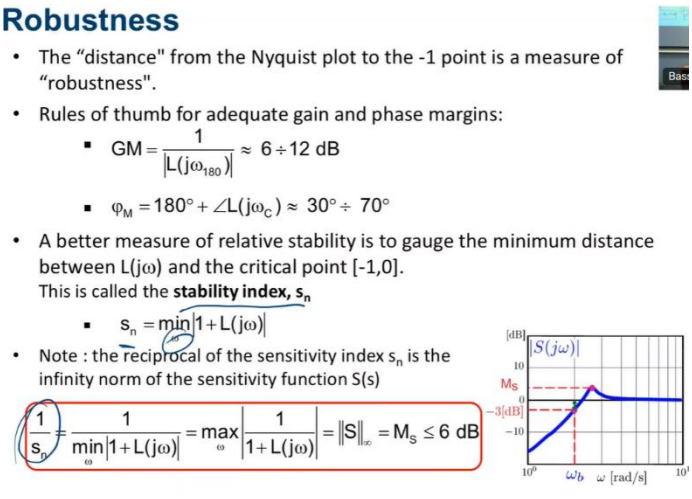
1. We are not taking the output impedance bump into the closed loop bandwidth. Looking at the bottom right plot, the dashed line is the loop t.f., and the Zo,cl is nothing else than Zo,ol/(1 + L(s)). As expected, we are attenuating the magnitude of Zo,cl with respect to the Zo,ol in the red region, which is the region where the magnitude of the loop gain is much smaller than 1. However, as soon as we get out of this window we don't have attenuation anymore, and Zo,cl = Zo,ol. Hence if we are exciting the system with a step current variation we get the effect of the resonance peak x, which produces the ringing.

   This is the reason why, in general, when closing the loop, we have to pay attention to keep the resonance peak of the output impedance inside the bandwidth, so that we are attenuating the output impedance. If we don't do this, we have attenuations at the output even if the system is perfectly stable.



2. The gain margin. If the gain margin is small, the oscillations are stronger, i.e. the damping factor is reduced.

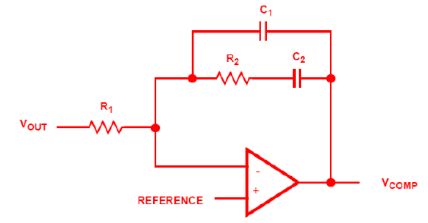




So for the type I we have to cut the 0dB axis at frequencies sufficiently small to have good phase and gain margins. However, we will never be fully capable of eliminating the oscillations in the response. Usually type I is not used in regulators.

Generic type II compensator

Widely used in DC/DC converter. It is implemented using the red network, adding a capacitor in series to the resistance in feedback. The blue one is the compensator t.f.; the mid frequency gain is computed with C1 open and C2 shorted.
The compensator t.f. Gc is define as the negative ratio between v_c at the output and v_o, because in our model the summing node was moved outside the compensator.
Hence Gc is the superposition of a summing node in cascade with a compensator. The summing node is the – sign and the compensator is producing the t.f. v_c/v_o.

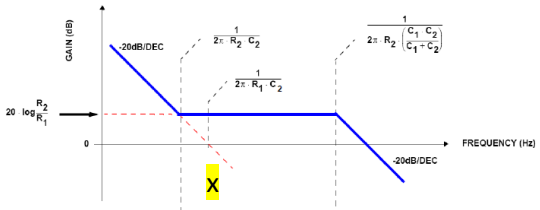


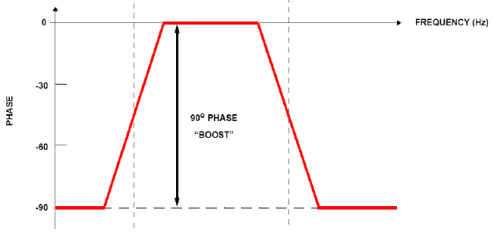
$$\text{note}: G_C(s) = -\frac{\tilde{v}_{COMP}(s)}{\tilde{v}_{OUT}(s)}$$

$$G_C(s) = \frac{1}{sR_1(C_1+C_2)} \cdot \frac{(1+sR_2C_2)}{\left[1+sR_2\left(\frac{C_1 \cdot C_2}{C_1+C_2}\right)\right]} =$$

$$\cong \frac{1}{sR_1C_2} \cdot \frac{(1+sR_2C_2)}{(1+sR_2C_1)} \quad \text{for } C_2 \gg C_1$$

$$G_C(s) = \frac{K_C}{s} \cdot \frac{\left(1+\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{\omega_p}\right)}$$





The Type II compensator provide up to 90° boost to the phase.

We are basically implementing a PI compensator plus an additional pole. The synthetic form for the t.f. of the compensator is in the yellow box. Kc is related to the crossing frequency at point x.

Simplified design and usage of the type II compensator

We want to define where to put the poles and the zeros of the compensator to compensate as best as possible the initial transfer function.
As general guidelines, let's start with a buck converter (black line t.f.) with a couple of c.c. poles and a negative zero due to the ESR. This is the t.f. to be compensated to have a desired loop shape, i.e. a loop t.f. with a pole in the origin, crossover at $1/10^{th}$ or $1/5^{th}$ of the switching frequency and then a second pole at half the switching frequency.

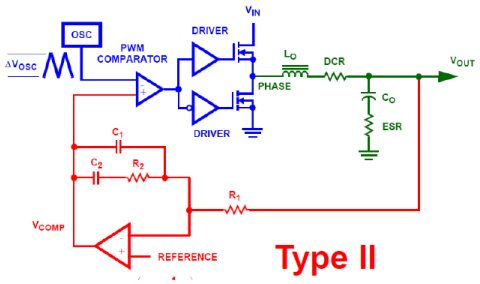
To do so, we take a type II compensator and we place the zero of the compensator at $1/10^{th}$ of the frequency of the c.c. poles (F_LC). The frequency of the pole is placed at half of the switching frequency fsw, the zero at $1/10^{th}$ of the c.c. poles and the mid frequency gain is regulated in order to compensate for the gain of the plant at the crossover frequency fc.
For instance, if we want to cross the 0dB axis at a certain frequency, we measure the plant gain at that frequency and correspondingly we provide a positive gain to get the loop t.f. to cross the 0dB axis in that point.

The shape of the loop t.f. (blue one) is not exactly ideal, we have a -20dB/dec rolloff, then a region with constant gain, a -40 rolloff and then -20 again. This means that we are introducing a zero-pole doublet in bandwidth. Its effect is to have a step response with a step at the beginning and then a slow component *(see Lacaita's notes)*. Here it is not a problem because we are not interested in t.f. between the reference and the output, because the reference is fixed.

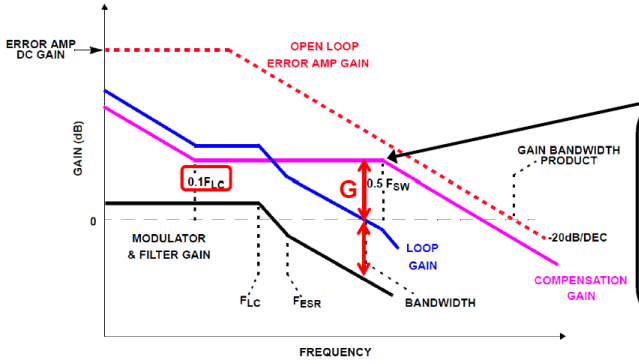
However, **pole-zero doublets can have a strong impact on the line to output t.f. and in the output impedance**.

- $f_z=0.1 \cdot f_{LC}$
- $f_p=f_{SW}/2$
- Mid frequency gain, G is calculated using:

$$G = \frac{1}{|G_{co}(j\omega_c)|}$$



Type II compensation network



Such pole is commonly used to attenuate the gain of the compensator at high frequencies and prevent the propagation of switching harmonics through the feedback loop.

Numerical implementation

In the upper right plot, the blue is the starting t.f. of the buck converter, the Goc to be compensated. The LF gain is 20dB and then we have a couple of c.c. poles at 3.4kHz and also a zero around 20kHz. We want to target a fc = 80kHz ($1/5^{th}$ of the switching frequency).
The gain of the plant at 80kHz is -24.7dB, so we have to boost the gain of +24.7dB, so this will be the mid gain of the type II compensator. As for the phase, it is -106°, so if we don't do anything, the phase margin is pretty large, 74°. Hence type II compensator is more than enough to compensate this situation because the starting phase margin is already high.

So the frequency of the zero is placed at $1/10^{th}$ of the frequency of the c.c. pole, and the pole frequency at ½ of the switching frequency. Then we compute Kc; the mid frequency gain G is Kc over w_z. Hence Kc/w_z must be the reciprocal of the gain of the plant.

- Target crossover frequency f_c = 80 kHz. Switching frequency, f_s= 400kHz.

1) Set $\omega_z = \frac{\omega_{LC}}{10} = 2.13 \text{ krad/s}$

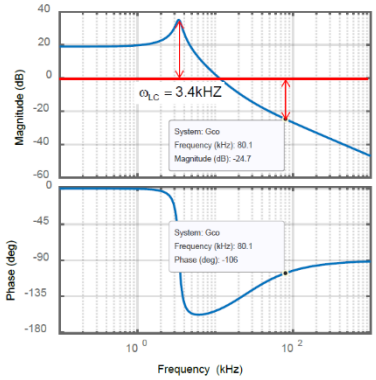
and $\omega_p = \frac{\omega_s}{2} = 1.25 \text{ Mrad/s}$

2) Calculate K_c:

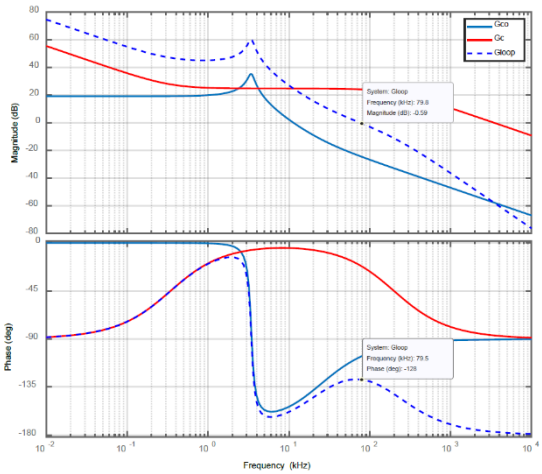
$$G = \frac{K_C}{\omega_z} = \frac{1}{|G_{co}(j\omega_c)|}$$

$$\rightarrow K_C = 36.7 \text{ krad/s}$$

$$G_C(s) = \frac{K_C}{s} \cdot \frac{\left(1 + s/\omega_z\right)}{\left(1 + s/\omega_p\right)}$$



The final result is in the following image, where we have the Bode plot of the Gco, starting plant t.f., the compensator t.f. Gc, and the loop gain Gloop. As expected, the Gloop crosses the 0dB axis at 80kHz and the phase margin that we have is still acceptable, 52°.



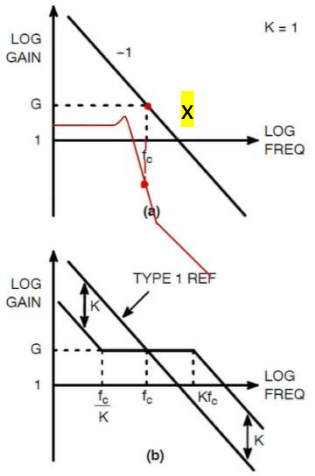
f_c= 80 kHz

φ_m= 52°

K factor design method

If we want to target a specific value of the phase margin, we need to tune the K factor. It is a method that provides us a tool to design the compensator so that the final phase margin is the one we are targeting.
Let's assume we are compensating a buck converter. The red curve is the magnitude of the control to output t.f. (Gco) of the buck. We want to set the crossover frequency at fc; the first thing to do is to evaluate the gain of the plant at that frequency and then apply a compensation gain which brings the crossover of the new compensated function at fc.

If we assume that we are using just a type I compensator, it is simply an integrator which as to intersect the 0dB axis in point x so that at fc the gain of the compensator balances the gain of the Goc.
The problem with this type I compensator is that it introduces a phase lag of -90°, and this is not good because around the fc the phase of Goc was already dropping toward -180°, and we are adding -90°, ending up in an unstable system.



- The K−factor is a measure of the reduction of gain at low frequencies and increase of gain at high frequencies achieved by controlling the location of poles and zeros of the compensation network in relation to the loop crossover frequency f_c.
- For type−I compensation network K is always 1.

To solve this issue, instead of using a type I compensator we can use a type II, introducing a zero and a pole. As we can see in the bottom plot, the zero is introduced at a frequency that is the fc/K, where K is a real number to be selected to obtain the desired phase margin. The pole is at a frequency given by K*fc. This because we want to put the maximum of the phase boost around the crossover frequency. The phase boost is in fact peaking at a frequency given by the geometric average between the zero frequency and the pole frequency, i.e. sqrt(fp*fz). In this way we get f_max = fc, with this sizing of the pole and zero frequencies.

The price to pay for including the zero and the pole to have a phase boost and a stable system with a desired phase margin is that the gain of the compensator at low frequency, and therefore the gain of the

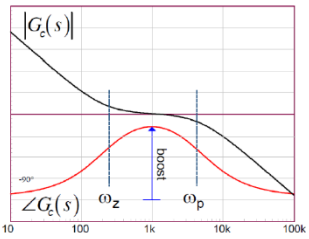
overall system when we put the compensator in series with the plant, is reduced, with respect to the type I compensator, by a factor K. At the same time, the gain is also increased by the same factor K at HF. This is a disadvantage because in principle we want to have the maximum gain at LF to minimize the static error, and in the same way we want the minimum gain at HF to get rid of the switching noise.

The target of the K factor methodology is to define the K value which is compatible with the phase margin we are targeting.

Properties of the type II compensator transfer function

The red line is the phase of the compensator, which provides a boost. It is easy to demonstrate that the maximum phase boost occurs at the geometric average between pole and zero. But which is the maximum phase boost? It can be computed with the arctangent.

$$G_C(s) = \frac{K_C}{s} \cdot \frac{\left(1+\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{\omega_p}\right)}; \quad \omega_p > \omega_z$$



Max phase boost occurs at:

$$\omega_{max} = \sqrt{\omega_z \cdot \omega_p}$$

and it is equal to*:

$$\varphi_{boost} = \text{atan}\left(\sqrt{\frac{\omega_p}{\omega_z}}\right) - \text{atan}\left(\sqrt{\frac{\omega_z}{\omega_p}}\right) = 2 \cdot \text{atan}\left(\sqrt{\frac{\omega_p}{\omega_z}}\right) - \frac{\pi}{2}$$

i.e., $$\sqrt{\frac{\omega_p}{\omega_z}} = \tan\left(\frac{\varphi_{boost}}{2} + \frac{\pi}{4}\right)$$

* use the following equality : $\text{atan}(x) + \text{atan}\left(\frac{1}{x}\right) = \frac{\pi}{2}$

Moreover, **sqrt(w_p/w_z) = k**.

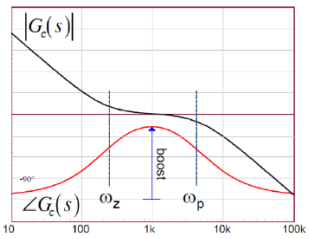
k-factor design method procedure

The first thing is to choose the crossover frequency, which is in general is $1/10^{th}$.
Then we want the maximum phase boost at w_c; the needed phase boost depends on how much phase margin we want to get in the final Gloop. This can be written with equation x. The maximum phase boost we can get from the type II compensator is 90°. Hence if from this computation I get a phase boost greater than 90°, the type II is not good for the job, I need to move to a type III compensator.

Procedure /1

- Choose suitable crossover frequency ω_c.
- We want : $\omega_{max} = \sqrt{\omega_z \cdot \omega_p} = \omega_c$
- Calculate the phase boost required by the compensator using:

x $\varphi_{boost} = \varphi_m - \angle G_{co}(j\omega_c) - 90°$ (note : $\varphi_{boost} < 90°$)



- Calculate the factor k using:

$$k = \sqrt{\frac{\omega_p}{\omega_z}} = \tan\left(\frac{\varphi_{boost}}{2} + \frac{\pi}{4}\right)$$ (see previous slide)

- Set:

$$\omega_z = \frac{\omega_c}{k} \text{ and } \omega_p = \omega_c \cdot k$$

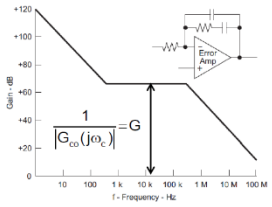
Now we can compute the factor k as seen before. Once k has been computed we can also compute the position of the pole and zero.

The missing thing is the mid frequency gain. How much large should it be?
We have to compute the mid frequency gain in such a way so that at fc this gain is exactly equal to the reciprocal of the gain of the control to output t.f., because I want to compensate the gain of the plant with the gain of the compensator. Hence G = 1/|Goc(jw_c)|.

Procedure /2

- Once k is known, the mid frequency gain can be calculated:

$$G = |G_c(j\omega_c)| = \frac{K_C}{\omega_c}\frac{\sqrt{1+k^2}}{\sqrt{1+1/k^2}} = \frac{1}{|G_{co}(j\omega_c)|}$$



- Therefore:

$$K_C = \omega_c \cdot \frac{\sqrt{1+1/k^2}}{\sqrt{1+k^2}} \cdot \frac{1}{|G_{co}(j\omega_c)|} \approx \frac{\omega_c}{k} \cdot \frac{1}{|G_{co}(j\omega_c)|} \quad \text{if } k \gg 1$$

- Now, ω_z, ω_p and K_C are known ⇒ $G_C(s) = \frac{K_C}{s} \cdot \frac{(1+s/\omega_z)}{(1+s/\omega_p)}$ is fully defined.

Then G, which is the gain of the compensator at mid frequencies, can be written as in the red box, which is obtained by replacing jw_c in the following equation, and then computing the magnitude.

$$G_C(s) = \frac{K_C}{s} \cdot \frac{\left(1+\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{\omega_p}\right)}; \quad \omega_p > \omega_z$$

The only unknown that remains is Kc.
Now Kc, w_z and w_p are known and the compensator t.f. is fully defined. We are just left with the network synthesis to produce the t.f..

Example

Buck converter. Top right plot is the Goc, which has a c.c. pole at 3kHz. Target fc = 80kHz, because fsw = 400kHz, and a phase margin of 60°.

- Target crossover frequency, f_c =80 kHz and phase margin, φ_m =60°.
- Required phase boost:

$$\varphi_{boost} = \varphi_m - \angle G_{co}(j\omega_c) - 90° = 76°$$

note: $\varphi_{boost} < 90°$ → ok!

- Calulate k:

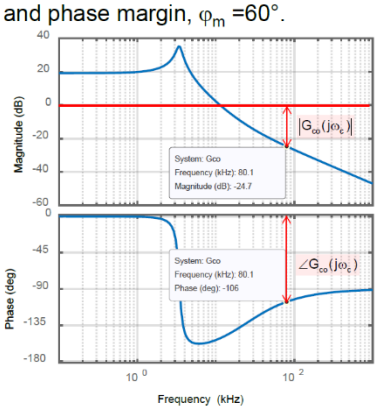
$$k = \tan\left(\frac{\varphi_{boost}}{2} + \frac{\pi}{4}\right) = 8.14$$

- Set:

$$\omega_z = \frac{\omega_c}{k} \text{ and } \omega_p = \omega_c \cdot k$$

- Calculate K_C:

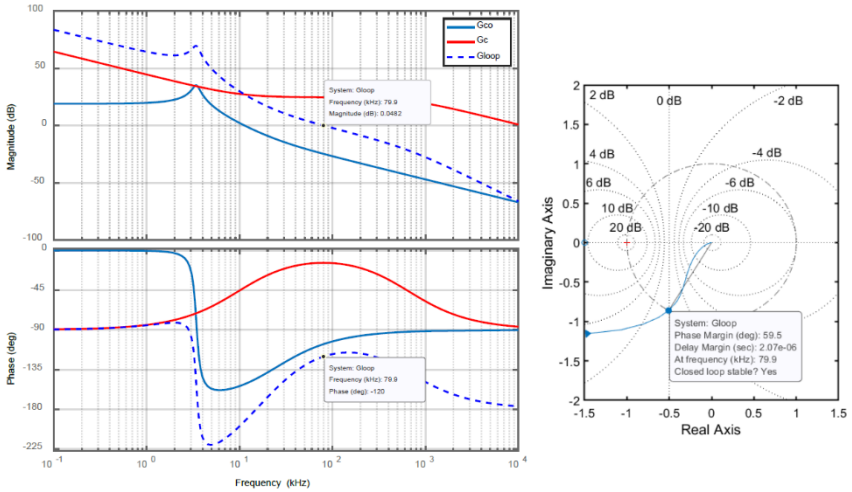
$$K_C = \frac{\omega_c}{k}\frac{1}{|G_{co}(j\omega_c)|} = 1.06\,\text{Mrad/s}$$



$$G_C(s) = \frac{K_C}{s} \cdot \frac{(1+s/\omega_z)}{(1+s/\omega_p)}$$

First thing is to look at the gain of the plant at 80kHz in the first plot, which is -24.7dB.

So we will need a compensator type II t.f. having a mid-gain of +24.7dB. The second important point is the phase of the plant at the desired crossover frequency. The phase is -106°. Since the target is 60°, we need a phase boost of 76°. Since the phase boost is < 90° we can proceed with a type II compensator. Then all the explained computations.

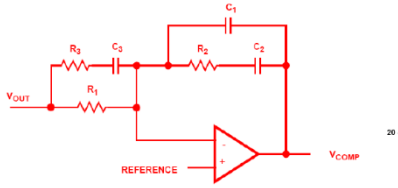


As a final result we see that, as expected, fc = 80kHz with a phase -120°, meaning that the phase margin is exactly 60°. The red line is the phase contribution of the compensator t.f., whose maximum is at fc. If we look at the Nyquist plot, as expected the Gloop crosses the unity circle with a phase margin of 59.5°.

Type III compensator

If we get the need of a phase boost larger than 90°, we need to resort to this compensator.
Type III compensator is a compensator with a pole in the origin (integrator action), a pair of zeros and a pair of poles.

The Gc(s) t.f. of the type III has still the minus sign outside it, and it can be simplified in a more usable way if we assume that the capacitor C2 >> C1, which is a condition that is almost always fulfilled in real compensator.

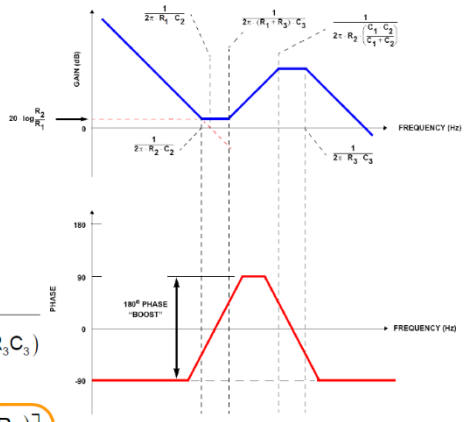


note : $G_c(s) = -\frac{\tilde{V}_{COMP}(s)}{\tilde{V}_{OUT}(s)}$

$$G_c(s) = \frac{(1+sR_2C_2)\left[1+sC_3(R_1+R_3)\right]}{sR_1(C_2+C_1)\left[1+sR_2\left(\frac{C_1\cdot C_2}{C_1+C_2}\right)\right](1+sR_3C_3)}$$

$$G_c(s) \cong \frac{1}{sR_1C_2}\cdot\frac{(1+sR_2C_2)\left[1+sC_3(R_1+R_3)\right]}{(1+sR_2C_1)(1+sR_3C_3)}$$

for $C_2 \gg C_1$

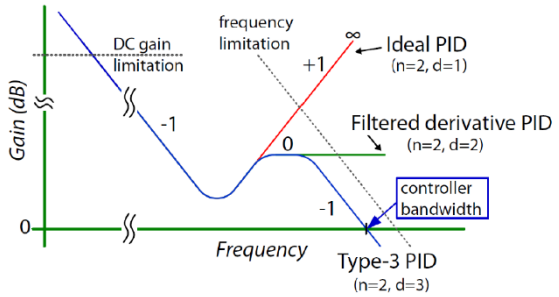


Type III compensator provides up to 180° phase boost.

More importantly, we notice that there are two regions in the Bode of the magnitude where the plot is flat. The gain in the first flat region sees C2 shorted and C3 open, so the gain is R2/R1. In the second flat region both C2 and C3 are shorted, and the gain is R2/(R3||R1).

The type III compensator is hence a PID compensator.

Type III as a filtered PID



Note: the type III compensator implements a filtered-derivative PID with an additional, high-frequency pole

n is the order of the numerator, d the order of the denominator in the PID. The red one is not a physical t.f. because it is anticausal (order of the numerator greater than the order of the denominator), so there is no ideal PID. Furthermore, there is no reason to use an ideal PID, because its gain at HF keeps increasing, while we want to decrease the gain at HF.

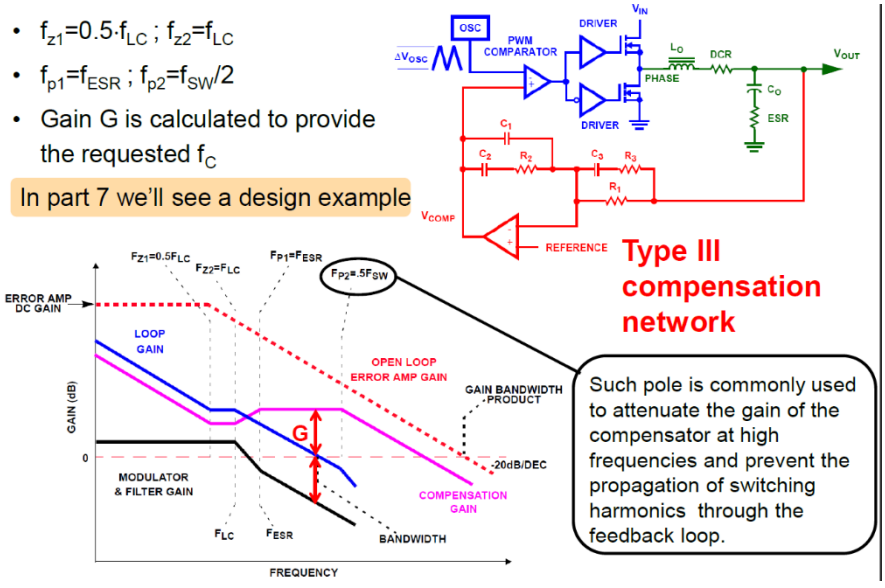
This is the reason why in the real world the PID is always a filtered derivative PID, meaning that the derivative term is filtered, i.e. there is an additional pole so that the order of the numerator is equal to the one of the denominator (green t.f.).

However, the type III compensator has an additional pole, which is useful because we want to attenuate the gain of the loop at HF, so it is a filtered PID compensator plus an additional pole.

Design guidelines for type III compensator

- $f_{z1}=0.5 \cdot f_{LC}$; $f_{z2}=f_{LC}$
- $f_{p1}=f_{ESR}$; $f_{p2}=f_{SW}/2$
- Gain G is calculated to provide the requested f_C

In part 7 we'll see a design example



The first zero of the compensator is typically (not compulsory) put at ½ of the frequency of the resonant filter, so the filter of the c.c. poles (f_LC). The second zero of the compensator is located exactly at f_LC. As for the first pole of the compensator (fp1), it is placed at the zero frequency of the ESR. The second pole is instead located at half of the switching frequency fsw in order to attenuate the Gloop (Gloop is the product between the pink line, compensator, and black line, plant).

The Gloop has a flat region because we are trying to compensate a pair of c.c. poles with two real zeros. Starting from these guidelines we size the component of the compensator.

Not only we have to locate the zeros and the poles of the compensator t.f., we also have to regulate the gain of the compensator t.f. in the mid frequency region. The gain in this region has to compensate the negative gain of the plant.

However, there is an issue when trying to compensate a couple of c.c. poles with two real zeros at the same frequency. Two c.c. poles contribute to the phase with a sharp decrease to -180°, while a real zero has a smoother increase in phase. So at f_LC the contribution of the zeros will be +90° (45° + 45° of two zeros), but at the same time we are dropping down abruptly of -180° due to the c.c. poles. So we have a drop of the phase around f_LC, because we are trying to compensate something that is dropping fast with something that is slower.

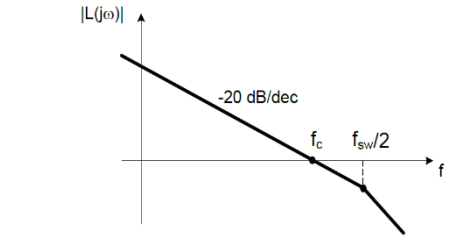
In order to adjust this, we put one zero at a smaller frequency and the other at f_LC, this is the reason why the first zero is at 0.5*f_LC.

As a second consideration, the error amplifier is an opamp that of course is not ideal. What's the impact of the limited GBWP of the amplifier and limited LF gain of the amplifier?
In fact, the LF gain of the loop transfer function cannot be infinite, it won't even be equal to the one of the amplifier, but scaled down by a voltage divider (done on purpose). Actually, in DC we don't have a local feedback because the capacitors are all open and we have the OL gain of the opamp.
If we want the compensator to work correctly, the GBWP of the opamp must be larger than the frequency at which the compensator t.f. crosses the 0dB axis, otherwise we are cutting the gain of the compensator and reducing the phase margin.

The third consideration is that there is again a zero-pole doublet in bandwidth. This is not a problem in general because we are not interested in the reference to output t.f., because the reference is a fixed DC value in a regulator. Anyway, the reference to output t.f. is not just L/(1+L), but it is H(s) * L/(1+L).
However, let's suppose we are interested in the transfer function L/(1+L), which is a t.f. with a feedforward gain of L and a feedback gain of 1.
The time response has two contributions, a fast and a slow one, but what is the relative amplitude of the two components? In fact, if the amplitude of the slow component is negligible with respect to the amplitude of the fast one, there is no problem. The relative amplitude depends on the gain, the larger the smaller the impact of the slow component. In our case the gain is more or less 20dB.

The problem caused by the doublet occurs in the load transient response, but will be investigated in part 7.

Phase margin

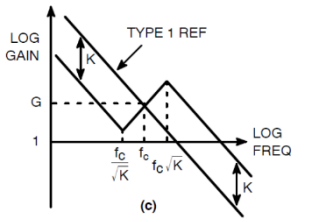


$$\varphi_m = 90° - \tan^{-1}\left(\frac{f_c}{f_{sw}/2}\right)$$

e.g., $\varphi_m = 68°$ for $f_c = 0.2\, f_{sw}$

In here, for the sake of simplicity, we are neglecting the bump in the -20 rolloff due to the non-perfect compensation of the two real zeros with the c.c. poles. We notice the Gloop is a piecewise t.f., one with rolloff -20 and the other -40 dB/dec.
The phase margin is related to the ratio between the position of the HF pole and the crossover frequency fc.

k-factor design method applied to the type III compensator



The following equations hold:

$$\omega_z = \frac{\omega_c}{\sqrt{k}} \text{ and } \omega_p = \omega_c \cdot \sqrt{k}$$

$$k = \tan^2\left(\frac{\varphi_{boost}}{4} + \frac{\pi}{4}\right) \qquad 90° < \varphi_{boost} < 180°$$

$$G = \left|G_c(j\omega_c)\right| = \frac{K_C}{\omega_c}\frac{1+k}{1+\frac{1}{k}} = \frac{1}{\left|G_{co}(j\omega_c)\right|}$$

$$K_C = \omega_c \cdot \frac{1+\frac{1}{k}}{1+k} \cdot \frac{1}{\left|G_{co}(j\omega_c)\right|}$$

$$G_C(s) = \frac{K_C}{s} \cdot \frac{\left(1+\frac{s}{\omega_z}\right)^2}{\left(1+\frac{s}{\omega_p}\right)^2}; \quad \omega_p > \omega_z$$

Coincident compensation poles and zeroes

We consider a compensator t.f. having a pair of zeros (at the same frequency) that is fc/sqrt(K) and a pair of poles at fc*sqrt(K). The provided phase boost is larger than 90°, up to 180°.

WHAT COMPENSATOR TO BE USED?

- Type I: is used where no phase boost is necessary at the crossover, in the case of current mode converters, in which we control both the current in the inductor and the voltage across the capacitor at the same time. With this controller, the control to output t.f. will be a first order t.f., even if the converter is working in CCM. So since we have just one pole, the type I is enough.
- Type II: is targeting applications where a phase boost is necessary. Most popular choice for current mode converters.

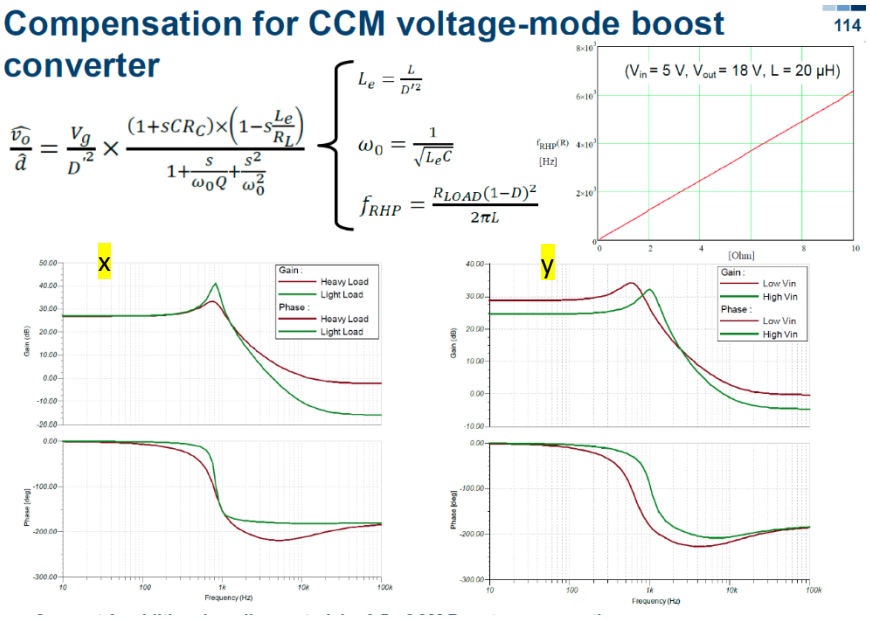
  We can also use a type II compensator to compensate complex transfer functions such as a buck converter working in CCM in voltage control mode. The type II can be used if we take advantage of the presence of the zero of the ESR which gives a +90° contribution that helps to get an acceptable phase margin. This happens only if the zero of the ESR happens around the crossover frequency, so if we use a large electrolytic capacitor with a large ESR value.

  However, if we have a buck converter in voltage control mode but using a MLCC (capacitor with small ESR), its zero is at HF and type II compensator is not a solution at all, so we have to go for a type III.
- Type III: is selected where a large phase boost is mandatory. This is the case for CCM voltage-mode converters. Generally, 2nd order and beyond types of transfer functions.

COMPENSATION OF CCM BOOST CONVERTERS

The CCM boost converter is not a minimum phase t.f., so the control to output t.f. of a boost converter in CCM is not minimum phase because we have a real positive zero (RHP zero) which is introducing a negative phase contribution.

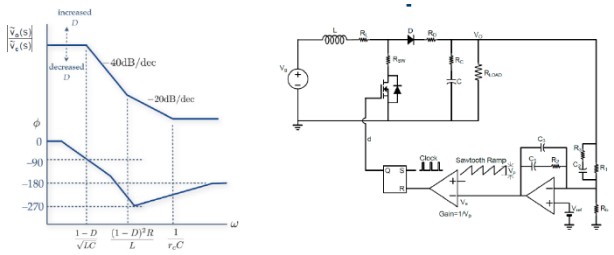
Not only we have e non-minimum phase t.f., but also the position of the RHP zero depends on the load (see f_RHP formula) and on the duty cycle. Since we have a dependance on the D and in general the output voltage of the converter is fixed because it's regulated, if we change the input voltage the D is changing and so the position of the RHP zero varies. This is also true for the equivalent inductance Le that appears in the w0 expression of the c.c. poles.



X is the Goc t.f. while we are changing the load R, y is the Goc changing the Vin voltage. How can we compensate such t.f.?

Firstly we need to understand which is the working condition, which in general is the one that gives us the worst phase margin, that is in the case of heavy load condition (red one) for the load, while for the voltage is the low input voltage (red one).

So combination of low input voltage and large load is the worst condition. The compensator design must be done in this condition.



- Select a Type III compensator with the following rules:
  - Compensator zeros are placed around the power stage resonant frequency.
  - The second pole of the compensator is placed coincident with the ESR zero.
  - The third pole of the compensator is placed coincident with the RHP zero frequency.
  - If the RHP zero or ESR zero is higher than half the switching frequency, the corresponding compensation pole is placed at half the switching frequency.

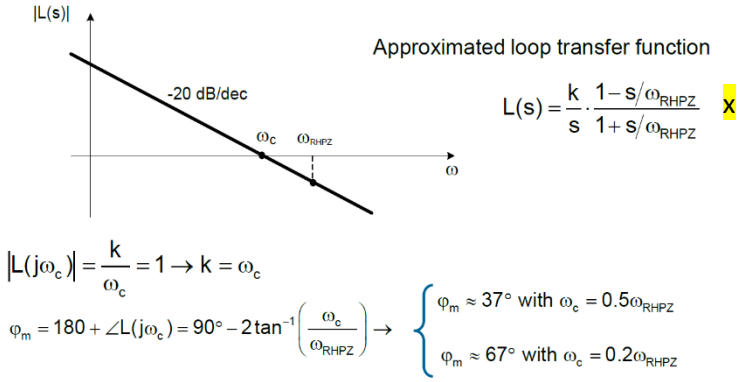
Note: design at the lowest input line and the maximum load.

Compensator zeros are placed around the power stage resonant frequency (i.e. c.c. poles frequency).

In general, the ESR zero falls at high frequency (MHz) when we are using a ceramic capacitor, because its ESR is very small, and this is far away from the crossover region. In this case we can take advantage of the type III compensator that we are using in a boost converter to introduce a pole at half of the switching frequency. The pole that was originally used to compensate the ESR zero, when the ESR zero is far away can be used to introduce a -40 db/dec rolloff starting at half the switching frequency.

The second pole of the compensator is placed in correspondence to the ESR zero of the plant to compensate it, and the third pole is placed a the same frequency of the RHP zero.

If we do so, we get the result below (bump is not represented).



The t.f. of the loop gain is x. We are compensating a RHP zero with a pole, so the magnitude is compensated, but not the phase (?).
In conclusion, L magnitude at fc must 1, and this allows us to compute the k.
Depending on the relative position of fc and RHP zero frequency we might have different phase margins. Since the RHP zero happens typically at kHz, we are forced to set fc at fractions of kHz, which is a big issue, because the bandwidth is shrank.

Moreover, the RHP zero frequency in the real case is not fixed, it moves, because ESR of the capacitor depends on the temperature.

FEEDFORWARD COMPENSATION

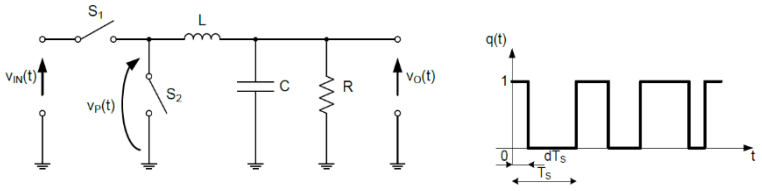
In general, we have a problem with the control to output t.f. (Goc), which is proportional to the DC input voltage. In the real world, however, v_in is not constant, so the Goc moves. This is a problem because compensation becomes not easy, since the t.f. we want to compensate moves up and down.
We might try to compensate for the worst case condition, but the best solution is to implement a feedforward compensation.

BUCK CONVERTER

Let's consider the feedforward compensation in a buck converter. Switches S1 and S2 are operated in antiphase. Let's focus on v_p(t), which is a square wave equal to the switching function q(t) multiplied by the amplitude of the input voltage → v_p(t) = q(t)*v_in.

Once again, let's apply a running average to v_p(t), assuming that v_in is varying slowly with respect to the switching function, which usually occurs in the real world.
<v_p> is d(t), continuous duty cycle, multiplied by the running average of the input <v_in>, which is actually v_in if it varies slowly.

- Control-to-output TF depends on V_{IN}, which usually ranges in a wide interval → $G_{loop}(0)$ varies with V_{IN}
- Solution: feedforward compensation
- Buck converter



$$\langle v_P(t)\rangle \cong \langle v_{IN}(t)\rangle \cdot d(t) \cong v_{IN}(t)\cdot d(t)$$ if v_{IN} is slowly varying

See part 6_additional reading materials: 6.6_Feedforward compensation

Second step

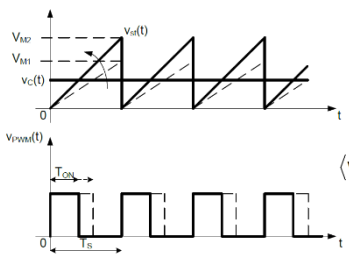
Let's assume that the amplitude of the sawtooth voltage waveform we are applying to the comparator that implements the PWM is not fixed, but it has an amplitude proportional to the input voltage. This is done by using a RC network. The switching period stays the same.

Starting from the previous <v_p(t)> equation, d(t) is the ratio between the control voltage and the peak value of the sawtooth voltage waveform, which is not constant but proportional to v_in. if we replace d(t) we can cancel out the v_in dependance, and <v_p(t)> depends just on the control voltage v_c.

- Amplitude of the sawtooth signal is proportional to V_{IN}

$$V_M(t) = \alpha \cdot v_{IN}(t)$$

- Frequency of the sawtooth signal is constant



$$\langle v_P(t)\rangle \cong v_{IN}(t)\cdot d(t)$$

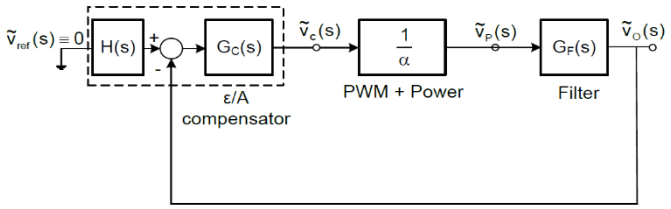
$$d(t) = \frac{v_C(t)}{V_M(t)} = \frac{v_C(t)}{\alpha \cdot v_{IN}(t)}$$

$$\langle v_P(t)\rangle = \frac{v_C(t)}{\alpha \cdot v_{IN}(t)} v_{IN}(t) = \frac{v_C(t)}{\alpha}$$

$$\tilde{v}_P = \frac{\tilde{v}_C}{\alpha}$$

Of course then we can linearize the equation.

If we apply the feedforward compensation the feedback is unchanged, but the power block can be replaced by a simple block with gain 1/alpha, where alpha is the proportionality coefficient between the input voltage and the amplitude of the sawtooth voltage, and therefore the power signal we are applying to the filter is just a proportional function of the control voltage → v_in has disappeared both in DC and AC, the behaviour of the system is completely independent on the fluctuations of the input voltage, but also independent on the fast variations of the input voltage.



$$G_{CO}(s) = \frac{\tilde{v}_O(s)}{\tilde{v}_C(s)} = \frac{1}{\alpha} \cdot G_F(s)$$

$$L(s) = G_C(s) \cdot \frac{1}{\alpha} \cdot G_F(s)$$

independent of V_{IN}

The new loop t.f. L(s) is the compensator t.f. Gc(s) multiplied by 1/alpha * Gf(s).

The advantage of the feedforward is that the don't have to wait for the signal to travel all around the loop to provide the error signal, now if the input changes, immediately we are changing the height of the sawtooth voltage waveform, so we are changing the duty cycle which is counteracting the variation of the input voltage.

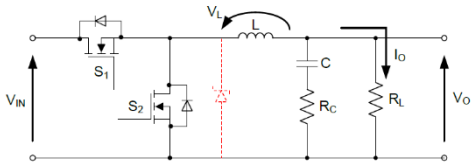
CASE STUDY – STEP-DOWN REGULATOR DESING

Specs

- Input voltage: V_{IN} = 5 ÷ 18 V
- Output voltage: V_O = 3.3 V
- CCM operation
- Output voltage ripple: ΔV_O = 30 mV
- Output current: I_O = 3 A → P_o ~10 W
- Switching frequency: f_S = 400 KHz.

SYNCHRONOUS BUCK CONVERTER

We introduce a slight modification to the standard buck converter. The diode that usually is in place of S2 is replaced by a switch (Schottky diode is not important) and the two mosfets are driven in antiphase.



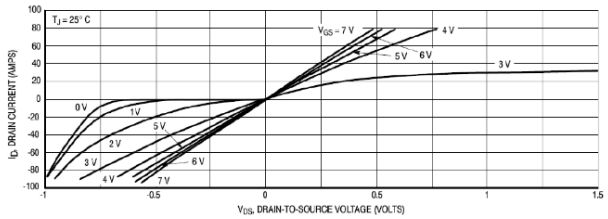
- A synchronous buck converter is a modified version of the basic buck converter circuit topology in which the freewheeling diode is replaced by a second switch, S_2.
- This modification is a tradeoff between increased cost and improved efficiency.
- Schottky diode sometimes used to further improve the efficiency

In general, besides the use of a mosfet we may want to use a Schottky diode in parallel to it.

Advantages

We have a smaller power dissipation in conduction (less conduction losses). Of course this is true if we properly select the synchronous mosfet S2.

The plot represents the drain current of a mosfet device as a function of the drain to source voltage around the zero. The parameter changing between the curves is the Vgs. Assuming Vgs is sufficiently high, the output characteristic of the mosfet is a line, so the mosfet behaves as a pure resistor in ohmic region. Since it's operated as a resistor it is a bidirectional device and current can flow from drain to source or from source to drain.



$$P_d = V_d \cdot I_o \cdot (1-D)$$

$$P_{MOS} = I_o^2 \cdot (1-D) \cdot R_{DSon}$$

$$\Rightarrow \quad \boxed{P_{MOS} < P_d \quad \text{if} \quad R_{DSon} < \frac{V_d}{I_o}}$$

Let's compute the conduction losses in the standard buck and in the synchronous one. In the former case losses are computed multiplying the forward voltage drop across the diode (assumed constant) and the average current across a switching cycle. Technically, the P should have been computed as 1/Ts multiplied by the integral from 0 to Ts of the instantaneous current multiplied by the forward voltage,

which is constant. In a buck converter, the instantaneous current in a diode, if the converter is working in CCM, is flat during the (1-D)*Ts time period, so the average current is the peak current (i.e. the inductor current, which in a buck is the output current) multiplied by 1-D.

As for the power dissipation in the latter case, the power dissipated by a resistor (i.e. the transistor in ohmic) is Rds(on) multiplied by the square of the root mean square value of the current flowing in the mosfet, Irms. Irms is Ipeak*(1-D). So we can replace the diode with a mosfet if the power dissipated by the mosfet is smaller. This happens if Rds,on < Vd/Io.

There is still something missing when making this comparison. In fact, in a standard buck we have to drive just a single device, while here two mosfets, so we are doubling, assuming the two mosfets are identical, the power dissipated by the gate driver. However, if the switching frequency is small and also the gate charge is, in general the dominant losses are the conduction losses.

Moreover, it is almost impossible to have a synchronous turn on and turn off of the two mosfets, because they are never exactly identical. To avoid cross conduction, that is a situation where both the mosfets are on, we introduce a **dead time**. Assuming we start from a situation where S2 is on and S1 is off, before turning off S2 and on S1 we will first turn off S2, we keep it off for a dead time and after it we turn S1 on. The same thing is done in the opposite case. If both switches are off and we don't use the Schottky diode, the current in the inductor will be flowing in the body-drain parasitic diode of S2, determining a certain power dissipation.

If we want to reduce the power dissipation we can put a Schottky diode in parallel to the mosfet. The advantage is that the voltage drop across it when it's on is very small (0.4V – 0.5V) and power dissipation is reduced at the same current.

Synchronous step-down regulator

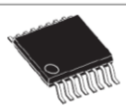
In order to close the loop and regulate the converter we are going to use this IC by ST.

L5989D

4 A continuous (more than 5 A pulsed) step-down switching regulator with synchronous rectification

Features

- 4 A output current (more than 5 pulsed)
- Operating input voltage from 2.9 V to 18 V
- External 1.8 V ± 2% reference voltage
- Output voltage from 0.6 to input voltage
- MLCC compatible
- 200 ns T_{ON}
- Programmable UVLO matches 3.3 V, 5 V and 12 V bus
- F_{SW} programmable up to 1 MHz
- Voltage feed-forward
- Zero load current operation
- Programmable current limit on both switches
- Programmable sink current capability
- Pre-bias start up capability
- Thermal shutdown



HTSSOP 16

Applications

- Consumer: STB, DVD, LCD TV, VCR, car radio, LCD monitors
- Networking: XDSL, modems, routers and switches
- Computer and peripherals: printers, audio / graphic cards, optical storage, hard disk drive
- Industrial: DC-DC modules, factory automation
- HC LED driving

This IC also includes power switches and their drivers.

There are external components to be added which are the filter inductor, the filter capacitor and several other capacitances and resistances. These externa C and R are the components of the compensation network, the one we use to design the compensator (or error) amplifier (region x). The opamp is inside the chip.

Figure 1. Test application circuit

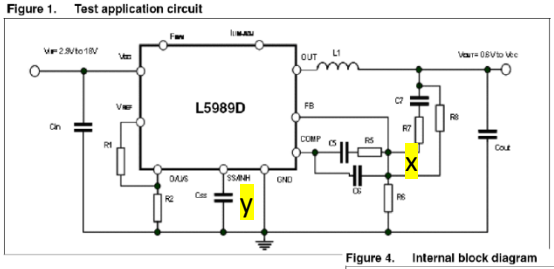
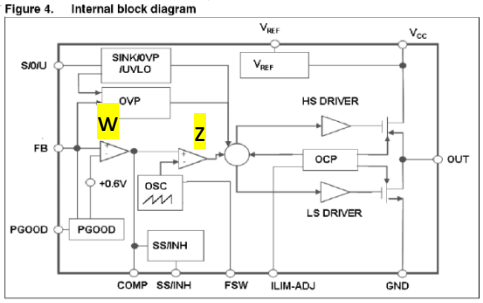


Figure 4. Internal block diagram



Instead, y is the so-called '**soft start capacitor**', it is used to determine the duration of the startup phase of the converter.
The other image displays the internal structure of the chip. We have two switches, a pMOS in high side position and a nMOS in low side position, with their corresponding drivers. OCP is the overcurrent protection to limit the current that can flow in the two switches. Z is the PWM and w is the error amplifier (in the image there's a mistake, + and – terminals are inverted).

One of the first thing to do is the inductor selection.

INDUCTOR SELECTION

The inductance value sets the current ripple in the output capacitor, so the larger the inductance the smaller the current ripple in the inductor and so the smaller the current ripple in the output capacitor.
In general, the rule of thumb is to **set the peak-to-peak current ripple from 20 to 40% of the average inductor current** (for a buck converter in CCM).
To compute the current ripple we compute the current variation during the on time period or during the off time period (as in the slide).

- The inductance value sets the current ripple flowing through the output capacitor.
- Rule-of-thumb: set the inductance value such that the current ripple is 20% - 40% of the average inductor current.
- The inductance value can be calculated by the following equations:

$$|\Delta i_L| = \frac{V_O + V_{LS}}{L}(1 - D_{min})T_S \le \Delta I_{max} \quad \text{with} \quad D_{min} \cong \frac{V_O + V_{LS}}{V_{INmax} + V_{LS} - V_{HS}}$$

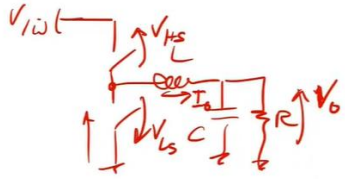
$$\Downarrow$$

$$L \ge \frac{V_O + V_{LS}}{\Delta I_{max} f_{SW}}(1 - D_{min}) = 8\,\mu H \quad \text{with} \quad \Delta I_{max} = 0.3 I_L \; ; \; D_{min} \approx 20\%$$

				Min	Typ	Max	Unit
Rdson HS	High side MOSFET on resistance	Iout = 1.0 A		75	85	95	mΩ
			(1)	111	120	132	mΩ
Rdson LS	Low side MOSFET on resistance	Iout = 1.0 A		62	67	72	mΩ
			(1)	92	100	106	mΩ

The formula is however a bit different, we don't have just Vo/L, but we have a Vls term, which is the voltage drop across the low-side switch when we are working in the off time period. Since the output voltage is relatively low, 3.3V, a voltage drop of few mV can be comparable with the output voltage, so better to take Vls into account.

Moreover, this voltage derivative is multiplied by the duration of the off time period. However, the D is changing because the input voltage is changing, and the worst case is the minimum vale of D. Also Dmin must be computed considering the parasitic components.



To compute the D we apply the V*s balance on the inductor.

$$(V_{in} - V_{HS}) \cdot D - (V_o + V_{LS})(1 - D) = 0$$

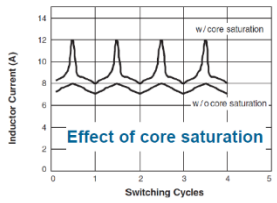
$$D = \frac{V_o + V_{LS}}{V_{in} + V_{LS} - V_{HS}}$$

The minimum D occurs with the Vin,max, minimum Vls and minimum Vhs.
Let's assume that we are targeting a delta(i_L) in the worst case of 30% of Io (Io and I_L are equal in a buck converter). The result is the one in the red box.
For the Rds,on we are putting the typical value and not the maximal one. This because if we perform a sensitivity analysis on Dmin, the typical is the best. Upper rows are at 25°C, lower rows (1) are stimated values in the whole temperature range.

Once we have computed the needed L value, we cannot pick any inductance with value 8 uH. In fact, there are potential problems. The magnetic core in the inductor, for instance, doesn't have to saturate.

We have 3 possibilities to select an inductor.

1) Find an off-the-shelf inductor having the desired inductance L with suitable rms and peak current ratings:
   - this may not be cost effective, however, if the standard types are not fairly close to your requirements.



2) Scan the literature for standard core types which you can custom wind to meet your particular requirement:
   - this is a quick way to get a prototype up and running;
   - it can also be very cost effective for some production situations.

3) Have the inductor custom wound by one of the many companies in the business:
   - cost might be an issue.

1. The easiest one, we select an inductor available on the market having the **desired inductance value**, but the inductor must have also a **suitable rms current** specification and a **suitable peak current** specification. If we don't respect them, the inductor is going to blow.
   In an inductor, the coil is always wounded around a magnetic core made with a ferromagnetic material which displays saturation. Hence we have to pay attention that the peak current must be sufficiently low to avoid the saturation of the magnetic core.

   When the magnetic core saturates, we have a situation like in the plot of the previous image. The triangular ripple is the expected current flowing in the inductor, and the peaking is the result of the saturation of the magnetic core.
   In a buck converter, the derivative of the current during Ton is (Vin – Vo)/L; current increases when we turn on the high side switch but if we come to a point where the magnetic core saturates, the inductance value L starts to drop quickly, and so the derivative of the current increases (saturation of the core is met when we have the corners in the plot). The peakings are not necessarily destructive, but we are reducing for sure the efficiency of the converter.

   The other important point is the rms rating. There is a maximum value for Irms for which, if overcame, the temperature at which the inductor is working is increasing over a specified maximum value. If we are using an insulated copper wire the risk is to melt the insulator of the inductor. The other potential risk of increasing too much the temperature is that the L drops and, once again, we have peaking as in the previous case.

   In the case where we don't find an off-the-shelf inductor with the characteristics we want the solution is to oversize the inductor, but this will give us higher cost and PCB area.

2. We scan the literature finding a magnetic core having the correct shape, volume and geometry and starting from the core we build the inductor. So the core is chosen from literature, the inductor is DIY. This is a quite effective solution.

3. A company custom-manufactures the inductor.

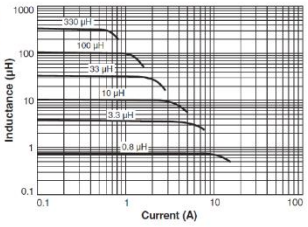
APPROACH 1

Manufacturer	Series	Inductor value (µH)	Saturation current (A)
Coilcraft	XPL7030	2.2 to 4.7	6.8 to 10.5
	MSS1048	2.2 to 6.8	4.14 to 6.62
	MSS1260	10	5.5
Panasonic	ETQP5M4R7YFM	4.7	8
Wurth	WE-HC/HCA	3.3 to 4.7	7 to 11
	WE-TPC type XLH	3.6 to 6.2	4.5 to 6.4
	WE-PD type L	10	5.6
Coiltronics	DR74	3.3 to 4.7	4.3 to 5.4
	DR125	10	5.3
BI	HM78-60	4.7 to 10	5.4 to 6.8

MSS1048

Part number[1]	Inductance[2] (µH)	DCR max (mΩ)	SRF typ[3] (MHz)	Isat (A)[4] 10% drop	20% drop	30% drop	Irms (A)[5] 20°C rise	40°C rise
MSS1048-682NL_	6.8 ±30%	16.3	35	3.60	5.00	5.94	4.52	6.01
MSS1048-822NL_	8.2 ±30%	20.0	28	3.14	4.14	4.84	4.38	5.71
MSS1048-103ML_	10 ±20%	23.0	24	3.08	3.84	4.32	3.99	4.79
MSS1048-153ML_	15 ±20%	36.0	20	2.46	3.06	3.44	3.51	4.26
MSS1048-223ML	22 ±20%	50.0	12	2.36	2.90	3.28	2.86	3.58

4) DC current at 25°C that causes the specified inductance drop from its value without current.
5) Current that causes the specified temperature rise from 25°C ambient.



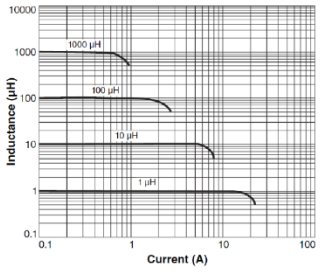
We open the catalog and check for specifications, retrieving the part number. The inductance is 10 uH +- 10%.
Saturation current is typically specified as the current that determines a reduction of the inductance value of a given percentage starting from the no-load inductance value or the low current inductance value. In the table we have 3 possibilities. If we target the 10% reduction, the maximum current is 3A, for instance.

As for the maximum Irms, there are a couple of specifications depending on the maximum variation in temperature we want to target.
Other specifications are the DCR, i.e. the parasitic resistance and SRF (self-resonant frequency). In fact, the real inductor has an inductance and a resistance in series, both in parallel with a parasitic capacitance coming from the windings. So we have a resonant network with a resonant frequency of 1/sqrt(LC). However, this in general is not a problem.

The plot on the left reports instead the inductance value versus the current. We have to look at the 10 uH curve in correspondence of 3A, the current at which we are working. Since we are working beyond the knee of the curve, we are in a dangerous situation. This because the knee moves back as we increase the temperature (the larger the temperature, the smaller the limit value for which the inductance value starts to decrease significantly), and so the inductance value with which we work is not the value we want, but smaller.

2nd try

Part number[1]	Inductance[2] (µH)	DCR[3] (mOhms) typ	DCR[3] (mOhms) max	SRF typ[4] (MHz)	Isat (A)[5] 10% drop	20% drop	30% drop	Irms (A)[6] 20°C rise	40°C rise
MSS1260-562ML_	5.6 ±20%	14.9	16.6	30.0	7.86	9.02	9.74	4.00	6.40
MSS1260-682ML_	6.8 ±20%	16.6	18.5	27.0	7.40	8.26	8.80	3.80	5.90
MSS1260-822ML	8.2 ±20%	17.0	20.0	26.0	7.10	7.96	8.50	3.40	4.80
MSS1260-103ML_	10 ±20%	21.5	23.9	22.0	6.18	6.92	7.40	3.00	4.00
MSS1260-123ML_	12 ±20%	24.5	27.3	20.0	5.18	5.94	6.42	2.80	3.70



MSS1260

We select a slightly larger inductor, still with the nominal inductance value as before. However, Ipeak and Irms are improved.

NB: the synchronous buck converter is not able to work in DCM because the switch S2 that replaces the diode is bidirectional. Is this an advantage? It is a disadvantage at high load. At high current the efficiency is better, because we are replacing the diode with a mosfet, but it is the opposite at low currents.

APPROACH 2

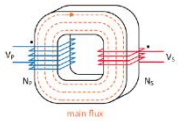
- Circuit designer must design and build the magnetic components (inductors, transformers) needed for the particular application.
- Design consists of:
    1. Selecting appropriate core material, geometry, and size (off the shelf standard parts).
    2. Selecting appropriate copper winding parameters: wire type, size, and number of turns.

Magnetic core

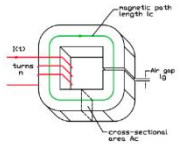
In transformers, the magnetic core is used to confine the magnetic flux within the core to get an optimal coupling between the windings. This is also why we want a magnetic core with a large permeability, the larger, the more confined the magnetic field.

The use of a ferromagnetic core in inductors allows to achieve large L values in small volumes. With the ferromagnetic material we want to confine the magnetic field within the core because if we don't use any core the magnetic flux line can interact with circuits located close to the inductor, which are subjected to a varying magnetic field, and we get EMI.

- Starting point: a magnetic core is required in most of the applications
    - in transformers: to define a path for the magnetic flux in order to get a good magnetic coupling between windings;



    - in inductors: to achieve large L values in a small volume, concentrate flux created by coil MMF to store the magnetic energy in airgap, spatially confine the magnetic flux (EMI avoidance).



Design challenges

A variety of factors constrain the design of a magnetic device.

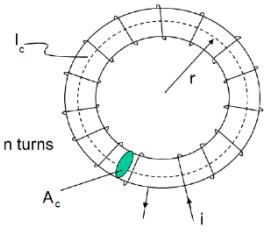
- The peak flux density must not saturate the core.
- The peak ac flux density should also be sufficiently small, such that core losses are acceptably low.
- The wire cross-sectional area should be as large as possible, to minimize the winding dc resistance and copper loss.
- Winding(s) should be arranged to minimize HF copper loss.
- An air gap is needed when the device stores significant energy. But an air gap is undesirable in transformer applications.

For a given magnetic device, some of these constraints are active while others are not significant

The peak AC flux density should be small because the losses in the magnetic core depend on the AC component of the flux density.

Inductor basics

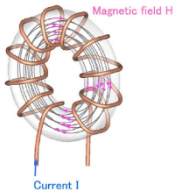
A coil of wire around a ferromagnetic core.



$\oint H \cdot dl = \sum i$ (Ampere's law)

$H\, l_c = ni = MMF$

$\Rightarrow \boxed{H = \frac{ni}{l_c}}$



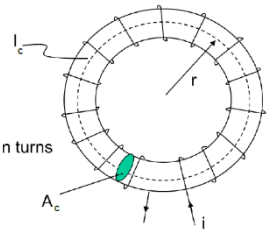
Magnetic flux density inside the core, $\boxed{B = \mu \frac{ni}{l_c}}$

The flux crossing the cross section area A of the core, $\boxed{\Phi = \int B dA = BA_c}$

An inductor is a simple coil of copper wire rounded around a toroidal core. We can compute the value of the inductance given a certain number of turns from the application of the Ampere's law. The net magnetomotive force calculated on a closed loop must be equal to the total current that crosses the inside part of the loop.
If we assume that the magnetic field strength is uniform inside the magnetic core, computations simplify as H*lc, where lc is the average length of the toroidal. n is the number of cores, i the current in the inductor. n*i is the magnetomotive force.

B is the magnetic flux density (magnetic induction), B = u*H.
The magnetic flux inside the core is the surface integral of the magnetic flux density over the surface and Ac is the area of the cross-section of the toroidal.

We can then define the flux linkage as the product of the magnetic flux and the number of turns in the inductor. The inductance is defined as the ratio of the flux linkage and the current that flows in the inductor.



$\lambda = n \cdot \Phi = n \cdot A_c \cdot B$ (flux linkage)

- We define *inductance* L as the ratio of the flux linkage, λ to the current, i: $L = \frac{\lambda}{i}$

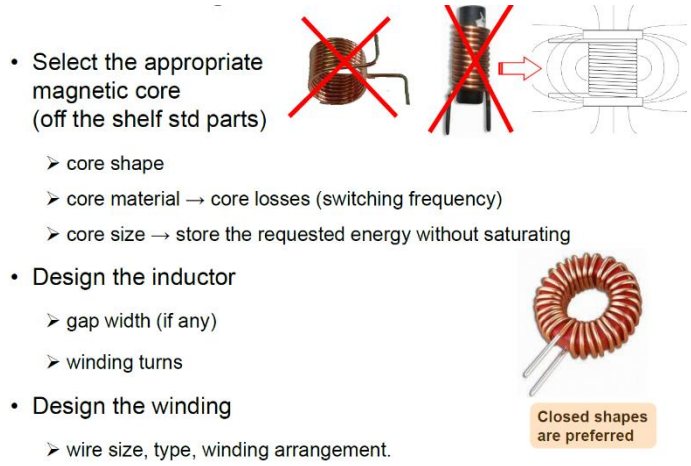
$\Rightarrow \boxed{L = n^2 \frac{\mu A_c}{l_c}}$ [Henry]

Note: $\boxed{L = \frac{n^2}{\mathcal{R}} = n^2 \cdot A_L}$

where $\mathcal{R} = \frac{l_c}{\mu A_c}$ and $A_L = \frac{1}{\mathcal{R}}$

magnetic reluctance — inductance factor

Inductor design steps



There are 3 steps:

1. Selection of the appropriate magnetic core: core shape, material and size. The core shape is important; typically a coreless material is never used, except for very large switching frequencies. Moreover, a cylindrical one is not used because in it the flux lines are everywhere around it, producing EMI to nearby circuits. The core typically has a closed shape to confine the magnetic flux.
   Instead, the core material determines the core losses, which are strongly dependent on the switching frequency.
   Moreover, the core size must be designed in a way that the inductor is able to store the required energy without making the core to saturate.

2. Once the core has been selected, we have to understand whether a gap is necessary (typically yes if we use ferrite, no if we use distributed gap material). We have also to select the number of winding turns.

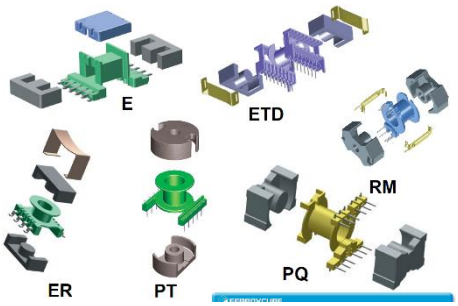
3. Winding design, peaking up the correct cross-section of the wire, the wire type and the winding arrangement.

Core shapes

These are examples of core shapes present on the market.

Core assembly

The core always comes in a split fashion (except for the toroidal one). The core is split on purpose to make the assembly easy.



Core geometry comparison

Aspect	Pot Core; RM Core	Double slab core	E core	Ec; ETD Cores	PQ Core	EP Core	Toroid
core costs	high	high	low	medium	high	medium	very low
bobbin costs	low	low	low	medium	high	high	none
winding costs	low	low	low	low	low	low	high
winding flexibility	good	good	excellent	excellent	good	good	fair
assembly	simple	simple	simple	medium	simple	simple	None
mounting flexibility	good	Good	good	fair	fair	good	poor
heat dissipation	poor	good	excellent	good	good	poor	good
shielding	excellent	good	poor	poor	fair	excellent	good

Magnetic core materials (review)

Magnetic units and definitions

H: magnetic field strength

Φ: magnetic flux

B: magnetic flux density (aka magnetic induction)

μ: permeability ;

μ_o: vacuum permeability

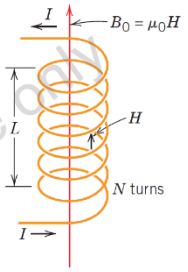
μ_r: relative permeability

quantity	*MKS*	*unrationalized cgs*	*conversions*
core material equation	$B = \mu_0 \mu_r H$	$B = \mu_r H$	
B	Tesla	Gauss	$1T = 10^4 G$
H	Ampere / meter	Oersted	$1A/m = 4\pi \cdot 10^{-3}$ Oe
Φ	Weber	Maxwell	$1Wb = 10^8$ Mx $1T = 1Wb / m^2$

Applied magnetic field

26

- The applied magnetic field, called the **magnetic field strength,** is designated by H. It can be generated by means of a cylindrical coil (or solenoid) consisting of N closely spaced turns, having a length L, and carrying a current of magnitude I



- Computation of the applied magnetic field, H:

$$H = \frac{N \cdot I}{L} \ (A/m)$$

- Computation of the magnetic flux density in a vacuum, B_0:

$$B_O = \mu_o \ H \ (Tesla)$$

Core material characteristics: B-H relation

- The magnetic field intensity H is in a sense a measure of the "effort" that a current is putting into the establishment of a magnetic field.
- The strength of the magnetic field produced in the core also depends on the permeability of the magnetic core.
- The relationship between the magnetic field intensity H and the resulting magnetic flux density B produced within a material is given by

$$B = \mu_r \mu_o H \quad Wb/m^2 \text{ or } T$$

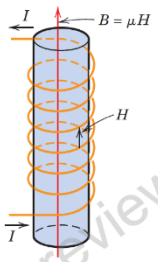
μ_o = permeability of a vacuum ($4\pi \times 10^{-7}$ H/m)

$\mu_r = \mu_r / \mu_o$ = relative permeability of medium

- ≈ 1 for air
- Several hundreds to thousands for ferromagnetic materials

Response to a magnetic field

- A magnetic field is induced in the material



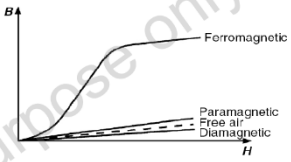
B = magnetic induction inside the material

$$B = \mu \cdot H \ (Tesla)$$

- The **magnetic induction**, or **magnetic flux density**, denoted by B, represents the magnitude of the internal field strength within a substance that is subjected to an H field.

Core materials: B-H relation

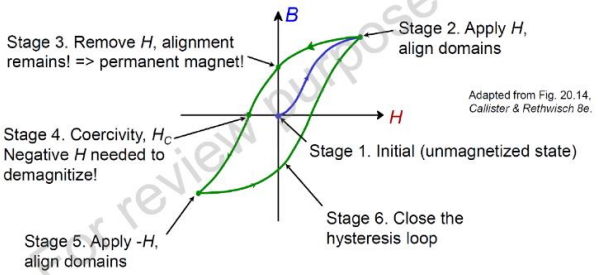
- Materials:
  - Diamagnetic → $\mu < \mu_0$
  - Paramagnetic → $\mu > \mu_0$
  - Ferromagnetic → $\mu >> \mu_0$



- Metals such as iron, nickel and cobalt are ferromagnetic; i.e. they produce a very high magnetic flux density, B, within them for a modest magnetising force, H. This may be shown on a magnetisation curve, i.e. a plot of B vs. H.

Hysteresis and permanent magnetization

- Ferromagnetic materials display a non linear B-H relationship with hysteresis and saturation.

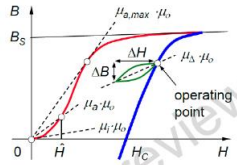


Definitions of permeability

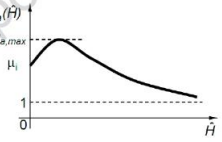
30

- There are several versions of μ_r depending on conditions. The index 'r' is generally removed and replaced by the applicable symbol e.g. μ_i, μ_a, μ_{rev} etc.

permeability from the initial magnetization curve



magnetic field intensity dependence of μ_a



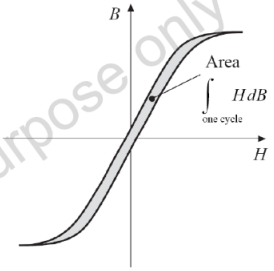
incremental permeability	initial permeability	amplitude permeability
$\mu_\Delta = \frac{1}{\mu_0}\cdot\left(\frac{\Delta B}{\Delta H}\right)_{OP}$	$\mu_i = \frac{1}{\mu_0}\cdot\left(\frac{dB}{dH}\right)_{H=0}$	$\mu_a = \frac{1}{\mu_0}\cdot\left(\frac{B}{H}\right)_{H=\hat{H}}$

Core losses: hysteresis loss

31

$$W = (A_c \ell_m)\int_{one\ cycle} H dB$$

The term $A_c \ell_m$ is the volume of the core, while the integral is the area of the B–H loop.



(energy lost per cycle) = (core volume) (area of B–H loop)

$$P_H = (f)(A_c \ell_m)\int_{one\ cycle} H dB$$

Hysteresis loss is directly proportional to applied frequency

Core losses: hysteresis loss

- Hysteresis loss varies directly with applied frequency
- Dependence on maximum flux density: how does area of B–H loop depend on maximum flux density (and on applied waveforms)? Empirical equation (Steinmetz equation):

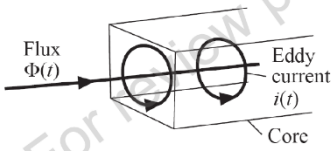
$$P_H = K_H f B_{max}^{\alpha} (core\ volume)$$

The parameters K_H and α are determined experimentally.

Dependence of P_H on B_{max} is predicted by the theory of magnetic domains.

Core losses: eddy current loss

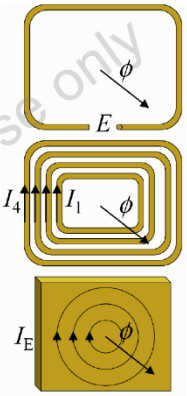
Magnetic core materials are reasonably good conductors of electric current. Hence, according to Lenz's law, magnetic fields within the core induce currents ("eddy currents") to flow within the core. The eddy currents flow such that they tend to generate a flux which opposes changes in the core flux $\Phi(t)$. The eddy currents tend to prevent flux from penetrating the core.



Eddy current loss $i^2(t)R$

Core losses: eddy current loss

- An ac flux flinking a rectangular-shaped conductor induces an ac voltage E across its terminals
- If the conductor terminals are shorted, a substantial current flows
- The same flux linking smaller coils induce lesser voltages and lower currents
- A solid metal plate is basically equivalent to a densely packed set of rectangular-shaped coils
- The induced currents flowing inside the plate are eddy currents, and flow to oppose the change in flux



Core losses: eddy current loss

- Ac flux $\Phi(t)$ induces voltage $v(t)$ in core, according to Faraday's law. Induced voltage is proportional to derivative of $\Phi(t)$. In consequence, magnitude of induced voltage is directly proportional to excitation frequency f.
- If core material impedance Z is purely resistive and independent of frequency, $Z = R$, then eddy current magnitude is proportional to voltage: $i(t) = v(t)/R$. Hence magnitude of $i(t)$ is directly proportional to excitation frequency f.
- Eddy current power loss $i^2(t)R$ then varies with square of excitation frequency f.
- Classical Steinmetz equation for eddy current loss:

$$P_E = K_E f^2 B_{max}^2 (core\ volume)$$

- Ferrite core material impedance is capacitive. This causes eddy current power loss to increase as f^4.

Core materials

Most common:

- High-permeability metals and alloys (not good for high frequencies)
- Ferrites (magnetic oxides mixed and assembled in ceramic form)
- Powdered iron and other ferromagnetic alloys in a ceramic or composite matrix.

Core type	B_{sat}	Relative core loss	Applications
~~Laminations iron, silicon steel~~	~~1.5 - 2.0 T~~	~~high~~	~~50-60 Hz transformers, inductors~~
Powdered cores powdered iron, molypermalloy	0.6 - 0.8 T	medium	1 kHz transformers, 100 kHz filter inductors
Ferrite Manganese-zinc, Nickel-zinc	0.25 - 0.5 T	low	20 kHz - 1 MHz transformers, ac inductors

We are focusing on ferromagnetic materials because we need a material with a high permeability. Iron-based materials are just suitable for low frequency applications (50Hz), so we cannot use them because we would end up with huge core losses due to the Eddy currents.

Instead, ferrites are the gold standard for high frequency applications, with a large magnetic permeability. A ferrite is basically made out of magnetic oxides, so the material is not electrically conductive, it has an almost 0 electrical conductivity. The result is that we don't have core losses due to Eddy currents, and this is perfect because Eddy currents depend on the square of the frequency.
There are two types of ferrite, the Manganese-zinc ferrite, which can be used up to few MHz, and the Nickel-zinc that can be used up to tens of MHz.
The typical **drawback is instead the low saturation flux density**, 0.25 to 0.5 T.

The third kind of material we might used is powdered iron or powdered iron alloys. These materials are the base of what we call **distributed gap cores**. With respect to the ferrite, the starting material of the powdered iron core are small iron particles (i.e. a metallic material); we take the bulk of iron, we mill it into a fine powder. The particles of the powder are covered with a non-magnetic material, i.e. a polymer, and the covered particles are pressed into a cast obtaining the desired shape, such as the toroidal shape.

Instead, for ferrite cores we start from the raw material that are magnetic oxides, we mill the material into a fine powder, but we don't need to cover the particles with a plastic material, we simply press the powder into a cast and we apply a high temperature process called sintering and we get a ceramic material.

The advantage of the powdered cores is the larger saturation magnetic flux density. Moreover, **powdered iron cores display a soft saturation, while ferrites a hard saturation**.
However, powdered iron cores have the disadvantage that, even if we are reducing the Eddy current losses, they are not completely eliminated, so we can use them up to few hundreds kHz.

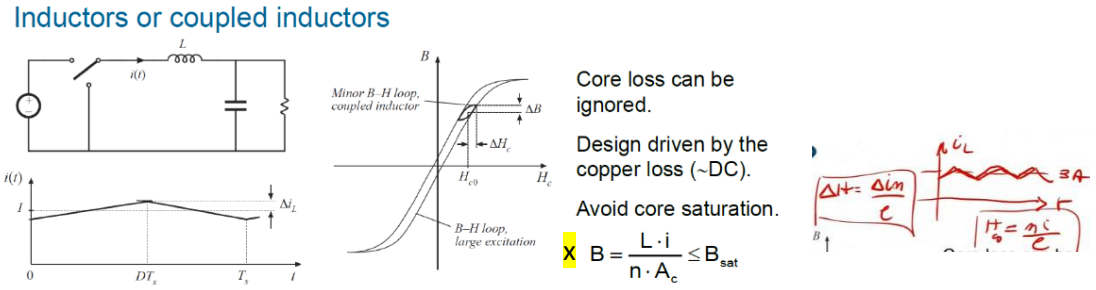
Ferrite or powder core?

There are some guidelines on which to choose. We have to distinguish two kinds of applications. In the case of inductors or coupled inductors, the best choice is still ferrite, even if we can also use powder iron.

Which is the difference between coupled inductors and a transformer?
Coupled inductors is a component that is supposed to store energy, while a transformer doesn't store energy, the power comes into it and immediately goes out.

The inductor we are focusing on is an inductor working almost in DC, in the sense that the current is constant around 3A (in the example) plus a ripple which is 30% of the value of the average current.

We can check the B-H loop for the inductor (right plot). We are moving along the **minor loop**, that is the bold one. In fact, the DC current is producing a H field that is n*i/L.
Then we have the ripple, so we have an additional delta(H) due to the ripple delta(i). However, since delta(i) is a small fraction of the average current, we are moving in the minor loop.



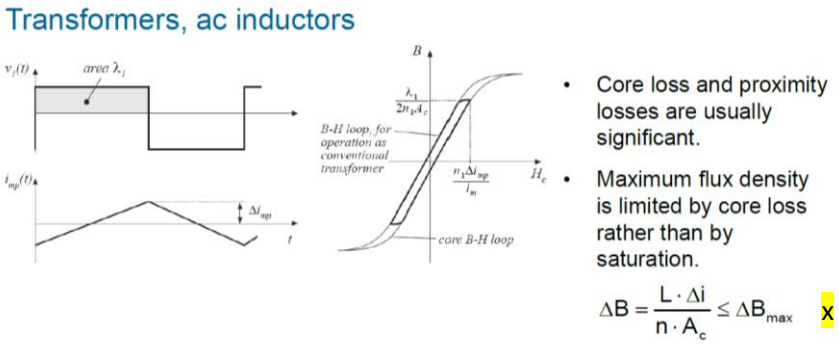
- A high-frequency ferrite material can be employed in this application
- However, powdered core materials having higher core losses and greater saturation flux density can be employead as well, leading to a physically smaller device.

If this is the case, the core loss can be ignored because it is not the dominant contribution since they depend on the AC value of the magnetic flux density H and if we apply and almost static magnetic field to a magnetic core the core is not dissipating any power.

The most important thing when designing this kind of inductors is to **avoid saturation**. In fact, the larger the current, the more we move towards saturation.
The condition to avoid saturation is x. Starting from the flux linkage lambda = L*i = n*phi, since phi = A*B (A is the area), if we want B < Bsat we need to verify inequality x.

In general, a high frequency ferrite is a good choice for this kind of application. However, since the core losses are not the dominant component because delta(i) is small, we can also use a powdered iron core. The advantage of the powdered iron core is that the saturation flux density for them is much larger than the one of the ferrite. The consequence is that Bsat is larger, and we can shrink the sizes of the inductor at the same value of inductance simply decreasing the cross-section of the toroid. The price to pay is a slightly large power dissipation with respect to the ferrite.

Transformers or AC inductors
In this case the current is a pure AC current and the B-H diagram has a magnetization loop that is bigger.



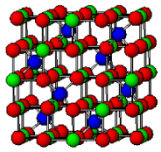
- A high-frequency material having low core loss (ferrite) is employed.
- Both core and copper losses must be accounted for in the design of the transformer.

The magnetic field is now swinging between two large values and therefore core losses and proximity losses in the copper wire will be dominating.
So the design of the transformer or of the AC inductor has to take another condition as a limit (x). We are working away from saturation, which is not a concern anymore, what is important is the delta(B)_max, which produces losses in the copper and in the core.

The ferrites are the only materials we can use if we are limited by the power dissipation and not by the saturation.

FERRITES

- Ferrimagnetic materials with spinel structure of the formula MFe_2O_4



Magnetite unit cell. Red atoms are oxygen, blue and green represent iron - different colors for Td/Oh sites respectively.

- M...divalent metal ion, where M = Mn, Fe, Co, Ni, or Cu, Zn, Mg, Cd
- With M=Fe → Magnetite: Fe_3O_4

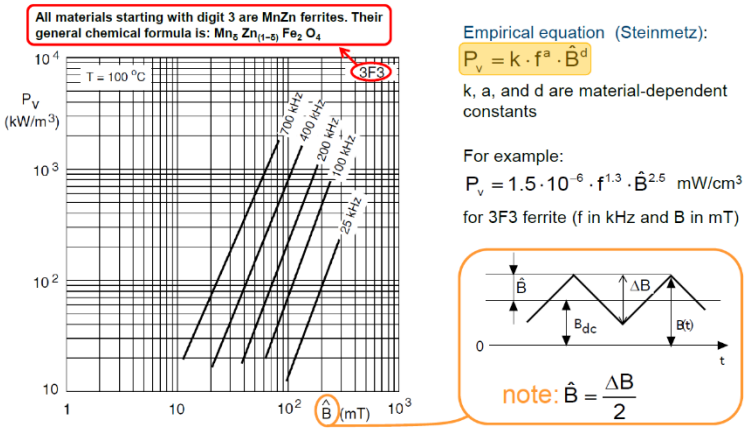
Basic Ferrite Material Properties					
Materials	Initial Permeability μ_i	Flux Density B_{max} Tesla	Curie Temperature, °C	dc, Coercive Force, H_c Oersteds	Resistivity Ω - cm
Manganese Zinc	750-15 K	0.3-0.5	100-300	0.04-0.25	10-100
Nickel Zinc	15-1500	0.3-0.5	150-450	0.3-0.5	10^6

The manganese-zinc ferrite shows a resistivity quite large compared to the one of a normal metal like iron, while the resistivity of the nickel-zinc is even higher, almost similar to an insulator resistivity.
As for the permeability (relative permeability) it is quite large. The flux density is the same for the two types of materials, around 0.3 or 0.5T, to be compared to the 0.9T.
The other parameter is the Curie temperature, which is eventually as low as 100°C for some types of ferrite.

Ferrites core losses

Ferrites are mainly prone to hysteresis losses, even if there is a small non-dominant component coming from the Eddy losses.

- Core loss incorporates *hysteretic* and *eddy current losses* and is a function of flux density amplitude, $\hat{B}$ and frequency, f.



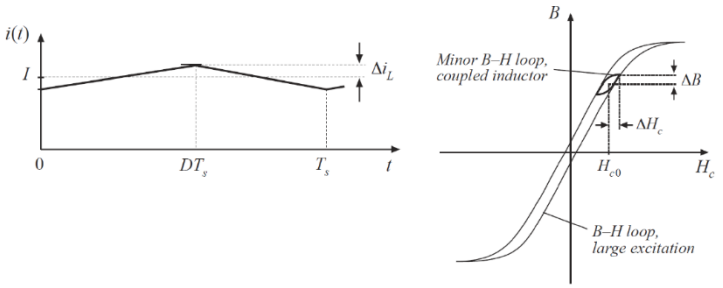
Steinmetz equation allows us to establish the power losses in a magnetic material.

Pv is the unit volume loss, which depends on the frequency f and on the peak AC value B of the magnetic flux density.
The plot on the left is obtained exciting the magnetic material with a sinusoidal magnetic field, so f is the frequency of the sinusoidal excitation.

There is an issue. If the AC component is not sinusoidal, but triangular, in principle we have a problem because the plot is obtained with a sinusoidal excitation. So in principle we should split the AC into spectral components and for each spectral component we should compute the corresponding power loss. In general this is not done.

How to compute delta(B)

If we want to use the previous plot we need to compute the B_hat, which is half the delta(B).
We start from L*i = n*phi, so L*delta(i) = n*delta(phi) = n*A*delta(B). Then the last computations are in the image. The value obtained from the formula in the blue box is the one we need to put in the formula to get the power dissipation.

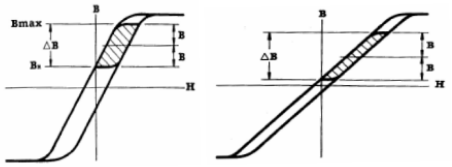


$$V = L\frac{di}{dt} = nA\frac{dB}{dt} \Longrightarrow \Delta B = \frac{L\cdot\Delta I}{n\cdot A} \Longrightarrow \boxed{\hat{B} = \frac{L\cdot\Delta I}{2\cdot n\cdot A}}$$

Ferrite cores: gapping

Ferrite core is supposed to store energy (LI^2)/2, but can ferrite store energy without saturating? Yes, if we use a huge core, otherwise if we want to keep dimensions small we need to resort to **gapping**.
A ferrite has a large value of relative permeability (that is the slope of the B-H curve) and a relatively low saturation magnetic flux density. The combination of these two results in the ferrite core being prone to saturation. In fact, it is sufficient to apply a relatively small H to fall in saturation. The larger the permeability, the steeper the B-H curve and the faster we end up in saturation for small H values.

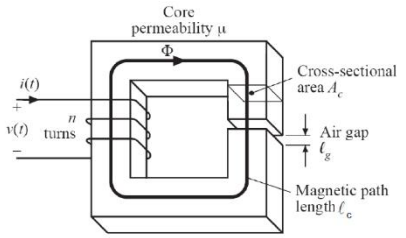
- Problem: high μ_r, relatively low B_s
- Ferrite cores are prone to saturation
- Gapping is mandatory
- **Gapping** of the magnetic core along the flux path causes a *"shearing over"* of the effective B-H loop (actually the Φ-$\mathcal{F}$ loop), which lowers the effective permeability and allows unsaturated operation at higher bias operating levels.

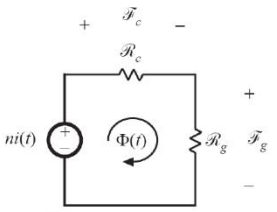


B-H Loop before gapping **B-H Loop after gapping**

The effect of gapping the core is to stretch in someway the B-H loop.
The gap is not modifying the characteristics of the magnetic material, the permeability is not affected, it is the effective permeability that changes.

Equivalent model for gapping





$$\Im = n \cdot i = \left(\Im_c + \Im_g\right) = \Phi \cdot \left(\Re_c + \Re_g\right) \qquad \Re_c = \frac{\ell_c}{\mu_o \cdot \mu_r \cdot A_c};\Re_g = \frac{\ell_g}{\mu_o \cdot A_c}$$

The equivalent reluctance of the gapped core is:

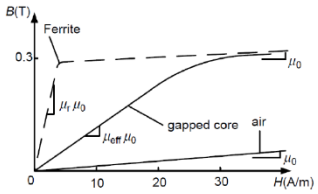
$$\Re_{eq} = \left(\Re_c + \Re_g\right) = \frac{\ell_c}{\mu_r \mu_o \cdot A_c} + \frac{\ell_g}{\mu_o \cdot A_c} = \frac{\ell_c}{\mu_o \cdot A_c}\left(\frac{1}{\mu_r} + \frac{1}{\ell_c / \ell_g}\right) = \frac{\ell_c}{\mu_o \cdot \mu_{eff} \cdot A_c}$$

We can get an electrical equivalent of the magnetic circuit, replacing the windings with a magnetomotive force generator n*i, which is equivalent to a voltage generator, and we can replace the reluctances of the core (reluctance of the magnetic core plus the reluctance of the gap) with two electrical resistances.
As for the magnetic flux, it is the equivalent of the current that flows in the equivalent circuit.

Effective permeability

This means that the reluctance of the gapped core is equivalent to the reluctance of a core of length ℓ_c and relative permeability μ_{eff}, with:

$$\mu_{eff} = \frac{1}{\frac{1}{\mu_r} + \frac{1}{\ell_c / \ell_g}} = \frac{\mu_r}{1 + \mu_r \frac{\ell_g}{\ell_c}}$$



If $\mu_r >> 1$, then the effective relative permeability is:

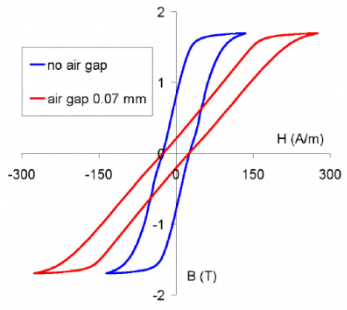
$$\mu_{eff} \approx \frac{\ell_c}{\ell_g}$$

One way to interpret this result is to say that if the air gap (length ℓ_g) is distributed over the whole core of length ℓ_c, the effective permeability of the core is reduced from μ_r to μ_{eff}.

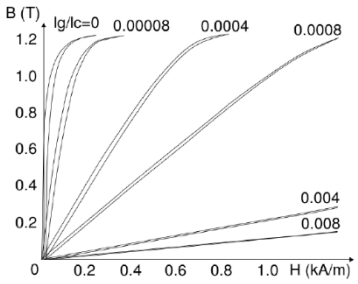
Having a gap in a core is like having a full core with a permeability which is reduced to the effective permeability. So a gapped core is like a bulk core but with a smaller permeability.
The larger the gap with respect to the length of the core, the smaller the effective permeability.

If the magnetic material has a relative permeability u_r >> 1, which happens for ferromagnetic material, we can approximate the effective permeability as the ratio between the core length and gap length. In this case we have a big advantage, because the gapped core displays a permeability which is linear. Of course we still have saturation but the permeability is constant, so B linearly depends on H.

Air gap



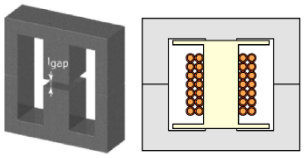
Air gap reduces the effective permeability



Changes of effective permeability and linearisation of the B-H loops caused by increasing air gap

In the right plot we are increasing the length of the gap compared to the length of the core.

Impact of the gap on the inductance value



$$L = \frac{n^2}{\left(\Re_c + \Re_g\right)} = \frac{n^2}{\Re_{eq}} = n^2 \cdot \frac{\mu_o \cdot \mu_{eff} \cdot A_c}{\ell_c}$$

Note:

$$\text{if } \Re_g \gg \Re_c \rightarrow \begin{cases} L = n^2 \dfrac{\mu_o A}{l_g} \\ A_L = \dfrac{\mu_o A}{l_g} \end{cases}$$

L is n^2 divided by the sum of the two reluctances. If the reluctance of the gap, as it is usually happening, it's much larger than the reluctance of the core we get a simple relationship for the inductance value. A_L is the **permeance** or inductance factor, i.e. the inductance we get in an inductor having a single turn of wire.

We notice that the larger the gap lg the smaller the value of the inductance if we leave all the other parameters unchanged.

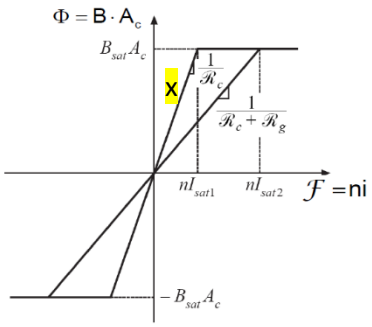
Impact of the gap on saturation

We can replace phi_sat in the first formula and solve it for the current. F = ni is the magnetomotive force. Line x is a situation where there is no gap, and we notice that the saturation magnetomotive force is quite small. If we apply a gap, the larger the thickness of the gap, the smaller the slope and the larger the saturation magnetomotive force, and so saturation current.

$$ni = \Phi\left(\mathcal{R}_c + \mathcal{R}_g\right)$$

$$\Phi_{sat} = B_{sat}A_c$$

$$I_{sat} = \frac{B_{sat}A_c}{n}\left(\mathcal{R}_c + \mathcal{R}_g\right)$$



Effect of air gap:

- decrease inductance
- increase saturation current
- inductance is less dependent on core permeability

Note:

$$\text{if } \mathfrak{R}_g \gg \mathfrak{R}_c \rightarrow I_{sat} = \frac{B_{sat}\cdot A_c}{\sqrt{L}}\cdot\sqrt{\mathfrak{R}_g}$$

I_{sat} can be increased keeping L fixed and increasing $\mathfrak{R}_g$

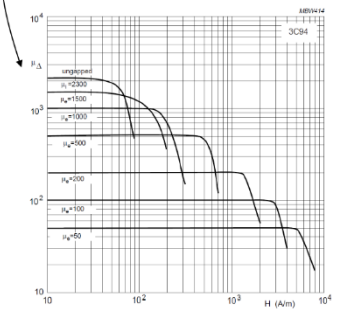
If Rg >> Rc, the saturation current can be rewritten as in the image. Hence in principle the saturation current can be increased keeping the same saturation value of the inductance by increasing Rg.

DISCRETE AIR GAPS

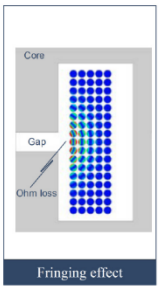
They are commonly used in ferrite cores, but there are some drawbacks:

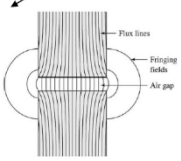
- The incremental permeability versus the H field strength shows sharp knees. If we increase the H field, which means that we are increasing the current in the inductor, and we reach and overcome the knee, the inductance value drops quickly. This is the reason why, if we use ferrite, we need to take some headroom far away from the knee. Furthermore, the knee position shifts leftward increasing the temperature.
- Discrete gaps result in inductors being vulnerable to eddy currents. In principle, flux lines should be straight in the gap region, but if the gap is consistent we have the fringing flux, so some flux lines run out of the gap region and they enter in contact with the copper wires, generating losses due to eddy currents and proximity losses.

- Discrete gaps are commonly used in ferrite cores
- The discrete gap structure results in an inductor that reaches a sharp saturation point, requiring lots of headroom in the design.
- Discrete gaps also result in inductors that are vulnerable to eddy current losses in the coil due to fringing, and to generating EMI.



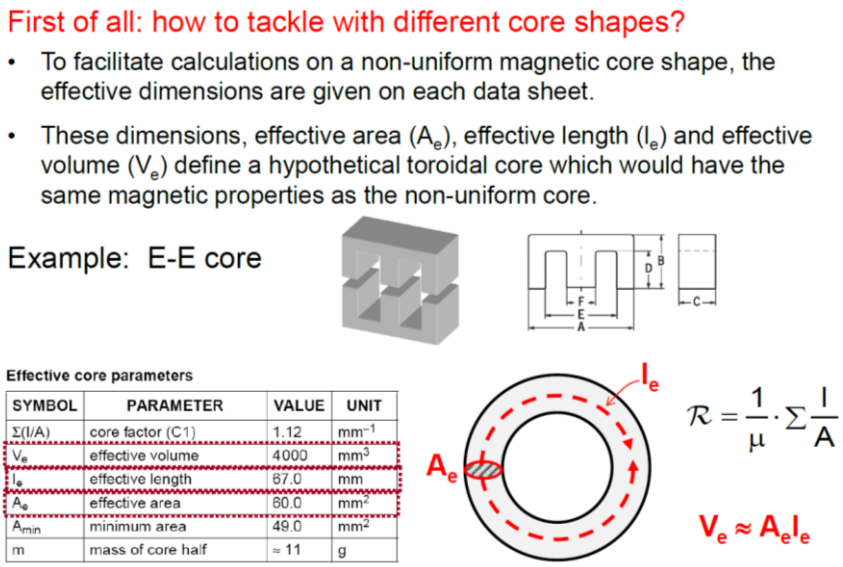
Incremental permeability vs magnetic field strenght





MAGNETIC CORE SIZE

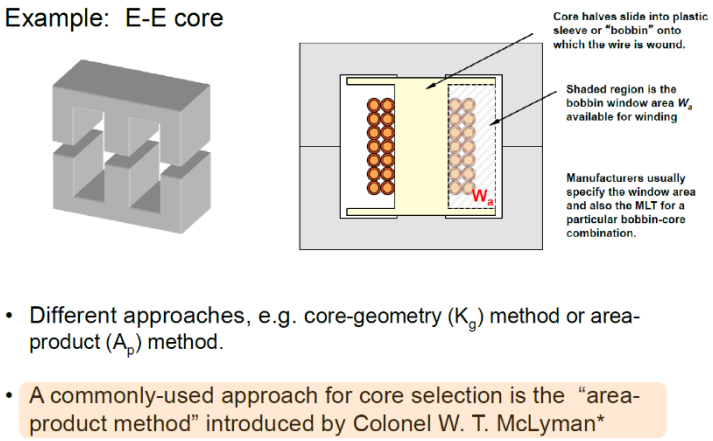
Once we have selected the material, we can focus on the core size. We might have different shapes, but we have the concepts of equivalent area, equivalent length of the magnetic core and equivalent volume. These equivalents define a prototypal toroidal core with the same properties of the core we are considering. We can use the same formulas seen for the toroidal core even if the shape of the core is different, we simply need to use these three equivalents that in general we find in the datasheets.



Another important thing in the datasheets is the core factor, which allows us to compute the reluctance of the core dividing the core factor by the magnetic permeability.

CORE SELECTION PROCEDURE

We have two different approaches to select the proper size of the core: the core-geometry method or the area-product method. The difference between the two is minimal.



As an example, we will consider the E-E core.

Area-product method

This method is based on the definition of some constraints the core has to fulfill to work properly. These constraints are: the core doesn't have to saturate and the area must be sufficiently big to allocate the wire. Then there is a third constraint that states that the current density in the copper wire must be limited to a certain value to avoid excessive power dissipation due to Joule losses in the copper.

Consider the constraints on the magnetic core.

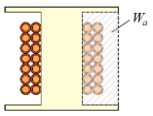
- First constraint (core saturation): $B < B_{max} < B_{sat}$

$$L \cdot \frac{di}{dt} = n \cdot A_e \cdot \frac{dB}{dt} \Longrightarrow L \cdot I_{max} = n \cdot A_e \cdot B_{max}$$

$$B < B_{max} \text{ if } \boxed{A_e > \frac{L \cdot I_{max}}{n \cdot B_{max}}} \quad (1)$$

Requested cross-sectional area to avoid core saturation

- Second constaraint (available window area): $A_{Cu} < W_a$



- Copper wire is round → dead space .
- Some insulating layers are sometimes added between windings to reduce capacitance or improve the voltage breakdown.

$$\Longrightarrow \boxed{n \cdot A_W \le K_u \cdot W_a} \quad (2)$$

- A_w = bare wire area
- K_u = window utilization factor (typ. 0.4 – 0.5)

First constraint

The amplitude of the magnetic flux density must be smaller than a maximum value that the designer is setting, which has to be sufficiently smaller than the Bsat. We need some margins because Bsat depends on the temperature.
We always start from the flux linkage.

$$\lambda = L \cdot i = n\phi$$
$$= n A_e B$$
$$B = \frac{L i}{n A_e} < B_{max}$$

This is the inequality to be satisfied. i is replaced by i_max in the worst scenario, which is the maximum current that flows in the inductor.

Second constraint

The window area is the area that is used to allocate the copper wire. The total area occupied by the copper must be smaller than the window area, otherwise we wouldn't be able to allocate the copper wire.
Usually, the copper wire is round, so we have some dead space in between the windings if we pack them. Moreover, the copper wires are never bare copper wires, otherwise we short all the windings, so copper wires come in a form where the surface is covered with an insulator that typically is a polymer.

Ku is a coefficient called **window utilization factor** and it takes the dead space and the space occupied by the insulator into account.

Third constraint

- The wire size is related to a maximum rms current density:

$$I_{rms} \le J_{max} \cdot A_W \quad (3)$$

- J_{max} = max current density, typ. 400 ~ 500 A/cm²

- Plugging eq. 3 into eq.2 we get:

$$\boxed{W_a \ge \frac{n \cdot A_w}{K_u} = \frac{n \cdot I_{rms}}{K_u \cdot J_{max}}} \quad (4)$$

Copper wire must be conveniently allocated in the bobbin

- Combining eqs. 1 and 4 we finally get:

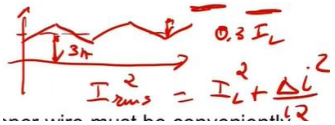
$$\boxed{A_p = W_a \cdot A_e \ge \frac{L \cdot I_{max} \cdot I_{rms}}{B_{max} \cdot K_u \cdot J_{max}} \cdot 10^4 \ [cm^4]}$$

The product of core cross sectional area and window area is the "area product" A_p

(B in Tesla, J in A/cm²)

- This gives a quick method for choosing cores. Some manufacturers specify the area product for various cores.

The wire size is related to the maximum rms current density. In general, as rule of thumb, the rms density current value Jmax is around 400 ~ 500 A/cm^2.
The area of the copper wire (Aw is the copper wire area) sets the relationship between the current and the current density.
Irms is a triangular current around a DC value of 3A with a ptp current ripple 30% of the nominal current.



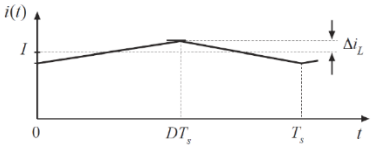
The red box has the final formula inside, where Ap is the area-product, which is the product of the area of the window Wa multiplied by the area of the cross-section of the core Ae. The factor 10^4 is a coefficient to use if we express the current density in A/cm^2.

Once we have selected the size of the magnetic core we still need to select the gap.

A note on area-product (A_p)

$$A_p \geq \frac{L \cdot I_{max} \cdot I_{rms}}{B_{max} \cdot K_u \cdot J_{max}}$$

- Assuming $\Delta i_L << I$, $I_{MAX} \approx I_{rms} \approx I$



- We can write the A_p as follows:

$$A_p \geq \frac{L \cdot I_{max} \cdot I_{rms}}{B_{max} \cdot K_u \cdot J_{max}} \approx \frac{2 \cdot W_{mag}}{B_{max} \cdot K_u \cdot J_{max}}$$

where W_{mag} is the magnetic energy stored in the inductor

→ Core size is eventually dictated by the energy that the inductor is required to store

GAP LENGTH DETERMINATION

In general, the parameter that is reported in the datasheets is not the gap length but the A_L, inductance factor. Of course, A_L and gap length are related.

Gapped cores are available from vedors with predetermined values of inductance factor, A_L .
The required A_L can be determined by rewriting eq. (1) as:

$$n > \frac{L \cdot I_{max}}{A_e \cdot B_{max}}$$

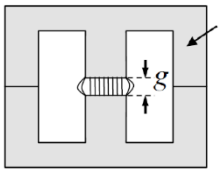
Plug into $L = n^2 \cdot A_L$ to obtain:

$$A_L < \frac{A_e^2 \cdot B_{max}^2}{L \cdot I_{max}^2} \cdot 10^{-8} \text{ [H]}$$

(B in Tesla, A_e in cm^2)

Remember that:

x $A_L \approx \frac{\mu_o A_e}{l_g}$



Fringing field can be ignored if g is small

GRADE	A_L (nH)	μ_e	AIR GAP (μm)
3C90	63 ±5%	≈ 120	≈ 320
	100 ±8%	≈ 190	≈ 175
	160 ±8%	≈ 305	≈ 100
	250 ±15%	≈ 480	≈ 55
	315 ±15%	≈ 600	≈ 40
	800 ±25%	≈ 1525	≈ 0

E 13/7/4 core. A_L measured in combination with a non-gapped core half

The coefficient 10^-8 comes from the fact that Ae is expressed in cm^2.

In general, in a gapped core the dominant reluctance is the reluctance of the gap, so we can write A_L as x.
Moreover, the inequality in the blue box sets an upper boundary for A_L. This inequality has a well-defined physical meaning; by setting this inequality we are saying that the volume of the gap (Ae*lg) confines most of the energy. Hence the volume of the gap must be sufficiently large to allocate the energy that has to be stored in the inductor.
The energy stored in the inductor is ½ * (B^2)/u0 * Ae*lg. The gap volume, and so the gap length, must be such to allocate the energy in it.

Winding turns calculation

Once A_L is known, the last step is to determine the number of turns in order to get the value of inductance we want. $L = n^2 * A_L$

Once A_L is known n can be easily calculated base on the desired L value:

$$n \geq \sqrt{\frac{L_{min}}{A_L}}$$

Further design steps:

- select copper wire with suitable AWG ;
- check A_{Cu}/W_a ;
- calulate the core loss;
- calculate the copper loss;
- calculate the inductor temperature increase;
- iterate design, if needed

➢ we'll see all these steps later...

So we can select the core size with the area product method, then we can select the inductance factor and finally the number of turns. There are further steps which will be investigated later on.

Example 1

Let's assume we are targeting the 3F3 ferrite, which is a medium frequency ferrite that can be used for applications with switching frequency between 0.2 and 0.5 MHz.

Design Example #1

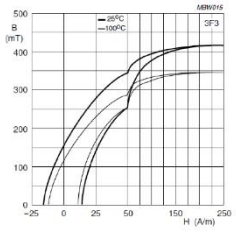
Material choice: 3F3

3F3 SPECIFICATIONS

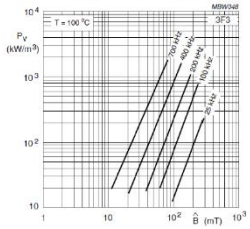
A medium frequency power material for use in power and general purpose transformers at frequencies of 0.2 - 0.5 MHz.

SYMBOL	CONDITIONS	VALUE	UNIT
μ_i	25 °C; ≤10 kHz; 0.25 mT	2000 ±20%	
μ_a	100 °C; 25 kHz; 200 mT	≈ 4000	
B	25 °C; 10 kHz; 1200 A/m	≈ 440	mT
	100 °C; 10 kHz; 1200 A/m	≈ 370	
P_V	100 °C; 100 kHz; 100 mT	≤80	kW/m³
	100 °C; 400 kHz; 50 mT	≤150	
ρ	DC; 25 °C	≈ 2	Ωm
T_C		≥200	°C
density		≈ 4750	kg/m³

FERROXCUBE



Typical B-H loops



Specific power loss as a function of peak flux density with frequency as a parameter

If we look at the upper BH diagram it seems that there is a discontinuity, but it is not. We have that notch because we are using different scales on the x axis.

The saturation flux density on the BH curve clearly depends on temperature. At low temperature we have 440 mT, but at 100°C the saturation flux density drops to 370 mT. We want to design the inductor in a safe way avoiding saturation, so we need to take some margins when selecting Bmax. In this example we select Bmax = 300 mT.

The magnetic core is made available by FERROXCUBE. Circuit specifications are in the image. Kw is the window utilization factor.

Design Example #1

60

Circuit spec:
- $L = 8\ \mu H$
- $I_{dc} = 3A$
- $\Delta i_L = 0.9\ A$

Assumptions:
- $B_{max} = 0.3\ T$ (3F3)
- $K_w = 0.5$
- $J_{max} = 500\ A/cm^2$

Required area-produc A_p
- $A_P > 110\ mm^4$

E13/7/4 core satisfies this criterion. Using this core, the design requires:
- $A_L < 150\ nH$

For 3F3 ferrite a std gapped core is available with:
- $A_L = 100\ nH$

From this we find:

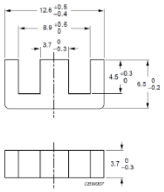
$$n \geq \sqrt{\frac{L_{min}}{A_L}} = 8.9 \rightarrow 9\ \text{Turns}$$

FERROXCUBE

Ferroxcube E13/7/4 bobbin and core parameters

Effective core parameters

SYMBOL	PARAMETER	VALUE	UNIT
Σ(l/A)	core factor (C1)	2.39	mm^{-1}
V_e	effective volume	369	mm^3
l_e	effective length	29.7	mm
A_e	effective area	12.4	mm^2
A_{min}	minimum area	12.2	mm^2
m	mass of core half	≈ 0.9	g



Winding data and area product for E13/7/4 6-pins coil former

NUMBER OF SECTIONS	WINDING AREA (mm^2)	MINIMUM WINDING WIDTH (mm)	AVERAGE LENGTH OF TURN (mm)	AREA PRODUCT Ae x Aw (mm^4)
1	11.6	7.1	24	144

Core halves

A_L measured in combination with a non-gapped core half, clamping force for A_L measurements, 15 ±5 N.

GRADE	A_L (nH)	μ_e	AIR GAP (μm)	TYPE NUMBER
3F3	63 ±5%	≈ 120	≈ 320	E13/7/4-3F3-A63
	100 ±8%	≈ 190	≈ 175	E13/7/4-3F3-A100
	160 ±8%	≈ 305	≈ 100	E13/7/4-3F3-A160
	250 ±15%	≈ 480	≈ 55	E13/7/4-3F3-A250
	315 ±15%	≈ 600	≈ 40	E13/7/4-3F3-A315
	700 ±25%	≈ 1330	≈ 0	E13/7/4-3F3

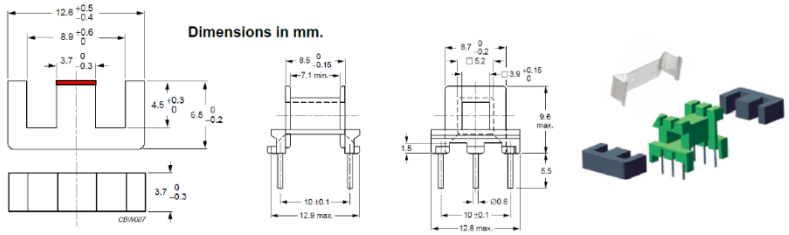
Beware of the tolerance: L =7.45 μH < L_{min} with $A_L = A_L - 10\%$

We put the specs and assumptions in the Ap equation. Then we need to go in the catalog of the manufacturer to find the core that provides an Ap larger than the calculated lower bound.
Then the other value to compute is the maximum value for the inductance factor A_L. The last step is the selection of the number of turns. If the number is discrete, we round it up to the next integer value.

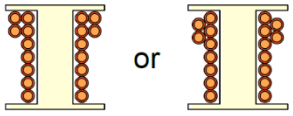
NB: A_L is not a fixed number, it has a tolerance that comes from the mechanical tolerance of how the gap is made. If we want to be precise, we have to check that in the worst case we still are satisfying the minimum value of inductance. In our case it leads to a L = 7.45 uH, which is still acceptable. Otherwise, we need to increase the number of turns.

Design Example #1

6



Further steps:
- select copper wire with suitable AWG size → AWG19
- check A_{Cu}/W_a → 50% → Warning: bobbin cannot accommodate a single layer! Not really an issue in DC inductors
- core loss → 17 mW
- copper loss → 55 mW
- core temperature increase → ~ 7 °C



using the empirical formula $R_\theta = 53 \cdot (V_e)^{-0.54}$ [°C / W] where V_e is in cm^3

→ Design is OK …..

The found 9 turns must be rounded around the bobbin, which is a plastic material associated and combined with the core. The problem is that the internal length of the bobbin is 7mm, so we have to check that 7mm is sufficient to accommodate 9 turns. If not, we have to use a multilayer winding. Of course we need also to select the copper wire before doing this check. The copper wire diameter has to be chosen so that the current density matches the specifications. If the diameter is 1mm we cannot accommodate 9 turns, just 7, so we need to go for a second layer. The orthocycled winding (right) reduces the dead space. The other advantage is that also the lateral space occupied by the windings is seriously reduced, which from the datasheet is (8.7 – 5.2)/2 and we should stay in this limit.

The next step is to compute the core loss. To compute it, we need the Pv-B_hat plot. So we need to compute delta(B) and then B_hat will be delta(B)/2. L*delta(i) = n*Ae*delta(B) is the formula we need for delta(B) computation. To get the value in tesla we need the Ae in m^2.
We get B_hat = 31mT, which corresponds to Pv = 40mW/cm^3. This is the core loss, but we want the power loss per unit volume, so we need to multiply this by the volume of the effective volume.

Instead, the core temperature increase is computed using an empirical formula.

Example 2

Design Example #2

FERROXCUBE 62

Circuit spec:
- L = 8 µm
- I_{dc} = 3A
- Δi_L = 0.9 A

Assumptions:
- B_{max} = 0.3 (3F3)
- K_w = 0.5
- J_{max} = 500 A/cm^2

Required area-produc A_p
- A_P > 110 mm^4

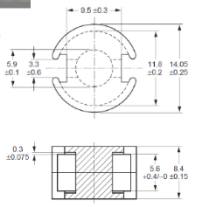
PC 14/8/I core satisfies this criterion.
Using this core, the design requires:
- A_L < 800 nH

For 3F3 ferrite a std gapped core is available with :
- A_L = 400 nH

From this we find:

$$n \geq \sqrt{\frac{L_{min}}{A_L}} \approx 4.47 \rightarrow \text{ 5 Turns}$$

P14/8/I bobbin and core parameters

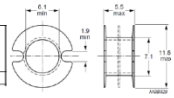


Effective core parameters

SYMBOL	PARAMETER	VALUE	UNIT
Σ(l/A)	core factor (C1)	0.700	mm^{-1}
V_e	effective volume	628	mm^3
l_e	effective length	21.0	mm
A_e	effective area	29.9	mm^2
A_{min}	minimum area	23.6	mm^2
m	mass of set	≈ 3.5	g

Winding data and area product (for P14/8/I) for CP-P14/8 coil former

NUMBER OF SECTIONS	WINDING AREA (mm²)	MINIMUM WINDING WIDTH (mm)	AVERAGE LENGTH OF TURN (mm)	AREA PRODUCT Ae x Aw (mm⁴)
1	8.8	4.1	28.9	263



Core sets for general purpose transformers and power applications
Clamping force for AL measurements, 60 ±20 N.

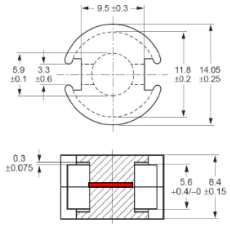
GRADE	A_L (nH)	μ_e	AIR GAP (µm)
3F3	100 ±3%	≈ 56	≈ 470
	160 ±3%	≈ 89	≈ 270
	250 ±3%	≈ 140	≈ 160
	315 ±5%	≈ 176	≈ 120
	400 ±5%	≈ 224	≈ 95
	2400 ±25%	≈ 1340	≈ 0

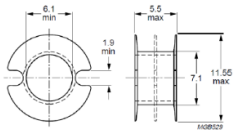
Beware: min. L =9.5 µH with A_L=A_L-5% → may want to get a custom gap!

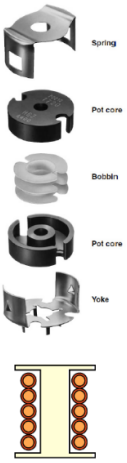
We use another type of core instead of the E-E core, the Pot core.

In this case we need 5 turns and the internal height of the bobbin is 5mm, so we can accommodate a single layer of winding. A single layer is better because if we have multiple layers we might have serious problems with proximity effects.

Design Example #2







Further steps:

- copper wire AWG size → AWG19
- check A_{Cu}/W_a → 37% (quite conservative)
- core loss → 20 mW
- copper loss → 35 mW
- core temperature increase → ~ 4 °C

  using the empirical formula $R_\theta = 53 \cdot (V_e)^{-0.54}$ [°C / W] where V_e is in cm^3

→ We can try the P11/7/I…..

Example 3

Design Example #3 (aggressive)

64

We go for the smaller pot core PC11/7/I, having an area product of 91.2 mm^4.

Using this core, the design requires:

- A_L < 340 nH

For 3F3 ferrite a std gapped core is available with :

- A_L = 315 nH

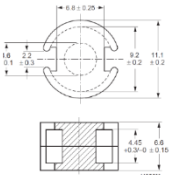
From this we find:

$$n \geq \sqrt{\frac{L_{min}}{A_L}} \approx 5 \rightarrow 5 \text{ Turns}$$

P11/7/I bobbin and core parameters

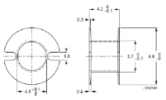
Effective core parameters

SYMBOL	PARAMETER	VALUE	UNIT
Σ(l/A)	core factor (C1)	0.860	mm^{-1}
V_e	effective volume	309	mm^3
l_e	effective length	16.3	mm
A_e	effective area	19.0	mm^2
A_{min}	minimum area	13.7	mm^2
m	mass of set	≈ 1.9	g



Winding data and area product (for P11/7/I) for CP-P11/7-1S coil former

NUMBER OF SECTIONS	WINDING AREA (mm^2)	MINIMUM WINDING WIDTH (mm)	AVERAGE LENGTH OF TURN (mm)	AREA PRODUCT Ae x Aw (mm^4)
1	4.8	3.1	22.6	91.2



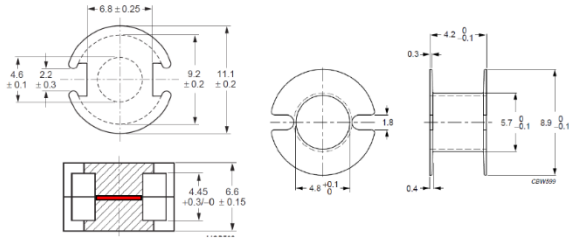
Core sets for general purpose transformers and power applications

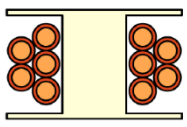
Clamping force for AL measurements, 60 ±20 N.

GRADE	A_L (nH)	μ_e	AIR GAP (µm)
3F3	63 ±3%	≈ 43	≈ 500
	100 ±3%	≈ 68	≈ 290
	160 ±3%	≈ 109	≈ 170
	250 ±5%	≈ 171	≈ 100
	315 ±5%	≈ 215	≈ 75
	1750 ±25%	≈ 1195	≈ 0

Beware of the tolerance: L =7.48 µH < L_{min} with A_L=A_L-5%

Design Example #3





Further steps:

- copper wire AWG size → AWG19
- check A_{Cu}/W_a → 68% !
- core loss → 30 mW
- copper loss → 27 mW (assuming proximity effect is negligible)
- core temperature increase → ~ 6 °C

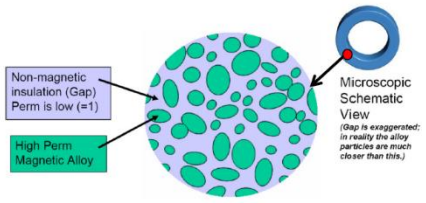
→ Very tight design. May want to replace AWG 19 with AWG 20…

POWDER CORES

It is an alternative to the ferrite core. It is also called distributed gap core.
If we look at the core at the microscopic level, we notice that there is a large quantity of tiny particles which are made of a high permeability material like iron or iron alloys which are dispersed in a non-magnetic insulation with low permeability.

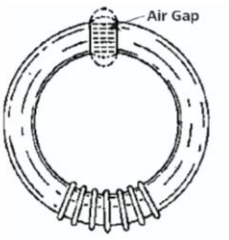
The advantage of powder cores is that there is no discrete gap, so the gap is distributed all along the magnetic core, not physically concentrated in just one location. As a consequence, the saturation is not sharp.
The other advantage is that the size of the particles is small, so eddy currents can only flow within the single particle and they are strongly reduced.

- At a microscopic level, magnetic alloy powder grains are separated from one another by binder insulation.
  (This is not at the magnetic domain level; domains are vastly smaller than powder core grains.)



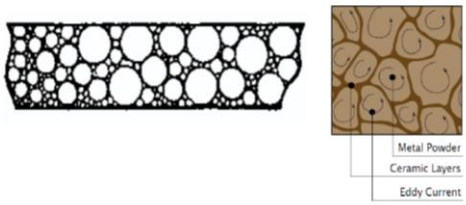
- Distributing the gap throughout the powder core structure serves two main purposes:
  1) eliminating the disadvantages of a discrete gap structure, which are sharp saturation, fringing loss,and EMI, and
  2) controlling eddy current losses so that higher B_{sat} alloys may be used at relatively high frequencies, despite comparatively low bulk resistivity in the alloy.

DISCRETE VS DISTRIBUTED GAP CORE



Discrete air gap

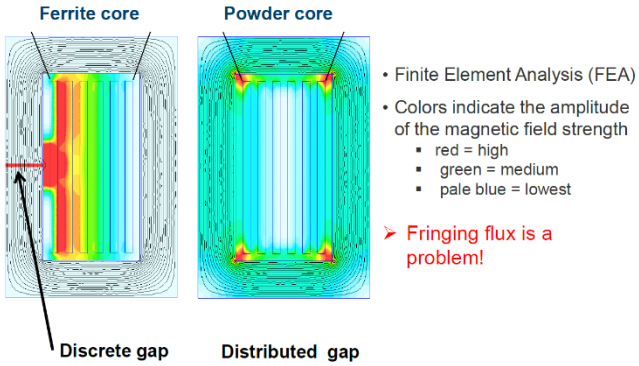
- A comparatively large air gap is introduced into a basically high-permeability material that is part of the path of the circuit.



Distributed air gap

- A very large number of small air gaps is distributed throughout the core.
- Distributed air gap allows the opportunity to obtain much larger effective gaps in the magnetic path, since it minimizes second-order effects such as leakage and fringing flux.

If we consider the fringing flux, powder cores are much better. In the discrete gap case we have a large magnetic field close to the gap and, the larger the magnetic field in this region the larger the losses due to eddy currents.
Instead, in a distributed gap core we don't have a high field region.



Disadvantages

We have reduced the eddy current losses thanks to the small particles but we still have them, so in general core losses in a distributed gap core are larger than in a ferrite core. The other disadvantage is the cost, which is in general higher.

	MPP	High Flux	Kool Mμ®
Permeability	14-550	14-160	26-125
Core Loss	Lowest	Moderate	Low
Perm vs. DC Bias	Better	Best	Good
Saturation (B_{sat})	7.5 Kilogauss	15 Kilogauss	10.5 Kilogauss
Nickel Content	80%	50%	0%
Curie Temp. (°C)	460	500	500
Relative Cost	High	Medium	Low
Available Permeability	14,26,60,125,160, 200,300,500	14,26,60,125,160	26,60,75,90,125
Frequency Range	60Hz - 1MHz	60Hz - 200kHz	60Hz - 500kHz

- High Flux: 50% Ni, 50% Fe powder core
- MPP: 80% Ni, 18% Fe, 2% Mo powder core
- Kool Mμ: Fe, Al, Si powder core
- Powdered Iron: Fe, Si powder core

MPP is the short for Moli-permalloy, High flux is nickel and iron, Kool Mu is made of iron, aluminum and silicon (also called sendust).

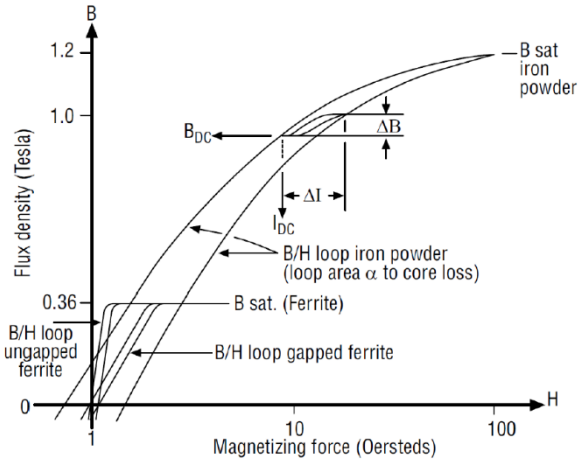
These three materials can be used as base for distributed gap cores.

SELECTION GUIDELINES FOR POWDER CORES

- For the lowest loss inductor, MPP material should be used since it has the lowest core loss.
- For the smallest core size in a DC bias dominated design, High Flux material should be used since it has the highest flux capacity.
- Kool Mu (sendust core) is well known for the moderate cost and has significantly lower losses and substantially better thermal properties when compared to powdered iron cores.

GAPPED FERRITE VS POWDER CORES

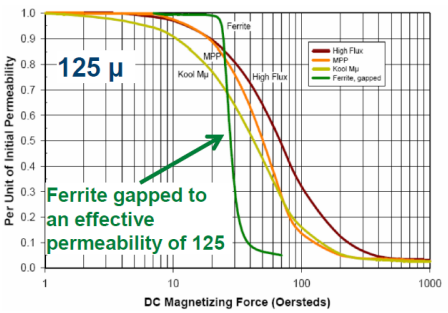
The ungapped ferrite saturates immediately at very low H. Furthermore, ungapped ferrite has also a low saturation flux density Bsat. If we introduce a gap we stretch the ungapped curve.
Instead, the B-H diagram for a distributed gap core has a much larger Bsat and it occurs very softly because of the distributed gap and the stochasticity of the distribution of the gap.

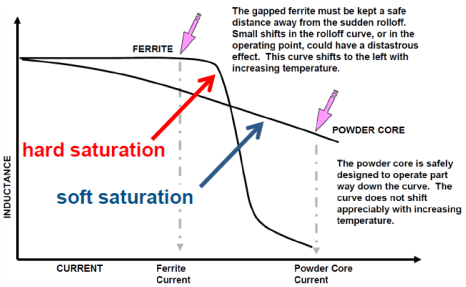


Saturation characteristics

The left plot reports the permeability u on the y axis as a function of the magnetic field strength normalized to the initial permeability on the x axis. So when the magnetic field is 0 we read 1 on the x axis. When we increase the magnetic field, the permeability stays the same for a while and then if we have a ferrite at a certain point we have a sudden drop.
The powder core materials don't have such a sharp saturation.

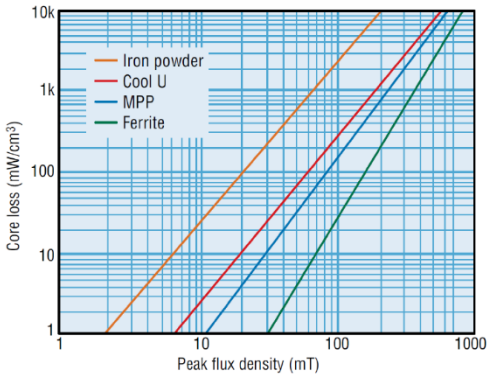
In general, with distributed gap cores even if we design the core in a bad way, we never really get into troubles.
Moreover, the other difference is that the knee of the ferrite moves back as the temperature increases. Instead, for distributed gap cores, the saturation curves don't depend on the temperature.





- Powder materials saturate gradually, even when the current load increases significantly.
- A gapped ferrite will maintain an inductance closer to the unbiased value until saturation occurs, where a sudden drop in inductance is seen.
- The flux capacity of any power ferrite is reduced significantly as temperatures rises, while the flux capacity of powder cores remains essentially constant over temperature.

Core loss comparison



@50kHz

Frequency	10 kHz	100 kHz	500 kHz	1 MHz
Ferrite	1×	1×	1×	1×
MPP	2×	5×	9×	12×
Kool Mμ	2×	9×	18×	20×
Amorph. Choke	2×	15×	25×	25×
High Flux	5×	15×	40×	40×
Powdered Iron	20-40×	25-60×	25-100×	13-21×

In the table we have a comparison between the loss in a ferrite core and the various distributed gap cores. Ferrite core is taken as a reference. If we focus on fsw = 100 kHz MPP are dissipating five times the ferrite cores.

In the plot we have the specific core loss, so the core loss per unit volume, as a function of the peak value of the flux density. These losses are measured at 50 kHz and the comparison is between iron powder core, Cool U, MPP and ferrite. The power losses in distributed gap cores are fairly larger than in the ferrite cores.

SUMMARY

- **Soft Saturation**: ferrite must be designed in the safe, flat area of the rolloff curve. Powder cores are designed to exploit the controlled, partial roll-off in the material.
- **Flux Capacity**: with more than twice the flux capacity of ferrite at 50% inductance roll-off, powder cores can provide a reduction in required core size of up to 35%.
- **Temperature**: flux capacity of ferrites decreases with temperature while powder cores stay relatively constant.
- **Fault-tolerance**: powder core designs are inherently fault-tolerant with soft saturation curves, whereas gapped ferrite is not.
- **Fringing Losses**: do not occur with powder cores, but can be excessive with gapped ferrites.

MATERIAL CHOICE

The starting point in our design was to build an inductor with an inductance value of 8 uH, which is the lower bound. The average current in the inductor is 3A.
We want to implement this inductor using a distributed gap core and the material is Kool Mu, which is a good compromise between cost and performance. It shows relatively low core loss compared to the high flux material, the saturation characteristic is soft and there are also other advantages summarized in the table.

Moreover, the alloy on which this material is based is an iron-aluminum-silicon alloy which shows an almost 0 magnetic striction (*it's a property of magnetic materials that causes them to change their shape or dimensions during the process of magnetization*).

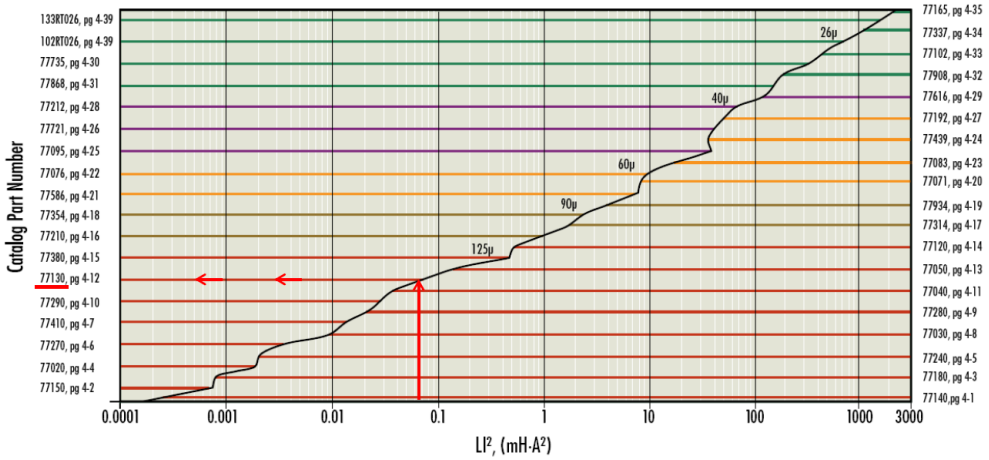
Magnetics Kool Mμ®

- Near zero magnetostriction alloy → ideal for eliminating audible frequency noise in filter inductors.

	Kool Mμ®
Permeability	26-125
Core Loss	Low
Perm vs. DC Bias	Good
Temperature Stability	Very Good
Temperature Rating	200° C continuous
Saturation Characteristic	Soft
Nickel Content	0%
Relative Cost	Low

SELECTOR CHART FOR KOOL MU TOROIDAL CORES

Once we have chosen the shape of the core, toroidal in our case, and the material, the manufacturer provides selection chart.



- Charts give the optimum permeability and smallest core size for DC bias applications.
- Charts are based on a permeability reduction of not more than 50% with DC bias, typical winding factors of 40% for toroids and an AC current that is small relative to the DC current.
- Charts are based on the nominal core inductance and a current density 500-600 A/cm^2.

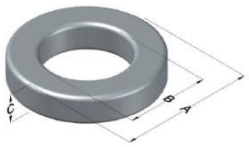
In principle, we can always use the Ap product procedure seen in the previous examples to pick up the distributed gap cores, but this is not necessary because of the existence of these selection charts. This chart is though to allow the designer to pick up the core with the optimum permeability and the smallest size compatible with the application. So **by using this chart we get the optimum core size and permeability**.

On the x axis there is the product of L, the inductance, and i^2, the current in the inductor. Li^2 is proportional to the energy stored in the inductor, so basically we are selecting the core on the base of the energy stored in the inductor.

In some cases, if we want to be extra-safe we might think to use the maximum current, which is the sum of the average current and half of the ripple.

Once Li^2 has been computed we look at the chart and we move vertically till we cross the black line. We crossed it in the region with red horizontal line, which corresponds to a relative permeability of 125. Once we have selected the permeability we have to continue vertically upwards until we cross the first colored horizontal line (in this case it is immediately above the black line). Then we move horizontally and we get the suggested part number of the toroidal core with the best sizes for the application we are targeting.

1. Compute the product of LI^2 where L = inductance required with DC bias (mH) and I = DC current (A)
   → $LI^2 = 0.008(mH) \times 9\ (A^2) = 0.072\ mH \cdot A^2$
2. Using the Kool Mμ LI^2 chart locate 0072 $mH \cdot A^2$ on the bottom axis. Following this coordinate vertically results in the selection of 77130 as an appropriate core for the above requirements.
3. From the 77130 core data page read the inductance factor, $A_L (nH/T^2)$
   - note: $L = n^2 \frac{\mu A}{l} = n^2 \cdot A_L$ …A_L is the inductance corresponding to a single turn



See part 7_additional reading materials: 7.4_77130 data sheet

Kool Mμ Permeability (μ)	A_L (nH/T²)	Core Marking			Coating Color
		Lot Number	Part Number	Inductance Grade	
125	53 ± 12%	XXXXXX	130A7	N/A	Black

Dimensions	Uncoated		Coated Limits			Packaging
	(mm)	(in)	(mm)	(in)		
OD (A)	11.2	0.440	11.9	0.465	max	Bulk Pack 4 bags/box Box Qty= 6000 pcs
ID (B)	6.35	0.250	5.84	0.230	min	
HT (C)	3.96	0.156	4.60	0.181	max	

Once we have the part number we look at the catalog and the most important parameter is the inductance factor A_L, that is the inductance we would get from the core by just using a single wire core. Using this parameter we can compute the inductance of the inductor with the formula in the red box. But since we know the target value for the inductance, we can solve the formula for n to get the number of turns.
Of course, A_L is a value with a tolerance, and we have to put ourselves in the worst case scenario, that is 53 – 12%, because the target inductance is a lower bound, we don't want to fall below 8uH, so we take the minimum A_L.

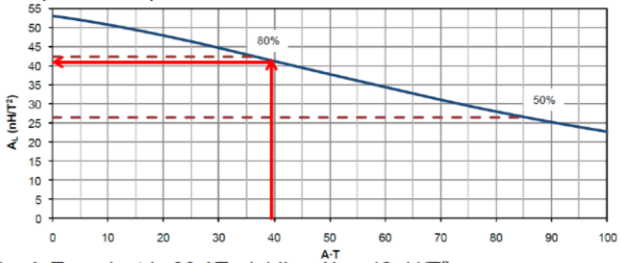
… A_L of the selected core is 53 nH/T^2 ± 12%. The minimum A_L (worst case) is core is $A_L \cdot 0.88\ nH/T^2$.

4. Compute the number of turns needed to obtain 8 μH at no load

$$n = \sqrt{\frac{L(\mu H) \cdot 1000}{A_L(nH/T^2) \cdot 0.88}} = 13$$

5. To calculate the number of turns required at full load, determine the DC Bias Ampere ·Turn product and check the effective inductance fator A'_L

x



77130 data sheet

… the A·T product is 39 AT, yielding $A'_l \approx 42 nH/T^2$

Looking at the image, we get n = 13, which is a first guess on the number of turns that has been calculated by considering a zero current flowing through the inductor. This is of course not true because the current in the inductor is not 0, we have 3A. These 3A will produce a magnetic field that will be pushing us toward saturation, so we need to understand the impact of this current on the permeability and, in turn, on the inductance factor.

To perform this check, the manufacturer is providing the plot x. A_L is displayed as a function of the A*T product, which is the ampere-turn product. The A*T product consists in the multiplication of the current that flows in the inductor by the number of turns, that in our case is 3*13 = 39. Then we go vertically in the plot until we cross the blue line and we get the corresponding A_L value, 42 (reduced from 53).

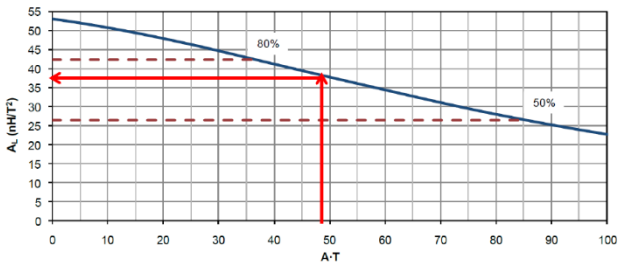
We are interested in working close to saturation because if we work at higher magnetic flux density close to saturation, the higher the flux density the smaller the size of the core. So we can take advantage of the soft saturation to get close to saturation with no risks because there is no sharp knee like in ferrite, the approach is very smooth, and doing this we are able to squeeze the size of the magnetic core.
In general, from the selection chart of page 241 we pick up a core where the effective inductance factor is between 50% and 80% of the nominal inductance factor.

However, there is a problem. Since the permeability at full load (3A) is not 53nH/T^2 but 42, we need to increase the number of turns to restore the desired inductance value, otherwise our final inductance will be smaller than the target. The new number of turns is adjusted with the empirical formula below.

6. Addjust the number of turn according to

$$n' = n\frac{A_L}{A_L} = 16 \qquad \text{empirical formula!}$$

7. Re-calculate the Ampere ·Turn product and check the new inductance fator A"$_L$.



… the A·T product is 48 AT, yielding A"$_L \approx 37.5$nH/T^2

We notice that the formula for n' doesn't have a sqrt dependance on A_L. This is due to the fact that, if we increase the number of turns, we are also increasing the magnetizing field (magnetic field strength) and, since we are working close to saturation, we are decreasing A_L.

Now we have to check if n' is a good result with the same procedure of before. A_L'' = 37.5 nH/T^2.
The final check is, by using A_L'', that the L value is correct.

8. Calculate the inductance of the core with 16 turns and with 48 AT DC bias level

$$L(\mu H) = \frac{n'^2 \cdot A_L''(nH/T^2)\cdot 0.88}{1000} = 8.45\mu H \quad OK!$$

… the minimum inductance requirement of 8.0 µH has been achieved at full load .
Should the final L be unsatisfactory, iterate again.

9. Select the wire size using the AWG wire table

AWG Wire Size	Resistance Ω/meter (±.305±Ω/ft)	Wire O.D. (in) Heavy Build	Wire Area sq. cm² (±0.001)	Current Capacity, Amps (Rated by columns of Amps/cm²) 200	400	500	600	800
6	.00130	.421	139.2	26.6	53.2	66.5	79.8	106
7	.00163	.376	111.0	21.1	42.2	52.8	63.3	84.4
8	.00206	.336	88.7	16.7	33.5	41.8	50.2	66.9
9	.00260	.299	70.2	13.3	26.5	33.2	39.8	53.1
10	.00328	.267	56.0	10.5	21.0	26.3	31.6	42.1
11	.00414	.238	44.5	8.34	16.7	20.8	25.0	33.3
12	.00521	.213	35.6	6.62	13.2	16.5	19.8	26.5
13	.00656	.1902	28.4	5.25	10.5	13.1	15.8	21.0
14	.00828	.1715	23.1	4.16	8.33	10.4	12.5	16.7
15	.01044	.1529	18.4	3.30	6.61	8.26	9.91	13.2
16	.01319	.1369	14.72	2.62	5.23	6.54	7.85	10.5
17	.01658	.1224	11.77	2.08	4.16	5.20	6.24	8.32
18	.02095	.1095	9.42	1.65	3.29	4.11	4.94	6.58
⇨ 19	.02640	.099	7.54	1.31	2.61	3.27	3.92	5.22
20	.03323	.0879	6.07	1.04	2.08	2.59	3.11	4.15
21	.04190	.0785	4.84	0.823	1.65	2.06	2.47	3.29
22	.05315	.0701	3.86	0.649	1.30	1.62	1.95	2.59
23	.06663	.0632	3.14	0.518	1.04	1.29	1.55	2.07

NB: with distributed gap cores we want to work on purpose close to the saturation, taking advantage of the soft saturation, in order to get the minimal size core.

If the L value we find after the first iteration is not satisfactory we have to go through another iteration.

The next point is to select the wire size. This is typically easy because the guideline is that the current density in the wire must be, from specs, around 500 A/cm^2, and the manufacturer provides a table where we find the current densities and the Irms that flows in the copper wire producing that current density.
The target Irms = 3A, so we look in the 500 A/cm^2 column until we find the closest value to 3A.

We can also read the specific resistance per unit length (Ohm/m) and the outer diameter of the copper wire, which has an insulating layer covering it.

Winding factor

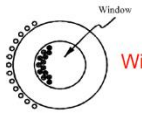
The last step is to compute the winding factor, which is the ratio between the area occupied by the copper inside the window and the area of the window. The window area of the toroidal core is found in the datasheet (x).

- The wire table indicates that 19 AWG is needed to carry 3.0 A with a current carrying capacity of 500 A/cm^2.
- 16 turns of 19 AWG (wire area = 0.754 mm^2) equals a total wire area of 12.06 mm^2.
- The window area of a 77130 is 26.8 mm^2. Calculating window fill, 12.06 mm^2/26.8 mm^2 corresponds to an approximate 45% winding factor.

Physical Characteristics						
Voltage Breakdown wire to wire min (V_{AC})	Break Strength min (kg)	Window Area W_A(mm^2)	Cross Section A_e (mm^2)	Path Length L_e (mm)	Volume V_e (mm^3)	Weight (g)
1000	12.0	26.8	9.06	26.9	244	1.5

77130 data sheet

x y z



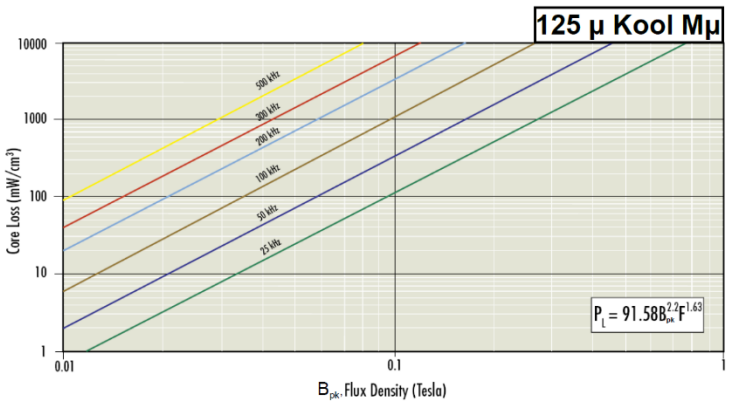
Winding factor usually limited to 50% !

- A 77130 with 16 turns of 19 AWG will meet all the requirements for this inductor.

Now the inductor is finally built, but there are some checks we need to perform, that are the computation of the core loss and copper loss and of the temperature increase.

CORE LOSS

Power dissipated inside the core, and they can be calculated using the graph provided by the manufacturer which is specific core loss vs peak value of the AC component of the flux density.
Bpeak is half of delta(B). To get Bpeak in T, we need the A in m^2. The area is found in the table of the previous image, Ae (y).



$$B = \frac{L \cdot \Delta I}{2 \cdot N \cdot A} = 0.0248\ T$$

$$P_C' = 91.58 \cdot B_{pk}^{2.2} \cdot f^{1.63} = 470\ (mW/cm^3)$$

note : f in kHz

$$Volume = 244\,mm^3$$

$$\Rightarrow \boxed{P_C = 114\,mW}$$

$$V_L = NA\frac{dB}{dt} = L\frac{dI}{dt} \Rightarrow \Delta B = \frac{L \cdot \Delta I}{N \cdot A} \Rightarrow \boxed{B_{pk} = \frac{L \cdot \Delta I}{2 \cdot N \cdot A}}$$

Once we have the specific core loss from the plot we have to multiply it by the volume (again we have to look at the previous table, z) to get the core loss.

COPPER LOSS

We are speaking about losses in the copper wire by which the inductor is made. If the current in the inductor is almost a DC current the computations are quite easy, while if the current has strong AC components at high frequency (big ripples), we will have skin effects and proximity effects that are increasing the resistance of the copper wire by orders of magnitude, thus increasing the copper loss.

Let's start from a simple situation where the current flowing in the inductor is either a DC current or an AC current with a sufficiently small frequency not to have skin effects. In this case the computation of the power dissipated by the copper is gained by computing the DC resistance of the wire and multiplying it by the squared rms value of the current.

DC resistance of wire

$$R = \rho \frac{\ell_b}{A_w}$$

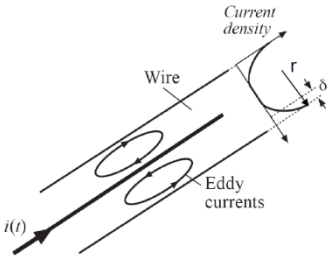
where A_w is the wire bare cross-sectional area, and ℓ_b is the length of the wire. The resistivity ρ is equal to $1.724 \cdot 10^{-6}\ \Omega$ cm for soft-annealed copper at room temperature. This resistivity increases to $2.3 \cdot 10^{-6}\ \Omega$ cm at 100°C.

The wire resistance leads to a power loss of

$$P_{cu} = I_{rms}^2 R$$

NB: the resistivity increases with temperature. Moreover, when we select the copper we have the resistance per meter value, and to get the total resistance we need to multiply it by the total length of the wire.

High frequency copper loss

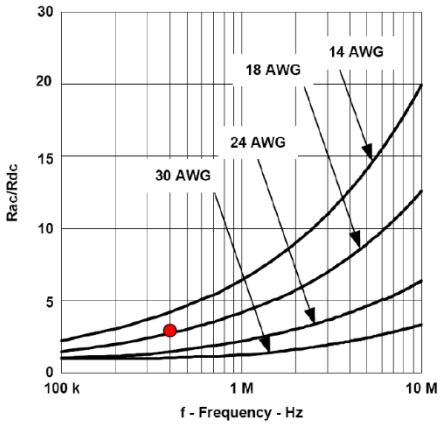


x $R_{ac} = R_{dc} \dfrac{r^2}{2r\delta - \delta^2}$ $\quad \delta \le r$

$\delta = \delta(f)$

For copper at room temperature, sinusoidal excitation:

$$\delta = \frac{7.5}{\sqrt{f}}\ \text{cm}$$



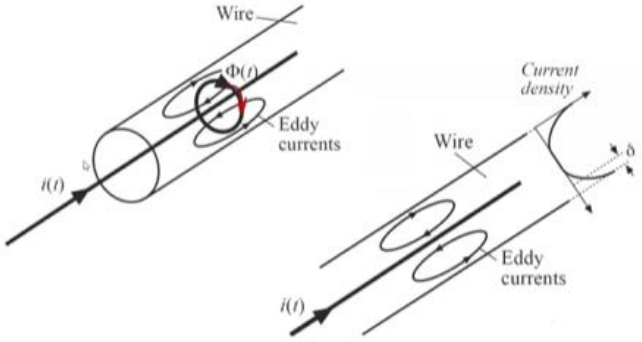
f expressed in Hz

Let's suppose to push a sinusoidal current with a given frequency in the copper wire. However, our current is made of a DC value to which it is superposed a triangular AC signal. Nevertheless, the we can decompose the triangular AC part in its spectral components and analyze them singularly (in fact they will be sinusoidal).

The first thing if we consider an isolated copper wire is the skin effect. It is an effect according to which if we increase the frequency of the current, most of the current will concentrate on the surface of the copper wire.

The skin effect is a consequence of the eddy currents that we are generating inside the wire. When we push a sinusoidal current in the wire, using the right hand rule we are generating a magnetic field that goes like loops inside the copper wire and due to the magnetic field we are inducing eddy currents.
The direction of the eddy currents is such that they produce a magnetic field that is opposing to the magnetic field established by the current.

Due to this effect, eddy currents in the central portion of the copper are opposing the current, while eddy current close to the surface will be adding to the current, increasing it close to the surface. The overall current is not changed, what we are changing due to the skin effect is the distribution of the current density.



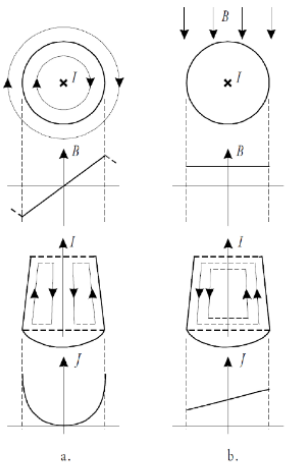
The skin effect has a consequence on the resistance of the copper wire and, in turn, on the power dissipation.
We can think as if the current is concentrated in an annular region close to the surface at high frequency, where the thickness of the annular region is equal to the penetration depth.
Since we are not using all the available copper wire, the resistance seen by the current will be large than the resistance in the case of a low frequency current.

The difference in resistance is given by the ratio of the total area of the copper wire and the area of the annular region, as in x. When the penetration depth delta is equal to the radius if there is no skin effect and AC and DC resistances are the same.

Proximity effect

When we speak of a skin effect we are considering an isolated copper wire where we are pushing HF current and due to the HF current the current density is concentrating towards the periphery of the copper wire. So it is the magnetic field generated by the current itself that determines the redistribution of the current itself.

Instead, in the case of proximity effect we have a copper wire with a current flowing in it. The current is a LF current, so there is no skin effect. However, there is a variable external magnetic field in proximity. The consequence is that we are generating inside the copper wire loops of eddy currents whose direction depends on the right hand rule.



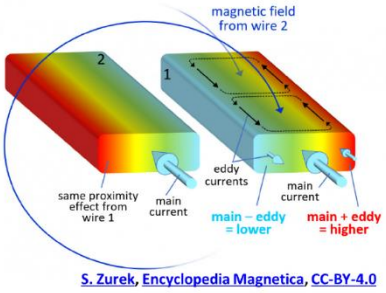
a. Skin effect and b. proximity effect in round conductors.

Two conductor proximity effect

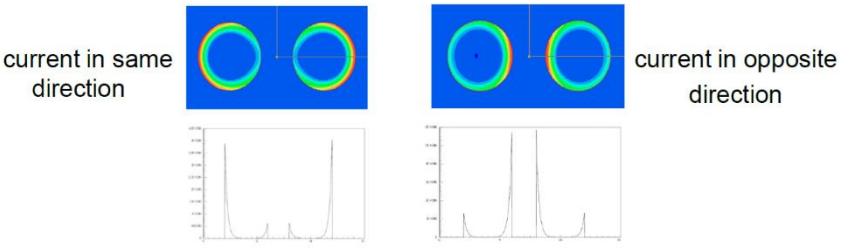
Let's we consider two copper wires close to each other and that there is current entering in one direction in both the wires. If we focus on the copper wire on the left, the current that flows in it is producing a magnetic field which is impinging on the adjacent copper wire, inducing eddy currents. The direction of the eddy current is so to generate a magnetic field that opposes to the one that has generated them.

Due to the generation of eddy currents we have an increase of the current close to the external surface of the copper wire and a decrease, since they counteract the external flow of current, in the internal volume of the copper wire.

- Ac current in a conductor induces eddy currents in adjacent conductors by a process called the *proximity effect*. This causes significant power loss in the windings of high-frequency transformers and ac inductors.



S. Zurek, Encyclopedia Magnetica, CC-BY-4.0

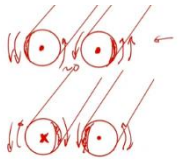


There is a simple way to understand where the surface current is concentrating in a copper wire in a situation where we have high frequency magnetic fields and high frequency currents.

Golden rule for HF current

Eddy currents are always acting in a way to prevent the penetration of magnetic field inside the conductor.

According to the golden rule, if we have two adjacent wires with current in the same direction, the current is concentrating on the surface because in the inside of the wire the H field tends to cancel out if the copper wires have the current in the same direction. Instead, if the currents are in two different directions, the H field tends to build up in

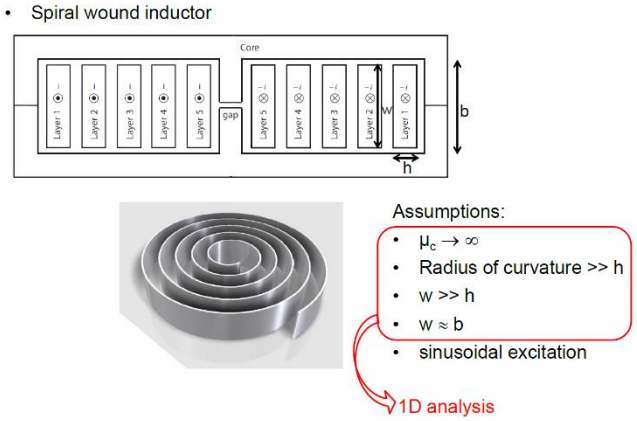


the region in between the copper wires and tends to cancel out in the outer region, so current density is concentrated towards the inner surface.

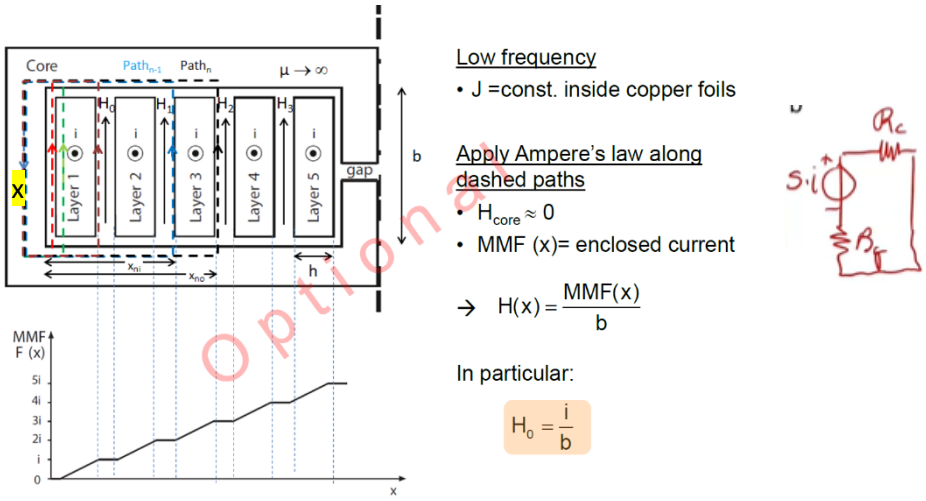
Example: proximity losses in a 5 layer inductor

The inductor includes 5 turns. We need to make some assumptions. b is the breath of the magnetic core, i.e. the height of the window.

If these assumptions are verified we can perform a 1D analysis.



In the next image we have half of the inductor and let's assume that the current flowing in the inductor, that enters in each layer (dots) is a LF one, so the current density is constant in each layer.
First of all, the H field inside the magnetic core is almost 0 (it is 0 if the permeability is infinite), and this can be demonstrated using and electrical equivalent of the magnetic circuit with a magnetomotive force of 5*i, the reluctance of the core and the reluctance of the gap.
We want to calculate the magnetomotive force drop Fc across the reluctance Rc. Fc = Hc*lc = 5i*Rc/(Rc+Rg). lc is the length of the magnetic core. If the reluctance Rg goes to infinite the permeability goes to 0, so the core permeability goes to zero and so Hc tends to zero.

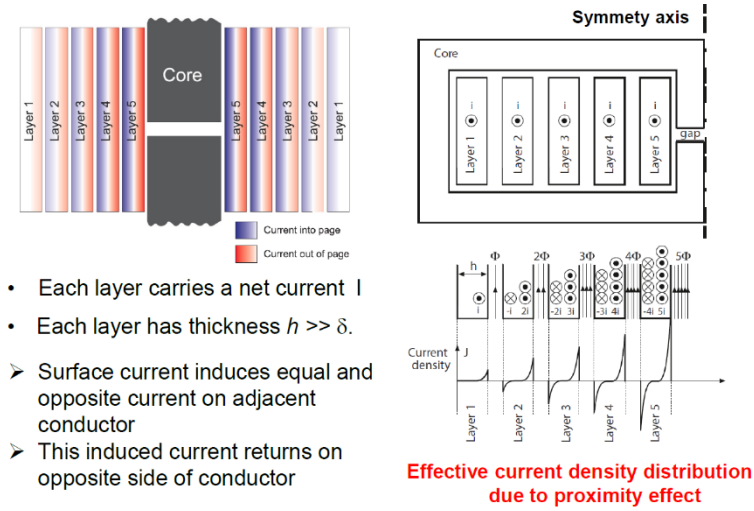


We can compute the magnetic field strength in the window between adjacent layers (H0, H1, …) using the magnetomotive force plot. It is a plot (the one in the bottom left) of the MMF as a function of the position x moving inside the window starting from the boundary of the core that can be drawn using the Ampere's law (the total magnetomotive force along a closed loop is equal to the current that crosses the loop). So let's consider an amperian loop, that is the red one in the image. One boundary is x and fixed, the other one is moving inside the layers.

If we move inside we are intercepting some current, but not all the one in the layer, just a slice of it; the more we slide into, the larger the current we intercept so the larger the magnetomotive force. Between one layer and the other the magnetomotive force is flat because we are not intercepting any current.

To compute the magnetic field inside the region between adjacent layers we consider the current enclosed by the amperian loop and the H field is just in the layer region. In particular, H1 = 2*H0, H2 = 3*H0 and so on.

Since current flowing on a surface is proportional to the H field next to that surface, if we consider layer 1 in the next image, we have 0 H field on the left and 1 H0 on the right. So the current will be concentrated next to the right surface of the layer, because since we have one unit of H0 we need one unit of current, which in this case corresponds to the net current entering the layer.
In the second layer we have 1 H0 on the left and 2 H0 on the right surface. So we will have two units of current close to the right surface and one unit of current close to the left surface. However, the net current must still be i, so current is entering on the right surface and exiting on the left surface. So the additional units of current are like an eddy current contribution recirculating in the layer. In the third layer we will have two eddy currents, and so on.



The problem is that the more we move to the left the more the parasitic currents increase, and this is a pain in the back because it increases power dissipation.

High frequency limit

What is the impact of the current concentrating on the surface?
This analysis is valid only for high frequency, where the penetration depth is negligible with respect to the height of the copper foil.

If we consider the first layer, most of the current is confined in a region close to the surface of the copper foil, with a thickness equal to the penetration depth (we can consider a constant current density extending over a thickness equal to the penetration depth). So this current is using just a fraction of the copper foil, hence we can understand what the AC resistance of the copper foil is at HF by making the ratio of the copper areas; the effective used copper area is delta/copper_size.

The current $i(t)$ having rms value I is confined to thickness d on the surface of layer 1. Hence the effective "ac" resistance of layer 1 is:

$$R_{ac} = \frac{h}{\delta} R_{dc}$$

This induces copper loss P_1 in layer 1:

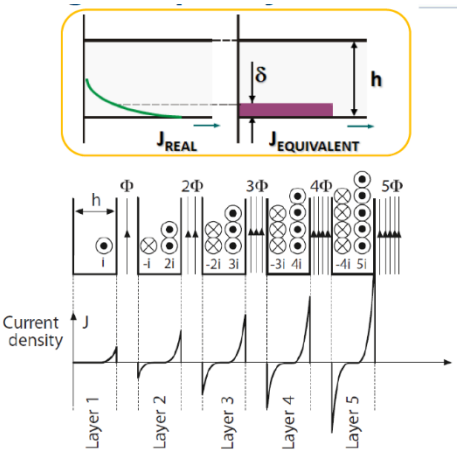
$$P_1 = I^2 R_{ac}$$

Power loss P_2 in layer 2 is:

$$P_2 = P_1 + 4P_1 = 5P_1$$

Power loss P_3 in layer 3 is:

$$P_3 = \left(2^2 + 3^2\right) P_1 = 13P_1$$



Power loss P_m in layer m is:

$$P_m = I^2 \left[(m-1)^2 + m^2 \right] \left(\frac{h}{\delta} R_{dc} \right)$$

Instead, in the second layer we have a current close to the left surface and two units of current close to the right surface, so the total power is P1 + 4P1.

NB: the current is negative (entering) on the right side and positive on the left side because the magnetic field strength inside the copper wires at HF must be 0, and this is an effect of the eddy currents, which tend to prevent the penetration of the magnetic field inside the conductors. Since the current inside the conductor is 0, we can apply the KCL to a loop comprising the right surface of layer 1 and the left surface of layer 2 and the net current crossing the loop must be 0 because the magnetic field is 0.

If we have M layers, the formula in the next image gives the total power dissipation.

Add up losses in each layer:

$$P = I^2 \left(\frac{h}{\delta} R_{dc} \right) \sum_{m=1}^{M} \left[(m-1)^2 + m^2 \right]$$
$$= I^2 \left(\frac{h}{\delta} R_{dc} \right) \frac{M}{3} \left(2M^2 + 1 \right)$$

Compare with dc copper loss:

If a dc or low-frequency ac current of rms amplitude I were applied to the M-layer winding, its copper loss would be

$$P_{dc} = I^2 M R_{dc}$$

So the proximity effect increases the copper loss by a factor of

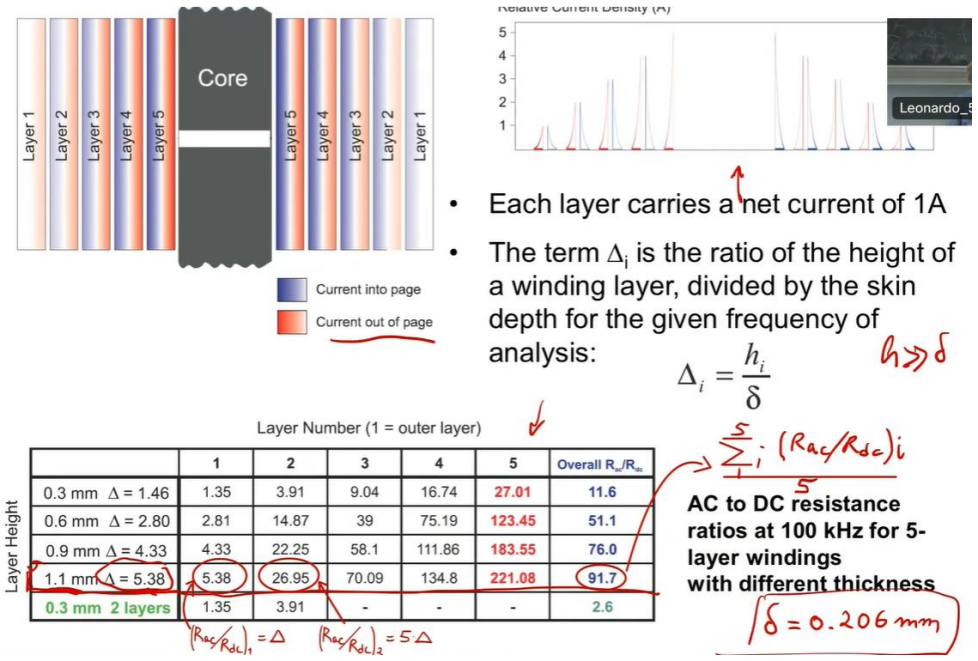
$$F_R = \frac{P}{P_{dc}} = \frac{1}{3} \left(\frac{h}{\delta} \right) \left(2M^2 + 1 \right) = R_{ac}/R_{dc}$$

This expression is valid for a foil winding having h>>δ.

If we were in DC, the power dissipated would be I^2*Rdc*M. Instead, the Pdc/Pac ratio is called proximity effect factor → the larger the number of layers the larger the power dissipated in AC with respect to the one dissipated in DC.

We can notice that the higher the frequency, the smaller the penetration depth and the higher the proximity effect factor.

The impact of this is dramatic. In the HF assumption (h >> delta), a 5-layer inductor has a ratio between AC layer resistance and DC layer resistance is 221.08.

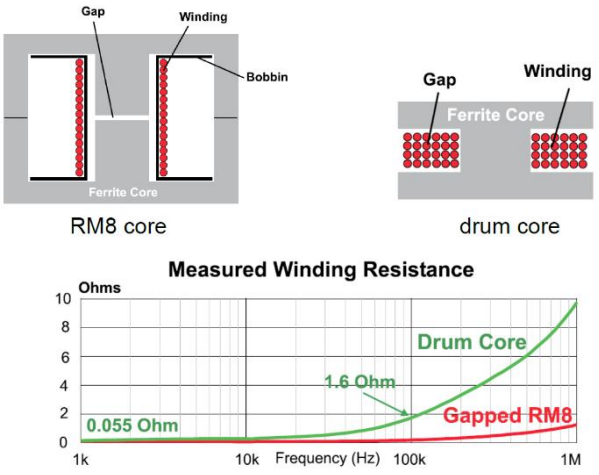


	1	2	3	4	5	Overall R_{ac}/R_{dc}
0.3 mm Δ = 1.46	1.35	3.91	9.04	16.74	27.01	11.6
0.6 mm Δ = 2.80	2.81	14.87	39	75.19	123.45	51.1
0.9 mm Δ = 4.33	4.33	22.25	58.1	111.86	183.55	76.0
1.1 mm Δ = 5.38	5.38	26.95	70.09	134.8	221.08	91.7
0.3 mm 2 layers	1.35	3.91	-	-	-	2.6

The overall Rac/Rdc is the sum of the values of all the layers divided by 5, and we get 91.7.
The conclusion is that the AC resistance of this 5-layer inductor with the copper height of 1.1mm and operation frequency of 1kHz is 90 times larger than the DC one. Hence the AC power dissipation is huge with respect to the DC one. This is the dramatic effect of proximity effect. In an inductor with multiple layers, the AC power dissipation is orders of magnitude greater than the DC one.

Proximity losses in inductors

In the image we have two inductors with the same DC resistance value, 0.8 Ohm. The difference is that in one case (left) we are using a pot-core and a single layer or windings, while in the other case we have 6 layers because the breath of the inductor is much smaller. At DC we don't have any difference, but if we increase the switching frequency up to 1MHz, the RM8 core has a limited increase in resistance, of more or less 1 Ohm, while the resistance of the drum core increases a lot (factor 200).



The important message we can take from this is that if we want to avoid serious proximity effect we should stick to an inductor having the maximum breath so that we can allocate a single layer of copper wire.

The thing is completely different in transformers, because in transformers we can fight the proximity effect with **winding interleaving**.

(For a deeper insight on proximity effect check part 7 from slide 98 to 108)

Total loss in M-layer winding

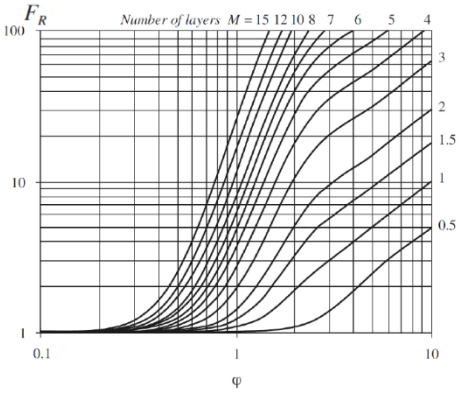
Starting from the following equation:

$$F_R = \varphi \cdot \left[G_1(\varphi) + \frac{2}{3}(M^2 - 1) \cdot (G_1(\varphi) - 2G_2(\varphi)) \right]$$

We can report the curves for the proximity effect factor Fr as a function of phi = h (copper foil height) / delta (penetration depth). What changes between one curve and the other is the number of layers. The curves are known as **Dowell's curves**.

Proximity effect factor F_R

- $F_R = R_{ac}/R_{Ddc}$
- φ= *layer thickness/penetration depht*
- Large diameter
  - Significant R_{ac}/ R_{dc}
- Large layer count
  - Require diameter << δ
- Non-sinusoidal waveforms
  - High frequency harmonics



R_{ac} to R_{dc} ratio for due to the proximity effect, as a function of φ =h/δ and layers number, for sinusoidal excitation

If we go to the region where h >> delta, so phi is more or less 10, the Fr we get is the Fr we can get in the simplified approach. Moreover, if the number of layers is fixed, e.g. 3, the larger the thickness of the foil or the diameter of the copper wire, the higher the Fr, so AC dissipation will be much larger than DC dissipation.

At a fixed h/delta, i.e. if we move vertically, **the larger the number of layers the larger the Fr**.
NB: Dowell's curves are obtained with a perfect sinusoidal excitation and assuming square copper wires, so we need some geometrical manipulations if we use round wires, according to the formula:

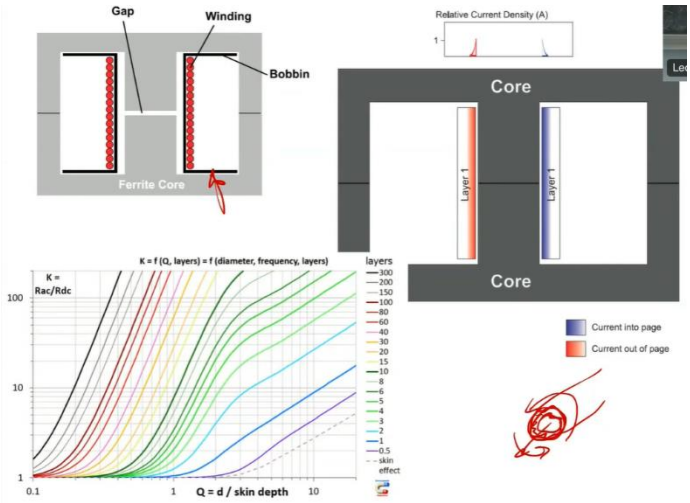
$$\varphi = \frac{h}{\delta'} = \sqrt{\eta}\sqrt{\frac{\pi}{4}}\frac{d}{\delta}$$

where d is the diameter and eta is the winding porosity, typically 0.8.

Skin effect or proximity effect?

Let's assume we have an inductor as in the image, with just one single layer of round copper wire. To have a pure skin effect, we would need a wire that is isolated in the universe. But when we put it inside the window of a magnetic core, the magnetic field is concentrating on the internal surfaces of the copper wires, current is concentrated close to the core.
This is much worse than skin effect, because in skin effect the current distributes all along the perimeter of the conductor, here it is just close to the internal surface.

Even with a single layer of turn the proximity effect is still there. Moreover, if in a region we have the superposition of fringing and proximity effect we might have serious troubles, the temperature might skyrocket and the copper wires melt together.



HF COPPER LOSS IN A TOROID WITH SINGLE LAYER

The simplified analytical treatment used for an E-E core or pot core is not possible in a toroidal inductor because it is a 2D structure that cannot be reduced to a 1D structure, so we have to compute the Fr with numerical methods.

We are interested in the red curve. In our case we are working at 400 kHz, and D/delta = 8. The red curve is intercepted at a factor Fr = 5.

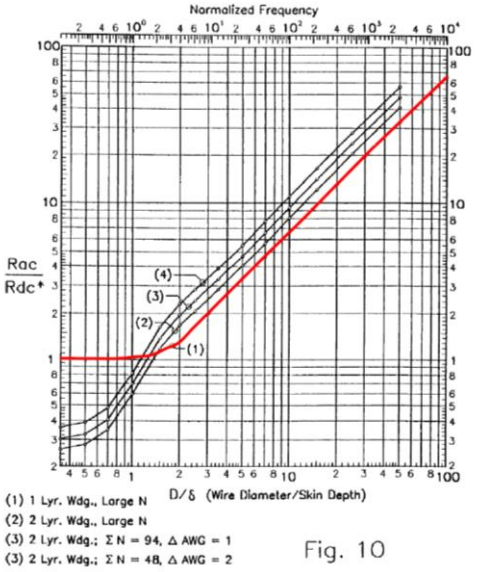
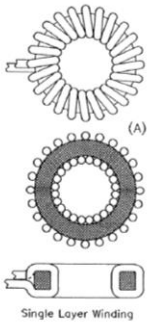


Fig. 10

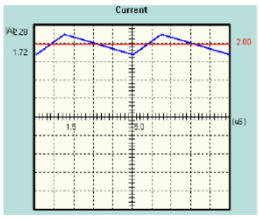


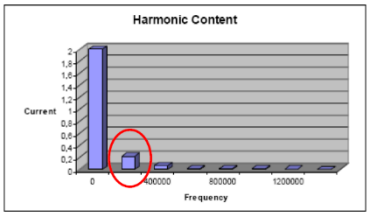
Curve (1) in figure is the AC/DC resistance ratio vs. D/ δ for a single layer toroidal winding, assuming;
- many turns of relatively fine wire;
- a core OD/ID ratio of nearly 2.0.

Winding loss

However, our current is not a sinusoidal one, it has a DC contribution plus a triangular ripple. Firstly we need to get rid of the DC component, which produces a loss that can be computed using the DC resistance of the copper wire. As for the AC contribution, if it is pure sinusoidal we just read the amplitude and we use the Dowell curve. But in this case we have a triangular contribution, and it can be split in its harmonic components. In principle we should take the rms value of the harmonics, multiply it by the Rac calculated at 400kHz fir all the harmonics.

However, the amplitude of second and third harmonics is negligible with respect to the fundamental component, so we can do some simplifications, and this is the reason why from the sum we are left with I1, that is like having a pure sinusoidal waveform.





- CCM operation with 30% peak-to-peak current ripple.
- We can just consider the fundamental harmonic (weigth of higher order harmonics is negligible)
- Split the dc and ac current components

$$I_{dc} = I_O = 3A$$

$$I_{ac} = \sqrt{I_{rms}^2 - I_{dc}^2} = \sqrt{\sum_{i=1}^{\infty} I_i^2} = 0.26A \approx I_1 \quad \left(\text{note: } I_{rms} = \sqrt{I_o^2 + \frac{\Delta i^2}{12}} \right)$$

So the rms value of the AC component can be ascribed fully to the first harmonic.
We can compute the winding loss as below.

$$P = I_{dc}^2 \cdot R_{dc} + I_1^2 R_{ac}$$

Rdc is the number of turns in our inductor multiplied by the specific resistance of the copper wire multiplied by the average length per turn. The average length per turn can be found in the datasheet and it depends on Fr and it is not just the perimeter of the toroidal because the winding it's a bit skewed.

$$R_{dc} = n_{turns} \cdot \frac{length}{turn} \cdot R_{specific} = 16 \cdot 18.6mm \cdot 2.64 \cdot 10^{-5} \Omega / mm = 7.8m\Omega$$

Winding Length Per Turn			
Winding Factor	(mm)	Winding Factor	(mm)
0%	15.2	40%	18.1
20%	16.7	45%	18.6
25%	17.0	50%	19.0
30%	17.4	60%	19.9
35%	17.8	70%	20.9

77130 data sheet

$$Q = \frac{D}{\delta} \sim 8 \ @ \ 400kHz$$

Using the the AC/DC resistance ratio graph for toroidal core with a single layer

$$R_{ac} = 5 \cdot R_{dc} \sim 40 \ m\Omega$$

Therefore:

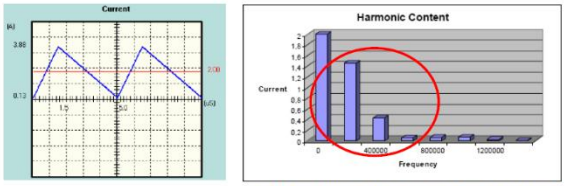
$$P_W = I_o^2 \cdot R_{dc} + I_{ac}^2 \cdot R_{ac} = 70mW + 2.7mW \approx 73 \ mW$$

In the end Rac = 5*Rdc and we can compute the power dissipated in the copper windings. We notice that, just by chance in our design, the AC contribution is negligible, but it is an expected result because we are designing a DC inductor, and most of the current is DC current.

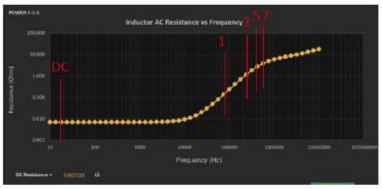
Current that is not a DC current

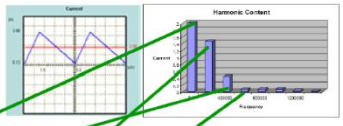
If we have an inductor that is working, for instance, at the boundary between CCM and DCM, we cannot neglect the second and third harmonics. Hence we have to split the current in its harmonic components and, for each component, by using the Dowell curve, we compute the Rac and then we compute the total power with the formula in the image. It is the DC power plus the rms of the fundamental current multiplied by Rac at the frequency of the fundamental plus all the other contributions.

- The current waveform should be split properly into its full spectrum of frequency components, and each of these used with the corresponding ac resistance calculation.



1) Extract harmonic content





2) Calculate ac resistance for each harmonic

$$P = I^2_{DC} * R_{DC} + I^2_{rms\_1} * R_{AC\_1} + I^2_{rms\_2} * R_{AC\_2} + I^2_{rms\_3} * R_{AC\_3} + \ldots$$

3) Calculate the winding loss

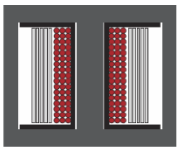
Non-sinusoidal waveforms – time domain approach

There is another approach, which is in the time domain. We firstly build the inductor, and once it is built, we can extract the corresponding Dowell curves. Then we can use a software to extract an electrical model of the inductor starting from the Dowell curve, taking the proximity effect into account. We can put it in Pspice and then working in the time domain to compute the power dissipation.

The model has a DC resistance plus a ladder network of resistances and inductances. At DC only the DC resistance survives, while if we increase the frequency, the most bottom inductor opens and we are adding a resistive contribution and so on with the increase in frequency.

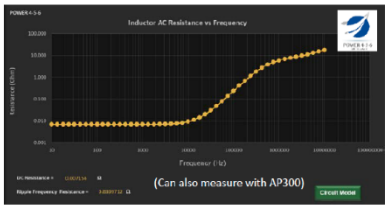
- Alternatively, a suitable equivalent circuit can be extracted from Dowell's curve to simulate proximity losses.

Step 1
Define the Winding Structure

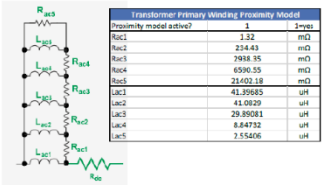


Find n, Rdc and L

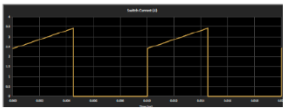
Step 2
Sweep Dowell's Eqs for AC Resistance



Step 3
Circuit Model to Match Predictions



Step 4
Simulate Waveforms - all Proximity Loss is Time Domain



Dr. Ray Ridley, ridleyengineering.com

CORE TEMPERATURE RISE

The last step is to compute the temperature rise of the inductor. Core power dissipation and copper dissipation has been computed in the previous steps, and the temperature rise can be computed with the formula x, where we are using R_theta which is the thermal resistance.

- **Total loss**

$P_{TOT} = P_C + P_W \sim 190\ mW$

- **Temperature rise**

x $\Delta T = P_{TOT} \cdot R_\theta \approx 20\,°C$ ☺

Using the empirical formula $R_\theta = 53 \cdot (V_e)^{-0.54}$ [°C / W], where V_e is in cm^3

Delta(T) is not the final temperature, but the temperature increase with respect to the ambient temperature.

The next image summarized the output of the software that each manufacturer of magnetic components provides.

Design Output

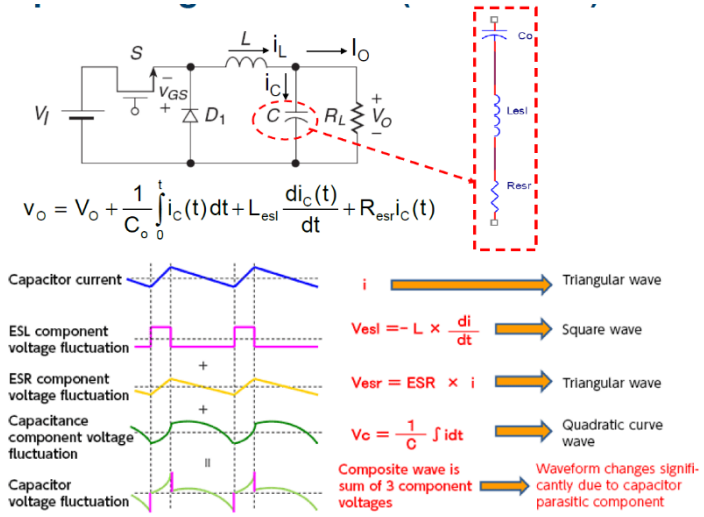
Core Information		Inductor Information	
Part Number	77130-A7	Inductance at Full Load (µH)	8.04
ID (mm, nominal)	6.3	Inductance at No Load (µH)	13.57
OD (mm, nominal)	11.2	Effective Permeability at Full DC	74.0
Height (mm, nominal)	4.1	Wound Core Dimen. (mm)	14.5 x 8.6
Permeability	125	Number of Turns	16
Core Area (sq cm)	0.093	Wire Size (AWG)	19
Path Length (cm)	2.75	Winding Factor	0.46
AL (mH/1000 turns)	53	DC Resistance of Winding (Ohms)	0.007
Recommended Header	TV-H2206-4A	Core Loss (mW)	95.1
Select Core from List	Select	Copper Loss (mW)	70.4
English / Metric		Total Losses (mW)	165.6
		Temperature Rise (degrees C)	13.7

OUTPUT CAPACITOR SELECTION

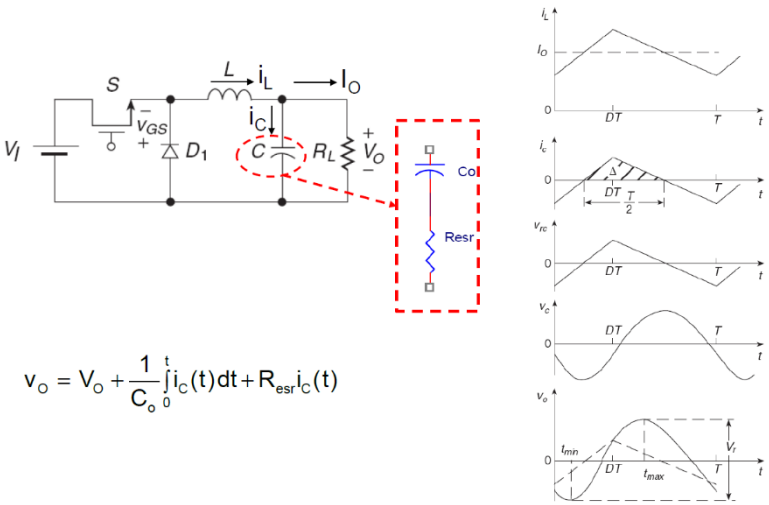
- **ESR** is an abbreviation for **E**quivalent **S**eries **R**esistance, the characteristic representing the sum of resistive (ohmic) losses within a capacitor.
- While ESR is undesirable, all capacitors exhibit it to some degree. Capacitor manufacturers look to ways to reduce the ESR of the capacitor…
- The **ESL** is rarely specified by manufacturers. Becomes significant at high frequencies.

The real capacitor is typically a series of a capacitance, a resistance ESR and an inductance ESL. In general, both ESL and ESR are unwanted contributions to the capacitance.
The ESR of a capacitor is always quoted in the datasheet, but the ESL not always. However, in general ESL is so small that comes into play only at very large frequencies, so we can neglect it.

BUCK CONVERTER



The current in the capacitor is the inductor current with the DC component that flows in the load removed, so we are left with a triangular current with 0 average. This current flows in the series of the three components of a real capacitor.

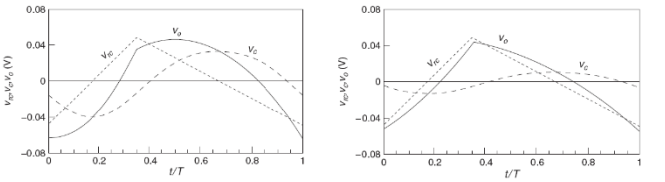


The ESL produces a square voltage waveform contribution because we have to derive the current to get the voltage on the ESL. The ESR produces a triangular contribution and the capacitance produces pieces of parabola.

$$\Delta v_{ESR} = \Delta i_L \cdot ESR$$

$$\Delta v_C = \frac{\Delta i_L}{8 C_O f_{SW}}$$

$$\Delta v_O \approx \frac{\Delta i_L}{8 C_O f_{SW}} + \Delta i_L \cdot ESR \Longrightarrow \text{Conservative estimation!!}$$

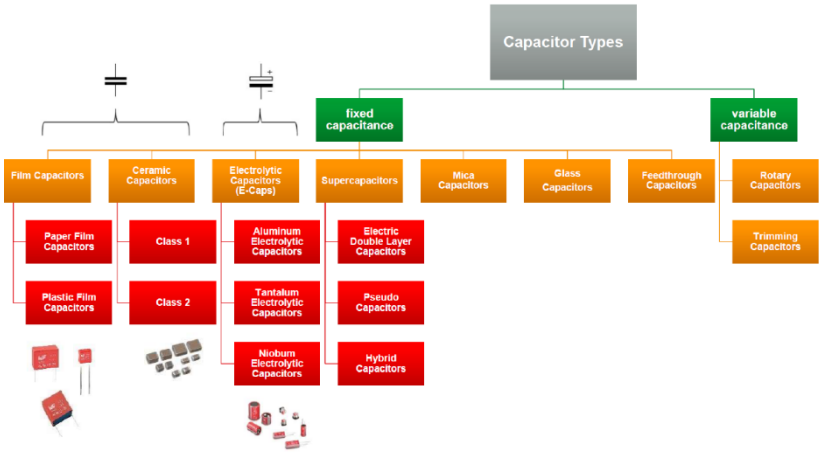


$\Delta v_{ESR} \sim \Delta v_c$ $\Delta v_{ESR} > \Delta v_c$

The overall peak to peak voltage ripple at the output is the sum of the ESR and capacitive contributions. This is a conservative estimation because capacitive and ESR contributions are shifted by 90°, so the real peak to peak ripple is smaller than the sum of the two contributions.

However, when delta(Vesr) >> delta(Vc), which is the typical situation for electrolytic capacitors, the total peak to peak ripple is dominated by the ESR contribution.

CAPACITOR TECHNOLOGIES



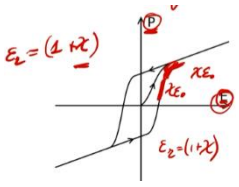
Typical capacitors in PE applications are film, ceramic and electrolytic capacitors (mainly the last two).

Ceramic capacitors

They are divided in two subfamilies, class 1 and class 2, and it depends on the type of dielectric material with which the capacitor is made. Class 1 has a dielectric with a relative permittivity epsilon_r which is typically a factor of 10. Class 2 has a epsilon_r as high as few thousands, so their volumetric efficiency is much larger than class 1, i.e. we can get a large value of capacitance in a small volume. We cannot go always for class 2, because there are some drawbacks. Class 1 provide a capacitance that is stable with applied voltage and temperature, class 2 provide a capacitar that strongly depends on the applied voltage and temperature.

This because the material out of which the class 2 capacitor is made is a ferroelectric material. It means that if we plot the electric polarization P as a function of the electric field E, which depends on the voltage applied between the plates of the capacitor, we get a curve which shows saturation and hysteresis.

The slope of the curve is the product of the electric susceptibility x and epsilon0. The more we increase E the closer we are to saturation. 1+x is equal to the relative electrical permittivity of the dielectric and, due to saturation, we have a reduction of the permittivity.
Ceramic capacitors have also a negligible ESR.

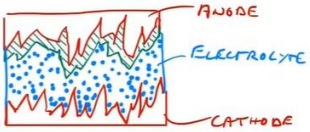


Electrolytic capacitor

The other capacitor family we can consider is the one of electrolytic capacitor, among which we have aluminum, tantalum and niobium electrolytic capacitors.
The difference between aluminum and tantalum is in the metal with which the metal plates are made. The difference is the dielectric, which is obtained by growing an oxide on top of the metal plate with an electrochemical process. If we use Al as starting metal we get an aluminum oxide, if we use tantalum we get tantalum pentaoxide. The difference is the permittivity, which is higher for tantalum pentaoxide, so volumetric efficiency of tantalum capacitors is higher than aluminum capacitors.

- Common capacitor types used in low voltage DC/DC converters are
  - aluminum electrolytic
  - tantalum electrolytic
  - ceramic

- Electrolytic capacitor
  - is a polarized capacitor;
  - the anode is made of a metal that forms an insulating oxide layer through anodization;
  - the oxide layer acts as the dielectric of the capacitor
    - note: $\varepsilon_{r Al_2O_3} \approx 9$; $\varepsilon_{r Ta_2O_5} \approx 25$
  - a solid, liquid, or gel electrolyte covers the surface of the oxide layer, acting as the cathode plate of the capacitor.



Furthermore, electrolytic capacitors are polarized capacitors, and if we reverse it we could blow the capacitor because we are reversing the electrochemical process with which the oxide layer is created, and if we do this hydrogen is produced and the capacitor might explode.

If we look at the structure, we have a top metal plate, the dielectric, then there is soft paper soaked with electrolyte (which is the reason for the high ESR), and then the other plate.
The surface of the Al place where the oxide has to be grown is not flat, it is etched with a chemical process to be very rough. In this way it's possible to increase the net surface without increasing the volume, and this is important to increase the volumetric efficiency. The top plate is the anode of the electrolytic capacitor (+).

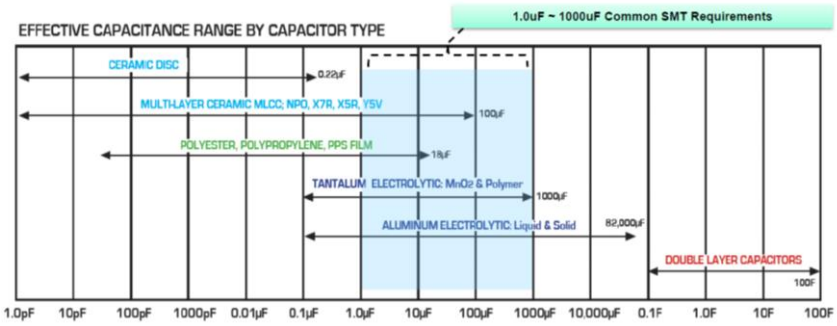
The oxide is grown with an electrochemical process called anodization, and we grow an oxide in the order of um. This is important because the thinner the oxide, the larger the capacitor. The oxide follows the roughness of the metal plate, and its thickness is also related to the maximum voltage the capacitor can withstand (the thicker the oxide, the smaller the specific capacitance), because the oxide has a dielectric strength and if we overcome it the dielectric geos in breakdown and the C goes in shortcircuit.

After the oxide has been deposited, we cannot simply attach the bottom plate (also roughened) and put in contact with the oxide, because peaks and valleys won't match the one of the oxide, and there could be voids filled by air, which is a bad conductor, so we would end up with a capacitor with a large parasitic resistance.

Hence we insert soft paper, which is a deformable material, soaked with a conductive medium, which is typically some salt (liquid electrolyte). The electrolytic layer can be done also with gel electrolyte or solid electrolytes.

The advantage of the electrolytic capacitor is the large volumetric efficiency, we can get a large specific capacitance in a small volume. The disadvantage is that, despite the use of an electrolytic layer, the series resistance is never 0.

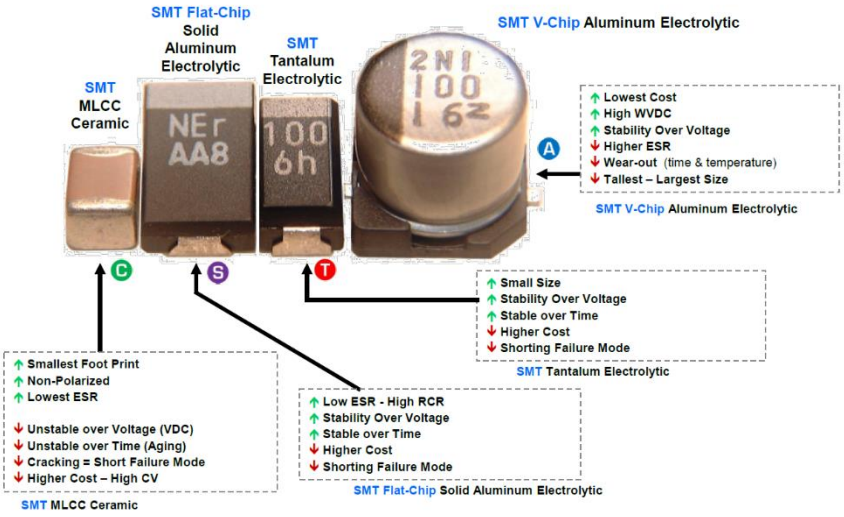
CAPACITANCE OFFERED BY TECHNOLOGY



- **Ceramic** (MLCCs) dominate lower range of capacitance values (0.5pF ~ 10µF)
- High Capacitance **X5R MLCCs** (Cap 10µF ~ 100µF)
- **Tantalum** capacitors cover 0.1µF up to 1000µF in low VDC ratings
- **Aluminum E-Caps** cover wide capacitance value range and supports higher VDC ratings.
- **Double layer** capacitors (Farad values) in use for back-up and short term power applications

In PE we are interested in the light blue window. Polyester capacitors are rarely used.
There are general guidelines to select the capacitor, for instance tantalum capacitor cannot be used at high voltages because its voltage rating is typically few tens of volts. Multilayer capacitors provide a limited value of capacitance, up to hundreds of uF, and aluminum capacitors have some other problems.

SMT technologies

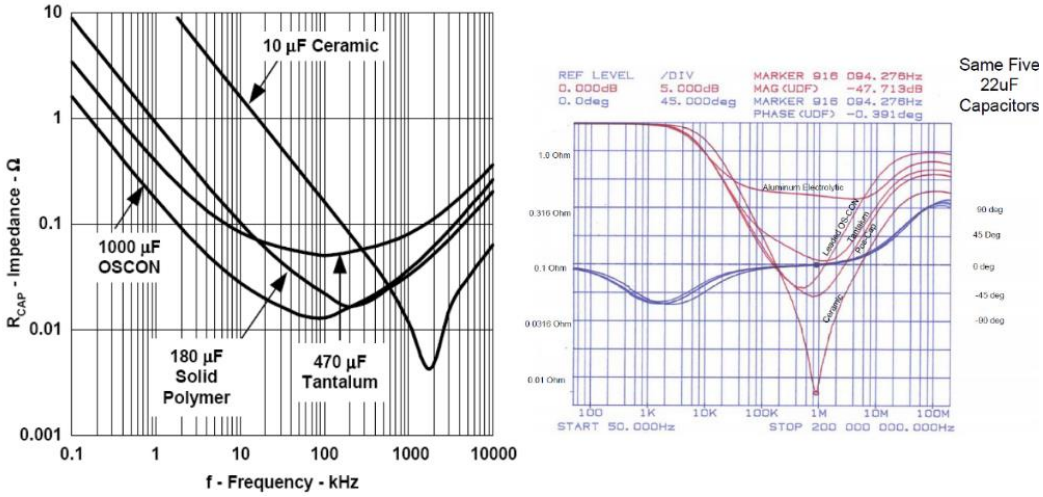


The capacitor providing the smallest footprint is MLCC, which are non-polarized and with small ESR, but they are unstable with voltage, time and with a high cost.
Next to it we have a solid electrolytic capacitor, also called OsCon. It offers a very low ESR (around 10 mOhm), stable with respect to voltage and time.
Tantalum capacitor are based on solid electrolyte, typically manganese oxide, but nowadays they are replaced by tantalum capacitor using a conductive polymer as dielectric (they are called POSCAP).

The advantage of the POSCAP is a really low ESR stable with time and voltage, but the problem is that if we reverse the bias it catches fire, and if the capacitor fails, the typical failure mode is a shortcircuit.

The last capacitor is the aluminum electrolytic capacitor with liquid electrolyte. It is the cheapest but also largest one. They can withstand hundreds of volts, stable over the applied voltage but they have a high ESR (highest compared to the other capacitor) and if we don't properly use them, so we are not respecting the specifications of the maximum rms current in the capacitor, we might end up with an excessive power dissipation in the capacitor which is translated into heat and the heating makes the liquid dielectric to evaporate. If so, ESR increases, power dissipation increases and we get into a positive loop and the capacitor explodes.

OUTPUT CAPACITOR IMPEDANCE



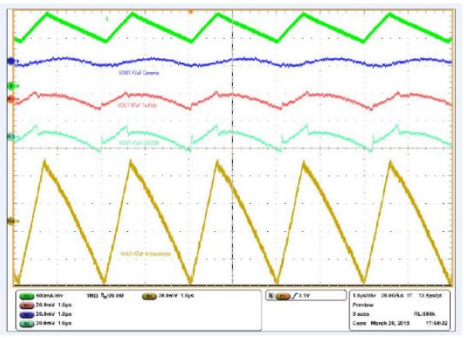
On the right plot we have the impedance of several capacitors having the same capacitance value but different chemistry or material. The aluminum electrolytic provides an impedance that rolls off with a slope of -20 dB/dec as expected in a capacitor, but at a certain frequency the impedance becomes flat until when it increases again. The reason for this is that the real capacitor is a series of 3 components, an ESL, the ESR and the capacitor itself.
There is a frequency, the self-resonance frequency, where the impedance of C and ESL cancel each other and we are left with the ESR. If we look at the plot in correspondence of where the minimum of the impedance occurs, we can read the value of the ESR, because it happens at the self-resonant frequency.

Al-electro capacitor has the largest ESR. The best is the ceramic capacitor.
The HF impedance of the Al-electro capacitor is really bad, and this is why in decoupling we use a large electrolytic capacitor and a small ceramic capacitor.

OUTPUT VOLTAGE RIPPLE BY CHEMISTRY

In principle we can understand which is the chemistry of the capacitor just by looking at the output ripple across it when we are injecting a triangular current, which is the case for instance of the buck converter.

In a ceramic capacitor the voltage ripple seems sinusoidal but it is not, it is a piecewise curve that composes pieces of a parabola. We can notice that in the case of electrolytic, having a large ESR translates in a large triangular ripple.

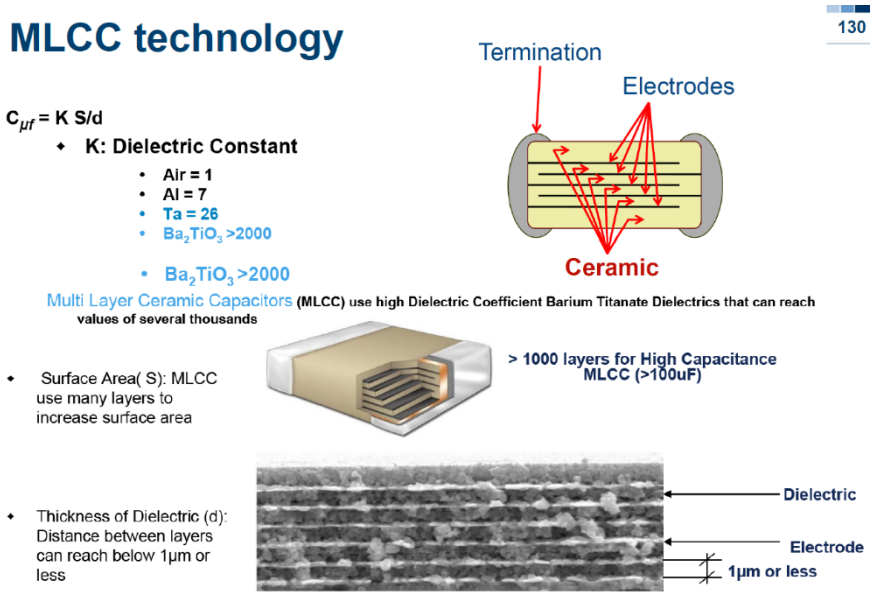


Comparison of the output voltage ripple of a buck converter using four different capacitor chemistries.

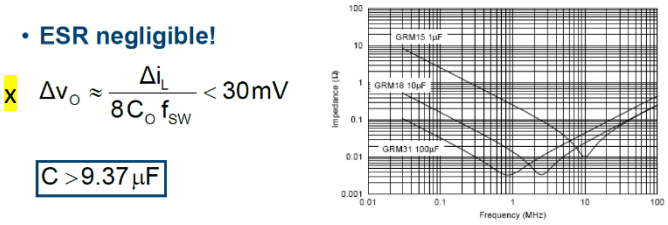
- All caps = 47µF
- Scale = 20mV/div

MLCC TECHNOLOGY

We are considering a ceramic capacitor. MLCC means multi-layer ceramic capacitor, and the structure is interdigitated, we have many metal plates that are interdigitated and the dielectric is between them.



The dielectric is usually barium titanate, which has a relative permittivity of 2000, and this is the reason why we can get up to um in a small volume. The main advantage is the negligible ESR.



Manufacturer	Series	Cap value (µF)	Rated voltage (V)	ESR (mΩ)
MURATA	GRM32	22 to 100	6.3 to 25	< 5
	GRM31	10 to 47	6.3 to 25	< 5
PANASONIC	ECJ	10 to 22	6.3	< 5
	EEFCD	10 to 68	6.3	15 to 55
SANYO	TPA/B/C	100 to 470	4 to 16	40 to 80
TDK	C3225	22 to 100	6.3	< 5

X is the formula we use in a buck converter to select the capacitance. Delta(i_L) is 30% the nominal current according to our design, and fsw = 400 kHz.

In selecting C, we might choose GRM31 with a C = 22uF; we are taking this and not 10 uF because we have to take into account the voltage derating due to the ferroelectric material.

ELECTROLYTIC CAPACITORS

The dominant contribution to the ripple will be the ESR contribution, so the choice procedure is different, I have to select a capacitance basing on the ESR value.
The starting point is that, assuming ESR is dominant, is:

$$\Delta V_o = ESR \cdot \Delta i_L$$

And hence we get to the inequality in the next image.

- ESR is the dominant contribution to the voltage ripple
- Design procedure:
  - select the capacitor such that

$$ESR < \frac{\Delta v_O}{\Delta i_L}$$

  - once C_o is known, check that

$$\left(\frac{\Delta i_L}{8\, C_O\, f_{SW}} + \Delta i_L \cdot ESR \right) \le \Delta v_O$$

  - check that I_{rms} is lower than the rated value

    Note: The RMS current rating of a capacitor is determined by the amount of current required to raise the internal temperature approximately 10°C above an ambient temperature of 105°C.

When selecting the ESR we have to take some margin so that it is not equal but smaller. This because once we have selected the capacitor according to the ESR, we have to compute the capacitive contribution, adding it to the ESR contribution and check that the sum of the two is actually smaller than delta(v_o), x.

There is also another check; the average current in the capacitor is zero by definition because at steady state there is no current, but there is an AC current that, combined with the ESR, produces a power dissipation inside the capacitor. There is an Irms in the capacitor that should not exceed a certain limit, otherwise the capacitor might explode.
The rated Irms in a capacitor is the amount of current required to raise the internal temperature of the capacitor of 10°C.

Example of choice – OS-CON

Size code	Part number	Rated voltage (V)	Rated capacitance (μF)	ESR(mΩ) (max) 100kHz to 300kHz/20℃	Rated ripple current 100kHz (mArms) at 105℃
	25SVP22M※1	25	22	50	2000
	20SVP56M	20	56	40	2400
	20SVP68M	20	68	40	2400
	16SVP100M	16	100	35	2670
	16SVP150M	16	150	30	3020
	16SVP180MX	16	180	30	3020
	10SVP150M	10	150	30	3020
F8 →	10SVP270M	10	270	25	3700
	10SVP330MX	10	330	25	3700
	6SVP220M	6.3	220	25	3700
	6SVP330M	6.3	330	25	3700
	6SVP470MX	6.3	470	25	3700
	4SVP680M	4.0	680	25	3700

- Select:

$$ESR = 25m\Omega < \frac{\Delta v_O}{\Delta i_L} = 33m\Omega$$

- Check:

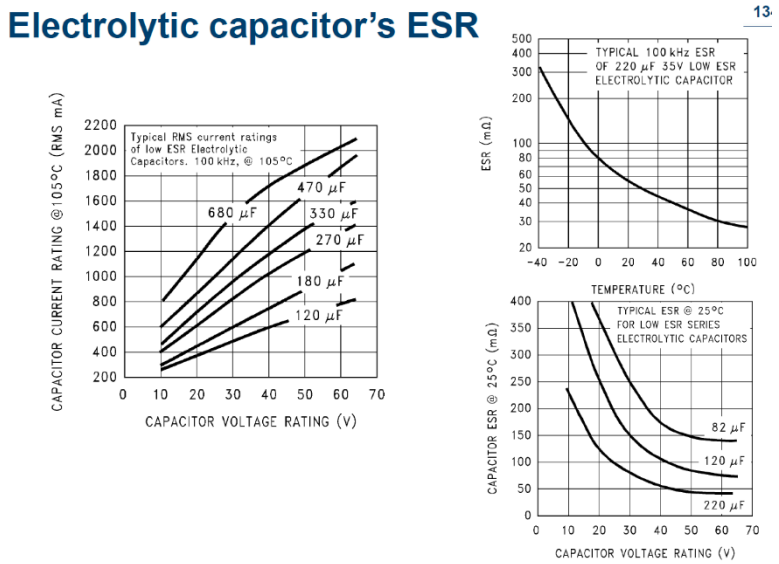
$$\frac{\Delta i_L}{8\, C_O\, f_{SW}} + \Delta i_L \cdot ESR = 23.5mv < \Delta v_O = 30mV \quad ☺$$

$$I_{rms} = \frac{\Delta i_L}{\sqrt{12}} = 0.26A < 3.7A_{rms} \quad ☺$$

Electrolytic ESR

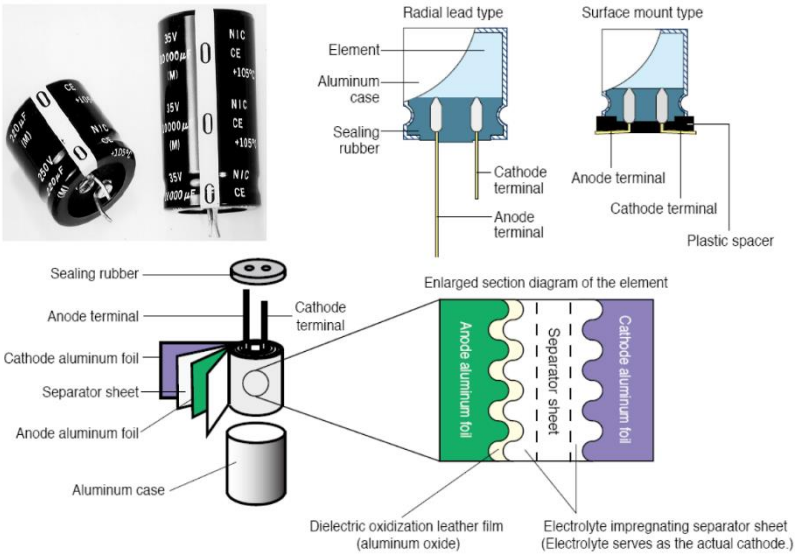
There are two interesting points for an electrolytic capacitor:

1. **ESR depends on the temperature**, the larger T the lower the ESR, because the conductivity of the electrolyte is enhanced by T, so we might have troubles in using the electrolytic capacitors at very low temperatures and also if we consider the compensation of a DC/DC converter, because the ESR introduces a zero; if T increases, ESR decreases and the frequency of the zero ESR*C0 is moving towards higher frequencies. And if we rely on the position of the ESR zero to have a good phase margin we might have problems.
2. **ESR depends on the capacitor rating**, i.e. maximum voltage we can apply to it. As we increase the voltage rating, ESR reduces because if we increase the voltage rating, it means that we have to increase the thickness of the oxide, so if we want to keep the capacitance constant we also have to increase the area, but if we increase the area we decrease the ESR. Similarly, if we fix the rating voltage but we decrease the capacitance value, the area is decreasing and so the ESR is increasing.

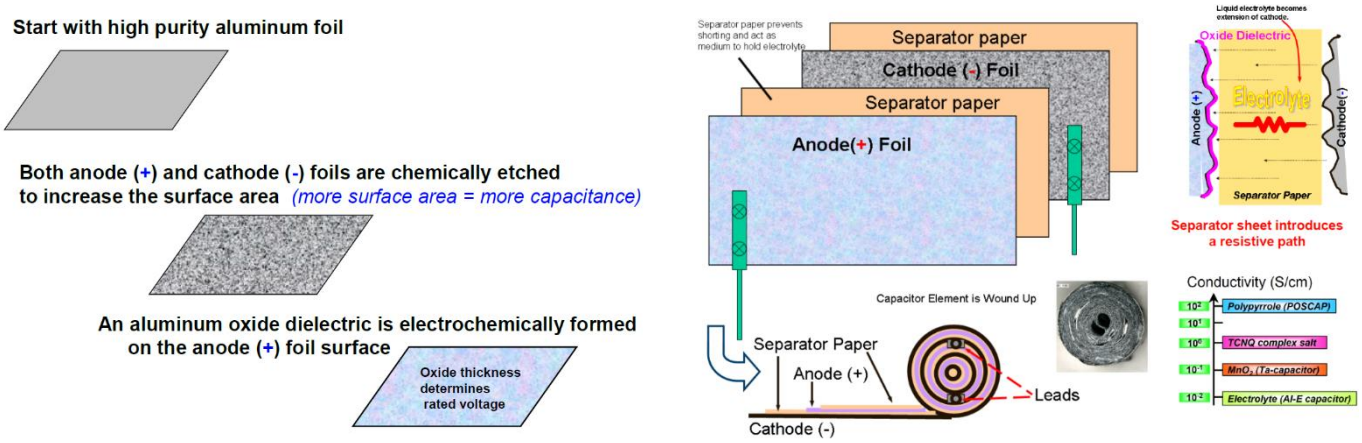


Where does the ESR come from?

It comes from the electrolyte layer.



CAPACITOR CONSTRUCTION



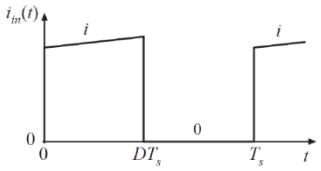
In the end we roll up the film in the capacitor.
In the bottom right angle we have the conductivity of the various electrolyte we can use. TCNQ are OsCon. Since POSCAP uses conductive polymer, the conductivity is higher.

INPUT CAPACITOR SELECTION

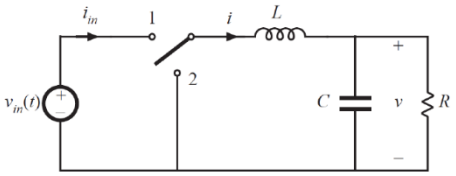
INPUT FILTER

The input filter is typically needed in a buck, buck-boost converter, but in general in all the converters where the shape of the input current is pulsating and not constant.
The square voltage waveform can be split in its harmonic component, where D*I is the DC component of the current multiplied by the duty cycle plus all the other harmonics components. Cos() is the harmonic, the part in front of it is the amplitude of the harmonic.

Input current $i_{in}(t)$ is *pulsating*:



Buck converter example



Approximate Fourier series of $i_{in}(t)$:

$$i_{in}(t) = DI + \sum_{k=1}^{\infty} \frac{2I}{k\pi} \sin(k\pi D) \cos(k\omega t)$$

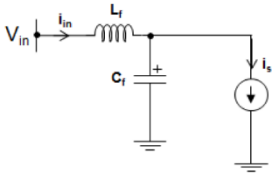
Note: $I=I_0$

High frequency current harmonics of substantial amplitude are injected back into $v_{in}(t)$ source. These harmonics can interfere with operation of nearby equipment. Regulations limit their amplitude, typically to values of 10 μA to 100 μA.

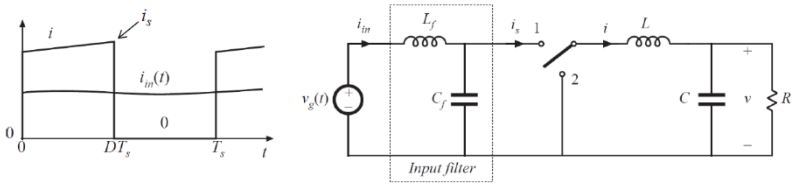
The problem is that the buck converter, during its normal operations, is injecting back in the voltage bus spurious harmonics that might disturb other circuits connected to the same voltage bus Vin.

Low pass filter

To reduce the harmonic components that are injected back we might use a second order filter. The buck converter is replaced with a current generator i_s. If we think at i_s as the superposition of a DC value plus harmonics, we have to design the LC filter in order to divert the harmonics towards ground, while the inductance is sufficiently high to filter out the harmonics and let just the DC component to flow back to Vin. So the impedance of the inductor increases with frequency and the one of the capacitor decreases with frequency, so the HF harmonics are going toward ground in the capacitor Cf, and the DC in the Lf.



$$H(s) = \frac{i_{in}(s)}{i_s(s)} = \frac{1/sC}{1/sC + sL}$$



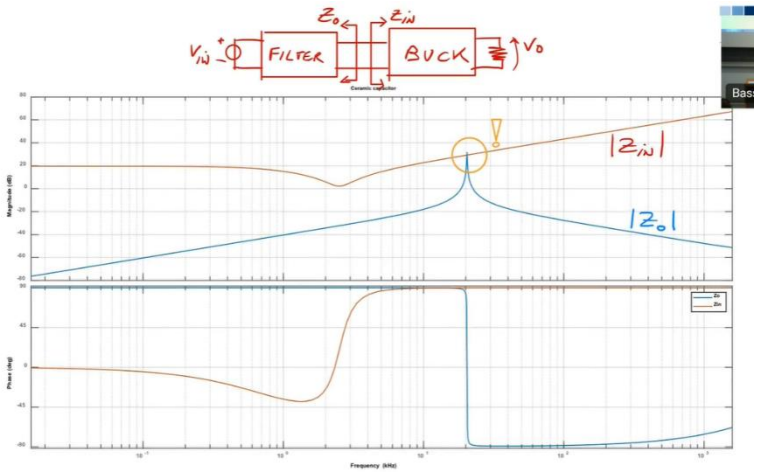
Magnitudes and phases of input current harmonics are modified by input filter transfer function $H(s)$:

x $$i_{in}(t) = \underbrace{H(0)DI}_{I_{IN}} + \sum_{k=1}^{\infty} \left\| H(kj\omega) \right\| \frac{2I}{k\pi} \sin(k\pi D) \cos\left(k\omega t + \angle H(kj\omega)\right)$$

The t.f. H(s) is a simple current divider.
There are two ways to size the input filter: we can either operate in the frequency domain, writing the input current as the DC signal plus the harmonics as in x, or in the time domain.

When designing the input filter, if we look at the input impedance looking at the input of the buck converter we see a negative resistance. In fact, if we assume that the output voltage is regulated, so the output voltage is constant, if we change the input voltage, e.g. increasing it, the power must remain the same, so the input current must decrease, and this makes a negative resistance to show up at the input of the buck converter.

The problem with negative impedances is the input filter, because the input filter is like a parallel of an inductor and a capacitor and if we add a negative resistance to the parallel, we might end up in oscillations.



There are some conditions to check to see if the additional input filter causes oscillations of not. We have to plot the input impedance of the buck converter, the output impedance of the filter and if the magnitude of Zo overcomes the magnitude of Zin we might end up in an unstable behaviour of the system.

If we have Zo overcome Zin we have to damp the peak, and this can be done adding a resistance in series to the capacitor of the filter, or by using a capacitor with a sufficiently high ESR. Hence it is a problem that might occur with MLCC rather than with electrolytic capacitors.

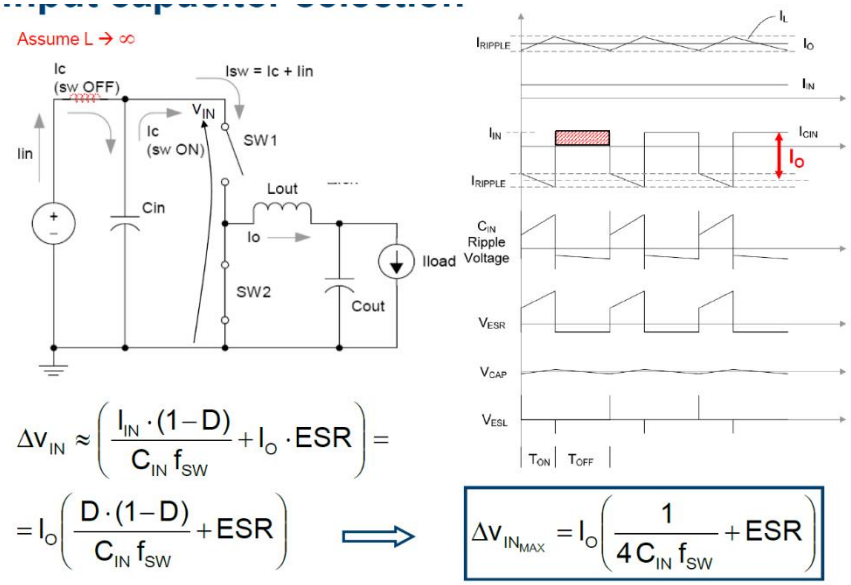
INPUT CAPACITOR SELECTION

We can select the filter with a two steps procedure.
Firstly we assume that the input inductance is very large so that all the AC signal is flowing through the capacitor and only the DC in the inductor. With this specification we can select the value for Cin, and once this is done we can select the value of the inductance by putting a limit to the maximum delta(i_in), so peak to peak current ripple in input. To do all this we need to understand the waveforms in the various components.

In the inductor, since the inductor is huge, we will have only the DC current (first waveform).
Instead, let's assume that the current is positive if it enters in Cin and let's start from a situation where the switch is off, open. All the DC current from the Lin flows in Cin, so the DC current is equal to the average input current: Idc = Iin = Po/Vin, if there are no losses. Moreover, in a buck converter, Iin/Io = D, so Iin = D*Io.

During the on time, the switch is closed. The converter is working at steady state, there is a current Io in Lout and so we have a current Io across SW1. But from Lin still comes a current equal to Iin, the previous one, so the net current in Cin is Iin – Io, which is Io(1 – D).
So the difference between the peak current in the on and off time periods in Cin is Io (second waveform).

Can we calculate the delta_V across the capacitor? It is the sum of two contributions, a capacitive on and the ESR one, according to x.



x $\Delta v_{IN} \approx \left(\frac{I_{IN} \cdot (1-D)}{C_{IN} f_{SW}} + I_O \cdot ESR \right) =$

$= I_O \left(\frac{D \cdot (1-D)}{C_{IN} f_{SW}} + ESR \right) \Rightarrow \boxed{\Delta v_{IN_{MAX}} = I_O \left(\frac{1}{4 C_{IN} f_{SW}} + ESR \right)}$

For the ESR component we have to consider the peak to peak current, that is Io in Cin. We can identify a maximum, a worst case, that is when D*(1-D) gets its maximum, that is at 50%, which is a value that might occur.

The guideline to pick up the capacitor is simple. In general, **we try to keep delta(v_in,max) less than 1% than the nominal input voltage**.
Once we pick up the capacitor Cin matching this condition, we have to compute the Irms of the current with the formula in the image below. It is a formula that is correct for waveforms that are rectangular, i.e. similar to square waveforms with zero average value.

Last step is to check that Irms is lower than the rated value and to check the power dissipation of the capacitor, which is a factor that enters in the calculations of the efficiency.

- A low ESR aluminum or tantalum bypass capacitor is needed between the input pin and ground pin to prevent large voltage transients from appearing at the input. Typically C_{IN} is designed to keep the maximum peak-peak voltage across the input filter in the order of 1% V_{IN_MAX}

$$\Delta v_{IN_{MAX}} = I_O \left(\frac{1}{4 C_{IN} f_{SW}} + ESR \right) \leq 1\% \cdot V_{IN\_MAX}$$

- The capacitor connected to the input has to be capable to support the maximum input operating voltage and the maximum RMS input current required by the device.

$$I_{rms} = I_O \sqrt{D \cdot (1-D)} \rightarrow I_{rms_{MAX}} = I_O / 2$$

- The input capacitor is subject to a pulsed current, the RMS value of which is dissipated over its ESR affecting the overall system efficiency.

$$P_{MAX} = I^2_{rms_{MAX}} \cdot ESR = \frac{I_O^2}{4} \cdot ESR$$

Numerical example

Part Number	Cap. (µF)	W.V. (Vdc)	Max. Tanδ	Max. Impedance 100KHz/20°C	Max. Impedance 100KHz/-10°C	Max. Ripple Current at 100KHz/105°C (mA rms)
NRE-HL330M35V5x11F	33		0.12	0.58	2.3	210
NRE-HL560M35V6.3x11F	56		0.12	0.22	0.87	340
NRE-HL151M35V8x11.5F	150		0.12	0.13	0.52	640
NRE-HL221M35V8x16F	220		0.12	0.087	0.35	840
NRE-HL221M35V10x12.5F	220		0.12	0.080	0.32	865
NRE-HL271M35V8x20F	270		0.12	0.069	0.27	1050
NRE-HL331M35V10x16F	330		0.12	0.060	0.24	1210
NRE-HL471M35V10x20F	470		0.12	0.046	0.18	1400
NRE-HL471M35V12.5x16F	470		0.12	0.049	0.16	1450
NRE-HL561M35V10x23F	560		0.12	0.042	0.17	1650
NRE-HL681M35V10x28F	680		0.12	0.031	0.12	1910
NRE-HL681M35V12.5x20F	680		0.12	0.035	0.12	1900
NRE-HL681M35V16x16F	680		0.12	0.042	0.12	1940
NRE-HL102M35V12.5x25F	1000	35V	0.12	0.027	0.12	2230
NRE-HL102M35V18x16F	1000		0.12	0.043	0.11	2210

$$I_{rms_{MAX}} = \frac{I_O}{2} = 1500\,mA$$

$$\Delta V_{IN_{MAX}} = I_O\left(\frac{1}{4\,C_{IN}\,f_{SW}} + ESR\right) = 96mV \le 180mV$$

$$P_{MAX} = \frac{I_O^2}{4}\cdot ESR = 70mW < 1\%\cdot P_O$$

NIC Components Corp.

We notice that power dissipation is 70mW, which is not negligible, but since it is 1% of the output power it is still acceptable.

Once we have the Cin and corresponding ESR values (680 uF and 31 mOhm in this case), we can select Lin so that the delta(i_in) is smaller than a target.

COMPENSATION NETWORK

We have selected the controller L5989D that implements a feedforward compensation, i.e. the t.f. between the power and the control voltage is 1/alpha, where alpha is the coefficient of proportionality between the height of the sawtooth voltage generator and the input voltage.

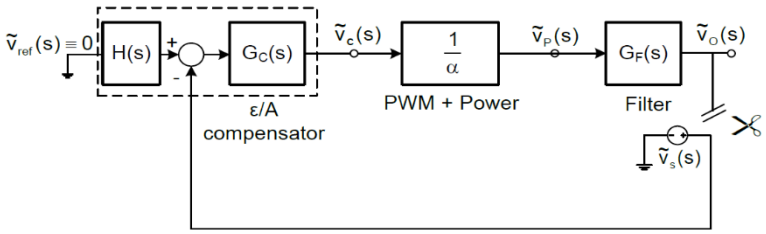
The error amplifier, which is internal, will be assumed as ideal (we will remove this assumption later on).

- The L5989D is a voltage mode controller. The control loop is a single voltage feedback path including error amplifier and error comparator..
- The error amplifier is a voltage operational amplifier with high bandwidth. So selecting the compensation network the E/A will be considered as ideal, that is, its bandwidth is much larger than the system one.
- The voltage feed forward generates a sawtooth amplitude directly proportional to the input voltage, resulting in (see Part 6):

$$\frac{\tilde{v}_P(0)}{\tilde{v}_C(0)} = \frac{1}{\alpha} = 9$$

RECAP – COMPENSATION NETWORK

The following is the block diagram of the regulator.



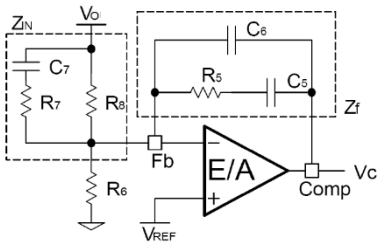
$$L(s) = -\frac{\tilde{v}_O(s)}{\tilde{v}_s(s)} = G_C(s) \cdot \frac{1}{\alpha} \cdot G_F(s) = G_C(s) \cdot G_{CO}(s)$$

We are left with the design of the compensator to get a well-behaving loop t.f., which means a high value of L(s) at LF in order to get a good line and load regulation, plus a rolloff with -20 dB/dec and crossing the zero dB axis at the desired CL BW, and then a second pole possibly around half of the switching frequency.

In L(s), the loop t.f. (not the Gloop), the summing node is not taken into account.
The question now is: am I going for a type II or a type III compensator? In general, we can use a type II compensator if the zero of the ESR falls near to the crossover frequency of the Gloop, so that we can take advantage of the phase lead of the ESR zero to improve the phase margin. This is if the frequency of the ESR zero falls between the frequency of the c.c. poles of the converter and the crossover frequency.

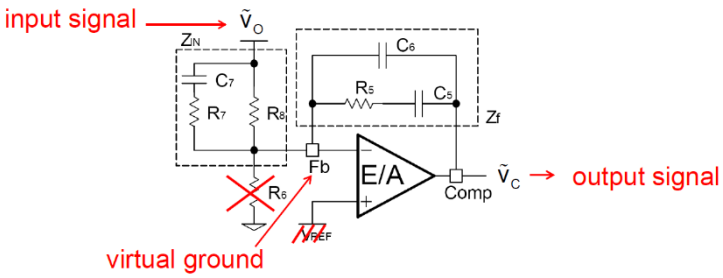
If this doesn't happen, e.g. if we use a MLCC with small ESR and hence a zero at very HF, we are forced to use a type III.
The bottom image is a typical implementation of the compensator, where the error amplifier is internal. The reference voltage is generated internally, in our case 1.8V, and we have to select the external components, which are 7, so in principle we would need 7 constraint for them. In our case we will have just 6 constraints, so we have a degree of freedom, i.e. we guess the value for one of the components and then we go for all the others with the constraints.

- Type II compensation might be suitable with an electrolytic filter capacitor, since $f_{LC} < f_{ESR} < f_C$
- For unconditional stability for any type of output capacitors, and a wide range of ESR values, we should implement local feedback with a type III compensation network.



AC transfer function

We are applying a small perturbation to the output and we compute the control voltage. In AC the reference voltage is ground, so the – of the opamp is virtual ground and R6 is between ground and VG, so it doesn't play any role in AC (it plays a role in DC to set the static value of the output voltage).
So we have the inverting t.f. using the -Zf/Zin and the Laplace domain.



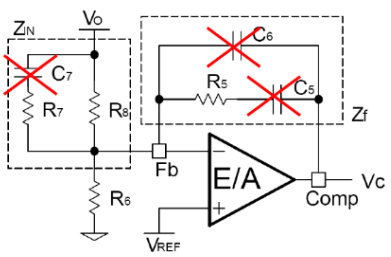
- Plain inverting amplifier
  - assume the OA ideal

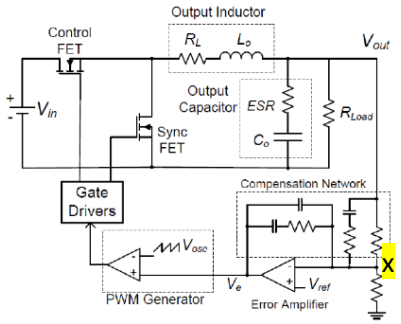
$$\tilde{v}_c = -\frac{Z_f(s)}{Z_{IN}(s)} \cdot \tilde{v}_o \quad \Rightarrow \quad \boxed{G_c(s) = -\frac{\tilde{v}_c(s)}{\tilde{v}_o(s)} = \frac{Z_f(s)}{Z_{in}(s)}}$$

In our interpretation the summing node is external, so the – sign in the t.f. is removed.

DC behaviour

At DC the capacitors are open circuits, so it seems that the error amplifier is working in open loop, and this is true, but there is still an external loop (right image) and the outer loop forces Fb (Vx) to be equal to Vref.





Note

- outer loop forces $F_b = V_{REF}$
- $Z_F \rightarrow \infty$; $Z_{IN} \rightarrow R_8$

⇒ The voltage divider R_6, R_8 and the rerefence voltage V_{REF} set the DC value of the output voltage.

If so, we can write the voltage divider equation on R6 and R8: Fb = Vout * (R6/(R6 + R8)) = Vref = 1.8V.
By regulating the ratio of R6 and R8 we can regulate the voltage at the output.

- Output voltage is programmed by reference voltage and external voltage divider.
- The F_b pin is internally referenced to 1.8V.
- The output voltage is defined by using the following equation:

$$V_O \frac{R_6}{R_6+R_8} = V_{ref} \rightarrow \boxed{V_O = V_{REF} \cdot \left(1+\frac{R_8}{R_6}\right)}$$

- For V_0=3.3 V it follows that:

$$\frac{R_8}{R_6} = \left(\frac{V_O}{V_{REF}} - 1\right) = 0.833 \qquad (1)$$

COMPENSATOR TRANSFER FUNCTION

The t.f. has a pole in the origin and other two poles and two zeros, which are both real negative zeros. The other constraints will be coming from the positioning of the various poles and zeros.

$$G_C(s) = \frac{Z_f(s)}{Z_{in}(s)} =$$

$$= \frac{(1+sR_5C_5)[1+sC_7(R_7+R_8)]}{sR_8(C_5+C_6)\left[1+sR_5\left(\frac{C_5 \cdot C_6}{C_5+C_6}\right)\right](1+sR_7C_7)}$$

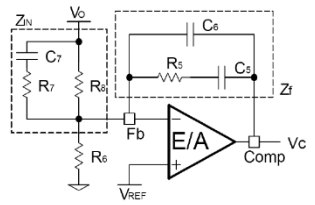
$$f_{p0} = 0$$

$$f_{p1} = \frac{1}{2\pi R_7 C_7}$$

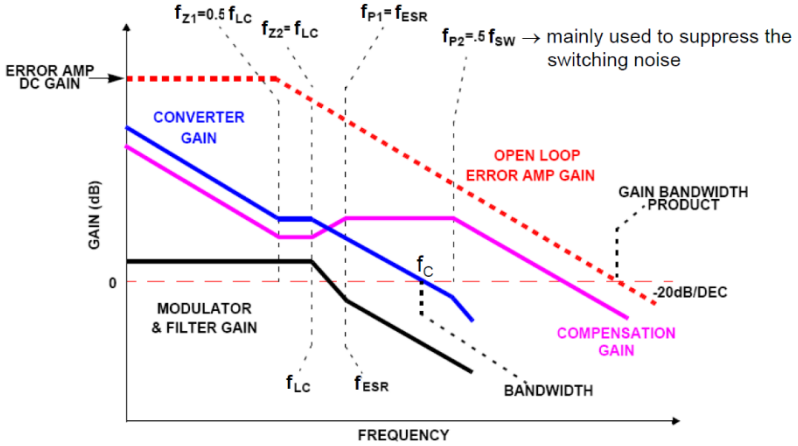
$$f_{p2} = \frac{1}{2\pi R_5 \left(\frac{C_5 \cdot C_6}{C_5+C_6}\right)}$$

$$f_{z1} = \frac{1}{2\pi R_5 C_5}$$

$$f_{z2} = \frac{1}{2\pi C_7 (R_7+R_8)}$$



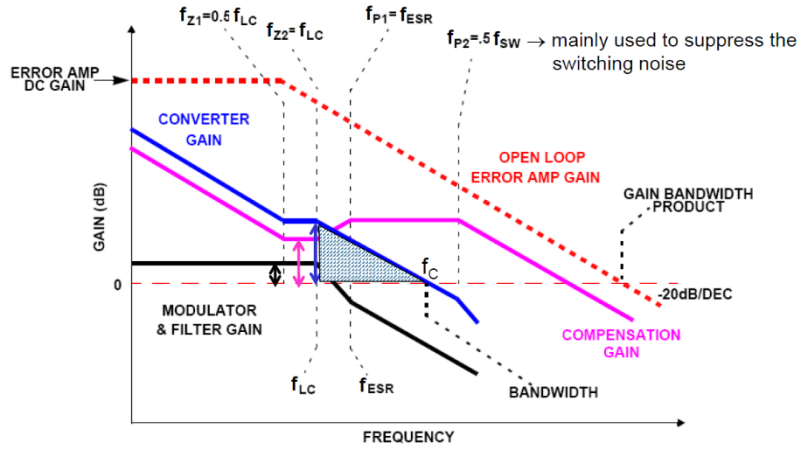
POLES AND ZEROS POSITIONING



The black line is |Goc|, pink line is |Gc| and the resulting blue line is the L(s), magnitude of the loop t.f..
We position the two zeros so that the first zero is typically at half of the frequency of the c.c. poles, and the second zero is located exactly at the same frequency of the c.c. pole pair f_LC.

The first pole fp1 is located in correspondence of the zero of the ESR to cancel it, and the second pole at half of the switching frequency to get a rolloff of -40 dB/dec and cut the switching noise.

However, we have also to position the crossover frequency. In general, it is located between $1/10^{th}$ and $1/5^{th}$ of the fsw. We will position it at $1/5^{th}$, being very aggressive, so we have to move properly the gain of the compensator in order for the loop gain to cut the zero dB axis at fc = 5*fsw.
To do so, we use the gain-bandwidth product. The gain at f_LC is the product of the black line gain (1/alpha) and the pink line gain (R5/R8). In this way we get another constraint (the 6^{th}, the other ones are in the zeros and poles positioning).



$R_5/R_8 \cdot \frac{1}{2} \cdot f_{LC} = f_C \cdot 2$

Guidelines for the design

Guidelines for positioning the poles and zeroes and for calculating the component values:

1. Choose a value for R_8, usually between 2kΩ and 5kΩ.
2. Pick a gain (R_5/R_8) that will shift the open loop gain up to give the desired bandwidth (suggested bandwidth: $f_C \sim 20 \div 30\%\ f_{SW}$). The following equation will calculate an R_5 that will provide the desired f_C:

$$R_5 = \frac{f_C}{f_{LC}} \cdot \frac{R_8}{1/\alpha} \qquad (2)$$

3. Calculate C_5 by placing the zero z_1 at 50% of the output filter double pole frequency:

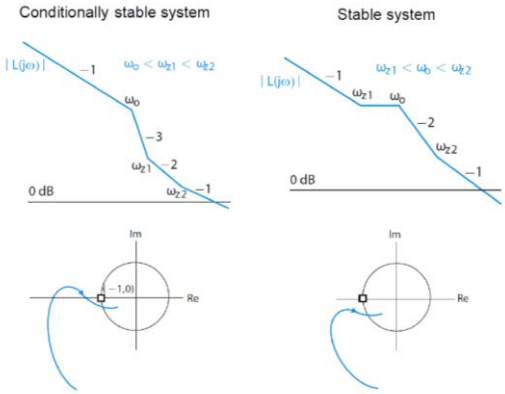
$$C_5 = \frac{1}{\pi R_5 f_{LC}} \qquad (3)$$

Position of compensation zeros and stability

We know where to place the two compensations zeros, but it is a loose guideline, we can place the in other ways if we want to optimize the transient response of the converter.
For instance, one zero might be placed before the frequency of the c.c. pair, and the second zero after the frequency of the c.c. pair.

What we don't have to do, never, is to place both the compensating zeros after the frequency of the c.c. pair (left). If we do this, we have a portion of -60 dB/dec rolloff, and if this happens the system is conditionally stable; if for any reason the gain of the loop t.f. L(s) reduces, we might end up in an unstable system.

This can be seen drawing the Nyquist plot. In fact, we have a region where the phase becomes larger than 180°, and if the magnitude of the gain reduce we might encircle the -1 critical point leading to an unstable system.



- The first compensation zero should be placed prior to the power stage double pole. This step is necessary to prevent the converter from being a conditionally stable system.
- If the compensation zero is not present until the power stage double pole appears, the initial -90° loop gain phase will dip down well below -180° over the power stage double pole.

Instead, the system is always stable if we place one zero before the c.c. pair and one after it.
We are left now with the other steps.

4. Calculate C_6 by placing the pole p_2 at the half the switching frequency:

$$C_6 = \frac{C_5}{\pi R_5 C_5 f_{SW} - 1} \qquad ④$$

5. Set the pole p_1 at at the ESR zero frequency, f_{ESR} and also set the zero z_2 at the output filter double pole. This combination will yield the following component calculations:

$$R_7 = \frac{R_8}{\frac{f_{ESR}}{f_{LC}} - 1} \qquad ⑤$$

$$C_7 = \frac{1}{2\pi \cdot R_7 \cdot f_{ESR}} \qquad ⑥$$

In most of the cases, C5 >> C6, and if this is true (good design), we can simplify the equation neglecting the -1 and cancelling out the C5 dependance.
Then the last step is to set the position of p1 at the zero of the ESR and the second zero at the frequency of the c.c. pair. These two additional conditions allow us to select R7 and C7.

Of course, we can start from the educated guess (educated = not the first value that comes into mind) on another parameter and then derive the others from the constraints.

Numerical example result

Converter parameters

- Input voltage: V_{IN} = 5 ÷ 18 V
- Output voltage: V_O = 3.3 V
- Reference voltage V_{ref} = 1.8 V
- Output current: I_O = 3 A → P_o ~10 W
- Switching frequency: f_S = 400 KHz.
- Total Output Capacitance: C_{OUT} = 270 µF
- Total ESR: ESR = 25 mΩ
- Output Inductance: L = 8 µH
- Desired Bandwidth: f_C = 80kHz

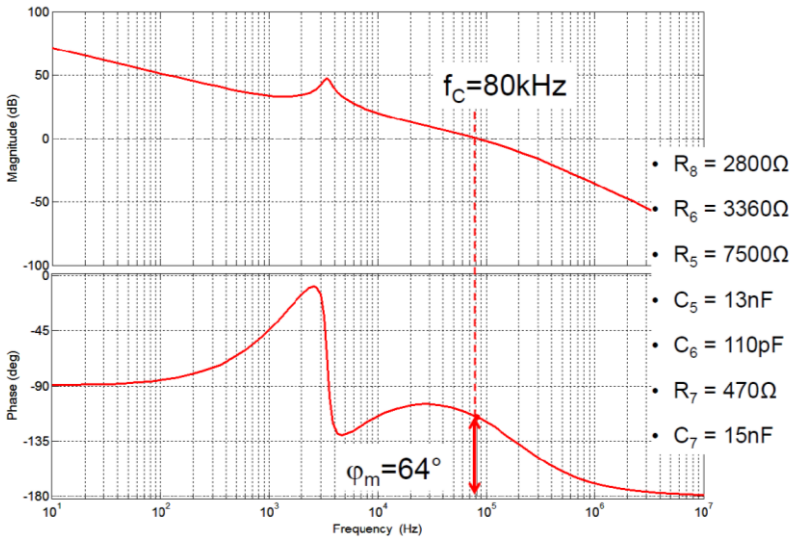
Component values

According to the guidelines for the Type III network

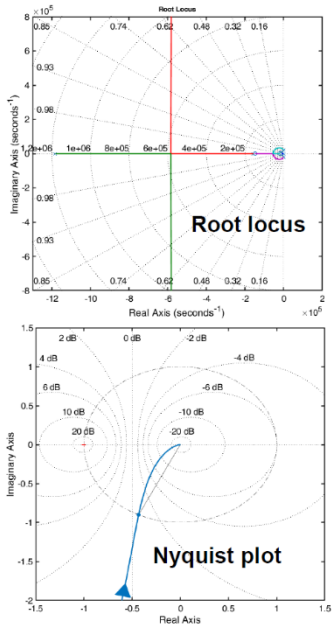
- R_8 = 2800 Ω (±0.5%)
- R_6 = 3360 Ω (±0.5%)
- R_5 = 7264 Ω
- C_5 = 12.8 nF
- C_6 = 110 pF
- R_7 = 475 Ω
- C_7 = 14.2 nF
- These calculated values need to be replaced by standard resistor and capacitor values before the gain and phase plots can be plotted and examined.

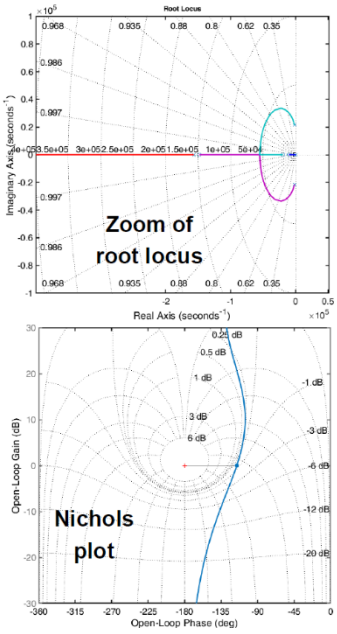
R8 and R6 have to be selected with special care because they determine the voltage divider to set the output voltage. So if we want a given precision of e.g. 1%, these two resistances must have a precision of 0.5%. Instead, the other components are not critical.
Of course, in the real world we have to pick the existing resistance value that is the closest to the one we need.

Gloop Bode plot

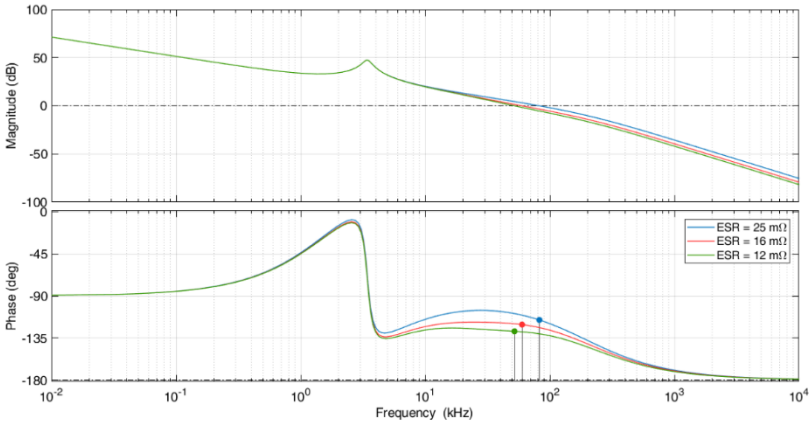


As expected, the crossover frequency is 80 kHz and the phase margin is 64°, pretty good.
The shape of the loop t.f. is not the ideal one with a constant -20 dB/dec rolloff and then -40 dB/dec, there is a bump that comes from the fact that we have two c.c. poles in the converter that we are trying to compensate with two real zeros. However, provided that the crossover frequency is sufficiently higher with respect to the resonance, the bump is not an issue.





Sensitivity of Gloop to ESR



- Minimum $\varphi_m \approx 52°$ for ESR = 12 mΩ
- Minimum $f_{-3dB} \approx 50$ kHz for ESR = 12 mΩ

The ESR depends on the temperature because it is related to the conductivity of the electrolyte in the capacitor, and it reduces with the increase in temperature. An increase in temperature is expected because there is some power dissipation in the capacitor.

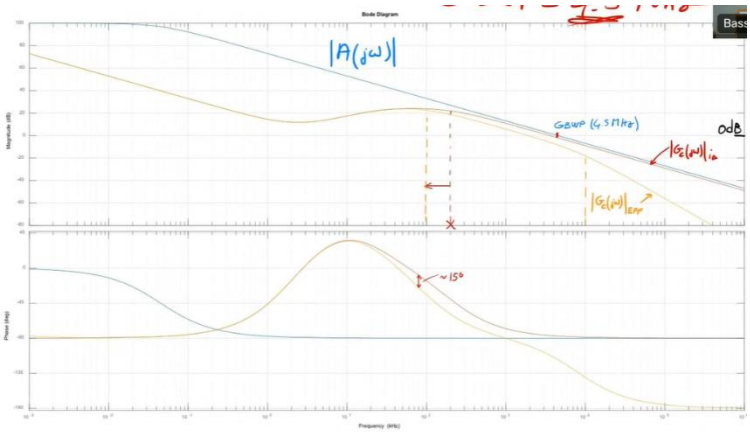
If the ESR decreases, the crossover frequency is reduced, so the closed loop BW is shrinking, and also the phase margin is reducing.

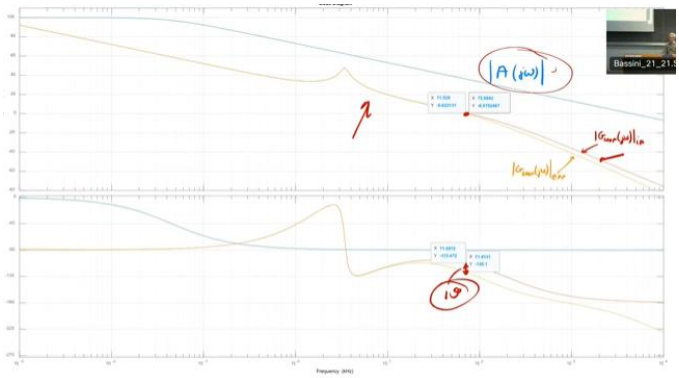
Finite GBWP of the error amplifier impact on the loop transfer function

We calculated the loop t.f. and designed the compensator assuming that the error amplifier was ideal, so the opamp that is included in the chip was ideal. In reality, the opamp is not ideal, and the LF gain A0 is 100dB more or less, and GBWP = 4.5 MHz. Red line in the following plot is the compensator.

The impact of this limited GBWP is negligible if the GBWP of the opamp is larger than the frequency at which the compensator t.f. crosses the 0dB axis.

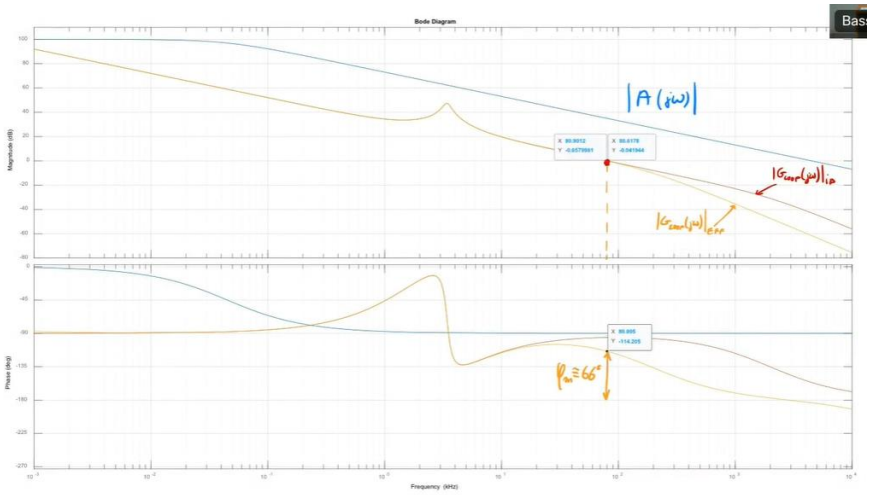
In the case of the image we are ok, because the GBWP is a bit larger. However, the non ideal opamp shifts the pole of the compensator t.f. back in frequency. The yellow line is the result; due to the shift of the pole, the phase margin is slightly reduced. Instead, we have almost no impact on the crossover frequency.





- f_{p2} **pushed back → φ_m degradation**

To restore the original phase margin, we should design the compensator t.f. shifting the second pole at larger frequencies so that the t.f. of the opamp crosses the yellow line, and this is the frequency where we want to place the second pole to restore the phase margin.



- φ_m **restored @ 66°**

Since the second pole depends only on C6, it is possible to reduce the value of C6 to shift the pole to higher frequencies.

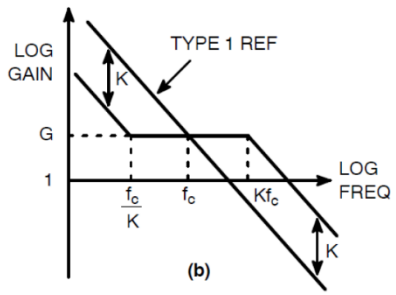
ALTERNATIVE COMPENSATOR WITH THE k-FACTOR METHOD

Alternative approach using a type II compensator. For phase boost < 90°.

- Targets: f_c =80 kHz and φ_m = 60°
- Required phase boost:

$$\varphi_{boost} = \varphi_m - \angle G_{co}(j\omega_c) - 90° = 76°$$

→ go for a type-II compensator →



Calulate k:

$$k = \tan\left(\frac{\varphi_{boost}}{2} + \frac{\pi}{4}\right) = 8.14$$

- Set:

$$\omega_z = \frac{\omega_c}{k} \text{ and } \omega_p = \omega_c \cdot k$$

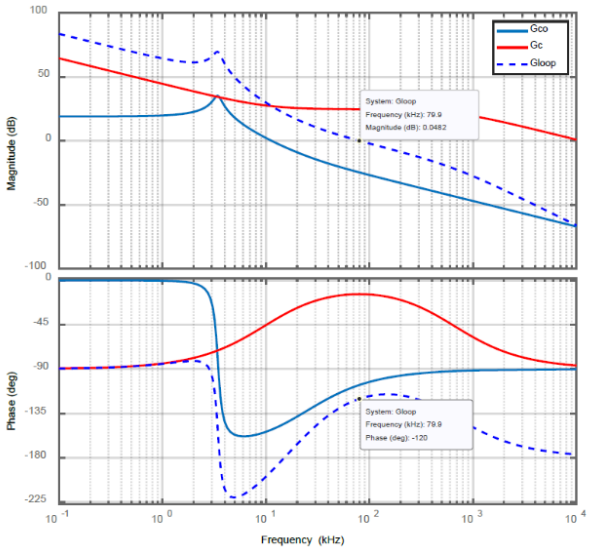
- Calculate K_C:

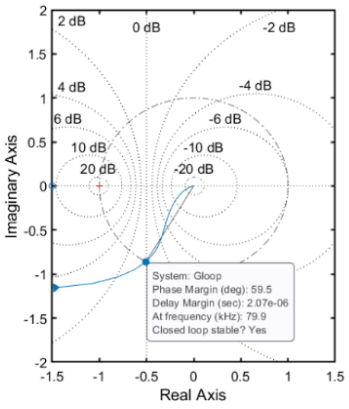
$$K_C = \frac{\omega_c}{k}\frac{1}{|G_{co}(j\omega_c)|} = 1.06\,\text{Mrad/s}$$

$$G_C(s) = \frac{K_C}{s}\cdot\frac{\left(1+\frac{s}{\omega_z}\right)}{\left(1+\frac{s}{\omega_p}\right)}$$

The procedure is already seen in part 6.

The final result is that the Bode diagram of the loop t.f. (dashed blue) has a pretty stable system with a phase margin of 60°. So type III is not necessary, type II is sufficient. However, **if the ESR of the capacitor is pretty small because we are using e.g. a ceramic capacitor, type III is the only way**.



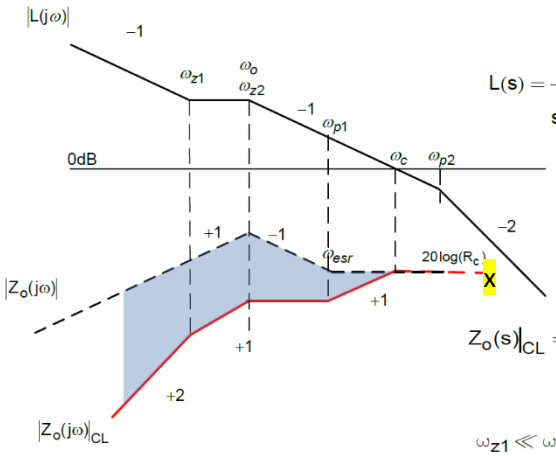


OUTPUT IMPEDANCE AND AUDIOSUSCEPTIBILITY

The output impedance is responsible for the load transient response, and we want a regulator with a good load transient response, fast and with no oscillations.
The load step response can be evaluated starting from the closed loop output impedance.

OUTPUT IMPEDANCE – ASYMPTOTIC ANALYSIS

We start from the knowledge of the loop t.f. and the OL output impedance (dashed line in the plot below). The relationship between CL and OL impedances is the red formula.



$$L(s) = \frac{K_T\left(1+\frac{s}{\omega_{z1}}\right)\left(1+\frac{s}{\omega_{z2}}\right)\left(1+\frac{s}{\omega_{esr}}\right)}{s\left(1+\frac{s}{Q\omega_o}+\frac{s^2}{\omega_o^2}\right)\left(1+\frac{s}{\omega_{p2}}\right)\left(1+\frac{s}{\omega_{p1}}\right)}$$

$$|Z_o(j\omega)|_{CL} = \frac{|Z_{o,OL}(j\omega)|}{|1+L(j\omega)|}$$

$$Z_o(s)|_{CL} = \frac{K_Z s^2\left(1+\frac{s}{\omega_{esr}}\right)}{\left(1+\frac{s}{\omega_{z1}}\right)\left(1+\frac{s}{\omega_{z2}}\right)\left(1+\frac{s}{\omega_c}\right)}$$

$\omega_{z1} \ll \omega_{z2} \ll \omega_{esr} \ll \omega_c$, negligible R_l

$$K_Z = R_c \frac{\omega_{esr}}{\omega_{z1}\omega_{z2}\omega_c}$$

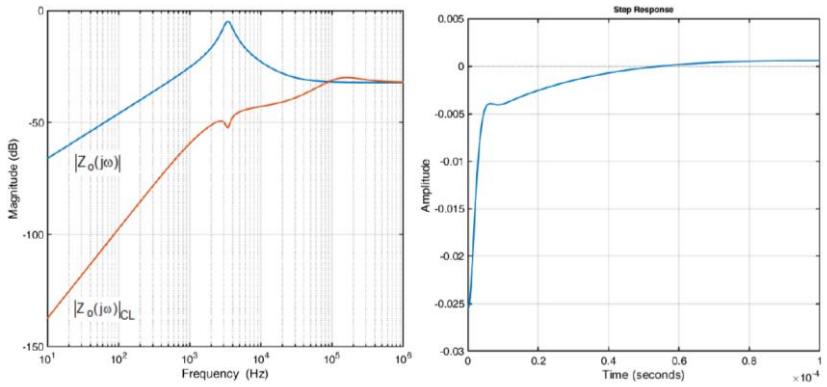
If the magnitude of the loop t.f. (that is the Gloop) is much larger than 1, |L(jw)|>>1, the output CL impedance is the Zo,OL divided by |L(jw)|, which means that in dB we simply have to subtract the magnitude of |L(jw)| to the magnitude of the Zo,OL. Thus we can build the graph of Zo,CL.

The result is the red line in the image above. In the L(s) t.f. we have the two poles, the two zeros, the real negative zero due to the ESR and the compensation poles plus the pole in the origin.

NB: at HF, the L(s) becomes smaller than 1 (in reality as soon as we cross the crossover frequency) and so there is no more feedback, so Zo,OL = Zo,CL. At HF the capacitor is a short and the inductor an open circuit, so the Zo is just the ESR. So x is the 20*log(ESR).

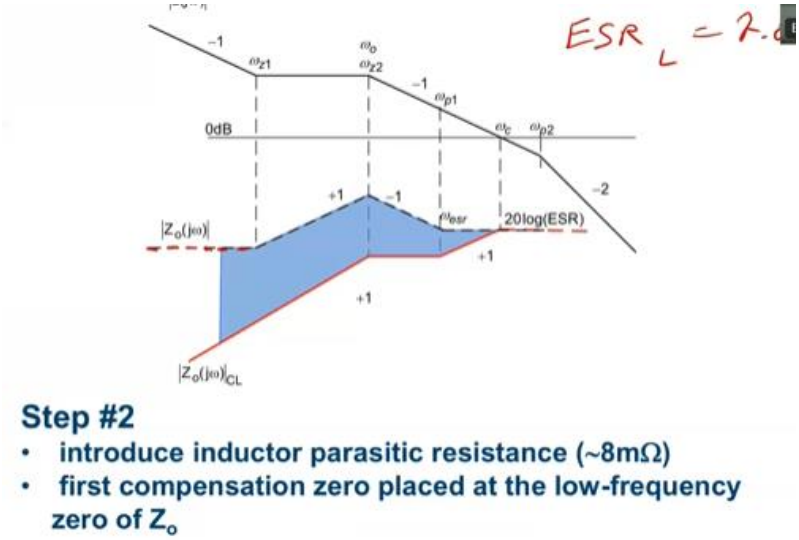
What's the shape of the time domain response, knowing that in Zo,CL we have pole, pole, zero and pole?
If we have a step in input, given this Zo,CL, we have the superposition of two exponential pulses. The fast one has the time constant that depends on the HF pole, the slow on the frequency of the LF pole and the zero sets the ratio of amplitude of the components depending on whether it is closer to the LF or HF pole, proportionally. In fact, if we shift the zero back keeping fixed the poles, we are decreasing the amplitude of the slow component and so approaching an ideal step response.

The real response is given in the following image.



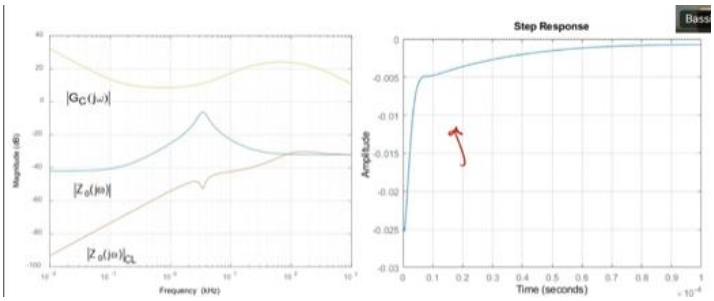
The step response has a fast component and a slow one which has also some resonances and both slow component and resonances are due to the zero pole doublet.
The initial value is negative and equal to 25 mV because if we look at the output network, with the current generator that emulates the load that increases the sunk current, if the sunk current increases by 1A, the step is so fast that the feedback cannot instantaneously react so all the current goes through the path that shows the minimal impedance. In the inductor there is no current because it is an open circuit at high frequencies, so all the current flows through the capacitor and its ESR. So we have a drop across the ESR, whose resistance is 25 mOhm.

How can I eliminate the slow component?
The inductor we are using is not ideal but it has a ESR, so the actual impedance has an additional pole at a time constant L/ESR.
We can take advantage of this pole in the Zo,OL by shifting the first zero of the compensator wz1 back in order to compensate this pole. If we do so, the Zo,CL becomes 'cleaner', the first LF pole is eliminated.



The blue region is the amplitude of the loop t.f.. When it is 0, the feedback is no more existing.
We are compensating a zero of the Zo,OL with a zero in L(s) because in the Zo,CL the Zo,OL is divided by L(s).

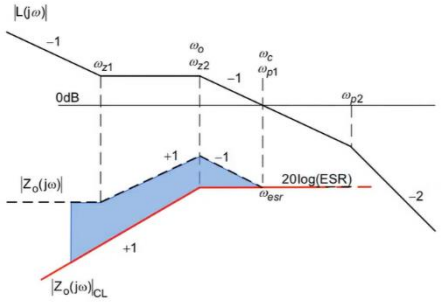
With this solution, the step response is cleaner, with a fast and a slow component still but with no oscillations.



The second step to eliminate the slow component leads to two possibilities:

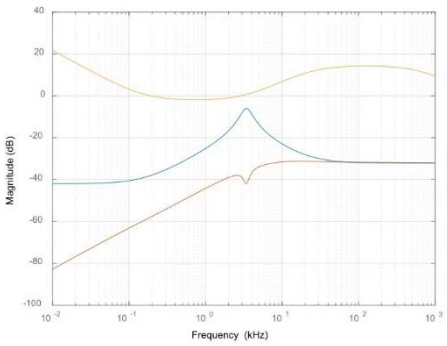
1. We shrink the BW, so we move the crossover frequency of the loop t.f. up until fc = f_ESR. At this point we have the compensation pole of Gc that compensate w_ESR. The Zo,CL has now the shape of a HP filter, whose time response is an exponential pulse.

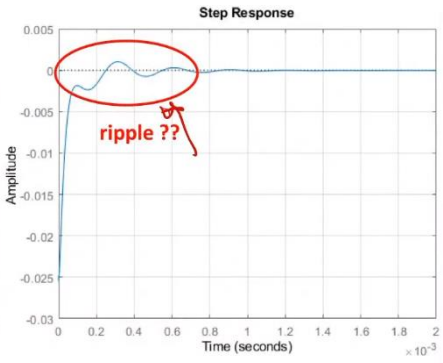
However, if we compute the response we will have some ripples, even if we don't have the small component anymore.



Step #3

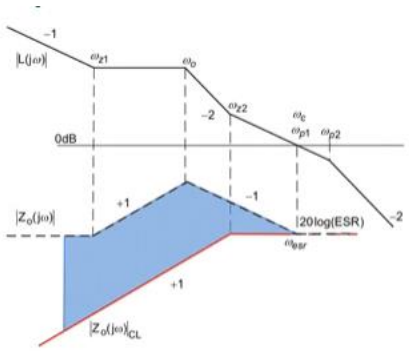
- **shrink BW until $f_c=f_{esr}$**
- **slower load step response but better shape**





The ripples come from the fact that we have reduced the bandwidth, and so the effect of the resonance peak is not reduced a lot because the loop gain has a small value at that frequency.

To further improve the situation, we have to change the filter capacitor using a capacitor with a smaller ESR so to place the ESR zero at the crossover frequency. Of course we also have to move the second compensation zero in the region between the w0 and w_c. Instead, the first zero wz1 has an impact on

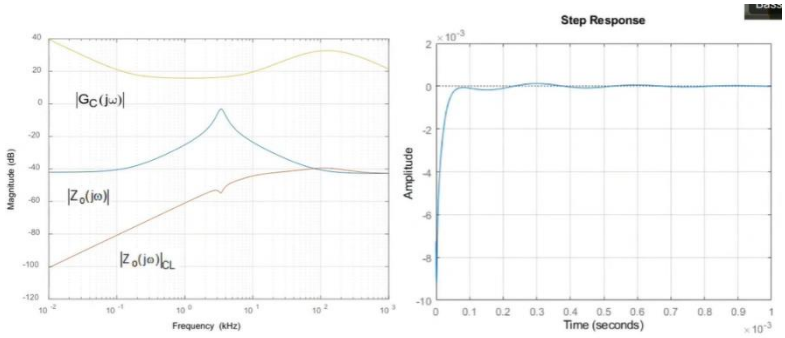


Step #4

- **reduce ESR until $f_{esr\_new} = f_c$**
- **place the second compensation zero between ω_o and ω_c**
- **faster and cleaner load step response**

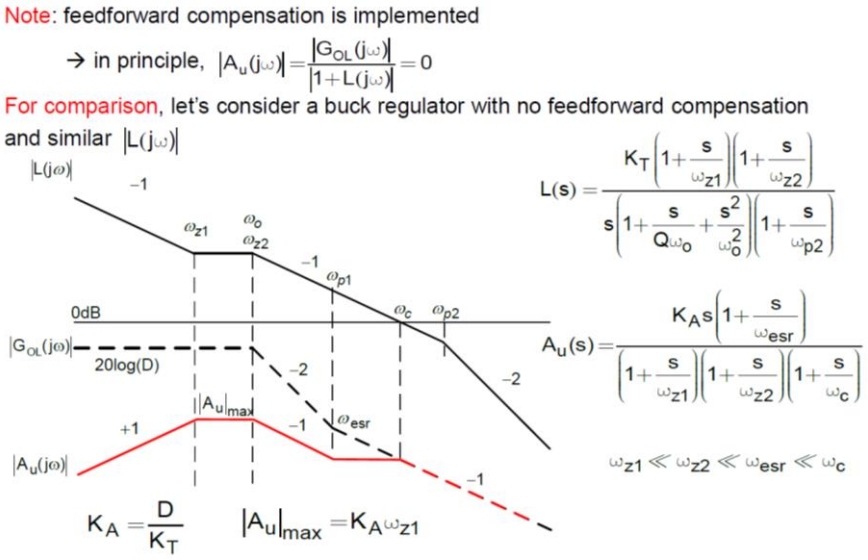
the line step response; if we shift it back, we are moving the LF pole of the audiosusceptibility back and if so the longer the exponential decay in the response (*see later*)

We still get a clean Zo,CL, but the amplitude of the loop at w0 is much larger, so we are attenuating the effect of the resonance and the time response is almost ideal.



AUDIOSUSCEPTIBILITY (CL line to output t.f.)

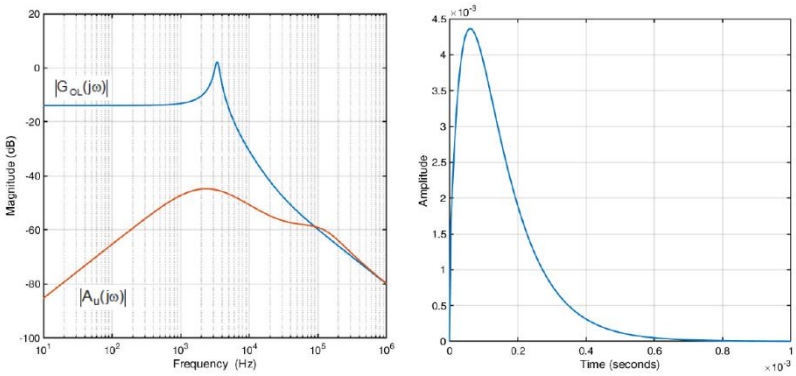
In our case it's 0 in principle, because the controller we are using has a feedforward compensation, which is ideal in the buck converter, so any variation of the input voltage is immediately compensated by the variation of the sawtooth voltage waveform and so basically we have no effect on the output.



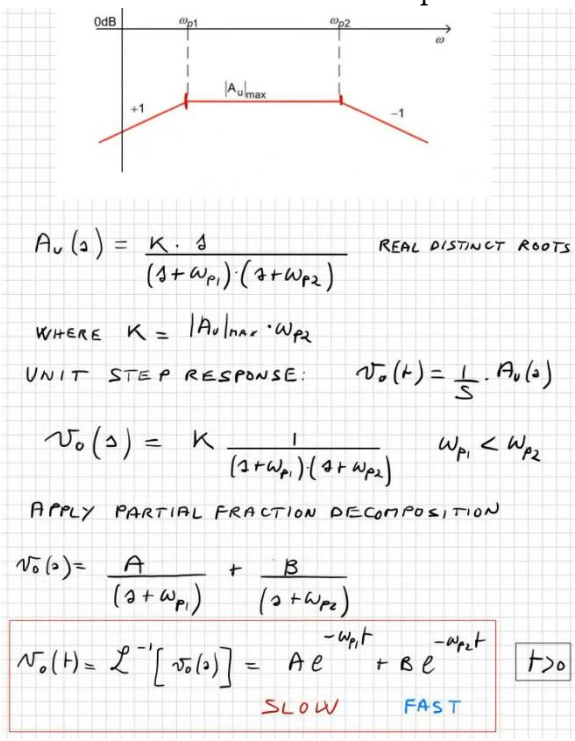
We can still consider a regulated buck converter which doesn't exploit feedforward compensation. In this case the audiosusceptiblity t.f. is not 0 and we can obtain the shape of the audiosusceptibility t.f. Bode use the asymptotic analysis. We start from the OL line to output t.f., and frequency by frequency we subtract the amplitude of the loop t.f. to it. The red line of the previous image is what we are left with.

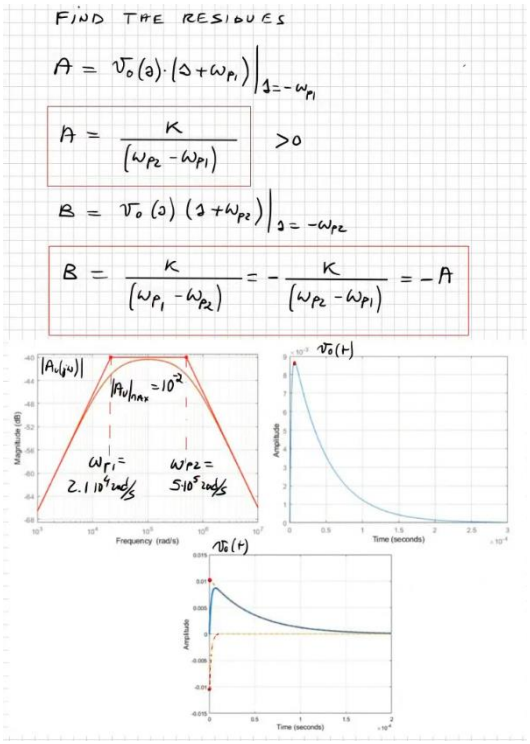
In the end we get a t.f. that resembles a bandpass filter t.f.. The response is the superposition of two exponential pulses, the slower one has a time constant related to the position of the LF pole.

Line step time response



The response shows an exponential increase which is very fast followed by a slow exponential decrease. The result is similar to the case-example.



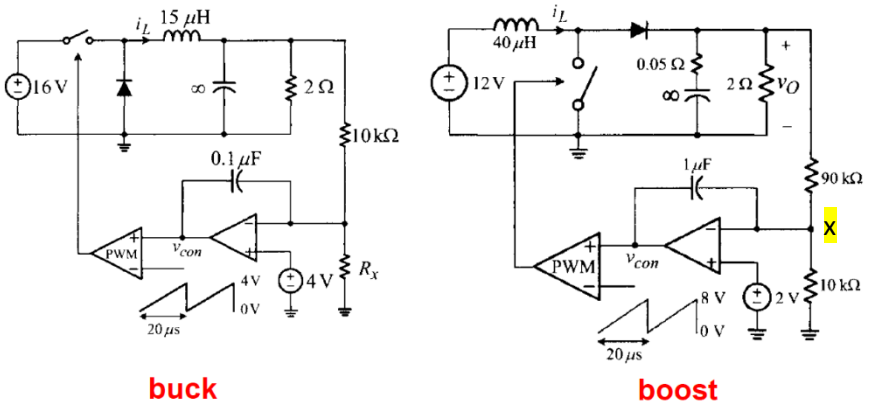


SOFT START CIRCUIT

Soft start is an auxiliary circuit we might need in our converter, that is actually necessary in a boost converter.

In the image we have a regulated boost converter with a simple implementation of the compensator, which is just an integrator, then the error is amplified and fed to the input of the PWM.
Soft start is needed to make the converter to properly start, it is useful in the startup phase. If we assume the converter has been off for a large time, the output capacitor is completely discharged (inf means very large capacitor), so the voltage across it is 0.

When at a given time we start switching the switch, initially the voltage across the output capacitor remains 0V because it is a very big capacitor, so the voltage of the resistive voltage divider x is 0V, this voltage is then compared with 2V making the opamp to saturate to the high level, so the PWM output is always 1 anf the switch is always close. If the switch is always close, the current flows to ground and never passes in the diode, so we end up in a situation where we are increasing indefinitely the current in the inductor until everything blows up.



buck **boost**

- Soft-start is mandatory for a boost converter.
- It is important for buck regulators to have an internal soft-start function that ramps the output voltage in a controlled manner upon startup to limit the inrush current.

We don't have such problem in buck converters but, in order to avoid large inrush currents that we might have at the startup, also in the buck converter it's advised to use a soft start.

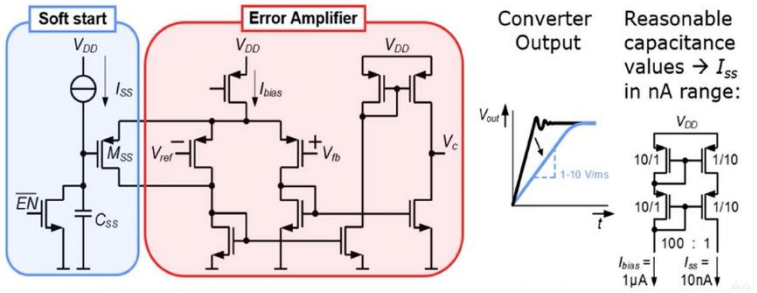
The soft start is something that allows the output of the converter to increase slowly from 0 up to the nominal voltage until we get to the final stable operational point.
In a buck converter the soft start simply makes the output voltage to increase very slowly until we reach the desired value, and this limits the inrush current. The typical slope of the startup is V/ms, so since we are considering V of increase, the duration of the soft start is few ms.

SOFT START IMPLEMENTATION – 1

The error amplifier in this case is an OTA. The soft start is an additional circuit and there is an additional mosfet MSS connected in parallel to the mosfet of the differential pair. Its gate is connected to the capacitor Css that is charged using a constant current generator.

At startup, the enable EN signal opens the transistor and Css is initially discharged, but then as soon as we start the converter and the mosfet is off, the current is able to charge the Css. The voltage across Css is a linear ramp starting from 0 and increasing with a linear slope I/Css.

Since the voltage applied to the Mss transistor is lower than the voltage applied to the internal transistor of the differential pair, the transistor of the differential pair is bypassed, the Mss takes over and hence the initial value of the reference is set by the voltage on Css, that initially is 0.
When the voltage reaches Vref and we overcome it, the Mss gets out of play and the internal transistor takes over.

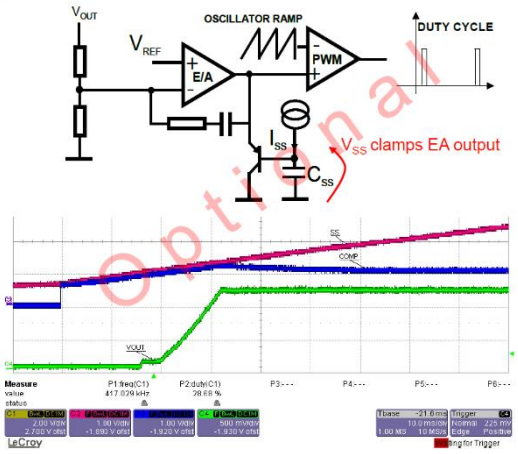


- Soft start input at error amplifier replaces fixed reference by slowly increasing ramp of ~ 1-10 V/ms
- Ramp voltage V_{SS} is generated from a capacitor charged by a costant current.
- V^+ increases linearly until it reaches V_{ref} at $t = t_{SS}$.
- Accordingly, the output voltage ramps up from 0 to the steady-state value in a time t_{SS}.
- Example: slope = 2V/ms, C_{ss} = 5 pF → I_{ss} = 2V/ms·5pF = 10 nA

The effect of the soft start is having the Vref that is not fixed, but slowly increasing from 0V up to Vref. In this case we don't have the saturation of the opamp because initially Vref = 0V. And as far as the Vref is increasing, the output voltage follows.

SOFT START IMPLEMENTATION – 2

- The **soft start** is implemented clamping the error amplifier output to the adjustable voltage ramp generated across the external soft start capacitor



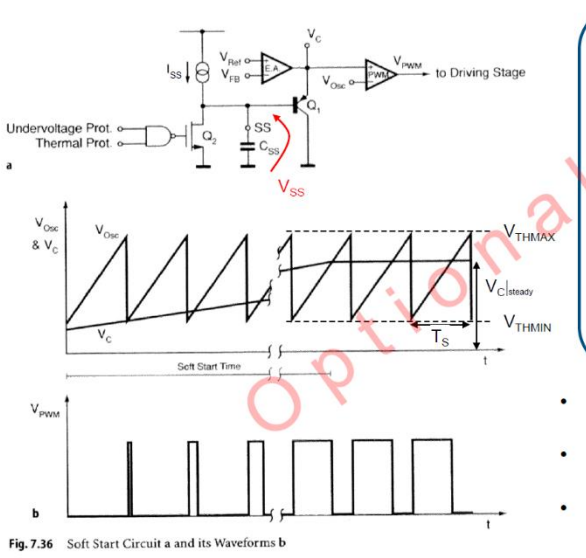


Fig. 7.36 Soft Start Circuit a and its Waveforms b

$$V_C = V_{SS} + 0.7\ V$$

$$= \frac{I_{SS}}{C_{SS}} \cdot t + 0.7\ V$$

$$\frac{V_C|_{steady} - V_{THMIN}}{T_{ON}} = \frac{V_{THMAX} - V_{THMIN}}{T_S}$$

$$V_C|_{steady} = \Delta V_{TH} \cdot D + V_{THMIN}$$

$$D = \frac{V_O}{V_{IN}}$$

$$t_{ss} = \left(V_C|_{steady} - 0.7\,V\right) \cdot \frac{C_{SS}}{I_{SS}}$$

- Reduce inrush current when the system starts up
- Capacitor voltage limits maximum duty cycle
- After capacitor charges, feedback loop takes over control of duty cycle

BUCK REGULATOR – EFFICIENCY ESTIMATION

STEP DOWN REGULATOR LOSSES

- Both power elements (high side and low side) are integrated in the L5989D
- Power dissipation limits the output current capability
- Main contributions to power dissipation:
  - low and high-side conduction losses
  - low and high-side switching losses
  - inductor losses (already evaluated !)
- other contributions
  - quiescent current losses
  - input filter capacitor losses
  - MOS driver losses

The main contribution to power losses are the losses in the low side and high side switches, which are the conduction losses and switching losses. Then we have the inductor losses.
Then there are other contributions that are in general less relevant:

- **Quiescent current losses**: we are using an IC with many internal blocks, so we need to bias all the internal circuits, so we are drawing from the PS a bias current that provides a power dissipation.
- **Input filter capacitor losses (and also output capacitor)**
- **MOS driver losses**: extremely important if we want a high efficiency for the system at low currents.

We can estimate all these contributions.

CONDUCTION LOSSES

In the high side, the Irms is the output current multiplied by the square root of D, because i_HS is a rectangular waveform where the peak is Io and the duration DTs.
For the low side, it is a complementary current, so Io*sqrt(1-D).

$$I_{rmsHS} = I_O \cdot \sqrt{D}$$

$$I_{rmsLS} = I_O \cdot \sqrt{1-D}$$

$$P_{cond} = I_{rmsHS}^2 R_{DSonHS} + I_{rmsLS}^2 R_{DSonLS}$$

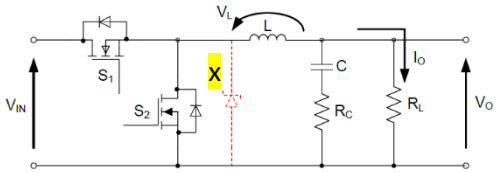
with V_{in} = 12 V → D ~ 0.3; R_{DSonHS}=120mΩ, R_{DSonLS}=100mΩ,
we obtain:

$$P_{cond} = 0.95\,W$$

The conduction losses are quite huge, causing a reduction of the efficiency of almost 1% (Po = 10W).

SWITCHING LOSSES

Let's suppose we are using hard switching, so no soft switching or resonant switching. The switching losses are the product of the maximum voltage between drain and source during turn on and turn off and the maximum current, that in this case is Io, multiplied by the factor in the image.



$$P_{sw} = V_{dsMAX} \cdot I_O \cdot \left(\frac{t_{on} + t_{off}}{2}\right) \cdot f_{SW}$$

$V_{dsMAXHS} = V_{in}$

$V_{dsMAXLS} = V_D$, due to the deatime

With V_{in}=12V; t_{on}~t_{off}~50ns, V_D=0.5V, f_{SW}=400kHz:

$$P_{sw} = V_{in} \cdot I_O \cdot \left(\frac{t_{on} + t_{off}}{2}\right) \cdot f_{SW} + V_D \cdot I_O \cdot \left(\frac{t_{on} + t_{off}}{2}\right) \cdot f_{SW} = 0.75\,W$$

negligible

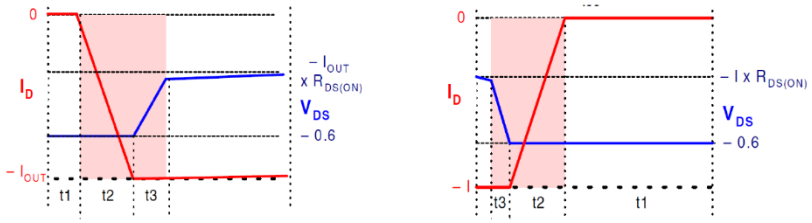
Which is the Vds,max for the HS switch and for the LS switch?
The Vds,max for HS is the Vin, but Vds,max for LS is just the forward voltage drop on diode x. This because of the presence of the dead time, which is much smaller than Vin, and because of this in general the switching losses of the LS switch is negligible with respect to the HS switch. This is due to the fact that the Vds of the LS switch swings between 0 and Vd, where Vd is the forward voltage drop of the Schottky diode, which is much smaller than Vin.

We use the Schottky diode in parallel to the LS switch for two reasons:

1. Smaller voltage drop with respect to the voltage drop across the intrinsic diode of the mosfet, so I decrease the conduction loss when the current flows through the diode.
2. The Schottky diode is a unipolar device, there is no storage of minority carriers and so we don't have reverse recovery happening.

When S1 is on again, we don't go through a reverse recovery phase because of the Schottky.

LS switching waveforms



TURN-ON **TURN-OFF**

The dead time is needed to avoid cross-conduction in the two mosfets. Let's start from a situation in the middle where both devices are off. The buck converter is working in CCM and so the current is flowing through the Schottky. When the LS switch is turned on we are applying a positive Vgs voltage, larger than the threshold and the LS mosfet works immediately in ohmic region as soon we cross the threshold voltage. This because the Vgs = 1V ore or less, Vds = 0.5V and so both Vgs and Vgs are above the threshold voltage. Since we are in ohmic, the mosfet behaves as a resistance.
When time passes, we are charging the input capacitance of the mosfet and so we are increasing the charge applied to the Vgs and so the Rds,on reduces. When it reduces, the inductor current is shared between the Rds,on and the Schottky diode. The smaller the resistance becomes, the larger the current in the mosfet. Once the mosfet current reaches the inductor current, which is equal to the output current, the diode disengaged and the voltage across the mosfet reduces down to the product between Iout and Rds,on, which is smaller than the voltage across the Schottky.

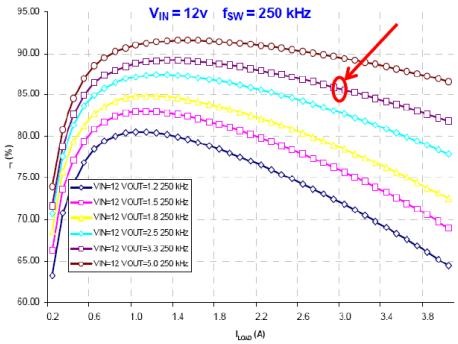
STEP DOWN REGULATOR EFFICIENCY

The efficiency is, by definition, the ratio between the output power and the output power plus all the losses. The main contributions to the losses are conduction, switching and inductor losses. Driving and quiescent losses are in general negligible at large output currents.

$$\eta = \frac{P_O}{P_O + \sum P_{loss}}$$

by considering only the main contributions to losses we get:

$$\eta = \frac{P_O}{P_O + P_{cond} + P_{sw} + P_{ind}} \approx 84\%$$



at V_{in} =12V, wich favourably compares to experimental reults (note: exp. results obtained @ 250kHz switching frequency).

Note: efficiencies exceeding 95% can be achieved with multi-chip packages including discrete MOS switches along with the PWM driver (see, for instance, International Rectifier *SupIRBuck* family)

We have to look at the curve corresponding to 12V and output voltage of 3.3V, which is the purple curve. At 3A the measured efficiency is 86%. It seems not a very good efficiency, but it is however acceptable. There are other approaches to increase the efficiency, but first we have to try to minimize the conduction and switching losses. So HS and LS switches are the main responsible for the losses, and this happens

because they are integrated within the same silicon of the rest of the circuit. If this is the case, in general the Rds,on of the transistor is not optimized and we don't have much area to be spent to reduce it.

To do better, instead of using fully integrated mosfet devices we can think of using discrete mosfet devices that are optimized for a minimal Rds,on, or a multichip module, with multiples chips in the same package.

More importantly, above the efficiency value, the efficiency is decreasing at low load currents, reaches a peak and then decreases again at high currents. This can be justified by rewriting the efficiency as below, considering the losses grouped depending on the dependance they have on the current, if on the square of the current (e.g. conduction losses), or linearly on the current (e.g. switching losses) and the third contribution are losses that are not depending on Io.

$$\eta = \frac{P_O}{P_O + \sum P_{loss}} = \frac{V_o I_o}{V_o I_o + A I_o^2 + B I_o + C}$$

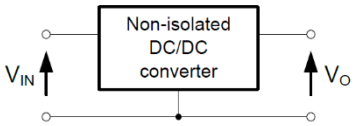
Then I divide everything at the denominator by VoIo.
At large currents, the term 1 + A*Io/Vo is dominant, so we have an efficiency that goes down. At low currents, the term that dominates is C/(VoIo).
C is a constant that defines power losses that are independent with respect to the output current. These losses are the quiescent losses, due to the current drawn from the PS to bias the system, and driving losses, which are more important and depend on linearly the frequency (P = Qg * Vgs,max * fsw).

To reduce these losses, we can do the power switch segmentation. In fact, the power mosfet has a cellular structure, so we can split the power mosfet in several segments and when the load current is large we turn on all the segments in parallel but, as far as the current reduces, and so the conduction losses are not the dominant component, we can turn off the largest segments in the power mosfet and just switch on and off the small segments, which provide a lower gate charge Qg.

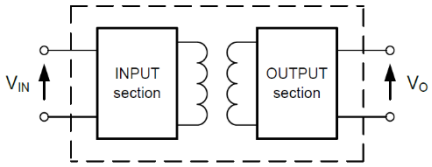
The other possible approach is to reduce the fsw or other approaches like pulse-width modulation.

ISOLATED DC/DC CONVERTERS

ISOLATED VS NON ISOLATED CONVERTERS



Non-isolated dc–dc converters may have *common negative* or *common positive* input and output terminals.



Isolated dc-dc converters use transformers to obtain dc isolation between the input and output.

→ The isolated output is floating, i.e. it doesn't have a defined or fixed voltage relative to input voltage nodes.

The typical reason for using isolation is safety.

ISOLATION

- **What is it?**
  - No direct conduction path between input and output (galvanic isolation).
  - Dangerous electrical shorts cannot pass through to either side in the event of component failure or fault.
- **Why it is needed?**
  - Protect the primary side from faults and components failures on the secondary side.
  - Protect the secondary side from line-level events, e.g. surges, lightening strikes, noise.
  - Protect the human operator and the equipment.
  - Regulatory standards require isolation.
- **How is it implemented?**
  - Power transferred through a transformer.
  - Secondary-to-primary signal sensing through an opto-coupler or a transformer.

ISOLATED DC/DC CONVERTERS

The one in the image below is the typical implementation of an isolated DC/DC converter. The input voltage from power lines is rectified and filtered (PFC circuit used nowadays) and in output we have 310V peak (220 rms * sqrt(2)). This voltage is converted in the desired output voltage using an isolated converter. The transformer that is used is not a low frequency transformer, it is a transformer placed between the switch network and the filter network, so it is operating at a switching frequency of hundreds of kHz, not to say MHz. At a given output power, **the higher the frequency the smaller the size of the transformer**, and this allows us to shrink the size of the transformer.

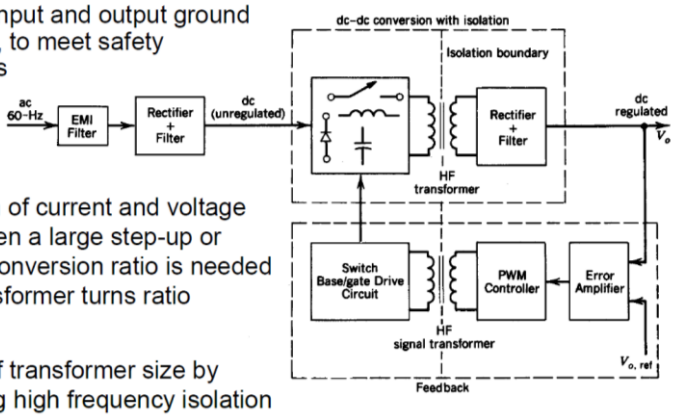
Most of the converter are isolated, so there is a feedback path and also in the feedback path we have a HF transformer. We can use a digital isolator (not optocouplers, which are slow).

Safety is not the only reason to use isolation, there are also other reasons, such has in the conversion from 48V to 1V in data centers. If we would use a buck converter with e.g. 50 V in input down to 1, it means that the buck should work at 2% duty cycle, that is unpractical, especially if the switches are operated at high frequencies (MHz). If e.g. Ts = 1us, 2% is 20 ns in which we have to turn on the mosfet and turn it off again.

With isolated topologies we can take advantage of the turn ratio of the transformer to relax the specification on the duty cycle.

Objectives

- Isolation of input and output ground connections, to meet safety requirements
- Minimization of current and voltage stresses when a large step-up or step-down conversion ratio is needed → use transformer turns ratio
- Reduction of transformer size by incorporating high frequency isolation transformer inside converter



Note: this is only one possible arrangement. Controller often on primary side with switch.

A third reason for isolated converters based on a transformer is that we can put many secondary windings on the magnetic core obtaining hence different output voltages from the same converter.

- Multiple output voltages can be obtained via multiple secondary windings and multiple converter secondary circuits
- Multiple outputs can be "quasi-regulated" using a single controller and feedback loop.

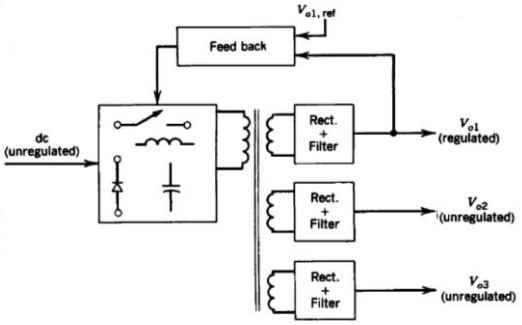
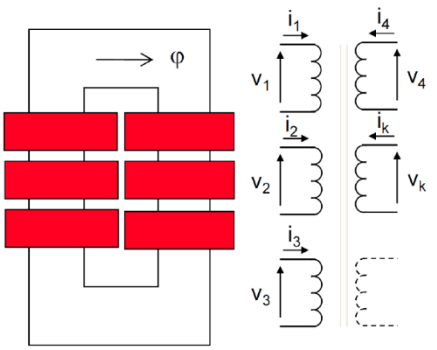


Figure 10-3 Multiple outputs.

TRANSFORMER POWER HANDLING CAPABILITY

Total apparent power, S, of n winding transformer (VA)



$$S = \sum_{k=1}^{n} V_k \cdot I_k \quad (1)$$

- n = total number of windings
- V_k = rated RMS voltage of winding k
- I_k = rated RMS current in winding k

We start by defining the total apparent power in a transformer, which is the product of the apparent power of the windings. Each winding has its own apparent power that is the product of the rms value of the voltage across the winding and the rms current that flows through the winding.

Let's focus on the generic winding k. What is the relationship between the voltage across the terminals of the winding of k and the magnetic flux inside the transformer? The Faraday law.

Starting from the Faraday law, we can compute the rms value of the voltage across the winding k, Vk.

Instantaneous voltage on winding k:

$$v_k = n_k \cdot \frac{d\varphi}{dt}$$

- φ = instantaneous flux in core section
- n_k = turn number of winding k

RMS voltage of winding k:

$$V_k = n_k \cdot K_v \cdot A_e \cdot B \cdot f \quad (2)$$

$\varphi = \varphi_0 \sin(\omega t)$

$\frac{d\varphi}{dt} = \omega \varphi_0 \cos(\omega t)$

$\omega = 2\pi f$

- A_e = transformer core section (m^2)
- f = operation frequency (Hz)
- B = core flux density (T)
- K_v = form factor of voltage waveform v_k (ex: $K_v = \frac{2\pi}{\sqrt{2}}$ for sinewave)

Ae is the area of the cross-section of the magnetic core. If we assume the flux is changing sinusoidally, when we apply the derivative, with respect to time, we have a frequency dependance, that is then inside the terms B*f.

We can combine equations 1 and 2 to get to equation 3.

By plugging equation 2 into eq. 1 we get:

$$S = \sum_{k=1}^{n} (n_k \cdot K_v \cdot A_e \cdot B \cdot f \cdot I_k) = K_v \cdot A_e \cdot B \cdot f \cdot \sum_{k=1}^{n} n_k \cdot I_k \quad (3)$$

We can now write $\sum_{k=1}^{n} n_k \cdot I_k$ as :

$$\sum_{k=1}^{n} n_k \cdot I_k = J \cdot \sum_{k=1}^{n} n_k \cdot A_{wk} = J \cdot K_u \cdot W_a \quad (4)$$

x

by forcing the same RMS current density J (A/m^2) in all windings to impose uniform distribution of copper loss volumic density & avoid hot spot.

- A_{wk} = conductor section of winding k
- K_u = window utilization factor = A_{Cu}/W_a
- W_a = effective section of core window

Awk is the area of the copper wire; if we sum all the Awk we get the total area associated to the copper in the primary winding or secondary winding, depending if k = 1 or 2. So summation x is the total copper area in the transformer. Then we can express it as the total window area multiplied by a utilization factor.

Apparently, I'm using the same current density J for all the windings and this is by design, it is forced to be equal in each winding. If we verify this condition, the total power dissipation in the copper is minimized and the power dissipation is evenly distribute among the windings.

Finally, we can express the area-product of a transformer Ap. 10^4 is to convert from A/m^2 to A/cm^2

By combining equations 3 and 4 setting $B < B_{max}$, we finally get:

$$A_p = A_e \cdot W_a \geq \frac{S}{J \cdot K_u \cdot K_v \cdot B_{max} \cdot f} \cdot 10^4 \ [cm^4]$$

Area-product of transformer

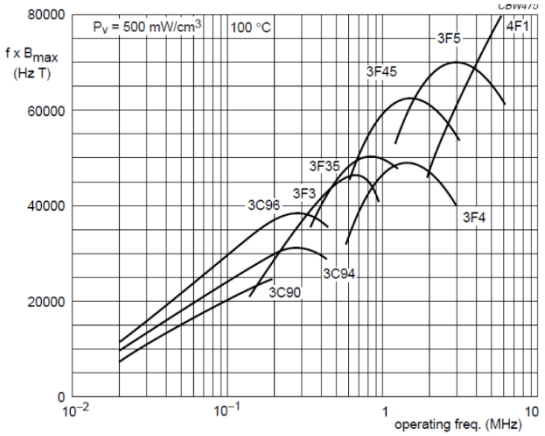
- S = total apparent power (VA)
- A_e = transformer core section (cm^2)
- W_a = effective section of core window (cm^2)
- K_u = window utilization factor
- K_v = form factor of supply voltage waveform v_k
- f = operation frequency (Hz)
- B_{max} = max core flux density (T)
- J = RMS current density (A/cm^2)

- A_p is proportional to the total apparent power S and inversely proportional to the product $B_{max} \cdot f$

When we are designing the transformer we know the total apparent power and we have to pay attention not to saturate the core, so we put Bmax in the formula and f is known. We use the formula to get the minimum Ap.
We can also use this formula solving for S; in this case we want to understand the maximum power the transformer can handle, and it is proportional to the frequency.

NB: the larger the f, the smaller Ap and the larger Bmax the smaller the Ap at a given apparent power, so the smaller the size of the transformer. In particular, Ap is inversely proportional to the product of Bmax and f.

This is important because the catalog typically reports the graphs below.



Performance factor (f × Bmax) at P_V = 500 mW/cm3 as a function of frequency for power ferrite materials.

- The performance factor (f × Bmax) is a measure of the power throughput that a ferrite core can handle at a certain loss level.

This is the performance factor of power ferrites, i.e. it reports the product Bmax*f as a function of the operating frequency, and it is measured at a given volume specific core dissipation Pv.

If we want to design a transformer that for instance works at 400 kHz, in order to minimize the Ap, so the core size, we have to maximizer the performance factor, so we have to select the ferrite that provides a maximum value around 400 kHz, that in this case is 3F3 or 3F35.

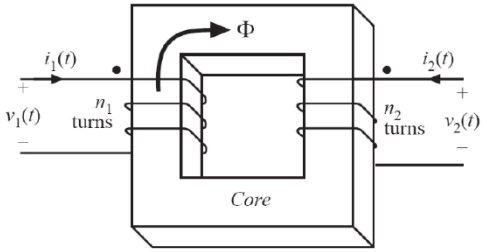
TRANSFORMER MODELING

Two windings, no air gap:

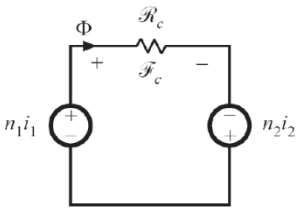
$$\mathcal{R}_c = \frac{\ell_m}{\mu A_c}$$

$$\mathcal{F}_c = n_1 i_1 + n_2 i_2$$

$$\Phi \mathcal{R}_c = n_1 i_1 + n_2 i_2$$



Magnetic circuit model:



Dot convention

The primary and secondary currents flowing into the winding terminals marked ● produce a mutually additive magnetic flux.

We have a prototype transformer with a square-shaped core and two windings with no air gap in the core, because the transformer doesn't have to store power, power goes in and immediately goes out.
The reluctance of the core is the average length of the core divided by the area of the cross-section of the core. The electric equivalent of the magnetic circuit comprises a generator for the total magnetomotive force that is n1*i1 + n2*i2.
The meaning of the dots is to identify where the current enters, in the sense that the current that enters in the dot produces a magnetic flux that is mutually additive in the two windings (check with the right hand rule).
So the magnetic flux phi is the magnetomotive force divided by the reluctance.

If the reluctance of the core is 0, which happens if we push the magnetic permeability of the core to infinite, **we get an ideal transformer, n1*i1 + n2*i2 = 0**. It means that the magnetic flux in the core tends to be 0.

The ideal transformer

In the ideal transformer, the core reluctance $\mathcal{R}_c$ approaches zero.

MMF $\mathcal{F}_c = \Phi \mathcal{R}_c$ also approaches zero. We then obtain

$$0 = n_1 i_1 + n_2 i_2$$

Also, by Faraday's law,

$$v_1 = n_1 \frac{d\Phi}{dt}$$

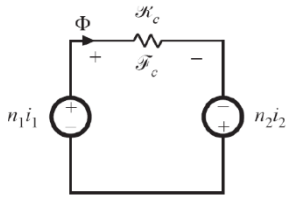
$$v_2 = n_2 \frac{d\Phi}{dt}$$

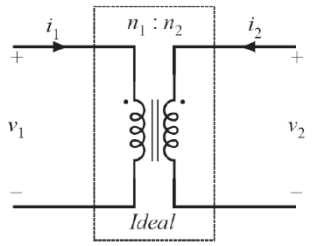
Eliminate Φ :

$$\frac{d\Phi}{dt} = \frac{v_1}{n_1} = \frac{v_2}{n_2}$$

Ideal transformer equations:

$$\frac{v_1}{n_1} = \frac{v_2}{n_2} \quad \text{and} \quad n_1 i_1 + n_2 i_2 = 0$$



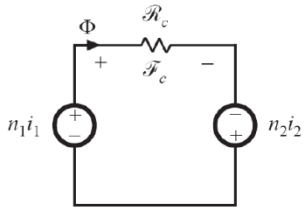


We can also write that i2/i1 = - n1/n2.
We have the – because in the example in both the windings the current is entering, but there is no storage of energy in a transformer, so if the physical current is entering in the primary, the physical current is exiting in the secondary, otherwise we violate the principle of conservation of energy.

The magnetizing inductance

For nonzero core reluctance, we obtain

$$\Phi \mathcal{R}_c = n_1 i_1 + n_2 i_2 \quad \text{with} \quad v_1 = n_1 \frac{d\Phi}{dt}$$



Eliminate Φ:

$$v_1 = \frac{n_1^2}{\mathcal{R}_c} \frac{d}{dt}\left[i_1 + \frac{n_2}{n_1} i_2 \right]$$

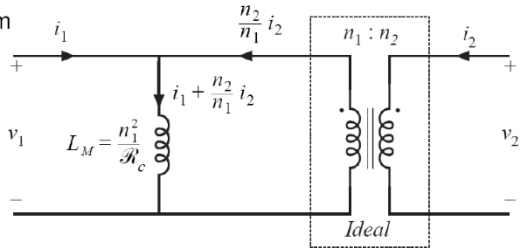
This equation is of the form

$$v_1 = L_M \frac{di_M}{dt}$$

with

$$L_M = \frac{n_1^2}{\mathcal{R}_c}$$

$$i_M = i_1 + \frac{n_2}{n_1} i_2$$



Magnetizing current $i_m(t)$: the current required to create $H_c(t)$ that couples primary and secondary windings through magnetic induction.

This is the most important real component to be added to the ideal model to approximate the real behaviour of the transformer. This comes from the fact that the reluctance of the magnetic core is not zero, so the permeability is not infinite, making the reluctance to be finite.

In the image, eliminate phi means putting the phi expression in the derivative for v_1. The second equation can be rewritten introducing the magnetizing inductance Lm.
Now in the real model there is the magnetizing inductance with a current i_m flowing in it. So the magnetizing inductance is a component to put in parallel with the primary side or secondary side, it doesn't matter. Of course depending on where we put it we have a value, if on the secondary side we need to multiply Lm by the square of the turn ratio.

As for the physical meaning, in an ideal transformer, all the current i1 that enters the primary side is reflected on the secondary side according to the turn ratio. This is not true in a real transformer, part of the input current flows in the magnetizing inductance and so it is not available on the secondary side. From a physical standpoint, some of the current in the primary side is used to magnetize the core establishing a magnetic flux that links the primary and secondary sides. From al electrical standpoint we can model this adding a magnetizing inductance. **It is a real inductance, it is not a fictional component**, and it exhibits saturation and hysteresis (if there is hysteresis we have to take into account the losses in the magnetic core).

- Models magnetization of core material
- A real, physical inductor, that exhibits saturation and hysteresis
- If the secondary winding is disconnected:
  - we are left with the primary winding on the core
  - primary winding then behaves as an inductor
  - the resulting inductor is the magnetizing inductance, referred to the primary winding
- Magnetizing current causes the ratio of winding currents to differ from the turns ratio

Magnetizing current causes the ration of winding currents to differ from the turn ratio. It means that i2/i1 != n1/n2.

Core saturation

- Saturation occurs when core flux density $B(t)$ exceeds saturation flux density B_{sat}.
- When core saturates, the magnetizing current becomes large, the impedance of the magnetizing inductance becomes small, and the windings are effectively shorted out.
- Large winding currents $i_1(t)$ and $i_2(t)$ ***do not*** necessarily lead to saturation. If

  $$0 = n_1 i_1 + n_2 i_2$$

  then the magnetizing current is zero, and there is no net magnetization of the core.
- Saturation is caused by excessive applied volt-seconds

Saturation has to be avoided because if the core saturates, the magnetizing inductance becomes a short and we are shorting the transformer.
Moreover, in a transformer a large primary or secondary current doesn't necessary produce a saturation of the core like happens in an inductor, because the magnetic flux generated by the primary current cancels with the magnetic flux generated by the secondary current.

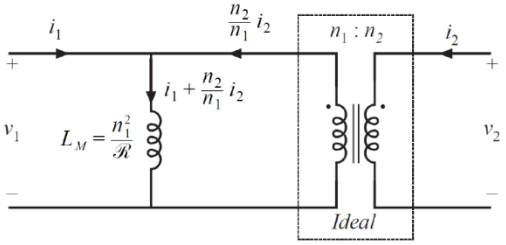
The transformer may become saturated due to the excessive volt-seconds applied to the core. The responsible is once again the current in the magnetizing inductance; we can compute the magnetizing current as x. The integral of v_1 is the flux density.

Magnetizing current depends on the integral of the applied winding voltage:

x $$i_M(t) = \frac{1}{L_M}\int v_1(t)\,dt$$

Flux density is proportional:

$$B(t) = \frac{1}{n_1 A_c}\int v_1(t)\,dt$$



Flux density becomes large, and core saturates, when the applied volt-seconds λ_1 are too large, where

$$\lambda_1 = \int_{t_1}^{t_2} v_1(t)\,dt$$

limits of integration chosen to coincide with positive portion of applied voltage waveform

- Saturation mechanisms in transformers differ from those of inductors, because transformer saturation is determined by the applied winding voltage waveforms, rather than the applied winding currents.

Since L*i = n*phi = n*Ae*B, we can arrive to the second equation solving for B. Hence if the integral of v_1 is too large we may end up in saturation, so B(t) < Bsat.

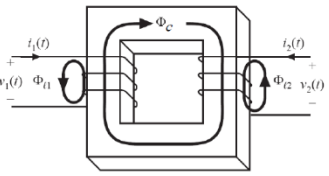
t_1 and t_2 depend on the voltage waveform we are applying on the primary. If for instance v_1 is a square wave, t1 is the beginning of the positive square and t2 it's the end of it, so the time interval where the voltage is always positive or always negative.

Leakage fluxes

We have to add additional components that makes the difference between a real transformer and an ideal one. We have to add leakage inductances that come from the leakage flux.
In fact, not all the flux linked with the primary windings is also linked with the secondary windings. This non ideal behaviour can be accounted for adding two leakage inductances, one on the primary side and the other on the secondary side.

$$\phi_c = \frac{N_1 i_1 + N_2 i_2}{\Re_c};\quad \phi_{l1} = \frac{N_1 i_1}{\Re_{l1}};\quad \phi_{l2} = \frac{N_2 i_2}{\Re_{l2}}$$

$$\lambda_1 = N_1(\phi_c + \phi_{l1});\quad \lambda_2 = N_2(\phi_c + \phi_{l2})$$



$$v_1 = \frac{d\lambda_1}{dt} = L_{l1}\frac{di_1}{dt} + L_m\left(\frac{di_1}{dt} + \frac{N_2}{N_1}\frac{di_2}{dt}\right) \quad (1)$$

$$v_2 = \frac{d\lambda_2}{dt} = L_{l2}\frac{di_2}{dt} + \frac{N_2}{N_1}L_m\left(\frac{di_1}{dt} + \frac{N_2}{N_1}\frac{di_2}{dt}\right) \quad (2)$$

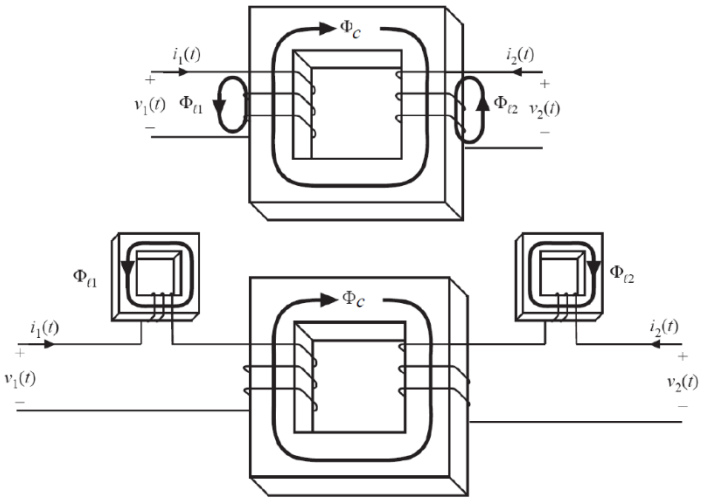
with

$$L_m = \frac{N_1^2}{\Re_c};\quad L_{l1} = \frac{N_1^2}{\Re_{l1}};\quad L_{l2} = \frac{N_2^2}{\Re_{l2}}$$

- Equations (1) and (2) provide a mathematical description of the input-output characteristics of a generic transformer

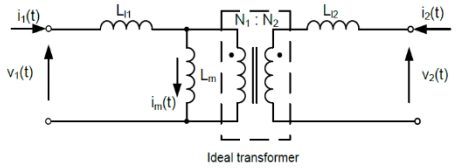
We start with the mutual flux (or common flux) phi_c, that is the one linked to both primary and secondary and we also have to link the leakage fluxes phil1 and phil2, where Rl1 and Rl2 are the reluctances of the leakage paths.

Then v_1 is by definition the time derivative of the flux linkage lambda1, and the same for v_2. Two new terms are appearing, the leakage inductances Ll1 and Ll2.
This system of equation provides a mathematical description that fully describes the real transformer, and the leakage inductances can be accounted for introducing two real inductances in the model in series with the primary and secondary sides.



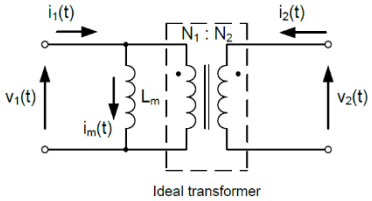
TRANSFORMER MODEL

An equivalent model that satisfies Eqs. (1) and (2) is illustrated below. This model is known as the T-model for a two-winding transformer.



Note: core and copper losses are neglected

In the initial analysis we'll use this simplified model, assuming an almost ideal magnetic coupling (k≈1)

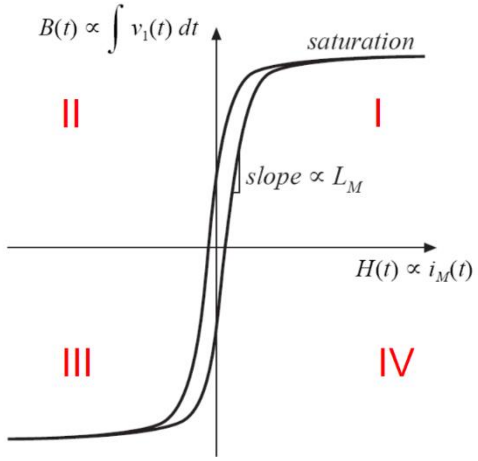


This is the physical model in which all the components that appear in the model are strictly related to the physical characteristics of the transformer. In our analyses we won't consider the leakage inductances, assuming there is a perfect coupling between the primary and secondary windings.

Moreover, there is something missing in this model, because we have also to consider the resistances of the windings to take the copper losses into account and we also need to put an additional resistance in series with the magnetizing inductance which accounts for the losses in the copper.

TRANSFORMER CORE B-H CHARACTERISTIC

In the image we have the B-H characteristic of a generic ferromagnetic material; we can split the characteristic in 4 quadrants.

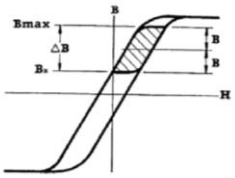


We split it because the typical isolated converter uses the magnetic cores by exciting them in two ways:

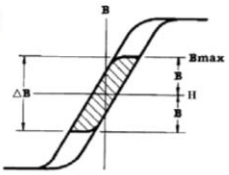
1. **Unidirectional core excitation**, and the magnetic core is used just in the first quadrant. The flyback converter and the forward converter use this approach.
2. **Bidirectional core excitation**, and we are using the 1st and 3rd quadrants. There are 3 isolated topologies that use the core in a bidirectional way: push-pull converters, half-bridge converters and full-bridge converters.

Two kinds of core excitation for the above converters

- Unidirectional core excitation, where only the positive part (quadrant I) of the B-H loop is used:
  - flyback converter (derived from buck-boost converter)
  - forward converter (derived from buck converter).
- Bidirectional core excitation, where the positive (quadrant I) and the negative (quadrant III) parts of the B-H loop are utilized alternatively
  - push-pull converter
  - half-bridge converter
  - full-bridge converters

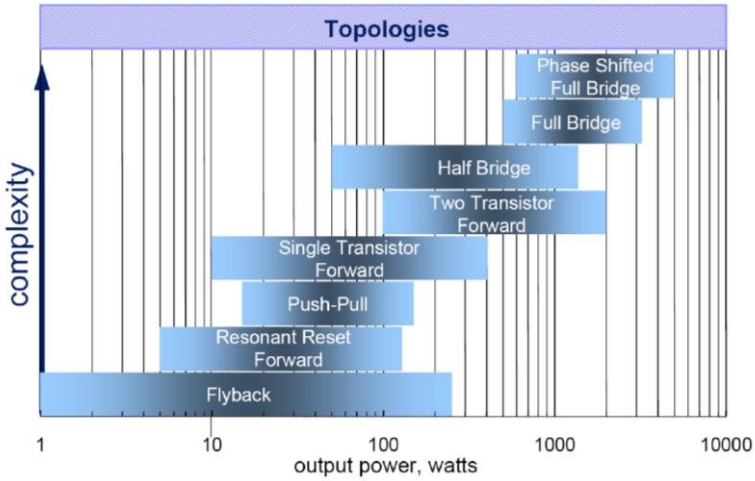


"Single-Ended"
AC operation with DC offset



"Double-ended"
AC-only Operation

The various isolated topologies have the following ranges of operations, depending on the output power.



The flyback topology is typically used for output power up to 100-150W, so low power applications. If we need to increase the output power we should move to other converters.

FLYBACK CONVERTER

The Flyback converter is derived from the buck boost converter. By replacing the inductor with a transformer in the buck converter we get the flyback.
It is the most commonly used in low power applications < 100W. We need a switch, a diode and the transformer. Dots are not on the same side.

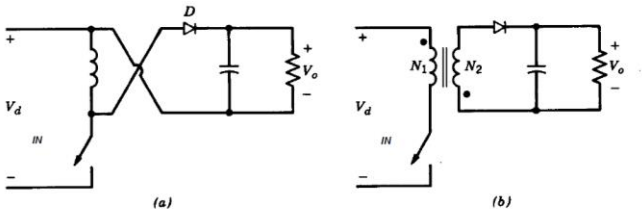


Figure 10-6 Flyback converter.

- Flyback is the most popular topology for low-power (<100 W) AC/DC conversion applications:
- a flyback transformer combines the actions of an isolating transformer and an output inductor into a single element hence, no separate inductor is needed;
- a flyback transformer provides galvanic (electrical) isolation up to ~5 kV DC, and it can provide multiple and/or negative outputs;
- simple design requires just one semiconductor switch (MOSFET) and one freewheeling diode.

CCM

Firstly we replace the transformer with the simplified model made of an ideal transformer and a magnetizing inductance. In CCM, the current in the magnetizing inductance never reaches zero.
Starting from this assumption and assuming that the diode is ideal, we have to understand the behaviour.

If the switch is on, a positive voltage is applied to the primary side that produces a voltage on the secondary side in the opposite direction and since the output voltage is positive upward, the diode is reverse bias and so no conduction on the secondary side and the output capacitor is providing all the current needed by the load. All the input current flows in the magnetizing inductance.
The physical meaning of this is that we are increasing the energy stored in the core of the transformer.

- During t_{on}, V_{in} is applied to primary and current ramps up in L_m.
- As switch opens, voltage "flies back" until caught by secondary diode / cap at V_{out} on secondary side, = ($-N_1/N_2\ V_{out}$) on primary winding (L_m).
- During t_{off}, current ramps down in L_m.

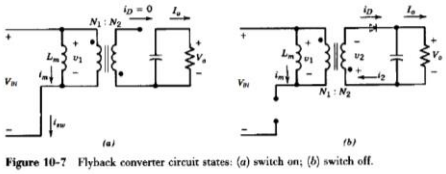
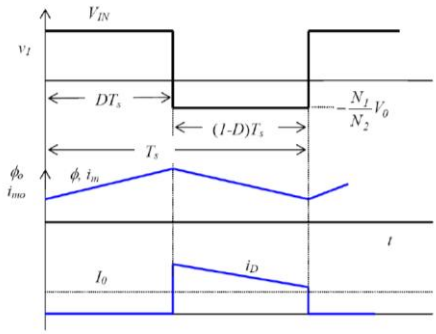


Figure 10-7 Flyback converter circuit states: (*a*) switch on; (*b*) switch off.



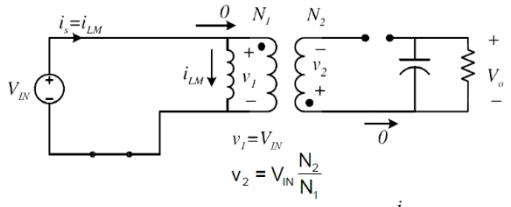
In the off time period, the current in the magnetizing inductance cannot be null instantaneously because there is energy stored in the magnetic core. The magnetizing current is exiting the dots in the primary side and so the physical current is entering the dots in the secondary side, making the diode conductive.

When the diode is conductive we are applying a voltage Vo to the secondary side. In the primary side we have a reflected voltage Vo * N1/N2 downward, so the voltage across the inductor has changed sign (it was upward in the Ton phase). This is the reason for the name 'flyback', since the voltage flies back.
The effect of the voltage inversion is to discharge the inductor, reducing the current in it → the energy is transferred to the output.

The first plot of the previous image is the voltage across the magnetizing inductance, v1. The first blue curve is either the magnetic flux or the magnetizing inductance current, which is triangular, as usual, with a certain average value. The last plot is the current that flows in the diode.

So during the on time period the energy stored increases and no current on the secondary side, while during the off time period the energy stored in the magnetic core reduces and power is transferred to the output. The current either flows in the primary during the on time period or in the secondary during the off time period, this is why there is no instantaneous power flow from the input to the output.

DC voltage transfer function



$$\Delta i_{L_M}^{+} = \frac{V_{IN} D T_S}{L_M}$$

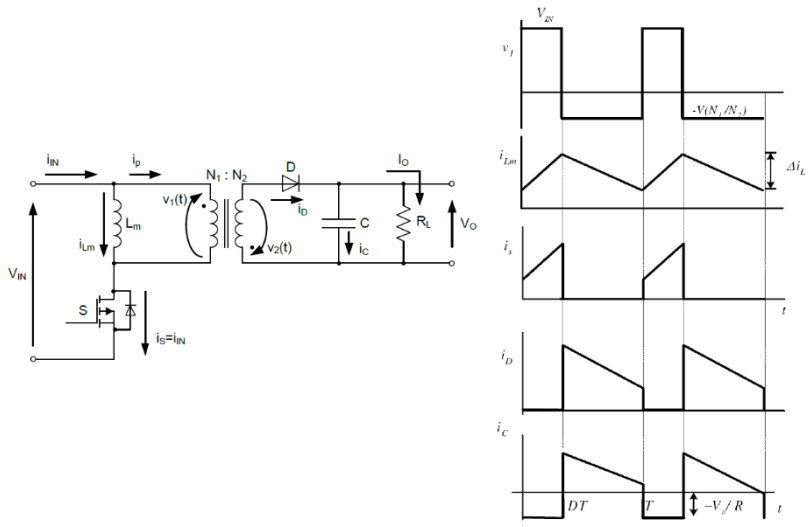
$$\Delta i_{L_M}^{-} = \frac{N_1}{N_2}\frac{V_O(1-D)T_S}{L_M}$$

$$\Delta i_{L_M}^{+} = \Delta i_{L_M}^{-} \Rightarrow \frac{V_O}{V_{IN}} = \frac{N_2}{N_1}\frac{D}{(1-D)}$$

Input output relationship is similar to buck-boost converter.

We apply the V*s balance on the magnetizing inductance. During Ton the voltage is Vin (Vin*D*Ts), while during the off time period it is reversed and Vo*N1/N2.
We get a result that is similar to the DC t.f. of a buck boost converter but there is the N2/N1 additional term, that is the turn ratio.

CCM waveforms



The first plot is the voltage applied to the magnetizing inductance (areas must be the same), then we have the current in the inductance, the switch current and the diode current.
The average diode current Id = Io, because the average current across the capacitor is 0. So if we shift the waveform i_D down by a quantity Io we get the i_c, instantaneous current that flows in the capacitor, and the positive area must match the positive on at steady state.

Recall

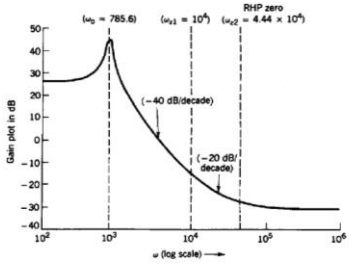
- **Conventional transformers store minimal energy**
  - E.g. Forward converter, Push-Pull converter etc.
  - Primary and secondary currents flow simultaneously.
  - Energy is transferred by the simultaneous flow of current.
- **Flyback "transformer" is not acually a transformer**
  - Primary and secondary windings do not conduct simultaneously
  → often called ***two-winding inductor***.
  - Current flows in primary while secondary diode is reverse-biased.
  - Primary current stores energy in air-gap.
  - When primary current stops, secondary current flows.
  - Air-gap energy is transferred to the secondary load:
    - **CCM** => only some of the stored energy is delivered to the load, primary current starts before secondary current has decayed to zero;
    - **DCM** => all stored energy is delivered to the load, followed by an interval of zero current flow in both windings.

A transformer is a magnetic component that doesn't store energy, current flows always simultaneously in the primary and in the secondary side. Since there is charge storage in the flyback the flyback core must be gapped.

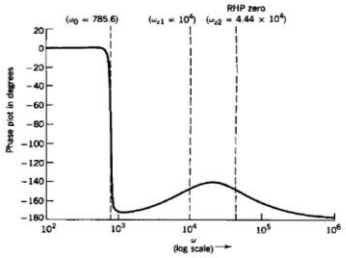
RHP ZERO OF THE FLYBACK CONVERTER IN CCM

$$G_{od}(s) = \frac{\tilde{v}_O(s)}{\tilde{d}(s)} = G_{od}(0)\frac{\left(1+\frac{s}{2\pi f_{z1}}\right)\left(1-\frac{s}{2\pi f_{z2}}\right)}{1+\frac{s}{2\pi f_O Q}+\left(\frac{s}{2\pi f_O}\right)^2}$$

A right half plane (RHP) zero is produced in the duty cycle-to-output transfer function when the flyback is operated in CCM.



gain

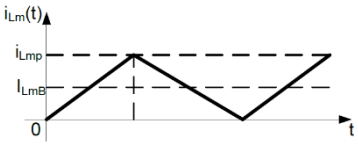


phase

The typical God for a flyback converter working in CCM has a pair of c.c. poles, which comes from the interaction of the magnetizing inductance and of the output capacitance, and a negative zero from the output capacitor and then a RHP zero.

A flyback converter for small power is typically used in DCM, not CCM, because in DCM we loose the RHP zero and this makes a big difference when closing the feedback loop, it is easier, and we also loose a pole.

CCM-DCM BOUNDARY



$$I_{LmB} = \frac{i_{Lmp}}{2} = \frac{V_{IN} D T_S}{2L_m}$$

Using $\frac{V_o}{V_{IN}} = \frac{N_2}{N_1}\frac{D}{(1-D)}$ we get:

$$I_{LmB} = \frac{V_O (1-D) T_S}{2L_m} \cdot \frac{N_1}{N_2}$$

As usual: If $I_L > I_{LB}$ → CCM; If $I_L < I_{LB}$ → DCM

The magnetizing inductor current starts from 0 at the beginning of the switching cycle, reaches a maximum at DTs and then drops back to zero reaching it at the end of the off period. The average current in this situation is the average boundary current i_LmB, which is half the peak current.

CONDITION FOR DCM OPERATION

We start from the definition of the boundary current.

Is there a way to connect I_Lm, current that flows in the magnetizing inductance, with the output current?
Yes, we can look at the two plots in the image. In the upper plot we have the instantaneous current flowing through the magnetizing inductance at the boundary, while the second one flows the current that flows through the diode. During the off time period, current in the diode is not simply equal to the current in the magnetizing inductance, because we have to consider the turn ratio of the transformer, so we have a factor N1/N2.

The average output current is then the average of the diode current, so it is the red area divided by Ts. In x we replace the Io = Vo/R.

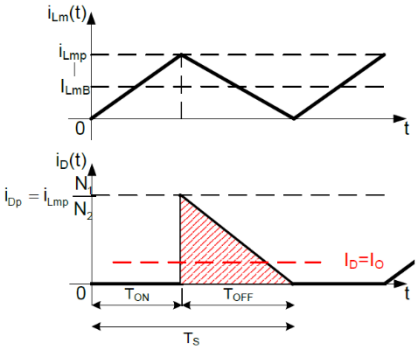
$$I_{LmB} = \frac{V_O (1-D) T_S}{2L_m} \cdot \frac{N_1}{N_2}$$

Note: in CCM, $I_O = I_D = I_{Lm} \cdot \frac{N_1}{N_2} \cdot (1-D)$

➢ DCM if $I_{Lm} < I_{LmB}$

x $\Rightarrow \frac{V_O}{R} \cdot \frac{N_2}{N_1} \cdot \frac{1}{(1-D)} < \frac{V_O T_S}{2L_m} (1-D) \cdot \frac{N_1}{N_2}$

i.e. $\frac{2 \cdot L_m}{R T_S} < (1-D)^2 \cdot \left(\frac{N_1}{N_2}\right)^2$



➢ This can also be expressed as:

$K < K_{crit}(D, N_1, N_2)$ for DCM

with $K = \frac{2L}{RT_S}$

and $K_{crit}(D, N_1, N_2) = (1-D)^2 \cdot \left(\frac{N_1}{N_2}\right)^2$

We can also introduce some different definitions to define the conditions of operations in DCM. For instance, the flyback is in DCM if the output current is smaller than the output boundary current, which is the current Io with I_LmB in place of I_Lm.

- It is natural to express the mode boundary in terms of the output power or magnetizing inductance rather than the dimensionless parameter K.

➢ DCM if $\frac{V_O}{R} < \frac{V_O T_S}{2L_m}(1-D)^2 \cdot \left(\frac{N_1}{N_2}\right)^2$

i.e.
$$\begin{cases} P_O < \frac{V_O^2 T_S}{2L_m}(1-D)^2 \cdot \left(\frac{N_1}{N_2}\right)^2 \Rightarrow P_O < \frac{V_{IN}^2 \cdot D^2}{2L_m f_S} \\ L_m < \frac{V_O^2 T_S}{2P_O}(1-D)^2 \cdot \left(\frac{N_1}{N_2}\right)^2 \Rightarrow L_m < \frac{V_{IN}^2 \cdot D^2}{2P_O f_S} \end{cases}$$

Note: in CCM , down to boundary between CCM and DCM ,

$$\frac{V_O}{V_{IN}} = \frac{N_2}{N_1}\frac{D}{(1-D)}$$

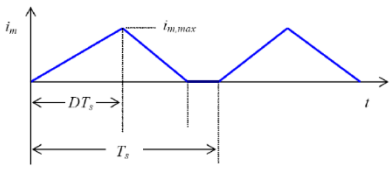
DCM

We want to compute the DC t.f. for the converter in DCM. We can derive it in two ways; the first one is using the approach of part 5, the other possibility is using the approach down in the slide, based on how the flyback works.

The flyback stores energy in the core during the on time period, and this energy is completely transferred to the output during the off time period. Completely because during DCM the current in the magnetizing inductance reaches zero before the end of the switching period. whSo in theory we can derive the DC t.f. applying an energy balance. The energy stored in the on time period is equal to the one released to the output in the off time period.
Assuming that there are no losses, all the energy W is released to the output, so we can compute the Po because the charge and discharge is performed every period.

The energy transferred from the input dc voltage source V_{IN} to the magnetizing inductance during one cycle for the DCM case is:

$$W = \frac{1}{2} \cdot L_m \cdot i_{m,max}^2$$



Assuming 100% efficiency, this results in dc output power given by:

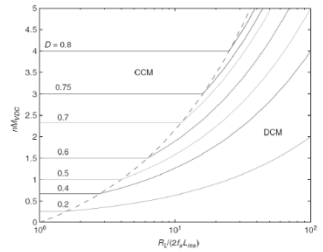
$$P_O = \frac{1}{2} \cdot L_m \cdot i_{m,max}^2 \cdot f_S = \frac{V_O^2}{R}$$

since

$$i_{m,max} = \left|\Delta i_m^+\right| = \frac{V_{IN} D}{L_m f_S}$$

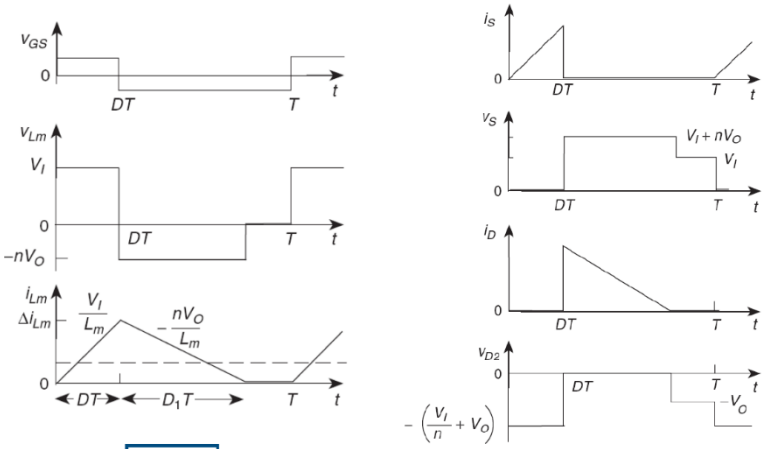
it follows that

$$\frac{V_O}{V_{IN}} = D\sqrt{\frac{R}{2L_m f_S}}$$



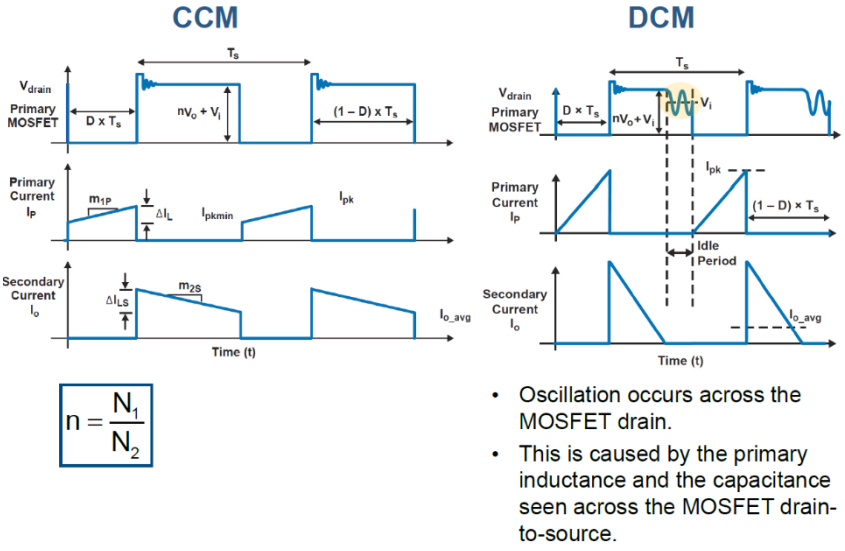
Thus the feedback loop will regulate the output by decreasing D as V_{IN} or R goes up, increasing D as V_{IN} or R goes down.

DCM waveforms



$$n = \frac{N_1}{N_2}$$

The only important thing to understand is the current in the magnetizing inductance, which starts from 0, reaches a maximum, goes back to 0 and then stays equal to 0 for a fraction of the switching period.
As for the voltage across the switch Vs (Vds), when the switch is closed there is no voltage drop across the switch (or small on the Rds,on). When the switch is open we have a first period where Vs is the input voltage plus the **reflected output voltage**, which his Vo*N1/N2 (n = N1/N2).
However, this voltage stays equal to this value as far as there is energy stored in the magnetizing inductance. When the current in the inductance reaches zero there is no more energy and the voltage drop across the inductance drops to 0, therefore the voltage across the switch remains just Vin.
In the real world we have resonances to reach the steady value Vin because immediately before the collapse the voltage stored on the parasitic capacitance of the switch is Vin+Vo*n, the final voltage is Vin and so there is an additional energy stored in the capacitor which bounces back and forth between the parasitic capacitor at the drain node and the magnetizing inductance.



$$n = \frac{N_1}{N_2}$$

(see part 8 slides for optional slides, 37-38)

We can take advantage of the resonances. When the device is in DCM, as soon as the energy stored in the inductor gets to 0, the resonances appear, so the switch voltage is reducing below the final voltage due to the ringing.

If we could sense the voltage and opening the switch exactly when the first ringing has reached the valley value, the advantage is that the voltage Vds is smaller to turn on the switch with respect to when we reach the steady state Vin, so we can spare some switching power, reducing the switching loss.

MAGNETIZING INDUCTANCE Lm DESIGN

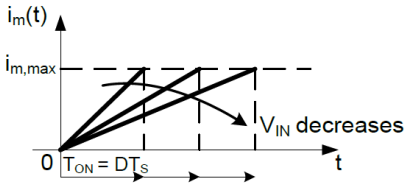
In a flyback converter we have to design the magnetizing inductance and the turn ratio. We want to design the Lm so that the converter works always in DCM regardless the input voltage or the output power.
The converter is in DCM if the condition in the blue box is fulfilled. It is an upper bound for Lm.
Any inductance value smaller than the upper bound provides operations in DCM. The problem is that in this formula neither Vin nor Po are constant, so we have to go for the worst case scenario and be sure that the device works in DCM or at the boundary between DCM and CCM. Hence the task is to identify the worst condition.

- The magnetizing inductance L_m must be limited to a maximum value, $L_{m,max.}$
- DCM ensured if $L_m \le \frac{V_{IN}^2 D^2}{2P_O f_S}$

note: flyback operates at CCM/DCM boundary when the "=" holds.

Worst conditions? V_{INwc}? D_{wc}? P_{Owc}?

- $P_{Owc} = P_{Omax}$
- $V_{INwc} = V_{INmin}$



- At a given P_O, $i_{m,max} = \text{const} = \sqrt{\frac{2P_O}{L_m f_S}}$
- The lower V_{IN}, the longer the T_{ON} needed to reach $i_{m,max}$, that is the larger D.
- $V_{IN,min}$ will make the flyback to operate at D_{max}

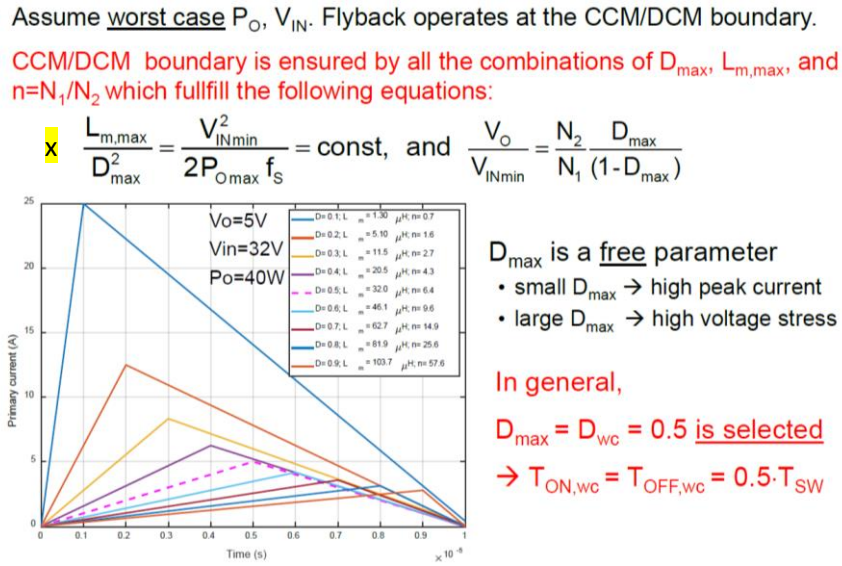
The worst case in terms of Po is when it is maximum, while for Vin we have to choose the minimum input Vin for the worst case.
However, a typical mistake is considering for D its minimum value as the worst situation. But this is not true, because the Vin and D are not independent variables; at a given output power Po we can compute the maximum current that is reached in the primary side at the end of the on time period (im,max) and if the input voltage is decreased we need more time, and therefore we need to apply a larger duty cycle, to reach the peak current. So the lower the input voltage, the larger the duty cycle. We cannot have minimum Vin and minimum D, otherwise we don't reach the minimum current to provide the output power.

Hence **the critical situation is defined by Vin minimum, maximum Po and maximum D**.

There is a problem with the relationship in the blue box. When designing a flyback converter, we know Vin,min because it is a design data, as well as for Po, but we don't know exactly what the maximal D is. We know that Vin,min results in Dmax, but we don't know the exact value for Dmax.

Dmax

Let's assume that we are in the worst case (Po,max and Vin,min). The CCM/DCM boundary is in principle ensured by all the combinations of Dmax and Lm,max that satisfy equation x.



Of course we have also to verify the equation on the right side. It looks like there is a free parameter that we can choose, e.g. the Dmax; once we have chosen D, we put it in and we can compute Lm,max.

So is there a smart way to choose Dmax?

The plot reports the primary current, i.e. the current that flows during Ton in the primary side, as a function of time for a flyback converter operated at the boundary between DCM and CCM.

There are several curves plotted for different D values. The converter is working with Vo = 5V and Vin = 32V, and Po = 40W.

We notice that if we increase the D from 0.1 to 0.9 the curve moves downward. If the D is too small, we have a large input current (25A) so high dissipation in the power mosfet, so we can increase the D, and at D = 0.9 the primary current is small, but we have another disadvantage, that is that the larger the D the smaller the peak current but also the larger the turn ratio N1/N2.

The disadvantage in having a large turn ratio is that when the mosfet is off the output voltage that is reflected on the magnetizing inductance is huge because of a large turn ratio. This forces us to use a mosfet device with a large breakdown voltage.

Hence the optimal point for D is exactly in the middle.

Lm design conclusions

The Lm design is done considering the Dmax = 0.5 and with the formula we can define the upper bound for the magnetizing inductance.

Any value smaller than the boundary is ok, but of course not too small because the smaller the larger the input current.

In the following image we have the plots for possible situations. In the worst case the Vin is minimum, the Po is maximum and D = 0.5, and we are at the boundary between DCM and CCM.

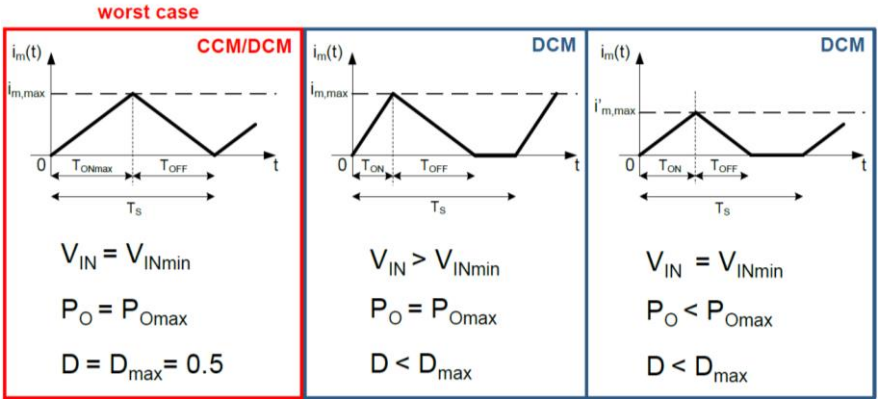
If the input voltage is increasing we are increasing the slope of the Ton time, and if so, if the Po is constant we still have to reach the same peak value of the current but at a smaller D*Ts, therefore we are decreasing D and we are entering the DCM. In the Toff phase the slope is still the same.

If Vin is kept minimum and Po is reduced, the slope during Ton is constant but the maximum current to be reached is reduced, therefore we reach it at a smaller DTs with respect to the worst case, so D < Dmax

and we are in DCM. It is the feedback loop that controls the D in a way that, upon a variation of the Vin or Po the D is set properly.

The point is that in any situation different than the worst case we are working in DCM.

- D_{max} is set by the designer around 0.5 (precisely 0.5 in our design)
- $L_{m,max}$ is selected from the equation $L_{m,max} = \frac{V_{INmin}^2 D_{max}^2}{2 P_{Omax} f_S}$
- N_1/N_2 is selected such to fully discharge L_m within T_S (see next slide)



Lm design recap

Use worst case $P_O = P_{Omax}$, $V_{IN} = V_{INmin}$, $D=D_{max}$

$$L_{m,max} = \frac{V_{INmin}^2 \cdot D_{max}^2}{2 P_{Omax} \cdot f_S} \xrightarrow{\text{Assuming } D_{max}=50\%} L_{m,max} = \frac{V_{INmin}^2}{8 P_{Omax} \cdot f_S}$$

If converter efficiency, η, is taken into account, we easily get:

$$L_{m,max} = \frac{\eta \cdot V_{INmin}^2}{8 P_{Omax} \cdot f_S}$$

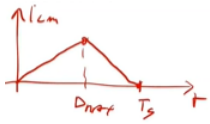
- Assume η ~ 0.75 - 0.85 as a first guess

However, in all this reasoning we never took the efficiency of the converter into account. If we take into account in the definition of Lm, we just have to have eta directly related to Lm.

Of course at the starting of the design phase we don't know the value for the efficiency eta, so in general we start with a guess, that is in general between 75% and 85%, then we perform the design and at the end of the design phase we compute the efficiency. If it is consistent with the guess the design is over, if not we have to iterate the design.

TURN RATIO DESIGN

It is defined in the worst condition by applying the V*s balance. In fact i_lm(t) increases in the Ton until Dmax and it goes down and reaches zero at the end of Toff. So we can apply a V*s balance to Lm obtaining that Vo/Vin,min = N2/N1 * Dmax/(1-Dmax).



$$\Delta i_{Lm}^{-} = \Delta i_{Lm}^{+} \Rightarrow V_O \frac{N_1}{N_2} T_{OFF} = V_{IN} T_{ON} \Rightarrow \frac{N_1}{N_2} = \frac{V_{IN}}{V_O} \frac{T_{ON}}{T_{OFF}}$$

For $D=D_{max}=0.5$, $V_{IN}=V_{INmin}$ (worst conditions) we get $T_{ON}=T_{OFF}=T_S/2$ at CCM/DCM boundary. Therefore:

$$\frac{N_1}{N_2} = \frac{V_{INmin}}{V_O}$$

Note: this is a lower bound for N_1/N_2. Any larger value of N_1/N_2 will make the flyback to operate in DCM

More accurately, taking into account non-idealities:

$$\frac{N_1}{N_2} = \left(\frac{V_{INmin}}{V_O + V_D} \right)$$

The Vin,min/(Vo + Vd) defines a lower boundary for the ratio N1/N2, it is a minimum value. We can use a turn ratio higher than it but not lower.

CASE STUDY – FLYBACK REGULATOR DESIGN

Flyback regulator is flyback converter plus the feedback.

SPECIFICATIONS

- Input voltage: V_{IN} = 22 ÷ 38 V (V_{IN_nom}= 28V)
- Output voltage: V_O = 5 V
- DCM operation
- Output voltage ripple: ΔV_O = 100 mV
- Output current: I_O = 10 A → P_o = 50 W
- Switching frequency: f_S = 100 KHz
- Closed loop bandwidth: f_C = 20 kHz

PRELIMINARY STEPS

Step 1: determine the turn ratio

$$\frac{N_P}{N_S} = \left(\frac{V_{INmin}}{V_O + V_D} \right) = 3.9 \rightarrow 4$$

Assume V_d ~ 0.6V
e.g., Vishay VS-62CTQ030-M3
Schottky Rectifier, 2 x 30 A

Step 2: determine the switch voltage rating

$$BV_{DSmin} > V_{INmax} + \frac{N_P}{N_S} \cdot (V_O + V_D) \approx 60 \text{ V}$$ Power MOS

Note: $BV_{DSmin} > (V_{INmax} + V_Z)$ if an active zener clamp is exploited

$$BV_D > \frac{N_S}{N_P} \cdot V_{INmax} + V_O \approx 14.5 \text{ V}$$ diode

We start by determining the turn ratio since we have all the information. Vd is the voltage drop across the diode, which we consider if we want to be accurate because the output voltage Vo = 5V is not very large, so Vd has the same order of magnitude of Vo. Vd depends on the diode we select in the flyback converter.
The smaller the voltage drop across the diode the better, because the average current that flows in the diode is equal to the average output current and so the average power dissipation in the diode is Io*Vd. If Vd = 1V, since Io = 10A, we end up with 10 W of dissipated power → better to use power Schottky diode to reduce the voltage drop across the diode. In our case we use two Schottky diodes in parallel because the peak output current is 40A, and with two diodes in parallel we can handle 60A at max.
Also in the case of Vd we need to take the maximum voltage Vd off the datasheet of the diode.

The second step is to determine the switch voltage rating, which is the minimum voltage that the diode and the mosfet must be able to sustain to avoid breakdown. Of course it is a lower bound.
If we start from the power MOS, the minimum voltage BVds,min must be bigger than the maximum Vds that might occur. Of course also in the BVds,min we need to take some margins.
The same reasoning applies for the diode, where we have to compute the maximum reverse voltage across it, and the breakdown must be larger than this. The maximum reverse voltage happens during Ton, where we are applying a voltage on the secondary side that is Ns/Np * Vin. Once again, this is the lower bound.

The third step is to estimate the converter efficiency. We need this estimation because the estimated efficiency is needed in the Lm,max formula.

To estimate the converter efficiency we have to highlight the major contributions to the power loss, that in our case come from the Schottky diode, the transformer (core and copper losses) and mosfet (conduction and switching losses).
So we have to estimate these three contribution to have a first guess of the efficiency.
Pschottky is already known from the beginning because the average power dissipated by it can be computed multiplying the forward voltage drop across the diode by the average current that is Io, because the average current in the capacitor is 0.

Of course we don't know the power losses in the transformer and switch, because we still have to design the transformer and choose the switch. So we start with a target losses for these two and we go on with the design, checking if they are met. In general, we try to be around 2% of the target output power with these losses.

Once the efficiency is known, we can compute the magnetizing inductance Lm, and the formula provides an upper bound.

Step 3: estimate the converter efficiency

$$P_{loss} = P_{schottky} + P_{transf} + P_{MOS}$$

Schottky diode loss is known:

$$P_{schottky} = V_d \cdot I_O = 6W$$

Set target max loss for transformer and power MOSFET, e.g. :

$P_{transf} = 2W$; $P_{MOS} = 2W$. This specification must be met by design.

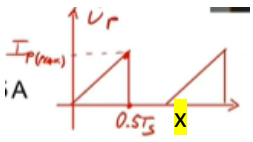
Therefore:

$$\eta = \frac{P_O}{P_O + P_{schottky} + P_{transf} + P_{MOS}} \approx 84\%$$

Step 4: determine the magnetizing inductance

$$L_{m,max} = \frac{\eta \cdot V_{INmin}^2}{8\, P_O \cdot f_s} \approx 10\,\mu H \qquad \text{(assume } \eta=0.84\text{)}$$

The fifth step is to determine the peak and rms value of the primary current, which is necessary what power mosfet to detect, because the conduction loss in the power MOS dependes on the rms current that flows in the primary side, while the switching loss depends on the peak current that flows in the primary side.



We put ourselves in the worst condition, with Vin,min and Dmax. The peak of the current can be determined using the delta(i) formula, so the voltage across the magnetizing inductance, Vin,min/Lm, multiplied by the duration of the Ton = Dmax/fs. The lower Lm, the larger the peak current, this is why we shouldn't exaggerate in decreasing Lm.

The conduction losses can be easily computed because they are given by the product of the Rds,on of the mosfer in ohmic region multiplied by the squared rms value of the current.
As for the switching losses, the switch is open at the end of Ton, when Ip goes down. In point x, at the beginning of the next switching period, the switch is turned on but there is no current in the magnetizing inductance, so the turn on transition is a **zero current switch**, so there is no energy dissipated during this transient. The only energy is dissipated during the turn off transition. This is why the average switching loss is computed considering only the t_off (turn off transient, not Toff), which is read on the datasheet.

Vds,max is the maximum voltage between drain and source during the turn off transition.

Step 5: determine the peak and the rms primary current in order to fully specify the power switch

$$I_{P(peak)} = \frac{V_{INmin} \cdot D_{max}}{L_m f_S} = 11\,A; \qquad I_{P(rms)} = \frac{I_{P(peak)}}{\sqrt{6}} = 4.5\,A$$

Switching loss is not negligible!

$$P_{sw} = \frac{V_{DSmax} \cdot I_{p,max}}{2} \cdot t_{off} \cdot f_S$$

Note: no energy dissipated during the turn-on transition

Conduction loss:

$$P_{cond} = r_{DSon} \cdot I^2_{P(rms)}$$

Select a power MOSFET such that:

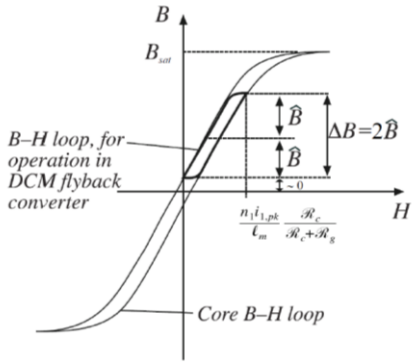
$$P_{MOS} = P_{cond} + P_{sw} \leq 2W$$

For example: Texas Instruments CSD19537Q3 → $P_{MOS} \approx 0.9$ W ☺
Infineon BSC252N10NSF G → $P_{MOS} \approx 1.1$ W ☺

The sixth step is the transformer design.

TRANSFORMER DESIGN

Step 6: select the core material



- In DCM, flux density swings all the way from zero to ΔB (flux remanence is negligible because of the gap).

- Ferrite is usually the best choice for inductors designed to operate in the discontinuous mode at frequencies above 50kHz, when core loss associated with large flux swing limits core utilization.

The flyback transformer is not actually a real transformer, but this means that the flyback uses the transformer as a two windings inductor, because we have to store energy in the magnetic core during Ton and release this energy from the core to the output during Toff. This is not the typical behaviour of a transformer, where power in immediately goes out. Nevertheless, the magnetic component in a flyback is still a transformer from a structural point of view, it is the usage that is not the classical one. the only difference is that we have a gap in the core to store energy.

The first step for the design of the transformer is the selection of the core material, which is important to prevent saturation and excessive power dissipation. Furthermore, we are design a flyback that works in DCM, so current in the primary side is a triangular one, so we have a large swing of current, which corresponds to a large loop in the B-H diagram of the magnetic core (bold one).
We have to pay attention not to saturate the core and to the power dissipated in the core, which is proportional to B_cap^alpha, where alpha is typically between 2 and 2.5, so strong dependance.

In general, the magnetic flux density B never reaches zero when the applied magnetic field strength H goes to 0, however, since our core is gapped, we are stretching the B-H loop and therefore we can assume

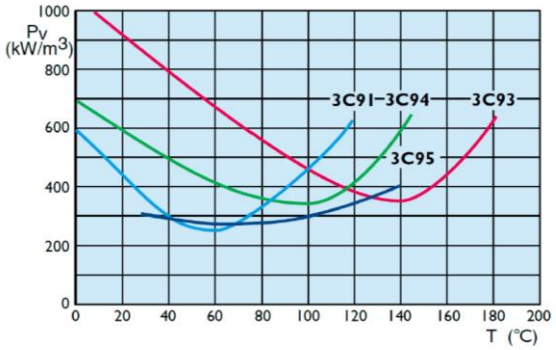
that the minimum B is equal to 0, because we are very close to it, it is useless to take the **remanence** into account.

Due to this large swing of the B, we can expect serious losses, and this is the reason why for flyback converter in DCM we never use distributed gap cores, because in these cores the losses due to eddy currents are not 0. Hence we have to go for ferrite.

In principle we should look at the performance factor and select a material which provides the best or maximum of it around 100 kHz. Then, to further differentiate, we can look at the specific losses of the material as a function of temperature. If we look at the plot below, most of the ferrite materials show a sharp minimum, so if we don't work at a temperature where the minimum occurs we might end up with a large specific power dissipation. Instead, the blue line is almost flat, so specific power dissipation is unchanged for a large range of temperatures.

E.g. Ferroxcube 3C95

FERROXCUBE



Power losses at 100 kHz and 200mT as a function of temperature

3C95 : a low to medium frequency power material with low power losses from 25 to 100 °C for use in power transformers at frequencies up to 0.5 MHz.

Once the material is chose, we have to select the core size. The flyback converter uses the transformer as a two windings inductor, so we can pick up the size of the transformer using the Ap formula seen when discussing the design of the inductor. Indeed, the magnetizing inductance Lm is actually an inductor.

Step 7: select the core size

The following area-product expressions are intended to help provide a rough initial estimate of core size for the flyback transformer (a two-winding inductor, actually).

$$A_p = W_a \cdot A_e \geq \left(\frac{L_m \cdot I_{max} \cdot I_{rms}}{B_{max} \cdot K_1} \right)^{\frac{4}{3}} [cm^4]$$

when core loss is not severe, so that flux swing is limited by core saturation

$$A_p = W_a \cdot A_e \geq \left(\frac{L_m \cdot \Delta I \cdot I_{rms}}{\Delta B \cdot K_2} \right)^{\frac{4}{3}} [cm^4]$$

when flux swing is limited by core loss

Application	K_{pri}	K_1	K_2
Inductor, single winding	0.7	0.03	0.021
Flyback transformer – non isolated	0.3	0.013	0.009
Flyback transformer – with isolation	0.2	0.0085	0.006

K_1 and K_2 have been calculated based on the window utilization factor K_{pri} reported in the table.

Note: K_{pri}=primary copper area/window area.

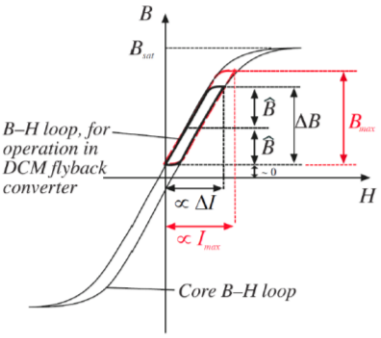
The terms Ku and Jmax of the inductor general formula have been included in the parameters K1 and K2. The other difference is the appearance of the power 4/3. This comes from a dimensional analysis.

We notice that we have two possible design equations depending on the fact that the core is limited by saturation (first formula, and never choose Bmax = Bsat, but smaller) or power dissipation (we use the maximum swing in amplitude). Coefficients K1 and K2 can be read on the tables and they are based on the window utilization factor, inside the term Kpri. Since we have a transformer, we have two windings, so Kpri must be divided by a factor 2.

Core size selection

List of symbols

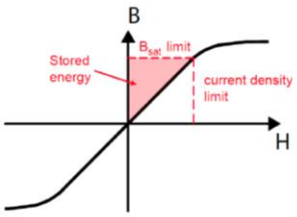
L_m = magnetizing inductance [H]
I_{max} = primary current swing (saturation limited) [A]
B_{max} = saturation limited flux density [T]
ΔI= primary current swing (loss limited) [A]
ΔB = flux density swing (loss limited) [T]
I_{rms} = primary rms current [A]



- The saturation-limited formula assumes winding losses >> core losses. K_1 is based on a hot spot temperature rise (at the middle of the center post) of 30°C.
- In the core loss limited formula, core and winding losses are assumed to be approximately equal. Therefore, the winding losses are halved by reducing the current density. Thus, K_2 equals $0.707 \cdot K_1$.

- The most important limiting factors in design are:
  a) temperature rise and efficiency considerations arising from core losses and ac and dc winding losses, and
  b) core saturation.

- The smallest size and lowest cost flyback transformer is achieved by fully utilizing the core, i,.e by operating the core at maximum flux density (limited either by saturation or by core loss), and also at maximum current density in the windings (limited by winding loss).

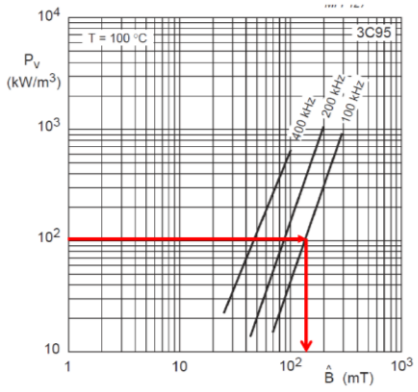


- We first need to determine whether the core is saturation-limited or core-loss limited.

We need to understand if the core is limited by saturation or power loss. We need to use a rule of thumb to do this; we set a maximum allowable core loss used as a target. In our case it is 100 mW/cm^3 allowed losses in a magnetic core.
Then we go to the plot of the specific loss for the selected ferrite material, the one in the image below. For the peak B_cap, we start from the rule of thumb value and we go horizontally until we reach the selected switching frequency, and then we go down to read B_cap.
As for the delta(B), this is the double of B_cap, sto 240mT, which is well below the saturation of the material.

So we can conclude that the flux density swing is not limited by saturation, it is limited by maximum power dissipation (**power limited design**).
So we have to use the formula 2 for core size selection, the one related to core loss.

- To determine if ΔB is core loss limited, enter the core loss curve for the selected core material at the nominal 100mW/cm³ loss limit (rule-of-thumb), and at 100kHz ripple frequency.



- The corresponding peak flux density, $\hat{B}$ is 120 mT.
- Therefore is ΔB is limited to 240 mT, well below the saturation limited Bmax (≈300mT).
- Core is loss limited.
- Use the loss-limited formula:

$$A_p = W_a \cdot A_e \geq \left(\frac{L_m \cdot \Delta I \cdot I_{rms}}{\Delta B \cdot K_2} \right)^{\frac{4}{3}} \text{ [cm}$$

- With ΔB = 0.24T, L=10 μH, I_{rms}=4.5 A, ΔI =11 A and K_2 =0.006, an A_p of 0.24 cm⁴ is required.

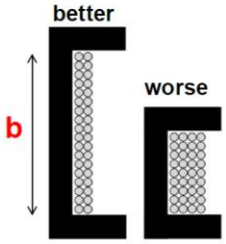
Core shape selection

Once we have the Ap (lower bound), the next step is to select the core shape.

The only important guideline is to choose a closed shape, not to have magnetic flux in the surrounding ambient producing EMI.

Step 8: select the core shape

- A critical step for inductors and especially for flyback transformers operating in DCM, because winding loss is expected to be large.
- The window should be as wide as possible to maximize winding breadth and minimize the number of layers. This minimizes ac winding resistance.



- For a flyback transformer, the wide window also minimizes leakage inductance, and the required creepage distance when line isolation is required has less impact.

  ➢ go for a E-E core

- For example, ferroxcube E 25/13/7, with 6-pin coil former, providing A_p= 2910 mm⁻⁴

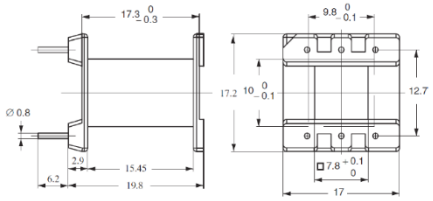
The other important guideline is to **choose a core shape that provides a large breath b**. It is advisable to do so because, if we look at the image above, the smaller the breath the larger the number of copper layer we have to use to get the desired inductance, and this is a problem because the larger the number of layers the worse the proximity effect. If we increase the breath, at the same turn number we can reduce the number of layers. In the example it is selected a double E core (E-E).

The most important characteristics of the core are the effective volume, the effective length and the effective area, together with the area product, the average length of turn, the minimum winding width (that is the effective breath of the bobbin) and the winding area.

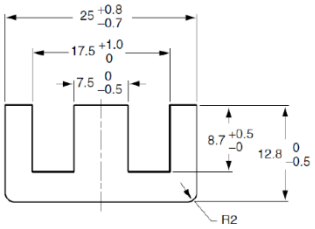
Step 8: select the core shape

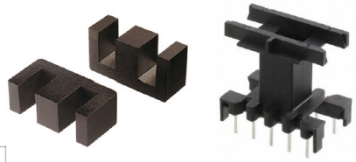
Effective core parameters

SYMBOL	PARAMETER	VALUE	UNIT
Σ(l/A)	core factor (C1)	1.11	mm^{-1}
V_e	effective volume	2990	mm^3
l_e	effective length	58.0	mm
A_e	effective area	52.0	mm^2
A_{min}	minimum area	52.0	mm^2
m	mass of core half	≈ 8	g



NUMBER OF SECTIONS	WINDING AREA (mm^2)	MINIMUM WINDING WIDTH (mm)	AVERAGE LENGTH OF TURN (mm)	AREA PRODUCT Ae x Aw (mm^4)
1	58	15.45	49	2910





window breadth: 15.45 mm
window height: 3.60 mm

Once the core has been designed, we can compute the number of turns to get the desired magnetizing inductance.

Primary turns calculation

Step 9: calculate the primary turns to provide the desired L

$$N_P = \frac{L_m \cdot \Delta I}{A_e \cdot \Delta B} = 8.8 \rightarrow 8 \text{ Turns}$$

Because N_p was rounded down from 8.8 to 8 turns, flux density swing is proportionately greater than originally assumed:

$$\Delta B_{EFF} = \Delta B_{ORIG} \cdot \frac{8.8}{8} \approx 0.27 \text{ T}$$

Using the core loss curves with the effective flux density swing, a specific core loss, P_C' of 145 mW/cm^3 is obtained at T = 100°C.

The core volume is 2.99 cm^3, thus providing a core loss of :

$$P_C = V_C \cdot P_C' \approx 430 \text{ mW}$$

Rounding up to 9 is not a good choice. We can go for 8 or 10, otherwise we get a fractional number of turns, since the turn ratio is 4.
If we choose 8, the price to pay is that we are choosing a number that is smaller than the target, so to keep Lm at the same value we need to increase delta(B), and so larger power dissipation in the core. If this increase in power dissipation is not acceptable, we have to go for 10 turns.

Delta(Beff) is 0.27, so B_cap = 135 mT, which corresponds, looking at the Pv vs B_cap plot, to a specific core loss of 145 mW/cm^3, which is larger than the rule of thumb we used to check if the core was power limited or saturation limited. To see if it is ok, let's compute the core loss to see if it is acceptable.
Since the core loss is 430 mW, which is ok because the target power loss for the transformer was 2W.

If the power we would have got from the computation is not acceptable, we would have the need of rounding to 10.

Gap length calculation

Last calculation as far the core is concerned.

Step 10: calculate the gap length to achieve the desired L

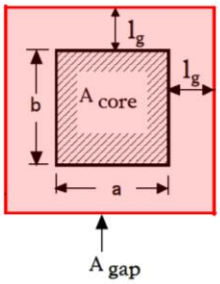
$$\ell_g = \mu_o \cdot N_p^2 \cdot \frac{A_e}{L_m} \approx 420\,\mu m$$ Note: fringing field has been neglected

A better estimation of the gap length including the fringing field correction, brings to:

$$\ell_g \approx 468\,\mu m$$

Gap area correction:

Because of the fringing field, the *effective* gap area is larger than the core center-pole area.
An empirical approximation of the effective gap is obtained by adding the *length* of the gap to the dimensions of the core center-pole cross-section.

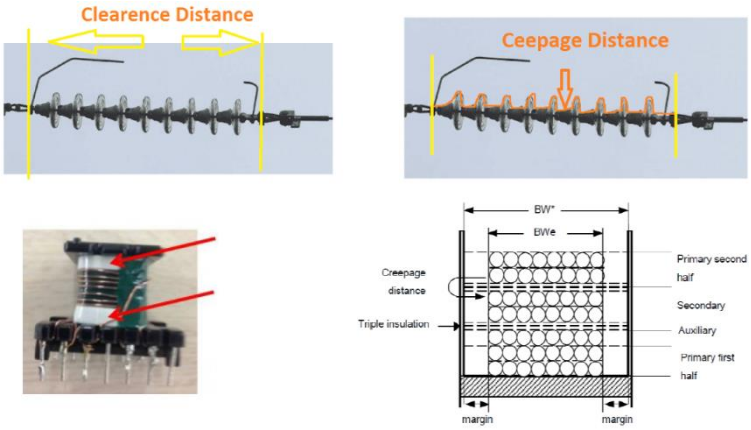


The inductance factor for a gapped core is dominated by the reluctance of the gap: A_L = u0*Ae/lg.
u0 = 4*pi*10^-2 H/m, and **Ae is in m^2, not mm^2**.
However, this is not a good calculation.

The next step is to compute the primary side conductor size. First of all we need to understand which is the window breath. The height h is the direction perpendicular to the breath. Furthermore, also **creepage** must be considered.

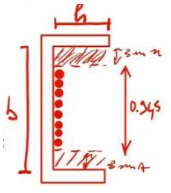
Creepage and clearance

The clearance is the shortest distance in air between two conductors. The creepage is the shortest distance between conductors along the surface of the insulating material, so the creepage is always larger than the clearance. In general, in a transformer we need to respect some minimal creepage depending on safety regulations.



Clearance is the shortest distance in the air between two conductors.
Creepage is the shortest distance to another conductor along the surface of the insulating material.

In our case, we have to consider the creepage of 3mm on the two sides, as in the image. Hence the available breath is not the total breath, but it is reduces.



As for the Irms, it has been already computed. For the selection of the copper size, it is based on the maximum current density we might allow in the copper wire.

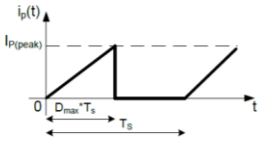
As usual, we stick to 500 A/cm^2. According to computations, we pick AWG#18.

Step 11: calculate the primary side conductor size

From Step 8, window breadth, b = 1.545 cm, and height, h = 0.36 cm.
A creepage allowance of 0.3 cm is necessary at each end of the windings.
Winding width is 1.545 minus (2 x 0.3) = 0.945 cm.

$$I_{P(rms)} = I_{P(peak)} \cdot \sqrt{\frac{D}{3}} = 4.5 \text{ A}$$

(worst case: $D=D_{max} \approx 0.5$)



- Primary conductor area for 500 A/cm^2 = 0.009 cm^2
  An AWG #18 wire would fit, but its diameter (0.1 cm) is about 4 D_{pen} !
- Possible solutions to mitigate proximity effects: Litz wire, multiple strands, foil.
- We implement the primary with a copper strip (foil) **0.945 cm wide** and **0.0127 cm thick** (5 mil), **8 turns**, spiral wound.
- Eight layers, including 0.005 cm (2 mil) low voltage insulation between layers results in a total winding height of 8 x 0.0177 = 0.141 cm.
- Eight turns with mean length/turn = 4.9 cm results in a total winding length of 39.2 cm.

As a conclusion, we can use 8 turns of AWG#18 wire to build the primary winding. In our case, the diameter of the wire is 1mm, so 8 turn occupy 8mm, which is ok also considering the creepage.
Again, this is not a good design because we are working at hundreds of kHz, and the penetration depth due to skin effects (proximity effects(?)) is, in AWG#18, 1/4th of the diameter, so most of the current will concentrate in the internal wall of the wire, so we are wasting most of the copper due to proximity effect. It is worse than the skin effect, because we are inside a magnetic core, so we have to consider proximity and current is not distributed all along the surface, but just in the internal wall of the coper wire (mezzaluna a sinistra del filo di rame).

To mitigate the proximity effects we can use different wires, or a copper foil. The height of the copper foil will be equal to the available breath, 0.945 cm, and its thickness is chose so that the current density is 500 A/cm^2. This is done with a thickness of 0.0127cm. The problem is copper foils are available in discrete thicknesses, we cannot cut them.

How many layers of foil do we need?
We have to go for a spiral winding where one layer corresponds to one turn. Since we need 8 turns, we have a spiral wound with 8 layers. But if we have 8 layers we might have proximity. In reality the thickness of the foil is comparable with the penetration depth, so even if we have several layers, the proximity effect is much lower than using a single layer of 8 turns of copper wire.

Secondary side conductor size
The first thing to do is to compute the rms value of the current that flows in the secondary side, and this can be done with the relationship in the following image, where Is(peak) is the peak value of the secondary current. In the worst case the maximum D = 0.5, so the number under the square root is 1/6.

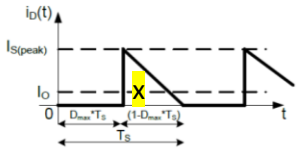
The peak value of the secondary current can be easily computed since the average current that flows in the secondary side is equal to the output current. This means that the area of the triangle x divided by the duration of the period must be equal to the output current, which in our case is 10A → (Iph*o,5Ts)/2Ts = Io.

The Io comes from the fact that the average secondary side current must be equal to the output current..
The area of the cross-section of the copper wire should be selected in such a way that the current density is typically around 500 A/cm^2 → AWG#12.

Step 12: calculate the secondary side conductor size

$$I_{S(rms)} = I_{S(peak)} \cdot \sqrt{\frac{1}{3} \cdot \frac{T_{OFF}}{T_S}} \approx \frac{I_O \cdot 4}{\sqrt{6}} \approx 17\,A$$

(worst case: $D=D_{max} \approx 0.5$)



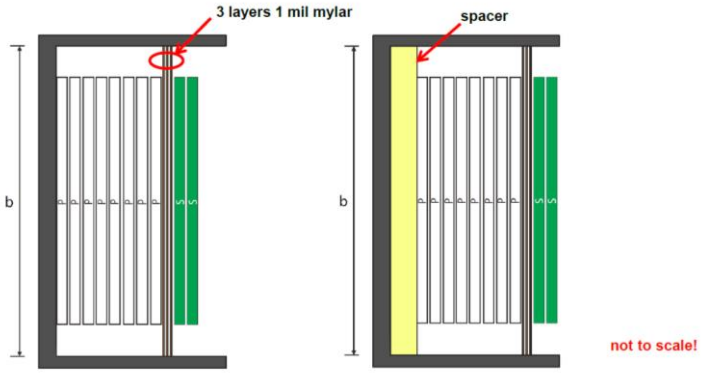
- Secondary conductor area for 500 A/cm^2 =0.034 cm^2
  An AWG #12 wire would fit, but its diameter (0.213 cm) is about 9 D_{pen} !
- Therefore, we implement the secondary with a copper strip (foil) **0.945 cm wide** and **0.0406 cm thick** (16 mil), **2 turns**, spiral wound.
- Two layers, including 0.005 cm (2 mil) low voltage insulation between layers results in a total winding height of 2 x 0.0456 = 0.09 cm.
- Two turns with mean length/turn = 4.9 cm results in a total winding length of 9.8 cm.

However, there is a problem. The diameter of AWG#12 copper wire is 0.2cm, which is about 9 times the penetration depth. We want to avoid this because we are wasting most of the area of the copper wire due to the skin and proximity effects.
We can apply the same considerations done for the primary side, we can implement the secondary side using copper foil, which takes the advantage of being much thinner. The section of the copper foil is a rectangle with an A equal to the available breath, 0.9cm and a width such that the area of the cross-section produces the needed current density of 500 A/cm^2.
We need two turns, so we will have a spiral with two layers because in a copper foil one turn is one layer.

The copper foil must be isolated, we have to consider an insulator layer between a layer and the adjacent one, and this increases the height of the winding. The insulating layer has a thickness of 2 mil (50um).
In the end we get a total height of 0.9cm. We have to check that the sum of the height of the primary side winding and of the secondary side winding is less than the height of the bobbin, otherwise we cannot allocate the copper windings inside the copper.
The other important number is the total length of the spiral winding, which can be calculated by multiplying the mean length per turn (MLT), which is provided by the datasheet of the core.

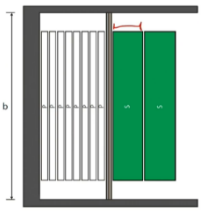
Transformer cross-section

Step 13: winding arrangement



- NOTE: winding build = 0.09 + 0.141+ 3·0.00025 = 0.238 cm < window height.
- If the windings do not occupy the entire available core window area, space the windings away from the centerleg gap to mitigate fringing field effects (see top right figure).

Once we have selected the type of copper conductor (in this case the foil) and the number of layers in the primary and secondary, we can understand the winding arrangement.



The white part of the transformer represents the primary side, the green on the secondary side. Between the primary and secondary side it is always requested to have an insulator layer, whose thickness depends on the safety requirements. The real scaled image of the transformer should be something like aside.

At this point we have to check that the total winding height, which is the sum of the height of the primary side, plus the sum of the height of the secondary plus the height of the insulator, is more or less the height of the bobbin, because we have to confine the windings inside the bobbin. The total height of the bobbin is found in the datasheet of the magnetic core.

There is a problem coming from the gap. In fact, the flyback converter has to store energy, so we need a gap in the core, and the problem with the gap is the fringing field. The fringing field interacts with the copper conductor and it produces eddy current losses inside the copper at HF, increasing the power loss. There are some approaches to mitigate the fringing field, like the distributed discrete gaps, where we have multiple gaps with smaller thickness. However, they cost a bit too much.

In principle, to avoid the fringing flux effect we should place the copper foil or the wires at a certain distance from the gap, in general a factor 3 or 4, that is usually 1.5mm. Thus the fringing flux is mitigated. If we are still not satisfied, we can include a spacer to further increase the total thickness.

If we introduce the spacer, we have to do two checks:

1. The total winding build plus the spacer must be smaller than the bobbin height.
2. With the spacer we are increasing the leakage inductance, so moving the winding away from the core is increasing it, because we are increasing the magnetic energy that is stored.

WINDING LOSSES

This is one of the last steps, and to calculate them we need to compute the winding resistances.

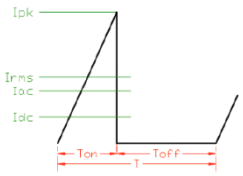
Primary side winding resistance

Step 14: primary side winding resistance

- Split the dc and ac current components.

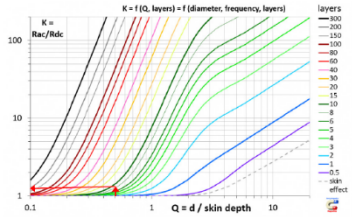
$$I_{P,dc} = \frac{I_{P(peak)}}{4} = 2.75A \quad (\text{worst case, } D_{max} = 0.5)$$

$$I_{P,ac} = \sqrt{I_{P(rms)}^2 - I_{P,dc}^2} = 3.55\,A$$



- **Primary side dc resistance:**

$$R_{P,dc} = \rho\frac{l}{A} = 2.3\cdot10^{-6}\cdot\frac{39.2}{0.012} = 7.5\ m\Omega$$



- **Primary side ac resistance:**

D_{PEN} = 0.024 cm @100 kHz.

Foil thickness = 0.0125 cm → Q =0.0125/0.024 = 0.52.

Entering Dowell's curve with Q =0.52 and 8 layers → $R_{ac}/R_{dc} \approx 1.3$

$$R_{P,ac} = R_{P,dc}\cdot 1.3 \simeq 10\ m\Omega$$

Note: for the sake of simplicity we are assuming that most of the ac current is concentrated in the first harmonic (check!!).

In the windings it flows a triangular current, it is not a DC current. It has a peak value and a rms value. This current must be known to understand the copper resistance, which comes from two contributions.

The first one is the DC resistance, which is related to the DC dissipation. In the worst case, i.e. the base of the triangle is 0.5Ts and the period is Ts, the average value of current is the area divided by the switching period. Hence the DC value of the current is the peak value in the primary divided by 4 (a factor 2 comes from the base of the triangle and the other factor comes from the area of the triangle formula).

As for the AC component of the current, it can be computed taking the square of the rms value of the current and subtracting from it the square of the DC current.
The rms value of the AC current is the summation of the squared rms values of the first component harmonic, second, third and so on.
Ip,ac will be used to compute the power dissipation in the primary winding associated to the AC current.

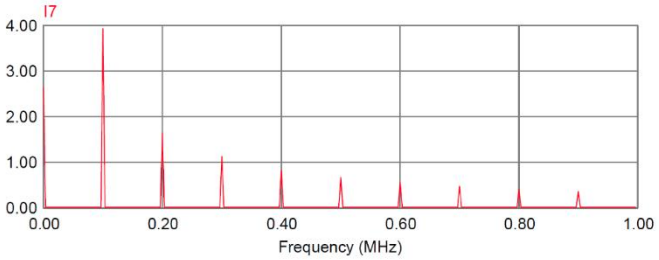
The DC resistance is the resistivity of the copper times the length of the wire (or copper foil) divided by the cross-section of the copper foil.

The AC resistance is computed considering the proximity effects. We have to use the Dowell's curve; the first thing to do is to compute the penetration depth at the switching frequency, which is 7.5cm/sqrt(f).
The foil thickness is 5 mil (0.0125 cm) and we can compute Q (ratio between thickness of the copper foil and penetration depth). Since Q is smaller than 1 we don't expect a strong proximity effect.
So we take the plot on the right, we enter in the computed Q, we go vertically until we intercept the curve corresponding to 8 layers and we get the Rac/Rdc ratio.

If instead of using 8 layers of copper foil we had used just one layer with 8 turns of round wire, in this case the round wire is AWG#18 and we would end up with a much larger Rac/Rdc ratio. This is a clear evidence that 8 layers of copper foil are much better than a single layer of copper wire.

However, in our calculation there is an issue, because the penetration depth is computed at the fsw, that is 100 kHz, but this is not correct. In fact, the waveform is not a sinusoidal waveform at 100 kHz, that is the assumption to build the Dowell's curves, but here the current is triangular, so we don't have a single harmonic, but multiple harmonics. With the calculation we did it is like considering only the first harmonic.
This is acceptable if most of the power of the triangular waveform is concentrated in the first harmonic, but this is not the case.

Primary side current spectrum in the worst condition



@ Vin = V_{INmin} = 22V, D = D_{max} = 0.5

Note: $THD = \frac{\sqrt{I_{ac}^2 - I_1^2}}{I_1} \approx 0.6$ → May want to check the effect of high-order harmonics

As expected, there is a DC component of 2.75, a first harmonic of 4 A (on the y axis we have the amplitude) and second harmonic at a 200 kHz of 1.8 A and so on.

In this situation, in principle we should consider the Rac/Rdc for each harmonic component and getting different Q.

The simplification according to which all the power of the signal is in the first harmonic can be used when the total harmonic distortion of the signal is small (the total harmonic distortion is the summation of the square of the rms values of the harmonics divided by the rms value of the first harmonic). Ideally it should be 0 if we have a single tone.

Secondary side winding resistance

Step 15: secondary side winding resistance

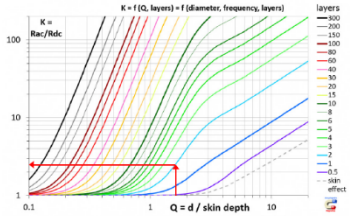
- Split the dc and ac current components.

$I_{S,dc} = I_O = 10A$

$I_{S,ac} = \sqrt{I_{S(rms)}^2 - I_{S,dc}^2} = 13.7\,A$ (worst case, D = 0.5)

- **Secondary side dc resistance:**

$$R_{S,dc} = \rho\frac{l}{A} = 2.3\cdot 10^{-6}\cdot\frac{9.8}{0.038} \approx 0.6\ m\Omega$$



- **Secondary side ac resistance:**

D_{PEN} = 0.024 cm @100 kHz.

Foil thickness = 0.0406 cm → Q =0.0406/0.024 = 1.7

Entering Dowell's curve with Q = 1.7 and 2 layers → $R_{ac}/R_{dc} \approx 2.5$

$$R_{S,ac} = R_{S,dc}\cdot 2.5 = 1.5\ m\Omega$$

Note: for the sake of simplicity we are assuming that most of the ac current is concentrated in the first harmonic (check!).

It is the same calculation than in the primary side. Also in this case we need to check the second and third harmonics.

Winding losses

Step 16: calculate the winding loss

- **Primary side loss**

$P_{P,dc} = I_{P,dc}^2\cdot R_{P,dc} = 56\ mW$

$P_{P,ac} = I_{P,ac}^2\cdot R_{P,ac} = 126\ mW$

$P_{P,tot} = P_{P,dc} + P_{P,ac} = 182\ mW$

- **Secondary side loss**

$P_{S,dc} = I_{S,dc}^2\cdot R_{S,dc} = 60\ mW$

$P_{S,ac} = I_{S,ac}^2\cdot R_{S,ac} = 280\ mW$

$P_{S,tot} = P_{S,dc} + P_{S,ac} = 340\ mW$

- **Total winding loss**

$P_W = P_{P,tot} + P_{S,tot} \approx 522\ mW$

TRANSFORMER LOSS

The total transformer loss is the sum of the winding loss (that is copper loss) and core loss.

Step 17: calculate the transformer loss and temperature rise

- **Winding loss**

$P_W = P_{P,tot} + P_{S,tot} \approx 520\ mW$

Might be underestimated!
Check the high-order harmonics

- **Core loss**

$P_C = V_C \cdot P_C' \approx 430\ mW$

- **Total loss**

$P_{TOT} = P_C + P_W = 950\ mW < 2W\ target\ loss$

- **Temperature rise**

$\Delta T = P_{TOT} \cdot R_\theta \approx 28\,°C$ ☺

Using the empirical formula $R_\theta = 53 \cdot (V_e)^{-0.54}$ [°C / W], where V_e is in cm^3

We notice that the copper loss is almost comparable with the core loss. This is not a coincidence. In fact, if the design is a good design, we might end up in a situation where the copper loss are equal to the core loss, because it is a situation that optimizes the total transformer losses.

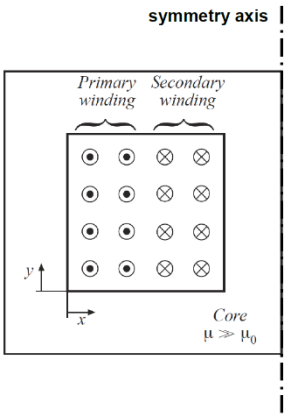
Losses are important because transformer loss reduces the efficiency of the converter, and furthermore the power dissipated by the transformer makes the temperature of the internal region of the transformer to increase. We can use an empirical formula to compute the thermal resistance of the transformer and get the heat dissipation. **The delta of increase with respect to the ambient temperature is equal to the total power dissipated inside the core multiplied by the thermal resistance**.

Moreover, often the datasheet provides a thermal resistance that is higher than the one we can get from theoretical calculations, so better to use the highest one.

In principle the design of the transformer stops here, but we might do something to improve the behaviour of the transformer. We start by consider a simple technique to reduce the proximity effect and the leakage inductance, which is the winding interleaving.

WINDING INTERLEAVING

We start by considering a simple transformer with a primary and secondary sides having the same number of turns, so turn ratio is 1, so the current is the same on primary and secondary sides. The starting point is the one aside.



A simple two-winding transformer example: core and winding geometry

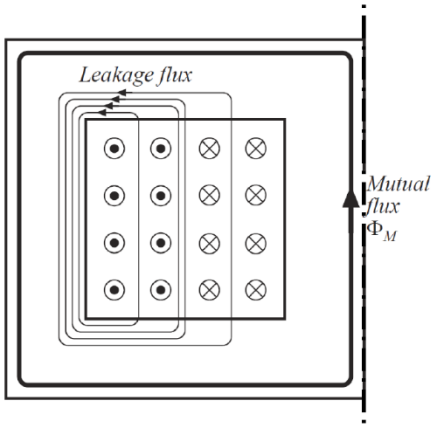
Each turn carries net current $i(t)$ in direction shown

There is a bigger flux, the mutual flux, that links the primary and secondary sides, but also flux lines that are linking into the window which generate the leakage flux, which is not completely linking the primary and secondary windings.

Mutual flux Φ_M is large and is mostly confined to the core

Leakage flux is present, which does not completely link both windings

Because of symmetry of winding geometry, leakage flux runs approximately vertically through the windings

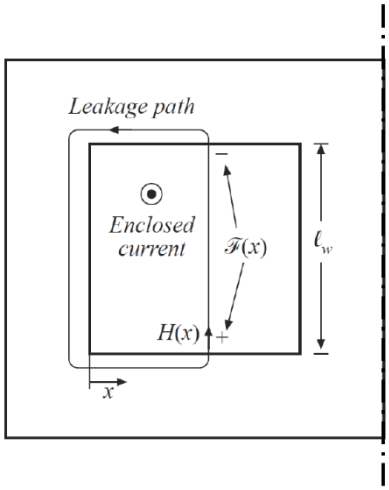


Due to the symmetry of the core and the symmetry of the windings we can say that the magnetic field inside the window is parallel to the winding, in the same direction of the flux lines (bottom to top, vertically).
We can calculate the magnetic field between the first layer of the primary and the second of the primary, between the first of the primary and the first of the secondary and so on using the Ampere's law.

Ampere's law, for the closed path taken by the leakage flux line illustrated:

$$\text{Enclosed current} = \mathcal{F}(x) = H(x)\ell_w$$

(note that MMF around core is small compared to MMF through the air inside the winding, because of high permeability of core)

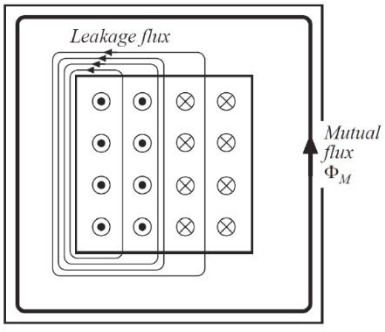


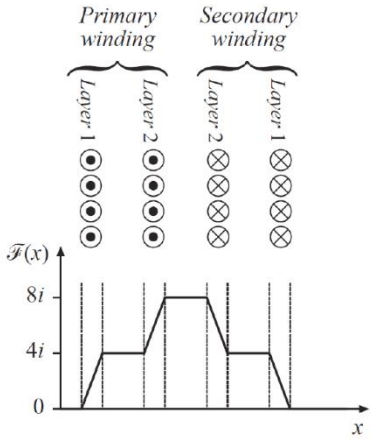
We identify an amperian loop and then sum(H*l) = i. Since the core is made by a material having a large permeability, the H field in the core tends to be 0, so we basically have no field in the core and we can write it just in the window region.

Let's apply this concept to the previous transformer and we want to plot the magnetomotive force as a function of x, the internal position in the window.

MMF $\mathcal{F}(x)$ across the core window, as a function of position x

$$\text{Enclosed current} = \mathcal{F}(x) = H(x)\ell_w$$

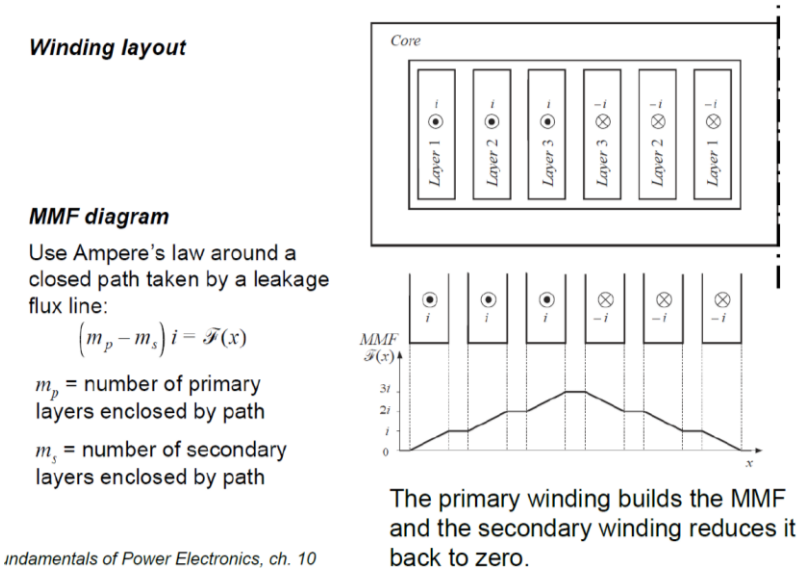




The magnetomotive force starts from 0 because there is no magnetic field in the core but, as soon as we slide the amperian loop inside the window, we have a magnetomotive force that is the number of copper wires multiplied by the current that flows in each winding. So between the first and second layer of the primary side we have 4*i.
If we move the amperian loop, current in the secondary is moving in the opposite direction, so magnetomotive force is reduced.

In the end **the flow of current in a transformer is such to compensate the magnetomotive force**, that is something expected because there is no gap in a transformer.

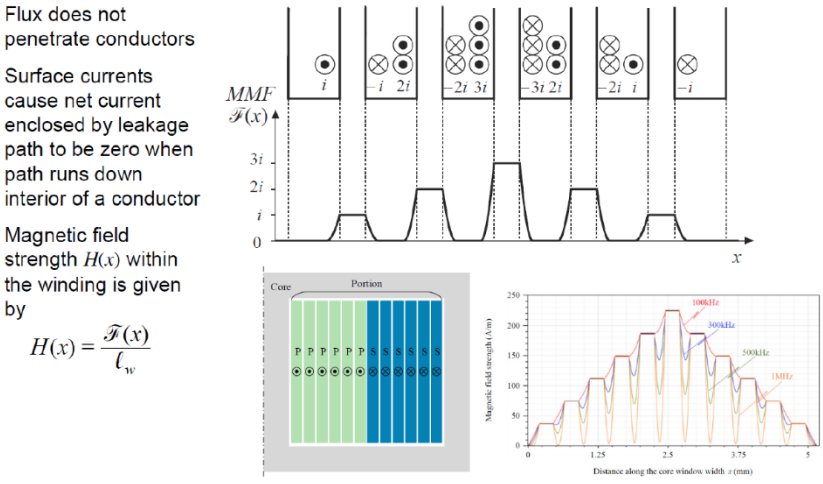
Instead, if the transformer is using a spiral winding we have a situation as below.



The analysis carried out so far was done at low frequency, so with no current concentrated towards the surfaces of the copper foil.

Increase in frequency

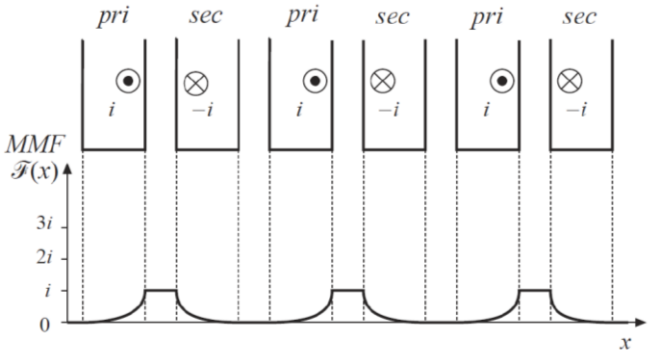
If we increase the frequency of the current flowing in a transformer, we are basically inducing eddy currents in the copper foils and the eddy currents prevent the penetration of the magnetic field inside the copper and the current is concentrating toward the surfaces of the copper foils.



This is a problem because the current density in the internal layers of copper can be quite large, with the current density that reaches high peak values.

Is there a way we can arrange the windings to avoid such huge build up of magnetic field or magnetomotive force inside the transformer? Yes, interleaving.

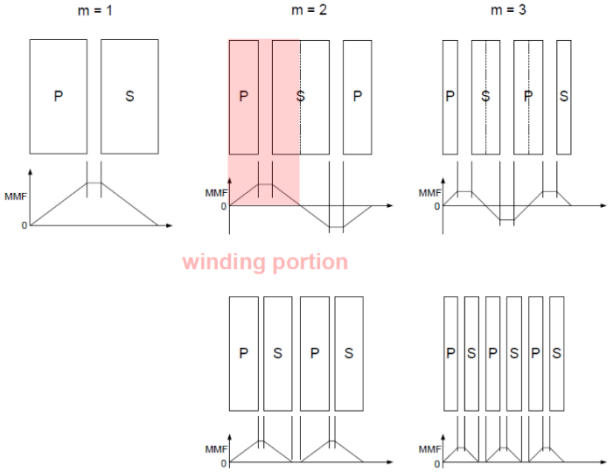
INTERLEAVING



- Reduce winding losses by reducing magnetic field (or equivently the mmf) seen by conductors in winding.
- Beware – increased inter-winding capacitance!

We interleave the primary and secondary sides. Now we are not building up the magnetomotive force, it increases but then it decreases immediately.
The price to pay is that we are increasing the inter-winding capacitance, which can be a problem in HF applications.

Example of interleaving design

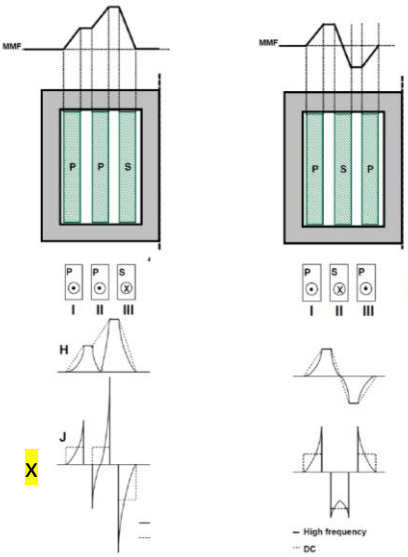


- Field intensity H at pri-sec interface is reduced in proportion to number of winding "portions," *m*

We can think of splitting the primary and secondary sides in two parts and interleaving them.

The reduction we can achieve depends on the winding portion, which is the part of the winding that is between two points where the magnetomotive force is zero. If we have two winding portions we are reducing the magnetomotive force by a factor of 2. Three portions → factor 3 with respect to the non-interleaved.

In the next image there is another example showing the distribution of the magnetic field strength in the window but also of the current density. On the left it is not interleaved, we notice that current concentrates in the surfaces and it builds up (x).



Eddy current loss proportional to H^2

⇒ loss reduced by a factor m^2

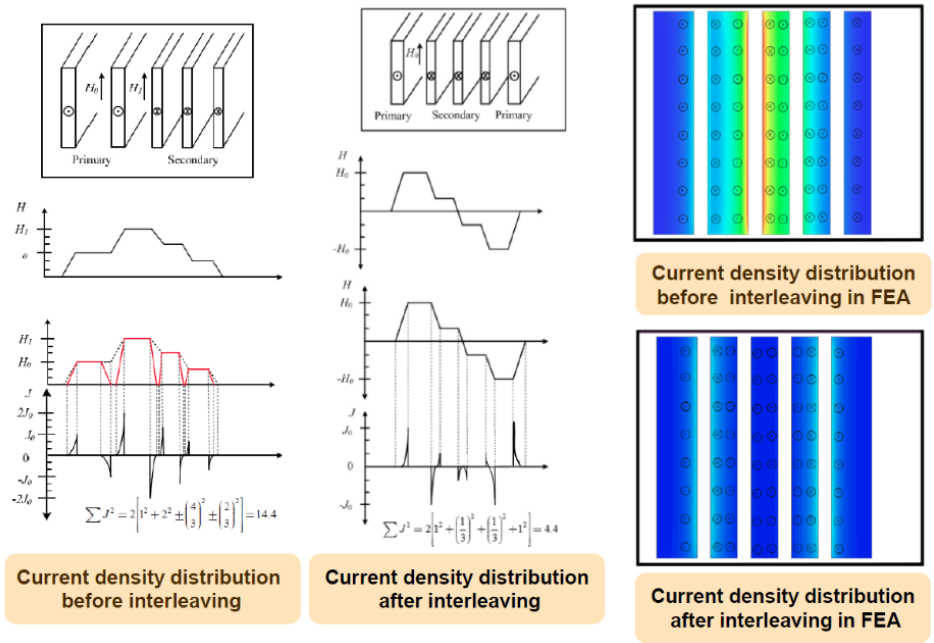
⇒ m=2 is the most practical for power transformers.

⇒ not for flyback converters! Currents only flow in the primary or secondary windings

In a flyback converter, the current never flows simultaneously in the primary and secondary sides, which is however fundamental to cancel the magnetomotive force. So the impact of interleaving in a flyback transformer is much less than in a real transformer like in a forward converter.
However, interleaving is still suggested in the flyback converter because it allows to reduce the leakage inductance.

We are interested in losses, which are proportional to the square of the current, and eddy current are proportional to the square of H, so in principle in splitting the winding in two portions we have an advantage m^2 in the losses, where m is the number of splits.

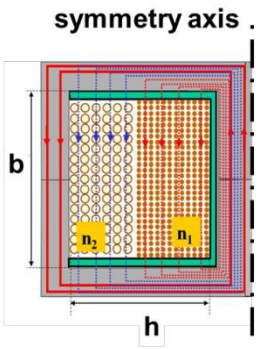
Example 2



LEAKAGE INDUCTANCE

We start by considering a transformer without interleaving. The red flux is the mutual flux that links the primary and the secondary and goes around the magnetic core, but some flux lines are in the window, and not linking all the primary windings and secondary windings, so we have some leakage flux that introduces a leakage inductance.

- In a typical transformer, the flux generated by ac current excitation in the primary winding follows the magnetic circuit and links with the secondary winding.
- Some flux will leak from the core and return to the air, winding layers, and insulator layers without linking the secondary windings.
- The partial coupling of the flux results in leakage inductance.



For symmetry reasons we can consider the leakage flux to go vertically in the window so that we can easily compute magnetomotive force inside the window.
We can compute the magnetomotive force (MMF) inside the window applying the same methodology seen before, considering amperian loops. The result is in the slide below, where b is the breath.

- Within the winding area, the leakage flux is approximately parallel to the winding interface.

  Apply Ampere's law along the dashed path:

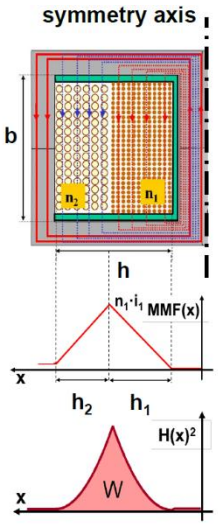
$$MMF(x) = H_{Fe} \cdot \ell_{Fe} + H_W(x) \cdot b \approx H_W(x) \cdot b \Rightarrow$$

$$\Rightarrow \boxed{H_W(x) = \frac{MMF(x)}{b}}$$

- The total leakage inductance referred to the primary side can be calculated by the energy stored in the magnetic field, i.e.,

$$E_W = \iiint_{V_W} \frac{\mu_0 \cdot H_W^2}{2} dV = \frac{1}{2} \cdot L_{lk1} \cdot i_1^2$$

energy approach

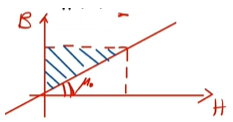


The MMF is equal to the current that enters the loop, so if the amperian loop is increased on the x axis, MMF increases linearly and then decreases because current direction is the opposite.

Then we want to compute the energy stored in the leakage inductance, supposing the leakage inductance in the primary side. It must be equal to the magnetic field energy store in the window. So if we can compute the energy stored in the window, since this energy is associated with the leakage flux, this energy must be the energy stored in the leakage inductance.

So we can write the balance x; the magnetic energy stored in the window must be equal to the magnetic energy stored in the leakage inductance referred to the primary side.
To compute the energy stored in the window we have firstly to understand what the specific energy is (energy per unit volume in the window).

Inside the window we have two things, air and copper, and both are not magnetic materials. Hence the B-H diagram inside the window is a simple straight line having a slope equal to the permeability of the air.



The specific energy stored in the window is the blue area in the image aside, which is ½ * B * H, with B = u0*H, where H is the H inside the window (Hw). This must be integrated over the volume of the window.

If we can compute the energy with the volume integral, we are then able to compute the leakage inductance.

NB: energy stored in the window is proportional to Hw^2, and the H field is proportional to the magnetomotive force, so it is like computing the area of the orange waveform.

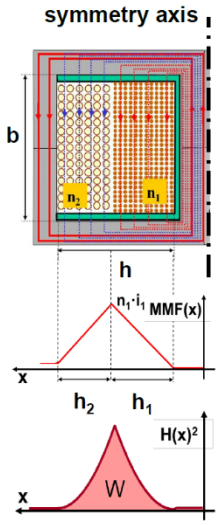
To solve the integral, we can consider the transformer in cylindrical geometry (bobbin) and the two dimensions are lumped in the term MLT (mean length per turn, that is the average length of the winding). What is left is the integral in the x dimension.

$$E_W = \iiint_{V_W} \frac{\mu_0 \cdot H_W^{\,2}}{2} dV = \frac{b \cdot MLT \cdot \mu_0}{2} \cdot \int_0^{x_w} H_W(x)^2 dx$$

where MLT = mean length per turn of the windings

- Considering a triangular MMF profile:

$$L_{lk1} = \frac{\mu_0 \cdot b \cdot MLT \cdot \int_0^h H_W(x)^2 dx}{i_1^2} = \frac{\mu_0 \cdot n_1^2 \cdot MLT \cdot (h_1 + h_2)}{3 \cdot b}$$



In the result, h1 and h2 are the heights of the primary and secondary windows.

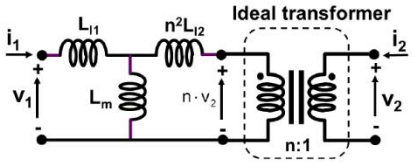
Important notes

The leakage inductance is proportional to the window breath b, and the larger b the better, because we are decreasing Llk. Furthermore, it is proportional to the square of the primary side number of turns and linearly on MLT and h1 + h2.

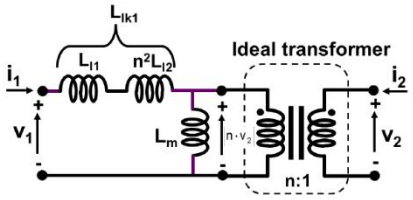
LEAKAGE INDUCTANCE REFERRED TO THE PRIMARY SIDE

We have to start from the initial model of the transformer, where we added a leakage inductance on the secondary side. But the leakage inductance on the secondary side can be moved to the primary side multiplying it (Ll2) by the square of the turn ratio.

Then, if the leakage inductance that is moved to the primary side is much smaller than the magnetizing inductance, we can move the transferred inductance before the magnetizing one, as in the image below.



- If $n_2L_{l2} << L_m$



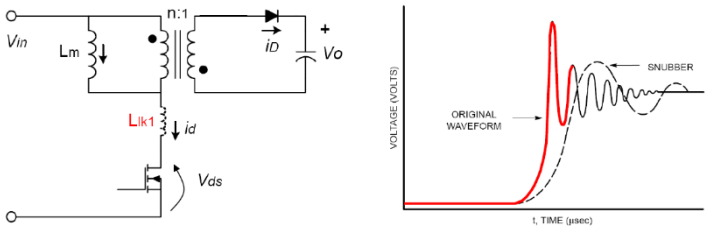
- L_{lk1} is (approximately) equal to the measured inductance of the primary winding with the secondary winding shorted.

So the total leakage current referred to the primary side is the leakage inductance of the primary side plus the moved inductance of the secondary side.
The previous formula for Llk1 can also be used for Ll2 simply changing n1 with n2. **The ratio between Ll2 and Ll1 is the square of the turn ratio**, as expected.

LEAKAGE INDUCTANCE IN THE FLYBACK CONVERTER

- Leakage inductance overvoltage

When the power MOSFET is turned off, there is a high voltage spike on the drain due to the transformer leakage inductance. This excessive voltage on the MOSFET may lead to an avalanche breakdown and eventually failure of power MOSFET. Therefore, it is necessary to use an additional network to clamp the voltage.



In the image it is present the total leakage inductance referred to the primary side. We want to understand what happens when we turn off the switch. Indeed, there are 2 main impacts of the Llk on the normal operation of the flyback converter:

1. If we don't do anything, every time we open the switch the switch may enter the breakdown (overvoltage problem).
2. The presence of Llk will steal energy from the magnetic core and so not all the energy stored in the magnetic core during the Ton is transferred to the output during the Toff.

Overvoltage

Let's assume that the switch is initially closed and the current in the primary side has reached a peak value at the end of Ton = DTs. At this point we open the switch. The primary current is also flowing through the leakage inductance, so we have energy stored in the leakage inductance and when we open the switch, energy cannot disappear. If the switch is ideal (infinite breakdown), there is still a parasitic capacitance between drain and source, Cds. When we open the switch, the Cds initially is discharged because previously the transistor was on. However, there is energy stored in the leakage inductance.

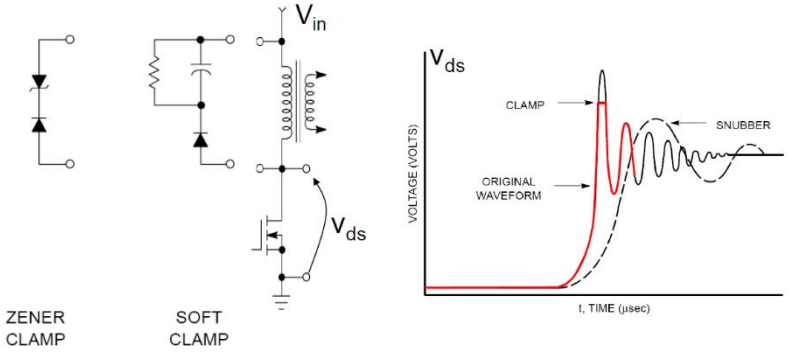
If we measure the v_ds(t) of the switch we see an oscillatory response as in the previous image. Resonances are damped because we also have parasitic resistances along the switch path.
The Lm is not entering into play in this transient because as long as there is energy in the magnetic core, the voltage between its nodes is clamped to Vo*n1/n2.

At the end of the transient, the steady state value of Vds is Vin + Vo*n1/n2 (inductor is a short circuit and capacitor is an open circuit).
The problem is that the mosfet is not ideal, it has a limited breakdown voltage, and if this voltage is smaller than the peak voltage reached due to resonances we have the breakdown. So we need to check that the maximum voltage reached during the transient is smaller than the breakdown voltage BVdss.

There are several possibilities to reduce the overvoltage. One possibility, that is not suggested, is to introduce a snubber network increasing the capacitance between the drain and the source, and we get the dashed curve. The typical solution is the introduction of a clamp network.

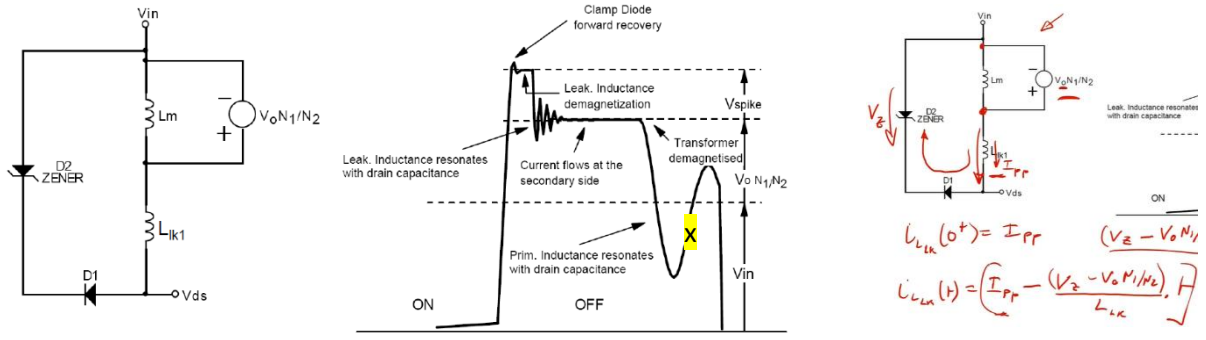
Clamp networks

Zener clamp or soft clamp, made of a parallel RC instead of the Zener diode.
The Zener diode is connected between the drain and the input voltage, so in parallel to the primary side.
There is also a standard diode back-to-back with the standard diode because if we turn on the mosfet and we don't have the standard diode we have a current across the Zener (forward bias), so a short to ground.



When the switch is turned off, the voltage increases but, as soon as Vds is larger than Vin + Vzener, the Zener diode starts to be conductive and clamps the voltage. Of course, Vin + Vz < BVdss.

We can draw and equivalent circuit during the avalanche, i.e. when the Zener diode is active. The magnetizing inductance is shorted by a voltage generator because as far as there is energy in the magnetic core the secondary side is pushing current in the diode, so we are reflecting a Vo on the secondary side as Vo*n1/n2.



If we assume that we are at the beginning of the transient and the mosfet is off, there is still a downward current in Llk1 that is Ipp (I-peak-primary, that is the current flowing in the Zener diode).
This current goes through the diode loop producing a voltage drop of Vz on the Zener.

The voltage across the Llk1 is Vz – Vo*n1/n2 downward.
If this voltage is positive, so if Vz > Vo*n1/n2, and this must be true, the Llk1 is discharged because we are reducing the initial current Ipp. Since we are applying a constant voltage, current is decreasing linearly.

As for the voltage Vds (right plot), once the switch is opened, it skyrockets up to the value Vin + Vz, then it stays constant for a time t_avalanche in which the Zener diode is avalanching and it is the time needed for the energy stored in the leakage inductance to be removed.
It can be computed from the i_Llk(t) formula putting i_Llk(t) to 0 and solving for t_av.

$$t_{av} = \frac{I_{pp} \cdot L_{lk}}{V_z - V_o N_1/N_2}$$

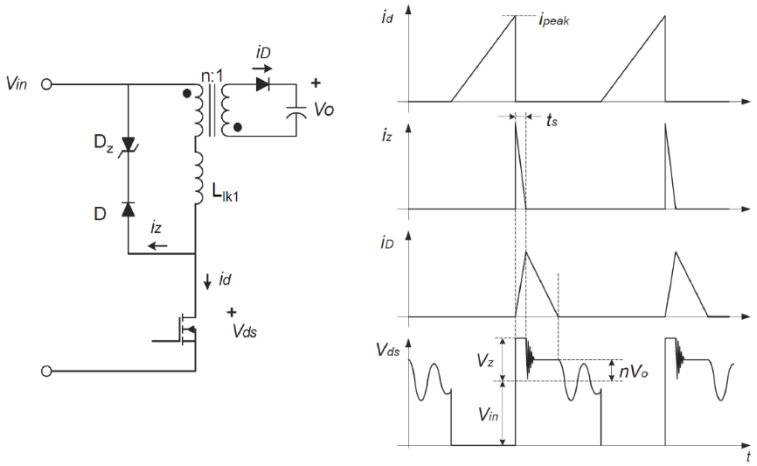
There is another thing to compute, which is the energy dissipated y the Zener diode. As long as the diode is avalanching, the voltage across it is Vz and there is an instantaneous current and so the power is Vz multiplied by the current. We integrate this instantaneous time over the avalanche and we get the total energy dissipated by the Zener during the on-off transition. The result is the following.

$$E = \frac{1}{2} L_{lk} I_{pp}^2 \frac{V_z}{V_z - V_o N_1/N_2}$$

It is not simply the energy stored in the leakage inductance, but there is a correction term. The larger Vz, the closer the total energy dissipated by the Zener to the energy stored in the leakage inductance.
However, if Vz and Vo*N1/N2 (reflected output voltage) are similar, we are dissipating an energy that is larger than the one stored in the leakage inductor. This energy comes from the magnetic core, so we are stealing energy from it. This is a problem because part of the stored energy is dissipated.

Then, once the t_av is over, the current in the leakage inductance reaches 0 and the voltage collapses to Vo*N1/N2. However, due to the parasitic capacitance that is connected between the drain of the switch and ground, we have some resonances with a natural frequency of 1/sqrt(Llk*Cpp).
The voltage across the magnetizing inductance stays equal to Vo*N1/N2 as long as there is energy stored in the magnetic core. When the energy is finished, the voltage between drain and source collapses back to Vin, because there is no more current in the magnetizing inductance and no more voltage across the Lm and no more voltage across Llk and so the final Vds is Vin.
However, we have slow resonances x that occur due to the interaction of the magnetizing inductance and the parasitic diodes (Lm >> Llk1).

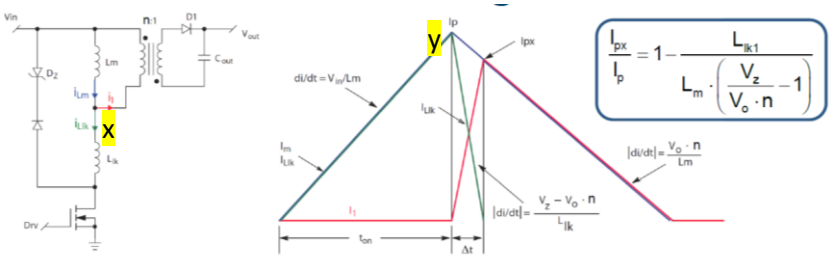
Current waveforms



i_D is the diode current and i_z is the i_leak. In an ideal flyback converter I would expect that when the switch is turned off the ideal diode current Ipp increases instantaneously up to a peak value that is Ipp*N1/N2 and then reduced linearly. i_D is the drain current, that has up to a peak value of Ipp.
This is what we expect in the ideal case. However, due to the leakage inductance and the clamping network, what happens is that the current i_d, so the current in the secondary, doesn't step up vertically instantaneously, but it increases linearly until it reaches a peak value, and the duration from the zero point to the peak is exactly the duration of the current in the Zener diode (t_av).

Due to this transient, the peak current in the secondary side is smaller than the ideal peak current, and this is the consequence of the 'energy stealing', not all the energy stored in the magnetic core is transferred to the output.

Adverse effect of the leakage inductance



- The leakage inductance L_{lk1} does not participate to the primary/secondary energy transfer.
- When the transistor opens, $I_{L_{lk1}}$ is routed to the zener diode transforming the leakage energy into heat.
- L_{lk1} delays the primary to secondary energy transfer (by Δt) until its current has dropped to zero.
- L_{lk1} diverts a substantial amount of the stored energy. To account for it, the primary current will be higher, forcing you to design a larger core transformer.

We can see the adverse effect with a simple KCL at node x. i1 is the current reflected to the secondary side producing i_D.
During Ton i1 is zero, while i_m and i_lk both increase with a slope Vin/Lm. The most interesting point is y, where I turn off the mosfet: the i_lk decreases with a high slope, and at the same time, due to the KCL, i1 is increasing. However, the red line i1 doesn't reach the ideal peak value Ip, but due to the finite slope, the peak value of the diode current is Ipx. The ratio between the real and ideal value is in the blue box. The larger the ratio, the less the energy stored during the Ton is transferred to the secondary side → the larger the Llk1 the worse the loss. Instead, the larger Vz with respect to Vo*N1/N2, the better.

Problems

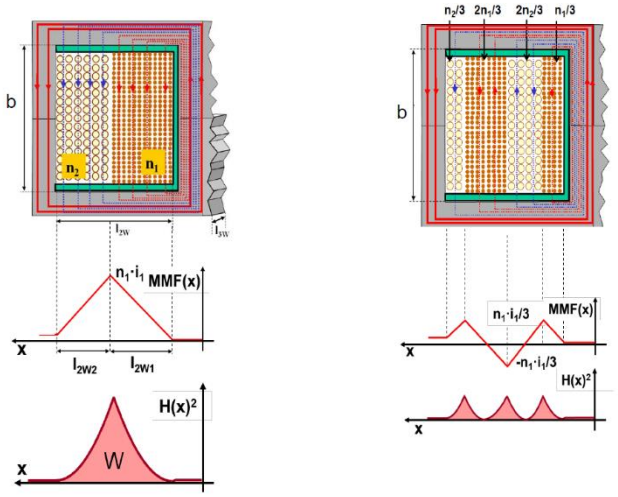
The leakage inductance doesn't participate to the primary/secondary energy transfer, the energy stored in Lm during the Ton is lost, but not only this, during this time period we also have some energy stored in the magnetic core that is stolen and dissipated on the Zener diode. If this happens we have a problem because the flyback converter is working with a limited efficiency.

LEAKAGE INDUCTANCE MINIMIZATION

Leakage inductance is a function of winding geometry, number of turns and separation between primary and secondary.

- Minimize the separation between the primary and main secondary winding(s).
- Interleave the primary and main secondary. It also reduces the proximity effects.
- Select a core with a long and narrow window.

Interleaving effect on leakage inductance



In the image we have an interleaved winding with 3 portions, so the peak magnetomotive force is reduced by a factor of 3, and the energy stored in the window is the sum of the three red areas, much is much smaller than the total area.

- Energy storage in leakage proportional to H^2

  ⇒ each interleaving portion can reduce leakage energy
- Leakage drops in proportion to H^2, i.e. to m^2

  m = number of winding "portions
  - 1 step of interleaving, m = 2

    $L_{LEAK}/4$
  - 2 steps of interleaving, m = 3

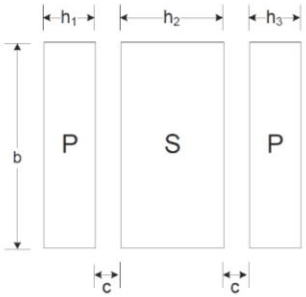
    $L_{LEAK}/9$
- However, extra gaps c between windings limit leakage reduction
  - In practice m=2 ⇒ ~$L_{LEAK}/2$
  - *Beware – increased inter-winding capacitance!*

General formula for Lleak calculation

- General formula to estimate the leakage inductance*

$$L_{leak} = \frac{\mu_0 \cdot N^2 \cdot MLT \cdot (\sum h + 3\sum c)}{3 \cdot b} \cdot \frac{1}{m^2}$$

N = number of turns in the winding to which the leakage is referred
- *MLT* = Mean length per turn of the windings
- Σ*h* = sum of the heights of all winding layers
- Σ*c* = sum of the heights of the spacing gaps between winding layers
- *m* = level of interleaving (number of winding "portions")
- *b* = bobbin winding breadth (width of the winding layer)



c is the spacing between adjacent windings due to the presence of the insulator. We have a factor of /3 for the h and not for the c because the MMF increases linearly in the winding and it stays constant in the insulator layer.

We can also say that Lleak is inversely proportional to the frequency. This because the higher the frequency, the less the magnetic field strength is able to penetrate the copper, so we don't have magnetic field strength in the copper at HF and so when we perform the integration we don't have much magnetic field in the copper, so the integral becomes smaller and Lleak decreases with the increase in frequency.

Leakage inductance reduction – example

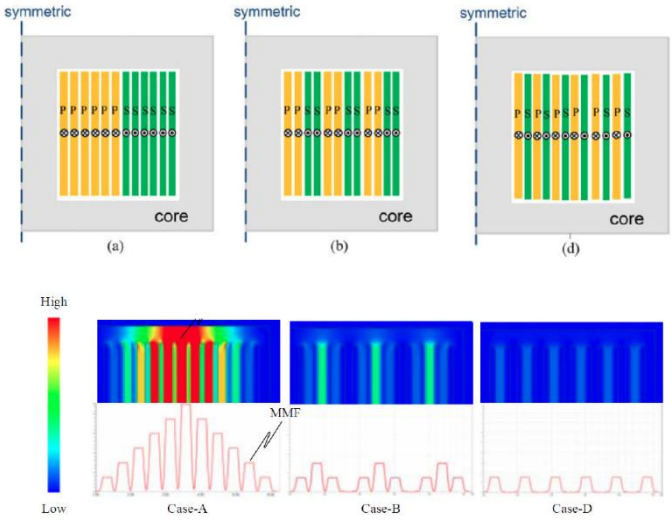


Fig. 4.10 FEA simulations results of *H* field in the window area

INTERLEAVING IN FLYBACK CONVERTERS

In general it seems useless in flyback converters because current either flows in the primary side during the Ton or in the secondary side during Toff, there is no simultaneous flow of current in both the windings. This is true but not exactly.

The working mode of flyback type transformers is clearly different than the one in conventional transformers. The main differences of these magnetic components can be summarized as follows:

- There are two magnetic roles: energy storage when the switch is on and energy transfer when the switch is turned off.
- Primary and secondary currents occur sequentially, *not simultaneously* as in a true transformer. Therefore they do not cancel.
- When current is flowing through one of the windings, there are very important induced currents in the "open" (passive) windings that should be accounted for losses calculation.

→ The effects of the application of interleaving techniques in Flyback transformers are different than in Forward type ones.

Interleaving or not the flyback is still a debated question. We can try to investigate its impact as long as copper loss and leakage inductance are concerned.

Copper loss

No effect in DCM in reducing the HF resistance to reduce copper losses, but in CCM it reduces the copper losses.

Leakage inductance

The concept of Lleak in a flyback makes sense when the mosfet is turned off, and around the turn off transient of the mosfet the current flows simultaneously in the primary and secondary sides.

Since we have simultaneous current and we want to minimize Lleak, it makes sense to interleave.

Copper loss

- Interleaving the windings in flyback converters working in DCM has little effect in reducing HF resistance [1,2].
- Interleaving the windings in flyback converters working in CCM has a stronger effect in in reducing HF resistance (the dynamic MMF must be considered) [3, 4].

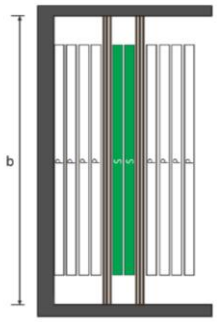
Leakage inductance

- The definition of leakage inductance in flyback transformer makes sense only during the commutation of the main switch, when primary and secondary current *flow simultaneously.*
- Depending on the transformer type, the leakage inductance can be reduced more than twice if interleaving [5].

FLYBACK WINDINGS DESIGN

The original design of the transformer is replaced with a new design which uses the interleaving.

- Interleaving is recommended to reduce the leakage inductance.
- Downside: interwinding capacitance is increased (can be easily calculated)



Interleaved flyback windings

- primary: copper strip 0.945 cm wide and 0.0127 cm thick (5 mil), 8 turns, spiral wound
- secondary: copper strip 0.945 cm wide and 0.0406 cm thick (16 mil), 2 turns, spiral wound
- isolation: 3 layers 1mil mylar

- Leakage inductance estimated using the method reported in slide 85 is 70 nH (primary side)
- Possibly move the windings away from the center leg gap in order to reduce the fringing field loss.

The 8 layers of the primary are split in 4 and 4 and in between we put the secondary side. The drawback in doing this is that we need to introduce another insulation layer which produces a larger height of the windings, so the advantage of interleaving is partially compensated by the increase in height of the windings.

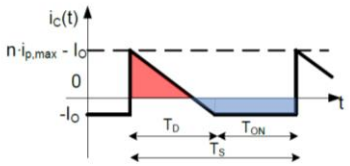
OUTPUT CAPACITOR SELECTION

We have to understand which is the current in the capacitor, which is the diode current shifted down by the output current.

The most important contribution to the ripple in the output capacitor comes from the ESR if we are using an electrolytic capacitor, and so we can pick up the capacitor considering the ESR alone.

Step 18: output capacitor

Neglecting the capacitive ripple:



$$\Delta V_O = ESR \cdot \Delta i_C = ESR \cdot I_{S(peak)} \leq 100mV \quad \Rightarrow \quad ESR < 2.5m\Omega$$

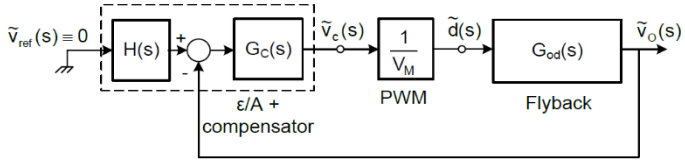
For instance, use 10 Sanyo "10SVP330M" OS-CON capacitors in parallel (C=330µF, ESR=17mΩ, rated ripple current=3.95 A_{rms}).

$$C_{TOT} = 3300\mu F, \; ESR_{TOT} = 1.7m\Omega$$

$$\Delta v_O = \Delta v_C + \Delta v_{ESR} \approx 74mV$$

CONTROLLER DESIGN

Step 19: closing the feedback loop (voltage mode control)



In DCM:

$$G_{od}(s) = \frac{\tilde{v}_O(s)}{\tilde{d}(s)} = G_{od}(0) \frac{1 + \frac{s}{2\pi f_z}}{1 + \frac{s}{2\pi f_p}} \qquad f_z = \frac{1}{2\pi \cdot ESR \cdot C} \cong 28KHz$$

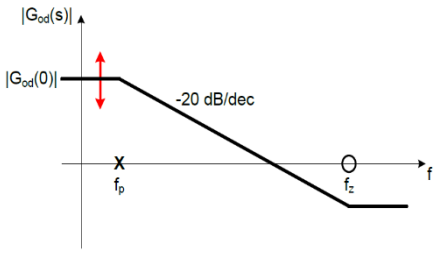
$$G_{od}(0) = \frac{\tilde{v}_O(0)}{\tilde{d}(0)} = V_{IN} \sqrt{\frac{R}{2L_m \cdot f_s}} \qquad f_p = \frac{2}{2\pi \cdot R \cdot C} = \frac{1}{\pi \cdot R \cdot C}$$

We have the classical block model with the difference that instead of the buck model we have to plug in the flyback model, the other blocks are identical.
The flyback is working in DCM and so we are loosing a pole, no more a c.c. pair of poles, and also the RHP zero, so we have just the zero due to the ESR of the output capacitor.

The problem is that God(0) depends on the input voltage, which is not constant.

Note

- $G_{od}(0)$ depends on V_{IN}
- the pole frequency, f_p, is fixed:
  - with I_O = 10 A we have R = 0.5 Ω, giving $f_p \approx 193$ Hz



Zero and pole frequencies are instead fixed.

CONTROL TO OUTPUT T.F.

It is the God multiplied by the t.f. of the PWM. We have to pick the controller we want to use, i.e. the IC integrating the PWM and the driver for the switch.

Control-to-output transfer function

$$G_{OC}(s)=\frac{\tilde{v}_O(s)}{\tilde{v}_C(s)}=G_{od}(s)\cdot G_M(s)=\frac{G_{od}(0)}{V_M}\frac{1+\frac{s}{2\pi f_z}}{1+\frac{s}{2\pi f_p}}$$

For instance, V_M=2.5 V for a UC1524A controller. Therefore:

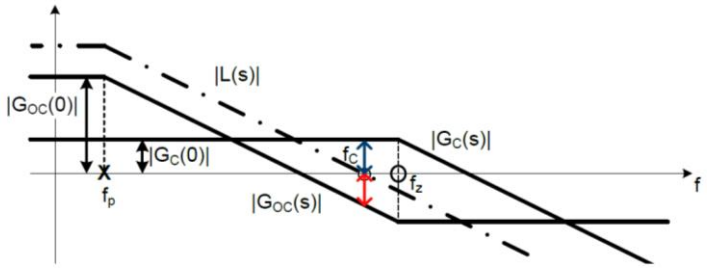
$$G_{OC}(0)=\frac{\tilde{v}_O(0)}{\tilde{v}_C(0)}=\frac{V_{IN}}{V_M}\sqrt{\frac{R}{2L_m\cdot f_s}}\cong 0.2V_{IN}$$

- $G_{OC}(0)\big|_{max}=7.6$ for $V_{IN}=V_{INmax}$

- $G_{OC}(0)\big|_{min}=4.4$ for $V_{IN}=V_{INmin}$

In this case the internal PWM generates a sawtooth waveform with a peak value of 2.5V.

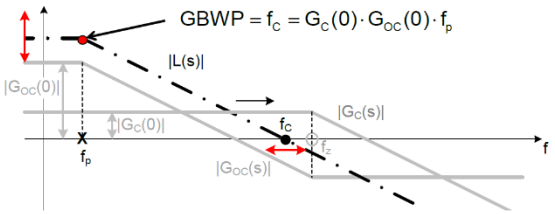
CONTROLLER DESIGN



Simple first-order low pass compensation network

- compensation pole used to cancel ESR zero
- set the crossover frequency (0 dB loop gain) at the desired value ($f_c = f_s/5$ = 20 kHz) by adjusting $G_C(0)$

We start from the God(s) t.f. that we have to compensate in order to obtain a well-behaving loop t.f., i.e. with a constant rolloff of -20dB/dec crossing the 0 dB axis at the desired crossing frequency.
In order to compensate the Goc we are using a simple LP filter t.f. with a gain at 0 Hz and a single pole.



➢ Worst condition for for $G_{OC}(0) = G_{OC}(0)\big|_{max} = 7.6$

➢ Using GBWP

$$f_{C_{MAX}} = G_C(0) \cdot G_{OC}(0)\big|_{max} \cdot f_p \le \frac{f_{SW}}{5} = 20kHz$$

$G_C(0) \le 13.6 = 22.7\,dB$

$G_{OC}(0)\big|_{max} = 7.6 = 17.6\,dB$

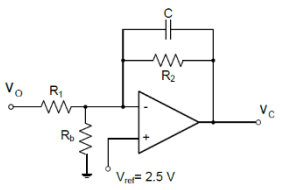
$|L(0)|_{max} = 40.3\,dB$

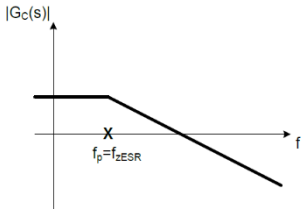
There is just one problem, that is that the starting function to be compensated, the Goc, is not fixed, so we have to identify the worst condition and compensate the Goc in the worst condition.
If Goc(0) moves up and down, the loop t.f. moves up and down accordingly and the crossover frequency left and right. We have never to overcome the target crossover frequency, especially if we are using an aggressive design putting fc at $1/5^{th}$ of the switching frequency.

In the end we get a loop t.f. with a magnitude of 40dB at LF. Is it good enough?

Controller implementation

- Compensation network





$$G_C(0) = \frac{R_2}{R_1} \qquad f_{zESR} = \frac{1}{2\pi R_2 C}$$

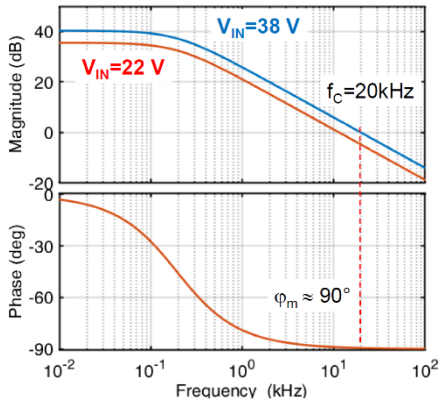
- Use for instance:
  - R_1 = 2.8 kΩ
  - R_b = 2.8 kΩ
  - R_2 = 39 kΩ
  - C = 150 pF

May want to consider a second capacitor in series with R_2

Rb is entering into play only in DC to set the output voltage thanks to the voltage partition with R1. The position of the pole must be such to compensate the zero.

If we are not happy with the result, we can introduce a capacitance in series with R2 introducing a pole in the origin and so a loop gain at LF.

L(s) Bode Diagram



If we decrease Vin the crossing frequency and so the closed loop BW is reduced. Furthermore, the ESR of the capacitor in the compensation network is not stable, it depends on the temperature.

FINAL CHECK

Step 20: check the minimum value of the duty cycle

From $\frac{V_O}{V_{IN}} = D\sqrt{\frac{\eta \cdot R}{2L_m f_S}}$, we get $\boxed{D_{min} = \frac{V_O}{V_{INmax}}\sqrt{\frac{2L_m f_S}{\eta \cdot R}} \approx 29\%}$

What is the effective efficiency?

We started our design by guessing it of 84%, and we can compute it now considering the main contribution to the losses. The diode was dissipating 6W, the transformer 1W, the switch 1W. A total of 8W is dissipated, the output power is 60W and so eta_eff = 86%.

It is slightly bigger than 84% but we don't have to iterate the design for two reasons:

1. Still good value
2. If the effective efficiency we are measuring at the end of the design is larger than the target one and the guess, there is no need to iterate, because it means that the flyback will work, in the worst condition with a Dmax < 0.5%.

However, if the final efficiency is smaller than the target or the guess, then in that case we need to re-iterate the design or we run the risk of working in CCM.

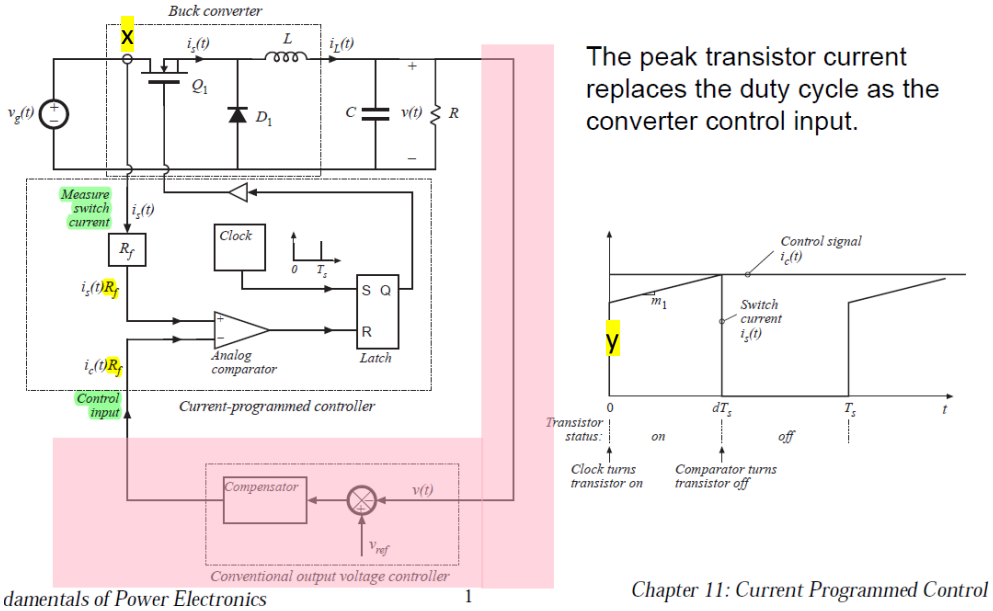
To increase the efficiency, we can do a lot in the diode, which is quite lossy → synchronous rectifier instead of a diode, or we replace the diode with a mosfet so that the Rds,on is smaller.

CURRENT PROGRAMMED CONTROL

It is a control modality that is much more robust than the voltage control and commonly used today. Let's consider a buck converter working in CCM.

In voltage mode control we are measuring the voltage across the capacitor, that is the output voltage of the DC/DC converter, this voltage is compared with a reference, the error is amplified and the amplified error is used to determine the DC at which the converter is operated.
As for the current mode control, we are controlling the current that flows in the inductor and we can do this in several ways. In the following, we won't be controlling the average current but the peak current that flows in the inductor.

Since we want to control the current in the inductor, we need to measure it. In the example the current is measured in point x. It is the same current in the inductor when the switch is on during Ton.



damentals of Power Electronics 1 Chapter 11: Current Programmed Control

This current, called i_s(t), is multiplied by a resistor obtaining the voltage i_s(t)*Rf. This voltage is fed to the non-inverting input of the comparator. The inverting input of the comparator has a control voltage v_c in applied. The control voltage v_c can be expressed as the product of a control current i_c(t)*Rf. In principle the comparator is comparing the voltages, but since Rf is the same, we are comparing the control current and the inductor current.

The shapes of the switch current and control current are reported in the plot on the right. The implementation of the peak current mode control, also called current programmed control, is very simple. We need to use a comparator to compare the i_s and i_c, a clock generator to generate a sequence of pulses at a frequency fs or at a period Ts and a simple set-reset (latch) flip flop.

When the clock generates a pulse, it is applied to the Set input of the flip-flop, so the Q output becomes high. At this point, the Q signal is a digital signal, then we have a gate driver and the mosfet is turned on in the plot, we are moving along the line y. So at t = 0 there is a clock pulse applied to the S producing the turn on of the power mosfet which makes the current to ramp up very quickly up to a point that is the inductor current immediately before the turn on of the switch. In fact, the converter is in CCM, so the current in the inductor is continuous, it never crosses the 0.

Then the mosfet is on and the current increases with a slope m1 = (Vin – Vo)/L. We are also assuming that everything is ideal.
Hence the singal i_s(t)*Rf increases up to when i_s becomes equal to i_c, or the sense signal becomes signal to the control signal, that is the same thing.

At this point, the output of the comparator becomes high and it is applied to the R input of the flip-flop, making Q = 0, so the mosfet is turned off and the current i_s immediately drops to 0, because we are measuring the switch current, not the inductor current. The Q remains zero until the next pulse generated by the clock occurs and it is applied to the S.
By doing this we are controlling the peak value of the current that flows in the inductor. In fact, if the control signal i_c is increased, the current i_s would increase more, so the D is increased so that the peak current in the inductor matches the control current.

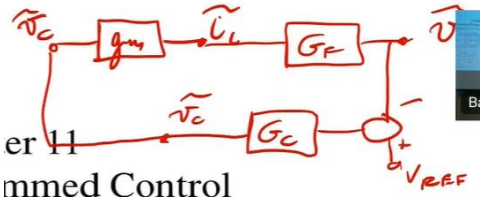
If we assume that i_c signal varies slowly with respect to the time period Ts we can very easily assume that the peak current value in the inductor follows perfectly the control current, switching period after switching period. This is the point of strength of this control modality.

In fact, by doing this we are forcing the current that flows in the inductor to be equal to the control current i_c, so we are transforming the inductor into a simple voltage controlled current generator. The current i_L of the inductor is proportional to the control voltage v_c. Hence using the inner current control loop we are forcing the current that flows in the inductor to be proportional to the control signal. The proportionality factor is Rf.

The advantage is that the inductor is no more an inductor, but a voltage controlled current generator. If we look at the circuit, the state variables are the current in the inductor and the voltage across the capacitor, so in principle the Goc has two poles (two state variables). However, if we use the internal current loop, we are forcing the current in the inductor to be proportional to the control voltage, so we are loosing a state variable and the dynamic of the converter is reduced from a second order dynamic to a first order dynamic. This has a great advantage.

If we use a peak current mode control, the following is the model of the inductor.

The current i_L coming from the inductor is fed to a filter Gf which is a simple parallel of C and R, producing the output voltage. If we want to control the output voltage, we measure it, we compare it with a reference voltage and amplify the error with a compensator, which produces the control voltage.

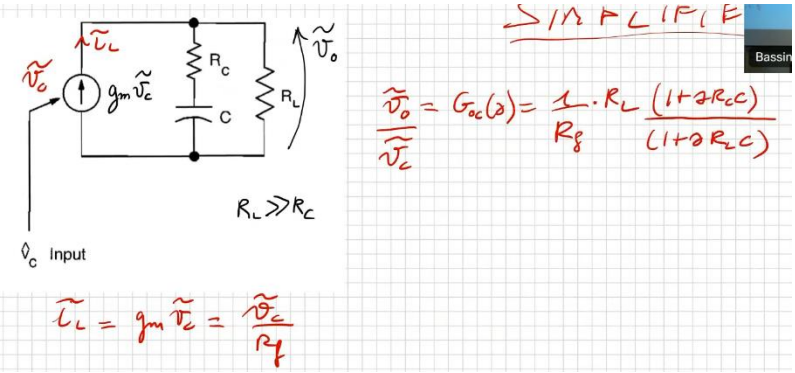


This loop is represented by the pink part of the circuit in the previous slide. So in the end we have a **multiloop system**; we have an inner loop with a large BW, comparable with the switching frequency, and a slower external loop. The inner loop regulates the current that flows in the inductor, the outer loop regulates the voltage across the capacitor, so we are regulating both the state variables in the buck converter.

The advantage is that Gf is a simple LP filter, so the duty cycle to output t.f., so the control to output t.f., is going to be a simple first order t.f., which is very easy to compensate without the need of a type III compensator.

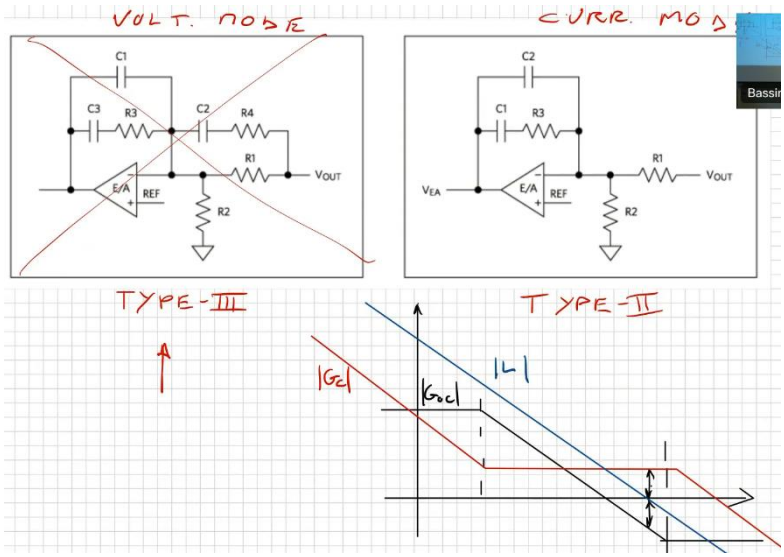
SIMPLIFIED MODEL

In reality, the inner current loop splits the c.c. poles of the buck converter, one pole is moved to LF and the second one to HF where it doesn't have any practical effect, so the final dynamic can be considered as 1st order.

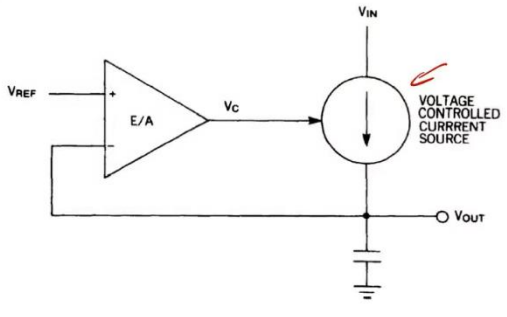


The control input v_c acts on the voltage-controlled current source. We can write the control to output t.f. Goc(s). The Goc(s) is a simple first order t.f. having a negative zero and a negative pole, of course in the assumption Rl >> Rc.

From a practical standpoint, compared to the voltage mode control, in a voltage mode control we typically have to use a type III compensator, especially if the ESR of the output capacitor is minimal. But this is not needed in current mode, we can use a simpler type II.

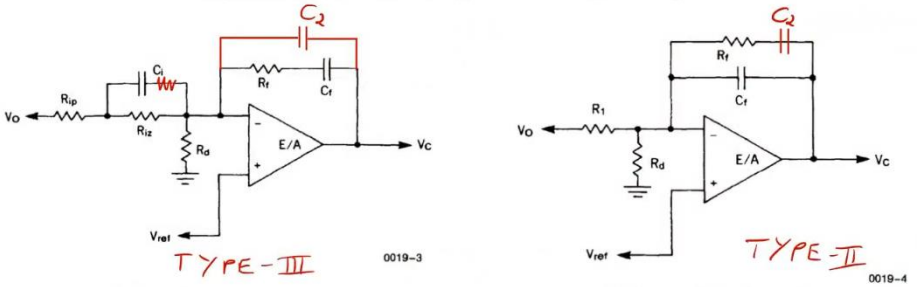


The advantage is that the Goc is a simple t.f. having a pole and a zero, so if we use a simple type II compensator which has a pole in the origin, a zero and a second pole we can place the zero of the compensator at the pole of the Goc and the pole of the Gc at the zero of the Goc ending up with a perfect L(s) having a constant rolloff of -20 dB/dec.
All we have to do is to regulate the mid-frequency gain of the compensator in order to regulate the crossover frequency.



The main idea behind CMC is that the inductor can be turned into a current source, thus eliminating the dynamics of the inductor in the loop. The controller sets a current reference and a fast inner-loop follows this reference cycle by cycle.

Figure 2. Inductor Looks Like a Current Source to Small Signals



A) Direct Duty Cycle Control

B) Current Mode Control

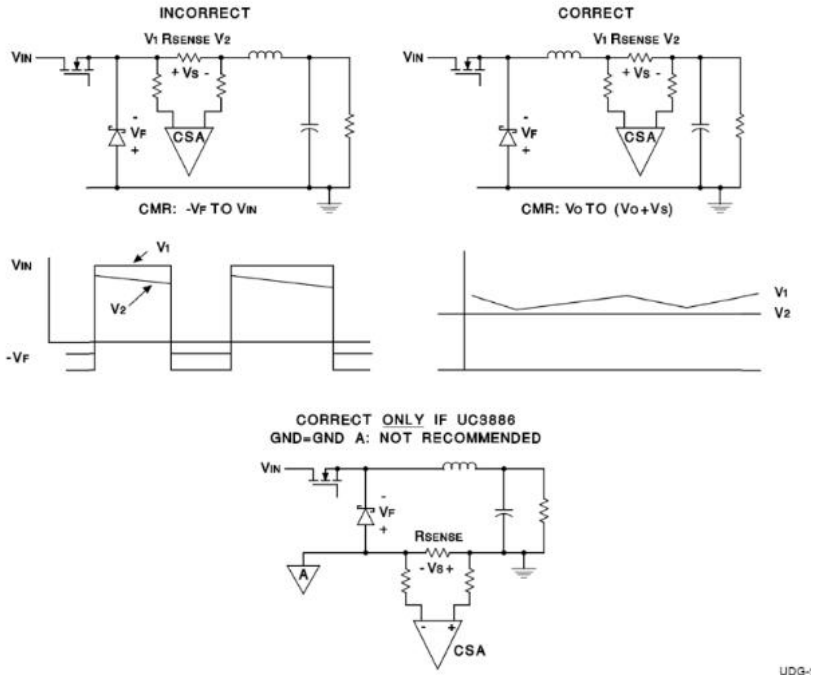
Figure 3. Required Error Amplifier Compensation for Continuous Inductor Current Designs

In principle, we can also use a simplified type II compensator where we drop the C1 capacitor of the previous image, having thus no pole at low frequency, so also at LF the L(s) gain is flat, so we need to check if it is sufficiently high to grant load and line regulations.

The current mode converter is not used just in DC/DC converter, but it is also used to control brushless or DC motors, eventually with 3 or 4 nested loop, always with an internal control loop.

CURRENT MEASUREMENT

Current is typically measured by putting a shunt resistance in series with the inductor. Of course it must be placed correctly; in the incorrect approach, it is before the inductance so we would end up with a current signal with a large swinging common mode, which we want to avoid. So the sensor resistor must be between the inductor and the output.



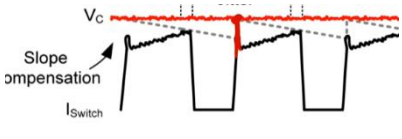
Pros and Cons

Advantages of current programmed control:

- Simpler dynamics: inductor pole is moved to high frequency.
- Simple robust output voltage control, with large phase margin, can be obtained without use of compensator lead networks.
- It is always necessary to sense the transistor current, to protect against overcurrent failures. We may as well use the information during normal operation, to obtain better control.
- Transistor failures due to excessive current can be prevented simply by limiting i_c(t).
- Transformer saturation problems in bridge or push-pull converters can be mitigated.
- Intrinsic feedforward compensation. In a peak current mode control we are not directly setting the duty cycle D; the D at which the converter is working is indirectly set by several parameters like the control current i_c, the current in the inductor and the input and output voltage in the slope term m1. Hence the resulting D in a current mode control is not directly proportional to the control voltage like in a voltage-mode control.

  The consequence is that if for instance we assume to have a step variation of the input voltage, the slope m1 of the current will increase. But if we increase the slope and keep the same control current i_c, the D is automatically reduced, compensating the variation of the input voltage. So the output voltage stays constant because the input voltage variation is absorbed by a reduction of the D.

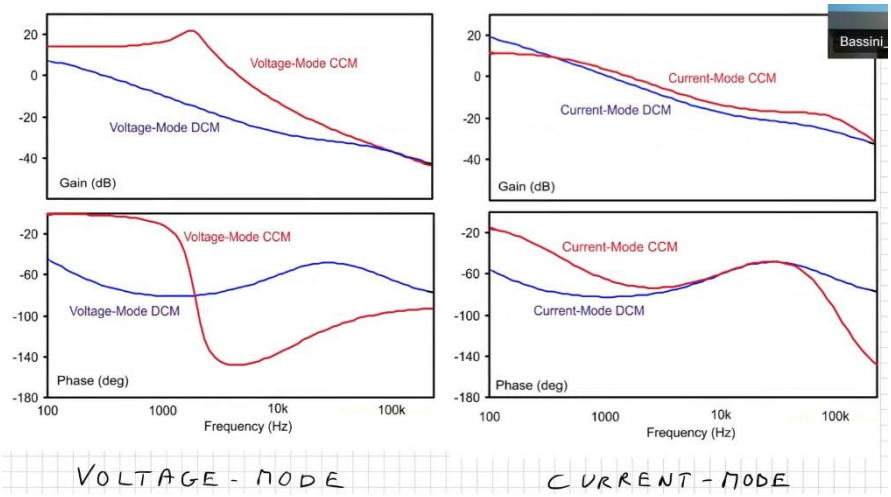
Disadvantage: susceptibility to noise. In fact, the real waveform i_s is noise with spikes when the mosfet is turned on, and this is due to small parasitic inductances in the high current path. The spikes can be high and overcome the control voltage. This can be counteracted with a blanking filter.



However, the major disadvantage are the intrinsic oscillations that occur at D > 0.5. The controller will become unstable for D > 0.5.

Comparison between voltage-mode control and current-mode control

In the plots we have the Goc for a buck converter working in CCM and DCM, modulus (top) and phase. There is a big difference in CCM (couple of c.c. poles) and DCM (single pole) in the voltage mode control. So if we have to design the voltage control loop for a controller that has to work in DCM and for any reason the buck converter starts working in CCM, the converter will very much probably become unstable, so it is difficult to design a control loop to have stable operations both in DCM and CCM with voltage mode.
However, in current mode control it is little difference in the Goc for CCM or DCM operations.



D > 0.5

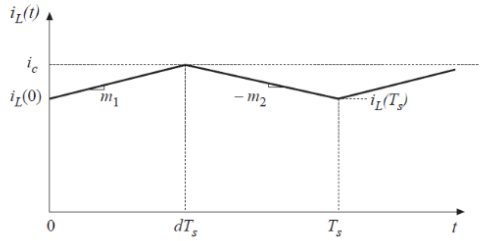
- The current programmed controller is inherently unstable for $D > 0.5$, regardless of the converter topology
- Controller can be stabilized by addition of an artificial ramp

Objectives of this section:

- Stability analysis
- Describe artificial ramp scheme

Why is the converter unstable for D > 0.5?

Let's start by considering the converter working in CCM and the steady state current flowing in the inductor (all transients are over).



Inductor current slopes m_1 and $-m_2$

buck converter

$$m_1 = \frac{v_g - v}{L} \qquad -m_2 = -\frac{v}{L}$$

boost converter

$$m_1 = \frac{v_g}{L} \qquad -m_2 = \frac{v_g - v}{L}$$

buck–boost converter

$$m_1 = \frac{v_g}{L} \qquad -m_2 = \frac{v}{L}$$

The black bold line is the inductor current. During the Ton = DTs the current increases with slope m1, and during Toff the slope decreases with slope -m2. m1 and m2 depend on the converter topology.

M1 and M2 relationship at steady state

First interval:

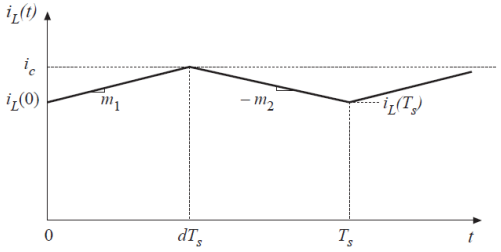
$$i_L(dT_s) = i_c = i_L(0) + m_1 dT_s$$

Solve for d:

$$d = \frac{i_c - i_L(0)}{m_1 T_s}$$

Second interval:

$$i_L(T_s) = i_L(dT_s) - m_2 d'T_s$$
$$= i_L(0) + m_1 dT_s - m_2 d'T_s$$



In steady state:

$$0 = M_1 DT_s - M_2 D'T_s$$

$$\frac{M_2}{M_1} = \frac{D}{D'}$$

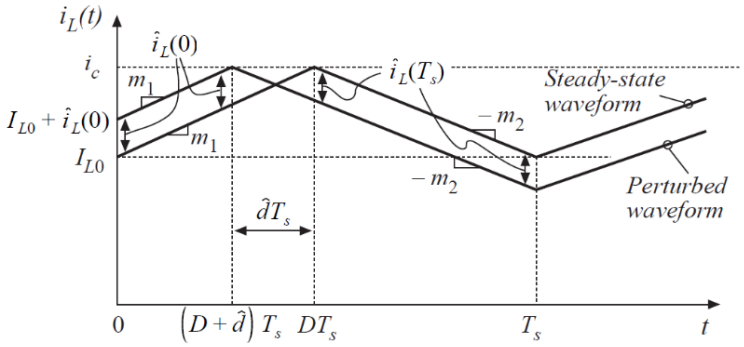
The balance to verify is that delta(i+) = delta(i-). d' = 1 – d.

NB: M2/M1 = D/D' for any converter topology.

Perturbated inductor current waveform

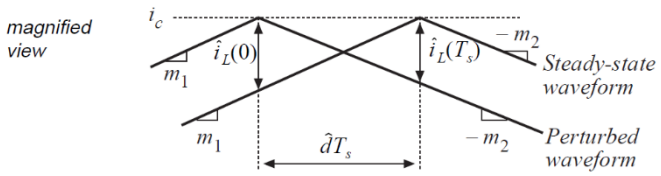
The one that starts form I_L0 is the steady state waveform. At t = 0 we apply a step perturbation to the current by a quantity i_tilde_L(0). We want to understand how this perturbation propagates along time, and if the converter becomes unstable or it returns to a steady state condition.

Once the perturbation has been applied the current starts from a different position but it increases with the same slope m1, so the current crosses the control value i_c sooner than the steady state. So we are also introducing a perturbation to the duty cycle, which is reduces (d_tilde < 0).



At this point the switch is turned off and the current goes down with the same slope m2 and we get a new level of current at the end of the switching period Ts. If we can compute this new value we can understand which is the impact of the perturbation at the end of the switching cycle, and see whether the system is becoming unstable or not.
So we want to compute the perturbation at the end of the switching cycle starting from the knowledge of the perturbation at the beginning of the switching cycle.

This can be done using some trigonometry.



$$\hat{i}_L(0) = -m_1 \hat{d} T_s$$

$$\hat{i}_L(T_s) = m_2 \hat{d} T_s$$

$$\hat{i}_L(T_s) = \hat{i}_L(0)\left(-\frac{m_2}{m_1}\right)$$

x $$\hat{i}_L(T_s) = \hat{i}_L(0)\left(-\frac{D}{D'}\right)$$

Assuming that the perturbation is small, we can replace m2 with M2 and m1 with M1, therefore getting the relationship x.

From this formula, we can apply it to the next switching cycle up to the n switching period.

$$\hat{i}_L(T_s) = \hat{i}_L(0)\left(-\frac{D}{D'}\right)$$

$$\hat{i}_L(2T_s) = \hat{i}_L(T_s)\left(-\frac{D}{D'}\right) = \hat{i}_L(0)\left(-\frac{D}{D'}\right)^2$$

$$\hat{i}_L(nT_s) = \hat{i}_L((n-1)T_s)\left(-\frac{D}{D'}\right) = \hat{i}_L(0)\left(-\frac{D}{D'}\right)^n$$

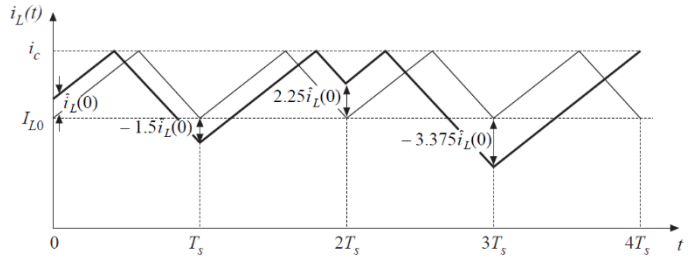
$$\left|\hat{i}_L(nT_s)\right| \rightarrow \begin{cases} 0 & \text{when } \left|-\frac{D}{D'}\right| < 1 \\ \infty & \text{when } \left|-\frac{D}{D'}\right| > 1 \end{cases}$$

For stability: $D < 0.5$

D/D' < 1 if D < 0.5, and in this case we are stable. So if the D > 0.5, cycle after cycle the amplitude of the perturbation becomes higher.

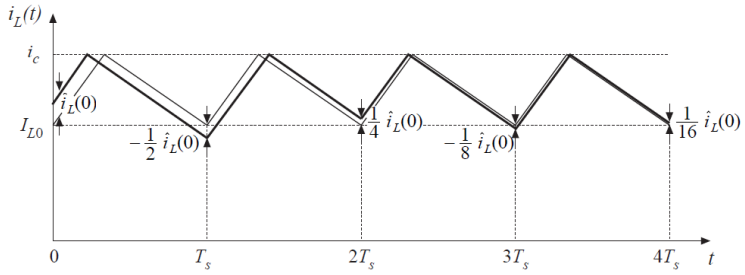
Example – D = 0.6

$$\alpha = -\frac{D}{D'} = \left(-\frac{0.6}{0.4}\right) = -1.5$$

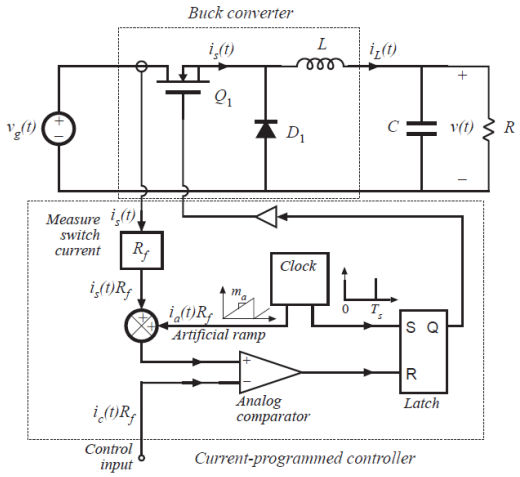


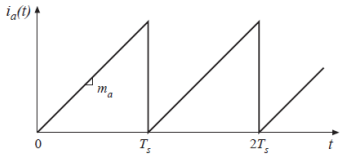
Example – D = 0.33

$$\alpha = -\frac{D}{D'} = \left(-\frac{1/3}{2/3}\right) = -0.5$$



STABILIZATION WITH AN ARTIFICIAL RAMP





Now, transistor switches off when

$$i_a(dT_s) + i_L(dT_s) = i_c$$

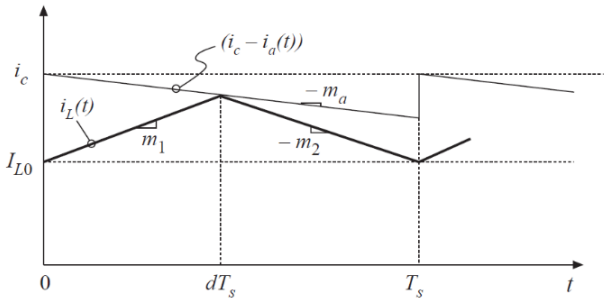
or,

$$i_L(dT_s) = i_c - i_a(dT_s)$$

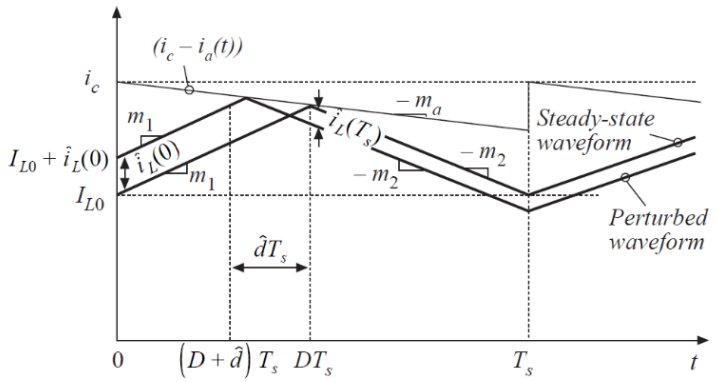
Instead of applying directly i_s(t)*Rf to the non-inverting input of the comparator, we add an artificial ramp, which is a current having a sawtooth voltage waveform.

This results in the following plot.

$$i_L(dT_s) = i_c - i_a(dT_s)$$



This simple trick solves the instability. Calculations are the same as before, what changes is the shape of the i_c waveform.



The application of a perturbation produces a perturbation in the D. Cycle after cycle the following happens.

First subinterval:

$$\hat{i}_L(0) = -\hat{d}T_s\left(m_1 + m_a\right)$$

Second subinterval:

$$\hat{i}_L(T_s) = -\hat{d}T_s\left(m_a - m_2\right)$$

Net change over one switching period:

$$\hat{i}_L(T_s) = \hat{i}_L(0)\left(-\frac{m_2 - m_a}{m_1 + m_a}\right)$$

After n switching periods:

$$\hat{i}_L(nT_s) = \hat{i}_L((n-1)T_s)\left(-\frac{m_2 - m_a}{m_1 + m_a}\right) = \hat{i}_L(0)\left(-\frac{m_2 - m_a}{m_1 + m_a}\right)^n = \hat{i}_L(0)\,\alpha^n$$

Characteristic value:

$$\alpha = -\frac{m_2 - m_a}{m_1 + m_a} \qquad \left|\hat{i}_L(nT_s)\right| \to \begin{cases} 0 & \text{when } |\alpha| < 1 \\ \infty & \text{when } |\alpha| > 1 \end{cases}$$

Is there a way to design the m_a so that alpha is always smaller than 1 for any duty cycle? Yes.

Alpha characteristics

$$\alpha = -\frac{1 - \dfrac{m_a}{m_2}}{\dfrac{D'}{D} + \dfrac{m_a}{m_2}}$$

- For stability, require $|\alpha| < 1$
- Buck and buck-boost converters: $m_2 = -v/L$

  So if v is well-regulated, then m_2 is also well-regulated
- A common choice: $m_a = 0.5\,m_2$

  This leads to $\alpha = -1$ at $D = 1$, and $|\alpha| < 1$ for $0 \le D < 1$.

  The minimum α that leads to stability for all D.
- Another common choice: $m_a = m_2$

  This leads to $\alpha = 0$ for $0 \le D < 1$.

  Deadbeat control, finite settling time

In a buck and buck-boost converter m2 is constant, so we can e.g. set m_a = 0.5*m2 and we get alpha = -1 for any 0 <= D < 1, therefore the converter will always be stable.

Upper left plot is of a system without artificial ramp (also called 'no slope compensation'). In this case not only the system is unstable for D > 50%, there is also another disadvantage, in fact the average inductor current is depending on the duty cycle D. This is a problem if the system we are working on can be controlled with a constant average current.

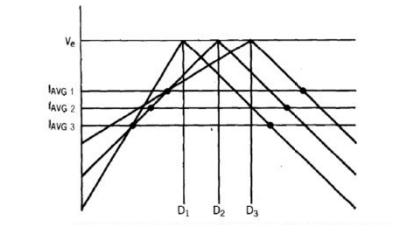


FIGURE 8 - PEAK CURRENT SENSING WITHOUT SLOPE COMPENSATION ALLOWS AVERAGE INDUCTOR CURRENT TO VARY WITH DUTY CYCLE

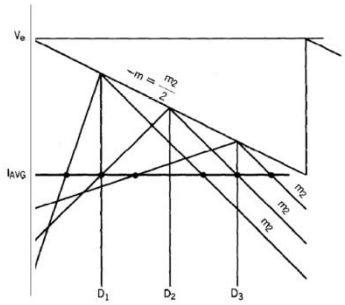
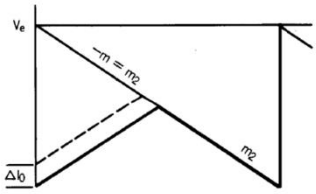


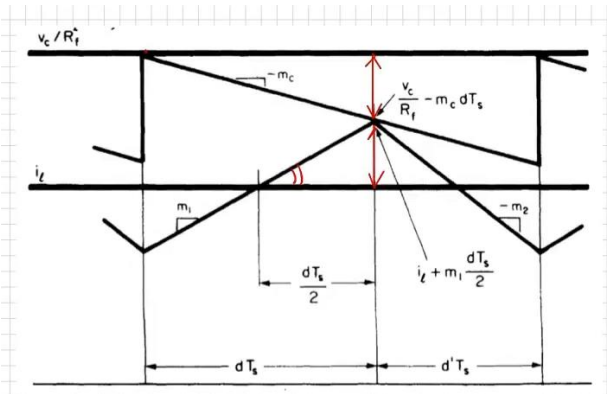
FIGURE 9 - AVERAGE INDUCTOR CURRENT IS INDEPENDENT OF DUN CYCLE AND INPUT VOLTAGE VARIATION FOR A SLOPE COMPENSATION OF m = -½ m2.



GURE 4 - FOR THE CASE OF m = - m2, A CURRENT PERTURBATION WILL DAMP OUT IN EXACTLY ONE CYCLE.

See part 9_additional reading materials: 9.2_Slope compensation

This isn't happening if we are using a slope compensation with m_a = m2*0.5 (bottom left plot). In this case the system works with a constant average current despite the D.
This can be demonstrated mathematically.



$$i_l + m_1 \frac{d T_s}{2} = \frac{v_c}{R_f} - m_c d T_s \quad (1)$$

REPLACING

$$\begin{cases} i_l = I_l + \tilde{i}_l \\ d = D + \tilde{d} \\ m_1 = M_1 + \tilde{m}_1 \\ v_c = V_c + \tilde{v}_c \\ m_c = M_c \end{cases}$$

INTO (1)

WE GET :

$$\begin{cases} I_l + \frac{M_1 D T_s}{2} = \frac{V_c}{R_f} - M_c D T_s \quad (2) \\ \tilde{i}_l + \frac{\tilde{m}_1 D T_s}{2} + \frac{M_1 \tilde{d} T_s}{2} = \frac{\tilde{v}_c}{R_f} - M_c \tilde{d} T_s \end{cases}$$

AT STEADY-STATE :

$$M_1 D = M_2 (1-D) \quad (3)$$

BY REPLACING (3) INTO (2)

$$I_L = \frac{V_c}{R_f} - M_c D T_s - M_2 (1-D) \frac{T_s}{2}$$

CHOOSING $M_c = \frac{M_2}{2}$ WE FINALLY HAVE:

$$I_L = \frac{V_c}{R_f} - M_c T_s$$

I.E. I_L IS INDEPENDENT OF D!

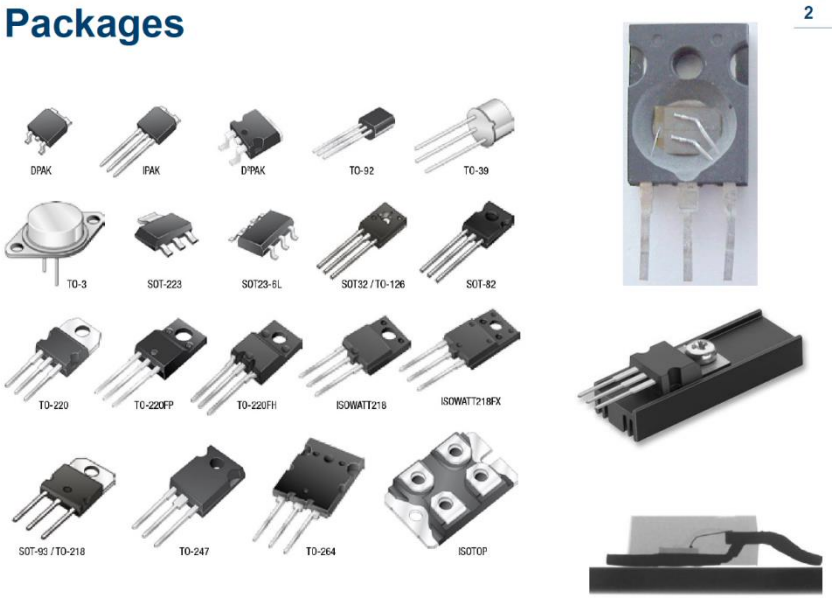
The last possibility is for m_a = m2 (right plot). If we apply a perturbation to the current at the beginning of the switching cycle, due to the equality of the compensation ramp and the slope of the current in the off time period, the perturbation is instantaneously absorbed in a single switching period, regardless the amplitude of the perturbation. This allows us to be very fast in recovering the perturbations on the current (Deadbeat control). This allows to have a large BW in the internal current control.

PEAK CURRENT MODE CONTROL IN A BOOST OR BUCK-BOOST CONVERTER

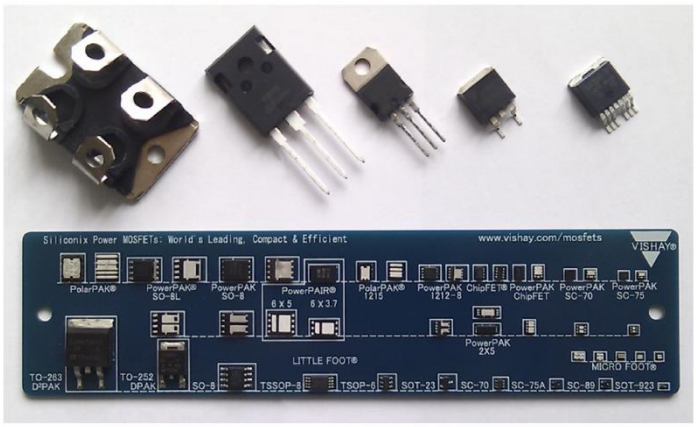
If we use this type of control with a boost or buck-boost converter working in CCM, the RHP zero is not removed. Instead, one of the two c.c. poles is removed.
Nevertheless, even if we have to live with the RHP zero, since the t.f. is a first order t.f. the compensation is much easier.

THERMAL MANAGEMENT

Our prototype of thermal system is the chip including the power electronic circuit packaged in a standard package. In the image we have the most common TH packages. The most common is the TO-220.



We might have also surface mounting packages if we want to shrink the size of the PCB. In this case, in general, we don't have an heatsink.



The case temperature is the temperature at a specified accessible reference point on the package in which the microelectronic chip is mounted.

THERMAL MANAGEMENT

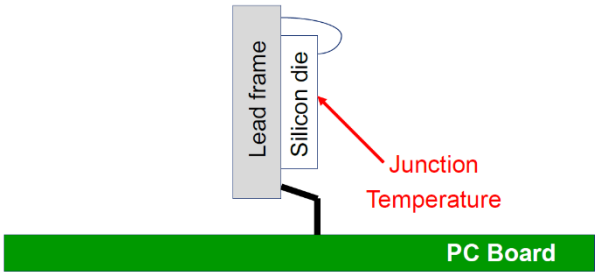
In all electronic applications, temperature becomes an important factor when designing a system. We are so concerned about temperature for at least two reasons:

- The failure rate R of an electronic component increases with temperature, and can often be described by the Arrhenius equation. We can also define the **mean time before failure** that is the reciprocal of the failure rate R.
- The maximum safe junction temperature for plastic encapsulated device is determined by the glass transition temperature of the plastic, approximately +150°C. Exceeding limit temperature for an extended period can result in device failure. If the package becomes soft we might have shear stress on the bonding wires and this may end up in breaking the wires.

- Therefore the temperature of the device must be calculated not to exceed the specified maximum junction temperature (Tjmax).
- To design a good *thermal management* solution, the Tj should always be kept at the lowest operating temperature.

JUNCTION TEMPERATURE

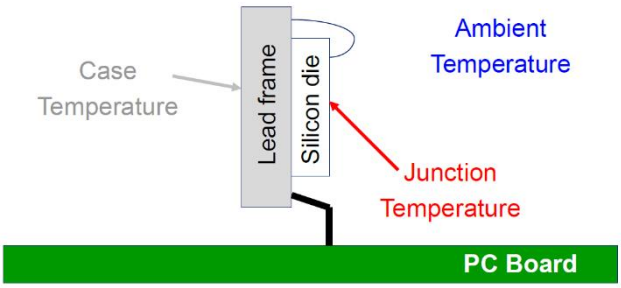
- Junction temperature is the temperature of the hottest place on the silicon die in an integrated circuit or device



If we consider for instance a power mosfet for low voltage. Most of the power is dissipated in the channel, so the hottest point will be close to the channel position on the surface of the silicon die.

AMBIENT AND CASE TEMPERATURE

- This is not the same as the case (or package) temperature or the ambient (or air) temperature



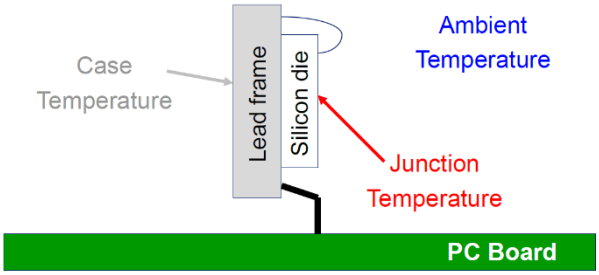
The ambient temperature is the temperature of the surrounding area, and the case temperature is the temperature on the external part of the package, in the case of TO-223 is in general the temperature on the surface of the metal pad.

JUNCTION, CASE AND AMBIENT TEMPERATURE

Let's start from a situation where there is no power dissipation because the device is off. Everything is at thermal equilibrium, so the three temperatures are equal.

- First, the system is off (no power is being dissipated)
- The ambient, package case, and silicon die junction temperatures are in thermal equilibrium

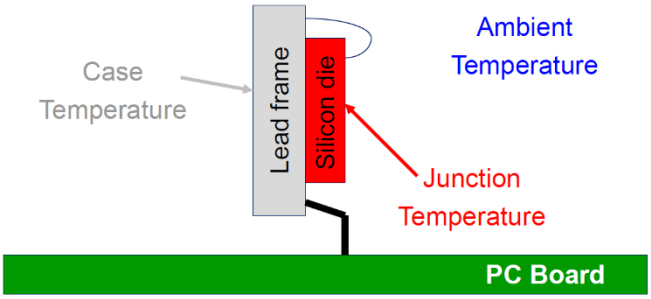
$T_{ambient} = T_{case} = T_{junction}$



If we turn on the power mosfet (or the device) we are dissipating power, so injecting heat in the thermal system. Power dissipation is occurring on the surface of the silicon die, close to the channel region (in a mosfet). So initially the temperature of the silicon chip (Tjun) is increasing, becoming greater than the temperature of the case, which is still at ambient temperature. Since Tjun > Tamb we have a temperature gradient that makes the heat generated close to the surface to transfer inside the thermal system, so heat is transferred from the silicon die to the lead frame and the lead frame starts to increase in temperature.

- Next, the system is turned on
- The silicon die heats up due to the absorbed power being dissipated as heat

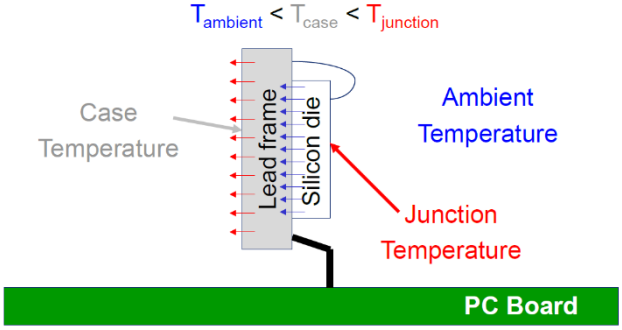
$T_{ambient} = T_{case} < T_{junction}$



So Tcase becomes hotter than the surrounding ambient and so we have a temperature gradient between the metal tab of the package and the surrounding air and so heat is transferred to the ambient.

- Some of the heat is transferred to the package (case)
- The case heats up, but not as much as the silicon die
- From the package (case), some of the heat is transferred to the ambient air

$T_{ambient} < T_{case} < T_{junction}$



After the end of the transient, at a steady state situation, we have a balance. We are balancing the power that is injected into the system, so dissipated inside the silicon chip, with the heat per unit time (so thermal power) that is transferred through the thermal system. In this thermal equilibrium situation we have Tamb < Tcase < Tjun.

Maximum ratings for junction temperature

- Semiconductor devices are specified by their manufacturers at a maximum temperature range:

Maximum Ratings

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ...+150	°C

  - T_J max = 200 °C (silicon devices in metal cases)
  - T_J max = 150 °C (silicon devices in plastic cases)

They can be read on the datasheet.

THERMAL SYSTEMS

- Thermal systems transfer or store thermal energy.
  - Thermal energy is associated with the translation, rotation, vibration and electronic states of the atoms and molecules that comprise matter.
- Thermal systems have a static and dynamic behavior similar to mechanical, electrical, and fluid systems, but in some ways they are quite different.
  - Thermal systems exhibit resistance and capacitance effects, can be analyzed by electrical circuit analysis techniques, and have a dynamic response that can be characterized by a time constant.
  - However, there is no thermal inductance.

Heat transfer

- Heat transfer is thermal energy in transit

- Heat transfer occurs when two surfaces have different temperatures, thus causing thermal energy to transfer from the hotter surface to the colder surface.
- If the temperature difference is increased, the amount of heat flow will be increased:

Heat transfer ∝ temperature difference

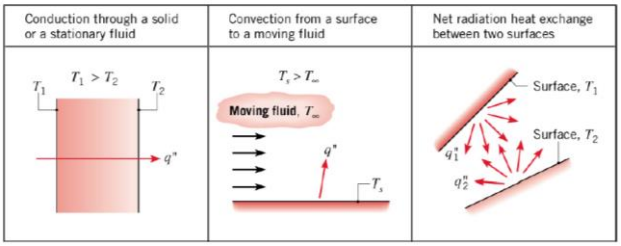
Terminology

Quantity	Meaning	Symbol	Units
Heat	Amount of thermal energy transferred over a time interval $\triangle t > 0$	Q	J
Heat flow (Heat transfer rate)	Thermal energy transfer per unit time	q	W
Heat flux	Thermal energy transfer per unit time and surface area	q"	W/m^2

Heat is an energy, so its unit of measure is the J. Heat flow is instead a power.

Heat transfer modes

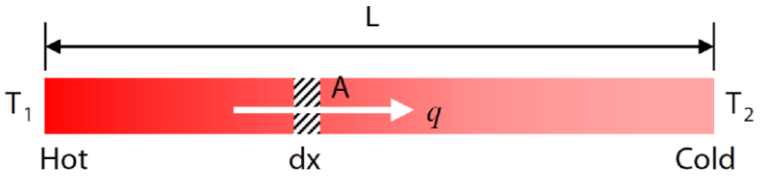
- **conduction** : heat transfer in a solid or a stationary fluid (gas or liquid) due to the random motion of its constituent atoms, molecules and /or electron
- **convection** : heat transfer due to the combined influence of bulk and random motion for fluid flow over a surface.
- **radiation** : energy that is emitted by matter due to changes in the electron configurations of its atoms or molecules and is transported as electromagnetic waves.



Conduction

Heat transfer driven by temperature gradients within a material:

- Vibration of atoms(lattice waves), electron movement : solid
- Molecule motions and impact : gas or liquid



- Heat flow: q [W] = λ A ΔT/L = (T_1 - T_2) / R_{thcond} (from Fourier's equation)
- Thermal resistance: R_{thcond} = L / λ A
  - Cross-sectional area: A
  - λ = Thermal conductivity [$W\text{-}m^{-1}\text{-}°C^{-1}$] ($\lambda_{Al}$ = 237 $W\text{-}m^{-1}\text{-}°C^{-1}$)
  - Units of thermal resistance are °C/W

Let's assume to have a bar or rod of a given material where the left surface has a temperature of T1 and the right surface a temperature of T2. The heat transfer rate in this case is proportional to the temperature difference between the two surfaces and it is inversely proportional to the length of the bar.

In particular, we can identify a thermal resistance, which is associated to the conduction mechanism, and it is the ratio between the length of the rod and the product of lambda (thermal conductivity coefficient) and A, the cross-section of the rod. Rth is very similar to the electrical resistance.

Typical thermal conductivity for heat sinks commonly used materials

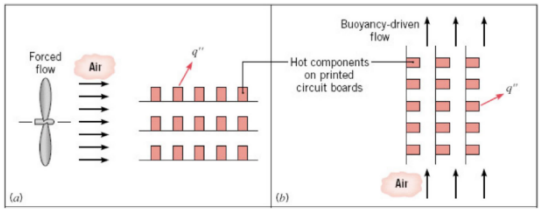
TABLE 8.2 Thermal Conductivity k in W/(m °C) for Some Materials

Aluminum	237	Zinc	116	Rubber	0.15
Brass	111	Epoxy	0.207	Plastic	0.25
Copper	401	Mica	0.71	Teflon	0.22
Iron	80	Mylar	0.19	Paper	0.15
Iron 1% silicon	42	Nylon	0.242	Air	0.0262
Iron 5% silicon	19	Bakelite	0.19	Water	0.597

Material	W/mK
Aluminum (pure)	216
Aluminum Nitride	230
Alumina	25
Copper	398
Diamond	2300
Epoxy (No fill)	0.2
Epoxy (High fill)	2.1
Epoxy glass	0.3
Gold	296
Lead	32.5
Silicon	144
Silicon Carbide	270
Silicon Grease	0.2
Solder	49.3

Convection

- Convection is the heat transfer mode between a surface and a fluid. The convection heat transfer can be divided in two main phenomena:
  a) forced convection, where the fluid motion is due to external forces (imposed for example by a fan)
  b) natural convection, where the fluid motion is due to buoyancy forces, because of the density modification of the fluid close to the surface



Heated air rises, cools, then falls. Air near hot surface is replaced by cooler air, and the cycle repeats.

The convection can be described using a convection thermal resistance.

Irrespective of the details of the mechanism, the rate of heat transfer by convection between a surface and a fluid can be calculated from the relation:

$$q = h \cdot A \cdot (T_s - T_\infty) \qquad (1)$$

- h is the convection heat transfer coefficient [W/m²K],
- T_s is the surface temperature,
- T_∞ is the temperature of the fluid at some specified location far from the surface.

Equation (1) can be written as:

$$q = (T_s - T_\infty)/ R_{thconv}$$

where

$$R_{thconv} = 1/(hA)$$

is the thermal resistance to convection heat transfer ([K/W] or [°C/W]).

However, using the Rth is a strong simplification because the A coefficient is not constant.

Typical values of convection coefficient

Air		*Water*
Natural convection	*5*	*200*
Forced convection	*10 - 100*	*500 - 10,000*
Condensing steam	*-*	*5,000 - 50,000*

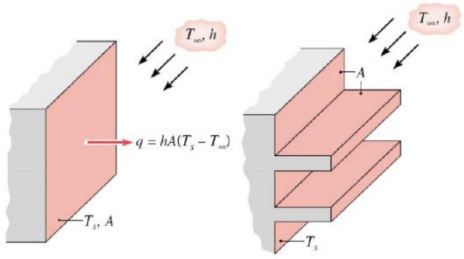
Decreasing Rth_conv in the convection mechanism

We have either to increase h, but this is not possible once selected the fluid and the regime of the flow, or we have to increase the surface area. The idea is to increase the heat exchange surface without increasing the volume too much. We can do this using **fins**.

There is a strategy to optimize the heat sink, and it is based on adopting the minimum distance between the fins (so that the density is large) but still above the minimum optimum distance, which is the one that grants a turbulent flow of the air between fins (in case of natural convection, fins distance is more or less 0.5 cm). In fact, if the flow regime is turbulent we are maximizing h.

$R_{thconv} = 1/(hA)$

- Fins are used to add a secondary surface to the primary surface and thus increasing the heat transfer area.



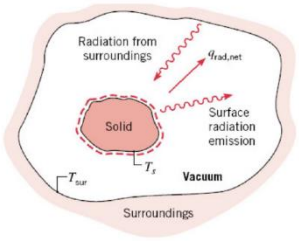
- The critical issue in the application of finned sinks is to ensure that air flow through the fins is turbulent rather than laminar.

However, if the heat sink is intended to work in forced convection, we can decrease even more the distance between fins (0.1 ~ 0.2 cm). Never use a heat sink designed for forced convection in a natural convection case.

Heat radiation

Not so important to the overall heat propagation because the surfaces of the packages are in general very small. However, most of the heat sinks are black not because black is cool and we are in a fashion contest, but because black aluminum has a very large emissivity, so we are maximizing the heat transfer rate contribution due to radiation, even if contribution is not dominant. It is not black-painted aluminum, the black comes from the aluminum oxide, derived anodizing the aluminum.

Radiation heat transfer occurs as a result of radiant energy emitted from a body by virtue of its temperature.



Typical ε (dimensionless)

- Polished Aluminium 0.04
- Black Paint 0.98
- Skin 0.95

Rate of heat transfer:

$$q = \sigma \cdot \varepsilon \cdot A(T_s^4 - T_{surr}^4)$$

where

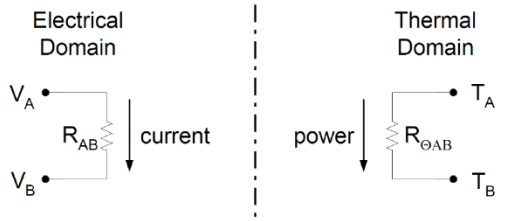
σ = Stefan-Boltzmann constant
(5.669 x 10^{-8} W/m^2.K^4)
ε= emissivity
T= absolute temperature

Summary of heat transfer processes

Mode	Mechanisms	Rate Equation	Transport Property or Coefficient
Conduction	Energy Diffusion due to random molecular motion	$q_x''(W/m^2) = -k\frac{dT}{dx}$	k (W/m·K)
Convection	Diffusion + Bulk motion (advection)	$q''(W/m^2) = h(T_s - T_\infty)$	h (W/m^2·K)
Radiation	Electromagnetic waves	$q''(W/m^2) = \varepsilon\sigma(T_s^4 - T_{sur}^4)$ or $q(W) = h_r A(T_s - T_{sur})$	ε h_r (W/m^2·K)

THERMAL – ELECTRICALL DUALITY

- Both have an "across"variable, a potential
- Both have a "through"variable, something that flows
- Both have a resistive element that impedes flow



Comparison between thermal and electrical circuit

Electrical Circuit Properties	Thermal Circuit Properties
Voltage : V	T : Temperature
Current : I	q : Heat Flux
Resistance : R	R_{th} : Thermal Resistance

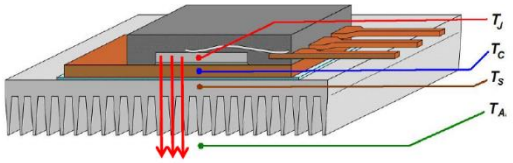
$$\frac{dQ_{charge}}{dt} = -\sigma A\frac{dV}{dx} \equiv \frac{\Delta V}{R} \quad \longleftrightarrow \quad \frac{dQ_{heat}}{dt} = -\kappa A\frac{dT}{dx} \equiv \frac{\Delta T}{R_{th}}$$

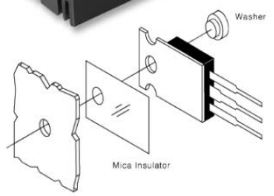
Summary of the duality

	Electrical Domain			Thermal Domain		
	Variable	Symbol	Units	Variable	Symbol	Units
Through Variable	Current	I	Amperes or Coulombs/s	Power or Heat Flux	P_D	Watts or Joules/s
Across Variable	Voltage	V	Volts	Temperature	T	°C or K
Resistance	Electrical Resistance	R	Ohms	Thermal Resistance	$R_{\Theta AB}$	°C/W or K/W
Capacitance	Electrical Capacitance	C	Farads or Coulombs/V	Thermal Capacitance	C_Θ	Joules/°C
"Ohm's Law"	$\Delta V_{AB} = V_A - V_B = I * R_{AB}$			$\Delta T_{AB} = T_A - T_B = P_D * R_{\Theta AB}$ (derived from Fourier's Law)		
	Electrical impedance			*Thermal impedance*		

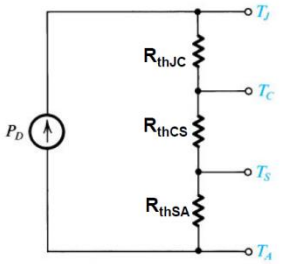
Of course we need to consider a thermal resistance if the system is at thermal equilibrium, so there is no transient and everything is at steady state. When we are considering transient situations we have to introduce another element in the system which is the thermal capacitance.

THERMAL CIRCUIT





- Conduction path leads to a convective interface.
- Heat is removed from a finned sink by air flowing over the fins.



- Steady-state transport
  - thermal equilibrium, i.e. T_x=const.
  - heat flow = q = P_D = dissipated power
- Heat transfer equation

$$(T_J - T_A) = P_D (R_{thJC} + R_{thCS} + R_{thSA})$$

Our prototype thermal system is the chip inside the package, the package is mounted on a heatsink and there is an insulator layer between the package and the heatsink because the heatsink is conductive and the metal tab has the same potential of the drain. If the system has the load connected to Vdd, when the switch is off the drain of the mosfet is at Vdd which might be very high and we don't want to have an electric shock when touching the heatsink.

However, just putting a layer of insulator is not sufficient, because there's a screw which goes through the hole of the TO-220 and we have to tighten the screw to attach it to the heatsink. If we don't use a insulating washer on one side and the insulating bushing on the other, otherwise we still get the electric shock because the screw is still metal.

When we turn on the mosfet and we wait for a while until the transient is over, at steady state the heat is generated where we have the red arrows, it propagates by conduction through the metal tab and the heatsink and then by convection to the surrounding ambient.

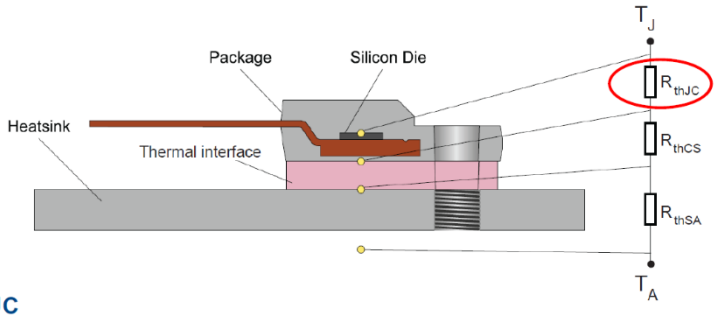
So the heat is moving along a path that includes three components: silicon, glue, package.

We can lump these three contributions into a single thermal resistance, the thermal resistance R_thJC between the junction and the case. Then the heat is moving through the insulator layer, which has its own thermal resistance R_thCS (case to sinker thermal resistance). The third contribution is a convection contribution, which can be modelled with a thermal resistance between sinker and ambient, R_thSA.
So we have three resistances in series and the heat generated inside the system is modelled with a current generator.

The voltages at the nodes of the network correspond to the temperature: ambient temperature Ta, sinker temperature Ts, case temperature Tc and junction temperature Tj. Once Ta is known we can derive all the other temperatures.

At steady state, so when the temperatures are constant, the heat flow is equal to the power dissipated, and therefore the total temperature drop between the junction and temperature Tj – Ta is as in formula x.

Junction-to-case thermal resistance



R_{thJC}

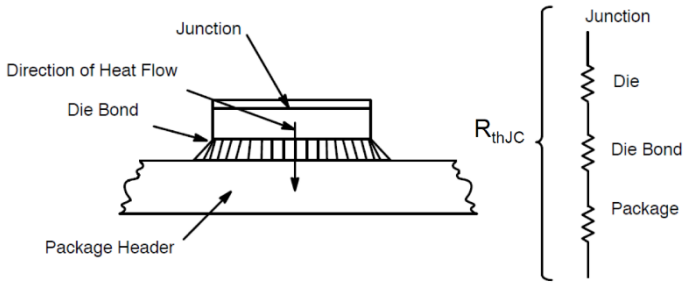
- Junction-to-case thermal resistance
- Defined by the manufacturer, e.g.

Junction−to−Case (Drain) − Steady State (Note 1)	$R_{\theta JC}$	1.4	°C/W

for the NTP5864N

- The circuit designer's only influence on R_{thJC} is in choosing an appropriate device

It is something we cannot change once we bought our device. R_thJC is the lump of three contributions, as below.



The R_thJC, combined with the case temperature, determines the maximum power dissipation that we have in the datasheets.

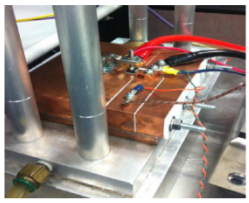
Maximum power dissipation ratings

- Junction-to-case thermal resistance, R_{thJC} and case temperature, T_c determine the max power dissipation.

MAXIMUM RATINGS

Parameter			Symbol	Value	Units
Continuous Drain Current – $R_{\theta JC}$ (Note 1)	Steady State	T_C = 25°C	I_D	63	A
		T_C = 100°C		45	
Power Dissipation – $R_{\theta JC}$ (Note 1)	Steady State	T_C = 25°C	P_D	107	W
		T_C = 100°C		54	
Operating Junction and Storage Temperature			T_J, T_{STG}	−55 to 175	°C

NTP5864N

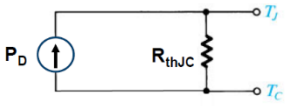


Note: the device is typically mounted on a water-cooled heat sink to set T_C at 25 °C.

Therefore:

$$T_j - T_C = R_{thJC} \cdot P$$

x $$P_{MAX} = \frac{T_{jmax} - 25°C}{R_{thJC}}$$

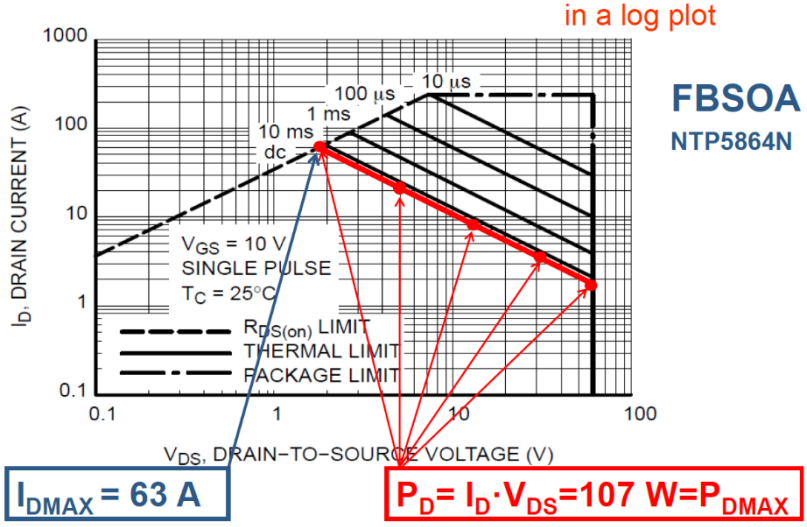


The numbers in the table are measured by the manufacturer by putting the package on top of an infinite heatsink, which is a copper block where there is a flow of water to keep the temperature of the block at 25°C, so the case temperature is maintained at 25°C.

The temperature drop between the junction and the case is Tj – Tc = R_thJC*P, with fixed Tc = 25°C. We notice that the maximum power dissipation decreases if the temperature increases simply because we are increasing Tc = 25°C.

FBSOA is the forward bias safe operating area and below we have a plot of the drain current as a function of Vds in a log-log plot. If we plot the equation x of the previous image in the graph below we get a line, because the power dissipation is Vds*Id = Pmax in the worst case.
In the log-log plot, log(Vds) + log(Id) is constant, so it is a line in a log-log plot and the maximum power dissipation is defined by a line.
Moreover, the maximum current that can flow in a mosfet at steady state is determined by the maximum power dissipation which, in turn, is depending on the maximum junction temperature.

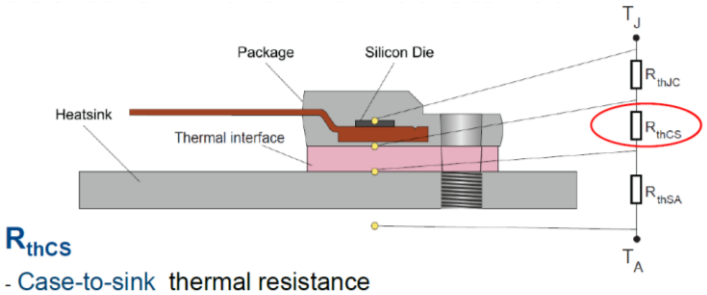
$\log(P_{max}) = [\log(I_D) + \log(V_{DS})]_{MAX} = \log(const)$ ⇨ line with slope -1 in a log plot



FBSOA
NTP5864N

Case-to-sink thermal resistance

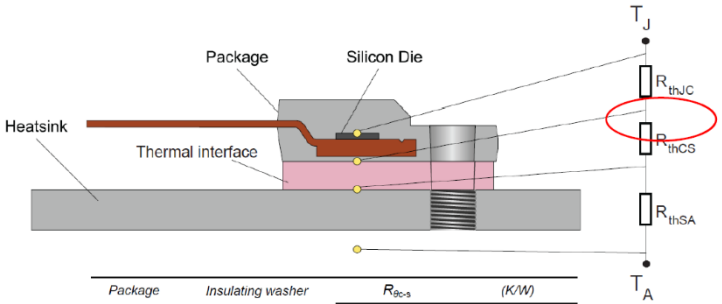
This is the second contribution, and it comes from the insulating layer. We can decide the material to use for the insulator and its thickness.



R_{thCS}

- Case-to-sink thermal resistance
- Designer's influence on R_{thCS}:
  - Choosing device with the desired type of case.
  - Applying a thermally-conductive compound between case and sink.
  - Because the collector (or drain) is usually electrically connected to the case, and the collector is not usually at ground potential:
    - The case must be electrically insulated from the sink (usually with a washer of mica or similar material), or
    - The sink must be electrically insulated from the chassis
  - Without additional information, it is usually assumed that $R_{thCS} \cong 1$ °C/W.

The thermal resistance of the insulating layer depends on: thickness L of the insulating layer, area A of the insulating layer and thermal conductivity lambda of the insulating layer.

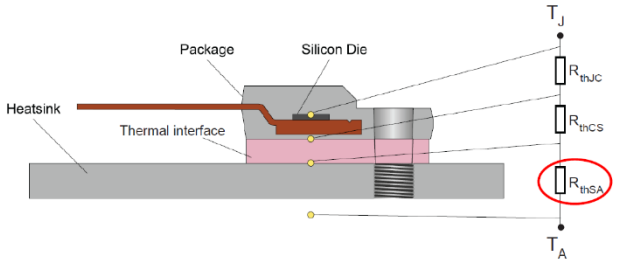


Package	*Insulating washer*	$R_{θc-s}$ *(K/W)*	
		Silicone grease	
		with	*without*
TO-3	No insulating washer	0.10	0.3
	Teflon	0.7-0.8	1.25-1.45
	Mica (50 -100 µm)	0.5-0.7	1.2-1.5
TO-66	No insulating washer	0.15-0.2	0.4-0.5
	Mica (50 -100 µm)	0.6-0.8	1.5-2.0
	Mylar (50 -100 µm)	0.6-0.8	1.2-1.4
TO-220	No insulating washer	0.3-0.5	1.5-2.0
	Mica (50 -100 µm)	2.0-2.5	4.0-6.0
TO-247	No insulating washer	0.1-0.2	0.4-1.0
	Mica (50 -100 µm)	0.5-0.7	1.2-1.5

Typical case-to-sink thermal resistances value for various packages

The thermal resistance between case and sinker is in the order of 2.0 to 2.5 °C/W for TO-220.
Sometimes we need to use Silicone grease (silicone sigillante) because the surface of the heatsink is not flat and to maximize the match between the rough surface of the heatsink and the rough surface of the metal tab we can use it. If we put too much Silicone grease we increase too much the distance and the effect is worsened. So we need to put the grease and tighten the screw and check that the amount that grease that comes out from the boundary is small.

Sink-to-ambient thermal resistance



R_{thSA}

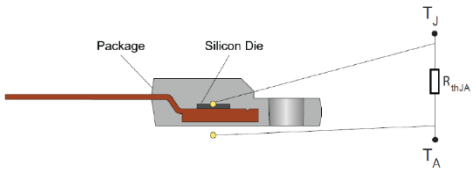
- Sink-to-ambient thermal resistance
- The circuit designer's influence on R_{thSA} is by selection of the heat sink

This contribution strongly depends on the heatsink we are using. The thermal resistance between sinker and ambient depends on many factors, like the shape, size and material the heatsink is made of.

Junction-to-ambient thermal resistance

Sometimes, in the datasheets of thermal components, we find this thermal resistance quoted. It means that if we are using the device with no heat sink at all we have to consider this total thermal resistance between junction and ambient, which is larger than the one between junction and case.

Junction−to−Ambient − Steady State (Note 1)	$R_{\theta JA}$	33	°C/W	**NTP5864N**



R_{thJA}

- Junction-to-ambient thermal resistance
- Measure of heat flow between the chip junction and the air
- Relevant for packages used ***without external heat sinks***

If the system dissipates a small amount of power we can avoid using an heatsink.

JUNCTION TEMPERATURE PREDICTION

Example:

- MTP15N06V in TO-220 package, R_{thJC} = 2.73 °C/W
- R_{thCS}= 1.2°C/W
- FischerElektronik SK 409 63,5 heat sink; R_{thSA}= 5.6 °C/W
- P_D = 8 W
- T_A = 40 °C

- $T_C = T_A + P_D (R_{thSA} + R_{thCS}) = 94.4$ °C
- $T_J = T_A + P_D (R_{thJC} + R_{thSA} + R_{thCS}) = 116.2$ °C

We have a nMOS in package TO-220.

HEATSINK SELECTION

We want to select the heatsink to keep the maximum junction temperature within specified limitations. The design equation is the one below.

- The following equation applies:

$$R_{thSA} < \frac{T_{J_{MAX}} - T_A}{P_D} - R_{thJC} - R_{thCS}$$

- Heat sinks range . . .

. . . from none (device case *is* the heat sink),
. . . to clip-on units,
. . . to large extruded aluminum units,
. . . to forced-air cooled arrangements,
. . . to water-cooled copper units.

So the heatsink can be selected by calculating the needed thermal resistance between the sinker and the ambient from the formula above. Once the upper bound to the thermal resistance is known, we have to pick up the heatsink.

HEAT SINK CATEGORIES

Heat Sink Category	Advantages	Disadvantages	Applications	Examples
Passive	• Widely available • Cost • Ease of use	• Limited in power dissipation capability	• Natural convection / systems that do not depend on air flow • Low power density applications	Metal plate
Semi–Active	• Lower thermal resistance for the same volume compared to passive heat sinks	• Limited in power dissipation capability	• Systems with existing air flow • Medium power density applications	Fin heat sink
Active	• Incorporates forced convection and heat sink into a single unit • Provides much greater heat dissipation capability compared with passive and semi–active heat sinks	• Long term reliability • Cost	• High power density applications	Fan heat sink
Liquid Cooling	• Provides much greater heat dissipation capability compared with passive and semi–active heat sinks	• Complexity • Cost	• High power density applications • Low profile applications requiring constant heat cycling	Liquid cold plate
Phase Change Cooling	• Heat is spread evenly • The cooling liquid very effectively removes heat	• Complexity • Cost • Requires additional board space and height	• High power density applications	Vapor compression phase–change cooler

- Passive: we have no mechanical fan, and the typical mechanism is natural convection. The cost is very low and easy to find in the market. Furthermore, they are also very simple to use. However, a typical disadvantage is the limited amount of power dissipation capability, so they are typically used when the application doesn't need a large power density.
- Semi-active: passive heatsink with an additional blower, so we put a fan next to it. With the fan we can reduce the thermal resistance between sinker and ambient. Typical applications are medium power density applications, or situation where we have already a passive heatsink and we want to improve the thermal resistance by putting a fan or blower close to it.
- Active: it incorporates a blower or a fan, it is a single unit. They are typically used when the room where to place the heatsink is pretty limited and power density requirement are quite demanding.

We will neglect the other categories because they are used in high voltages applications.

HEAT SINK TYPES

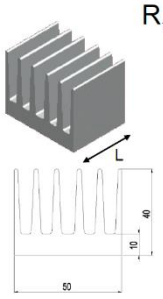
Heat Sink Category	Description	Advantages	Disadvantages	Applications	Examples
Stamped	• Sheet metal stamped into desired shapes	• Cost	• Limited power dissipation capability	• Low power applications • Natural convection applications	Metal plate
Extruded	• Heat sink with extruded fins	• Fins align in direction of air flow • Cross-cut fins for omni-directional air flow • Cost • Sizing flexibility	• Limited power dissipation capability	• Medium power density applications • Natural or forced convection applications	Extruded fin heat sink
Bonded Fin	• Thermally conductive aluminum-filled epoxy fins attached to a base plate	• Effectively increases the surface area of heat sink for the same form factor due to a much higher fin height-to-gap ratio	• Cost	• Forced convection applications • Higher power applications for smaller form factor	Aluminum filled epoxy fins
Casting	• High density fins or pins produced with die casting process	• Higher performance and heat sink shape flexibility compared with standard extrusion or bonded fin heat sinks	• Cost	• High density forced convection applications	Pin heat sink

Example of passive heatsink selection

Given

- R_{thJC} = 2.73 °C/W; R_{thCS}= 1.2 °C/W
- P_D = 15 W
- T_A = 50 °C, T_{JMAX}=150 °C

$$R_{thSA} < \frac{T_{J_{MAX}} - T_A}{P_D} - R_{thJC} - R_{thCS} = 2.73\,C/W$$



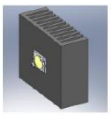
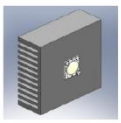
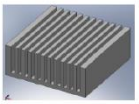
Example: Alutronic PR289/50/AL

Pv (W)	R thk (K/W)		
10	3,60	2,70	2,20
20	3,40	2,45	1,90
30	3,27	2,26	1,83
40		2,12	1,71
50		2,00	1,56
l (mm)	50	100	150
A (cm²)	186	351	516
V (cm³)	52	104	155
m (g)	140	280	420

The data are in the image, and the DC power dissipation of the mosfet is 15W. Ambient temperature is 50°C because maybe we are in a box, not in open air.
Once we have R_thSA we have to understand the best shape for us, that in this case is an extruded aluminum heatsink.

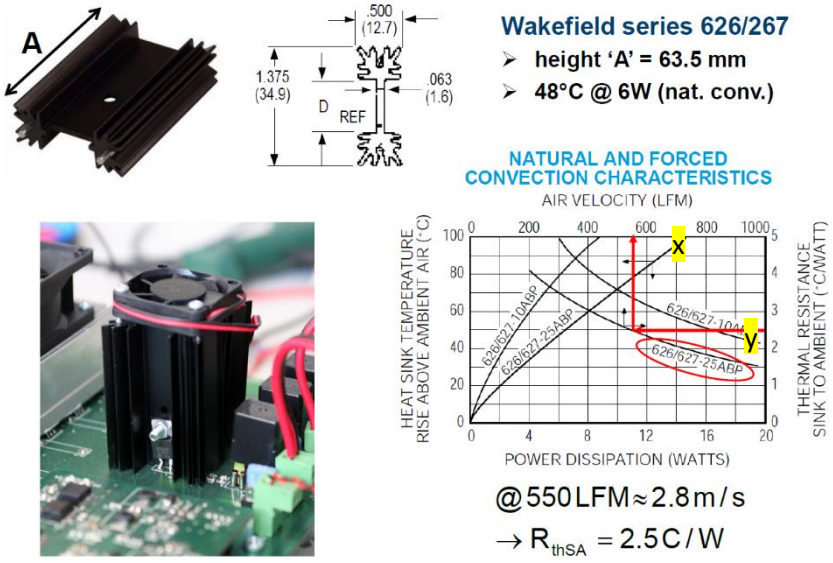
We can notice that the R_thSA (thermal resistance between sinker and ambient) is not constant but it depends on the power that we are dissipating. If we are dissipating 10W we have 2.70, 20 we have 2.45 and so on.

Moreover, when using a passive heatsink, the orientation of the fins must be such that the flow of the air moves through the fins along the longest direction (L in the image) and from top to bottom (vertical orientation of the next slide).

Fin Orientation	Illustration	Guideline	Experimental Results		
		Typical Relative Effectiveness	Measured T_h	$R\Theta_{ha}$ (°C/W)	Typical Relative Effectiveness
Vertical		100%	65°C	2.2	100%
Horizontal		85%	69°C	2.5	88%
Horizontal Up		70%	84°C	3.2	61%

The consequence is that the effectiveness of the heatsink is reduced down to 85%, for example. The thing is even worse if the heatsink is mounted flat.

Example of semi-active heatsink selection



We have a power mosfet which is mounted on the heatsink, so we haven't selected it, but we need to check if it is enough to remove the power.
The thermal resistance is reported in the plot aside, which reports the heatsink temperature above the air temperature (curve x) as a function of the power dissipation. Having 48°C @ 6W means that R_th = 48/6.
For a power of 15W this is not enough, we would go above ambient temperature of 100°C. So we can add a fan. To get the correct velocity of the fan, we need to look at curves y, which report the thermal resistance as a function of the air speed, where LFM is linear feet per minute. So we start by setting the needed thermal resistance, we move horizontally until we intercept the curve and then we go vertically to read the air velocity needed.

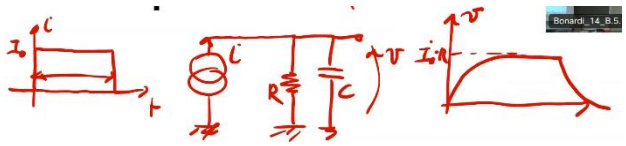
TRANSIENT THERMAL ANALYSIS

In this case we need to compute the junction temperature not at steady state but during a transient. There are 3 types of transients: **step like pulse**, **rectangular pulse** and **train of rect power pulses** with a given period and a given duration. More complex transients can be analyzed recalling them back to these three transients.

Upon the application of a rect power pulse, typically the following mistake is made. Let's assume that we are applying the power pulse and we want to understand the temperature of the junction. In general, we are interested in the maximum temperature reached by the junction, which usually it's at the end of the power pulse. The most common error is to compute this maximum temperature using the steady state thermal resistance, but this is wrong, we are severely overestimating the junction temperature.

Moreover, if I want to compute the temperature of the junction upon the application of a train of power pulses, the typical mistake is to compute the average power and then multiply the average power by the thermal resistance to get the junction temperature. In this case we are underestimating the junction temperature.

In both cases we are neglecting the thermal capacitance of the thermal system; in fact, the thermal system is able to store energy, and this ability is represented by the introduction in the thermal model of a thermal capacitance.
For instance, let's assume we are injecting a current in a RC cell and we are measuring the voltage. The current is a rectangular current; if the duration of the rect is much larger than the time constant we see an exponential increase of the output voltage and then an exponential decrease. This is exactly the same thing that happens in a thermal system when we account for the thermal capacitance.



However, if we reduce the duration of the rect and we make it smaller than the time constant, the voltage increases, it cannot reach the steady state because at a given time it stops, and therefore the peak voltage drop we measure is much smaller than Io*R (asymptotic value) and the reason for this is the capacitance. The same happens in a thermal system.

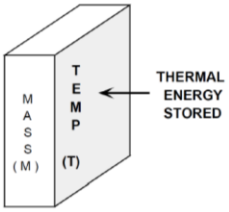
Transient thermal analysis:

- Involves determining the junction temperature under conditions that vary over a period of time.
- Typical transients are induced by the application of single or multiple power pulses.
- Use of steady-state thermal resistance is not satisfactory for finding peak junction temperatures for pulsed applications:
    - Plugging in the peak power value results in overestimating the actual junction temperature
    - Using the average power value underestimates the peak junction temperature at the end of the power pulse.
- The reason for the discrepancy lies in the thermal capacity of the semiconductor and its housing, i.e., its ability to store heat and to cool down before the next pulse.

THERMAL CAPACITANCE

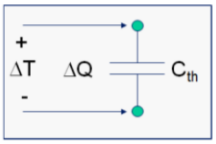
- In the transient case, the thermal capacitance C_{th} must be considered along with the thermal resistance.
- C_{th} is a quantity that represents the capacity to store heat. It is given by:

$$C_{th} = \rho \cdot c \cdot V = \frac{\Delta Q}{\Delta T} \quad [C_{th}]=(\text{Joules}/^\circ\text{C})$$



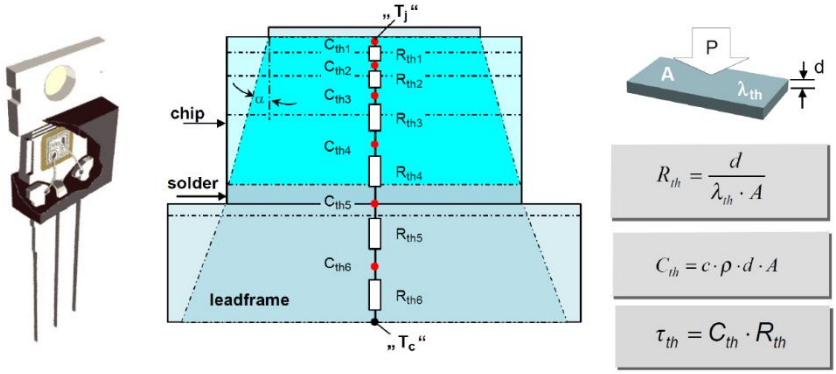
where

- ΔQ is the quantity of heat absorbed in Joules for a temperature rise ΔT, after temperature has stabilised;
- V = volume of sample material;
- ρ= density of material;
- c = specific heat of material, that is the amount of heat per unit mass that must be added to the material to raise its temperature one degree.



It refers to the ability of the system to store energy. It can be computed with the formula in the image. We inject a given quantity of heat in the system and we get a given variation in the temperature, and the ratio between energy injected and temperature variation is the thermal capacitance.

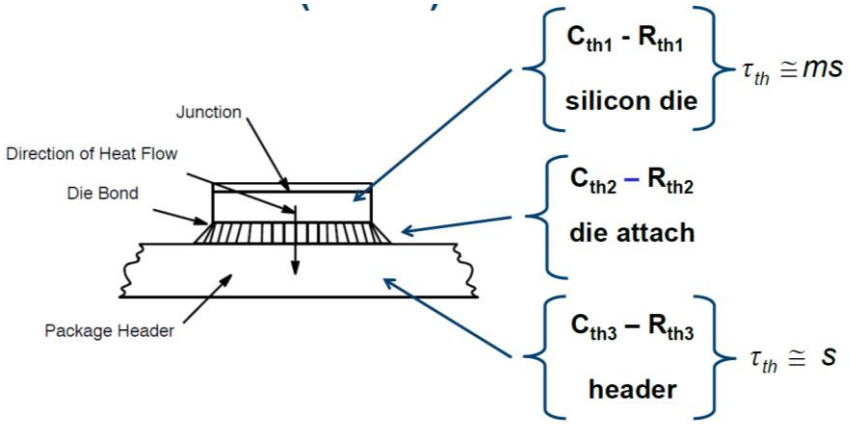
THERMAL MODEL



- Ttypical power transistor in a package with a solid cooling tab – e.g. TO-220 or D-Pak.
- In practice, it has been shown that the segmentation of the structure into partial volumes is not critical

In principle, to properly analyze the system from a thermal point of view, we should develop a distributed model and associate a thermal resistance and thermal capacitance to each part of the model. We can also define a thermal time constant.

However, in practice, it can be demonstrated that the segmentation of the structure into many layers is not really necessary, we can split it into 3 slices and it is more than enough to get an accurate modelling of the thermal system. This is done as in the next slide.

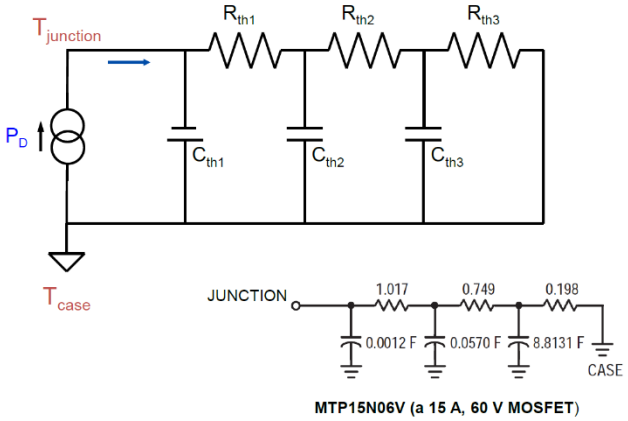


- The normally distributed thermal capacitances have been lumped into three single capacitors.
- This simplification assumes current is evenly distributed across the silicon chip and that the only significant power losses occur in the junction.

We are considering 3 different thermal cells.

THERMAL RC NETWORK

This is the electrical network for modelling the thermal impedance between junction and case, with the previously mentioned 3 cells.



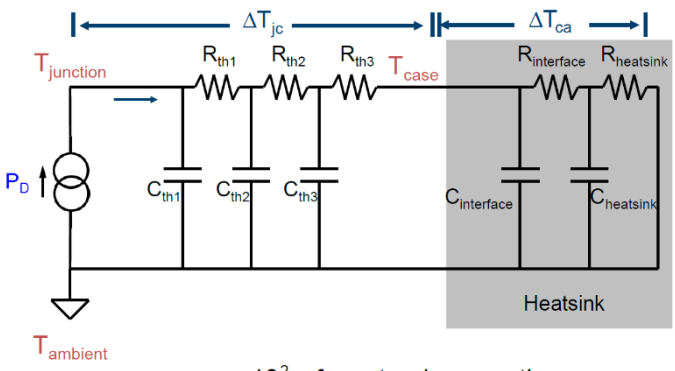
MTP15N06V (a 15 A, 60 V MOSFET)

The thermal network is usually provided in the datasheets.

The way of the image to represent the impedance of the system is called Cauer network, where the capacitors are connected between nodes and ground. This is a model that can take the heat propagation into account, so it is a physical model. There is another model, based on the Foster network, which is easier to use but it is not a physical model.

What happens if the package is mounted on a heatsink?

If so, we need to develop the Cauer network so to include the thermal properties of the interface and the thermal capacitance of the sinker itself, so we are adding two additional RC cells.



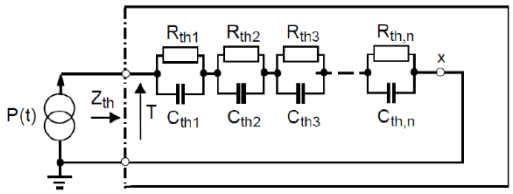
$\tau_{th\,heat\,sink} \approx 10^2$ s for natural convection

$\tau_{th\,heat\,sink} \approx 10$ s for forced convection

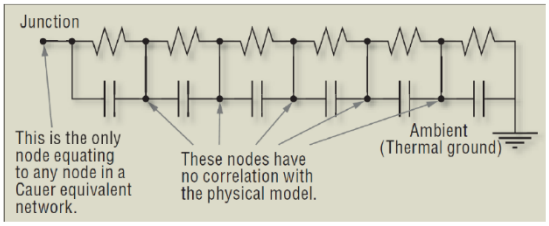
We can notice that the thermal time constant of the heatsink is huge, despite the use of natural or forced convection. It is in the order of seconds, while the package time constant is

in the order of ms. This is useful because it allows us to separate the two networks in the calculations of the junction temperatures.

Foster network



- The most frequently used thermal equivalent circuit diagram is mathematically very simple to use but as a purely formal description has no relation to the real physical structure.

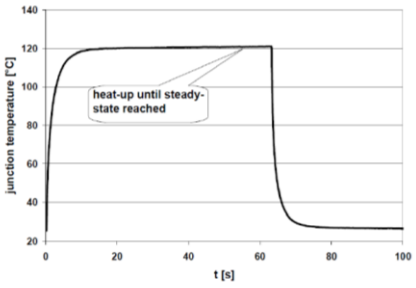


In this case we have a sequence of thermal cells in sequence. The Foster network is non-physical, i.e. if we calculate the temperatures in the intermediate points, those temperatures are not related with the effective physical temperatures. However, if we compute in a proper way Rth1, Cth1 and so on, we get a model able to correctly predict the junction's temperatures. So it is a model that allows only to compute the junction temperature, not the internal temperatures.
Since we are not always interested in the internal temperatures, this model can be helpful, in fact it is a sequence of RC cells and hence the total impedance is just the summation of the impedances of the single cells, and the response in the time domain is the superposition of exponential responses.

THERMAL IMPEDANCE

Development of the Z_{th} Diagram

- The device is mounted on an infinite (water cooled) heat sink to fix the case temperature to a known value.
- A step pulse of heating power, P, is introduced at the junction.
- T_J will rise to some steady state value dependent upon the response of the thermal network (heating curve).
- When the power input is terminated, T_J will decrease back to its initial value (cooling curve).
- The heating curve or the cooling curve can be used to determine $Z_{th}(t)$ for devices.



The thermal impedance can be defined in an operative way. We consider a device mounted on an infinite heatsink that keeps the case temperature fixed, typically at 25°C. Then we apply a step power pulse, turning on the mosfet, so the mosfet is dissipating a constant power. The junction temperature will increase in some ways reaching the final steady state temperature (plot on the right).

The thermal impedance is defined as below.

- The heating curve thermal impedance is

$$Z_{thJCh}(t) = \frac{T_J(t) - T_c}{P}$$

- The cooling curve thermal impedance is

$$Z_{thJCc}(t) = \frac{T_{Js} - T_J(t)}{P}$$

where

- $T_J(t)$ is the junction temperature of the device;
- T_{Js} is the junction temperature of the device at stedy state;
- T_C is the case temperature

Note: $\boxed{Z_{thJC}(t \to \infty) = R_{thJC}}$

$\boxed{Z_{thJC}(t) = r(t) \cdot R_{thJC}}$ where $\boxed{r(t) = \frac{Z_{thJC}(t)}{R_{thJC}}}$

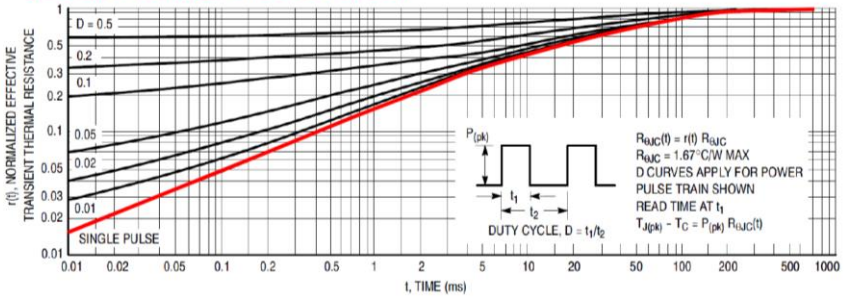
The thermal impedance between the junction and the case as a function of time is the junction temperature function of time minus the case temperature divided by the power we are dissipating in the device (cooling curve formula is not important).

NB: for t → inf, the junction-case thermal impedance becomes equal to the thermal resistance. This because for t that tends to infinite the system reaches the steady state. The thermal impedance is a 'transient quantity'.

Moreover, in general the thermal impedance is quoted in the datasheets, but typically we won't find the thermal impedance itself, but the normalized one. The normalized thermal impedance is the ratio between the thermal impedance and the thermal resistance between junction and case.

Datasheet curves

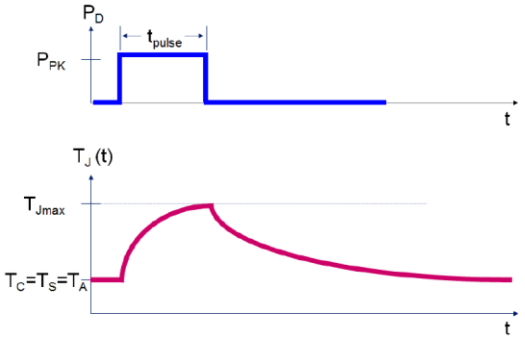
- Transient thermal response curves for power transistors have been around for a couple of decades.
- One can develop a thermal R–C network from a transient thermal response curve.
- These curves are perfectly adequate for analysis — if the power transistor's case temperature is known and the power dissipation waveform is relatively simple (e.g. single rectangular power pulse or periodic rectangular power pulses).



We are interested in the red curve, which is exactly the normalized thermal impedance. The behaviour of the curve is exactly the behaviour of the junction temperature upon the application of a constant power inside the device. We can use the red curve to develop an equivalent thermal model either with a Cauer or Foster network. This curve is perfect for the thermal analysis if the case temperature is known and the power waveform is relatively simple (rect pulse or train of rect pulses).

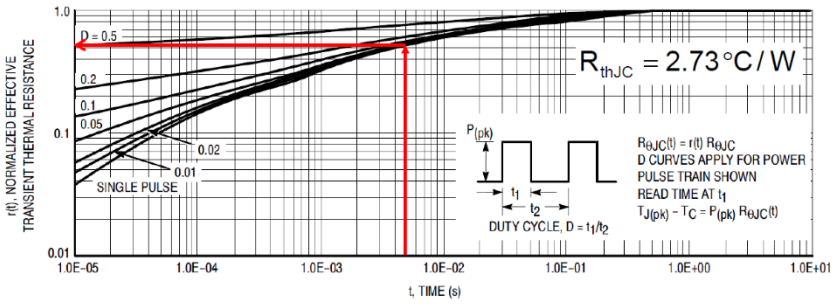
Example of transient thermal analysis – single pulse

- A MTP15N06V (a 15 A, 60 V MOSFET) is driven by a single 5ms, 50W pulse as shown below.
- Assume that the heat sink is large enough such that the case temperature does not have time to rise substantially within 5 ms. Ambient temperature is 40°C.
- Determine T_{Jmax}.



The system is subjected to a single rectangular power pulse with a duration of 5 ms and amplitude of 50W. The mosfet is mounted on a heatsink (whose time constant is pretty large). Since the power pulse has a duration of ms, the temperature of the heatsink doesn't change, we filter out completely the power pulse, so the only change in temperature occurs in the junction.

Ambient temperature is at 40°C, that is the external temperature of the sinker. The temperature of the junction starts from the temperature value of the case, it increases, reaches a maximum at the end of the power pulse and if the power pulse is off it will reduce down to the steady state value, so the ambient temperature. We are interested in the maximum junction temperature.
To compute it we use the single pulse curve. We read the duration of the pulse, we set it on the x axis, we move up vertically until we cross the single pulse curve and then we read the normalized thermal impedance.



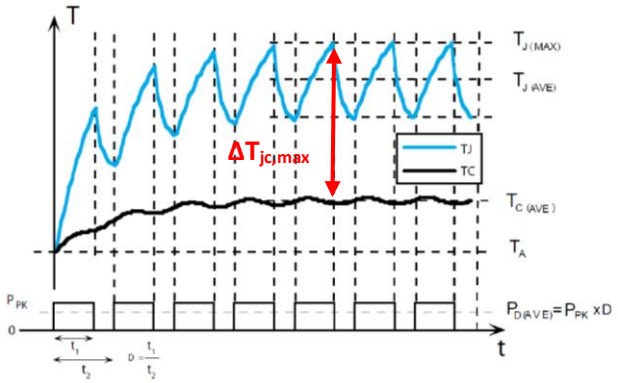
- We can read r(t) = 0.5 at = 5 ms on the transient thermal impedance curve (single pulse).

  Therefore:

$$T_{J_{max}} = T_C + \Delta T_{JC} = T_C + P_{PK} \cdot r(5ms) \cdot R_{thJC} = 108.25\ °C$$

The maximum junction temperature is the case temperature plus the delta temperature. The delta temperature is the peak power dissipated multiplied by the thermal impedance, which is the product of the normalized thermal impedance calculated at the end of the pulse, where the temperature will be maximum, multiplied by the thermal resistance.

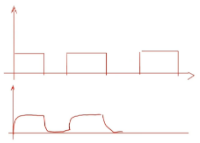
Thermal transient analysis of periodic pulses



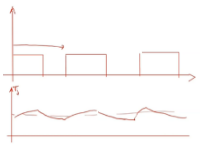
- If t_2 is much greater than the thermal time constants of the device than we have a situation similar to the "single pulse" case.
- Conversely, if t_2 is comparable to the thermal time constants of the device, the junction temperature of the device will vary as shown above.
- Because of thermal capacitance, the heat sink only respond to average power.

The system is subject to a train of identical rectangular pulses with an amplitude of Ppk, duration of t1 and period of t2. When we have repeated identical power pulses we can distinguish 3 different situations:

1. **The duration of the pulses is much larger than the duration of the time constants**. In this case the temperature goes up, reaches a steady state and then goes down, so we can treat the system like a system subjected to a train of isolated pulses, because two adjacent pulses don't interact.



2. **The distance between two adjacent pulses is much shorter than the time constant of the system**. In this case the system is not able to follow the instantaneous value of power dissipation, it is only sensitive to the average of the pulses, so after the initial transient we will have something moving around an average. If the duration of the single pulse is much shorter than the time constant, the amplitude of the ripple is negligible, so we can understand the junction temperature just by considering the average power that is dissipated.



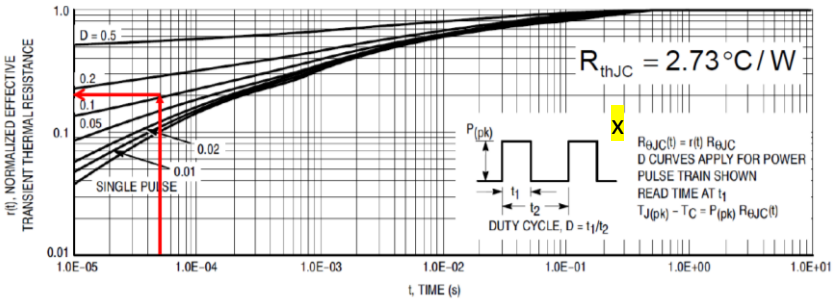
When we are in an intermediate situation, i.e. the duration of the pulses is comparable with the thermal time constant of the package, the temperature of the junction will follow the blue line of the image above. Hence the temperature will swing between a minimum and a maximum. If we consider just the average power we would make an error. We need to understand how to compute the peak value in this situation.

NB: the black curve is the temperature of the case. Apart from the initial transient, it reaches a steady state where temperature moves around a steady state value with practically no ripple. This happens because the thermal time constant of the heatsink is huge with respect to the times of the power pulses, so the heatsink, and as a consequence the case that is connected to it, it's only sensitive to the average power, because it cannot follow the fast variations of the injected power.
So we are in a situation where the case is sensitive to the average power, but the junction temperature may swing up and down around the average value.

To compute the peak value of the junction temperature, we are interested in computing the red delta after we have computed the case temperature. The delta can be easily computed using the thermal impedance curves previously introduced.

The one below is an example.

- A MTP15N06V is driven by repetitive pulses at 2 kHz frequency and 10% duty cycle (t_1 = 50 µs, t_2 = 0.5 ms).
- The peak power dissipation P_{PK} is 80 W.
- We can use the average power and use only the thermal resistances to calculate the case temperature T_C.
- Ambient temperature is 40°C; R_{thCS}=1.2°C/W; R_{thSA}=2.5°C/W
- Determine T_{Jmax} at the end of a pulse under steady–state conditions



Since we know the peak power value (Ppk) and the duty cycle of the train of pulses, we can compute the average power dissipated in the system. Since we know the thermal resistance and the temperature of the ambient, we can also compute the case temperature. The second step will be to calculate the delta(T_jc,max).

The first calculation is the one below. The case temperature is the ambient temperature plus the temperature drop between ambient and the case, which is the average power times the sum of the thermal resistances (resistances and not impedances because we are considering an average situation).

- The following equations apply:

$$T_C = T_A + P_{AV} \cdot (R_{thCS} + R_{thSA}) = T_A + P_{PK} \cdot D \cdot (R_{thCS} + R_{thSA})$$

$$T_{J_{max}} = T_C + \Delta T_{JCmax} = T_C + P_{PK} \cdot r(t_1, D) \cdot R_{thJC}$$

- We can read $r(t_1, D) = 0.2$ at t_1 = 50 µs, D=10% on the transient thermal impedance curve (10% curve).

  Therefore:

$$T_C = T_A + P_{PK} \cdot D \cdot (R_{thCS} + R_{thSA}) \approx 70°C$$

$$T_{J_{max}} = T_C + P_{PK} \cdot r(t_1, D) \cdot R_{thJC} = 113°C$$

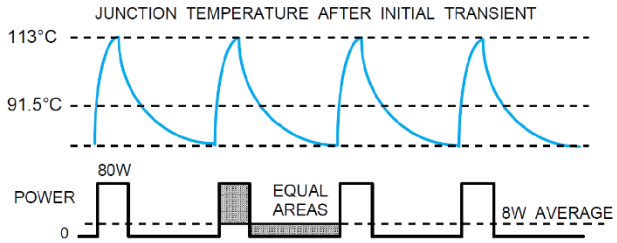
- Note:

$$T_{J_{av}} = T_A + P_{PK} \cdot D \cdot (R_{thJC} + R_{thCS} + R_{thSA}) \approx 91.5°C$$

Known the case temperature, we want to understand the maximum delta T between case and junction. To get this, we use the thermal impedance curves above. In this case we don't have to use the single pulse curve, but the other ones, in particular the one that corresponds to the duty cycle we are operating (10%, i.e. 0.1). From the plot we get the normalize thermal impedance as a function of time and duty cycle. Then, delta(T_jc,max) = T_j,max – Tc, so we can solve for T_j,max, since delta(T_jc,max) has been computed starting from the thermal impedance curves.

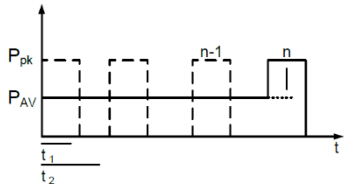
If we simply calculate the average junction temperature (red box), we get a smaller value, and the difference is almost 20°C between the average and the peak.

Moreover, if we look at the thermal impedance plot we have the insert x. This insert is a help to the user to compute the delta(T_jc,max) for a train of identical pulses.



DERIVATION OF r(t,D) FROM A SINGLE PULSE

We can derive the normalized thermal impedance from a single pulse in the following way.



- Pulses previous to the n^{th} pulse are averaged. Temperature due to the n^{th} pulse is then calculated and combined properly with the average temperature.

$$\Delta T_{JC_{max}\,n^{th}} = \Delta T_{JC_{av}} + \Delta T_{JCn^{th}} = P_{av}\cdot R_{thJC} + (P_{PK} - P_{av})\cdot r(t_1)\cdot R_{thJC}$$
$$= \left[D + (1-D)\cdot r(t_1)\right]\cdot P_{PK}\cdot R_{thJC} = r(t_1,D)\cdot P_{PK}\cdot R_{thJC}$$

$$\hookrightarrow \boxed{r(t_1,D) = \left[D + (1-D)\cdot r(t_1)\right]}$$

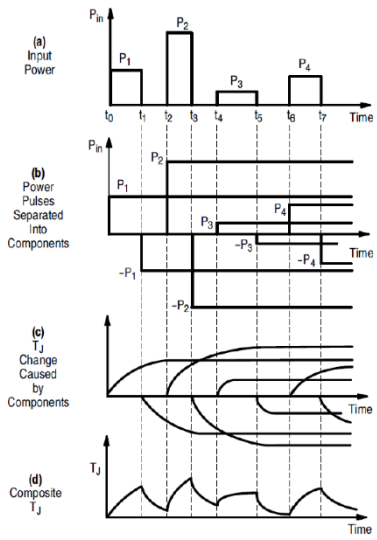
We consider a train of identical pulses of duration t1 and period t2, so D = t1/t2 and we want to consider the maximum junction temperature at the end of the generic n power pulse.
To do so, we can lump the previous power pulses that are occurring before the n power pulse we are interested in in a single power pulse (or step) having an height equal to the average power, as in the image. In order to compute the temperature at the end of the n power pulse we use the single pulse formula, so we consider the n-th power pulse as a single power pulse but we use the single power pulse formula considering a rectangular pulse having an amplitude equal to the delta(P), where delta(P) is the difference between the peak power of the n-th power pulse and the average power of the previous lumped power pulses.

Once again, we are assuming the case temperature is flat and known. We have to calculate the delta(T_jc,max) at the n-th power pulse, and this can be done considering two contributions. One is due to the average power, the other is the term computing using the single power pulse formula for an amplitude delta(P) and duration t1.

NB: in all these calculations the only assumption we have to make is that the case temperature is known.

ARBITRARY PULSE TRAINS WITH RANDOM VARIATIONS

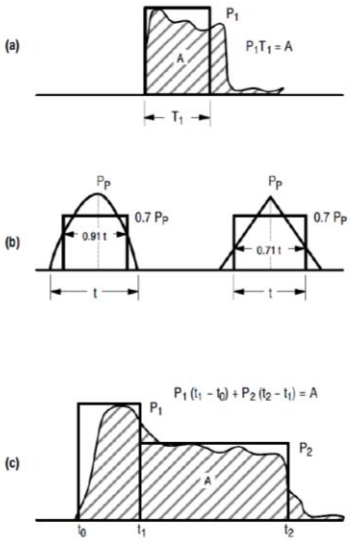
Thermal systems are linear systems, so we can always apply the principle of superposition of effects.
We have a sequence of thermal power pulses (a)) and we can split each pulse with a pair of step functions like in point b) made of a positive and a negative step (P1 and -P1, for instance). Now we know the thermal response of the system to a thermal variation of the power (c)). Then we sum all the responses and we get the d) plot.



- Application of superposition principle

NON-RECTANGULAR PULSES

In this case, in general, the simplest approach to slightly overestimate the temperature is approach a). We are approximating a non-rectangular pulse with a rectangular one having a peak value equal to the peak value of the original power pulse and an effective duration t_eff. t_eff is such that, once multiplied by the peak value, is equal to the energy of the real pulse.



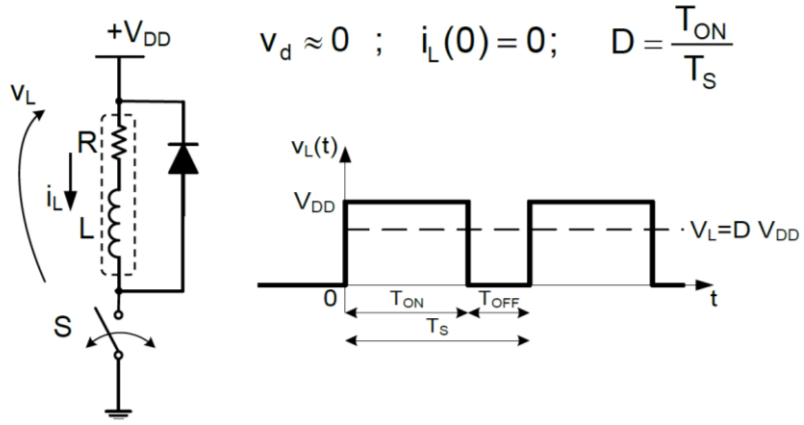
- For a pulse that is nearly rectangular, a pulse model having an amplitude equal to the peak of the actual pulse, with the width adjusted so the energies are equal, is a conservative model.

One typical situation is the exponential power pulse. In this case the typical approximation is to use a rectangular power pulse with a duration t_eff respecting the previous condition.

If we have a more complex shape, we can split the pulse in a superposition of rectangular pulses.

TUTORIAL 1

Inductive load

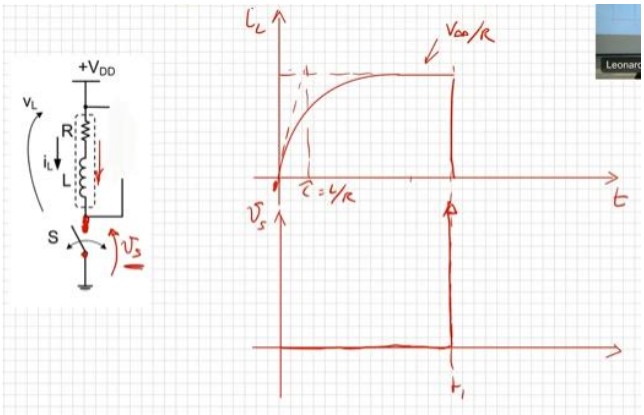


* *freewheel: a device in the transmission of a motor vehicle that automatically disengages the drive shaft whenever it begins to turn more rapidly than the engine.*

The inductive load is modelled with a resistor in series to an inductor, there is no ideal inductive load, they all have a series resistance. It is connected between PS and ground with an ideal switch. To the load it is applied a square wave voltage because the switch is turn on and off iteratively. Also, we assume the diode ideal (no voltage drop) and that there is no energy stored in the inductor.

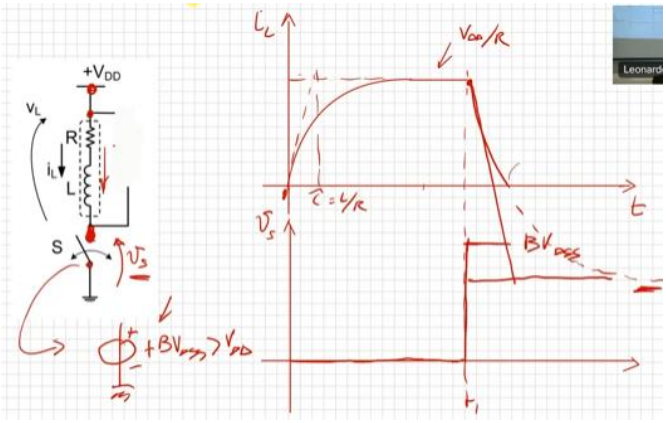
When the switch is closed we are applying Vdd to the load, when we open the switch the current is diverted into the diode, which is ideal. The diode is called **freewheeling diode** because it is used to allow the current to flow and avoid that the switch is damaged.
Let's suppose that at t = 0 the switch is closed, we keep it closed and then we open it. The initial inductor current is 0 and then we have an exponential growth with an exponential tau = L/R. The current reaches a steady state value of Vdd/R.

As for the voltage across the switch, if it is ideal, it stays equal to 0 as long as the switch is on. Then at t1 the switch is opened. If there was no diode, current was flowing in the inductor and it is pushed towards the high impedance node of the switch. The voltage rises until the dielectric strength of the air is reached. The air becomes conductive and we have plasma that emits photons. So the current drops quickly to 0.

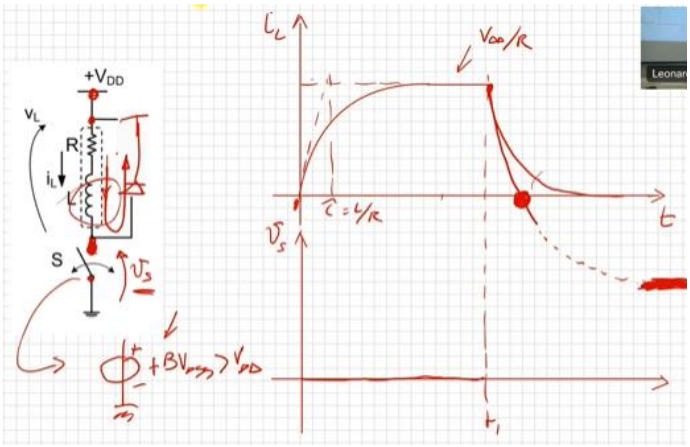


If instead of an ideal mechanical switch we have a power mosfet, when we turn off the switch the voltage increases fast as before until we reach the breakdown voltage BVdss. From this point on we can replace from an electrical point of view the switch with a constant voltage generator of +BVdss.

The load is now clamped between Vdd and BVdss, with the latter that is bigger, so we are reversing the voltage across the load and the inductive load is discharging.
The current decreases with an exponential discharge and we should ideally end up at negative currents.



Breakdown is however something we have to avoid except for a specific category of mosfet that can be operated in breakdown. But with a standard mosfet, to prevent it from damaging, we need a freewheeling diode. The current is redirected through the diode.
In this situation the current transient is still exponential, and the final value is 0 because we are shorting the inductive load.



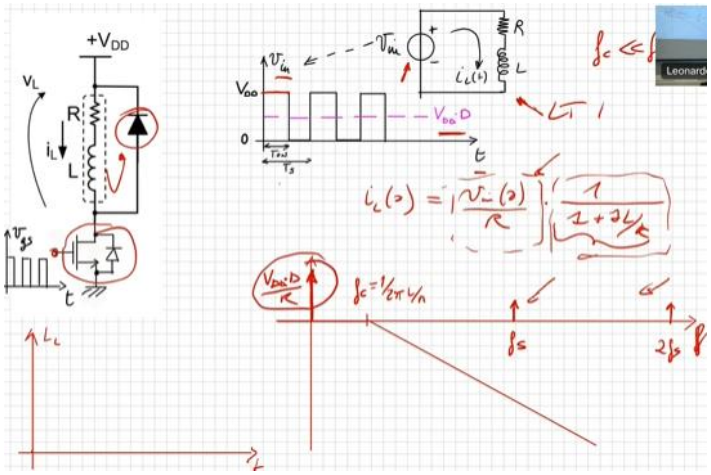
So the function of the freewheeling diode has been clarified. Now we want to understand what happens to the current in the inductive load when the switch is periodically turned on and off.
Again, in place of the switch we put a mosfet, so turning it on and off means that we are applying a Vgs. We can replace the diode with a voltage generator generating a time varying voltage Vin, which is either Vdd or 0V. The average voltage we are providing is Vdd*DC, where DC is the duty cycle. All this because it is a linear time invariant circuit, so we could also use the Laplace domain.

Basically we have a LP filter with a time constant L/R if we consider the circuit in the upper right.

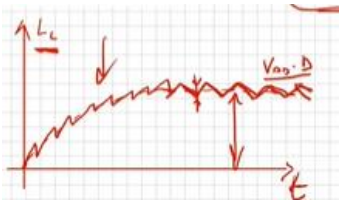
Then, Vin is a square voltage waveform that can be easily represented in the frequency domain because it is periodic; hence it will be the sum of sinusoidal signals at discrete frequencies.

The average value of the square waveform is not zero, so we have a DC component, then one at fs (switching frequency) and at all its multiple.
If fc << fs, the effect of the filter is to preserve the DC component and attenuate all the other harmonics.

What about the steady state of the current after the transient?



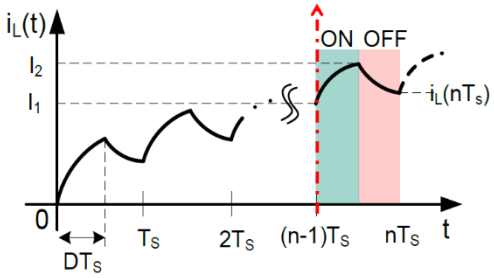
Firstly, the average value of the current at steady state will be Vdd*DC/R. But upon the average we see some current ripples. They come from the harmonics components that are not completely filtered out by the filter.
The relationship between the amplitude of the ripples and the cut-off frequency is that the larger fs with respect to fc the smaller the peak to peak ripple, because the larger the switching frequency the more the attenuation of the harmonics.



As for the transient, it is exponential with a ripple superimposed to it.

Let's consider a situation where the time constant L/R is comparable with the switching period (so fs is comparable with fc).
We want to compute the effective current during the transient and at the end of the transient.

The equation in the upper blue box is the voltage drop across the load, which is the sum of the two contributions.



$$v_L(t) = R \cdot i_L(t) + L\frac{di_L(t)}{dt}$$

During the switch-ON period:

$$L\frac{di_L(t_a)}{dt} + R \cdot i_L(t_a) = V_{DD} \quad \text{for } 0 \le t_a \le DT_S$$

Note: 't_a' is defined as the elapsed time during the ON period of a cycle.

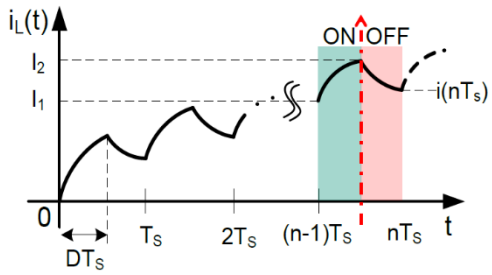
yielding $i_L(t_a) = \frac{V_{DD}}{R}(1 - e^{-\frac{t_a}{\tau}}) + I_1 \cdot e^{-\frac{t_a}{\tau}}$ for $0 \le t_a \le DT_S$)

i.e., $I_2 = \frac{V_{DD}}{R}(1 - e^{-\frac{DT_S}{\tau}}) + I_1 \cdot e^{-\frac{DT_S}{\tau}}$ at $t_a = DT_S$ (1)

We can analyze the transient focusing on a generic switching cycle, in this case from (n-1)Ts and Ts. The switching cycle is split in two parts, switch on and switch off. t_a is the time elapsed during the on time.
We can write the differential equation that describes the current evolution.
i_l(inf) = Vdd/R and i_l(0) = I1, where I1 is the current at the beginning of the switching cycle.

At the end of the on time period the current is described by the bottom blue box.

During the off time period the equation is still the same, the only difference is that the current is now directed to the freewheeling diode, so the voltage drop across the load is 0. Once again the time origin is moved. The solution is that i_1 drops exponentially and we can write the i_1 at the end of the generic switching cycle just by replacing t_b = (1 – DC)Ts.



$$v_L(t) = R \cdot i_L(t) + L\frac{di_L(t)}{dt}$$

During the switch-OFF period:

$$L\frac{di_L(t_b)}{dt} + R \cdot i_L(t_b) = 0 \quad \text{for } DT_S \le t_b \le T_S$$

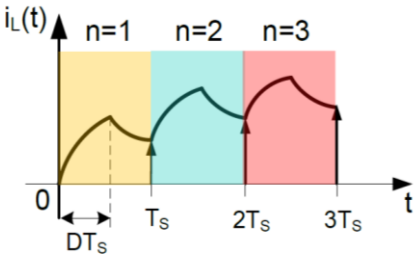
Note: 't_b' is defined as the elapsed time during the OFF period of a cycle.

which yields:

$$i_L(t_b) = I_2 \cdot e^{-\frac{t_b}{\tau}} \quad \text{for} \quad 0 \le t_b \le (1-D)T_S$$

i.e., $$i_L(nT_s) = I_2 \cdot e^{-\frac{(1-D)T_s}{\tau}} \quad \text{at} \quad t_b = (1-D)T_S \qquad (2)$$

Then, by combining equation 1 and equation 2 and replacing i1 with i_1 at the time instant (n-1)Ts.
In the end, the current at a generic n switching period is equal to the current at the end of the first switching period plus the current at the beginning of the generic n switching period.



NOTE:
Equation (3) represents a linear recursion between two successive values of the inductor current.

Combining eq.1 with eq.2 and replacing I_1 with $i_L((n-1)T_S)$ we get:

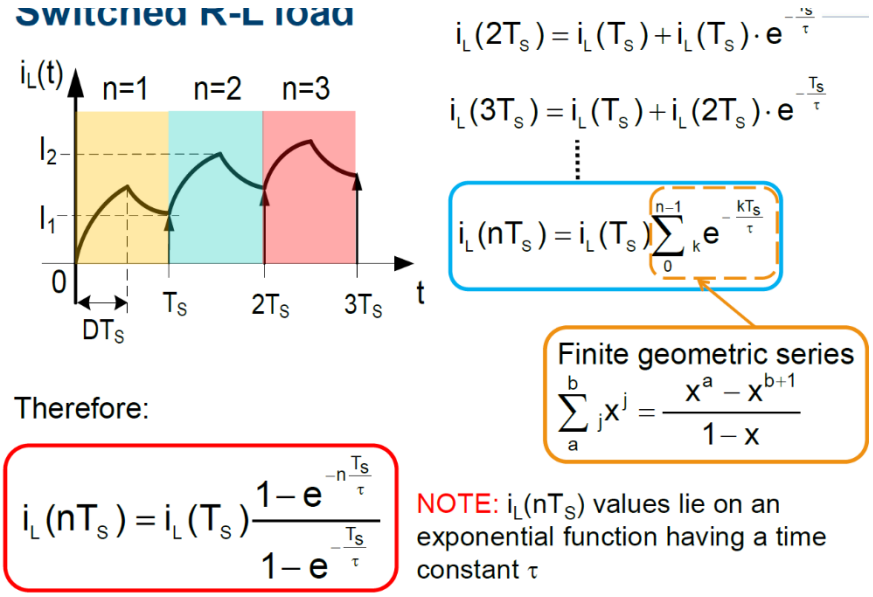
$$i_L(nT_S) = \frac{V_{DD}}{R}\left(1 - e^{-\frac{DT_S}{\tau}}\right) \cdot e^{-\frac{(1-D)T_S}{\tau}} + i_L((n-1)T_S) \cdot e^{-\frac{T_S}{\tau}}$$

that is,

$$i_L(nT_S) = i_L(T_S) + i_L((n-1)T_S) \cdot e^{-\frac{T_S}{\tau}} \qquad (3)$$

So equation 3 is a linear recursion between two successive values of the inductive current. We start by calculating the current at the end of the first switching cycle, we use this current in equation 3 and we compute the current at the end of the successive cycle and so on.

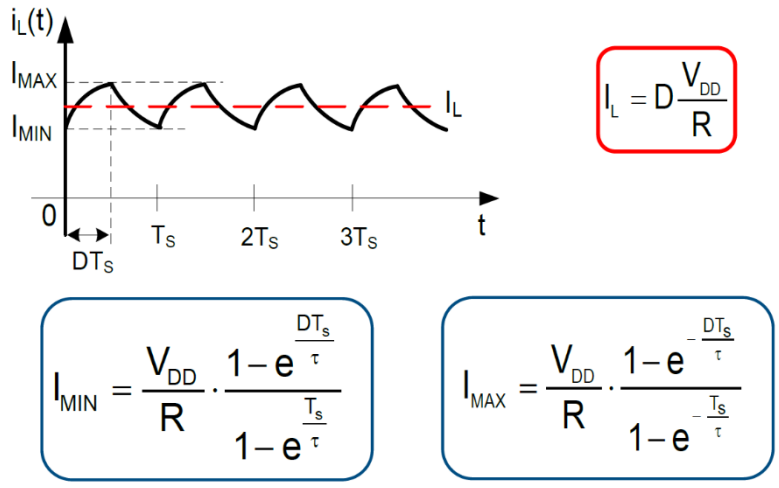
This is what is done in the next image.



The final more compact form is the one in the light blue box.
If we consider the finite geometric series up to n, it has a solution that it is the one in the yellow box. Of course, x = e^(-Ts/tau).
In the end, the current at the end of a switching period is the current at the end of the first switching period multiplied by a fraction (red box).
At the denominator we have a discrete exponential function that starts from 0 and runs up to a steady state value that can be computed with n = inf.
So the i_1 at a generic cycle lie on an exponential function whose time constant is tau = L/R.

NB: this equation provides the current at the end of the switching period, not at every time, but it is not a problem because in the switching period we know which is the shape of the current.

At steady state we reach the steady state value, a situation where the current starts from a minimum value, reaches the maximum one and then goes back to the initial value. The average value of the current (red box) is the one expected.



from eq.(3) with $I_{MIN} = i_L(nT_S)$ for $n \to \infty$

These results are typical of any first order system driven with an intermittent force, either electrical, thermal, mechanical ecc.

We want to understand what happens at steady state using some simple techniques.
The average inductor current at steady state can be easily calculated by considering:

- Properties of current and voltage waveforms in periodic steady state (PSS).
- Principle of inductor volt-second balance in PSS.
- KVL and KCL in PSS.

Steady state

- steady-state condition implies that voltage and current waveforms repeat with a time period T. Therefore:

$$v(t+T) = v(t)$$

and

$$i(t+T) = i(t)$$

Periodic steady state doesn't mean that voltage or current are constant, because that is the concept of DC. Steady state means that current and voltage are variable but they repeat with a period.

Principle of inductor volt-second balance

Inductor defining relation:

$$v_L(t) = L\frac{di_L(t)}{dt},$$

Integrate over one complete switching period:

$$i_L(t_o+T) - i_L(t_o) = 0 = \frac{1}{L}\int_{t_o}^{t_o+T} v_L(t)dt$$

In periodic steady state (PSS), the net change in inductor current is zero:

$$0 = \int_{t_o}^{t_o+T} v_L(t)dt$$

Hence, the total area (or flux linkage) under the inductor voltage waveform is zero whenever the converter operates in PSS.
Equivalently:

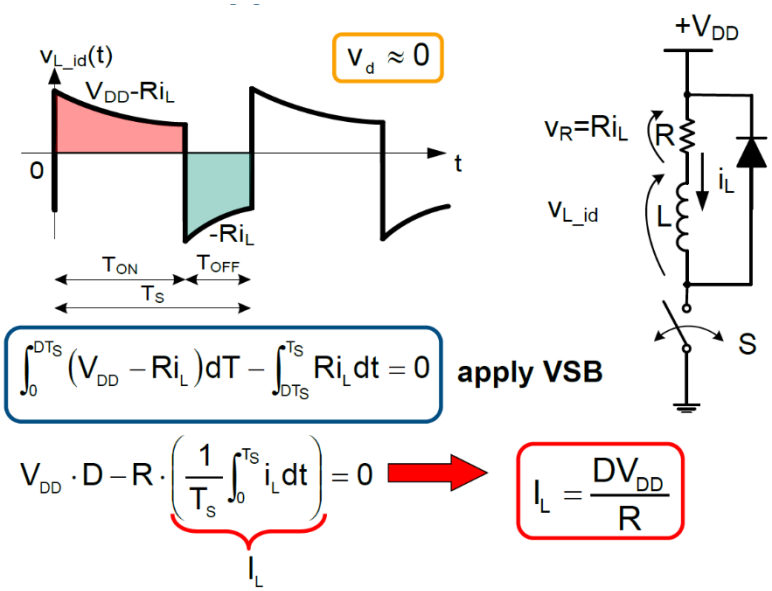
$$0 = \frac{1}{T}\int_{t_o}^{t_o+T} v_L(t)dt = \langle v_L(t)\rangle$$

The average inductor voltage is zero in PSS (volt-second balance).

We start from the voltage-current relationship in an inductor (ideal inductor). Then we integrate over a switching period between the generic t0 and t0 + T.
So the integral is 0. We can rewrite it by multiplying everything by 1/T; if we do so, we get the average voltage across the inductor, that is zero at periodic steady state.
The average voltage across an inductor that is operated in periodic steady state is zero, and this is called volt-second balance.
This is the dual of the average current flowing through a capacitor operated in periodic steady state that is zero (ampere-second balance).

We can see the voltage second balance on the example we were considering before. Let's consider the voltage drop across the ideal part of the inductive load and the one below is its plot. When the switch is

on, the voltage drop across the ideal component is Vdd minus the voltage across the resistor. Since the current is increasing during the on time period, the voltage drop is decreasing.
When we turn off the switch , the voltage drop across the ideal part of the inductor changes its sign and it is -i_L*R. once again, i_L is decreasing.
We apply the volt-second balance. The integral between 0 and Ts of the voltage across the inductor must be zero, we can split the integral in two parts, from 0 to DTs and from DTs to Ts. The first integral is the area during the on time period, the other during the off period. So the red area must be balance by the green area. The areas are the products of the voltage and time.



Just by applying the volt-second balance we get the value of the average current i_L (steady state average current).

KVL and KCL in periodic steady state

- Since KVL and KCL apply at any instant, they also apply to averages.
  Consider KVL:

$$\sum_{\text{Around loop}} v(t)=0, \qquad v_1(t)+v_2(t)+v_3(t)+\cdots+v_N(t)=0$$

$$\frac{1}{T}\int_{t_o}^{t_o+T}v_1(t)dt+\frac{1}{T}\int_{t_o}^{t_o+T}v_2(t)dt+\frac{1}{T}\int_{t_o}^{t_o+T}v_3(t)dt+\cdots+\frac{1}{T}\int_{t_o}^{t_o+T}v_N(t)dt=\frac{1}{T}\int_{t_o}^{t_o+T}(0)dt=0$$

$$\langle v_1\rangle+\langle v_2\rangle+\langle v_3\rangle+\cdots+\langle v_N\rangle=0$$

KVL applies to average voltages

The same argument applies to KCL:

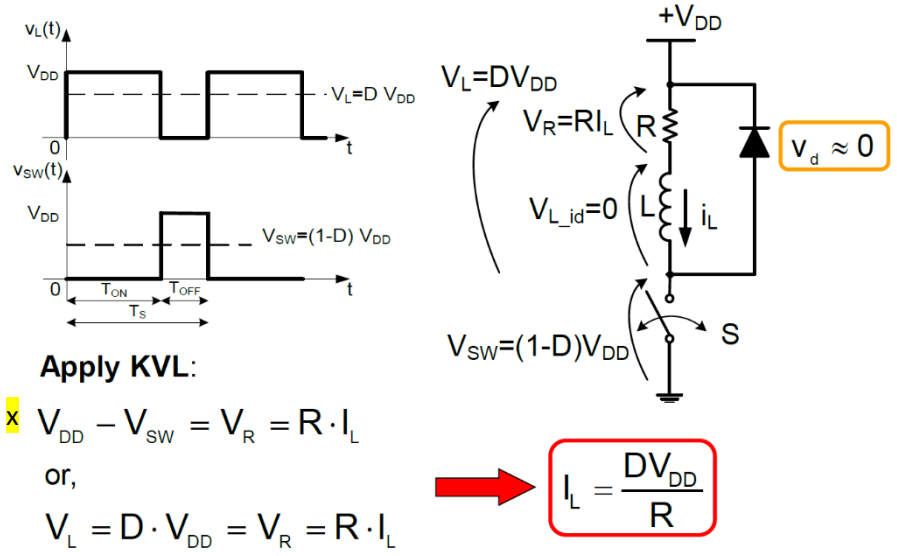
$$\sum_{\text{Into a node}} i(t)=0, \qquad i_1(t)+i_2(t)+i_3(t)+\cdots+i_N(t)=0$$

$$\langle i_1\rangle+\langle i_2\rangle+\langle i_3\rangle+\cdots+\langle i_N\rangle=0$$

KCL applies to avereage currents

We can rewrite the summation of voltages performing an integration from a generic t0 to t0+T because the integration operator is a linear operator. In the end we get the average voltages (assuming we are operating the circuit in periodic steady state) → the KVL applies also to the average voltages if the circuit is operated in periodic steady state.
The same conclusion can be drawn for the KCL.

KVL applied to averaged voltages in PSS



Apply KVL:

x $V_{DD} - V_{SW} = V_R = R \cdot I_L$

or,

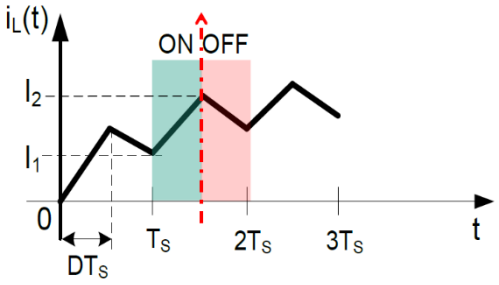
$V_L = D \cdot V_{DD} = V_R = R \cdot I_L$

$$I_L = \frac{DV_{DD}}{R}$$

The average voltage across the switch is: when the switch is closed the voltage drop is ideally zero, when the switch is open it is Vdd, so the average Vsw = (1 – DC)Vdd. So Vdd during the Toff and 0 during the Ton.

We can apply the KVL and we get equation x. Furthermore, the average voltage across the ideal inductive component is 0 by definition, so Vr = V_L and we get to the second equation.

SWITHCED R-L LOAD: SHORT Ts



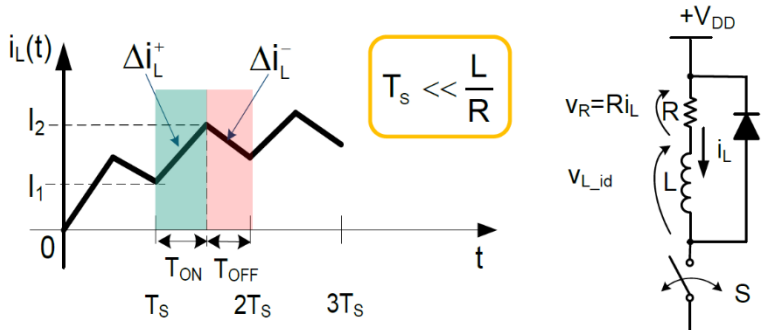
$$T_S << \frac{L}{R}$$

$$i_L(t_a) = \frac{V_{DD}}{R} \cdot \frac{t_a}{\tau} + I_1 \cdot (1 - \frac{t_a}{\tau}) = \text{for } 0 \le t_a \le DT_S \quad \text{(switch ON)}$$

$$i_L(t_b) = I_2 \cdot (1 - \frac{t_b}{\tau}) \text{ for for } \quad 0 \le t_b \le (1-D)T_S \qquad \text{(switch OFF)}$$

We are in a situation where the switching period is much smaller than the time constant L/R, that is the switching frequency is much larger than the corner frequency of the filter (fs >> fc). It is a situation that is very common in any power converter.

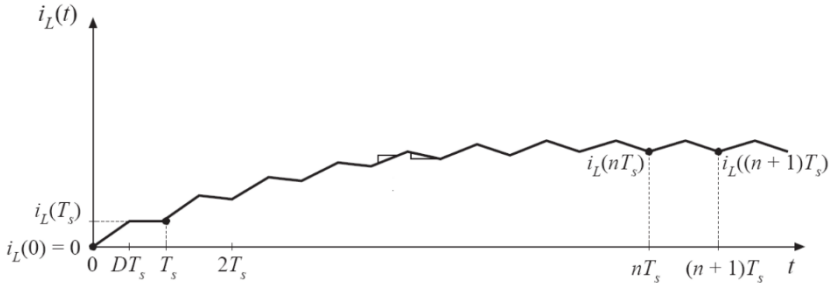
The only difference is that in place of pieces of exponentials, we can linearize and we have pieces of lines.



$$\Delta i_L^+ = \frac{V_{DD}}{R}\frac{T_{ON}}{\tau} - I_1\frac{T_{ON}}{\tau} = \frac{V_{DD} - RI_1}{L}T_{ON}$$

$$\Delta i_L^- = -I_2\frac{T_{OFF}}{\tau} = -\frac{RI_2}{L}T_{OFF}$$

The one below is the turn on transient, in a situation where the switching period is much slower than the time constant; the current increases more or less exponentially until we get a steady state. At steady state the ripple is a linear one.



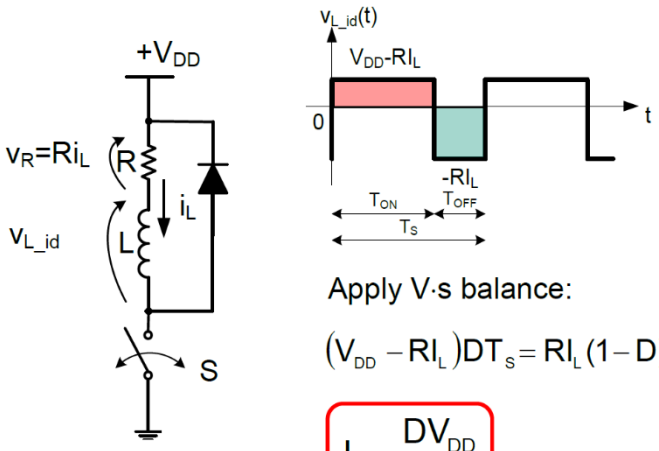
When the converter operates in equilibrium:

$$i_L((n+1)T_s) = i_L(nT_s)$$

i.e., $\Delta i_L^+ = -\Delta i_L^-$

If Ts is much smaller than L/R, then delta_i_L is much smaller with respect to the average current, so we can replace the instantaneous current with the average one, because the instantaneous current is very close to the average one.

If $T_S << \tau$ then $\Delta i_L << I_L$ i.e., $I_{MAX} \approx I_{MIN} \approx I_L$ (small ripple approximation)



$v_d \approx 0$

$i_L \approx I_L$

Apply V·s balance:

$$(V_{DD} - RI_L)DT_S = RI_L(1-D)T_S$$

$$I_L = \frac{DV_{DD}}{R}$$

This is the so called small ripple approximation, the amplitude of the ripple is negligible with respect to the DC component of the current.
With respect to the previous calculations, if we replace i_L with the average current, the voltage Vl_id is flat, and so the volt-second balance is the balance between the area of two rectangles. The result we get is the same, but we don't have to play with integrals because the voltage waveforms are square waveforms.

We can also calculate the peak to peak ripple, which is important because we need to check that the ripple is negligible. This is done calculating the current variation in the on and off time periods.

Under small ripple approximation, i.e. $I_{MAX} \approx I_{MIN} \approx I_L$

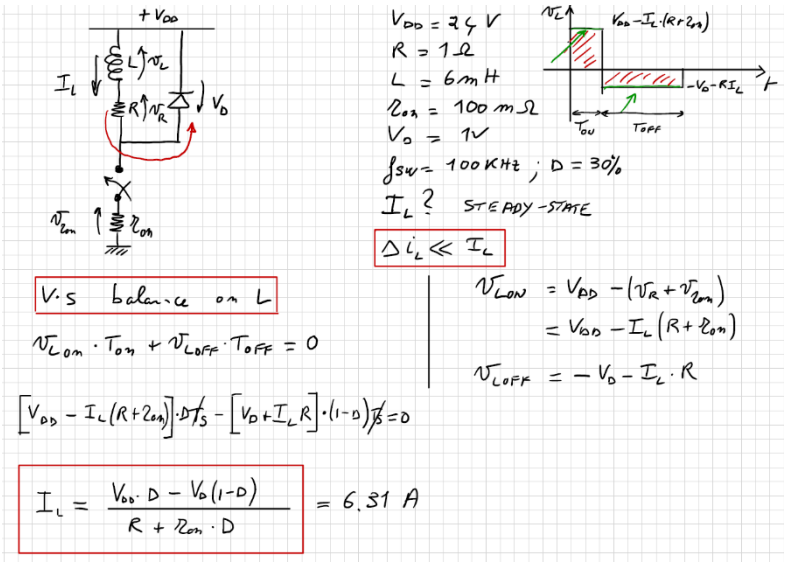
$$\Delta i_L^+ = \frac{V_{DD} - RI_L}{L} DT_S$$

$$\Delta i_L^- = -\frac{RI_L}{L}(1-D)T_S$$

$$\boxed{\Delta i_L^+ = -\Delta i_L^-}$$

In general, in any situation where the ripple is linear we can apply the small ripple approximation.

Example

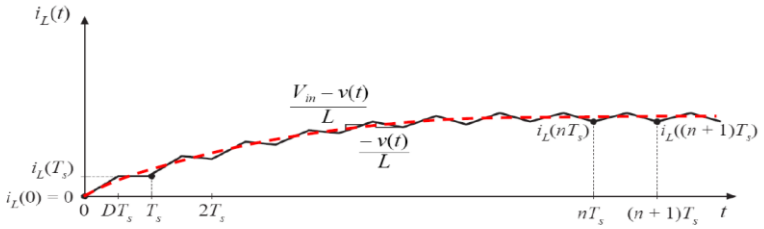


We can get to the same result applying the KVL applied to the average voltage.

Average model of the turn on transient

In the previous calculations we were interested in the steady state calculation. Turn on transient is a current increasing in an exponential fashion. Upon this exponential increase we have the ripple. When we want to understand the dynamical behaviour of the circuit we are not interested in the ripple, the exponential is the important part, the underline average behaviour. We want to get rid of the ripple.

This can be done by applying a moving average.



$$\langle i_L(t)\rangle_{T_S} = \frac{1}{T_S}\int_t^{t+T_S} i_L(\tau)d\tau$$

- Ignore the switching ripple
- Inductor current is averaged over one switching period (running average)
- Switching harmonics are removed (see part 6)

→ This model replicate the average behavior of the system state

State-space averaging

A state space is a space having n dimensions where n is the number of state variables of the systems. The state variables of the system are related to the energy we can associate to any independent conservative element → independent inductors and capacitors.
In our case we have just a single state variable, which is the current in the inductor, associated with the energy in the inductor.

The state space equation has to be split in to parts: one when the switch is closed, the other one when the switch is open (blue system of equations).

$$L\frac{di_L(t)}{dt} = -R\cdot i_L(t) + V_{DD} \qquad \text{for } 0 \le t \le DT_S$$
$$L\frac{di_L(t)}{dt} = -R\cdot i_L(t) \qquad \text{for } DT_S \le t \le T_S$$

$$L\frac{di_L(t)}{dt} = -R\cdot i_L(t) + q(t)\cdot V_{DD}$$

q(t) = switching function
- $q(t) = 1$ for $0 \le t \le DT_S$
- $q(t) = 0$ for $DT_S \le t \le T_S$

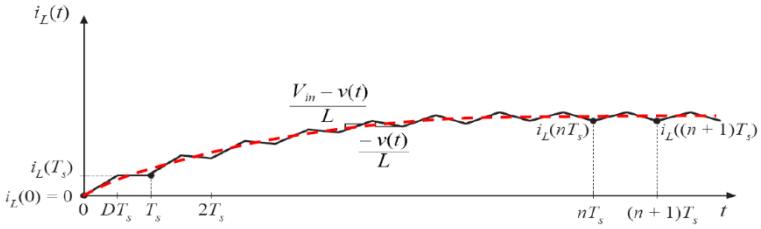
By averaging (running average) the differential equation we get:

$$L\frac{d\langle i_L(t)\rangle_{T_S}}{dt} = -R\cdot\langle i_L(t)\rangle_{T_S} + D\cdot V_{DD} \quad \text{with} \quad d(t) = \frac{1}{T_S}\int_t^{t+T_S} q(\tau)d\tau = D \quad \text{for } t \ge 0$$

At turn-on, the average inductor current increases exponentially

We can write the two spaces equations in a more compact form by introducing the switching function, which is a two levels function equal to 1 in the on time period and zero in the off time period.
Then to the differential equation in the blue box we apply the running average. In the end we get an average state space equation of the red box. In this expression, the term D*Vdd is the running average of the switching function, also called **duty cycle function** (in our case it is a constant value because the DC is constant).

In the end we get a first order equation where the variable is the running average current, that is the red dashed line of the next image. The equation in the red box provides both the behaviour during the transient, as far as the average current is considered, and the steady state value.



$$\langle i_L(t)\rangle_{T_S} = \overline{i}_L(t) = \frac{V_{DD}D}{R}\left(1 - e^{-\frac{t}{\tau}}\right) \quad \text{with} \quad \tau = \frac{L}{R} >> T_s$$

→ we'll see this in detail in part 6…

x x