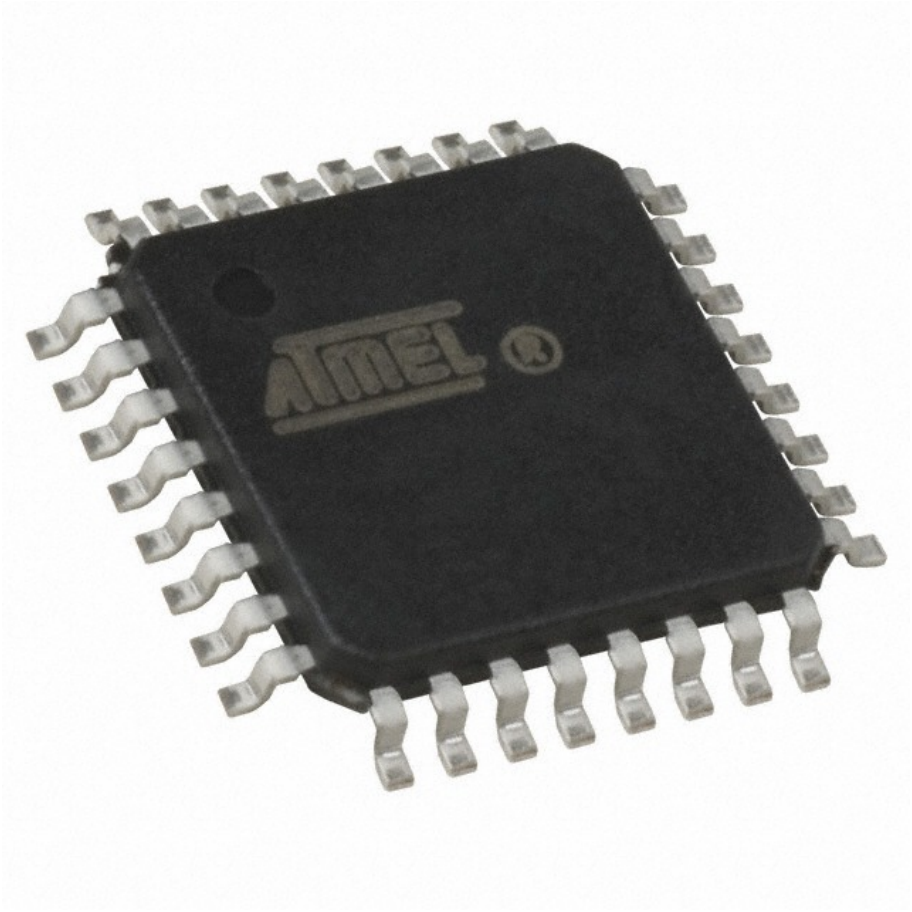
DR. MICHAEL P. HAYES

ENCE461

EMBEDDED SYSTEMS II LECTURE NOTES

HIGH-SPEED DIGITAL DESIGN V62



ELECTRICAL AND COMPUTER ENGINEERING

Contents

0 Introduction

So you want to build an embedded system? And you want to make the hardware reliable? Oh, and you want to be able to meet EMC[1] regulations so you can sell it without affecting someone's pacemaker? Unfortunately, a consequence of improved integrated circuit technology is that they are much more complicated, switch much faster, and use many different voltage levels. So it is not just a case of choosing your microcontroller and then connecting inputs to outputs. We have to worry about the parasitic components not shown on the schematic. We also need to understand the advanced peripherals of modern microcontrollers and consider voltage levels, timing requirements, and how to lay out a printed circuit board to make them work reliably.

[1] Electromagnetic compatibility

Once upon a time, there was digital electronics and there was analogue electronics. But, somewhere along the line the distinction got blurred. While digital signals have two states, in practice, they need to be treated as analogue signals. This is the focus of my lectures. The topics I cover are:

PIO ports Most MCU pins are input/output ports (PIO) a.k.a. GPIO. This lecture describes how they work; hopefully this should be a refresher. But there are some gnarly aspects if we want to minimise power consumption or avoid blowing things up on power-up.

The CMOS inverter PIO pins on MCUs are internally connected to CMOS inverters. Thus understanding the CMOS inverter is important when interfacing with PIO pins.

CMOS inputs CMOS has a high impedance and is thus susceptible to electrostatic discharge (ESD). Manfacturerers add protection diodes but these can lead to unexpected behaviour.

CMOS outputs Conceptually the CMOS inverter is just two switches, but these switches have non-linear behaviour. This lecture introduces some simple models.

Logic levels Here we look at noise margins and how to deal with logic running on different voltages.

Logic switching Modern logic devices switch rapidly and we find that we need better I/O models to deal with parasitic components.

Common path noise In which we find that pesky parasitic inductance gets in the way of fast changing signals.

Power supply decoupling Here capacitors come to the rescue to deal with the pesky inductance but we find that capacitors are far from ideal and that there are several types to choose from.

Crosstalk This is where those parasitic circuit components not shown on the schematic cause signals to interfere with each other.

PCB stackup In which we look at how a multilayer PCB is constructed and how we use the different layers wisely.

Transmission lines Hopefully this is a refresher to remind you that wires are far from ideal and that signals take time to propagate (even short distances across a PCB).

Transmission line reflections Here we consider what happens when we do not correctly terminate a transmission line.

Transmission line termination There are many ways to terminate a transmission line; here we look at a number of common techniques.

Differential signals When we want super fast signals we use differential signalling. This helps to cancel the crosstalk but what happens if there is timing skew?

Clock generation Microprocessors all need a clock but most run much faster than common crystal frequencies. Phase locked loops come to the rescue since they can multiply the frequency of a clock. We can also add some jitter to reduce spectral peaks that violate EMC compliance.

1 Microcontrollers

Billions of microcontrollers (MCUs) are sold each year. There are dozen main manufacturers producing a wide variety from low-performance 8-bit microcontrollers to powerful 64-bit systems on chip (SOC).

1 Packaging

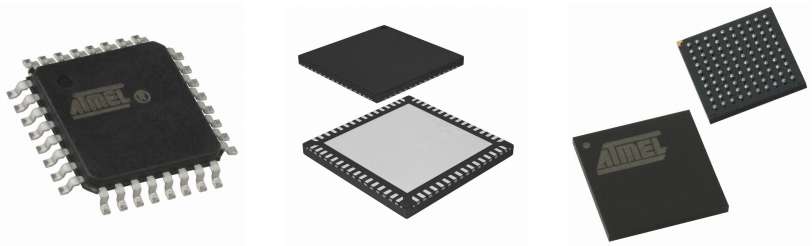
MCUs and other integrated circuits come in a wide-variety of packaging. The common packages (see Figure 1.1) are:

Quad flat pack (QFP) This package is popular for prototyping since the pins can be hand soldered. At a pinch, the pins can be probed with a scope[1].

Quad flat no leads (QFN) These are difficult to hand solder. However, they have a smaller package and smaller lead inductance.

Ball grid array (BGA) This provides the highest pin density and smallest lead inductance with a matrix of solder blobs on the bottom. These packages are difficult for prototyping since the pins cannot be probed with a scope or unsoldered. X-ray imaging is required to determine if the pins are properly soldered.

[1] It is better to provide test points.



(TQFP) (QFN) (BGA)

Figure 1.1: (TQFP) Thin quad flat pack; (QFN) Quad flat no-leads; (BGA) Ball grid array.

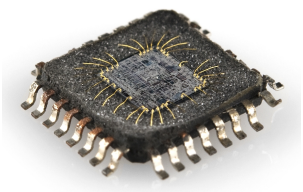


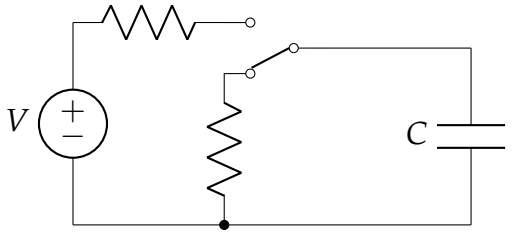
Figure 1.2: Bond wires interconnects between die and package pins.

10 Pins

Most MCU pins are input/output pins called PIO or GPIO. These are shared with peripheral signals (e.g., SPI, PWM, I2C, ADC, etc.). The second largest category are power supply pins. Finally, there are miscellaneous pins for:

- programming/debugging
- clocks
- reset

11 Power supplies



Whenever a digital signal switches high, some capacitance needs to be charged. If a capacitance C is charged to a voltage V, then the energy stored in the capacitor is

$$E = \frac{1}{2}CV^2. \tag{1.1}$$

However, in the process of charging, an equivalent amount of energy is dissipated in the resistance of the driving transistors[2]. When the digital signal switches low, the same amount of energy is dissipated in the driving transistors when discharging the capacitance. Thus if the signal switches at a frequency f, the power dissipated is

$$P = fCV^2. \tag{1.2}$$

[2] No matter how good the transistors are.

An MCU can have many thousands of interconnected transistors switching at once and so the total power consumption can be large. All the MCU designer can do to reduce power consumption, is to reduce the number of transistors, reduce the capacitance by making the transistors smaller, or by reducing the supply voltage. The latter strategy has the most effect due to the square of the voltage, but this will reduce the noise margins for external signals.

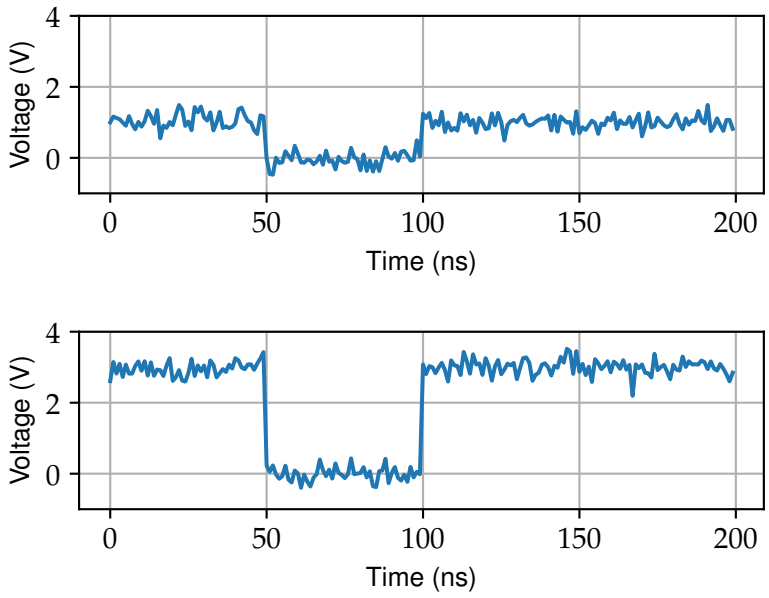


Figure 1.3: Example MCU output signals. Which do you think would be more reliable?

To minimise power dissipation and to ensure a good noise margin for external signals, is common to have two or more power supply connections:

VDDCORE to power all the internal logic: CPU, peripherals, memory, etc.

VDDIO to power all the I/O pin drivers.

Most MCUs require multiple VDDCORE[3] and VDDIO pins as well as associated grounds for signal integrity reasons.

[3] VDD stands for voltage to drains but for CMOS this is a misnomer.

11.1 *Power supply naming conventions*

VDD Voltage to drains; positive supply for CMOS (and NMOS) circuits.

VSS Voltage to sources; ground connection for CMOS circuits.

VCC Voltage to collectors[4]; positive supply for BJT circuits.

VKK Voltage to cathodes (for valves/tubes).

[4] Old habits die hard so VCC is often used in CMOS device datasheets.

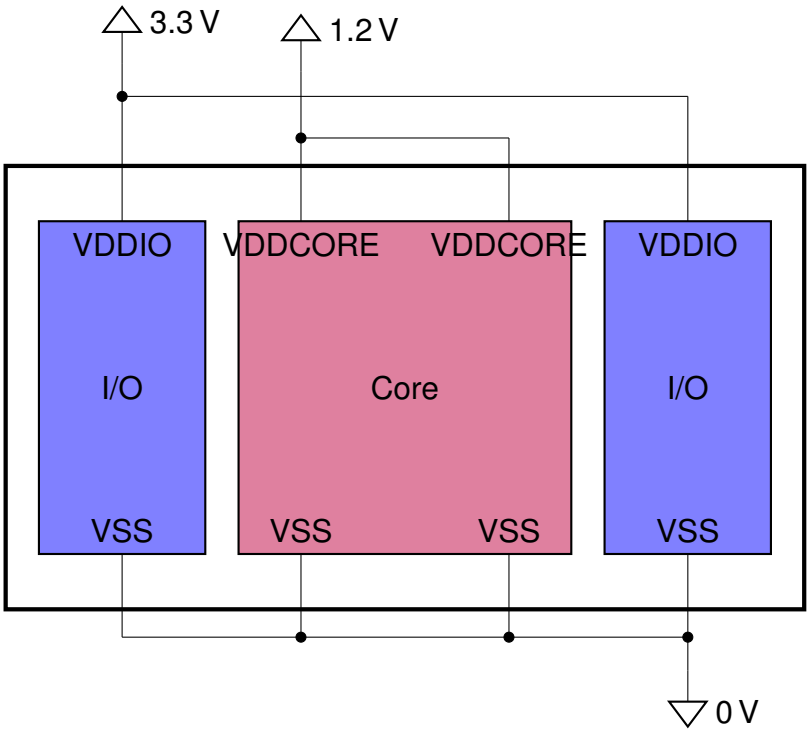


Figure 1.4: Typical power supply connections for a MCU. In practice, there are also separate power pins for analogue components such as ADCs and clock oscillators.

100 Clocking

All MCUs require a clock. For frequency stability a crystal oscillator is used but this requires an external crystal. Otherwise an RC oscillator is used, see Figure 1.5.

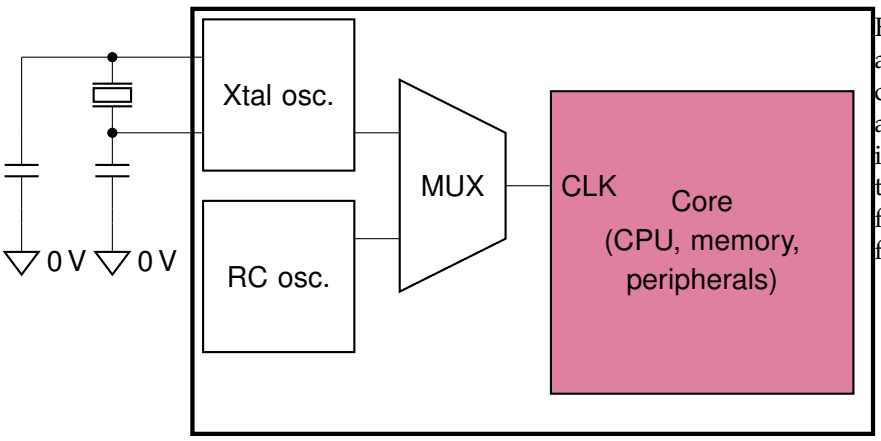


Figure 1.5: Most MCUs have an internal RC oscillator and a crystal oscillator that requires an external crystal. Not shown, is a phase-locked loop (PLL) that multiplies the oscillator frequency to the required clock frequency.

101 Programming

All MCUs need a way to program the internal flash program memory. This is achieved using a special purpose serial bus. The common ones, JTAG and SWD (serial wire debug), can also be used to debug a running program.

110 Reset

All MCUs have a power-on-reset circuit. Most MCUs also have a special reset pin (usually active low with an internal pull-up resistor). This can be connected to a push-button to reset the MCU during development.

111 Datasheets

Every MCU datasheet has an electrical characteristics section. The first part is the absolute maximum ratings, for example, the maximum voltage that can be applied to any pin without destroying the device.

1000 Exercises

1. What are most of the pins on a MCU used for?
2. Do you need to connect all the ground (GND) pins on a MCU?
3. Why do many MCUs have multiple power supply pins of different voltages?
4. Why do simple MCUs have a single power supply voltage?
5. Why does a MCU need a clock?
6. What is a PLL and what is it used for?
7. Would a crystal oscillate if a potential of 5 V was applied across it?
8. Can you get away with powering a MCU at a voltage greater than the absolute maximum value specified in the datasheet?

2
Assignment

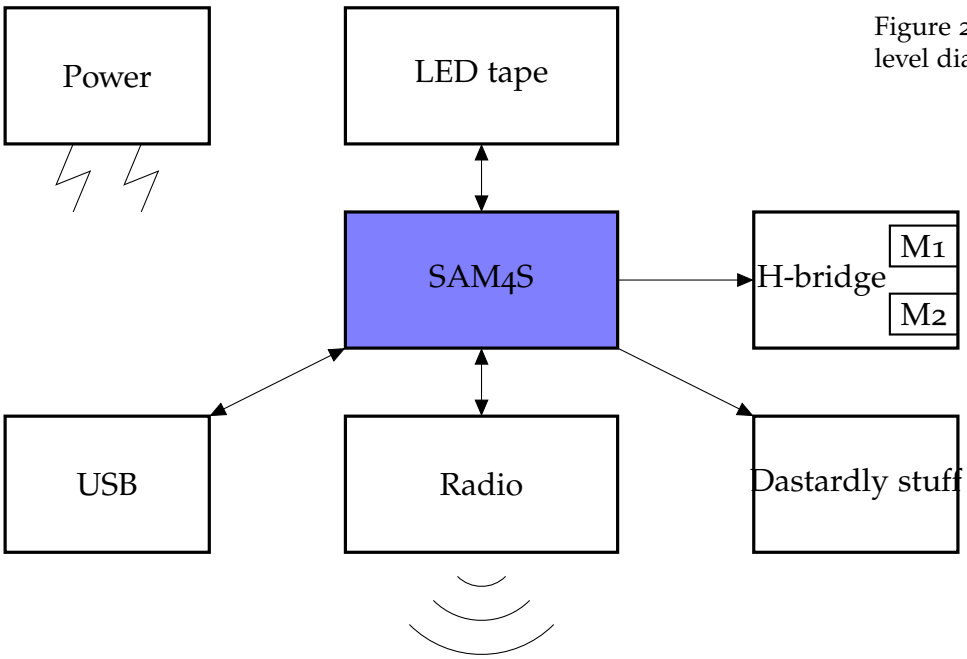


Figure 2.1: Racer board top level diagram.

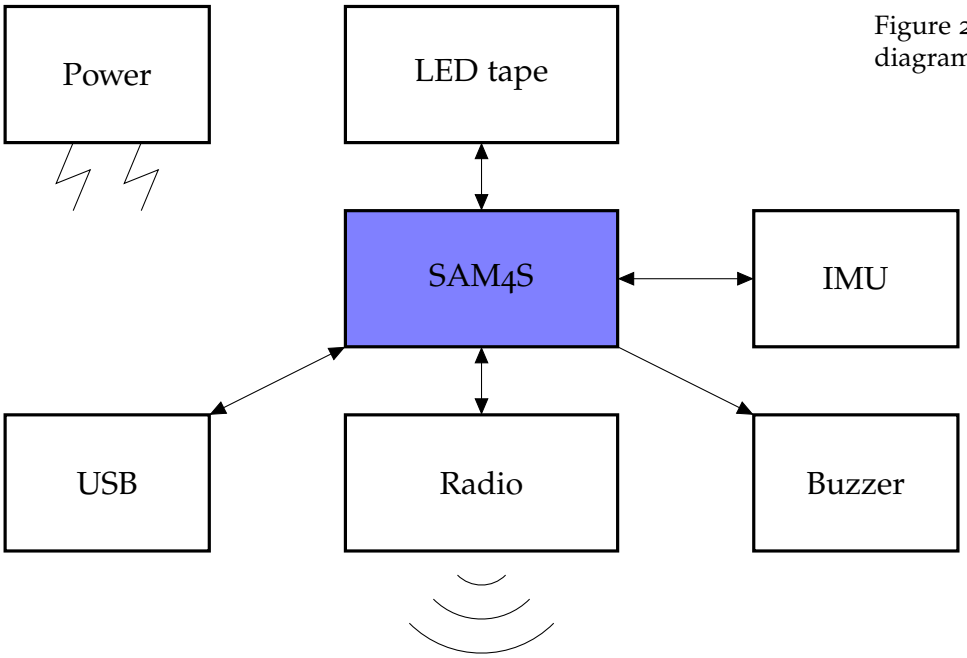


Figure 2.2: Racer hat top level diagram.

3
I/O ports

To save pins and to provide greater flexibility, most microcontrollers use software configurable bi-directional parallel input/output (I/O) ports. These are often called PIO (parallel I/O ports) or GPIO (general purpose I/O) ports.

1 PIO pins

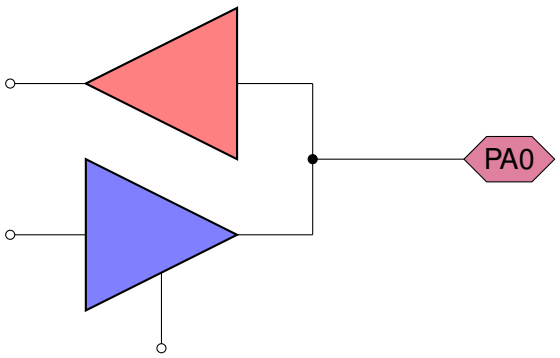


Figure 3.1: PIO pin buffers for PIO pin `PA0`.

If you follow a bond wire connecting a PIO pin to the silicon die you will get to a pair of buffers (see Figure 3.1). One of these buffers is an input buffer; the other is an output buffer[1]. The output buffer is a tri-state buffer with an enable pin to disable the output.

[1] This is physically much larger than the input buffer since it needs to handle larger currents.

1.1 PIO output operation

The output state is held by a flip-flop[2] and the data direction state is held by another flip-flop[3], see Figure 3.2. On reset, the data-direction flip is reset so the output buffer is disabled.

[2] Part of the port register.

[3] Part of the data direction register.

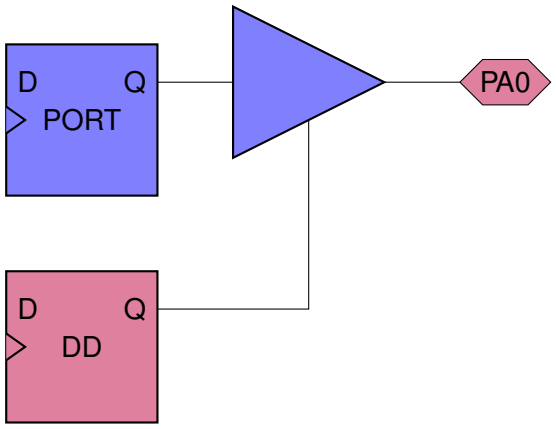


Figure 3.2: Simplified PIO output circuitry showing the port and data-direction flip-flops.

1.10 *PIO input operation*

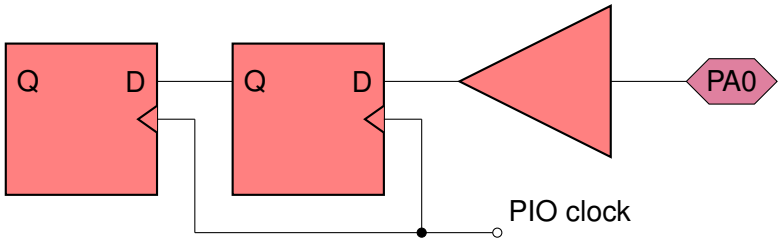


Figure 3.3: Simplified PIO input circuitry showing synchronising flip-flops.

The input buffer is typically a schmitt trigger[4]. It is then connected to a chain of flip-flops to reduce the chances of metastability. Note, these flip-flops are clocked and thus the input state cannot be read if this clock is stopped.

[4] This has hysteresis to reduce switching jitter for slowly changing input signals around the CMOS midpoint voltage.

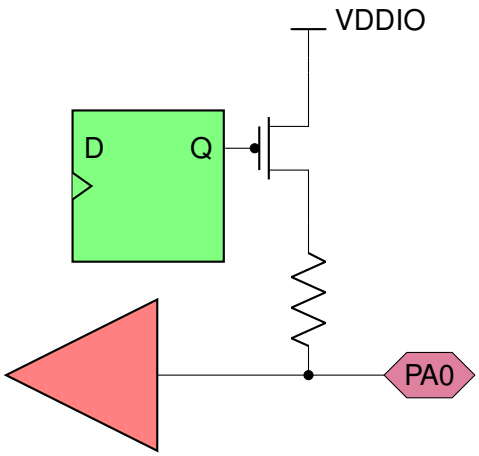


Figure 3.4: Simplified PIO pull-up circuitry. The resistor is typically 10-100 kΩ.

PIOs also have programmable pull-up resistors (see Figure 3.4) and sometimes programmable pull-down resistors. For most MCUs, the pull-up resistors are enabled on reset[5].

[5] On reset the AT91SAM processor configures the PIO pins as inputs with pull-up resistor enabled. The SAM4S uses 100 kΩ but the SAM7S appears to use a much stronger resistor, more like 10 kΩ. AVR MCUs do not enable the pull-up resistor.

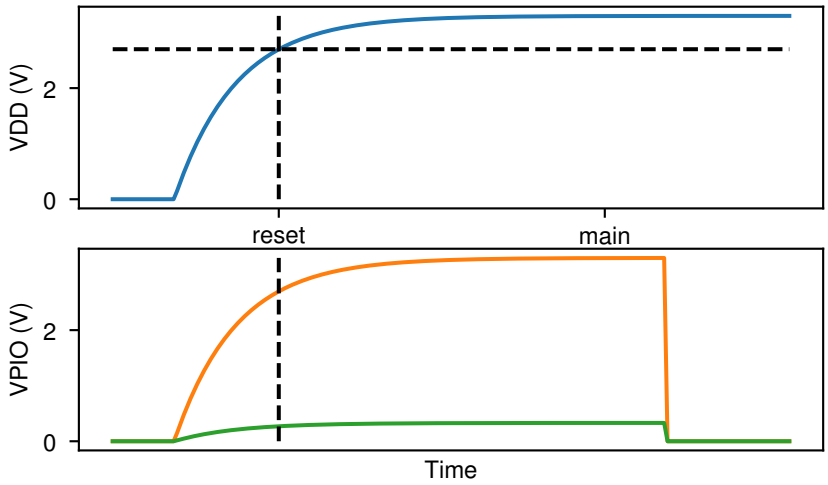


Figure 3.5: VDD voltage and PIO voltage with and without external pull-down resistor. The horizontal dashed line denotes the power-on-reset threshold. Initially, the pull-up resistor for the PIO is enabled; some time after main(), the PIO is configured as a low output.

1.11 *Metastability*

Metastability in a digital system is a temporary unstable state other than logic 0 or logic 1[6]. It can occur when the input to a flip-flop is not stable for the required setup time before the flip-flop is clocked (see Figure 3.6 and Figure 3.7). In the metastable state, the flip-flop output can take many times longer than normal to settle to logic 1 or 0; it may even oscillate before settling.

Metastability is unavoidable whenever there are asynchronous inputs. However, by using a chain of synchronising flip-flops, the probability that the last flip-flop is in a metastable state can be reduced to negligible levels.

[6] As an analogy, consider tossing a coin. It should land either heads or tails—the two possible stable states. But what happens if the coin lands on its edge? This is a metastable state. The coin will not stay in this state for long, since any slight perturbation will cause it to topple and land in a stable state. Obviously, the more often the coin is tossed, then the shorter the time before metastability is observed.

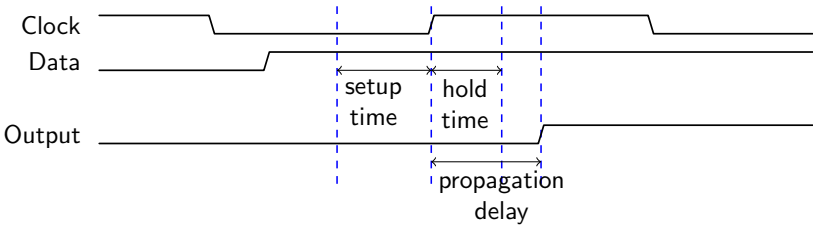


Figure 3.6: A flip-flop has setup and hold times for the data input with respect to the clock that must be observed to avoid a metastable state.

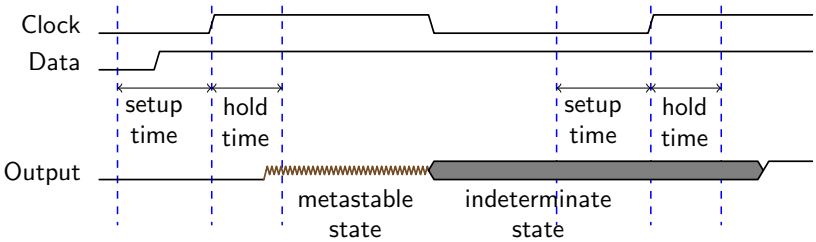


Figure 3.7: Example of when the data signal violates the setup time for a flip-flop. The output oscillates in a metastable state before resolving to an indeterminate stable state.

Metastability is unavoidable with asynchronous inputs. However, it can be mitigated using a chain of synchronising flip-flops. This reduces the probability of the metastable state propagating through the chain of synchronising flip-flops.

The mean time between failure for a flip-flop due to metastability is given by

$$\text{MTBF} = \frac{\exp(t'/\tau)}{T_0 f_c f_i} \tag{3.1}$$

where T_0 is the likelihood that a flip-flop enters a metastable state, τ is the metastability time constant, t' is the sampling time after the clock, f_c is the clock frequency, f_i is the average input frequency.

10 Multifunction PIOs

One of the most limited resources on a microcontroller is the number of pins for input/output. Thus microcontrollers share the PIO pins with peripherals using pin multiplexers[7] (pin muxes) that can be configured by software.

[7] Implemented using tristate buffers.

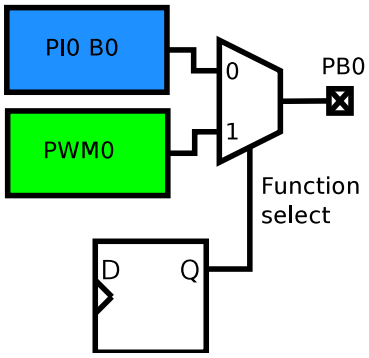


Figure 3.8: Multifunction PIO implementation example. Here the pin PB0 is shared with PWM channel 0.

Using pin multiplexers allows the designer to:

- route a peripheral output to multiple PIOs
- simplify a PCB layout

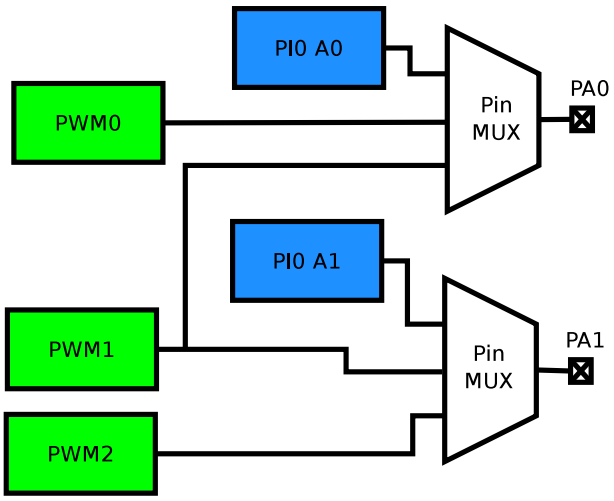


Figure 3.9: Multifunction PIO implementation example with two peripheral signals multiplexed to each pin. In this example, PWM channel 1 can be routed to either PA0 or PA1.

11 PIO programming models

There are two programming models for PIOs:

Read-modify-write Here you read a port register, modify the value, and write it back to the port register.

For example, a C code snippet to make pin B0 on an ATmega8 an output is

```
DDRB = DDRB | 0x01;
```

Here DDRB refers to the data direction register for PORTB. To make B0 an output it is necessary to *read* DDRB, *modify* the value by or'ing a bit mask, then *write* the new value back to DDRB. The C code to make B0 an input is

```
DDRB = DDRB & ~0x01;
```

Similarly, a C code snippet to drive B0 low is,

```
PORTB = PORTB & ~0x01;
```

Again, this is a read-modify-write operation, this time to the PORTB register.

Write-to-modify Here there are different registers[8] for setting, clearing, and sometimes toggling the PIO state.

For example, a C code snippet to make pin PA0 on an AT91SAM4S an output is

```
PIOA_OER = 0x01;
```

To make PA0 an output you just write the desired bit mask to PIOA_OER, the output enable register for

[8] These are really pseudo-registers that manipulate common physical registers.

PIOA. To make `PA0` an input you write the desired bit mask to the output disable register `PIOA_ODR`, for example,

```
PIOA_ODR = 0x01;
```

Note each of these operations is a single write. This has two advantages:

- It is faster (single cycle operation)
- It avoids race conditions with interrupt handlers[9]

[9] Consider what happens if a non-atomic operation that manipulates a PIO register is interrupted and the interrupt service routine modifies the same PIO register.

100 PIO considerations

1. Since PIO pins power up tristate (usually with internal pull-up resistors enabled), external pull-down (or pull-up) resistors should be used to ensure that devices remain disabled before the PIO is configured.
2. Configure PIO pins as low outputs[10] when sleeping to avoid powering external devices through pull-up ESD[11] diodes[12].
3. Set PIO pin output state before configuring as an output to avoid possible glitch[13].
4. The PIO clock must be enabled if want to use a PIO pin as an input.

[10] Except when there are external pull-up resistors

[11] Electrostatic discharge.

[12] And parasitic diodes.

[13] A glitch is unavoidable with some designs, such as the AVR, that use the output state flip flop to control the internal pull-up resistor.

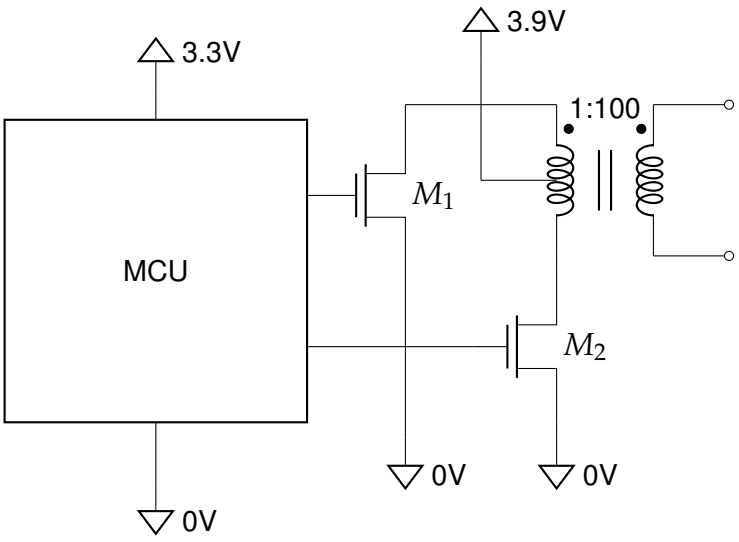


Figure 3.10: What does this do? How does it work? Why did the MOSFETs blow up when the power was switched on? How would you prevent the MOSFETs from blowing up?

101 Summary

1. On reset, all PIOs are configured as inputs; usually with an internal pull-up resistor. Ensure you have suitable pull-downs to avoid selecting devices before they are configured.
2. Use current limiting resistors during development in case of unintentional misconfiguration of a PIO.
3. Some PIOs need to be enabled before use[14].
4. It is becoming more common for microprocessors (like the AT91SAM) to have write-only registers for clearing, setting, and perhaps toggling selected PIO pins. This avoids read-modify-write operations on the port.
5. It is common for microprocessors[15] to have many more peripheral signals than available PIO pins. One strategy is use pin multiplexers[16]. and to map peripheral signals to multiple PIO pins. This gives the system designer more flexibility. A register is configured to choose the desired pin

[14] Input does not work on the AT91 PIO unless the PIO clock is enabled.

[15] Like the AT91SAM.

[16] The AT91SAM4S has A, B, and C options for most peripheral I/O.

110 Exercises

1. When a microcontroller first powers up what state are the PIO pins?
2. Give two advantages of using write-only registers rather than read-modify-write registers for a PIO port.
3. The PIO pins on Atmel AT91SAM4S ARM processors can be mapped to one of three peripheral signals (A, B, or C). What is the advantage of this?
4. Why should a PIO pin used as an input have a series resistor during development?
5. When putting a microcontroller-based embedded system to sleep to save energy, what state should you set the PIO pins?
6. It is common to see pull-up or pull-down resistors connected to PIO pins even though the PIO port has internal pull-up and pull-down resistors. Why?
7. Many PIO ports have Schmitt trigger inputs. Why?
8. Why do PIO inputs have one or more flip-flops cascaded together?
9. To save power, some microcontrollers can disable the clock driving the PIO logic. With the clock disabled, do you think the PIO will work as an input, an output, or neither?
10. A MCU uses 3.3 V for its I/O and on reset it enables 100 kΩ pull-up resistors for each PIO pin. If the program takes 0.1 ms to configure to PIO pin `PA0` as a low output, sketch the voltage waveform on pin `PA0`.
11. Modify the previous sketch if there is a 10 kΩ external pull-down resistor connected to `PA0`.
12. Consider a PIO synchronising flip-flop. If it is driven from a 5 V 100 MHz clock, what is the average current required to charge the clock input if it has 1 pF capacitance. State any assumptions. *Hint the energy stored on a capacitor is* $0.5CV^2$.

111 Supplementary information

There is no standard PIO programming model among microprocessor manufacturers. Some use read-modify-write data registers while others use write-to-clear and write-to-set data registers.

111.1 ATmega AVR PIO programming model

The ATmega AVR processors from Atmel have a number of 8-bit PIO ports (PORTB, PORTC, PORTD, etc.). Each port has three registers:

DDR R/W data direction register; 1 for output.

PORT R/W output state register. When port configured as input this controls whether pull-up resistor enabled[17].

PIN RO pin state register.

[17] Unfortunately, this can produce a glitch when switching from an input to an output or vice-versa.

111.10 AT91SAM PIO programming model

The AT91SAM ARM processors from Atmel have one or more 32-bit PIO ports. Each port has many registers; the key ones are:

PDSR RO port input state register

ODSR RO port output state register

SODR WO port output state set register

CODR WO port output state clear register

PER WO port enable register

PDR WO port disable register

OER WO port output enable register

ODR WO port output disable register

PPUDR WO pull-up enable register

PPUER WO pull-up disable register

4
Inverters

Most digital integrated circuits inputs and outputs are connected to CMOS[1] inverters, so to interface to them we need to know something about how they behave.

[1] Complementary metal oxide semiconductor; complementary refers to both NMOS and PMOS devices.

1 *The ideal inverter*

The ideal inverter transfer function is shown in Figure 4.1. Voltages below the switching (mid-point) voltage produce a high output; voltages above the switching voltage produce a low output[2].

[2] At the switching voltage there is infinite gain so any noise can cause rapid switching of the output voltage.

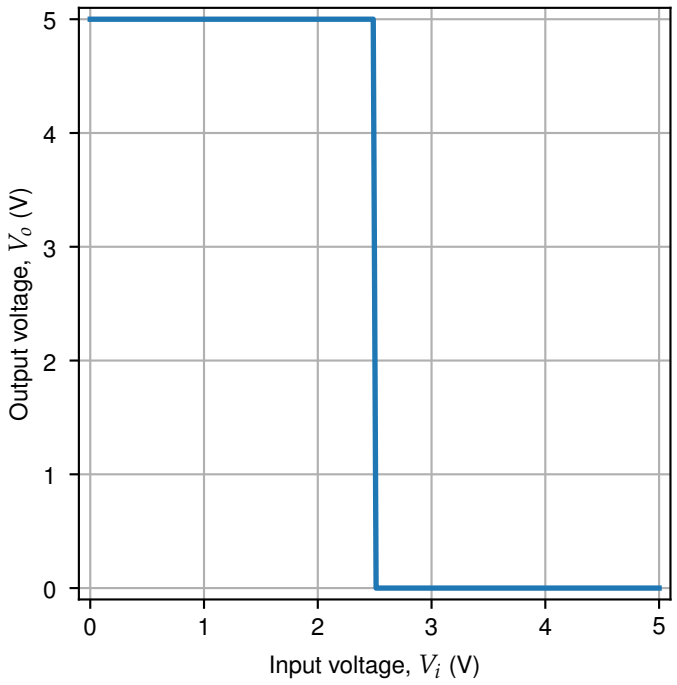


Figure 4.1: Ideal inverter transfer function for 5 V logic.

10 The NMOS inverter

Before CMOS, inverters were made with PMOS and then NMOS FETs. The NMOS inverter consists of an N-type MOSFET (NMOS) and a resistor. To be able to source sufficient current for fast switching the resistor needs to be small. However, this results in a static power loss when the N-type MOSFET is on, driving the output low.

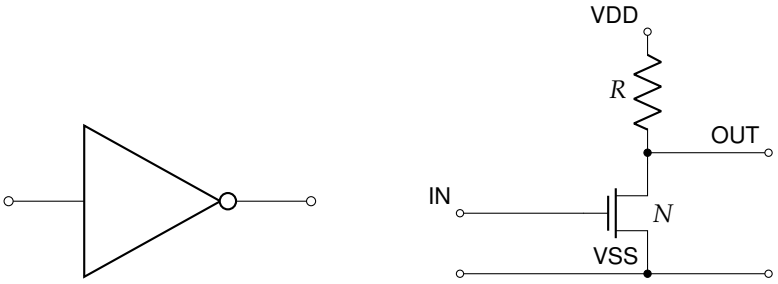


Figure 4.2: The NMOS inverter: symbolic form and implementation. The MOSFET acts like a voltage controlled resistor. With no applied gate-source voltage, the drain-source resistance is incredibly high and so the MOSFET acts like a switch.

11 The CMOS inverter

The CMOS inverter[3] consists of two MOSFET transistors in a totem-pole arrangement. The bottom transistor is N-type (NMOS) and the top transistor is P-type (PMOS). The drains are connected together and provide the output voltage.

The advantage of the CMOS inverter over the NMOS inverter is that there is no static power loss and a large current can be sourced for fast switching high.

[3] This was first proposed by Frank Wanlass in 1963 but took about 20 years to make.

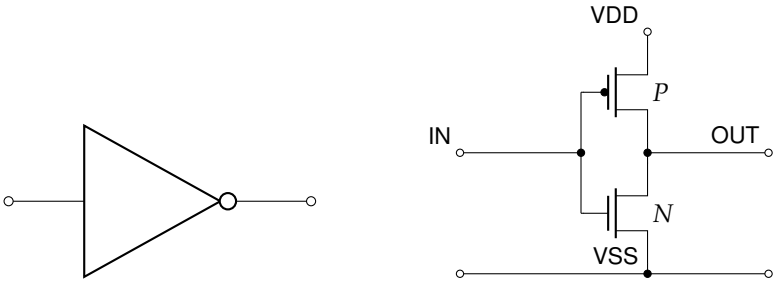


Figure 4.3: The CMOS inverter: symbolic form and implementation with complementary MOSFETs.

11.1 The CMOS inverter transfer function

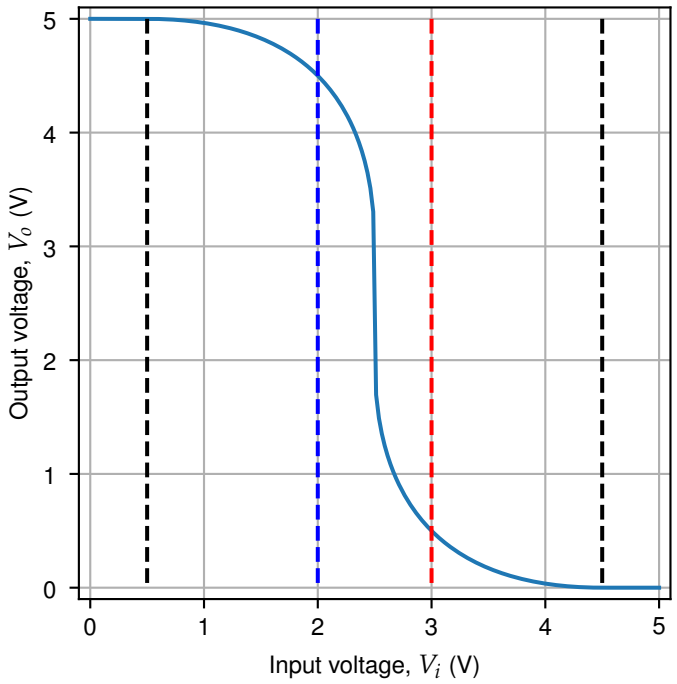


Figure 4.4: CMOS inverter transfer function.

Consider the transfer function of an CMOS inverter, relating output voltage, V_o, to the input voltage, V_i. There are a number of operating regimes depending on the input voltage and the threshold voltages[4] (V_{tn}, V_{tp}) of the MOSFETS comprising the inverter:

[4] When the gate-source voltage is below the threshold voltage the current through a MOSFET drops exponentially.

Region	NMOS	PMOS	Input	Static current
$V_i < V_{\text{tn}}$	Cutoff	Linear	Low	None
$V_{\text{tn}} \leq V_i < V_{\text{il}}$	Saturation	Linear	Low	Some
$V_{\text{il}} \leq V_i < V_{\text{ih}}$	Saturation	Saturation	Transition	Max.
$V_{\text{ih}} \leq V_i < V_{\text{DD}} + V_{\text{tp}}$	Linear	Saturation	High	Some
$V_{\text{DD}} + V_{\text{tp}} \leq V_i < V_{\text{DD}}$	Linear	Cutoff	High	None

Table 4.1: CMOS inverter regions. Note, the NMOS threshold voltage, V_{tn}, is positive. and the PMOS threshold voltage, V_{tp}, is negative.

11.10 CMOS inverter static power consumption

The NMOS transistor is fully turned off if the input voltage is below the transistor's threshold voltage, V_{tn}. Thus no static current flows through the transistors and there is no static power loss.

Once the input voltage rises above V_{tn} the NMOS transistor starts to turn on. Meanwhile the PMOS transistor is fully turned on and thus static current flows. At the switching voltage both the MOSFETs are fully

turned on and the static current[5] is a maximum, see Figure 4.5. Therefore, it is undesirable for the input voltage to loiter around the switching voltage.

[5] This is called short-circuit or crowbar current.

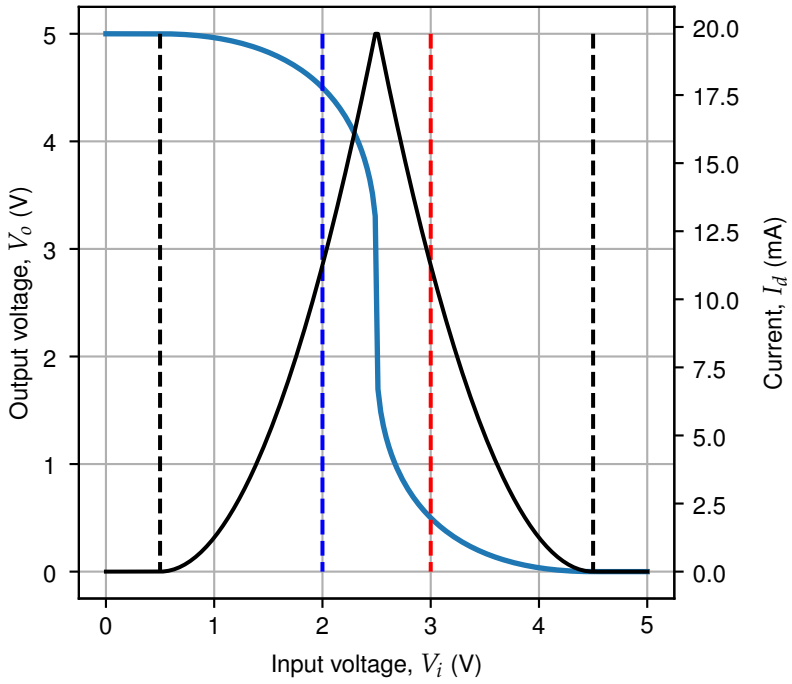


Figure 4.5: CMOS inverter transfer function and static power-supply current. The current depends on the size of the MOSFETs. Note, there are two regions where there is no static current.

11.11 CMOS inverter dynamic power consumption

CMOS power dissipation is dominated by the charging/discharging of load capacitances:

$$P = CV^2 f, \tag{4.1}$$

where C is the capacitance, V is the supply voltage, and f is the switching frequency.

11.100 CMOS inverters as analogue amplifiers

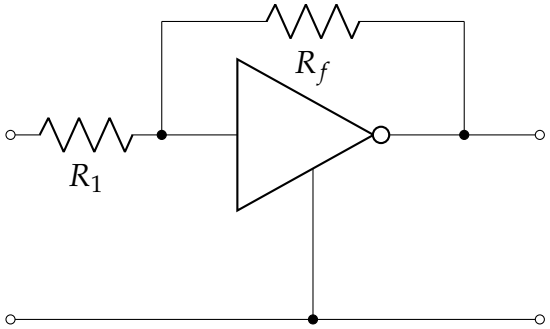


Figure 4.6: CMOS inverter biased as an analogue amplifier. Note, several inverters can be connected in parallel for greater current drive.

CMOS inverters can be used as analogue amplifiers by connecting a feedback resistor[6] between the input and output, see Figure 4.6. The highest gain occurs when the input signals are biased around the switching voltage, see Figure 4.7.

[6] Typically 1–10 MΩ.

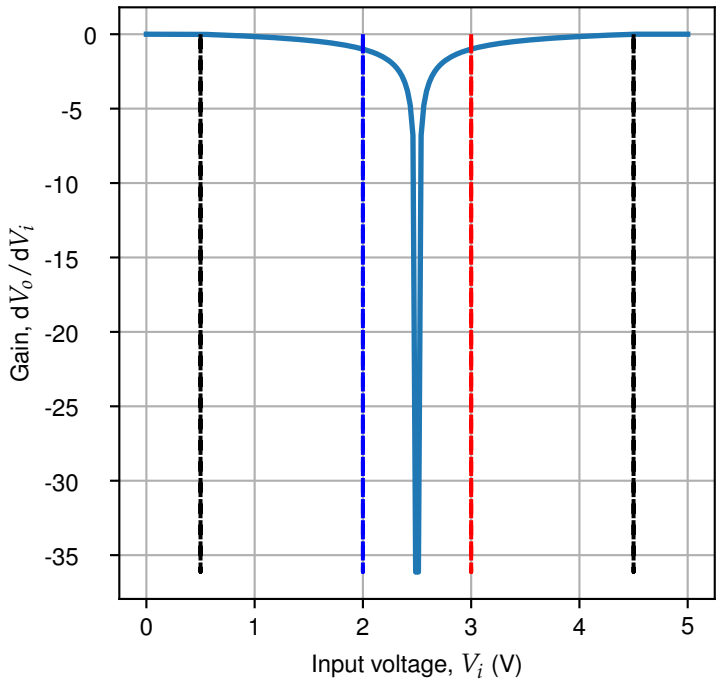


Figure 4.7: CMOS inverter gain. The input logic levels, V_{il} and V_{ih}, are defined where the gain is -1.

11.101 *CMOS inverters as oscillators*

Since a CMOS inverter has gain, it can be configured as an oscillator. Microcontroller clocks are usually generated using a Pierce oscillator using the circuit shown in Figure 4.8. The crystal in the feedback loop determines the oscillation frequency.

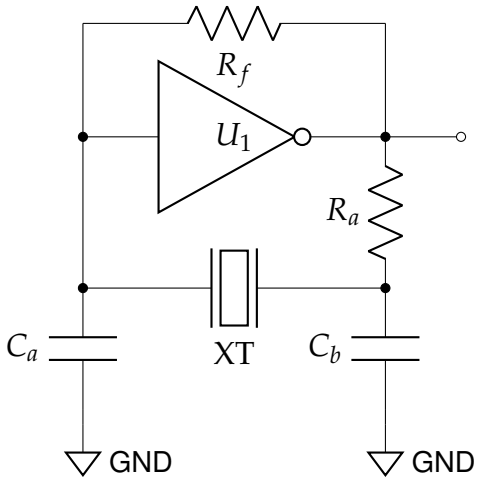


Figure 4.8: The gated Pierce oscillator circuit. R_f is a feedback resistor to ensure the CMOS inverter is biased in the linear region. R_a is a current limiting resistor to prevent over-heating of the crystal.

100 Further reading

http://www.ittc.ku.edu/~jstiles/312/handouts/Binder1.pdf. This has an excellent analysis of CMOS inverter and its transfer function.

Logic guide, TI, 2009, sdyu001z.pdf. The bees-knees of the different logic families and all the variations such as buffered outputs, partial power down, hot insertion, bus-hold inputs, etc.

101 Exercises

1. How should the unused inputs of a CMOS device be connected?
2. How is a CMOS inverter modified to produce a high impedance (tristate output)?
3. Are both transistors in a CMOS inverter ever fully on?
4. Is there any static current drawn by CMOS inverter when the input voltage is V_{SS}?
5. Is there any static current drawn by CMOS inverter when the input voltage is V_{DD}?
6. How can a CMOS inverter be configured as an amplifier?
7. Does the transfer function of a CMOS inverter convey any information about the switching speed?
8. What two advantages does a CMOS inverter have over an NMOS inverter?
9. What happens when a CMOS inverter input is not connected?

5
CMOS inputs

Weird things happen with CMOS but these can be simply explained with an understanding of some simple models.

1 CMOS simplified input model

A simplified low frequency model for a CMOS input signal is shown in Figure 5.1. The input resistance of a CMOS device is extremely high, the order of TΩ (10^{12} Ω). This in parallel with a small capacitance[1].

[1] Typically 3–5 pF.

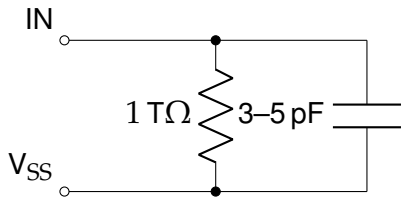


Figure 5.1: CMOS simplified input circuit showing input resistance and capacitance. The resistance can be neglected in practice.

10 CMOS electrostatic damage

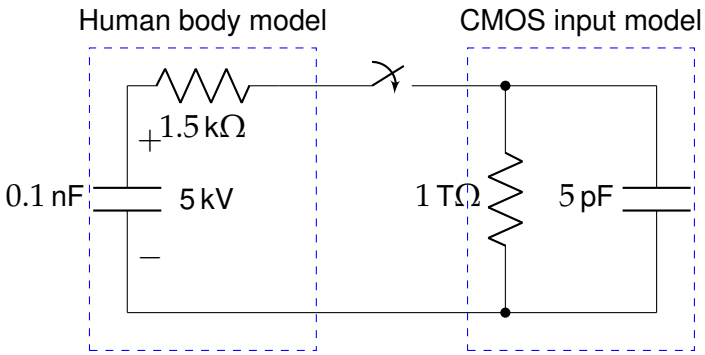


Figure 5.2: Schematic showing why the ESD protection diodes are necessary since the human body can be charged to several kV, say when scrunching a plastic bag.

The high resistance and low capacitance of CMOS signal pins make CMOS devices susceptible to electrostatic damage. Consider the circuit shown in Figure 5.2.

This shows a model of the human body connected to a model of a CMOS input. Due to the high CMOS input resistance, the electrostatic voltages (easily 5 kV) built up on a human body can easily destroy the very thin insulation layers of the MOSFETs in CMOS devices.

11 CMOS electrostatic protection

To protect CMOS devices from ESD during handling, manufacturers include protection circuits. The simplest and most common protection is a pair of clamping diodes for each signal pin[2], see Figure 5.3.

[2] Inputs and outputs.

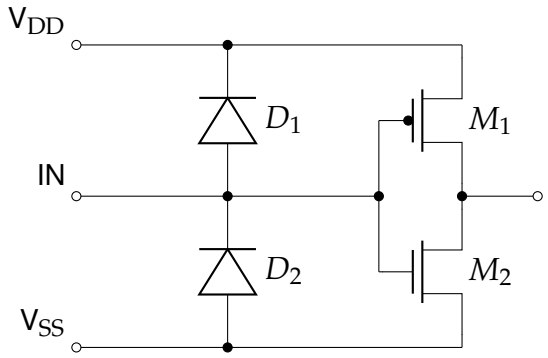


Figure 5.3: CMOS input circuit showing ESD protection diodes.

100 CMOS back drive

Consider the circuit in Figure 5.4. What does it do?

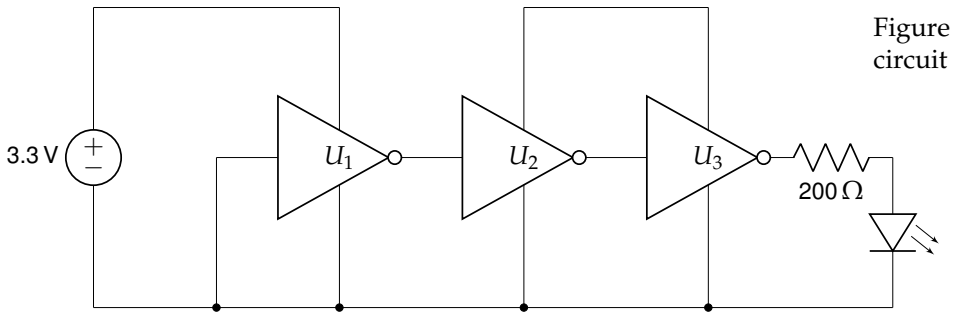


Figure 5.4: What does this circuit do?

Figure 5.4 shows that a CMOS device can be inadvertently powered by one of its signal lines. In this circuit, the input of U1 is grounded and its output is 3.3 V. Since U2 is not powered by its VDD pin, its upper ESD clamping diode conducts[3]. As a consequence there will be a 0.3 V drop across the upper ESD clamping diode, and the VDD pin of U2 will be driven to 3.0 V. This provides power for U3 which drives the LED on.

[3] since its input voltage is greater than the voltage on its VDD pin

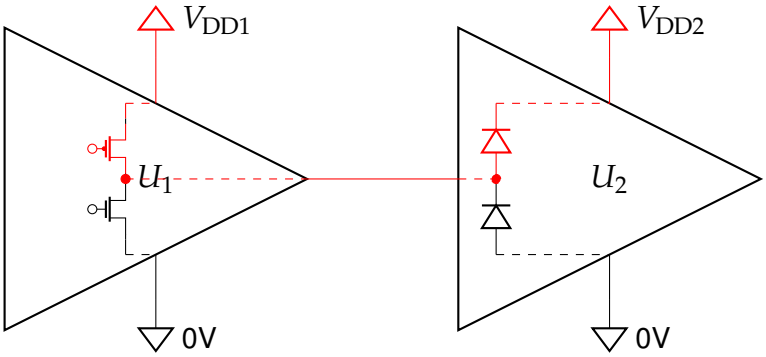


Figure 5.5: The back drive current path shown in red when $V_{DD2} < V_{DD1}$.

101 CMOS HV protection

For serious ESD protection[4], circuits should include external clamping diodes (see Figure 5.6), or transient voltage suppressor (TVS) diodes.

[4] especially with external cables.

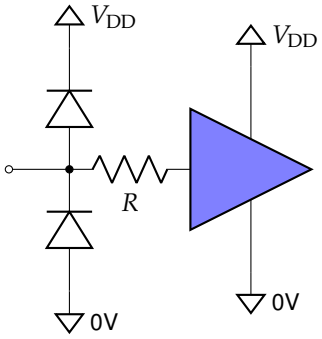


Figure 5.6: CMOS high voltage protection using external clamping diodes and resistor to limit current passed by the internal clamping diodes.

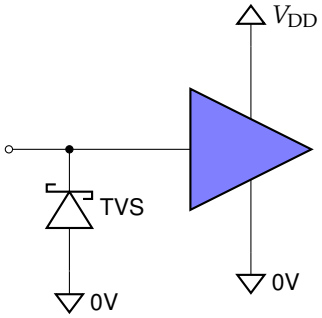


Figure 5.7: CMOS high voltage protection using external TVS clamping diode. These are similar to zener diodes but are faster acting and designed to suppress high voltage transients.

110 CMOS input model

A more accurate low frequency model for a CMOS input signal is shown in Figure 5.8. It's Thévenin equivalent is shown in Figure 5.9.

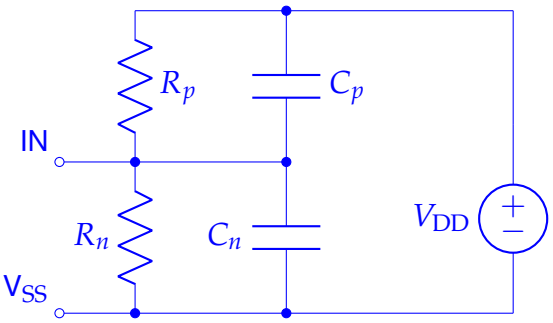


Figure 5.8: CMOS input circuit showing parasitic capacitance (these vary with the input voltage due to the Miller effect). Note, this model assumes that the ESD protection diodes are not conducting.

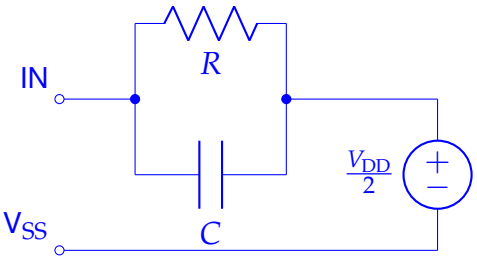


Figure 5.9: Thévenin input model of Figure 5.8. R is typically 1 TΩ, C is typically 3–5 pF.

111 Further reading

"Latch-up, ESD, and other phenomena", Texas Instruments, 2000, www.ti.com.cn/general/cn/docs/lit/getliterature.tsp?baseLiteratureNumber=slya014&fileType=pdf. This has an in-depth analysis of ESD and latch-up.

"Electrically Induced Damage to Standard Linear Integrated Circuits: The Most Common Causes and the Associated Fixes to Prevent Recurrence", Analog Devices Application Note AN-397, 2012, www.analog.com/media/en/technical-documentation/application-notes/AN-397.pdf

"Application Hints for Transient Voltage Suppression Diode Circuits", On Semiconductor, 2009, www.onsemi.com/pub/Collateral/AND8230-D.PDF. This has a good descriptions of robust interfaces and back drive protection.

1000 Exercises

1. How can a chip be powered inadvertently through a PIO pin when a microcontroller is sleeping?
2. Why do the inputs of CMOS devices have diodes connect to the power supply and ground?
3. What will happen if you apply a 6 V input to a 5 V CMOS device?
4. A Bluetooth module has some apparently unused metal pads on its underside. From what you know about CMOS models, suggest how you would determine if these pads are electrically connected inside the module?
5. Why is the Bluetooth module in Figure 5.10 not reset when its power is cycled using the regulator U3?
6. What is the purpose of the resistor in Figure 5.10?
7. What value do you recommend for the resistor in Figure 5.10? State assumptions.
8. Turning off the voltage regulator U3 in Figure 5.10 does not result in the expected energy saving. Why?

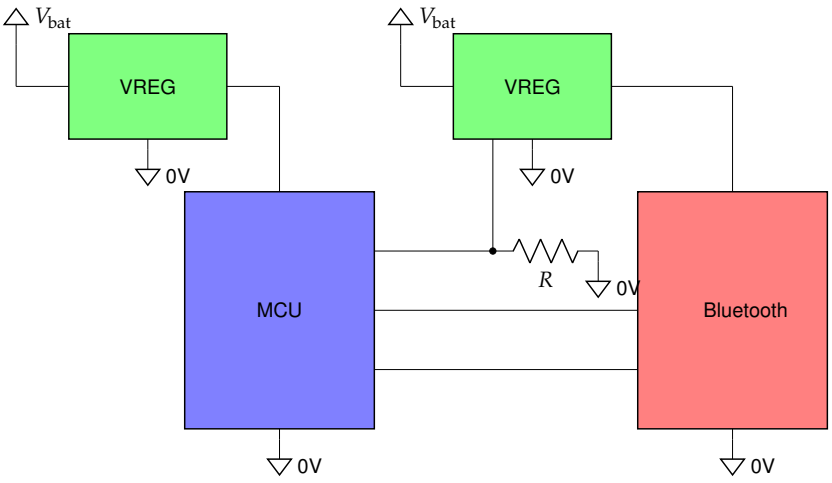


Figure 5.10: A Bluetooth module interfaced to a MCU using a UART.

6
CMOS outputs

The output voltage of a CMOS device, and thus the noise margins, depends on the resistive loading. To understand what happens it is useful to construct simple[1] electrical models of a CMOS device.

[1] These models are valid for low frequency operation.

1 CMOS output model

CMOS outputs have a totem-pole driver comprised of a PMOS pull-up transistor and an NMOS pull-down transistor (Figure 6.1). There are three useful[2] modes of operation:

[2] Having both transistors on is undesirable since it shorts out the power supply and will cook the device.

1. When only the pull-up (PMOS) transistor is enabled the output is high and can source current.
2. When only the pull-down (NMOS) transistor is enabled the output is low and can sink current.
3. When both transistors are disabled the output appears as a high-impedance (tri-state).

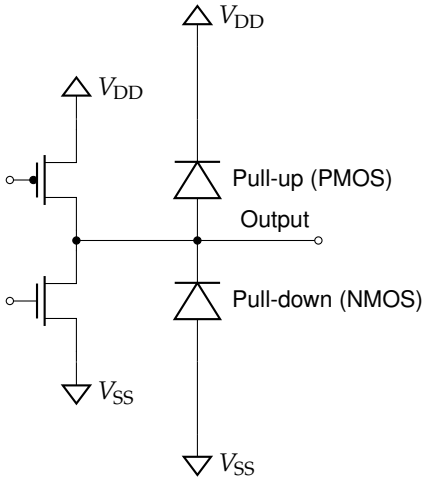


Figure 6.1: Totem-pole output driver showing the pull-up and pull-down transistors and the ESD protection clamping diodes.

10 CMOS simplified output model

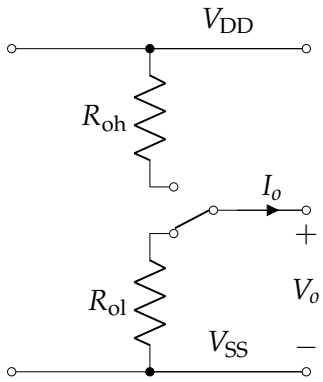


Figure 6.2: Simplified PIO pin output electrical model. In practice, the output resistances vary with current).

A simplified model for a CMOS output is shown in Figure 6.2. The relation between the output voltage and current is shown in Figure 6.3.

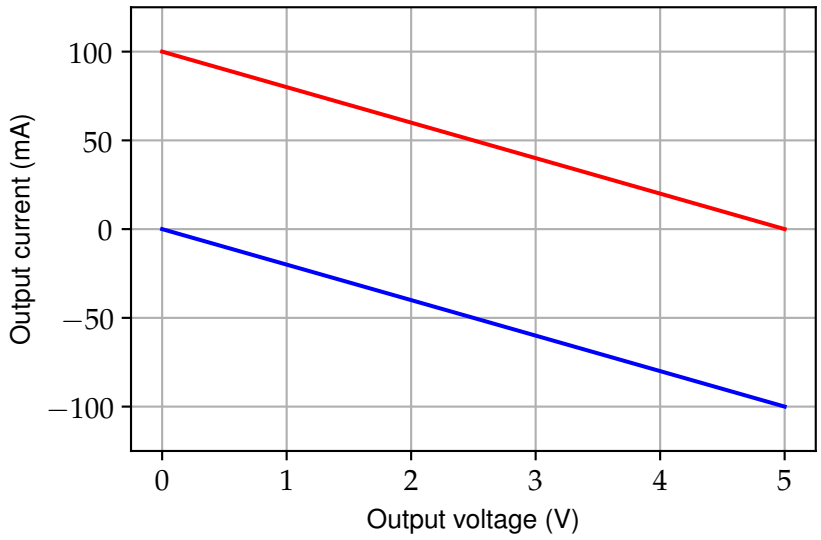


Figure 6.3: IV curves for CMOS simplified output model: (red) 'high'; (blue) 'low'. Here $V_{DD} = 5\,V$, $R_{oh} = R_{ol} = 50\,\Omega$.

11 CMOS IV curves

MOSFETS are non-linear and the output resistance of CMOS varies with current. Thus the IV curve is more complicated (see Figure 6.5). Note, the IV curve does not imply that the device can provide the specified current continuously due to thermal reasons.

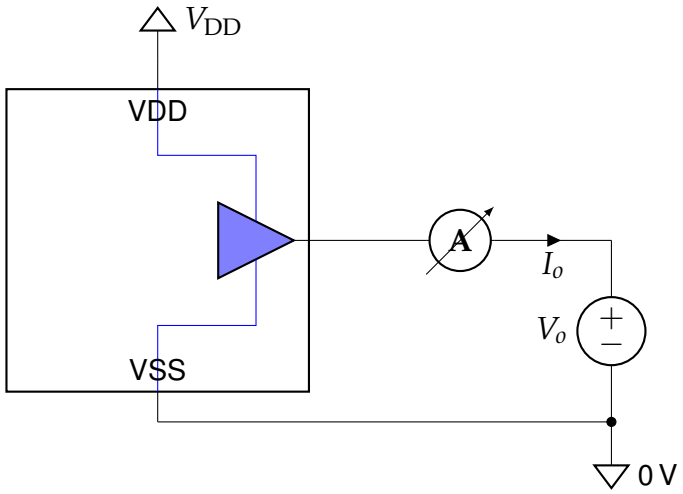


Figure 6.4: Measurement circuit for the I/V curve of a CMOS output.

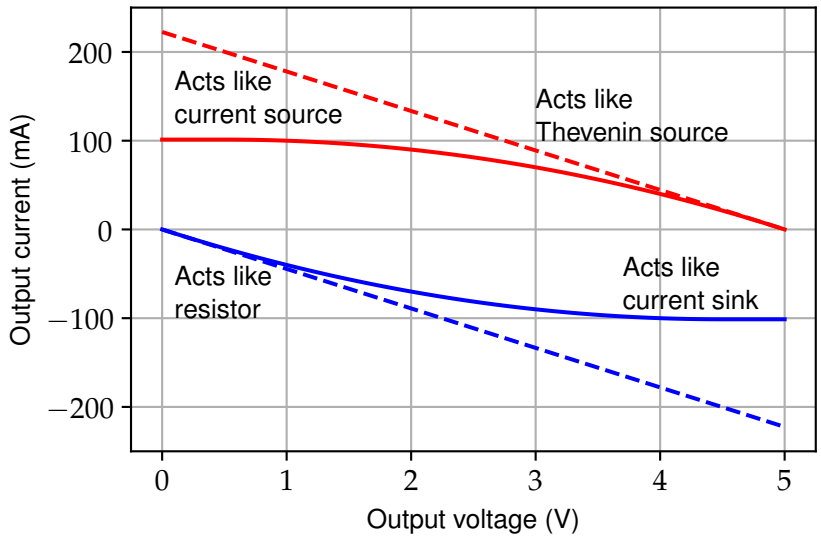


Figure 6.5: Example CMOS output characteristic curves: (red) 'high'; (blue) 'low'.

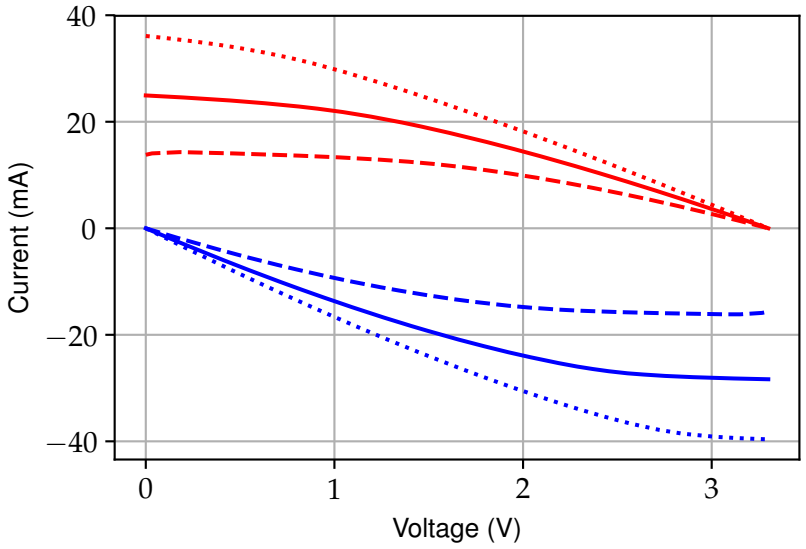
11.1 Maximum output current

While CMOS devices can source large currents to quickly charge load capacitance, for thermal reasons the datasheet specifies a lower static current. For example, most SAM4S pins can only handle 2 mA of static current[3].

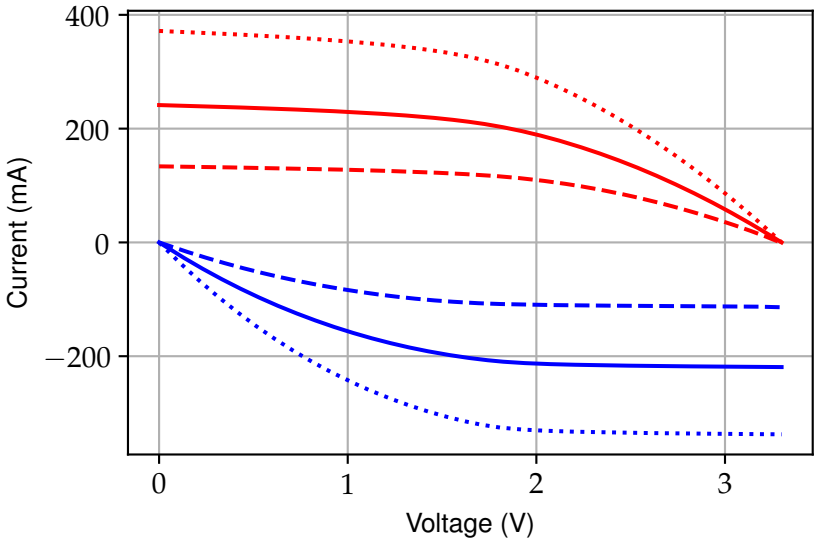
[3] In comparison, the ATmega8 datasheet states that the maximum current per I/O pin is 40 mA and the maximum current per VCC or GND pin is 200 mA.

11.10 Overvoltage

When a CMOS pin voltage is forced beyond the power supplies, the ESD diodes conduct and large currents can flow, see Figure 6.7.



(a)



(b)

Figure 6.6: SAM4S IV curves: (a) `PA0` pin (b) USB `UDP` pin. Note the USB pin can switch greater currents; it achieves this with larger MOSFETs. Also note the large variations in output current due to process variations.

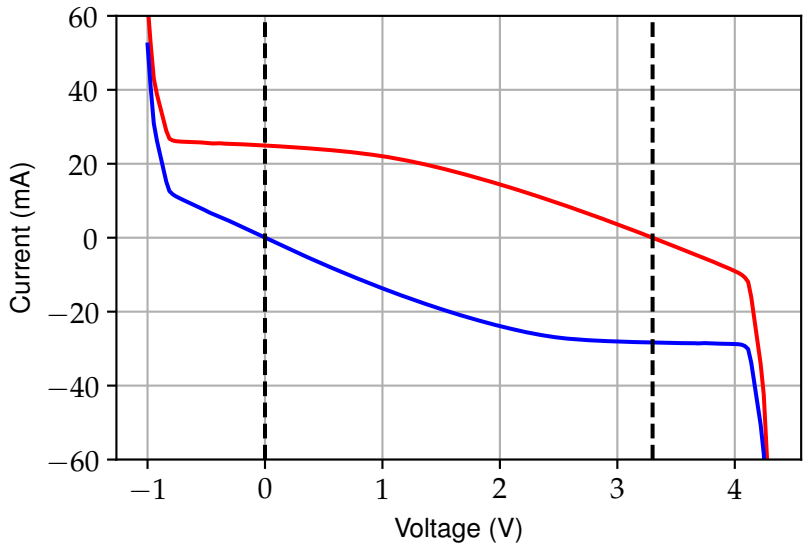


Figure 6.7: SAM4S IV curve showing region beyond power supply voltage where ESD diodes start to conduct.

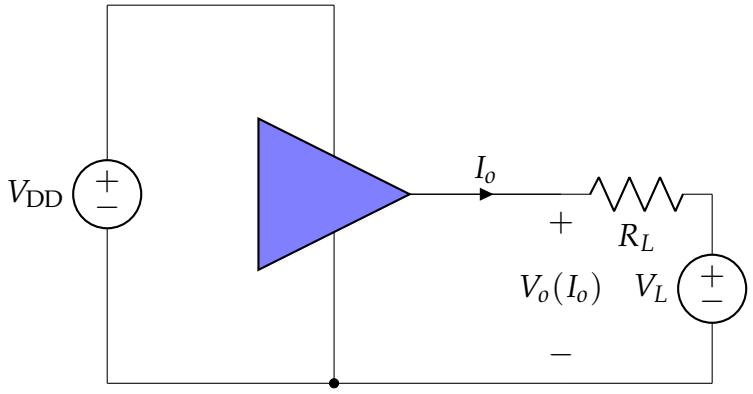


Figure 6.8: CMOS with a Thévenin DC load.

11.11 DC loading

If a CMOS driver is loaded, the output voltage changes. In general, with an arbitrary load (see Figure 6.8),

$$V_o(I_o) = I_o R_L + V_L. \tag{6.1}$$

Here $V_o(I_o)$ is the current dependent output voltage characteristic described by the IV curve. The solution for I_0 (and V_0) can be found graphically by annotating the IV curve with a load line. The point where the IV and load line curves intersect is the solution.

11.100 *Pull-down resistor*

A parallel pull-down resistor reduces the output high voltage[4] and creates a static output current. In this case $V_L = 0$ and the load line passes through the origin with a slope of $1/R_L$, see Figure 6.9.

[4] This reduces the logic-high noise margin.

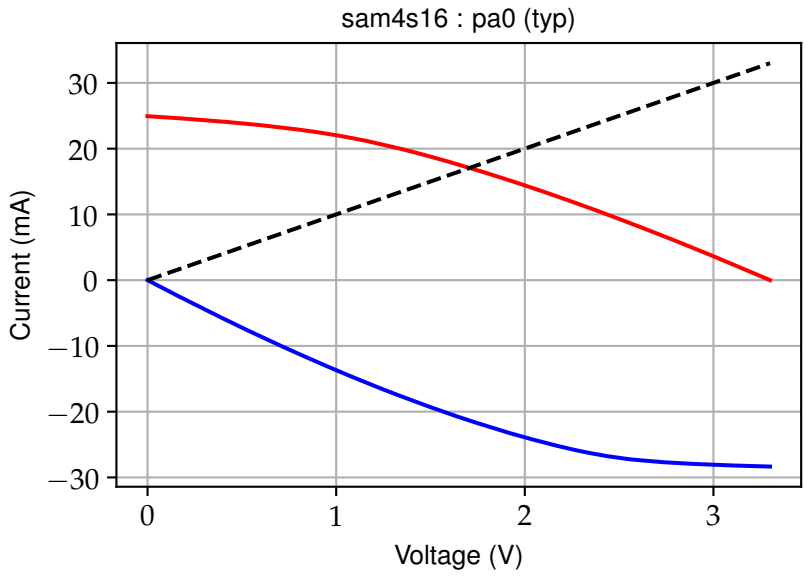


Figure 6.9: Example CMOS output characteristic curves with load line due to a parallel 100 Ω pull-down resistor. The logic-high output voltage is reduced from 3.3 V to about 1.7 V and there is a static current of 17 mA. The logic-low voltage is unaffected.

11.101 *Pull-up resistor*

A parallel pull-up resistor increases the output low voltage[5] and creates a static output current. In this case $V_L = V_{DD}$ and the load has a slope of $1/R_L$, see Figure 6.10.

[5] This reduces the logic-low noise margin.

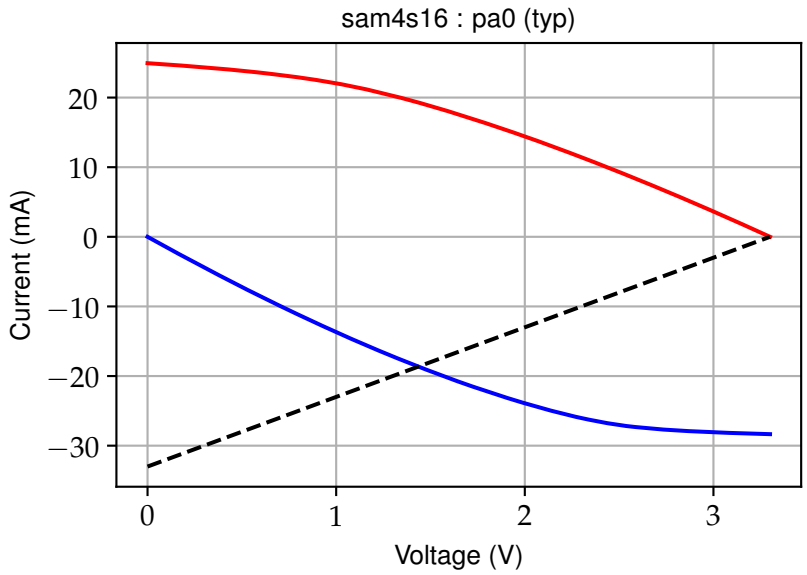


Figure 6.10: Example CMOS output characteristic curves with load line due to a parallel 100 Ω pull-up resistor. The logic-low output voltage is increased from 0.0 V to about 1.4 V and there is a static current of 18 mA. The logic-high voltage is unaffected.

100 *Open-drain outputs*

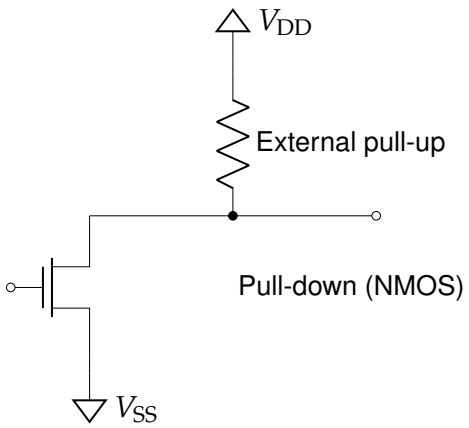


Figure 6.11: Open-drain driver showing external pull-up resistor.

Open-drain (open-collector) outputs do not use the totem-pole configuration; they only have a pull-down transistor (see Figure 6.11). A high output requires an extrenal pull-up resistor connected to V_{DD}. This simplifies interfacing to higher voltage logic levels or when multiple outputs share the same signal.

Open-drain outputs have a static power loss in the low state and thus the pull-up resistor cannot be strong[6]. As a consequence, they are slow when switching from low to high[7].

[6] A low resistance.

[7] The time constant is the product of the pull-up resistance and the load capacitance.

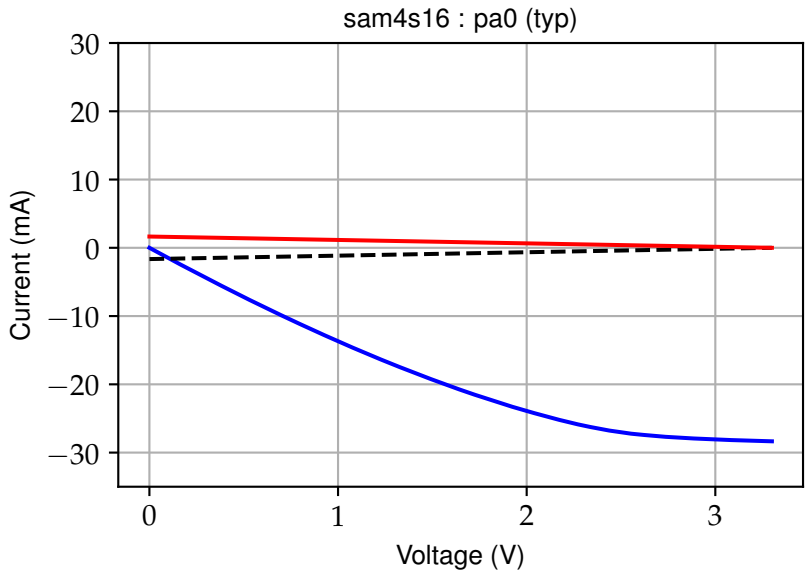


Figure 6.12: Example CMOS output characteristic curves for open drain output with a parallel 2000 Ω pull-up resistor.

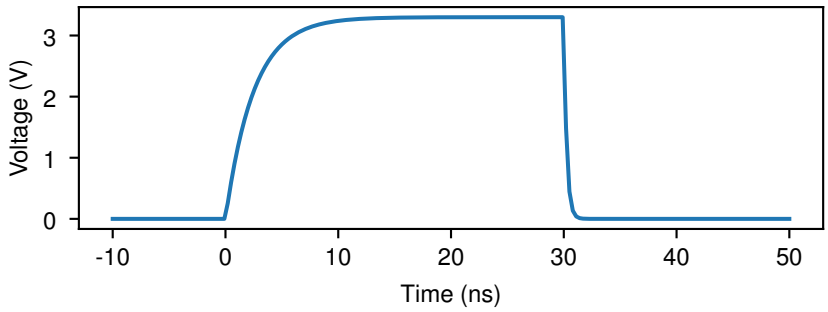


Figure 6.13: Open-drain waveform showing slow rising transisition and fast falling transisition.

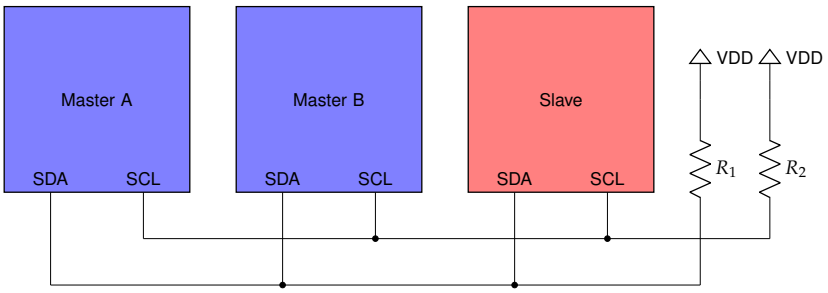


Figure 6.14: I2C interface between two masters and one slave device.

100.1 Open-drain contention

Open-drain outputs are used for bidirectional buses where there is a chance of contention; say one device wants to drive a signal high while another device wants to drive it low. An example is the I2C and CAN buses.

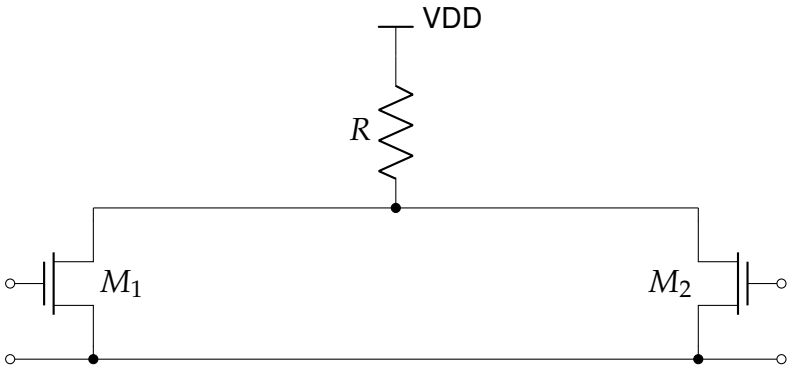


Figure 6.15: Open-drain drivers with bidirectional bus.

100.10 *Open-drain level-shifting*

Another use for open-drain outputs is voltage level shifting. The output high voltage depends on the voltage the external pull-up resistor is connected to[8].

[8] Be careful if the output device has a clamp diode and the pull-up resistor is small.

101 *LED driving*

Let's consider driving a LED from a PIO of a SAM4S microcontroller. This is shown schematically in Figure 6.16. The problem is how to choose the resistor R.

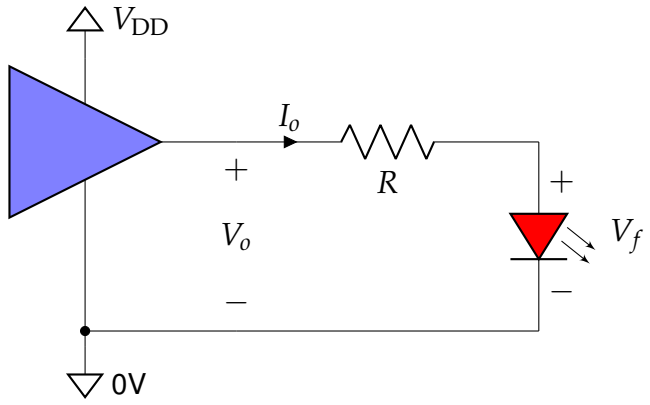


Figure 6.16: CMOS LED driving.

First we need to choose the current through the LED. For thermal reasons, this cannot exceed the maximum continuous current that a PIO pin can supply. Let's choose 2 mA[9]. Given a specified current, the forward voltage drop for the LED can be found from its datasheet. Let's assume a red LED with $V_f = 2.2\,\text{V}$ at 2 mA.

[9] All the SAM4S PIO pins can drive 2 mA continuously, some can drive 4 mA.

From the equivalent circuit shown in Figure 6.17, the required resistance, R, can be found using

$$R = \frac{V_o - V_f}{I_o}. \tag{6.2}$$

Since we know I_0 and V_f all we need is V_o. This can

can be found from the IV curve for the SAM4S, see Figure 6.18. From inspection of the curve, for $I_o = 2\,\text{mA}$, $V_o \approx 3.15\,\text{V}$, and so

$$R = \frac{3.15 - 2.2}{2 \times 10^{-3}} = 475\,\Omega. \tag{6.3}$$

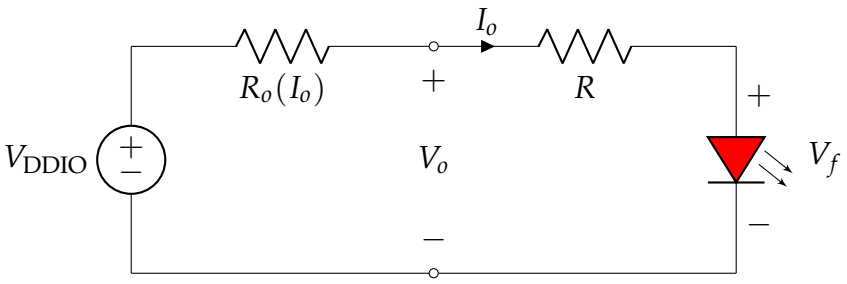


Figure 6.17: CMOS LED driving circuit model.

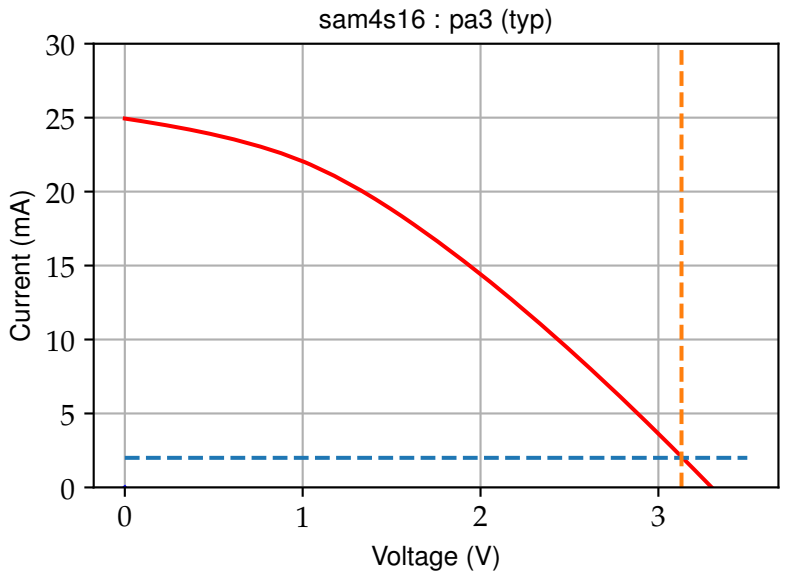


Figure 6.18: SAM4S typical output current versus voltage for pin PA3 showing output voltage for 2 mA source current.

110 Exercises

1. With reference to Figure 6.5, when it first switches from a low output to a high output does it behave more like a current source or voltage source?
2. With reference to Figure 6.5, when it first switches from a high output to a low output does it behave more like a current source or voltage source?
3. Consider Figure 6.5, what is the open-circuit output voltage in the high output state?

4. Consider Figure 6.5, what is the short-circuit output current in the low output state?

5. Consider Figure 6.5, what is the output voltage in the high output state if a 50 Ω load is connected to ground?

6. Consider Figure 6.5, what is the output voltage in the high output state if a 100 Ω load is connected to ground?

7. Consider Figure 6.5, what is the output voltage in the low output state if a 50 Ω load is connected to ground?

8. Consider Figure 6.5, what is the output voltage in the low output state if a 50 Ω load is connected to 5 V?

9. Consider Figure 6.5, what is the output voltage in the high output state if a 50 Ω load is connected to 2.5 V?

10. Sketch the current/voltage characteristic for a CMOS output assuming 5 V operation, an output resistance of 25 Ω (high and low).

11. Sketch the current/voltage characteristic for a CMOS output assuming 5 V operation, an output resistance of 25 Ω (high and low) that is current limited at ±100 mA.

12. With reference to Figure 6.18, determine the required resistance R to set the current to 5 mA through an LED with a forward drop $V_f = 2.5$ V.

13. With reference to Figure 6.18, what is the maximum current that can be sourced to an LED with a forward drop $V_f = 2.5$ V.

14. Sketch the IV curve for a 5 V voltage source in series with a 100 Ω resistor. Then draw the load line for a 400 Ω resistor and determine the voltage across the load resistor.

15. Sketch the IV curve for a 0 V voltage source in series with a 100 Ω resistor. Then draw the load line for a 400 Ω resistor in series with a 5 V voltage source. Finally, determine the current through the load.

16. Sketch the IV curve described by the relation $I(V) = (5 - V)/100$ for the region $0 \leq V \leq 5$. What is the output resistance?

17. Consider Figure 6.5, what is the minimum output resistance when the PMOS transistor is on?

18. Consider Figure 6.5, what is the minimum output resistance when the NMOS transistor is on?

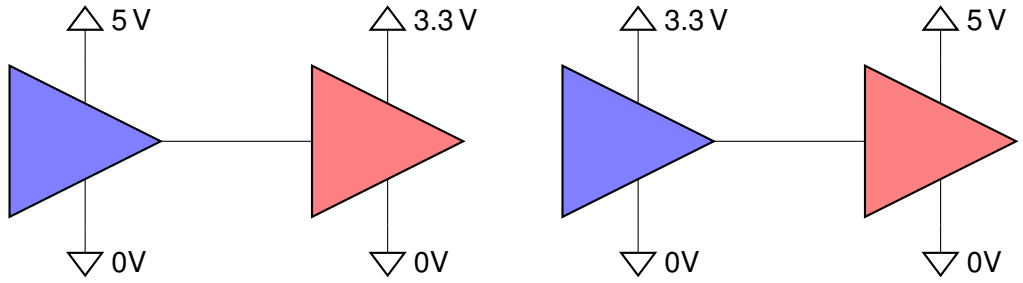
7
Logic levels

It is common for different ICs to have different power supply requirements. When they are interconnected, it is important for the logic levels[1] to be compatible.

[1] This information is usually specified in the electrical characteristics section of a datasheet or buried in an IBIS model.

1 Logic levels

Figure 7.1: Logic level compatibility.



Consider the schematics in Figure 7.1 where logic devices powered from different voltage levels are interconnected. How do we know if these will work?

The key parameters are the input and output logic levels:

$V_{\text{oh min}}$ Output high logic level[2]
$V_{\text{ol max}}$ Output low logic level[3]
$V_{\text{ih min}}$ Input high logic level
$V_{\text{il max}}$ Input low logic level

These are specified in the electrical characteristics section of a datasheet[4].

[2] V_{DD} for CMOS with no resistive load.
[3] 0 for CMOS with no resistive load.

[4] For CMOS, $V_{\text{oh}} = V_{DD}$ and $V_{\text{ol}} = 0$ but these values will change with DC loading, i.e., with the output current.

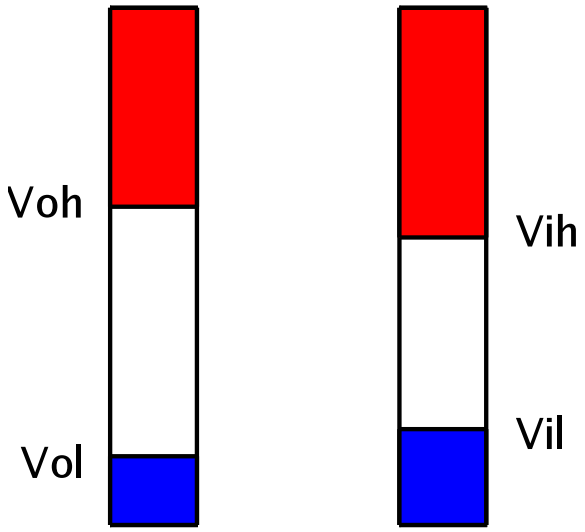


Figure 7.2: Logic level noise margins.

1.1 *Logic level noise margins*

It is important to realise that digital signals are really analogue signals contaminated by noise. The signal levels depend on loading and fluctuations present on the power supply rails. The output signal also gets contaminated with crosstalk and common-path noise from other signals plus reflections. The amount of noise that a digital system can tolerate depends on the logic level noise margins (Figure 7.2). These are defined by

$$\begin{aligned} \text{logic high noise margin} &= V_{\text{oh min}} - V_{\text{ih min}}, \\ \text{logic low noise margin} &= V_{\text{il max}} - V_{\text{ol max}}. \end{aligned}$$

1.10 *Logic level compatibility*

The following conditions must be satisfied:

1. *The logic high and logic low noise margins are adequate.* It is essential that both these noise margins are adequate, especially when driving 5 V logic from 3.3 V logic as shown in Figure 7.3(b)[5]. A rule of thumb[6] is that the noise margins should be at least 0.2 V.
2. *The logic high and low output levels do not exceed the maximum voltages for the input device since this can cause latch-up.* Typically, logic devices only can tolerate voltages that exceed[7] its power supply by ± 0.3 V or sometimes ± 0.5 V. This can cause problems when driving 3.3 V logic from 5 V logic unless the 3.3 V logic is specified as 5 V tolerant.

[5] Typically for CMOS $V_{\text{il}} = 0.3V_{\text{DD}}$ and $V_{\text{ih}} = 0.7V_{\text{DD}}$ although some microprocessors use $V_{\text{il}} = 0.2V_{\text{DD}}$ and $V_{\text{ih}} = 0.6V_{\text{DD}}$ to be TTL compatible.

[6] In practice, a noise budget should be determined.

[7] Otherwise the ESD protection diodes start conducting and latch-up can occur.

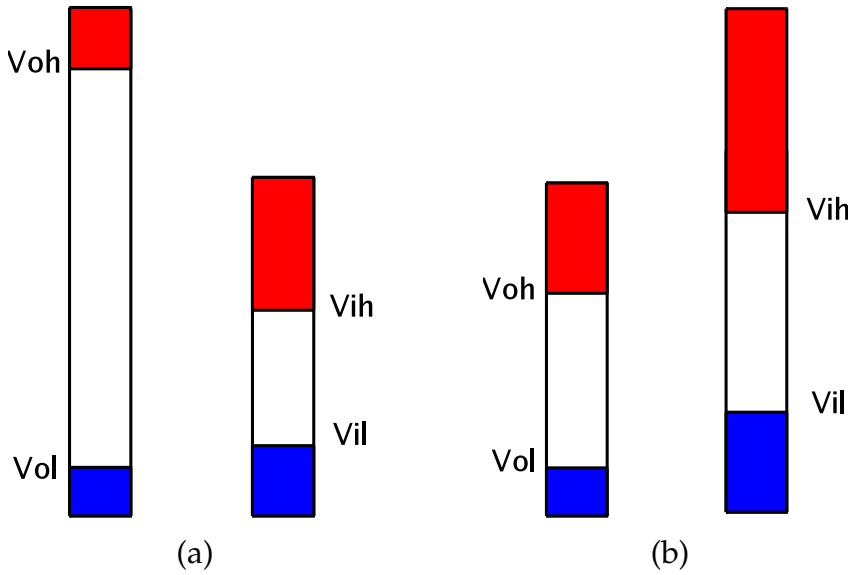


Figure 7.3: (a) Logic level compatibility for 5 V CMOS driving 3.3 V LVCMOS (here there is a chance of latchup). (b) Logic level compatibility for 3.3 V LVCMOS driving 5 V CMOS (here there is no guarantee that the logic high output level is sufficient).

1.11 *Latch-up*

Latch-up of CMOS devices occurs with unavoidable parasitic n-p-n-p structures within the IC. It can be triggered by:

1. Overvoltage on an input causing the ESD protection diodes to conduct.

2. Spikes on the power supply.

3. Incorrect sequencing of signals and power supplies.

Latch-up can cause the device to get hot. The device has to be repowered to reset the condition.

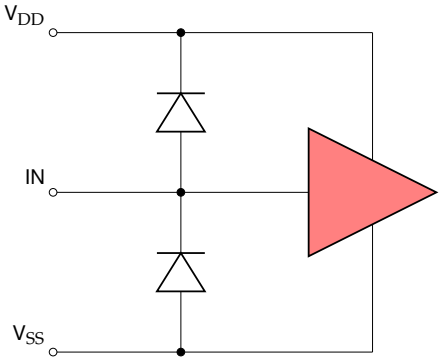


Figure 7.4: CMOS input circuit showing ESD protection diodes.

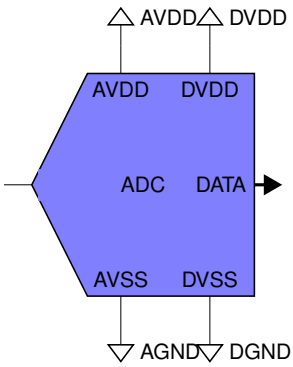


Figure 7.5: Many analogue to digital converters have separate analogue and digital power supplies since digital supplies are usually noisy. For some devices, the sequencing of the power supplies is important.

10 Level translation

When the maximum input voltages are exceeded use a level translation chip[8]. Level translation chips are also available for bidirectional signals.

[8] For example, the SN74LVC4245A octal transceiver translates from 5 V CMOS to/from 3.3 V LVTTL/LVCMOS.

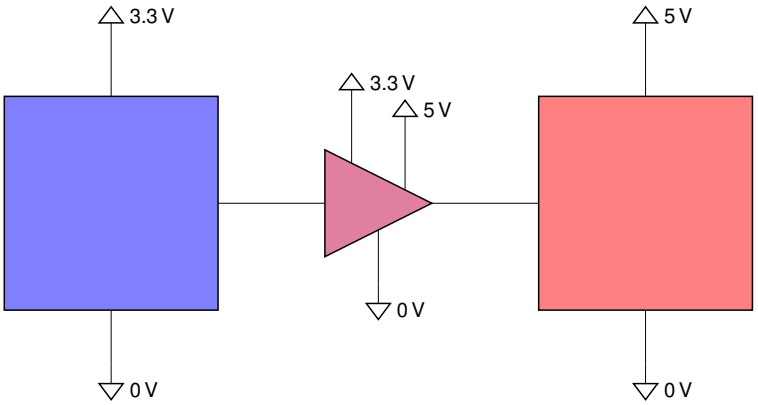


Figure 7.6: Application of a logic level translation chip.

Sometimes a resistor divider can be used[9] to reduce the signals levels although care is needed to ensure that the logic-high noise margin is still maintained.

[9] This will produce a static output current when the output is high.

11 Making a logic level decision

Consider two ICs connected as shown in Figure 7.7. The output driver will try to make the output voltage close to V_{DD} for a logic-high and close to ground for a logic-low. The receiver input then makes a decision: is the output is greater than V_{ih} or less than V_{il}. Unfortunately, the decision making is made unreliable by:

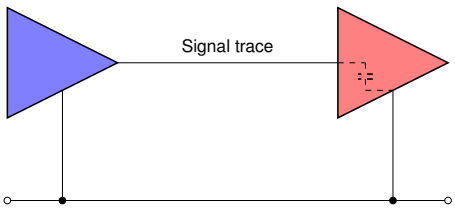


Figure 7.7: A CMOS buffer driving a another CMOS buffer.

1. Thermal noise from driver and receiver.
2. Signal losses due to loading[10].
3. Crosstalk induced by nearby switching signals.
4. Power supply fluctuations of the driver.
5. Power supply fluctuations of the receiver[11].

[10] For example, other ICs and termination.

[11] This can modulate the switching voltage used to make a logic-level decision.

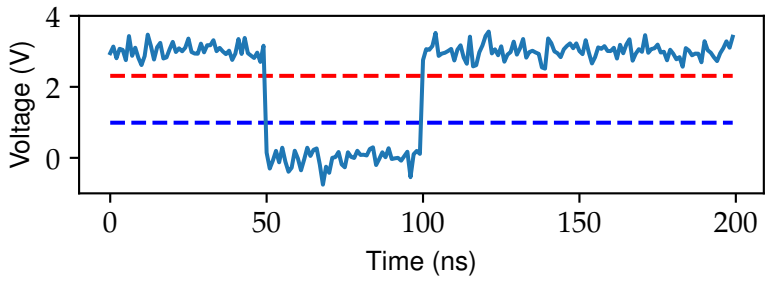


Figure 7.8: Digital signal with noise and the logic thresholds shown as dashed lines.

100 *Example—interfacing an image sensor*

Let's consider the interfacing of an image sensor to a microcontroller as shown in Figure 7.9. We need to consider the input signals to the image sensor and the input signals to the microcontroller. This requires some delving into the electrical characteristics section of the datasheets.

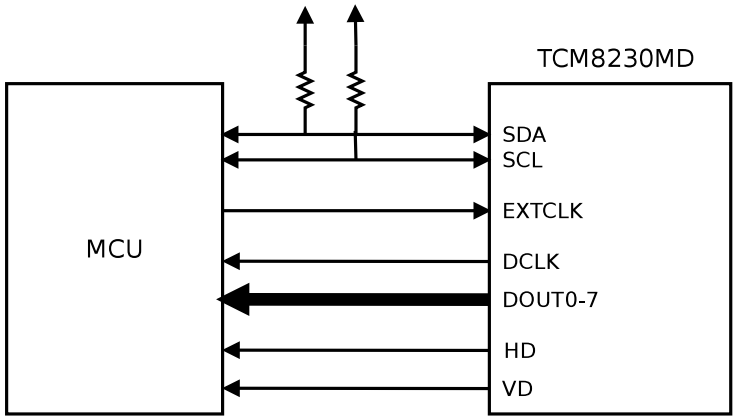


Figure 7.9: A TCM8230MD image sensor interfaced to a SAM7S microcontroller.

	SAM7S	TCM8230MD
V_{DDIO}	3.0–3.6	2.6–3.0
$V_{\text{ol min}}$	0.0	0.0
$V_{\text{ol max}}$	0.4_1	0.4_3
$V_{\text{oh min}}$	$V_{\text{DDIO}} - 0.4_2$	2.4_4
$V_{\text{oh max}}$	V_{DDIO}	V_{DDIO}
$V_{\text{il min}}$	-0.3	-0.3
$V_{\text{il max}}$	0.8	$0.2V_{\text{DDIO}}$
$V_{\text{ih min}}$	2.0	$0.8V_{\text{DDIO}}$
$V_{\text{ih max}}$	5.3	3.0

Table 7.1: V_{DDIO} is the voltage for the `VDDIO` pin on the SAM7S or the `IOVDD` pin on the TCM8230MD. Notes: (1) $I_{\text{ol}} = 8\,\text{mA}$, (2) $I_{\text{ol}} = -8\,\text{mA}$, (3) $I_{\text{ol}} = 2\,\text{mA}$, (4) $I_{\text{ol}} = -2\,\text{mA}$.

100.1 Inputs to the microcontroller

The specifications for the DOUT, HD, VD, and DCLK pins output from the image sensor are given in column 2 of Table 7.1. Let's assume that VDDIO for the SAM7S has a voltage $V_{\text{DDIO}} = 3.3\,\text{V}$ and IOVDD for the image sensor has a voltage $V_{\text{DDIO}} = 2.8\,\text{V}$. The noise margins are:

$$V_{\text{oh min}} - V_{\text{ih min}} = 2.4 - 2.0 = 0.4\,\text{V}, \quad (7.1)$$
$$V_{\text{il max}} - V_{\text{ol max}} = 0.8 - 0.4 = 0.4\,\text{V}. \quad (7.2)$$

These noise margins are both positive and greater than 0.2 V so it appears the signals from the image sensor can be reliably sensed by the SAM7S.

The other consideration is not to exceed the input voltage limits for the SAM7S:

$$V_{\text{oh max}} = 2.8 \leq V_{\text{ih max}} = 5.5, \quad (7.3)$$
$$V_{\text{ol min}} = 0.0 \geq V_{\text{il min}} = -0.3. \quad (7.4)$$

These constraints are true so there is no problem.

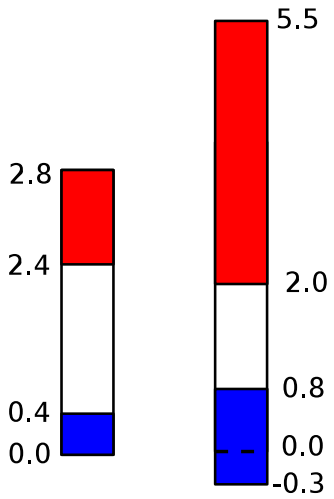


Figure 7.10: Voltage levels for image sensor driving SAM7S.

100.10 Inputs to the image sensor

The EXTCLK signal is a clock input to the image sensor. Again assuming that VDDIO for the SAM7S has a voltage $V_{\text{DDIO}} = 3.3\,\text{V}$ and IOVDD for the image sensor has a voltage $V_{\text{DDIO}} = 2.8\,\text{V}$, then

$V_{\text{il max}} = 0.2 \times 2.8 = 0.56\,\text{V}$
$V_{\text{ih min}} = 0.8 \times 2.8 = 2.24\,\text{V}$
$V_{\text{ol max}} = 0.40\,\text{V}$
$V_{\text{oh min}} = 3.3 - 0.4 = 2.90\,\text{V}$

and thus the noise margins are

$$V_{\text{oh min}} - V_{\text{ih min}} = 2.9 - 2.24 = 0.66\,\text{V}, \quad (7.5)$$
$$V_{\text{il max}} - V_{\text{ol max}} = 0.56 - 0.4 = 0.16\,\text{V}. \quad (7.6)$$

Yikes, the low noise margin is less than 0.2 V.

The other consideration is not to exceed the input voltage limits for the image sensor:

$$V_{\text{oh max}} = 3.3 \leq V_{\text{ih max}} = 3.0, \quad (7.7)$$
$$V_{\text{ol min}} = 0.0 \geq V_{\text{il min}} = -0.3. \quad (7.8)$$

These constraints cannot be met, an thus there is also a chance that we will exceed the maximum input voltage of 3.0 V if the output is 3.3 V.

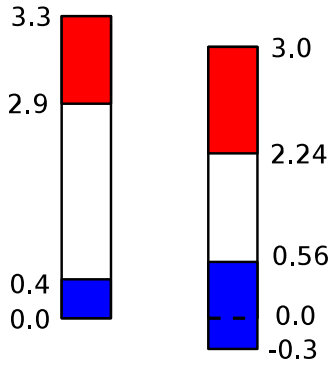


Figure 7.11: Voltage levels for SAM7S driving image sensor.

The SAM7S datasheet states $V_{\text{ol max}} = 0.4\,\text{V}$ for a current of 8 mA, however, with less current it is likely to be closer to zero. So there is a good chance that it will work but an engineer has to play safe. Let's consider some options:

1. Reduce V_{DDIO} on the SAM7S. With $V_{\text{DDIO}} = 3.0\,\text{V}$ then we won't exceed the maximum image sensor input voltage. However, all the I/O from the SAM7S will have to work with this reduced voltage.
2. Use a resistor divider[12]. This will reduce the high noise margin but this is more than adequate.
3. Use a logic level translation chip.

[12] This will increase power consumption and slow down the waveform.

101 CMOS logic levels

The text book definition of the logic levels is where the noise margins are maximised. They are found where the voltage gain is -1, see Figure 7.12.

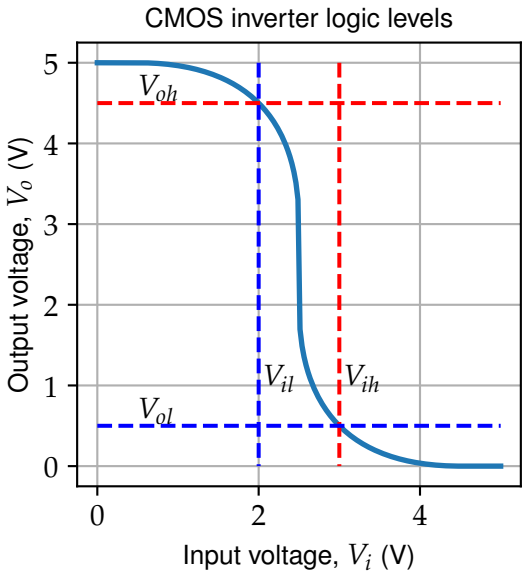


Figure 7.12: CMOS logic levels defined where the voltage gain is -1.

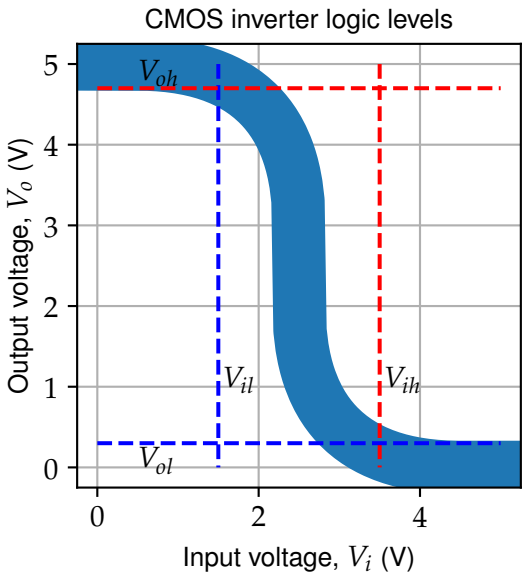


Figure 7.13: CMOS logic levels defined by chip manufacturers. The broad curve represents the variation in transfer function due to manufacturing process variations.

In practice, manufacturers use more conservative logic levels to allow for manufacturing variations, see Figure 7.13. Typically, standard CMOS devices specify V_{il} as $0.3V_{DD}$ and V_{ih} as $0.7V_{DD}$. Thus for 5 V logic, $V_{il} = 1.5\,V$ and $V_{ih} = 3.5\,V$.

110 *Schmitt trigger inputs*

With a non-Schmitt trigger input the device will switch at the same voltage for both a rising edge and falling edge. However, problems can occur if there is noise on the signal or on the power supply and the input voltage is slowly changing around the switching voltage[13].

Schmitt trigger devices use hysteresis[14] to avoid false switching around the switching threshold; they are useful to translate slow edges to accommodate the input rise and fall specifications of the following device.

[13] More subtly, oscillations can occur due to ground bounce since this changes the switching voltage. For example, an output may switch high causing the power supply voltage to momentarily sag. This causes the switching voltage to also sag and thus the input may no longer be considered to be high. The output then switches low, etc.

[14] The switching thresholds are different for rising and falling signals.

111 Further reading

Designing with logic, Eilhard Haselhof, Texas Instruments, 1997, www.ti.com/lit/an/sdya009c/sdya009c.pdf

Low voltage logic devices, Peter Forstner, Texas Instruments, 1997, www.ti.com/lit/an/scvae01a/scvae01a.pdf

Interfacing Between LVPECL, VML, CML, and LVDS Levels, TI, 2002, www.ti.com/lit/an/slla120/slla120.pdf

Sorting through the low voltage logic maze, Philips, ics.nxp.com/support/techdocs/logic/pdf/an10156.pdf

How to select little logic, TI, 2010, www.ti.com/lit/an/scya049/scya049.pdf

Designing with logic, TI, sdya009c.pdf 1997

Understanding and interpreting standard-logic data sheets, TI, 2003, www.ti.com/lit/an/szza036b/szza036b.pdf

Logic guide, TI, 2009, www.ti.com/lit/an/sdyu001z/sdyu001z.pdf

Understanding schmitt triggers, TI, 2011, www.ti.com/lit/an/scea046/scea046.pdf

http://www.allaboutcircuits.com/vol_4/chpt_3/10.html

1000 Exercises

1. What is meant by the noise margin of a logic device?
2. Why is it undesirable to drive 3.3 V CMOS logic from 5 V CMOS logic?
3. Suggest how 3.3 V CMOS logic can be safely driven from 5 V CMOS logic?
4. For logic level compatibility, what two conditions must be satisfied?
5. The AT91SAM7S256 microcontroller has VDDIO of 3.3 V. Its datasheet says that its inputs are 5 V tolerant. What does this imply?
6. What is latch-up?

7. What is a Schmitt trigger?
8. Consider the example in Section 100 but with $V_{\mathrm{DDIO}} = 2.5\,\mathrm{V}$ for the SAM7S. Without changing V_{DDIO}, how would you ensure reliable operation?
9. What is meant by power supply sequencing?

	SAM7S	SAM4S
V_{DDIO}	3.0–3.6	1.62–3.6
$V_{\mathrm{ol\,min}}$	0.0	
$V_{\mathrm{ol\,max}}$	0.4_1	0.15
$V_{\mathrm{oh\,min}}$	$V_{\mathrm{DDIO}} - 0.4_2$	$V_{\mathrm{DDIO}} - 0.4$
$V_{\mathrm{oh\,max}}$	V_{DDIO}	
$V_{\mathrm{il\,min}}$	-0.3	-0.3
$V_{\mathrm{il\,max}}$	0.8	$\max(0.8, 0.3V_{\mathrm{DDIO}})$
$V_{\mathrm{ih\,min}}$	2.0	$\min(2.0, 0.7V_{\mathrm{DDIO}})$
$V_{\mathrm{ih\,max}}$	5.3	$V_{\mathrm{DDIO}} + 0.3$

Table 7.2: Comparison of logic levels between the SAM7S and the newer SAM4S. V_{DDIO} is the voltage for the `VDDIO` pin. Notes: (1) $I_{\mathrm{ol}} = 8\,\mathrm{mA}$, (2) $I_{\mathrm{ol}} = -8\,\mathrm{mA}$. The SAM4S datasheet does not specify load conditions.

8
CMOS switching

When a digital logic device switches high it needs to source current to charge the load capacitance[1]. When it switches low, it needs to discharge the load capacitance. The switching speed depends on how much current the driver can source (or sink) and the total load capacitance.

[1] The output capacitance of the device plus the capacitance of the loads and connecting traces.

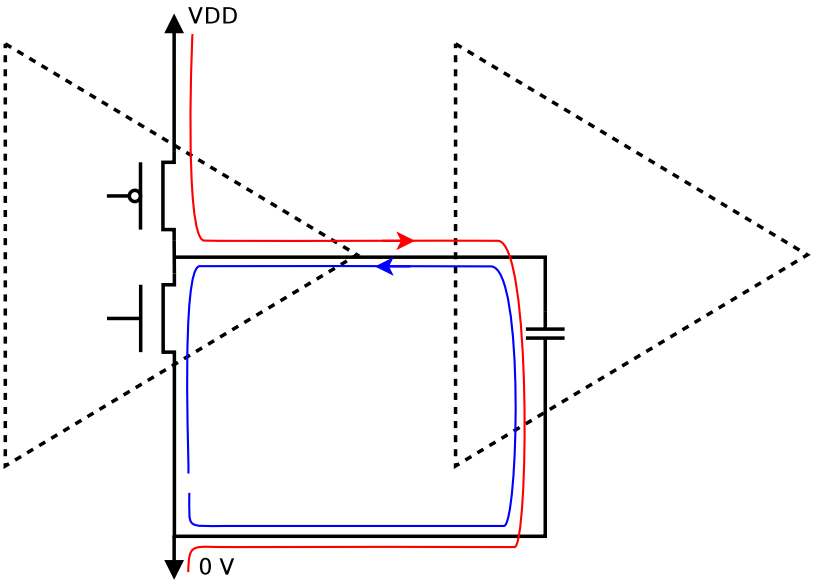


Figure 8.1: Switching current paths. Current is required from the power supply to charge the load capacitance (and the output capacitance of the driver).

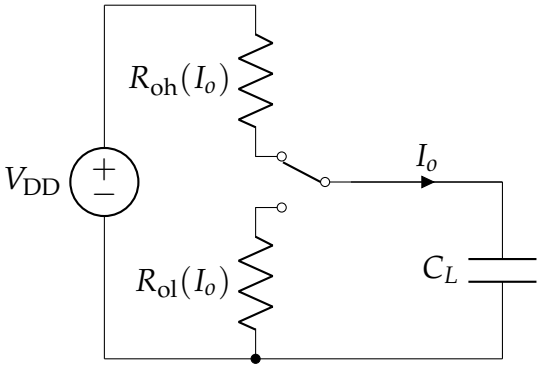


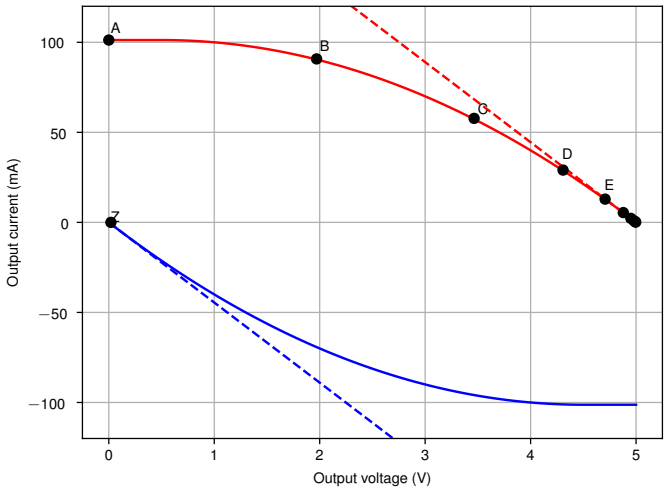
Figure 8.2: Simplified electrical model for CMOS output driving a load capacitor. Note R_{oh} and R_{ol} both vary with current.

1 CMOS switching time

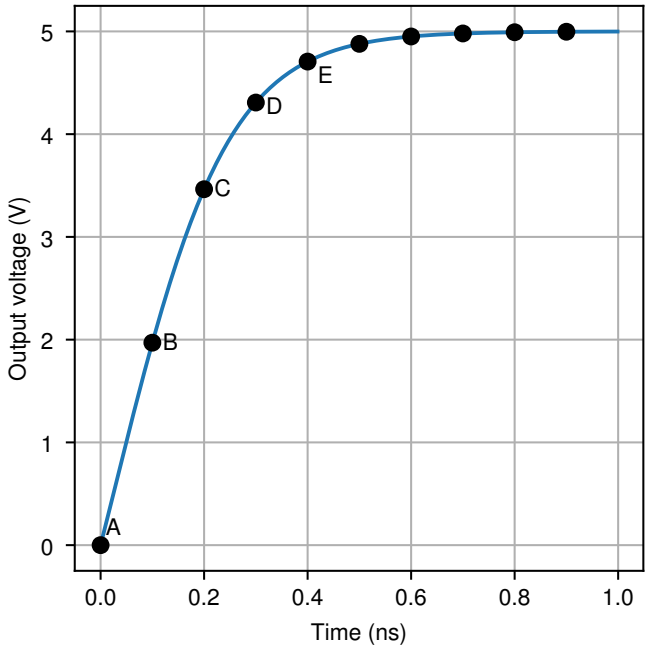
When a CMOS inverter switches from low to high the PMOS transistor switches on and the NMOS transistor switches off. However, the output voltage will not change as rapidly due to capacitive loading[2].

Initially the voltage rises linearly since the output current is saturated; it then rises more slowly as the PMOS transistor moves into the linear region (see Figure 8.3).

[2] This is the sum of intrinsic capacitance due to the output driver MOSFETs and connections and extrinsic capacitance due to input capacitance and stray capacitance.



(a)



(b)

Figure 8.3: (a) CMOS output IV curves (b) CMOS low to high switching response for 5 pF load. The labelled points are separated by intervals of 0.1 ns. It is assumed that the transition from Z to A (when the PMOS transistor switches on and the NMOS transistor switches off) has a negligible time.

10 Resistive loading

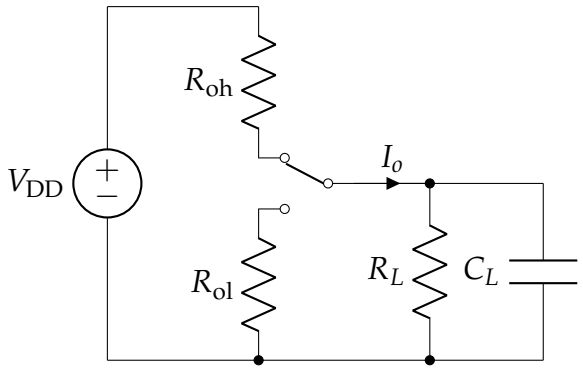


Figure 8.4: Simplified electrical model for CMOS output driving a load comprised of a capacitor in parallel with a resistor.

If a CMOS driver is loaded with a parallel pull-down resistor, then current is diverted from charging the load. This slows down the charge rate but speeds up the discharge rate. As can be seen from Figure 8.5(a), the output high voltage is decreased (reducing the high noise margin) and the driver needs to supply a static current.

11 Switching waveforms

In practice, the output of a logic device does not change linearly with time, as illustrated in Figure 8.6. Consequently, it is common to specify the rise and fall times between the 10% and 90% levels[3].

[3] This also ignores over and undershoot. Manufacturers typically specify the switching times for one output changing at a time. If multiple chip outputs switch at the same time then rise and fall times are slower. This is due to the inductance of the power supply pins and die interconnect wires limiting the current that can be supplied.

11.1 Signal bandwidth

The bandwidth of a signal is inversely proportional to the rise/fall times. One measure of the signal bandwidth is the *knee frequency*, commonly defined as

$$f_{\text{knee}} = \frac{0.5}{t_{10-90\%}}, \tag{8.1}$$

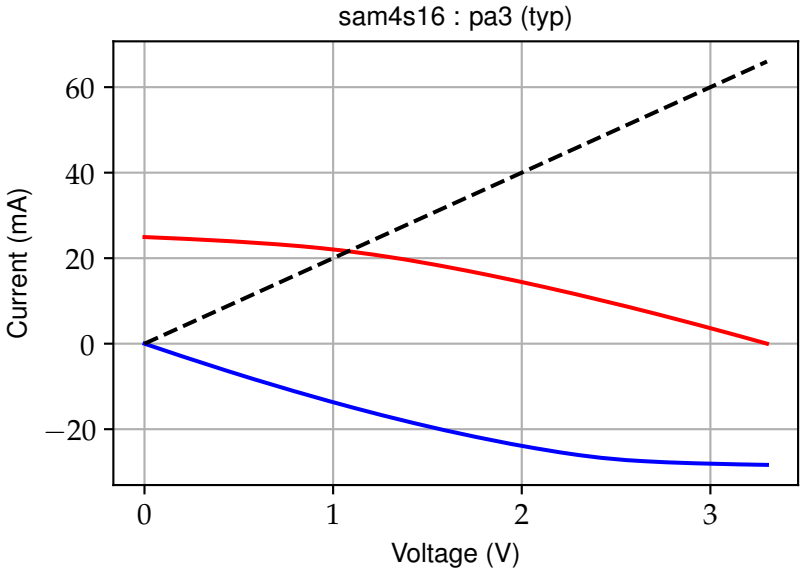
where $t_{10-90\%}$ is the 10% to 90% signal rise (or fall) time[4]. The knee frequency approximates the highest frequency component for sinusoidal analysis. It is a conservative overestimate of the signal bandwidth.

[4] This is usually the rise/fall time specified by the manufacturer.

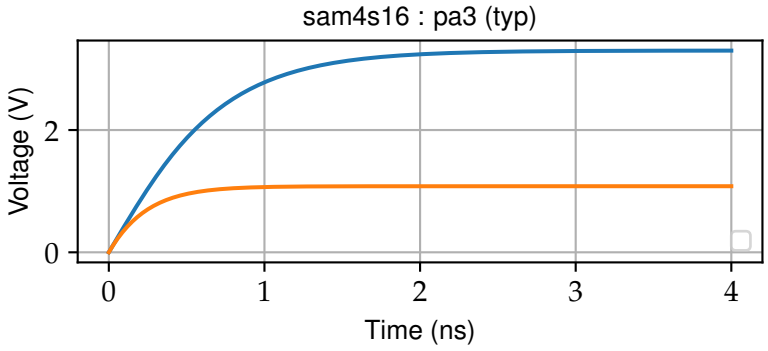
For example, consider a 50 MHz clock signal with 1 ns rise/fall times driven through a SMD pin adding 4 nH of inductance. The knee frequency is 500 MHz[5] and the pin reactance is

[5] Independent of the switching frequency.

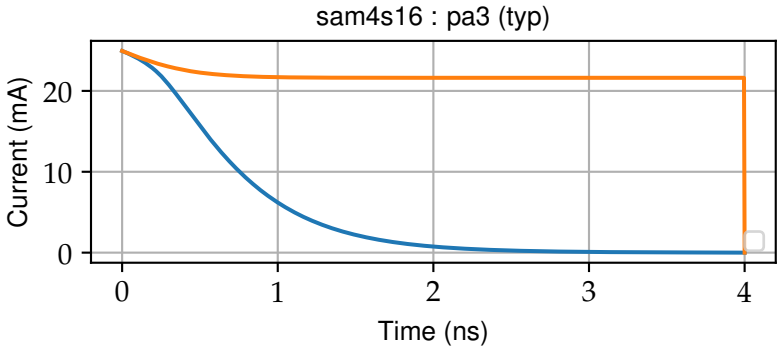
$$\begin{aligned} X_L &= 2\pi f_{\text{knee}} L, \\ &= 2\pi \times 500 \times 10^6 \times 4 \times 10^{-9}, \\ &= 12.6\,\Omega. \end{aligned} \tag{8.2}$$



(a)



(b)



(c)

Figure 8.5: SAM4S switching from low to high with and without 50 Ω load in parallel with 3 pF load (in addition to package and component capacitance of the pin): (a) IV curve, (b) time response, (c) output current. The dashed curve is the load line for a load resistance of 50 Ω.

Rise time (ns)	Knee frequency (MHz)
1	500
2	250
5	100
10	50

Table 8.1: Relation between rise time and knee frequency using (8.1).

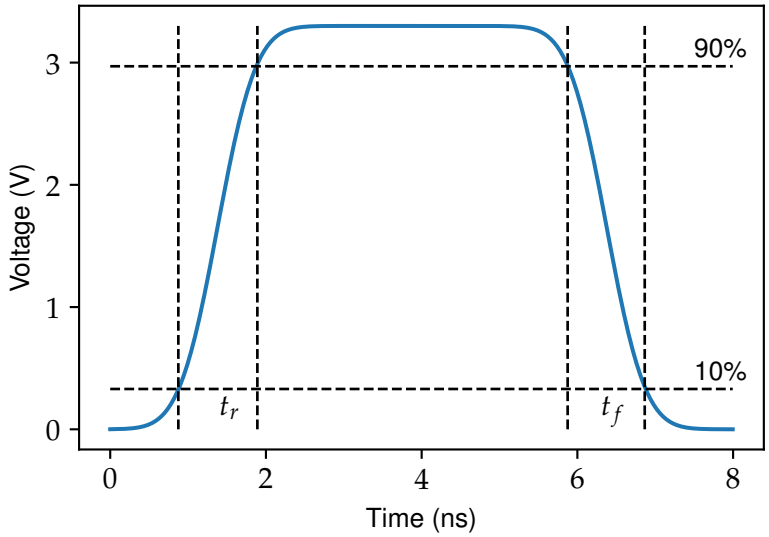


Figure 8.6: CMOS switching waveforms showing rise and fall times t_r and t_f, typically defined between the 10% and 90% levels.

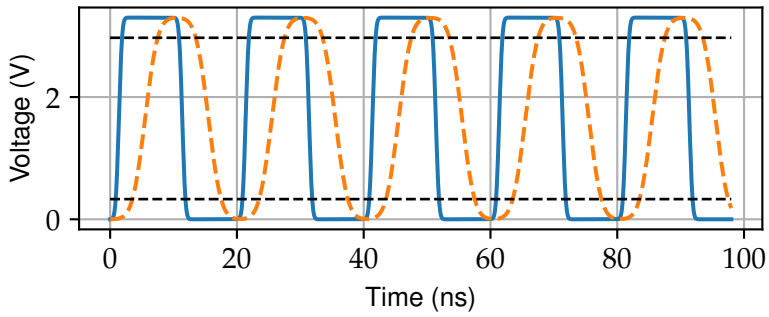


Figure 8.7: CMOS 50 MHz clock waveform with rise/fall time 1 ns (4 ns dashed).

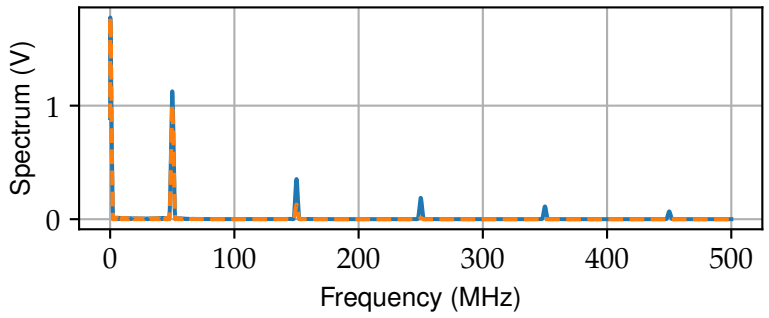


Figure 8.8: Spectrum of CMOS 50 MHz clock waveform with rise/fall time 1 ns (4 ns dashed). Note, the slower signal has reduced harmonics.

100 IBIS models

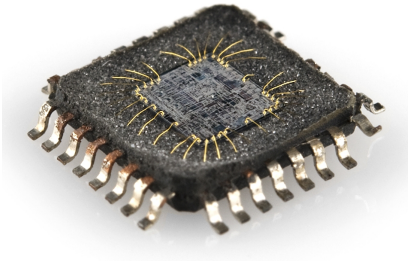


Figure 8.9: Bond wires interconnects between die and package pins.

High frequency analysis of a logic device requires a model of the parasitic components. This information is specified in an IBIS[6] text file that contains:

- parasitic components due to packaging (such as interconnect wire inductance, see Figure 8.9)
- a model for each pin, this includes:
  - component capacitance,
  - a table showing the relationship between pin voltage and current[7],
  - a table showing the time response for an output switching high and low.

[6] I/O buffer information specification. Manufacturers are too cagey to supply full SPICE models.

[7] For each possible pin state, logic-high, logic-low, tristate.

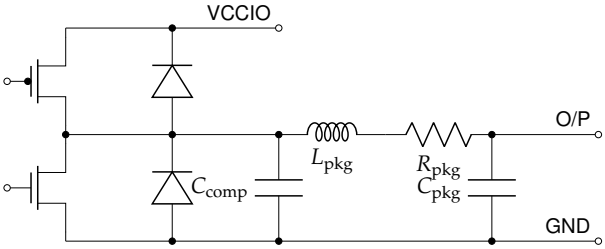


Figure 8.10: IBIS output model. The parasitic components are necessary for high frequency modelling. The diodes are for electrostatic protection.

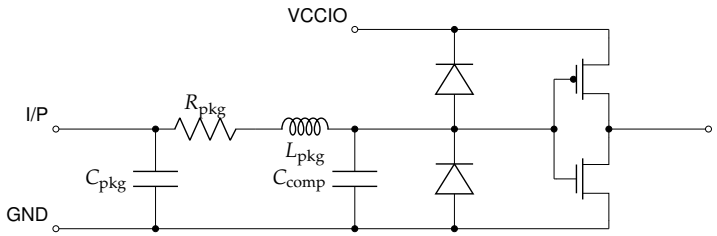


Figure 8.11: IBIS input model.

101 Example IBIS model

Snippets of an IBIS model for an AT91SAM4S microcontroller are shown in Figure 8.12 and Figure 8.13.

An IBIS file contains four types of current/voltage curve:

GND clamp This is measured when the buffer is tristate and illustrates the behaviour of the ground clamp ESD structure. The voltage range is from $-V_{CC}$ to V_{CC}.

Power clamp This is measured when the buffer is tristate and illustrates the behaviour of the power clamp ESD structure. The voltage range is from V_{CC} to $2V_{CC}$ but is referenced to V_{CC}[8].

Pulldown This is measured when the buffer is a logic low (the NMOS pull-down transistor is on). Confusingly, this curve is the logic low I/V curve with the GND clamp curve subtracted. The voltage range is from $-V_{CC}$ to $2V_{CC}$.

Pullup This is measured when the buffer is a logic high (the PMOS pull-up transistor is on). Confusingly, this curve is the logic high I/V curve with the power clamp curve subtracted[9]. The voltage range is from $-V_{CC}$ to $2V_{CC}$.

Note, the voltage range for these curves is from $-V_{CC}$ to $2V_{CC}$ since this is the possible range of voltages when reflections from a transmission line are present.

[8] The output voltage is $V_o = V_{CC} - V_{table}$, where V_{table} is the voltage tabulated in the file.

[9] The output voltage is $V_o = V_{CC} - V_{table}$, where V_{table} is the voltage tabulated in the file.

Component	SAM4S	SAM7
C_{comp}	2.8 pF	5 pF
C_{pkg}	0.5 pF	0.5 pF
R_{pkg}	110 mΩ	110 mΩ
L_{pkg}	3.8 nH	3.8 nH

Table 8.2: Typical values of IBIS model components for PIO pin PA0 on the AT91SAM7S and AT91SAM4S MCUs (with same LQFP package). The SAM4S is newer than the SAM7 and uses smaller MOSFETs.

```
[IBIS Ver]      3.2
[Comment Char] |_char
[File Name]     sam4s16_lq64.ibs
[File Rev]      1.0
[Date]          Mar 07 2012
|
[Component]     SAM4S16_LQ64
[Manufacturer] ATMEL Corporation
|
[Package]
| variable    typ        min        max
R_pkg         110mOhm    20mOhm    300mOhm
L_pkg         3.80nH     1.00nH    7.49nH
C_pkg         0.500pF    0.100pF   1.300pF
|
[Pin] signal_name  model_name      R_pin L_pin
   C_pin
|
1     NC           NC
2     GND          GND
3     PB0          b3ha40bp2ruarm_cfg1
4     PB1          b3ha40bp2ruarm_cfg1
5     PB2          ha40bp2ruarm_cfg1
6     PB3          ha40bp2ruarm_cfg1
7     VDDIN        POWER
8     VDDOUT       POWER
9     PA17         ha40bp2ruarm_cfg1
10    PA18         ha40bp2ruarm_cfg1
11    PA21         ha40bp2ruarm_cfg1
12    VDDCORE      POWER
13    PA19         ha40bp2ruarm_cfg1
14    PA22         ha40bp2ruarm_cfg1
15    PA23         hb40bp2ru_cfg1
16    PA20         ha40bp2ruarm_cfg1
17    GND          GND
18    VDDIO        POWER
<snip>
47    PA1          hb40bp2ru_cfg1
48    PA0          hb40bp2ru_cfg1
<snip>
[Model Selector] hb41bp2ru_cfg1
hb41bp2ru3v3           Model 3.3V
hb41bp2ru1v8           Model 1.8V
```

Figure 8.12: Snippet from IBIS model for a SAM4S microcontroller. This shows the values of the parasitic components and the model to use for each pin.

Figure 8.13: Snippet from IBIS model for a SAM4S microcontroller showing part of model (hb41bp2ru3v3) that describes most of the PIO pins.

```
[Model] hb41bp2ru3v3
Model_type     I/O
Polarity       Non-Inverting
Enable         Active-Low
Vinl = 0.8
Vinh = 2.0
Vmeas = 1.65
Vref = 1.65
Cref = 15pF
Rref = 50
|
|                         typ      min      max
|
C_comp                    2.474pF  2.403pF  2.637pF
|C_comp (ON state)        2.099pF  1.998pF  2.386pF
|C_comp (OFF state)       2.849pF  2.808pF  2.887pF

[Voltage Range]           3.3      2.85     3.6
[Temperature Range]       25       125      -40
[Pullup Reference]        3.3      2.85     3.6
[Pulldown Reference]      0        0        0
[Power Clamp Reference]   3.3      2.85     3.6
[GND Clamp Reference]     0        0        0
|
|*******************************************************************************

|
[Pulldown]
|
|
|    Voltage   I(typ)       I(min)       I(max)
|
     -3.3      -0.00938453  -0.00184193  -0.0142503
     -2.6445   -0.0116379   -0.0036782   -0.0142677
     -2.0031   -0.0131129   -0.00679987  -0.0141747
     -2.0004   -0.013116    -0.00681389  -0.0141735
     -1.7736   -0.0132891   -0.00799059  -0.0140411
     -1.3662   -0.013221    -0.00991479  -0.0136106
<snip>
[Pullup]
|
|
|    Voltage   I(typ)      I(min)      I(max)
|
     -3.3      0.0101941   0.00254488  0.0156987
     -2.8041   0.0121699   0.00432029  0.0170764
     -2.4036   0.0137279   0.00635637  0.0180544
     -2.1351   0.014716    0.0077917   0.0185027
     -1.9716   0.0152698   0.00865912  0.0185264
     -1.6137   0.0161431   0.0104716   0.0168654
     -1.6128   0.0161441   0.0104759   0.0168598
     -1.3794   0.0158603   0.0114987   0.0157037
     -1.34775  0.015722    0.0116182   0.0155715
     -1.3464   0.0157158   0.0116232   0.015566
<snip>
```

110 RC circuit theory revisited

Consider a capacitor, of capacitance C, connected through a resistor, of resistance R, to a voltage source of voltage V through a switch. Initially the capacitor is discharged and thus there is no voltage across it[10]. At time $t = 0$, the switch is closed. The voltage across the resistor is now V and thus an initial current $I_0 = V/R$ flows to charge the capacitor. This current causes the voltage across the capacitor to rise. However, as the voltage across the capacitor rises, the voltage across the resistor drops, and thus the current into the capacitor drops. As a result the current decreases exponentially with a time constant $\tau = RC$. Mathematically, the circuit current can be expressed as,

[10] The voltage, V, across a capacitor is given by Q/C, where Q is the charge on the capacitor. Charging a capacitor is analogous to filling a swimming pool with buckets of water. The more buckets of water in the pool (the charge), the higher the water level (the voltage). A pool with a greater capacity, C, requires more bucketfuls to raise the level.

$$i(t) = \frac{V}{R} \exp\left(-\frac{t}{\tau}\right) u(t), \tag{8.3}$$

where $u(t)$ is the unit step. The voltage across the capacitor is given by

$$\begin{aligned} v(t) &= \frac{1}{C}\int_0^t i(t')\mathrm{d}t', && (8.4)\\ &= V\left(1 - \exp\left(-\frac{t}{\tau}\right)\right) u(t), && (8.5) \end{aligned}$$

where the integral gives the accumulated charge, $q(t)$. The total charge stored on the capacitor is

$$\begin{aligned} Q &= \int_0^\infty i(t)\mathrm{d}t, && (8.6)\\ &= \frac{V}{R}\int_0^\infty \exp\left(-\frac{t}{\tau}\right)\mathrm{d}t, && (8.7)\\ &= \frac{V}{R}\left[-\tau \exp\left(-\frac{t}{\tau}\right)\right]_0^\infty, && (8.8)\\ &= VC. && (8.9) \end{aligned}$$

The instantaneous power stored in the capacitor is given by

$$\begin{aligned} p(t) &= v(t)i(t), && (8.10)\\ &= \frac{V^2}{R}\left(1 - \exp\left(-\frac{t}{\tau}\right)\right)\exp\left(-\frac{t}{\tau}\right) u(t) && (8.11) \end{aligned}$$

and the total energy stored in the capacitor is

$$\begin{aligned} E_c &= \int_0^\infty p(t)\mathrm{d}t, && (8.12)\\ &= \frac{V^2}{R}\int_0^\infty \exp\left(-\frac{t}{\tau}\right) - \exp\left(-\frac{2t}{\tau}\right)\mathrm{d}t, && (8.13)\\ &= \frac{1}{2}V^2C. && (8.14) \end{aligned}$$

The energy supplied by the voltage source is

$$\begin{aligned} E &= V\int_0^\infty i(t)\mathrm{d}t, && (8.15)\\ &= \frac{V^2}{R}\int_0^\infty \exp\left(-\frac{t}{\tau}\right)\mathrm{d}t, && (8.16)\\ &= V^2C, && (8.17) \end{aligned}$$

and thus the energy dissipated by the resistor is equal to the energy stored on the capacitor[11]. Interestingly, the energy dissipated by the resistor when charging the capacitor is independent of the resistance. All the resistance changes is how long it takes to charge the capacitor. But then again, is the capacitor ever fully charged? As its voltage gets closer to that of the voltage source the charging current gets smaller.

So what happens if we make the resistance zero? Well the capacitor will charge instantaneously[12] due to the infinite current that exists for zero time. This leads us into the murky world of impulses[13]. The current can be described by

$$i(t) = VC\delta(t). \qquad (8.18)$$

The product VC is the charge, Q, that gets stored on the capacitor and since $\delta(t)$ has units of $1/s$ both sides of the equation have the units of amps[14].

The energy delivered from the power supply again is V^2C and the energy stored in the capacitor is $0.5V^2C$. So what happened to the other $0.5V^2C$ of energy? How is this dissipated if there is no resistance? Fun with impulses, eh!

[11] This result can also be obtained by integrating the power $Ri^2(t)$ dissipated by the resistor.

[12] The time constant is zero.

[13] This is mathematical shorthand for an implicit limit. They are not functions.

[14] The unit of charge is the coulomb equivalent to the product of amps times seconds.

111 *An idealised IV curve for a CMOS driver*

An idealised IV curve for a CMOS driver with the PMOS pull-up transistor on can be expressed mathematically as

$$I_{\text{sat}} = \begin{cases} I_{\text{sat}} & V < V', \\ I_{\text{sat}}\left(\frac{V_{\text{DD}}-V}{V_{\text{DD}}-V'}\right) & V > V'. \end{cases} \qquad (8.19)$$

Here I_{sat} is the saturation current and V' is the output voltage below which current saturation occurs. When $V > V'$ the output buffer acts like a Thévenin source with output resistance,

$$R_o = \frac{V_{\text{DD}} - V'}{I_{\text{sat}}}. \tag{8.20}$$

If the pull-up transistor instantaneously switches on at $t = 0$, the voltage on a load capacitance, C, will linearly rise due to the constant current output. Once, $V > V'$ the current will exponentially decrease. The output current can thus be expressed as

$$i(t) = \begin{cases} 0 & t < 0, \\ I_{\text{sat}} & 0 \leq t < t', \\ I_{\text{sat}} \exp\left(-\frac{(t-t')}{R_o C}\right) & t \leq t', \end{cases} \tag{8.21}$$

where

$$t' = \frac{CV'}{I_{\text{sat}}}. \tag{8.22}$$

In this interval, the energy extracted from the power supply is

$$E_s = V_{\text{DD}} I_{\text{sat}} t' = V_{\text{DD}} V' C, \tag{8.23}$$

the energy stored in the capacitance is

$$E_c = \frac{1}{2} V'^2 C, \tag{8.24}$$

and thus the energy converted to heat in the PMOS transistor is

$$\begin{aligned} E_p &= E_s - E_c, & (8.25) \\ &= CV'\left(V_{\text{DD}} - \frac{V'}{2}\right). & (8.26) \end{aligned}$$

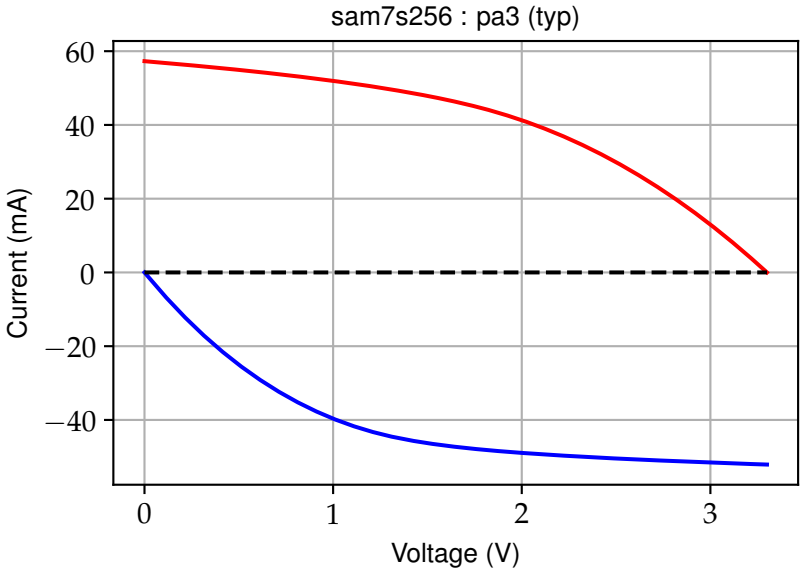
The efficiency[15] in charging the capacitance is

$$\eta = \frac{E_c}{E_s} = \frac{V'}{2V_{\text{DD}}}. \tag{8.27}$$

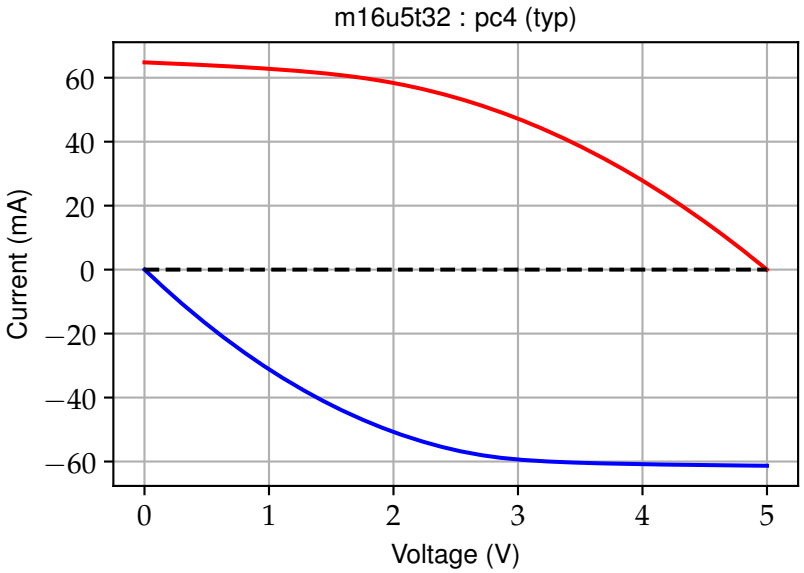
[15] The best efficiency of 0.5 occurs when $V' = V_{\text{DD}}$.

1000 *Further reading*

A First Approach to IBIS Models: What They Are and How They Are Generated, Application Note AN-175, Mercedes Casamayor, Analog Devices, 2004. This is an excellent simple introduction to IBIS models.



(a)

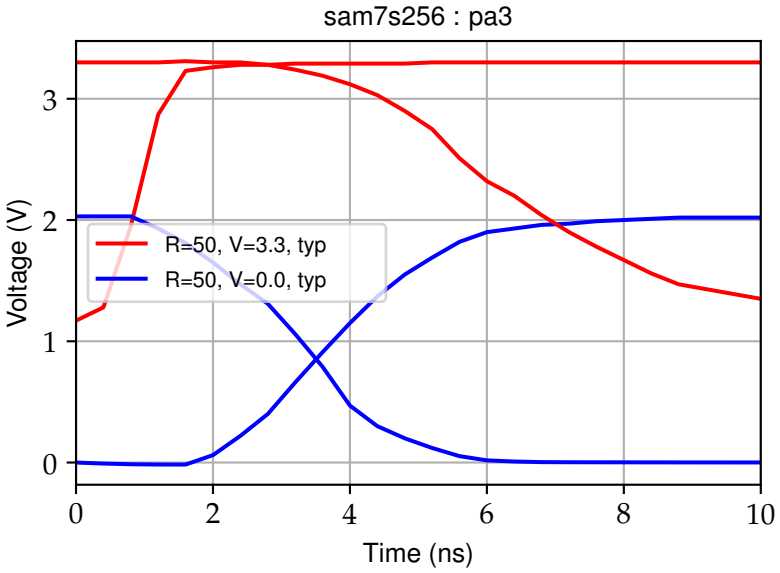


(b)

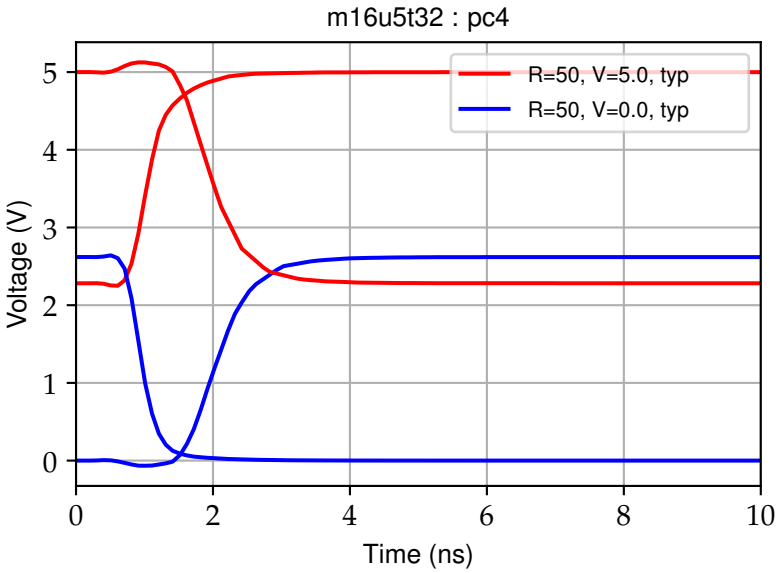
Figure 8.14: Typical IV curves: (a) ATSAM7S256 (b) ATmega32u2.

1001 *Exercises*

1. What is the purpose of an IBIS model?
2. When can the parasitic components in the IBIS model be ignored?
3. What information is tabulated in an IBIS model?
4. How are rise and fall times usually defined?
5. What is meant by knee frequency?



(a)



(b)

Figure 8.15: Typical switching waveforms driving into 50 Ω load in parallel with C_{comp}: (a) ATSAM7S256, $C_{comp} = 5.0, pF$ (b) ATmega32u2, $C_{comp} = 2.8\,pF$. Notice how both chips struggle to drive a 50 Ω load. The peak voltages can be found by plotting a resistive load line on the IV curves.

6. Why does the output voltage from a CMOS driver not immediately rise linearly?
7. What factors dominate the rise time of a CMOS driver output?
8. What is the primary cause of switching noise in a high-speed digital circuit?
9. Does a CMOS buffer with greater current drive produce more or less switching noise?
10. If a SMD lead adds 4 nH of inductance, what is its

reactance for a switching signal with 0.5 ns rise and fall times?

11. If a power supply circuit adds 40 nH of inductance, what is its reactance for a switching signal with 1.0 ns rise and fall times?

12. If the rise time for a 5 pF load is 2 ns, what is the approximate rise time for a 10 pF load?

13. The performance of a high-resolution ADC can be improved by slowing down the output signals. This can be achieved by adding series resistors or parallel capacitors to the output pins. Which of the two is better for battery life?

14. Consider the IV curve in Figure 8.16 with a superimposed load line. Determine the loading, the maximum voltage swing, and the steady state current when the output is high.

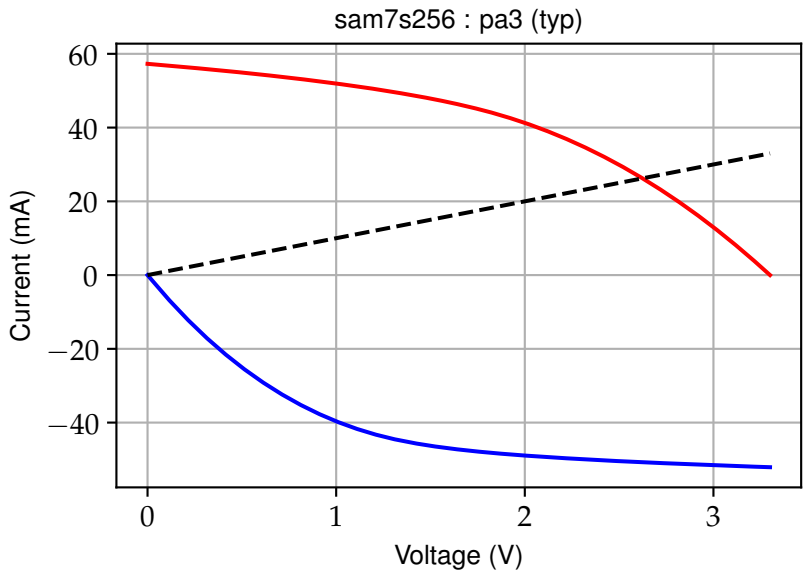


Figure 8.16: SAM7S PIO IV curve with load line (dashed).

15. Consider the IV curve in Figure 8.17 with a superimposed load line. Determine the loading, the maximum voltage swing, and the steady state current when the output is low.

16. The SAM4S datasheet says that maximum output frequency for the PA14 pin is 70 MHz with $V_{DDIO} = 1.62$ V and a load of 10 pF. If $V_{DDIO} = 3.3$ V, would the maximum frequency be higher, lower, or the same?

17. The SAM4S datasheet says that maximum output frequency for the PA14 pin is 70 MHz with $V_{DDIO} =$

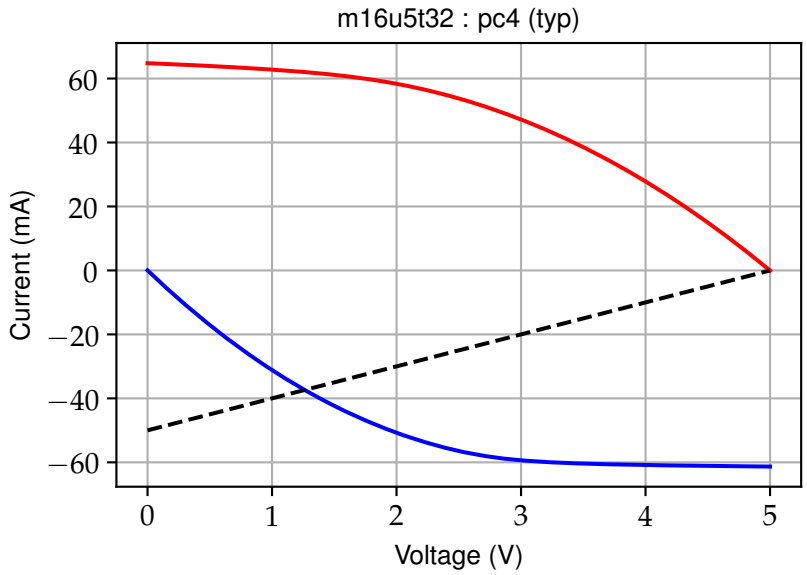


Figure 8.17: ATmega32u2 PIO IV curve with load line (dashed).

1.62 V and a load of 10 pF but 45 MHz with a load of 30 pF at the same voltage. Explain why the speed is not three times slower with three times the load capacitance.

9
Common path noise

Power supplies are a critical component of high-speed digital and mixed-signal systems since they need to:

1. Provide stable voltage references[1].
2. Distribute power uniformly to all devices.

[1] Necessary for logic level decisions.

Consider Figure 9.1 showing circuits with different power supply connections for an radio module with an MCU. Which circuit is likely to work best?

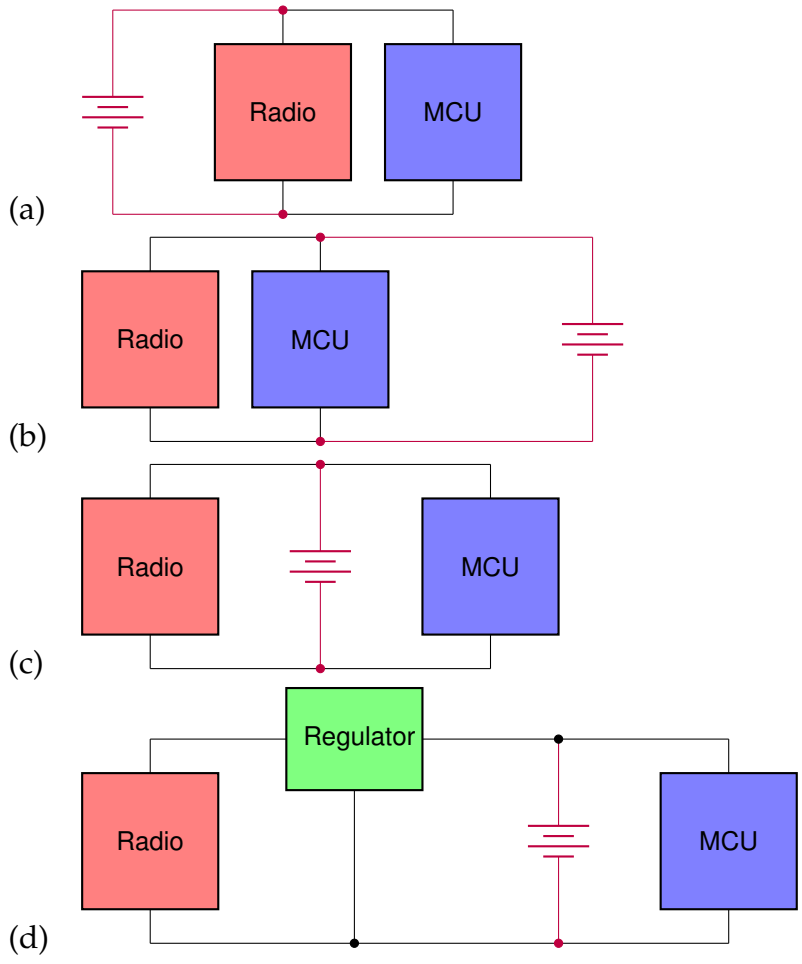


Figure 9.1: Alternatives circuit configurations of an radio module and MCU connected to a battery. Which is worst for the radio performance?

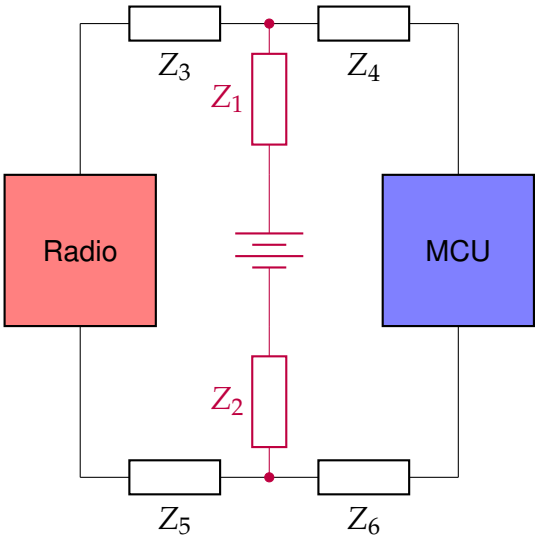


Figure 9.2: To determine which circuit has the least common path noise it is necessary to consider the impedance of each power supply loop; for high frequencies this is dominated by inductance. In this example, the common-path is shown in purple. The goal is to minimize Z_1 and Z_2.

To find the best PCB layout we must realise that each wire or PCB trace has impedance. *Common path noise* results from switching currents flowing through the impedance of *shared* power supply connections[2], see Figure 9.2. It reduces noise margins and can introduce false clocking.

At low frequencies the impedance of wires and traces is primarily resistive but at high frequencies it is dominated by inductance. Even small inductances can induce large noise voltages, since the voltage drop across an inductor depends on the rate of change of current,

$$v(t) = L\frac{\mathrm{d}}{\mathrm{d}t}i(t), \tag{9.1}$$

and $\mathrm{d}i/\mathrm{d}t$ can be large due to fast switching rates.

[2] This is analogous to house plumbing systems where turning on a tap can affect the shower temperature.

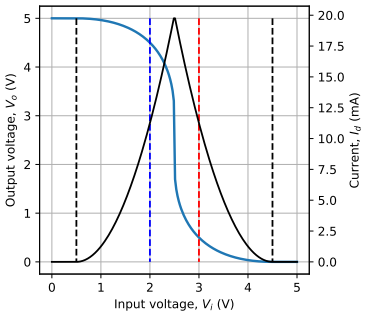


Figure 9.3: CMOS inverter transfer function with short-circuit current.

1 Common path noise example

Consider a MCU powered via a 75 mm long, 3 mm wide microstrip on a standard two-layer PCB[3]. This microstrip has an inductance of 0.27 μH/m, so a length 75 mm has an inductance of about $L = 20$ nH. Assuming a knee frequency of 250 MHz[4], the reactance is

$$\begin{aligned} X_L &= 2\pi f_{\text{knee}} L, \\ &= 2\pi \times 250 \times 10^6 \times 20 \times 10^{-9}, \\ &\approx 30\,\Omega. \end{aligned} \tag{9.2}$$

[3] $H = 1.6$ mm, $T = 33\,\mu$m.

[4] 2 ns rise/fall times.

If the current ripple amplitude is 50 mA, then the voltage drop has an amplitude

$$\begin{aligned} V &= I X_L, \\ &\approx 50 \times 10^{-3} \times 30, \\ &\approx 1.5\,\text{V}. \end{aligned} \tag{9.3}$$

This is larger than the noise margins.

10 Common path noise reduction

Common path noise can be reduced by:

1. Careful separation of high current and high frequency components from low level analogue components.
2. Separate power supplies[5].
3. Minimising inductance of power supply (by minimising loop area, preferably using power and ground planes; where this is not possible, make traces as wide as possible and run power and ground close together).
4. Use of 'star' points (when planes cannot be used).
5. Power supply decoupling.
6. Slowing down rise/fall times of output signals with series resistors[6].

[5] Regulators are good for low frequencies; at high frequency ferrite beads are required.

[6] This is a good idea for mixed-signal devices such as the outputs of ADCs.

Note that planes for power and ground connections provide the lowest impedance. At low speeds current follows the path of least resistance. At high speeds current follows the path of least inductance[7] since the reactance dominates the resistance.

[7] When using power and ground planes, the return current flows under the supply current since this minimises the loop area.

11 Ground/power bounce

The inductance introduced by leads of devices causes ground (and power) bounce whenever the device output switches low and discharges the load capacitance. The changing current induces a voltage across the lead inductance[8]. With many outputs switching, the induced ground bounce voltage modulates the voltage reference used to make an input logic level decision. Note, ground bounce cannot be measured with a scope—the induced voltage is internal to the package.

[8] This is a form of common path noise on the ground lead of a chip.

11.1 *Ground/power bounce mitigation*

Techniques to mitigate ground/power bounce include:

1. Packages with short pins[9] (to reduce the pin inductance)
2. Small packages (to reduce the bond wire inductance)
3. Multiple power/ground pins
4. Ground sense pins
5. Differential signal inputs
6. Avoidance of sockets

[9] For example, BGA and QFN.

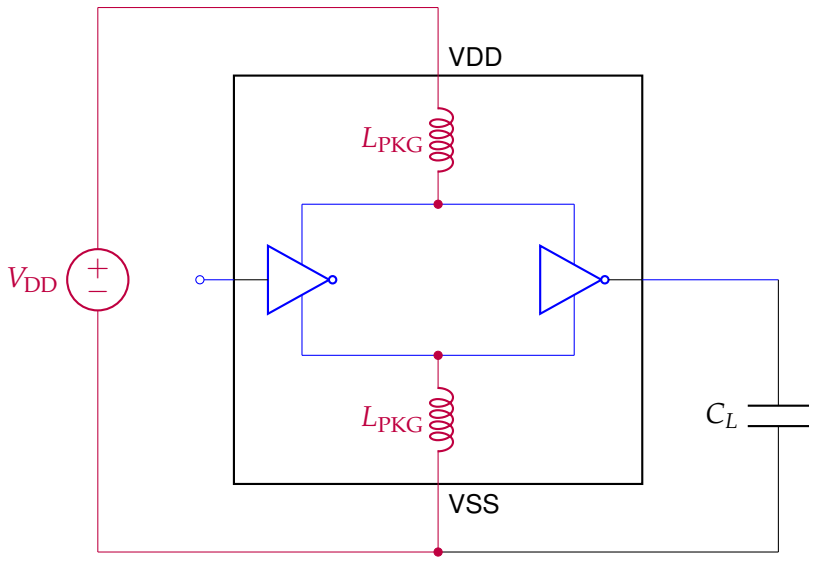


Figure 9.4: Ground/power bounce due to common lead inductance. Charging/discharging currents for C_L will modulate the power supply for the input buffer.

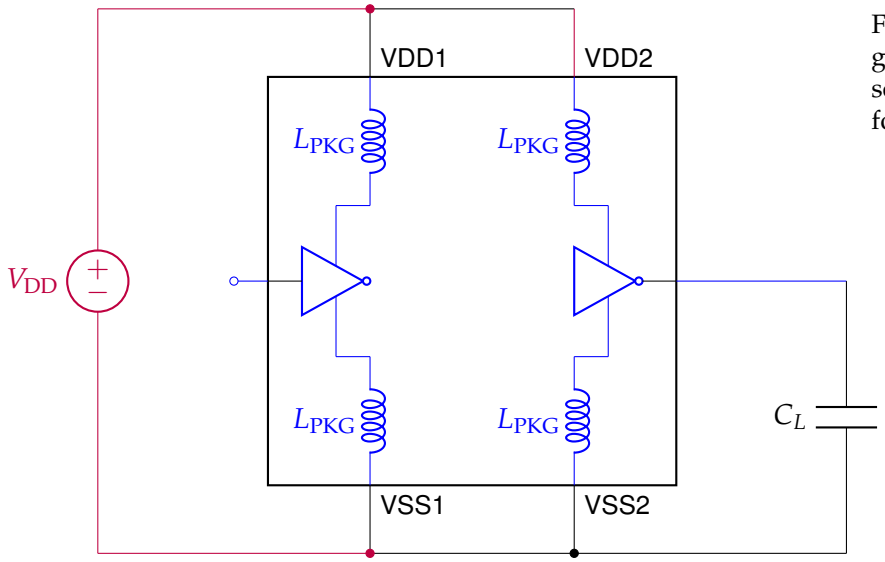


Figure 9.5: Reducing ground/power bounce with separate power/ground pins for input and output buffers.

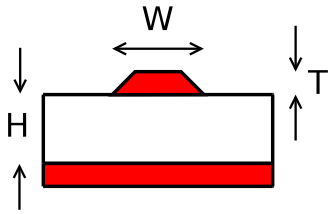


Figure 9.6: Trapezoidal model of microstrip trace cross-section. The inductance is reduced by increasing the width W, increasing the thickness T, or reducing the separation H.

100 Impedance of a PCB microstrip

At low frequencies a PCB trace is resistive. Although the resistance increases with frequency[10], at high frequencies the impedance is dominated by inductive reactance ($X_L = 2\pi f L$) (see Figure 9.7 for a 4-layer PCB).

Inductance is a property of a magnetic circuit and is dominated by the loop area of the signal and return path circuit. To minimise the loop area the signal and return traces must be routed close together. The minimum loop area occurs when we form a *microstrip*, consisting of a PCB trace above a ground plane (see Figure 9.6).

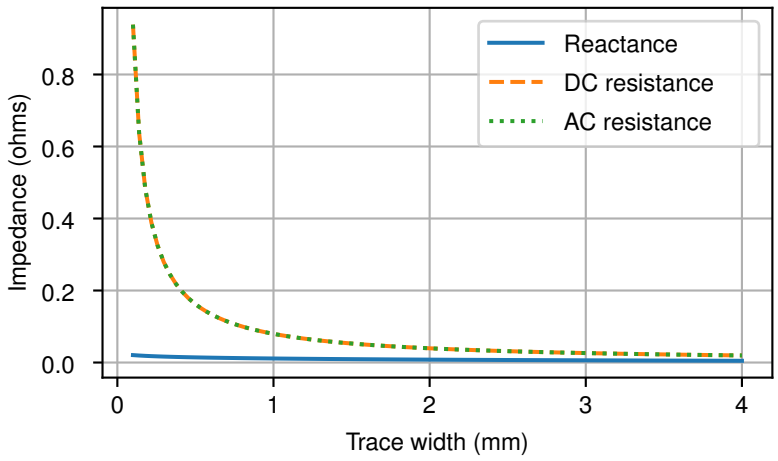
The inductance per unit length of a microstrip is

$$L' = 0.2 \times 10^{-6} \ln\left(\frac{5.98H}{0.8\,W + T}\right) \quad \text{(H/m)}. \qquad (9.4)$$

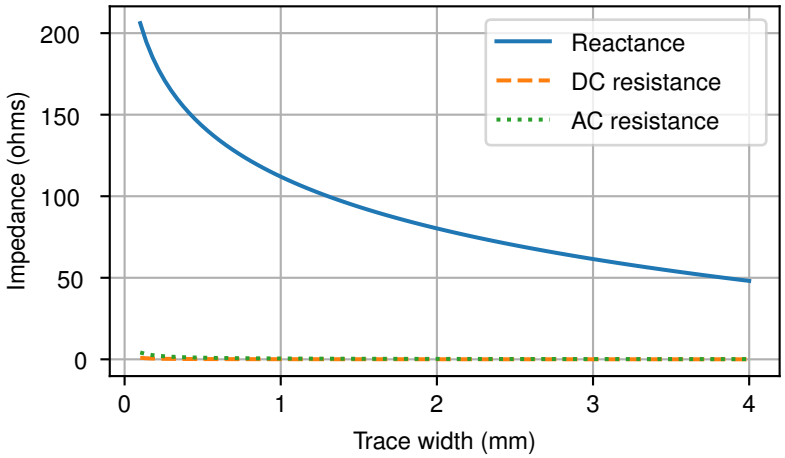
[10] Due to the skin effect where the current is concentrated by magnetic fields to be on the outer surface of a conductor, thus reducing the effective cross-sectional area of the conductor and increasing the resistance.

101 Summary

1. Common path noise results when switching currents induce power supply voltage variations due to shared power supply connections. This analogous to your shower becoming too hot when someone turns on a cold water tap in the house.
2. Power supply noise results from switching currents required to charge/discharge capacitive loads.



(a)



(b)

Figure 9.7: Microstrip trace DC and AC resistance (including skin effect) compared to inductive reactance for a 2-layer PCB ($T = 33\,\mu$m, $H = 1.5$ mm, length 15 cm): (a) $f = 25$ kHz, (b) $f = 250$ MHz.

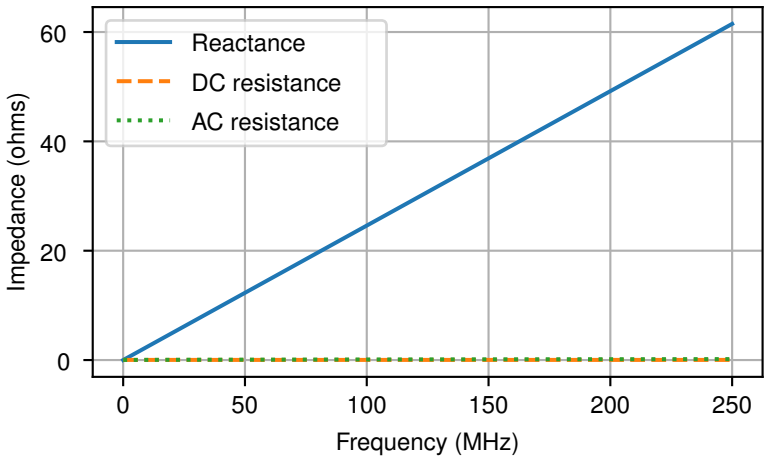


Figure 9.8: Microstrip trace impedance for a 2-layer PCB ($W = 3$ mm, length 15 cm).

110 Exercises

1. Why is it advisable to twist long power leads together when powering a circuit from a power supply?
2. What is a star-point?
3. Why does the skin-effect increase the resistance of a PCB trace?
4. Why should power supply traces on a high-speed digital PCB be as fat as possible?
5. Why should power supply traces be run over a ground plane?
6. What is the advantage of using power and ground planes?
7. Why use separate voltage regulators for digital and analogue components (even though the voltage may be the same)?
8. Discuss five techniques for reducing common path noise.
9. On a single-sided PCB would it be advisable to run the power supply traces close together or along opposite sides of the PCB?
10. Why cannot ground or power bounce be measured with an oscilloscope?
11. Describe five techniques to reduce ground or power bounce.
12. Describe how ground bounce can cause oscillations for a CMOS inverter. *Hint, consider the CMOS transfer curve.*

10
Power supply decoupling

Power supply (power bus) decoupling is important since it reduces both radiated and conducted electromagnetic interference (EMI) and improves signal integrity. Bypass (decoupling) capacitors provide a low impedance path between power and ground.

There are three levels of power supply decoupling:

1. Bulk bypass capacitors for low frequency decoupling. These are placed at the star point (usually close to the power supply connector).
2. Local bypass capacitors, placed to minimise loop area between IC VDD and VSS pins.
3. Interplane capacitance (for very high frequency decoupling).

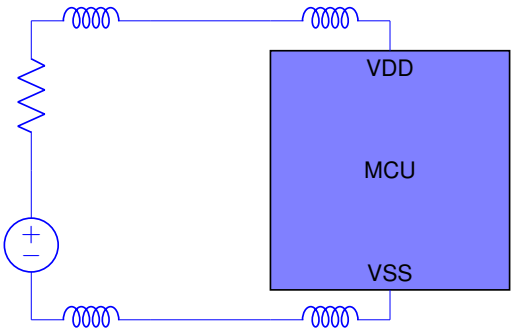


Figure 10.1: MCU power supply showing common path in blue.

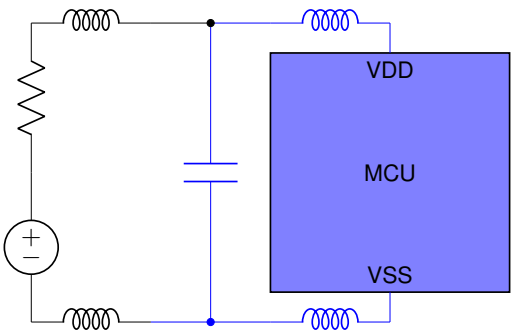


Figure 10.2: MCU power supply showing shorter common path in blue when a local decoupling capacitor is employed.

1 Local decoupling

Each power supply pin of an IC should have a local decoupling capacitor[1]. The capacitors need to be placed as close as possible to the IC's power and ground pins[2].

1.1 Capacitor placement

When placing local decoupling capacitors on a PCB there two main considerations: signal integrity and thermal.

The goal for good signal integrity is to reduce the impedance of the power supply to the chip by:

1. Placing the capacitor as close as possible to the power supply pins of the chip to reduce the loop area and thus the inductance.
2. Make the traces as wide as possible from the capacitor to the chip (again to lower the inductance).
3. Using multiple vias from the power and ground planes to the capacitor pads to reduce inductance to the planes.

[1] While they are not always effective due to stray inductance it is good design practice to play safe and make provision for them.

[2] The best placement will minimise the area of the loop formed by the capacitor and the traces to the IC pads.

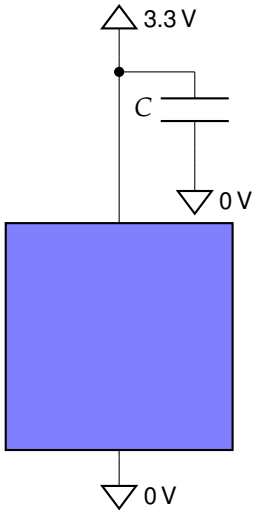


Figure 10.3: Schematic representation of IC and decoupling capacitor. The capacitor is placed to minimise the loop between it and the IC's power pins.

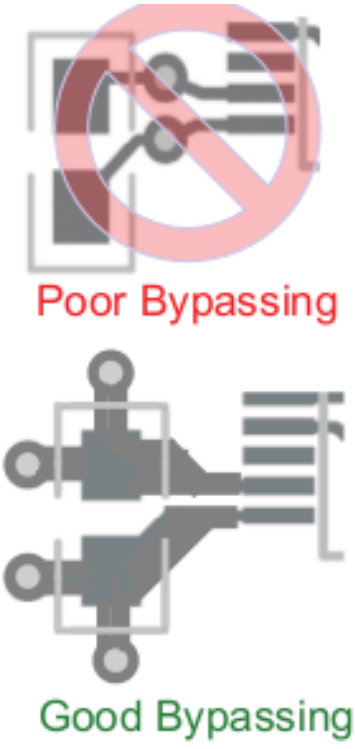


Figure 10.4: Bad and good decoupling capacitor placement. In the bad case, the capacitor is too far from the pins creating a large loop. In addition, the traces are too thin. The good case reduces the inductance by placing the capacitor closer to the chip, using wide traces, and multiple vias. Note the neck-down to the IC pads.

Thermal considerations are important for manufacturing. Ideally, the pads of a component should heat up at the same rate in the oven. For example, a component will tombstone (see Figure 10.5) if the solder on the pads melt at different times. The key considerations are:

1. Use thermal vias to reduce heat flow to the power and ground planes.
2. Keep the pad connections symmetrical.
3. Neck-down traces where they join to a chip pad[3].

Figure 10.5: Example of tombstoning.

[3] This reduces heat flow away from the chip pin.

1.10 *Capacitor types*

MLCC[4] capacitors are used due to their low ESR and low ESL. Typically 100 nF capacitors are used for frequencies below 20 MHz and 10 nF capacitors are used for higher frequencies. Surface mount packages are best since they have a lower ESL but are only effective up to 50 MHz. Above this they behave like inductors due to packaging (see Figure 10.7).

MLCC capacitors ae designed for power supply bypassing and should not be used for analogue circuits since their capacitance varies with voltage and temperature. The best MLCC capacitors use a X7R dielectric but these are more expensive than X5R and Y5V.

[4] Multilayer ceramic chip.

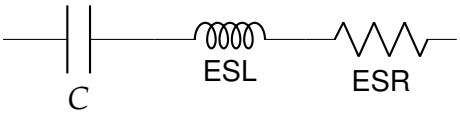


Figure 10.6: Real capacitor simplified electrical model.

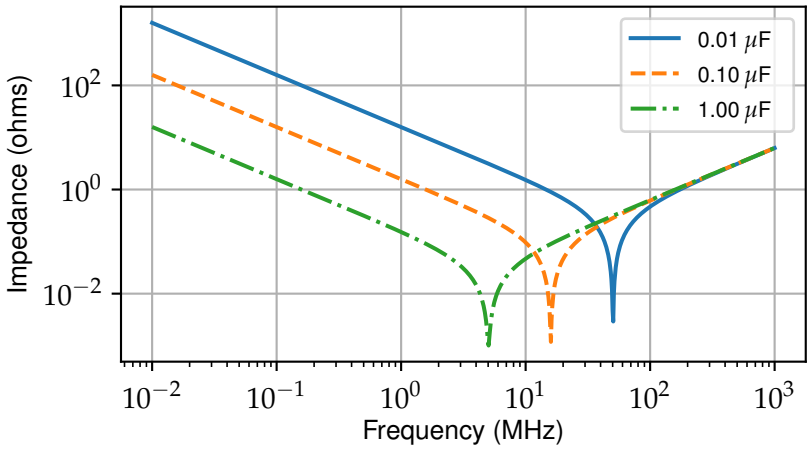


Figure 10.7: Surface mount capacitor impedance showing how capacitors look inductive at high frequencies. $L = 1$ nH, $R = 0.1$ ohm.

10 Multiple bypass capacitors

Mixed-signal devices (ADC, DAC, PLL, etc.) require well decoupled power-supplies. It is not uncommon to see 1 μF, 100 nF, and 10 nF local decoupling capacitors in parallel to improve the decoupling over a wide frequency range, see Figure 10.8. The smallest value capacitors should be placed closest to the power supply pins[5]. Usually, the manufacturer of these devices has a recommended layout.

[5] Since these respond the fastest.

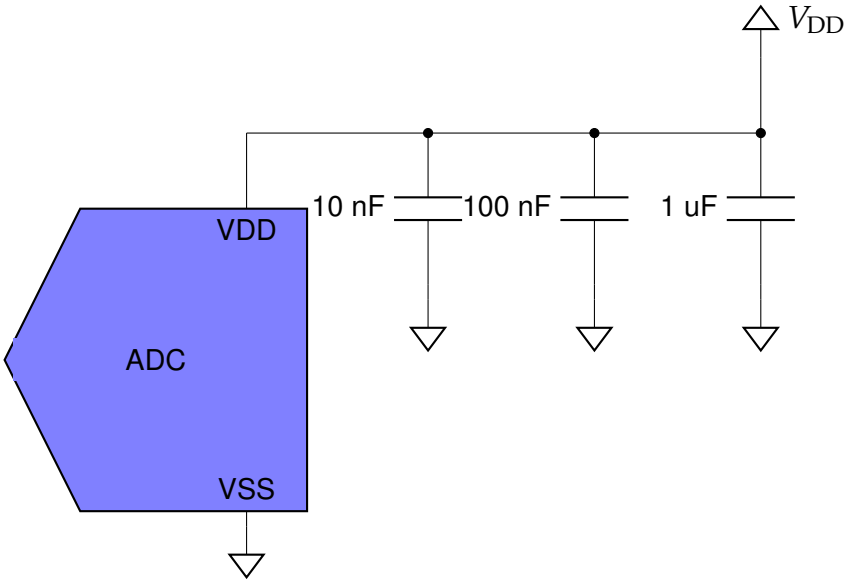


Figure 10.8: Multiple bypass capacitors for an ADC.

10.1 Anti-resonances

Unfortunately, the approach of multiple bypass capacitors in parallel does not always work since the ESL of one capacitor can form an anti-resonance with the capacitance of another capacitor. So sometimes a resistor is added in series with a capacitor to damp the resonance. Thus, paradoxically, a decoupling capacitor with a higher ESR can be better.

11 Further reading

"High Speed Analog Design and Application Seminar", Texas Instruments, www.ti.com/lit/ml/slyp173/slyp173.pdf. See p.22 for good and bad examples of decoupling capacitor placement.

100 Exercises

1. What is the point of local decoupling?
2. What sort of local decoupling capacitors should be used and why?
3. Where should local decoupling capacitors be placed?
4. Why are 100 nF decoupling capacitors not as effective as 10 nF capacitors at frequencies above 20 MHz?
5. Consider an 0603 surface mount capacitor. Would its inductance be smaller if its metallic contacts ran along the short edges or the long edges?
6. What is ESR?
7. Why is common to see multiple decoupling capacitors in parallel for an ADC?

11
Power supply decoupling II

1 *Interplane capacitance*

For high switching frequencies local decoupling capacitors have no effect. Instead, high frequency bypass is achieved by using PCB ground and power planes spaced close together. Thus the planes act like a distributed capacitor with little inductance that provides effective decoupling from 500 MHz to 5 GHz.

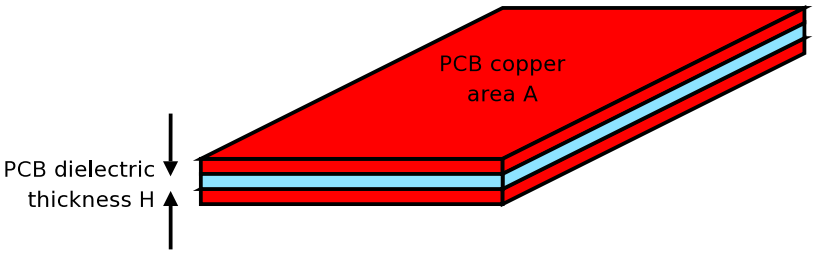


Figure 11.1: Capacitor formed by closely separated parallel planes of copper.

Two parallel plates of area A, separated a distance H, has a capacitance

$$C = \epsilon_r \epsilon_0 \frac{A}{H}. \tag{11.1}$$

where $\epsilon_r \epsilon_0$ is the permittivity of the dielectric. For example, a 2-sided FR4 PCB has a typical thickness of $H = 1.6\,\text{mm}$ and a relative permittivity $\epsilon_r = 4.5$. If the planes are 10 cm by 10 cm, then

$$C = 4.5 \times 8.8542 \times 10^{-12} \times \frac{0.01}{1.6 \times 10^{-3}}, \tag{11.2}$$

$$= 250\,\text{pF}. \tag{11.3}$$

With a multilayer board, the ground and power planes can be positioned as close 0.25 mm (0.01 in), increasing the capacitance for a 0.01 m^2 area to 1.6 nF.

10 Bulk decoupling

For low frequency decoupling a large bypass capacitor should be installed between power and ground at the point where wiring inductance becomes a problem. The bypass capacitor should be at the 'star point' (typically at the power supply connector, see Figure 11.2). Its capacitance depends on the amount of charge that it needs to supply to minimise voltage droop[1].

There are three types of SMT capacitor for bulk decoupling:

MLCC These are small with a low ESR[2] and ESL. Maximum capacitance[3] is $\approx 220\,\mu$F in 1206 package.

Tantalum electrolytic These have a lower ESR than aluminium electrolytic capacitors but are expensive. They are polarised. Maximum capacitance 2,000 μF.

Aluminium electrolytic These have a poor ESR but are cheap. They are polarised. Maximum capacitance 10,000 μF (surface mount).

Additional bulk decoupling capacitors may be required near chips with high current requirements (FPGAs, MCUs, etc.).

[1] Warning, bulk capacitors can cause a significant in-rush current when powered.

[2] Equivalent series resistance.

[3] This is increasing every year.

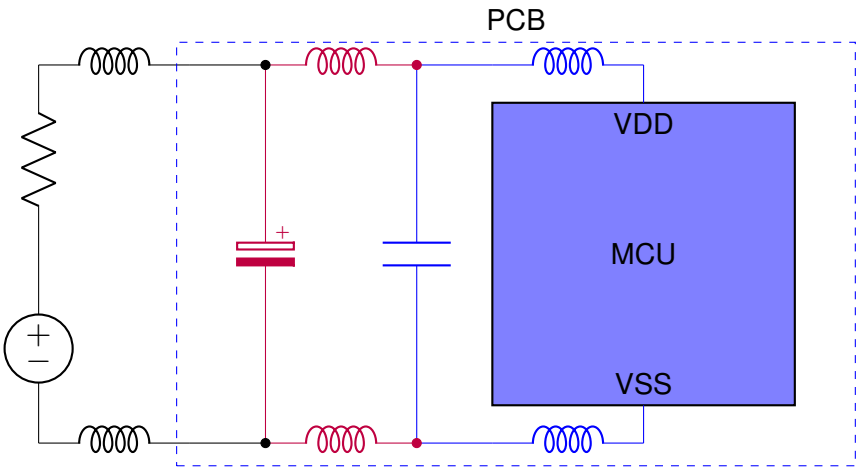


Figure 11.2: Reduction of common path noise using a bulk bypass capacitor in conjunction with a local bypass capacitor.

Figure 11.3: 22 μF MLCC capacitor in 0805 SMT package (0.08 in $\times$ 0.05 in).

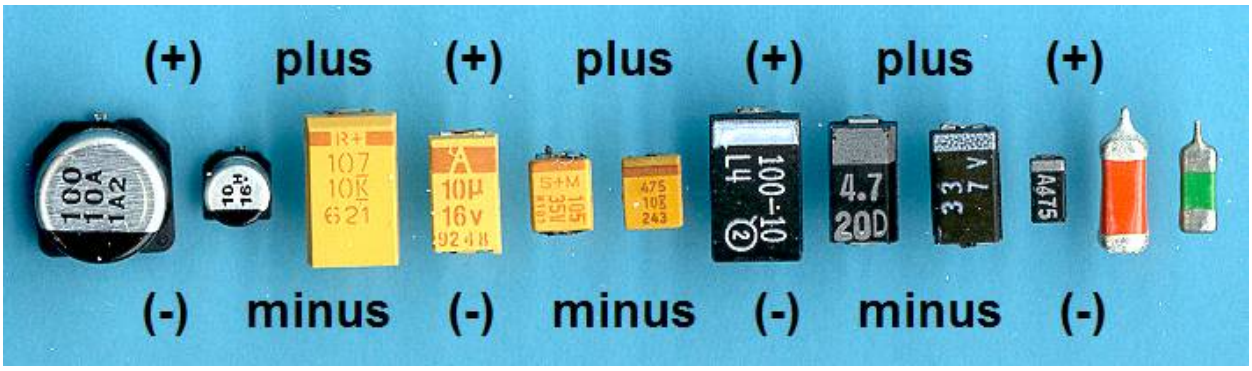


Figure 11.4: SMT electrolytic capacitors for bulk decoupling. The two on the left are aluminium electrolytic capacitors; the others are tantalum electrolytic capacitors.

10.1 *In-rush current*

The in-rush current depends on the supply voltage, supply resistance, supply lead inductance, and the ESR of the bulk decoupling capacitance.

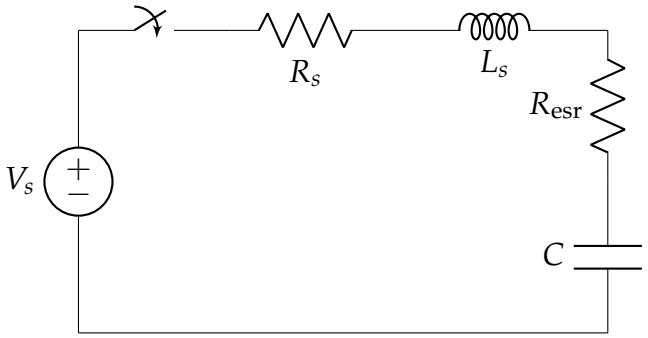


Figure 11.5: Equivalent circuit for calculating in-rush current.

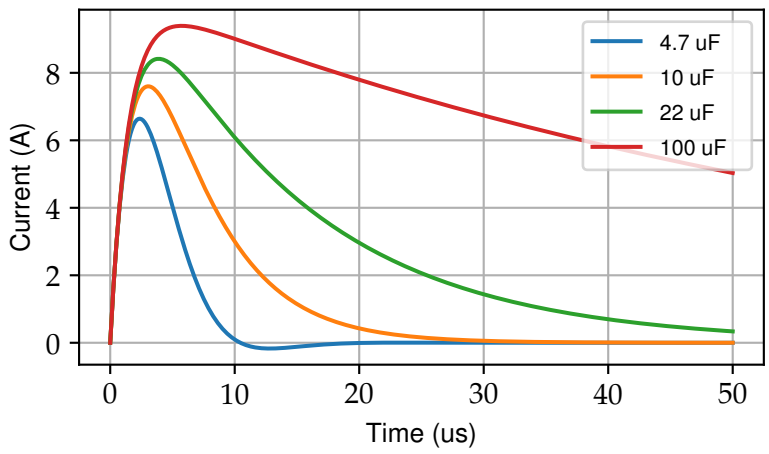


Figure 11.6: In-rush current. $V_s = 7\,\text{V}$, $R_s + R_{esr} = 0.7\,\Omega$, $L_s = 1\,\mu\text{H}$.

11 Analogue power supply decoupling

Analogue[4] components often require a clean power supply. Thus it is wise to have separate voltage regulators for analogue and digital components. Unfortunately, voltage regulators have a limited frequency response and additional low-pass filtering of the power supply is required.

One approach is an R-C low-pass filter as shown in Figure 11.7. A better approach is to use a ferrite bead as shown in Figure 11.8.

[4] And mixed signal components, such as ADC, DAC, PLL, radios, IR receivers, etc.

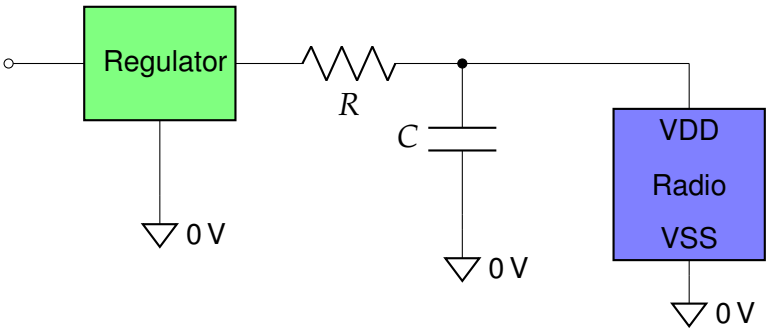


Figure 11.7: RC low-pass power-supply filter.

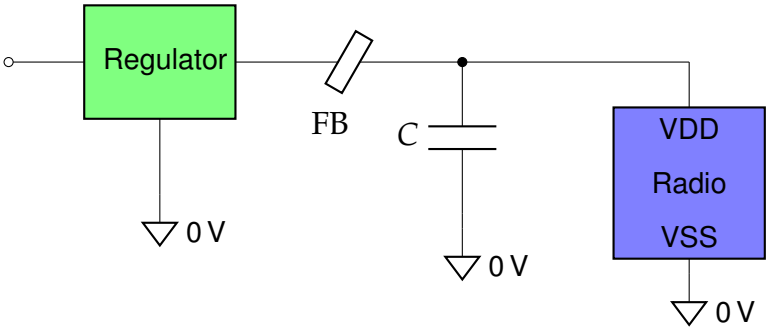


Figure 11.8: Low-pass power-supply filter using a ferrite bead.

11.1 Ferrite beads

A ferrite bead is an inductor with a lossy ferrite core. It has a low DC resistance (useful to minimise power losses), acts like an inductor (up to say 10 MHz), and then behaves like a higher value resistor (say 200 Ω) at higher frequencies. At high frequencies (> 500 MHz) the impedance drops due to parasitic capacitance.

Figure 11.9: SMT ferrite bead.

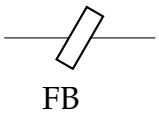


Figure 11.10: IEEE symbol for a ferrite bead. Often a ferrite bead is drawn as an inductor but it is not an inductor.

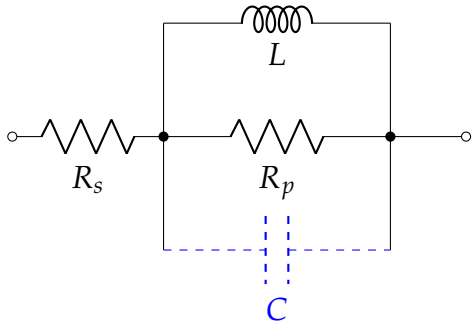


Figure 11.11: Electrical model of a ferrite bead. Low frequency energy passes through L; high frequency energy is absorbed by R_p. C represents the inter-turn parasitic capacitance.

There are two caveats using ferrite beads:

- The ferrite needs to be specified to handle the peak current; if the core of an inductor saturates then it no longer acts like an inductor[5].
- All inductors have inter-turn capacitance; this can create a resonance. Sometimes a parallel resistor is required for additional damping of the resonance.

[5] As the bias current is increased, its impedance drops so it does not reject high frequency noise as well. A design rule is to operate at 20% of the rated DC current.

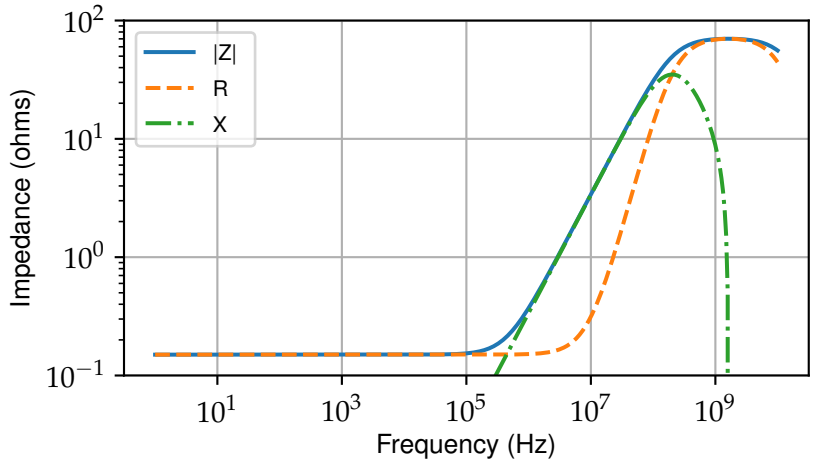


Figure 11.12: Ferrite bead impedance with frequency. $R_s = 0.15\,\Omega$, $C_p = 0.18\,\text{pF}$, $R_p = 70\,\Omega$, $L_p = 54\,\text{nH}$.

100 Power supply decoupling analysis

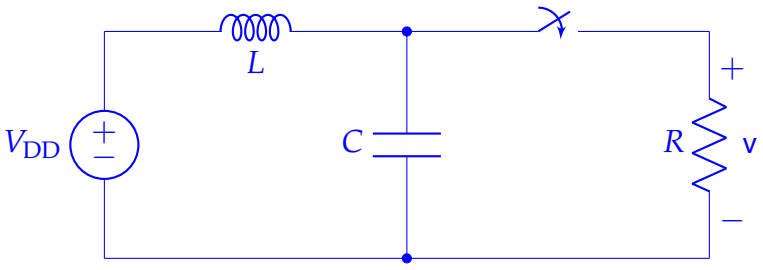


Figure 11.13: Power supply decoupling model.

The power supply decoupling capacitors attempt to reduce the droop in the power supply voltage when devices consume power. For example, consider Figure 11.13 if the capacitor is not present. When the switch closes, the inductance will resist the change in current and so the initial voltage across the resistor is zero. The voltage will then rise to V_{DD} with a time-constant L/R.

When the capacitor is present it will be charged to the power supply voltage, V_{DD}, and so when the switch closes the voltage across the resistor will be V_{DD}. As charge is removed from the capacitor the voltage will droop while the inductor gets its act together. The bigger the capacitor the smaller the droop[6].

[6] Assuming an ideal capacitor with no inductance.

To analyse the response let's consider a decoupling capacitor C for a chip of power consumption P powered with a supply voltage V_{DD}. Let's assume that the chip is drawing negligible current up until time $t = 0$ when it wants to draw a current

$$I = \frac{P}{V_{DD}} = \frac{V_{DD}}{R}, \tag{11.4}$$

where $R = V_{DD}^2/P$ is the internal resistance of the chip. Now for a first-order approximation the reactance of L dominates the reactance of C so we can simplify the model of Figure 11.13 to Figure 11.14. The voltage v across the chip is thus

$$v(t) = V_{DD} \exp\left(-\frac{t}{RC}\right) u(t). \tag{11.5}$$

Let's now consider the time Δt it takes for the supply voltage to sag by 5%. From (11.5),

$$\exp\left(-\frac{\Delta t}{RC}\right) = 1 - 0.05, \tag{11.6}$$

and then taking logs,

$$\Delta t = -RC \log(1 - 0.05), \tag{11.7}$$

$$\approx 0.05RC, \tag{11.8}$$

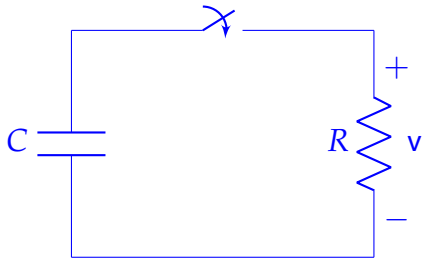


Figure 11.14: Power supply decoupling approximate model.

since $\log(1+x) \approx x$ for $x << 1$. Substituting for the internal resistance R yields the result:

$$\Delta t = 0.05C\frac{V_{DD}^2}{P}. \tag{11.9}$$

For example, let's consider a typical local decoupling capacitor of $C = 100\,\text{nF}$, $V_{DD} = 3.3\,\text{V}$, and $P = 0.2\,\text{W}$. The equivalent internal resistance is

$$R = \frac{V_{DD}^2}{P} = \frac{3.3^2}{0.2} = 54\,\text{ohms}, \tag{11.10}$$

giving a time-constant

$$\tau = RC = 54 \times 100 \times 10^{-9} = 5.4\,\mu\text{s}. \tag{11.11}$$

The time it takes for the power supply to droop 5% is thus

$$\Delta t = 0.05\tau = 0.05 \times 5.4 \times 10^{-6} = 272\,\text{ns}. \tag{11.12}$$

This simplified analysis ignores the inductance of the decoupling capacitor.

In comparison, let's assume that the power supply has an inductance of 80 nH. Without the decoupling capacitor, the time constant is

$$\tau = \frac{L}{R} = \frac{80 \times 10^{-9}}{54} = 1.5\,\text{ns}. \tag{11.13}$$

101 *Capacitor datasheets*

Manufacturers do not usually state the ESR; instead they state $\tan\delta$ (tan-delta dissipation factor), where

$$\tan\delta = \frac{\text{ESR}}{|X_c|}, \tag{11.14}$$

where X_c is the reactance[7] of the capacitor,

$$X_c = \frac{1}{2\pi f C}. \tag{11.15}$$

[7] Typically calculated at 1 MHz for $C < 1\,\text{nF}$ and 1 kHz for $C < 10\,\mu\text{F}$.

The tan-delta is usually specified as a percentage. This model predicts that the ESR increases with frequency,

$$\text{ESR} = \frac{2\pi f C}{\tan \delta}. \tag{11.16}$$

In practice, capacitors require a model more sophisticated than a series R-L-C model.

The ESL is often not stated but this is a function of the package size and contact geometry.

110 Further reading

AN 583: Designing Power Isolation Filters with Ferrite Beads for Altera FPGAs, Altera, 2009. https://cdrdv2-public.intel.com/654404/an583.pdf This has an excellent summary of using ferrite beads.

Ferrite Beads Demystified. Analog Devices, 2016. https://www.analog.com/en/analog-dialogue/articles/ferrite-beads-demystified.html This has an excellent section on ferrite beads as low-pass filters for power supply filtering.

111 Exercises

1. When is interplane capacitance required?
2. Why is interplane capacitance required?
3. What is the point of bulk decoupling?
4. Where should bulk decoupling capacitors be placed?
5. What sort of bulk decoupling capacitors should be used and why?
6. The USB specification says that the maximum capacitance between VBUS (+5 V) and ground is 10 μF. Postulate why.
7. Why did the tantalum electrolytic capacitor go bang when power was applied?
8. What considerations are required for analogue power supplies?
9. What is a common purpose of ferrite beads?

12
Crosstalk

Crosstalk is unwanted signals induced in *victim traces* from an *aggressor trace* due to mutual capacitance and mutual inductance. Examples of victim traces that are considered under high threat include:

- clock signals
- differential pairs
- video, audio signals
- reset lines

Crosstalk limits the distance of high trace density parallel buses and is the reason why differential serial buses are becoming more prevalent for high-speed communication. Crosstalk is reduced by keeping loops small and far apart. The larger the loop the larger the mutual inductance and thus the magnetic field is better coupled[1]. The closer the separation the larger the mutual capacitance and thus the electric field is better coupled.

[1] You have a more efficient loop antenna.

Figure 12.1: Schematic of aggressor and victim traces. The victim trace could travel in the same direction as the aggressor trace.

1 Crosstalk coupling

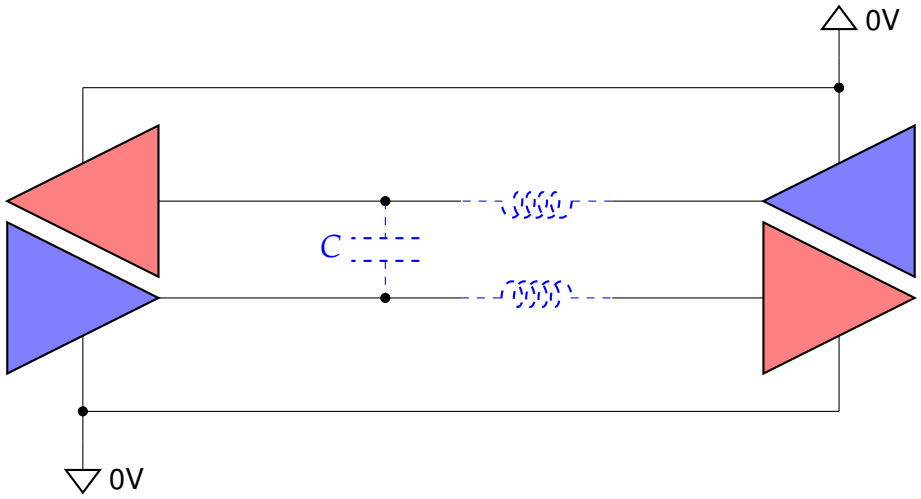


Figure 12.2: Simplified crosstalk model showing parasitic capacitance mutual inductance between traces.

There are two crosstalk mechanisms, see Figure 12.2:

- A changing aggressor voltage induces a current through capacitive coupling[2]. The induced current is

$$i_V(t) = C\frac{\mathrm{d}}{\mathrm{d}t}v_A(t), \tag{12.1}$$

where C is the parasitic capacitance between the traces.

- A changing aggressor current induces a voltage through inductive coupling[3]. The induced voltage is

$$v_V(t) = M\frac{\mathrm{d}}{\mathrm{d}t}i_A(t), \tag{12.2}$$

where M is the parasitic mutual inductance between the trace loops.

[2] Electric field coupling.

[3] Magnetic field coupling.

10 Crosstalk between microstrips

Crosstalk on a PCB is reduced by running traces over a ground (or power) plane forming *microstrips*[4].

[4] An even better strategy is to run traces between two planes forming *striplines*.

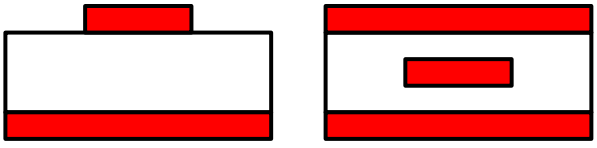


Figure 12.3: Cross-section of (a) microstrip and (b) stripline PCB traces.

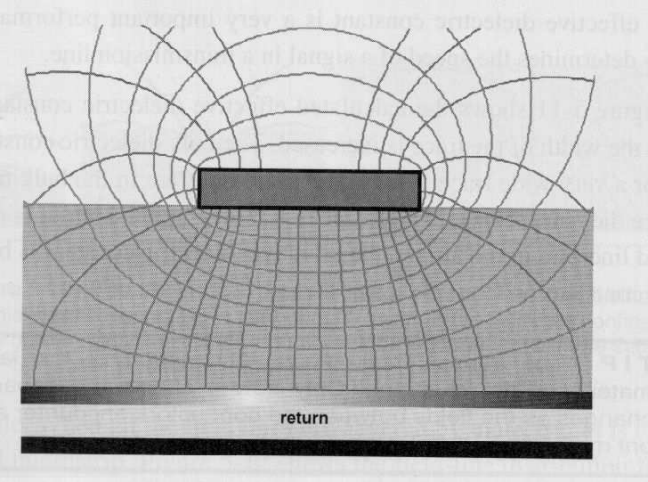


Figure 12.4: Cross-section of a microstrip trace showing electric and magnetic field lines. The fringing fields cause trouble.

A microstrip minimises the loop formed by a signal trace and its return path[5]. However, crosstalk also occurs between microstrips. To see how, consider the electric and magnetic fields for a microstrip in Figure 12.4. If a victim microstrip is placed parallel to the aggressor microstrip, both the changing electric and magnetic fields due to the aggressor will induce a crosstalk signal in the victim[6].

[5] Most of the return current will flow under the trace since this where the inductance is minimised.

[6] To complicate matters, if the victim microstrip is not terminated then crosstalk will bounce back and forth along it. See transmission line termination, Chapter 16.

11 *Crosstalk level*

The crosstalk level needs to be smaller than the noise margin[7]. The level of crosstalk can be approximated by

$$\text{Crosstalk} \propto \frac{BL}{1+\left(\frac{D}{H}\right)^2}, \tag{12.3}$$

where B is the signal bandwidth[8], L is the common length of the aggressor and victim microstrips, D is the separation between microstrips, and H the height of the traces above a common reference plane (see Figure 12.6).

[7] In practice, a total noise budget is required that also includes power supply noise.

[8] Inversely proportional to the signal rise/fall times.

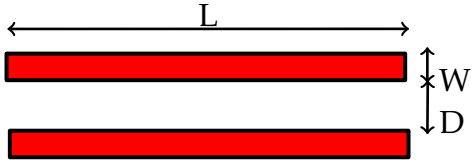


Figure 12.5: Parallel microstrips.

In practice, the level of crosstalk is difficult to estimate; the best approach is to use a field solver program[9].

[9] For example, HyperLynx from Mentor Graphics, Sigrity, and Spectraquest.

100 Crosstalk reduction

The most practical ways to control crosstalk on a PCB are:

1. Run traces over planes[10] to form microstrips[11].
2. Do not run traces over gaps in planes[12].
3. Keep traces close to their reference planes (since crosstalk proportional to the square of this distance).
4. Increase separation between traces. A rule of thumb is to apply the 3-W rule. This states that the separation between trace centres D should be at least 3 times the width of a trace W.
5. Avoid running high-threat traces alongside aggressive traces (run at right angles instead).
6. Avoid changing layers with fast signal traces.
7. Slow down the signal rise/fall times (use series resistors and slower logic devices).

[10] Either ground or power plane.
[11] This minimises forward crosstalk since Eddy currents induced in the plane under the signal trace create a field that decays rapidly with distance.
[12] This forces the return current to find a more distant path, increasing the inductance and thus crosstalk. See https://incompliancemag.com/article/slots-in-gnd-planes/ for a demo. Also see https://www.signalintegrityjournal.com/articles/692-split-planes-and-what-happens-wher

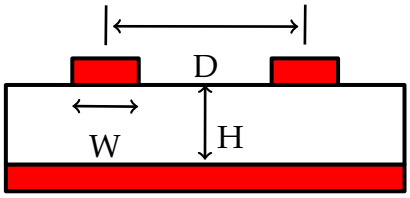


Figure 12.6: Separation of microstrip traces. With the 3-W rule $D > 3W$. Note the trace should be at least a distance W from the edge of the reference plane.

101 Crosstalk directivity

Crosstalk is highly directional. For a microstrip, the inductively and capacitively coupled crosstalk currents tend to cancel in the forward direction but add in the backward direction[13].

- A changing electric field couples a current into the victim microstrip proportional to the frequency[14] and mutual capacitance. This current flows in both directions along the victim microstrip from the point of coupling.
- A changing magnetic field will induce a voltage in the victim microstrip loop proportional to the frequency and mutual inductance (the mutual inductance acts like a poor transformer). The resulting current in the victim microstrip flows in the opposite direction to the current in the aggressor microstrip.

[13] The backward crosstalk grows fairly rapidly to a constant amplitude pulse with a duration twice that of the propagation time of the aggressor signal through the coupled region. The forward crosstalk pulse has an amplitude proportional to the length of the coupled region and a duration equal to the transition time of the aggressor pulse.
[14] There is no crosstalk at DC.

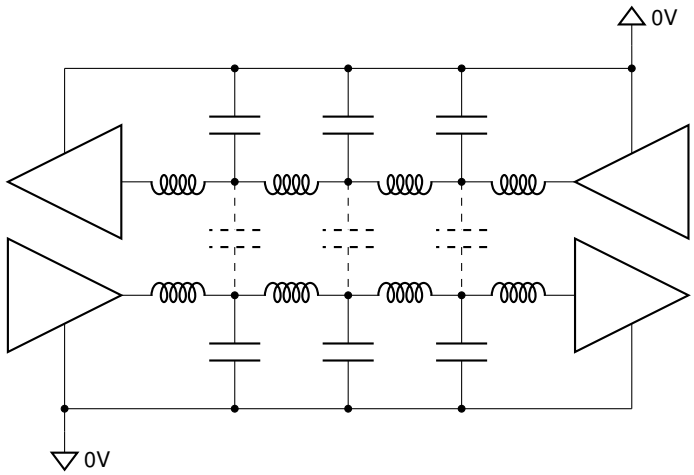


Figure 12.7: Model of two lossless transmission lines showing parasitic mutual capacitances (dashed) that couple the electric field from the aggressor to the victim trace. The magnetic field is coupled due to mutual inductance (not shown).

110 *Microstrip return current*

When laying out a microstrip trace it is important to consider the return current. This current flows in the plane close as possible to the trace, since this minimises the loop area and thus is the path of least impedance. It does not matter if the plane is a ground or power plane; see Figure 12.8 to Figure 12.12.

Fast signals ($>$ 50 kHz) should not be routed over a plane with a split, such as a partitioned power plane. The split forces the return current to take a tortuous path through the nearest local decoupling capacitors. This increases crosstalk and degrades EMC.

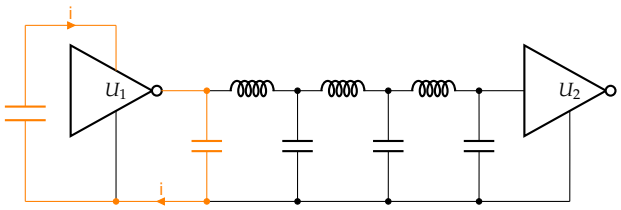


Figure 12.8: Current path for microstrip formed with a ground plane when U1 switches high. The dotted capacitors are parasitic.

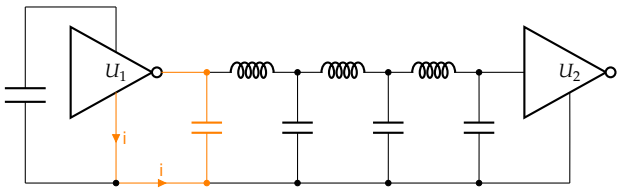


Figure 12.9: Current path for microstrip formed with a ground plane when U1 switches low.

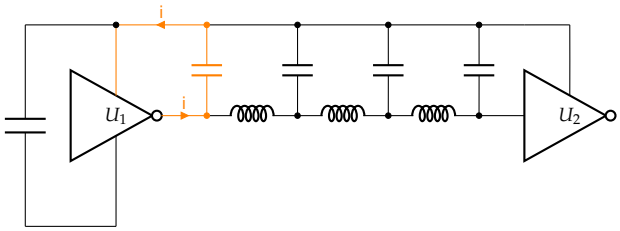


Figure 12.10: Current path for microstrip formed with a power plane when U1 switches high.

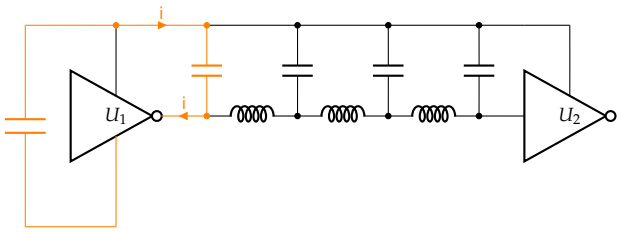


Figure 12.11: Current path for microstrip formed with a power plane when U1 switches low.

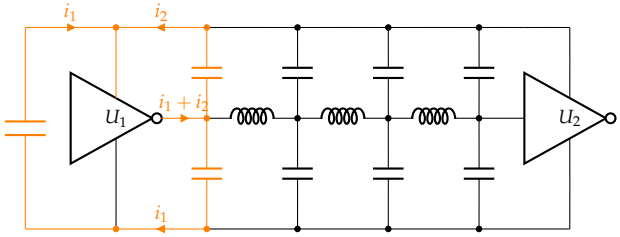


Figure 12.12: Current path for stripline when U1 switches high.

111 *Biomedical instrumentation case study*

Digital ICs produce fast transitions, 5 ns or faster. These fast transitions cause problems even for circuits where the clock speed is low. An example of this concerns some biomedical instrumentation for measuring EEG and EMG signals. The instrumentation consisted of a board with an amplifier and an ADC connected to a MCU board using an SPI bus. However, when a sinewave was input into the ADC board, the waveform recorded by the MCU had significant glitches.

The initial hypothesis for the glitches was that the SPI bus was configured in the wrong mode and so the data was being sampled on the wrong edge[15]. However, this was not the problem. To confuse the issue, another type of MCU recorded the signal without glitches.

The problem was with poor signal integrity[16] and in particular, crosstalk. Although the serial clock rate was only 32 kHz, the two boards were connected with a 10 cm long flexible ribbon cable with the SCLK and MOSI signals beside each other, see Figure 12.13. The fast edge rates of the ADC transitions on the MOSI signal generated crosstalk on the SCLK signal (and vice-versa). This lead to false clocking.

The problem is that there are two coupled signals,

[15] This is a common problem with SPI since there are four modes and no universal standard.

[16] The MCU that worked had a sightly higher logical high voltage than the one that was unreliable.

see Figure 12.13. Since the length of the lines cannot be reduced for the application, it is necessary to slow down the edge rates using series resistors[17] (see Figure 12.14).

The choice of resistor value[18] depends on the amount of driven capacitance since it forms a RC filter with this capacitance.

[17] These are not termination resistors since they have a value much larger than the characteristic impedance of the transmission line.

[18] Some devices cannot tolerate slow edge rates at their inputs. See `https://www.ti.com/lit/an/scba004d/scba004d.pdf`.

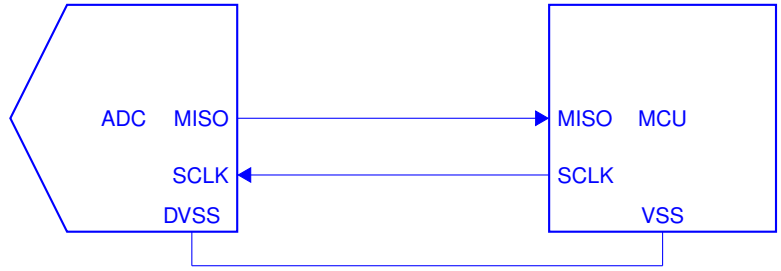


Figure 12.13: Simple schematic showing SPI interface between ADC and MCU with ground connection.

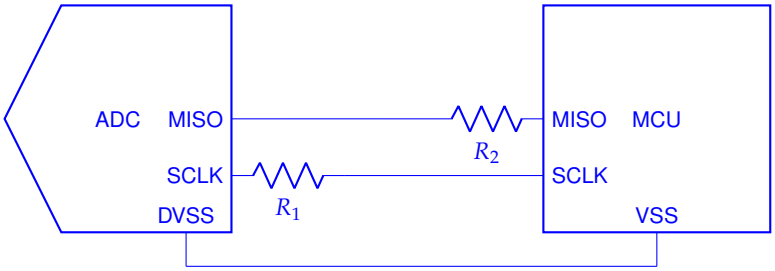


Figure 12.14: SPI interface with series resistors to slow the transitions and thus reduce crosstalk between the signals.

1000 Exercises

1. Why is crosstalk undesirable?
2. How can crosstalk be reduced?
3. Why is forward and backward crosstalk different?
4. What is a field solver program?
5. Why should traces not be run along the edge of a PCB?
6. How small does the crosstalk need to be?
7. Why is crosstalk worse with faster signal transitions?
8. How can you slow down the signal transitions?
9. How does a microstrip help reduce crosstalk?
10. Why does a multilayer PCB help reduce crosstalk?
11. Why should a trace not run over a gap in a plane?
12. If you have a ground plane on layer 2 and a partitioned power plane on layer 3 of a four layer board, on which layer should you route fast signals?
13. A synchronous memory bus such as DDR can have high levels of crosstalk and still operate. How?

13
Printed circuit boards

The PCB is usually the most critical component in a high-speed or mixed signal system. There are many considerations:

Manufacturability minimum trace size[1], minimum clearances, minimum hole/via size, etc.

Thermal even heating of pads to avoid tombstoning, heat dissipation through thermal vias, etc.

Signal integrity crosstalk, EMC[2], power supply decoupling, controlled impedance.

[1] The etchant concentration varies depending on how much copper is being removed.

[2] Electromagnetic compatibility. The goal is to reduce electromagnetic emissions and electromagnetic susceptibility.

1 PCB construction

A multilayer PCB is a stack of alternating layers of *prepreg mats* and *laminate sheets* (cores) (see Figure 13.2 and Figure 13.3):

Prepreg mats are a weave of glass fibre yarns preimpregnated with partially cured resin[3].

Laminate sheets (cores) are fully cured prepreg sheets with copper foil attached to one or both sides (just like a normal single or two-sided PCB).

[3] The FR4 epoxy resin system is the most popular for prepreg and laminate sheets. The FR4 name does not specify the chemistry but the fire-retardant level!

After the cores have been etched, they are joined with prepregs. Copper foils are added to the outside and the sandwich is heated under pressure[4]. The partially cured prepreg flows and bonds to the outside of the copper foils and laminate sheets. Holes are then drilled, the outer layers etched, vias electroplated, and a protective coating is electroplated to reduce corrosion.

[4] The prepregs squeeze much more than laminates since laminates are pre-cured.

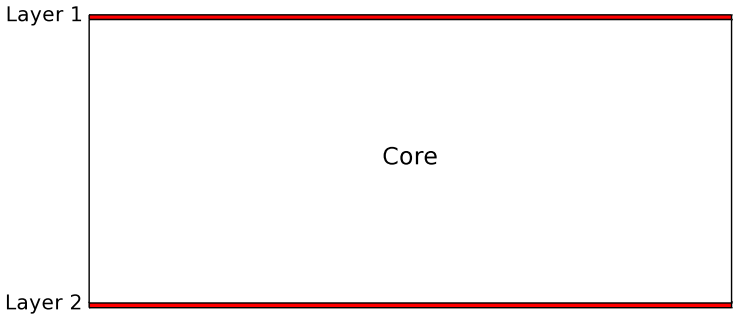


Figure 13.1: Two layer PCB construction.

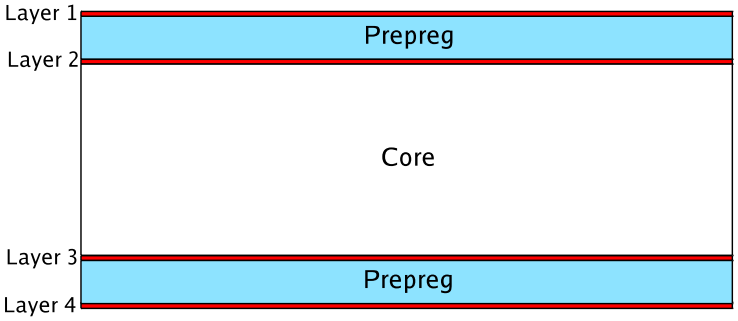


Figure 13.2: Four layer PCB construction. Note the core is thicker than the prepreg layers.

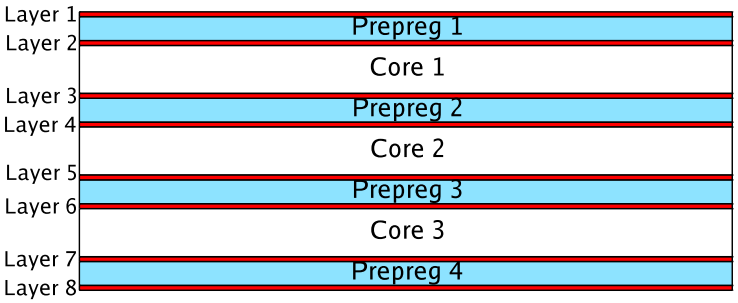


Figure 13.3: Eight layer PCB construction.

Layers	Cores	Prepregs	Core thickness	Prepreg thickness	Total thickness
2	1	0		1.52	1.59
4	1	2	0.99	0.23	1.59
6	2	3	0.36	0.23	1.61
8	3	4	0.24	0.13	1.52
10	4	5	0.20	0.115	1.74

Table 13.1: Stackup of typical multilayer PCBs. Dimensions in mm.

10 PCB stackup rules

There are two primary rules to consider for a PCB stackup:

Rule 1: Every high-speed signal trace should be routed over a continuous plane to reduce the loop area, control impedance, and control crosstalk. The separation should be as small as possible to minimise crosstalk.

Rule 2: Every power supply should form a parallel power-ground plane capacitor for high-speed decoupling. The separation should be as small as possible to maximise the interplane capacitance.

In addition, the following rules are desirable:

Rule 3: Route high-speed signals on buried layers[5] between planes. This reduces crosstalk since the planes act as shields.

[5] Forming strip lines.

Rule 4: Use multiple ground planes to reduce the ground plane impedance and reduce the common-mode radiation.

Rule 5: Use a symmetrical stackup. This is good from a mechanical point of view since it minimises warping in the oven.

Layers	Routing layers	Planes	Trace close to plane	Power planes close
2	1	1	No	No
4	2	2	One or other	
6	2	4	Yes	Yes
6	4	2	Yes	No
8	4	4	Yes	Yes
10	6	4	Yes	Yes

Table 13.2: Stackup comparison for different number of PCB layers.

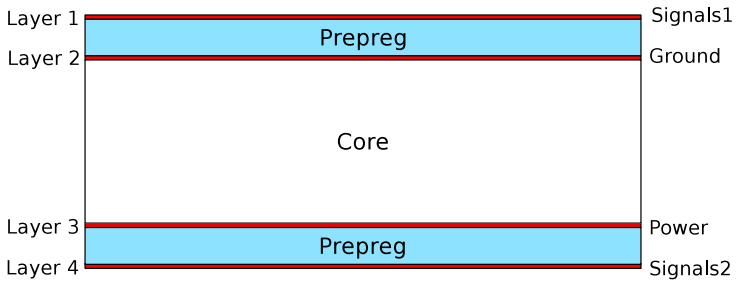


Figure 13.4: 4 layer PCB stackup with 2 signal layers. Note, the power and ground planes are not close.

10.1 PCB stackup examples

A four layer board with two signal layers cannot meet the two primary rules of PCB stackup. While the signal traces can be kept close to a ground (or power plane), the interplane capacitance is low due to the large separation (see Figure 13.4).

To meet the design rules more layers[6] are required. For example, a six layer configuration with good EMC for two signal traces is (see Figure 13.5):

Ground / Signal / Power / Ground / Signal / Ground

It keeps the signals buried between planes but makes it impossible to do rework. In addition, the ground on the top layer is not so effective since it has holes for the components. Another six layer stackup for three signal traces is (Figure 13.6):

Signal / Ground / Signal / Power / Ground / Signal

but here the Power and Ground planes are on the same core and thus are not so closely spaced.

[6] Adding layers costs more...

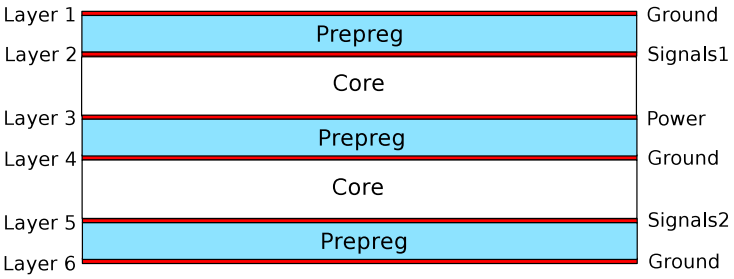


Figure 13.5: 6 layer PCB stackup for two buried signal layers.

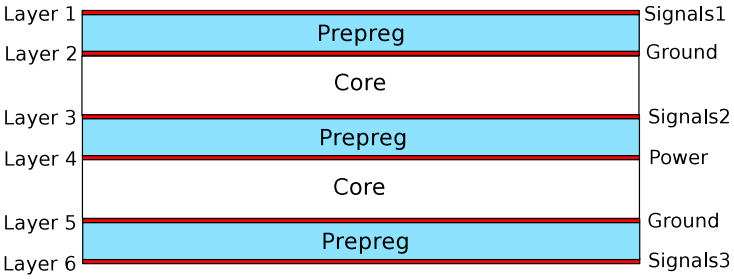


Figure 13.6: 6 layer PCB stackup for three signal layers.

In general, 8 layers is considered the minimum for a high-speed digital design with good EMC[7]. See for example Figure 13.7 where the fast signals are embedded between ground planes for maximum shielding; each signal layer is associated with a ground plane; and the power and ground planes are close.

[7] See http://www.hottconsultants.com/techtips/pcb-stack-up-1.html.

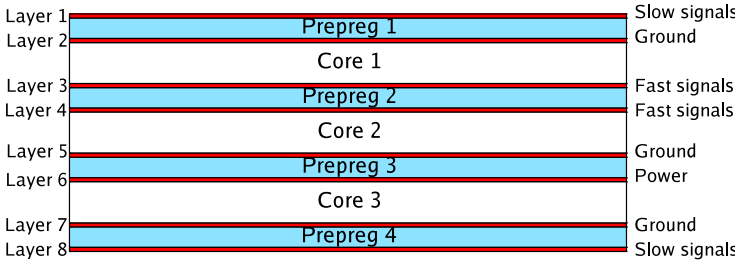


Figure 13.7: 8 layer PCB stackup for two buried high speed signal layers and two slower signal layers.

With large BGA chips additional signal layers are required to escape the signal traces. In addition, additional power planes are required for different voltage levels. A typical 10 layer stackup with multiple power planes is:

Signal (slow) / Power 3.3 V / Ground / Power 2.5 V / Signal (fast) / Signal (fast) / Power 1.5 V / Ground / Power 1.2 V / Signal (slow)

11 *Vias*

Vias are required to connect tracks on different layers. Blind and buried vias are more expensive to manufacture (see Figure 13.8).

Vias should be avoided for high speed signal traces unless there is a neighbouring via[8] for the return current. Without a return via, the return current will take a torturous path increasing the loop area and thus the crosstalk. Obviously you cannot use a return via when the planes are of differential potential (for example, a power plane and a ground plane). In this case the return current flows through the nearest decoupling capacitor, again increasing the loop area.

[8] Preferably multiple ground vias around the signal via.

For very high speed signals, the geometry of vias must be chosen to match the characteristic impedance of the signal microstrip[9] to prevent a reflection.

[9] A single via in a standard (1.57 mm thick) PCB can add 1.2 nH and 0.5 pF.

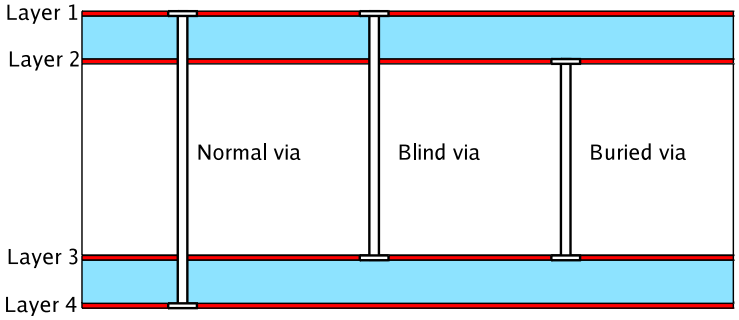


Figure 13.8: Types of vias. Blind and buried are more expensive.

100 PCB example

Good PCB manufacturers publish their PCB specifications and design rules. For example, Advanced Circuits state that their standard 0.062 inch, four layer PCBs are made with 1 oz copper, 2 sheets 2116, 1 oz copper, 0.039 inch core, 1 oz copper, 2 sheets 2116, then finally 1 oz copper. Let's make sense of this. Now 1 oz copper refers to the weight of the copper foil per square foot; it is approximately equivalent to 0.035 mm. The sheets of 2116 are sheets of prepreg, each sheet is 0.115 mm thick making each prepreg layer 0.23 mm thick. The 0.039 inch core is a laminate of cured prepreg coated on each side with copper foil. Since the core thickness 0.039 inch (0.99 mm) does not usually include the thickness of the copper foil, the total PCB thickness is

$$4 \times 0.035 + 2 \times 2 \times 0.115 + 0.99 = 1.59\,\text{mm}(0.061\,\text{inch}). \tag{13.1}$$

In comparison, a two layer FR-4 PCB with 1 oz copper is usually made up of eight sheets of 7628 prepreg. Each sheet is 0.19 mm thick so 8 sheets is 1.52 mm thick[10], giving a total PCB thickness of

$$2 \times 0.035 + 8 \times 0.19 = 1.59\,\text{mm}. \tag{13.2}$$

[10] The red UL/manufacturers logo is in the middle (layer 4).

Other typical multilayer PCB stackups are shown in Table 13.1.

101 Further reading

`https://emcfastpass.com/emc-cost-reduction/` presents an excellent case study of signal integrity and how the author managed to convert a four layer design to two layer design to save costs.

110 Exercises

1. Why are more than PCB four layers required to achieve high signal integrity?
2. What are the purposes of the multiple internal planes of a multilayer PCB?
3. What parameters control the characteristic impedance of a trace?

4. Is it better to run a trace over a ground plane or a power plane?
5. What are the disadvantages of having a ground plane on the top layer?
6. Why do BGA chips with many pins require many internal PCB layers?
7. Why are buried vias expensive?
8. It is a bad idea to jump layers for high speed signals. Why?

14
Transmission lines

In this lecture we find that KCL[1] does not always apply and wires are not simple electrical components.

[1] KVL is violated when a changing magnetic flux passes through a circuit loop.

1 *Introduction*

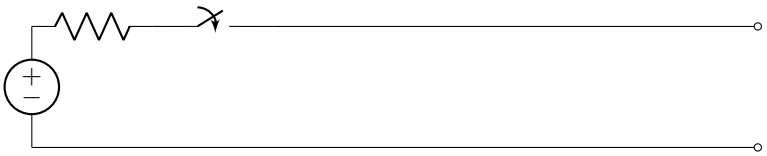


Figure 14.1: What happens when the switch is closed?

Consider what happens when a voltage source is connected to a pair of wires (see Figure 14.1). From simple circuit theory (KCL) you might say that the current is zero. But a pair of wires has capacitance and thus current is required to charge this capacitor.

What if we short circuit the wires? What effect will this have? Surprisingly, the initial current is the same as if the wires were open circuit! From a causality point of view this makes sense. Energy cannot flow faster than the speed of light. As a consequence:

1. The electric field propagates in the dielectric between the wires with a finite speed.
2. If a changing potential difference is applied to the wires, the potential difference varies along the wires.
3. The current flowing into one end of the wires can be different from the current flowing out the other end.
4. The current flowing into one end of the wire is unaffected by the impedance connecting the far end of the wires. This violates KCL[2].

Fortunately, these affects are only observed for rapidly changing signals or for long wires.

[2] KCL assumes that charge does not accumulate and thus the current flowing into the wire is the same as the current flowing out of the wire. But it does not account for the time delay. Thus if we wish KCL to be valid we need a better electrical model for the wires.

10 Transmission lines

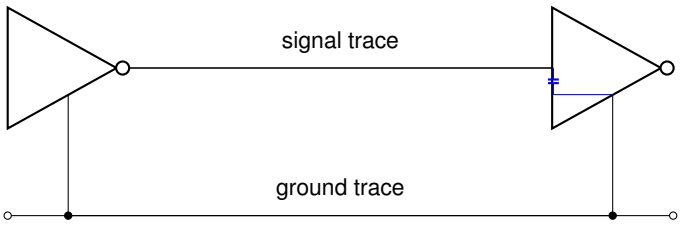


Figure 14.2: Signal connection with ground return.

When considering electrical propagation for high-speed signals we need to consider the signal and the return path, since the electric and magnetic fields propagate in the dielectric between the wires[3]. With a twisted pair or coaxial cable the return current flows in the other wire or the shield. With a microstrip the return current flows in the closest plane (either power or ground) taking the path of least impedance[4].

A rule of thumb is: *If the propagation delay, T, is longer than 1/6 of a signal transition time[5], T_t, then the voltage is not constant along the wires and thus a transmission line model is required*. In terms of the length, L of the transmission line,

$$L > \frac{1}{6} v T_t, \tag{14.1}$$

where v is the speed of propagation.

[3] The electrons are constrained to move along the conductors and move at a much slower speed.

[4] Flowing as close as possible to the signal trace to minimise the loop area and thus the inductance.

[5] A rising or falling edge.

T_t	L
2 ns	5 cm
2 μs	50 m
2 ms	50 km

Table 14.1: Wire length to be considered a transmission line, assuming v is half the speed of light.

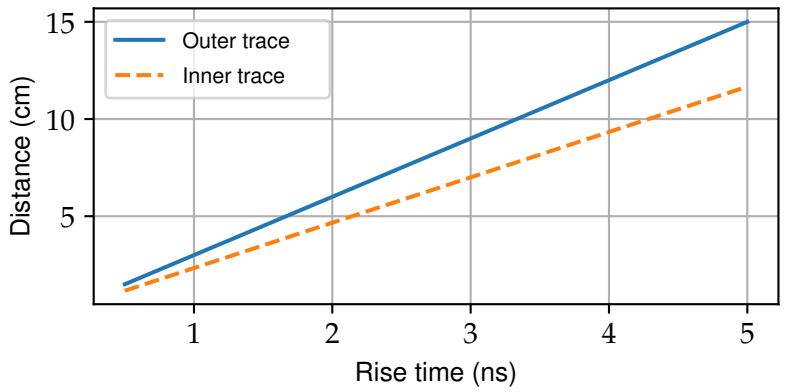


Figure 14.3: Maximum PCB trace length as a function of transition time before transmission line effects cannot be ignored. Note that fields within the PCB dielectric propagate more slowly than fields on the outside of the PCB.

11 Distributed transmission line model

We know that a transmission line formed by a signal and return conductor has capacitance, inductance, and resistance. We also know that a voltage change at one end does not cause an instantaneous change at the other end since it takes a finite time for the voltage and current signals to propagate. So how do we model this?

One solution is to break the transmission line of length, l, into a distributed model comprised of many little lumped models of length Δx, with each lumped model representing an incremental section of transmission line.

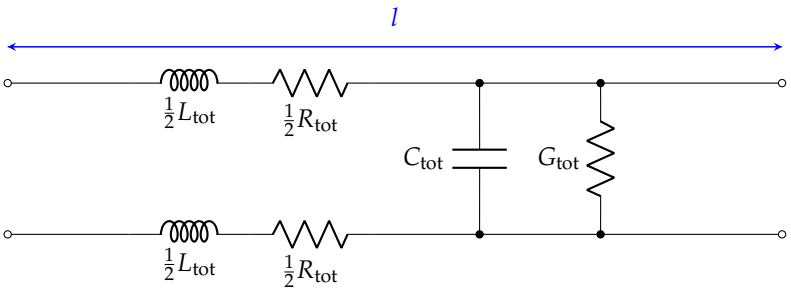


Figure 14.4: Lumped balanced transmission line model.

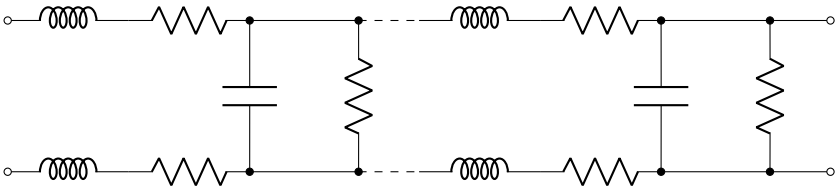


Figure 14.5: Distributed, lossy, balanced transmission line model, comprised of many identical sections.

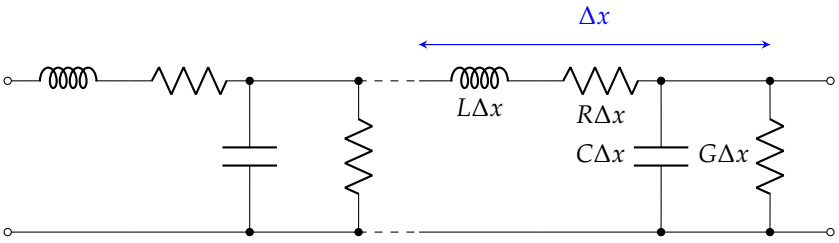


Figure 14.6: Distributed, lossy, unbalanced transmission line model. L, R, C, and G are per unit-length, i.e., $L = L_{\text{tot}}/l$.

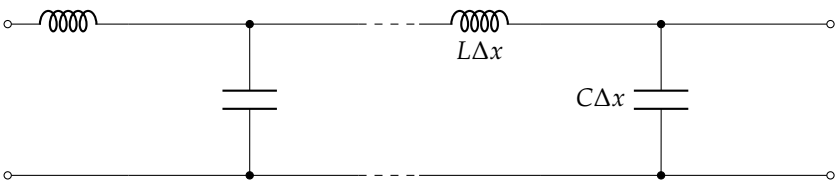


Figure 14.7: Distributed, lossless, unbalanced transmission line model.

It is common to model a transmission line with a resistance per unit length of R (Ω/m), an inductance per unit length of L (H/m), a capacitance per unit length of C (F/m), and a conductance per unit length of G (S/m).

100 Lossless transmission line

At high frequencies[6] the transmission line resistance is small compared to its reactance and a good approximation is to ignore losses. The resulting lossless transmission line model is shown in Figure 14.7.

Let's now connect a 5 V voltage source to an uncharged transmission line. Current will flow from the source to charge up the first part of the transmission line (see Figure 14.8). As this charges, current will flow to charge up the next part of the transmission line but not instantly due to the series inductance. This process continues until the entire transmission line charges to 5 V. The speed[7] that the voltage pulse propagates along the transmission line is given by

$$v = \frac{1}{\sqrt{LC}}. \tag{14.2}$$

Wow this is amazing! We've got all these inductances and capacitances and somehow magically their frequency dependence cancels[8].

The speed of propagation depends on the geometry of the signal and return conductors (in particular the width and separation) and the dielectric constant of the insulation in between. The speed of propagation is dominated by the permittivity of the dielectric and is typically about half to two-thirds the speed of light, see Table 14.2.

[6] At microwave frequencies the conductance cannot be ignored.

[7] This is the speed of the electromagnetic fields, not the electrons. The drift velocity of electrons is only about the speed of an ant.

[8] Thus there is no dispersion (or attenuation) and hence the voltage wave shape we apply to the transmission line does not change as it propagates.

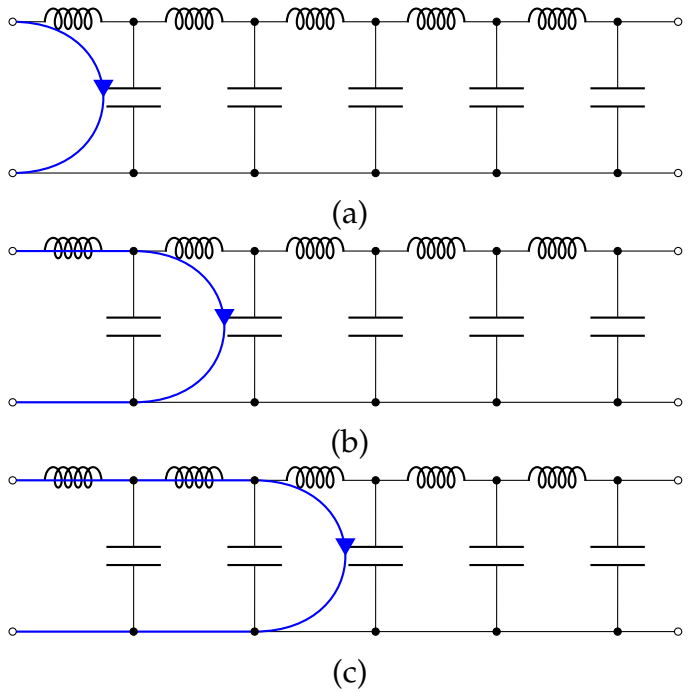


Figure 14.8: Transmission line charging. If the transmission line has a characteristic impedance, $Z_0 = 50$ ohm, and a 5 V pulse is applied, then a 100 mA current flows to charge the line. As each capacitor charges, the voltage and current pulse moves along the line. Note, the capacitors cannot charge simultaneously due to the inductances.

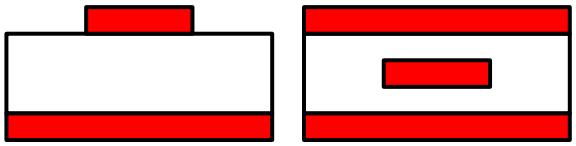


Figure 14.9: Cross-section of (a) microstrip and (b) stripline PCB traces. Signals are slightly slower in striplines since the field is constrained within the PCB dielectric.

Medium	Dielectric const.	Speed (mm/ns)	Delay (ps/mm)
Air	1.0	300	3.3
FR4 (outer trace)	2.8–4.5	140–180	5.5
FR4 (inner trace)	4.5	140	7.0

Table 14.2: Speed of propagation for different dielectrics. FR4 is the type of glass epoxy commonly used for PCBs. It is a blend of glass with $\epsilon_r = 6.0$ and a resin with $\epsilon_r \approx 3.0$ giving an effective dielectric constant $\epsilon_r \approx 4.5$.

100.1 Characteristic impedance

The other remarkable thing about a lossless transmission line is that the ratio of the voltage to the current along the line is a constant, given by

$$Z_0 = \sqrt{\frac{L}{C}}, \tag{14.3}$$

where Z_0 is called the *characteristic impedance*[9].

While Z_0 has units of ohms, it is not a measure of loss. It just relates the voltage to current of the propagating electric field at any point on a transmission line. Remember the propagating current has no idea what the load impedance is until it gets there. Once steady state is reached and the transmission line is charged, normal circuit theory will apply[10].

[9] It is difficult to make a transmission line with an impedance less than 30 Ω and more than 600 Ω.

[10] The current will depend on the load and not Z_0.

Transmission line	Z_0 (ohms)
Telephone wires in air	600
Free space	377
TV twin-lead cable	300
CAT5 unshielded twisted pair	100±15%
SATA/firewire cable	100
Ribbon cable 0.1″	100
USB shielded twisted pair	90±15%
TV coax cable	75
Ethernet coax cable	50

Table 14.3: Characteristic impedances. Like the speed of propagation, the characteristic impedance depends on the geometry of the conductors and the permittivity of the insulating dielectric.

100.10 Measuring characteristic impedance

Let's ponder for a moment what would happen if we try to measure the impedance of a 10 m length of CAT5 cable with an ohmmeter. What would we get? Would we see the transient response or just the steady state response? How about extending the cable to the moon? What would we measure[11]?

There are several ways to measure the characteristic impedance:

1. Connect a pulse generator to the transmission line and adjust the load impedance until there is no reflected pulse.

2. Connect a frequency generator to the transmission line and a known load, then calculate impedance from voltage drop at each frequency.

[11] Light takes about 2 s to come from the moon so an electric signal takes about 4 s. Thus we will measure 100 Ω (the characteristic impedance) of CAT5 until a reflection comes back in about 8 s. After a few bounces the cable will be charged and the ohmmeter will measure open circuit.

100.11 Controlled impedance

The characteristic impedance of stripline and microstrip traces depends on:

1. The trace thickness.

2. The trace width.

3. The separation of the trace from the return plane.

4. The dielectric constant of the insulation.

The trace thickness depends on the copper foil used and whether it is plated. The trace width can vary during manufacturing due to etching. To compensate this, PCB manufacturers can control the PCB thickness if they are told the desired characteristic impedance. However, the local etchant concentration varies (due to the layout) and thus the trace width can vary. Note if you choose trace widths for a certain characteristic impedance then decide to use more PCB layers then the characteristic impedance will change since the trace/plane separation is smaller.

The impedance of a PCB trace can be checked using time domain reflectometry equipment; this indicates where there are discontinuities, say when passing through vias or connectors. Some manufacturers put test traces on their boards so they can check the manufacturing process.

101 Summary

1. The signal and return conductors act as a waveguide for the propagating electromagnetic field. This can be modelled as a transmission line.
2. Electrical signals take time to propagate. The speed depends on the conductor geometry and dielectric constant of the insulation but is roughly half the speed of light in free space.
3. If the propagation delay T is longer than $1/6$ of a signal transition (a rising or falling edge) then the voltage along the transmission line is not constant and a distributed model is required.
4. The characteristic impedance is the frequency domain ratio of the voltage to current of a pulse propagating wave along a transmission line. It is also a function of geometry and dielectric constant of the insulation.
5. A lossless transmission line has a purely resistive characteristic impedance.
6. A lossy transmission line attenuates high frequencies more than low frequencies.
7. A lossy transmission line is dispersive; different frequencies have different phase speeds.
8. An infinite lossless transmission line looks like an impedance equal to its characteristic impedance.
9. Transmission line theory is applicable to any wave; electrical, optical, acoustical, elastic, etc.[12]

[12] One of my favourite books is "Waves" by Frank Crawford Jr. published in the 1960's. I found it in a second hand shop. It has many fun home experiments using light and sound waves.

110 Exercises

1. CAT5 unshielded twisted pair (UTP) cable has a characteristic impedance of 100 ohms. What is meant by characteristic impedance.
2. Why can't you measure the characteristic impedance of a 10 m length of CAT5 UTP cable with a multimeter.
3. Suggest a method where you can measure the characteristic impedance of a 10 m length of CAT5 UTP cable.
4. Why do signals on traces inside a multilayer PCB travel more slowly than on outer traces?

5. If electrical signals take 5 ns to propagate along 1 m length of RG58 coaxial cable how many metres of cable would be required for a 100 ns delay?

6. If a 1 m length of RG58 coaxial cable has a characteristic impedance of 50 Ω, what is the characteristic impedance of 2 m length of the same cable?

15
Transmission line reflections

When an EM disturbance (wave) travels along a transmission line, the ratio of voltage to current is determined by the *characteristic impedance* of the transmission line[1]. However, when the disturbance reaches the end of the line there can be an impedance mismatch. This causes the disturbance to reflect back along the transmission line, ensuring that Kirchhoff's laws are not violated.

[1] In the frequency domain.

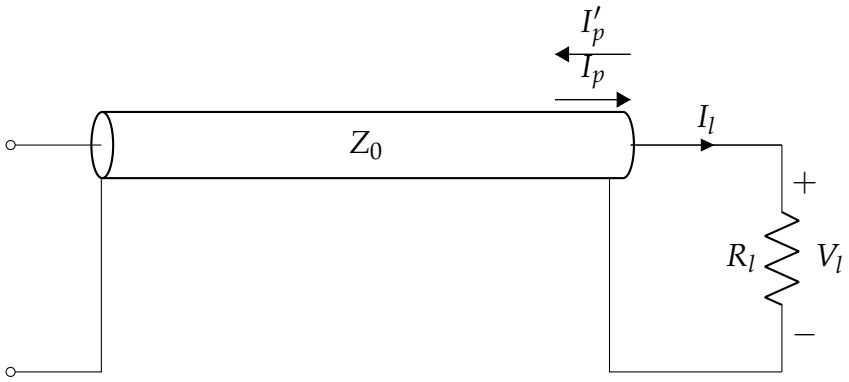


Figure 15.1: Lossless transmission line of characteristic impedance Z_0 terminated with load resistance R_l. Note, I'_p is defined in the opposite direction to I_p.

The voltage reflection coefficient for a lossless transmission line and resistive load is calculated by

$$\Gamma_l = \frac{R_l - Z_0}{R_l + Z_0}. \tag{15.1}$$

See Table 15.1 for a number of common scenarios.

Scenario	R_l	$\Gamma_l = \frac{V'_p}{V_p}$
Open circuit	∞	1
Short circuit	0	-1
Matched	Z_0	0

Table 15.1: Voltage reflection coefficients for a lossless transmission line.

1 Open-circuit lossless transmission line

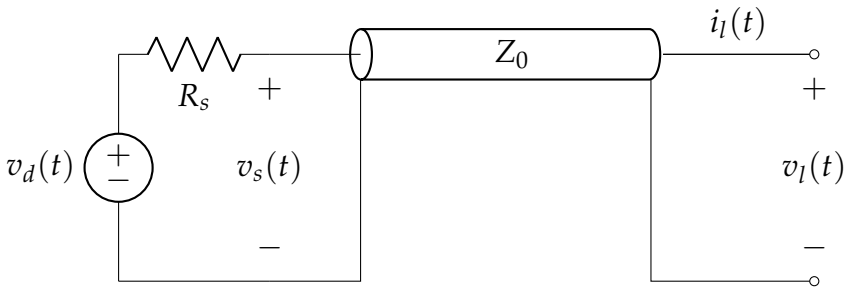


Figure 15.2: Open-circuit transmission line example where $\Gamma_l = 1$. $v_d(t)$ is the driver voltage, $v_s(t)$ is the source voltage, and $v_l(t)$ is the load voltage.

Consider a lossless transmission line of length $l = 15\,\text{cm}$. Since it is lossless then both the speed of propagation and characteristic impedance are both frequency independent constants. Let's say:

$$v = 0.5c = 1.5 \times 10^8\,\text{m/s}, \tag{15.2}$$
$$Z_0 = 60\,\Omega. \tag{15.3}$$

Now let's apply a $4\,\text{V}$ step to the transmission line through a source resistance $R_s = 20\,\text{ohm}$,

$$v_d(t) = 4u(t). \tag{15.4}$$

The step change produces an EM wave[2] that propagates along the transmission line. To make sense of the confusion that ensues, let's consider a number of time steps:

[2] This wave is not like an idyllic ocean wave, it has a step shape.

$t = 0$ From Figure 15.3, the wave has an initial voltage,

$$V_p = \frac{Z_0}{Z_0 + R_s} V_d = \frac{60}{60 + 20} \times 4 = 3\,\text{V}. \tag{15.5}$$

Note, this voltage will change when a reflection returns[3]. The EM wave has an initial current I_p, that can be found using the characteristic impedance, Z_0,

[3] Unless the transmission line is correctly terminated or infinitely long.

$$I_p = \frac{V_p}{Z_0} = \frac{3}{60} = 50\,\text{mA}. \tag{15.6}$$

$t = T$ The wave arrives at the load, where the delay T is given by the length of the line divided by the speed of propagation along the line, $T = l/v$. In this example,

$$T = \frac{l}{v} = \frac{15 \times 10^{-2}}{1.5 \times 10^8} = 1\,\text{ns}. \tag{15.7}$$

The propagating wave now faces a boundary condition. Since the transmission line is unterminated,

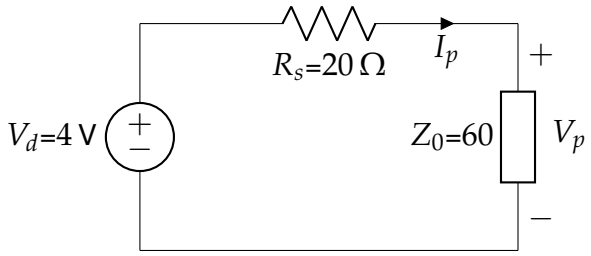


Figure 15.3: Circuit model to determine initial transmission line current and voltage.

zero current must flow in the load and so the wave is reflected. The current of the reflected wave is $I_p' = I_p = 50\,\text{mA}$ so that $I_p - I_p' = I_l = 0$. The voltage of the reflected wave is $V_p' = Z_0 I_p' = V_p = 3\,\text{V}$, corresponding to a voltage reflection coefficient of

$$\Gamma_l = \frac{V_p'}{V_p} = 1. \tag{15.8}$$

This result also can be found from the expression,

$$\Gamma_l = \frac{R_l - Z_0}{R_l + Z_0} = \frac{\infty - 60}{\infty + 60} = 1. \tag{15.9}$$

So at time $t = T$ at the load end of the transmission line we have an incident wave of 3 V and a reflected wave also of 3 V, giving a total voltage of $V_p + V_p' = 6\,\text{V}$.

$t = 2T$ The reflected 3 V wave arrives back at the source after charging the transmission line to 6 V. Since R_s does not match the characteristic impedance of the transmission line we get another reflection, where the source voltage reflection coefficient is

$$\Gamma_s = \frac{R_s - Z_0}{R_s + Z_0} = \frac{20 - 60}{20 + 60} = -0.5. \tag{15.10}$$

The wave reflected from the source has a voltage and current

$$V_p'' = -0.5 V_p' = -1.5\,\text{V}. \tag{15.11}$$
$$I_p'' = 0.5 I_p' = -25\,\text{mA}. \tag{15.12}$$

So at $t = 2T$ the voltage at the source is the initial $V_p = 3\,\text{V}$ plus the load reflection $V_p = 3\,\text{V}$ plus the source reflection $-0.5 V_p = -1.5\,\text{V}$, giving $3 + 3 - 0.5 \times 3 = 4.5\,\text{V}$.

$t = 3T$ The wave $V_p'' - 1.5\,\text{V}$ arrives back at the load. This will be re-reflected and so-on until the voltages on the transmission line converge to the steady-state voltage of 4 V that you would expect from a lossless pair of wires, see Figure 15.5.

1.1 *Open-circuit transmission line lattice diagram*

A lattice diagram is shown in Figure 15.4 for a mismatched transmission line. This is a graphical means of determining the signals on a transmission line with reflections. Each point on the blue curve shows where the wavefront is as a function of distance, x, and time, t.

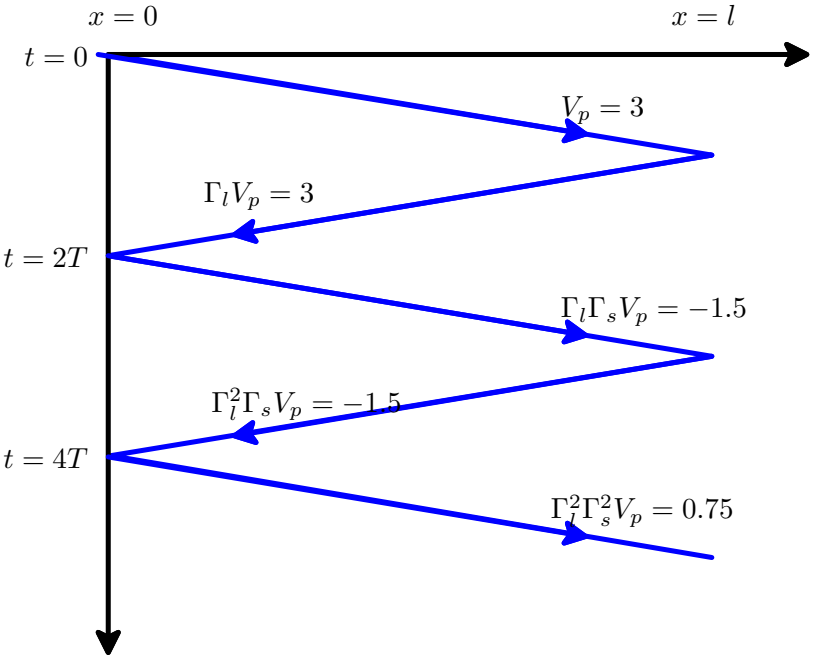


Figure 15.4: Mismatched transmission line reflection timeline (lattice or bounce diagram). Note the voltage reflection coefficients are real since the load, source, and characteristic impedance are all resistive.

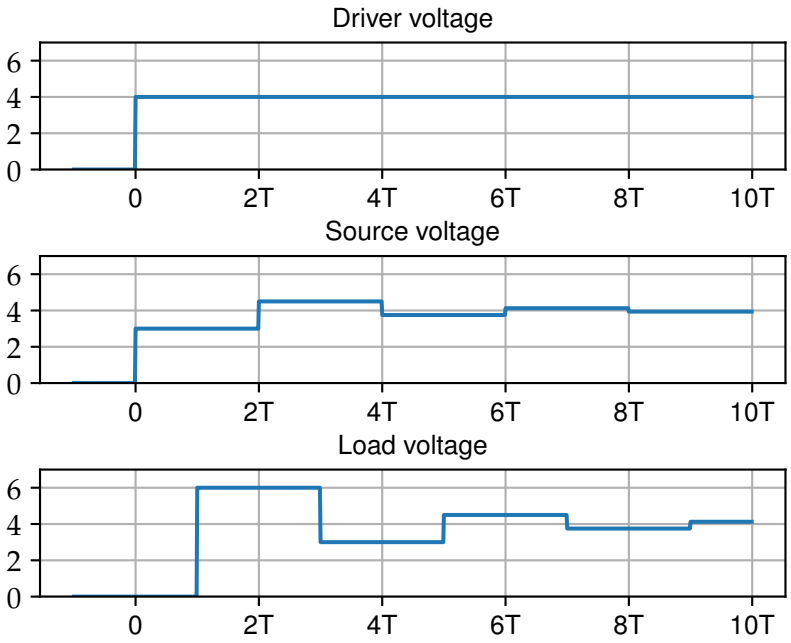


Figure 15.5: Open-circuit lossless transmission line signals for $Z_0 = 60\,\Omega$ and $R_s = 20\,\Omega$. Note how all the signals converge to the steady-state of 4 V.

10 Lossless transmission line reflections

Lossless transmission lines are simpler to model since the characteristic impedance, Z_0, is a real constant. This simplifies the circuit analysis since it can be performed in the time domain[4].

For example, consider Figure 15.1 where a wave with voltage amplitude V_p is propagating from left to right. The amplitude of the propagating wave is found from the characteristic impedance to be

$$I_p = \frac{V_p}{Z_0}. \tag{15.13}$$

But what happens if we connect a load impedance R_l to the end of the transmission line? From Ohm's law we expect

$$I_l = \frac{V_l}{R_l}. \tag{15.14}$$

So unless $Z_0 = R_l$, Kirchhoff's current law will be violated since $I_p \neq I_l$.

To avoid the end of the universe, a wave of voltage amplitude V_p' and current amplitude I_p' is reflected back to the source to ensure the boundary condition imposed by the load impedance is met. The ratio of the voltage of the reflected wave to the voltage of the incident wave is called the voltage reflection coefficient[5] and can be shown to be,

$$\Gamma_l = \frac{V_p'}{V_p} = \frac{R_l - Z_0}{R_l + Z_0}. \tag{15.15}$$

Note, that when R_l is chosen to be equal to Z_0, then there is no reflected wave.

[4] With lossy transmission lines, Z_0 is frequency dependent and Laplace or Fourier analysis is required.

[5] The current reflection coefficient depends on the definition for the direction of I_p'. If I_p is defined in the opposite direction to I_p, then the voltage and current reflection coefficients are the same. Moreover, $V_p = Z_0 I_p$ and $V_p' = Z_0 I_p'$.

10.1 Voltage reflection coefficient analysis

The amplitude of the reflected voltage wave, V_p', can be found by realising that

$$V_p + V_p' = V_l, \tag{15.16}$$

and for the continuity of current,

$$I_p - I_p' = I_l. \tag{15.17}$$

Using the definition of the characteristic impedance,

$$\frac{V_p'}{I_p'} = Z_0, \tag{15.18}$$

thus the current continuity equation (15.17) can be written

$$\frac{V_p - V_p'}{Z_0} = \frac{V_l}{R_l}, \tag{15.19}$$

or equivalently,

$$\left(V_p - V_p'\right)\frac{R_l}{Z_0} = V_l. \tag{15.20}$$

After equating (15.16) and (15.20), the voltage reflection coefficient[6] is

$$\Gamma_l = \frac{V_p'}{V_p} = \frac{R_l - Z_0}{R_l + Z_0}. \tag{15.21}$$

[6] At the load.

When the transmission line is lossless and the load impedance is purely real then the reflection coefficient Γ_l is also real and frequency independent.

10.10 Multiple reflections

When a reflected wave reaches the source end of a transmission line, there will be another reflected wave if the source impedance does not match the characteristic impedance. The source voltage reflection coefficient is given by

$$\Gamma_s = \frac{V_p''}{V_p'} = \frac{R_s - Z_0}{R_s + Z_0}. \tag{15.22}$$

The wave reflected from the source will be re-reflected from the load and so on until eventually steady state is reached.

10.11 Steady state

After a while the sum of all the reflections reaches a steady state value[7]. Interestingly, the steady state result is the same as given by conventional circuit analysis with the transmission line replaced by a pair of ideal wires.

[7] Especially, if there is some loss in the transmission line.

10.100 Multiple reflections example

Consider a driver, modelled as a voltage source $v_d(t)$ with output resistance R_s, that drives a lossless microstrip of characteristic impedance Z_0 and electrical delay T. The microstrip is terminated by a resistance R_r.

The initial voltage signal that is driven onto the microstrip is reduced by the voltage divider comprised of the source and characteristic impedances:

$$V_p = \frac{Z_0}{Z_0 + R_s} V_d = \frac{1 - \Gamma_s}{2} V_d. \tag{15.23}$$

Since all the impedances are real, the reflection coefficients are also real, and thus

$$V_p = \frac{Z_0}{Z_0 + R_s} V_d = \frac{1 - \Gamma_s}{2} V_d. \tag{15.24}$$

After a period T the voltage wave reaches the load and is reflected back with a reflection coefficient Γ_l. The reflected voltage signal is $\Gamma_l V_p$. This arrives at the source after another delay of T where it is re-reflected by a factor Γ_s giving $\Gamma_s \Gamma_l V_p$. This then propagates back down the microstrip to the load where it gets reflected again, and so on. The voltages along the microstrip are thus an infinite sum of reflected components. For example, the voltage signal at the source end of the microstrip is

$$\begin{aligned} v_s(t) &= v_p(t) + \Gamma_l v_p(t - 2T) + \Gamma_s \Gamma_l v_p(t - 2T) \\ &\quad + \Gamma_l \Gamma_s \Gamma_l v_p(t - 4T) + \Gamma_s \Gamma_l \Gamma_s \Gamma_l v_p(t - 4T) \\ &\quad + \cdots, \end{aligned} \tag{15.25}$$

and the voltage signal at the load end is

$$\begin{aligned} v_l(t) &= v_p(t - T) + \Gamma_l v_p(t - T) + \Gamma_s \Gamma_l v_p(t - 3T) \\ &\quad + \Gamma_l \Gamma_s \Gamma_l v_p(t - 3T) + \cdots. \end{aligned} \tag{15.26}$$

10.101 Mismatched transmission line example

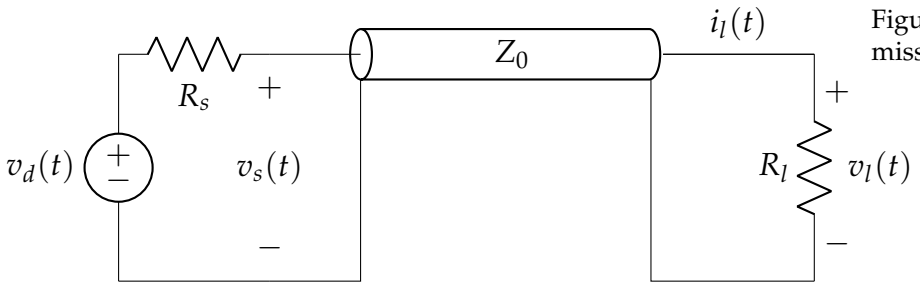


Figure 15.6: Mismatched transmission line example.

Let's now consider what happens if the transmission line in the previous example is terminated with a $20\,\Omega$ load, i.e., $R_l = 20\,\Omega$, $Z_0 = 60\,\Omega$, $v_d = 4u(t)$ V, and $R_s = 20\,\Omega$. Again, the impedances are resistive so we

can take some mathematical shortcuts and avoid using the Laplace domain.

In the steady-state, we can replace the transmission line with a pair of ideal wires and will find that the load voltage is

$$V_l = \frac{R_l}{R_l + R_s} V_d = \frac{20}{20 + 20} \times 4 = 2\,\text{V}, \tag{15.27}$$

and the load current is

$$I_l = \frac{V_l}{R_l} = \frac{2}{20} = 100\,\text{mA}. \tag{15.28}$$

Let's see how the voltage waves bouncing up and down a transmission line converge to these steady-state values:

$t = 0$ The initial voltage wave that appears on the line has an amplitude given by

$$V_p = \frac{Z_0}{Z_0 + R_s} V_d = \frac{60}{60 + 20} \times 4 = 3\,\text{V}, \tag{15.29}$$

with an associated current wave of

$$I_p = \frac{V_p}{Z_0} = \frac{3}{60} = 50\,\text{mA}. \tag{15.30}$$

$t = T$ The wave reaches the load but the current is too small (or the voltage is too high) to meet the boundary condition imposed by the load. We need a current of

$$I_l = \frac{V_p}{R_l} = \frac{3}{20} = 150\,\text{mA}, \tag{15.31}$$

but only 50 mA has propagated. Thus there is mismatch and a reflection is generated. The voltage reflection coefficient is

$$\Gamma_l = \frac{R_l - Z_0}{R_l + Z_0} = \frac{20 - 60}{20 + 60} = -0.5, \tag{15.32}$$

and so the reflection has a voltage

$$V_p' = \Gamma_l V_p = -0.5 \times 3 = -1.5\,\text{V}, \tag{15.33}$$

and a current

$$I_p' = \Gamma_l I_p = -0.5 \times 50 = -25\,\text{mA}. \tag{15.34}$$

Thus the total voltage at the load is

$$V_l = V_p + V_p' = 3 - 1.5 = 1.5\,\text{V}, \tag{15.35}$$

and the total load current is

$$I_l = I_p - I'_p = 50 - (-25) = 75\,\text{mA}. \tag{15.36}$$

As a check,

$$\frac{V_l}{I_l} = \frac{1.5}{75 \times 10^{-3}} = 20\,\Omega. \tag{15.37}$$

This is the same as the load resistance so we have not made a mistake.

$t = 2T$ At the source there is also a mismatch with a reflection coefficient

$$\Gamma_s = \frac{R_s - Z_0}{R_s + Z_0} = \frac{20 - 60}{20 + 60} = -0.5, \tag{15.38}$$

(which coincidentally is the same as the load reflection coefficient for this example). The reflection gets re-reflected with a voltage amplitude,

$$V''_p = \Gamma_s V'_p = -0.5 \times -1.5 = 0.75\,\text{V}, \tag{15.39}$$

and a current amplitude

$$I''_p = \Gamma_s I'_p = -0.5 \times -25 \times 10^{-3} = 12.5\,\text{mA}. \tag{15.40}$$

The total voltage at the source is now

$$V_p + V'_p + V''_p = 3 - 1.5 + 0.75 = 2.25\,V, \tag{15.41}$$

and the current is

$$I_p - I'_p + I''_p = 50 - (-25) + 12.5 = 87.5\,\text{mA}. \tag{15.42}$$

$t = 3T$ The re-reflected wave is reflected again. This time we have

$$V'''_p = \Gamma_l V''_p = -0.5 \times 0.75 = -0.375\,\text{V} \tag{15.43}$$

and

$$I'''_p = \Gamma_l I''_p = -0.5 \times 12.5 \times 10^{-3} = -6.25\,\text{mA}. \tag{15.44}$$

The load voltage is now

$$V_l = V_p + V'_p + V''_p + V'''_p = 3 - 1.5 + 0.75 - 0.375 = 1.875\,\text{V}, \tag{15.45}$$

and the load current is

$$I_l = I_p - I'_p + I''_p - I'''_p = 50 - (-25) + 12.5 - (-6.25) = 93.75\,\text{mA}. \tag{15.46}$$

Again if we check the ratio of the load voltage to the load current we should get the load impedance,

$$\frac{1.875}{93.75 \times 10^{-3}} = 20\,\Omega. \tag{15.47}$$

11 Lossy transmission line reflections

When a transmission line is lossy and/or when the load is reactive, the reflection coefficient is a function of frequency. For example, in the Laplace domain

$$\Gamma_l(s) = \frac{V_p'(s)}{V_p(s)} = \frac{Z_l(s) - Z_0(s)}{Z_l(s) + Z_0(s)}. \tag{15.48}$$

While $\Gamma_l(s)$ is called the reflection coefficient, it should be considered a transfer function. Only when both Z_0 and Z_l are purely real is $\Gamma_l(s)$ a constant.

100 Summary

1. The amplitude of the initial voltage pulse on a transmission line[8] is found from the voltage divider formed by the source resistance and characteristic impedance.
2. If the load impedance on a transmission line is different from the characteristic impedance there will be a reflection.
3. If the source impedance line is different from the characteristic impedance, the reflected pulse will be re-reflected at the source.
4. Once all the reflections have died out we get the steady-state response predicted by simple circuit theory.
5. Time domain analysis is difficult for lossy transmission lines or if the load and source impedances are not real (purely resistive).
6. RF electronics classes look at transmission lines in the frequency domain.
7. RF impedance matching is primarily concerned with the maximum transmission of energy over a narrow frequency band.
8. Impedance matching is important in the design of musical instruments.

[8] Before any reflections.

101 Further reading

Howard Johnson and Martin Graham, "High-speed digital design: A handbook of black magic", Prentice Hall, 1993.

Howard Johnson and Martin Graham, "High-speed signal propagation: Advanced black magic", Prentice Hall, 2003.

Donald Dearholt and William McSpadden, "Electromagnetic wave propagation", McGraw-Hill, 1973.

Simon Ramo, John Whinnery and Theodore van Duzer, "Fields and waves in communication electronics", John Wiley and Sons, 1965.

Justin Davis, "High-speed digital system design", Morgan and Claypool, 2006.

110 *Exercises*

1. A 500 mm length of ribbon cable with a 120 ohm characteristic impedance is open-circuit. What is the voltage reflection coefficient? State any assumptions.
2. A 500 mm length of ribbon cable with a 120 ohm characteristic impedance is short-circuit. What is the voltage reflection coefficient? State any assumptions.
3. A 500 mm length of ribbon cable with a 120 ohm characteristic impedance is terminated with 80 ohms. What is the voltage reflection coefficient? State any assumptions.
4. A 500 mm length of ribbon cable with a 120 ohm characteristic impedance is terminated with 80 ohms. What is the current reflection coefficient? State any assumptions.
5. A PCB microstrip trace with a 50 ohm characteristic impedance is driven from an IC with a 25 ohm output impedance, an open-circuit logic low voltage of 0 V, and an open-circuit logic high voltage of 3 V. If the IC output switches from logic low to logic high in 1 ns, what is the amplitude of the voltage and current at any point on the trace prior to any reflections? State any assumptions.
6. If the microstrip trace in the previous question is terminated by a 75 ohm resistor, what is the steady-state voltage and current at the load?
7. A 50 ohm coaxial cable is connected to a 75 ohm coaxial cable. What is the reflection coefficient and how could the cables be connected to avoid a reflection?
8. Why do you think a trumpet has a flared bell?

16
Transmission line termination

There are several undesirable problems caused by reflections on poorly terminated transmission lines::

1. False clocking due to spurious clock edges.
2. Timing jitter.
3. Reduced noise margins.
4. Charge injection onto the die substrate (via ESD protection diodes on CMOS gates)[1].
5. CMOS latch-up.
6. Degraded EMC.

[1] This can cause memory errors (DRAM) and degraded performance of mixed signal devices (ADCs/DACs/PLLs).

There are many termination methods with tradeoffs in power consumption, effect on noise margins, and component counts.

1 Receiver termination

Most termination methods are applied at the receiver. These include parallel termination, passive and active Thévenin termination, AC termination, and Schottky diode termination.

1.1 Parallel termination

This uses a pull-up or pull-down resistor[2]. The disadvantage of parallel termination is that a strong driver is required to provide the large drive current to maintain the active logic level. This produces a high static power requirement when maintaining a logic low with a pull-up or a logic high with a pull-down.

[2] Generally pull-ups used for TTL since TTL poorer at sourcing current.

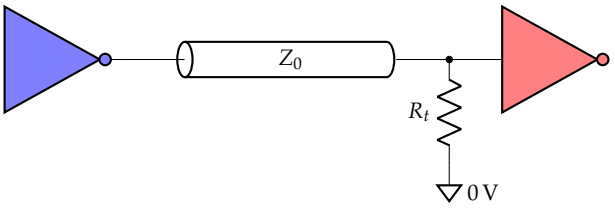


Figure 16.1: Parallel termination with a pull-down resistor.

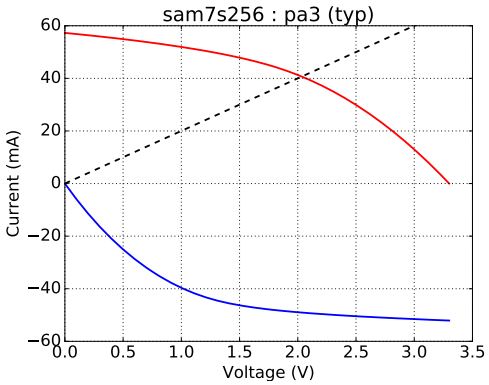


Figure 16.2: Parallel pull-down termination loading (50 Ω) on a SAM7S PIO pin. Note, the driver needs to source 40 mA in the logic high state and the output high voltage is decreased to 2.0 V, reducing the low noise margin.

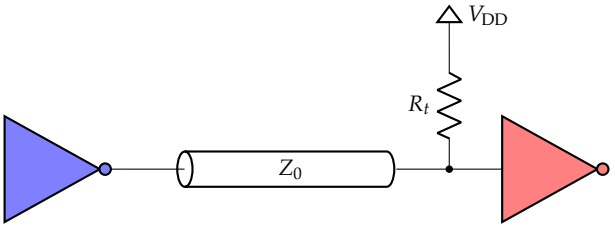


Figure 16.3: Parallel termination with a pull-up resistor.

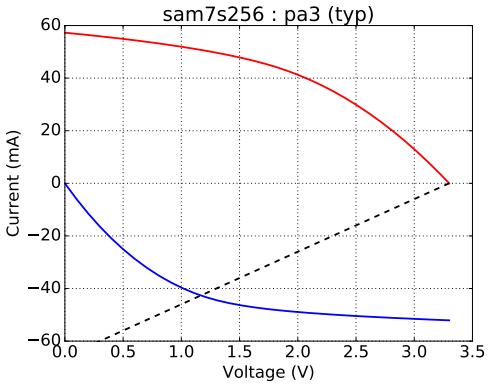


Figure 16.4: Parallel-pull-up termination loading (50 Ω) on a SAM7S PIO pin. Note, the driver needs to sink -43 mA in the logic low state and the output low voltage is increased to 1.2 V, reducing the high noise margin.

1.10 *Passive Thévenin (split) termination*

This is similar to parallel termination but uses two resistors as a voltage divider[3]. Compared to parallel termination, this reduces the peak drive current and improves the noise margin but suffers a static power loss regardless of logic level. The parallel combination of resistors

[3] This was common for old buses, e.g., SCSI-1 (330/220 Ω) and VME (470/330).

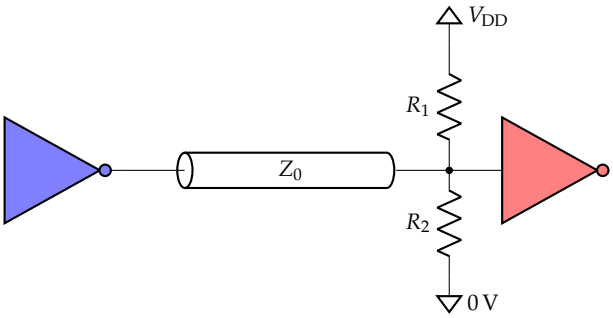


Figure 16.5: Passive Thévenin termination using a resistor voltage divider.

is chosen to match Z_0, i.e.,

$$Z_0 = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2}. \tag{16.1}$$

For CMOS the Thévenin voltage V_t is chosen to be $V_{DD}/2$ to equalise the high and low current requirements. Since

$$V_t = \frac{R_2}{R_1 + R_2} V_{DD}, \tag{16.2}$$

then $R_1 = R_2$[4].

[4] For TTL $R_2 > R_1$ since TTL better at sinking current.

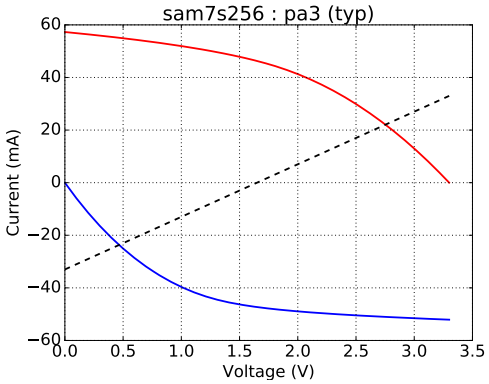


Figure 16.6: Split termination loading (50 Ω) on a SAM7S PIO pin.

1.11 *Active Thévenin termination*

Similar to passive Thévenin termination but uses a voltage regulator[5] to set the Thévenin voltage to avoid static power consumption in a resistive divider. The voltage regulator can be shared over many signals. Used by SCSI-2 bus, DDR memory.

[5] This must be a special regulator that can sink and source current.

1.100 *Fly-by termination*

When it is not possible to get the termination resistors close enough to the receiver pin, it is better to continue the trace past the receiver pin to the termination resistors. This is called *fly-by termination*.

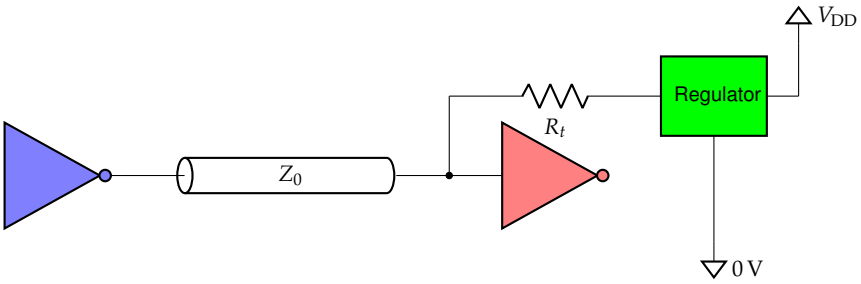


Figure 16.7: Active Thévenin termination showing a voltage regulator (that can sink and source current) to set the Thévenin termination voltage.

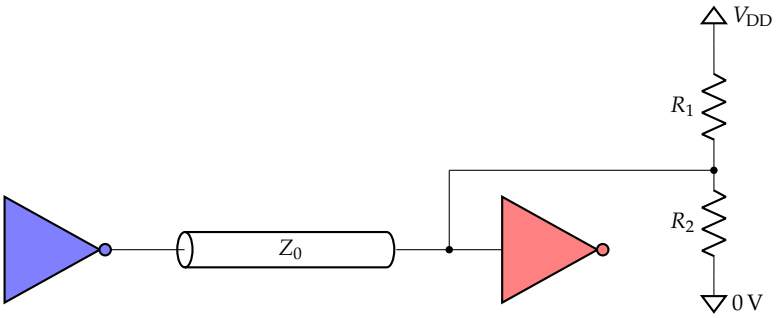


Figure 16.8: Fly-by termination, here using passive Thévenin termination.

1.101 *Schottky-diode termination*

Hot-carrier Schottky diodes can clamp the signal to reduce undershoot and overshoot[6]. It is useful when the characteristic impedance is unknown or when the trace has long stubs.

[6] This technique is also called forced perfect termination.

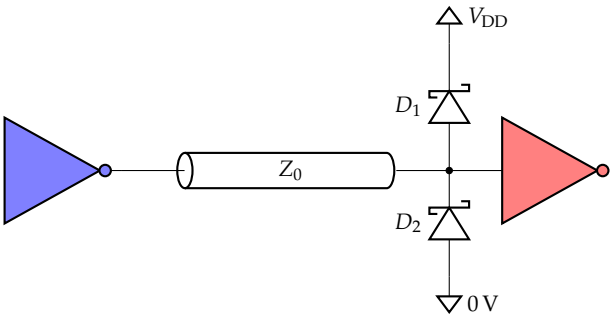


Figure 16.9: Schottky-diode termination.

1.110 *AC termination*

AC termination is a form of parallel termination with a capacitor in series to avoid static power dissipation. The rule of thumb for the capacitor value is

$$C = \frac{Z_0}{6\pi f T_r}, \tag{16.3}$$

where T_r is the signal rise/fall time and f is the switching frequency. The bigger the value the smoother the

waveform but the greater the power consumption (it tends to the parallel termination case). If the capacitor is too small then there is overshoot.

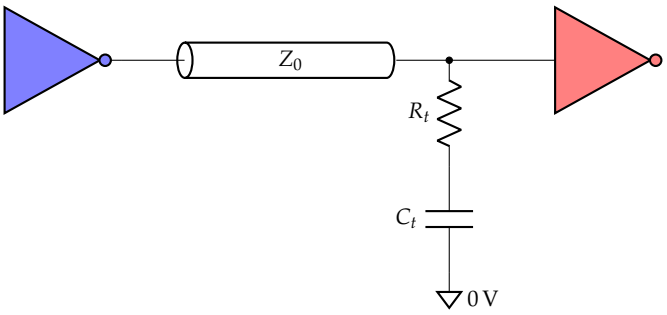


Figure 16.10: AC termination.

10 *Source termination*

Source termination uses a resistor in series with the transmission line, thus it is commonly called *series termination*. This resistor has a value $Z_0 - R_o$ (where R_o is the driver output resistance) to ensure a zero reflection coefficient at the source. While this halves the voltage driven onto the line, the unterminated receiver doubles the voltage a short time later.

Series termination is used for PCI and USB buses and is common with CMOS logic and battery powered systems since there is no static power loss. It can also be better than receiver termination since the source is resistive while the receiver usually is capacitive. However, the output impedance of a device can be different for different logic states and can dynamically change. Thus the termination resistor value is a compromise. It is also difficult to minimise the trace length before the terminating resistor (say for a signal from the middle of a BGA package[7]).

[7] Some MCUs, such as the Atmel SAM4s, use on die termination of a 36 Ω resistor in series with the output pin. The drivers have a nominal impedance of 10 Ω so the PCB trace should have $Z_0 \approx 50\,\Omega$.

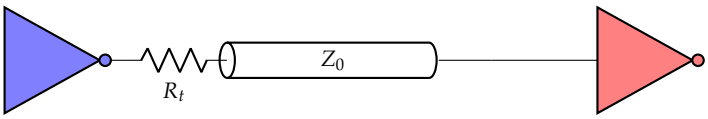


Figure 16.11: Series (source) termination.

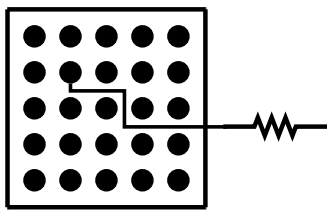


Figure 16.12: The problem of series (source) termination with BGA packages; it is hard to minimise the trace length before the resistor.

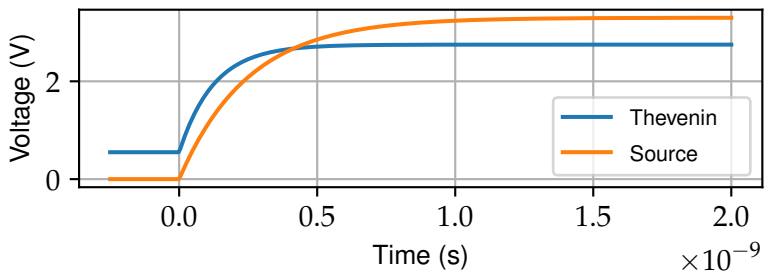


Figure 16.13: A comparison of the rise-time for source and Thévenin termination.

Source termination has slower switching speeds than Thévenin termination since the voltage pulse applied to the transmission line is halved. For Thévenin termination it is higher[8], since

$$v_p(t) = \frac{Z_0}{Z_0 + R_o} v_d(t). \tag{16.4}$$

A lower voltage pulse results in a lower current and hence the load capacitance takes longer to charge. On the other hand, Thévenin termination reduces the voltage swing, see Figure 16.13.

[8] Assuming $R_o < Z_0$.

11 Source and receiver termination

Some applications[9] use both source and receiver termination[10] $\Gamma_s = \Gamma_r = 0$, and

$$v_s(t) = v_p(t), \tag{16.5}$$

$$v_r(t) = v_p(t - T) = \frac{1}{2} v_d(t - T). \tag{16.6}$$

Note how the voltage signals are halved. This compromises noise margins but provides some immunity to reflections from discontinuities such as vias and stubs.

[9] DDR RAM uses this technique (a combination of series and active Thévenin termination called *stub series termination* (SSTL)) with receivers designed for the reduced voltage levels.

[10] Johnson (2003) calls this the axe-murderer approach since an axe-murderer never just takes one whack at the problem; he keeps whacking until the problem is solved.

100 Branches or Ys

When a transmission line branches there is an impedance mismatch. The propagating pulse takes both paths and thus sees two transmission lines in parallel. If each path has the same Z_0, the reflection coefficient is

$$\Gamma = \frac{Z_0/2 - Z_0}{Z_0/2 + Z_0} = -\frac{1}{3}. \tag{16.7}$$

This causes a reflection of $-v_p/3$ back toward the driver. Meanwhile a pulse of amplitude $2/3 v_p$ travels down each branch.

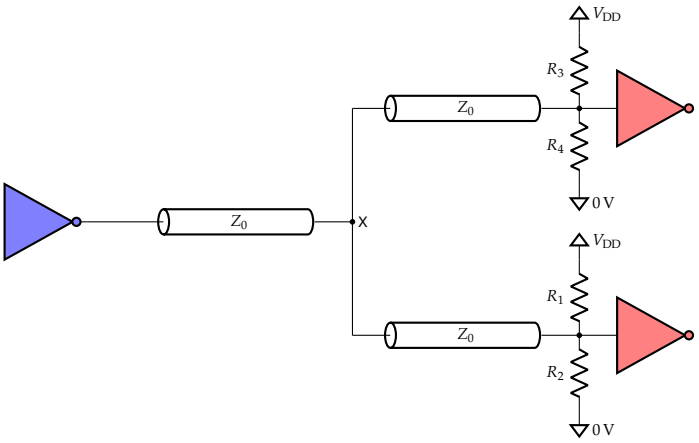


Figure 16.14: Branched transmission line. This has an impedance mismatch at point X even though each transmission line is correctly terminated.

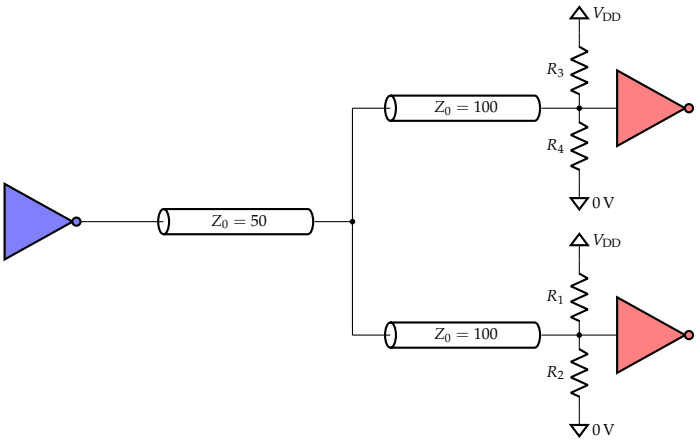
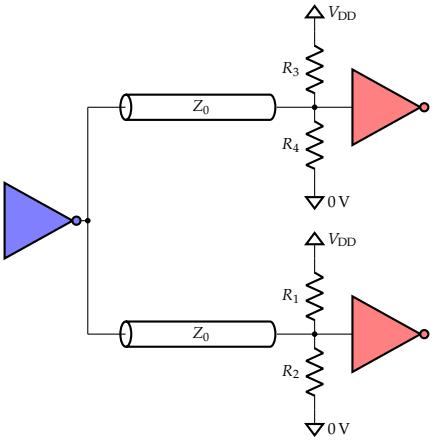


Figure 16.15: Correctly terminated branched transmission lines.



A typical bus has many distributed transceivers and thus there are many stubs. An analytic model of the bus soon becomes infeasible due to the multiple reflections from the stubs. The key is to keep the stubs very short and space them uniformly so that the capacitive load of each device is distributed evenly. This will slow the speed of propagation and reduce the characteristic impedance.

101 Bidirectional signal termination

With multiple devices driving a common transmission line, either series or parallel termination can be employed. Since only one device can drive the trace at a time (half-duplex), the other devices must be tristated[11].

[11] When using Thévenin termination and there is no active driver, the inputs will not be logic high or low and thus there will be a static current drain through the input transistors.

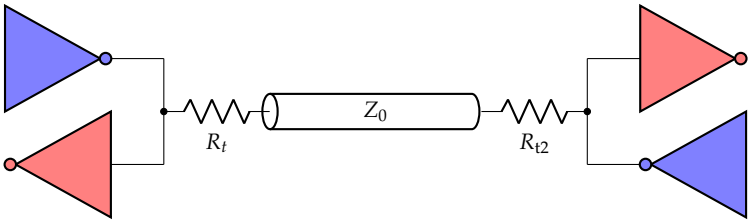


Figure 16.16: Half-duplex series termination.

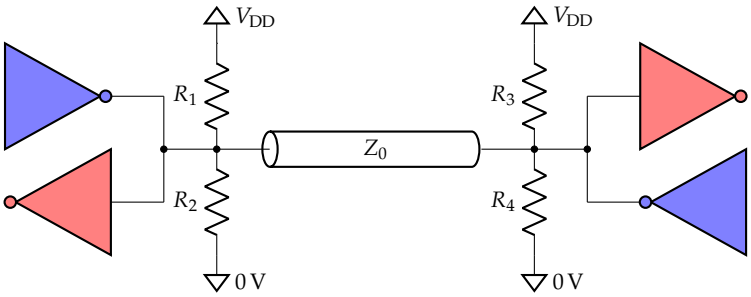


Figure 16.17: Half-duplex passive Thévenin termination.

110 Termination analysis

This section is for those who want more understanding.

110.1 Receiver termination

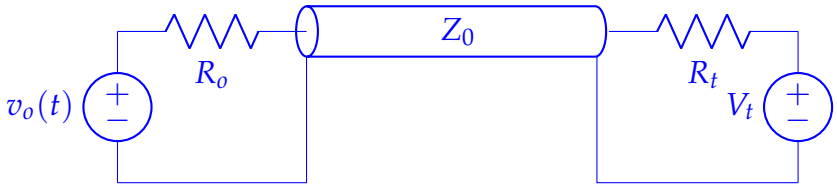


Figure 16.18: Receiver terminated transmission line schematic.

Consider a transmission line of characteristic impedance Z_0 driven by a driver with a voltage $v_o(t)$ and resistance R_o. Now the transmission line and driver resistance act as a voltage divider, so the voltage pulse v_p driven onto the transmission line is

$$v_p(t) = \frac{Z_0}{Z_0 + R_o} v_o(t). \tag{16.8}$$

Note, the input impedance of a lossless transmission line is Z_0. Also note that $v_s(t)$ is not necessarily the same as $v_p(t)$ since it depends on reflections.

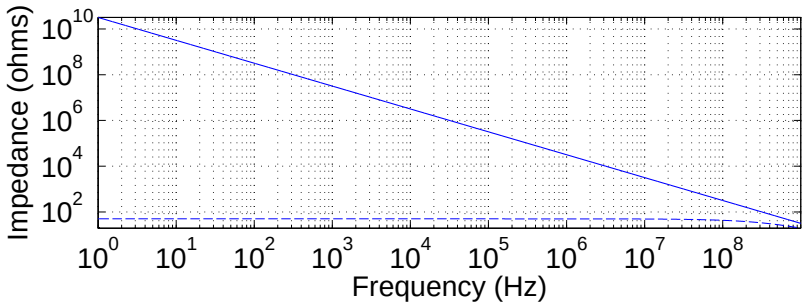


Figure 16.19: Impedance of a 5 pF capacitor and impedance of 5 pF in parallel with 50 Ω (dashed). Note the RC time constant is 0.25 ns and the capacitance can be neglected for frequencies up to 200 MHz.

If the receiver is correctly terminated[12] with the characteristic impedance of the transmission line then $\Gamma_r = 0$ and thus there is no reflected wave. Therefore, the voltage signals at the source and receiver ends of the transmission line simplify to

$$v_s(t) = v_p(t), \tag{16.9}$$
$$v_r(t) = v_p(t - T). \tag{16.10}$$

In practice, perfect termination cannot be achieved due to the receiver input capacitance.

[12] The impedance of the input capacitance is small compared to the termination resistance except at high frequencies.

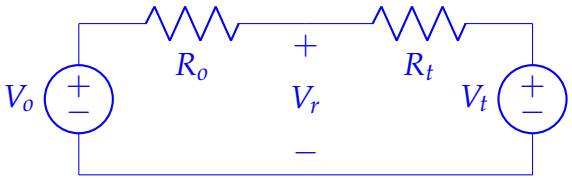


Figure 16.20: General receiver termination DC model.

In general, all resistive receiver termination techniques have the form shown in Figure 16.18. This uses a simple model for a CMOS driver where the output resistance is assumed constant[13]. For example, with pull-up parallel termination $V_t = V_{DD}$ and with pull-down parallel termination $V_t = 0$ (see Table 16.1). Under DC conditions we can simplify the model to Figure 16.20 (where $V_s = V_r$ since a lossless transmission line is simply a piece of perfect wire at DC). Applying KVL to the circuit,

$$V_r = \frac{R_t}{R_o + R_t} V_o + \frac{R_o}{R_o + R_t} V_t \tag{16.11}$$

and

$$I_s = \frac{V_o - V_t}{R_o + R_t}. \tag{16.12}$$

We need to be careful calculating the static power dissipation in the driver output resistance and termination resistor. For parallel termination, it is simply

$$P = \frac{(V_o - V_t)^2}{R_o + R_t}. \tag{16.13}$$

With passive Thévenin termination there is a gotcha since using (16.13) will under-estimate the power. This is a limitation of a Thévenin equivalent model[14]. Instead, we must consider the power dissipated in all the resistors of the original network,

$$P = \frac{(V_o - V_r)^2}{R_o} + \frac{(V_{DD} - V_r)^2}{R_1} + \frac{V_r^2}{R_2}. \tag{16.14}$$

For an example, let's consider a simple output model of a driver consisting of a voltage source in series with an output resistance $R_o = 20\,\Omega$. This gives a high voltage state $V_{oh} = 3.0\,V$ and a low voltage state $V_{ol} = 0.3\,V$.

[13] This is only valid for high resistance loads.

[14] If you're not convinced, try letting $R_2 = 0$. You will end up with two independent circuits with power being dissipated by R_o and R_1. Now since $R_2 = 0$ then $R_t = 0$, and thus (16.13) does not consider the power dissipated in R_1. If you are still not convinced, consider a resistor in parallel with a voltage source; this will dissipate heat but not affect the voltage of the source.

Method	R_t	V_t
Parallel pull-up	R_t	V_{DD}
Parallel pull-down	R_t	0
Passive Thévenin	$R_1 \parallel R_2$	$\frac{R_2}{R_1+R_2} V_{DD}$
Active Thévenin	R_t	V_t

Table 16.1: Termination resistances and Thévenin voltages for different resistive receiver termination methods.

Method	V_{rl}	V_{rh}	I_{sl}	I_{sh}	P_l	P_h
Parallel pull-up	1.2	3.1	-42.9	-4.3	128.6	1.3
Parallel pull-down	0.2	2.1	4.3	42.9	1.3	128.6
Passive Thévenin	0.7	2.6	-19.3	19.3	96.0	96.0
Active Thévenin	0.7	2.6	-19.3	19.3	26.0	26.0

Table 16.2: Static voltages at the receiver V_r (V), source currents I_s (mA), and power dissipation P (mW) for different resistive receiver termination methods. The subscript l denotes logic low and the subscript h denotes logic high. The power consumption for the active Thévenin method does not include losses in the regulator providing the termination voltage.

If the trace has a characteristic impedance $Z_0 = 50\,\Omega$ we need to select the termination resistor/resistors so that $R_t = Z_0 = 50\,\Omega$. The static receiver voltages, source currents, and power dissipation are tabulated in Table 16.2 with $V_t = 1.65\,\text{V}$ for the passive and active Thévenin methods.

110.10 Source termination

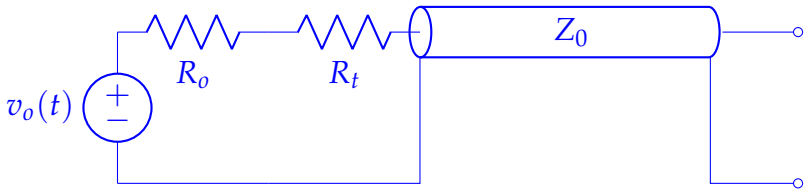


Figure 16.21: Series terminated transmission line schematic. R_o is the output resistance of the driver.

The effect of the series terminator resistor, with the output resistance of the source, is to halve the voltage that gets driven onto the transmission line,

$$v_p(t) = \frac{Z_0}{Z_0 + Z_0} v_d(t) = \frac{1}{2} v_d(t). \tag{16.15}$$

Now if the receiver is unterminated ($\Gamma_r = 1$) then signals incident at the receiver get reflected back toward the source. However, these do not get re-reflected since $\Gamma_s = 0$. Thus the source and receiver voltages are:

$$v_s(t) = v_p(t) + v_p(t - 2T), \tag{16.16}$$
$$v_r(t) = 2v_p(t - T) = v_d(t - T). \tag{16.17}$$

Note how the reflection has restored the full amplitude of the driver voltage.

Method	Noise margin (L)	Noise margin (H)	Rise time	Fall time
Parallel pull-up	Degraded		Faster	Slower
Parallel pull-down		Degraded	Slower	Faster
Passive Thévenin			Mixed	Mixed
Active Thévenin			Mixed	Mixed

Table 16.3: Relative merits of different resistive receiver termination methods. Here mixed means that the rise/fall times are faster when switching toward V_t but slower otherwise.

111 Termination example

Micro SD cards are a type of high-density flash memory that can be interfaced to a microcontroller using SPI with a maximum clock speed of 25 MHz. Let's consider terminating the clock line for an AT91SAM7S256 microcontroller. A crude model of the microcontroller output is a 3.3 V voltage source in series with a 22 Ω output resistance R_o[15].

[15] This is only valid for voltages close to 3.3 V.

The simplest approach is to use parallel pull-down termination at the receiver, i.e., by the SD card. The resistor needs to be placed as close as possible to the clock input of the SD card; alternatively, the clock signal can be run past the SD card to the terminator (fly-by termination). To achieve no reflection, the termination resistance needs to match the characteristic impedance of the microstrip clock trace. Assuming microstrip traces with a characteristic impedance of 50 Ω, the termination resistance must be $R_t = 50\,\Omega$.

Let's now consider the effect of termination on the logic levels. The input low logic level for the SD card is $V_{\text{il}} = 0.8\,\text{V}$ and the input high logic level is $V_{\text{il}} = 2.0\,\text{V}$ (for $V_{\text{DD}} = 3.3\,\text{V}$). Using the simple model for the microcontroller output, the no load output high voltage is 3.3 V. When loaded by the termination resistor, the output high voltage will drop since the termination resistor forms a voltage divider[16] with the microcontroller output resistance. The output high voltage[17] is thus,

[16] For an accurate answer, the I-V curve for the CMOS output driver should be used.
[17] In the steady-state.

$$V_{\text{oh}} = \frac{R_t}{R_t + R_o} V_{\text{DD}} = \frac{50}{50 + 22} \times 3.3 = 2.3\,\text{V}. \qquad (16.18)$$

Thus the logic high noise margin is $2.3 - 2.0 = 0.3\,\text{V}$. Note, the logic low noise margin is unaffected by the termination and is 0.8 V.

If the characteristic impedance was 70 Ω, then the output high voltage is,

$$V_{\text{oh}} = \frac{70}{70 + 22} \times 3.3 = 2.5\,\text{V}. \qquad (16.19)$$

If parallel receiver termination with a pull-up resistor to 3.3 V is employed, then the logic high output level will be 3.3 V. However, the logic low output level will be increased from 0.0 V to

$$V_{\text{ol}} = \frac{R_o}{R_o + R_t} V_{\text{DD}} = \frac{22}{22 + 50} \times 3.3 = 1.0\,\text{V}. \qquad (16.20)$$

This is greater than the logic low input voltage of the micro SD card so the noise margin is negative!

1000 Further reading

"DDR SDRAM termination, Fairchild Semiconductor, 1999.

Termination techniques for high speed buses http://www.ednmag.com/reg/1998/021698/04df_04.htm

Differential termination (common-mode requirements) http://www.sigcon.com/Pubs/edn/DifferentialTermination.htm

Differential termination http://www.national.com/an/AN/AN-903.pdf

LVDS termination http://www.national.com/appinfo/lvds/0,1798,100,00.html

Termination techniques for high speed buses http://www.ednmag.com/reg/1998/021698/04df_04.htm

Tom Granberg, "Handbook of digital techniques for high-speed design", Prentice Hall, 2004. Excellent study of modern computer techniques.

Eric Bogatin, "Signal integrity—simplified", Prentice Hall, 2004. Light on math derivations but easy to read and excellent summary.

Mark Montrose, "EMC and the printed circuit board", IEEE Press, 1998. Good comparison.

Mark Montrose, "Printed circuit board design techniques for EMC compliance", IEEE Press, 1996.

Douglas Brooks, "Signal integrity issues and printed circuit board design", Prentice Hall, 2003. Chapter 10 has a nice simple explanation.

Stephen Thierauf, "High-speed circuit board signal integrity", Artech House 2004. More analytical.

1001 Exercises

1. Why is it difficult to terminate a bus with stubs?
2. What sort of termination would you recommend for battery operation?

3. What effect does parallel pull-up termination have on noise margins?
4. What is the static power dissipation for a passive Thévenin terminator for a 5 V system with characteristic impedance of 50 Ω?
5. Why does source and receiver termination not work for ordinary CMOS logic?

17
Differential signals

Most modern high-speed digital buses use differential signals, for example, LVDS, PCI Express, Ethernet, USB, SATA. There a number of advantages:

- Immunity to common-mode interference (such as crosstalk, provided signals close together)
- Twice the noise immunity of single-ended (better for low voltage systems)
- Tolerant of ground offsets

The differential signals need to be closely separated with a constant geometry to ensure a constant characteristic impedance and good EMC.

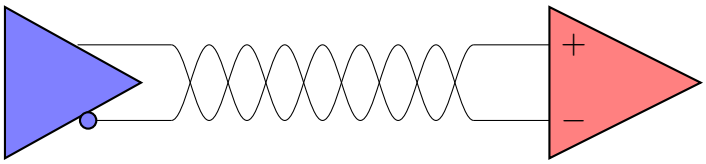


Figure 17.1: Differential serial bus.

1 *Differential and common-mode signals*

The output signals from a differential driver are of the form,

$$v_1(t) = \frac{V_{\mathrm{DD}}}{2} - v(t), \tag{17.1}$$

$$v_2(t) = \frac{V_{\mathrm{DD}}}{2} + v(t). \tag{17.2}$$

The differential signal[1] is the difference of the two signals

$$v_d(t) = v_2(t) - v_1(t) = 2v(t), \tag{17.3}$$

and the common-mode signal is the mean value

$$v_c(t) = 0.5\,(v_1(t) + v_2(t)) = \frac{V_{\mathrm{DD}}}{2}. \tag{17.4}$$

[1] Note, the amplitude of the differential signal is twice that of each of the singled-ended signals. This helps to increase the noise margin.

1.1 Common-mode rejection

If some common-mode interference is picked up, then it has no effect on the differential signal, *provided the input signals are within the operating range of the differential receiver*.

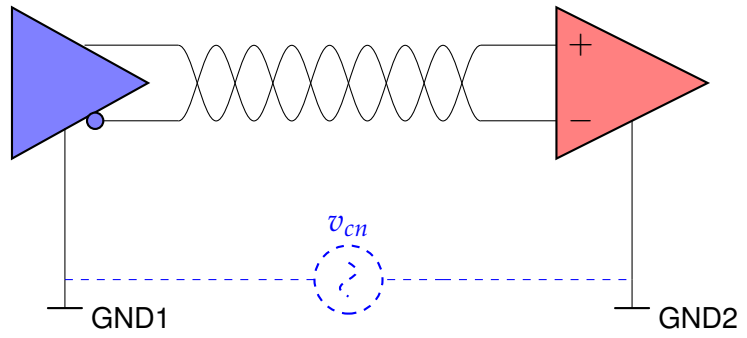


Figure 17.2: Common-mode interference.

1.10 Timing skew

The path lengths of the differential signals needs to be balanced to avoid timing skew between the signals, since any skew produces a common-mode signal. More importantly, unbalanced signals have poorer EMC since the fields are not equal and opposite.

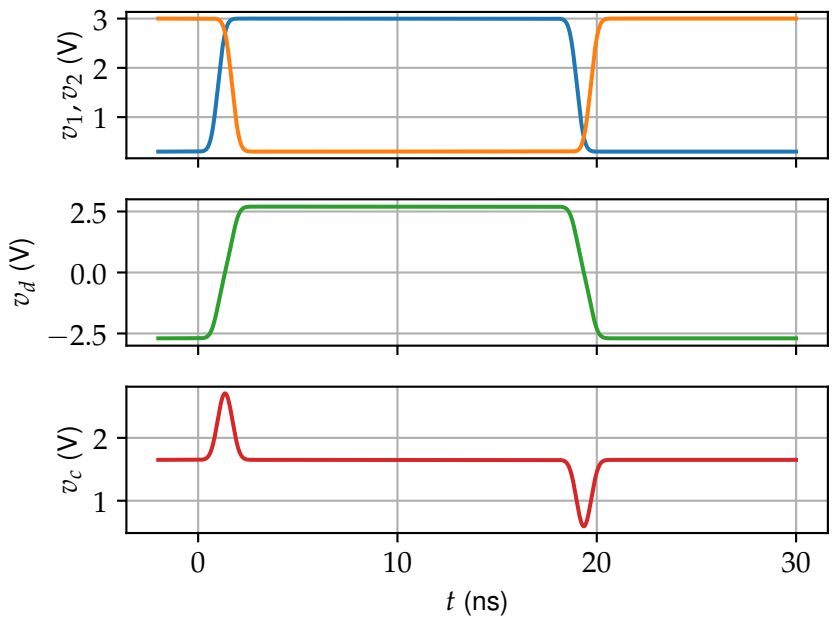


Figure 17.3: Differential signals with a skew giving rise to a changing common-mode signal $v_c(t) = 0.5(v_1(t) + v_2(t))$.

10 USB example

The USB common-mode voltage range is only 0.8–2.5 V and thus it requires an electrical shield to reduce common-mode interference[2]. USB uses a single twisted pair for bidirectional communication and is thus half-duplex. Series termination is used.

[2] USB sometimes uses external ferrite beads as common-mode chokes for the same purpose.

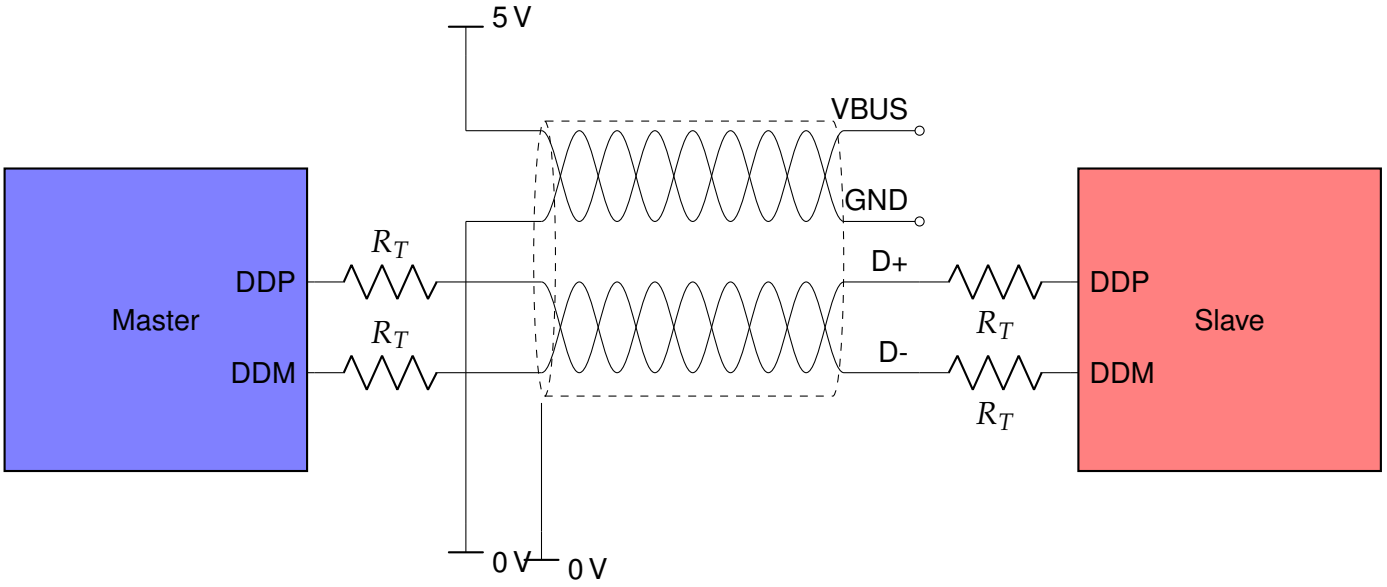


Figure 17.4: USB signals between master and slave.

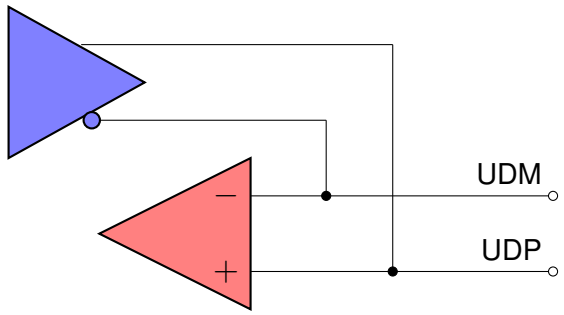


Figure 17.5: USB transceiver consisting of a differential driver and a differential receiver. The driver is tristated when receiving.

11 UTP Ethernet example

Ethernet[3] uses unshielded twisted pair (UTP) cables[4]. Separate pairs are used for full-duplex operation. Magnetics inside the connectors (a transformer and a common-mode choke) are used to reduce common-mode signals. The use of transformers make Ethernet immune to different ground potentials at different ends of a building.

[3] 10baseT, 100baseT, etc.
[4] CAT5, CAT6, etc.

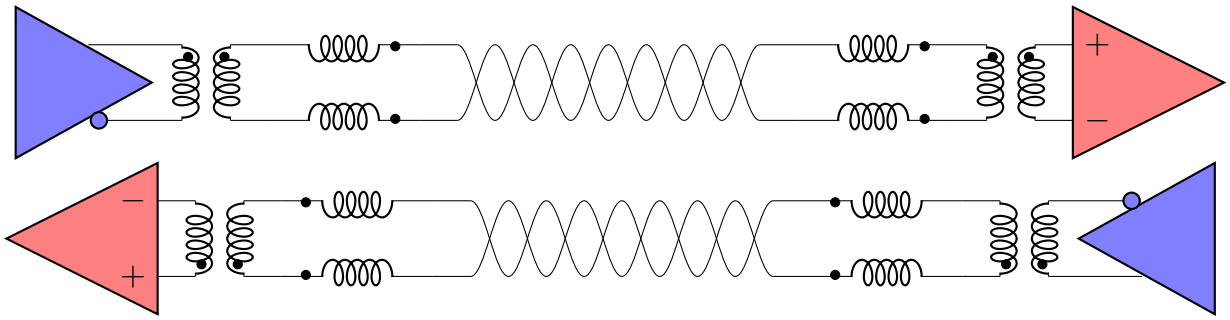


Figure 17.6: Ethernet over UTP. Note, separate circuits are required for full-duplex transmission. The passive Thévenin transmission resistors for each signal at the receiver are not shown.

100 Common-mode chokes

Common-mode chokes have two coils wound on a single core. They pass differential currents but block common-mode currents so are useful for blocking EMI radiated from cables attached to equipment, see Figure 17.8.

Figure 17.7: Common-mode ferrite choke.

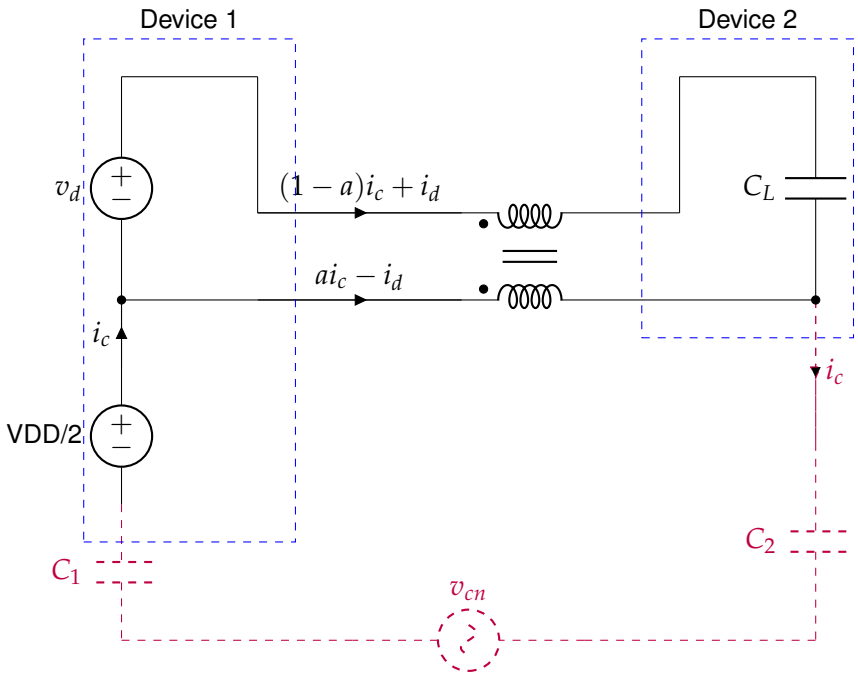


Figure 17.8: Use of a common-mode choke to limit the common-mode current i_c that is generated by the common-mode noise source v_{cn}. The parasitic capacitances C_1 and C_2 (or resistances) provide the current path. The desired differential current i_d passes through the common-mode choke unimpeded.

101 Termination

Differential traces should be independently terminated; this terminates both the common and differential-mode signals. However, with passive Thévenin termination (see Figure 17.9) this requires four resistors. While a single resistor between the traces is common (see Figure 17.11 as used by LVDS), this does not terminate the common-mode signal that can arise if there is any skew between the differential signals.

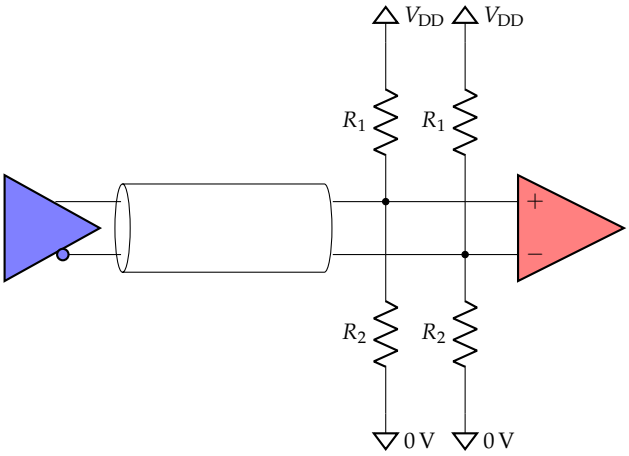


Figure 17.9: Passive Thévenin differential termination; $R_1 \parallel R_2 = Z_{odd}$.

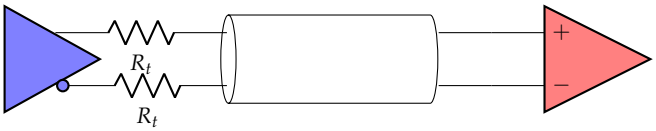


Figure 17.10: Differential series termination; $R_t = Z_{odd} - R_o$ where R_o is the driver output resistance.

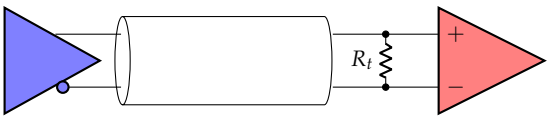


Figure 17.11: Simple differential termination as used by LVDS; $R_t = Z_{odd}$. This does not terminate common-mode signals since there is no path to the common ground.

110 Characteristic impedance

Differential signals need to be routed close together[5] but this changes the characteristic impedance due to coupling effects[6]. This coupling can be modelled using four characteristic impedances,

$$V_1 = Z_{11} I_1 + Z_{12} I_2, \quad (17.5)$$

$$V_2 = Z_{12} I_1 + Z_{22} I_2. \quad (17.6)$$

[5] Often as a twisted pair.

[6] The coupling impedance depends on the track geometry; edge coupled (microstrip or stripline) or broadside coupled (stripline).

If the transmission line is symmetrical, the characteristic impedance matrix is also symmetrical with the form

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \begin{bmatrix} Z_0 & Z_{\text{coupling}} \\ Z_{\text{coupling}} & Z_0 \end{bmatrix}, \quad (17.7)$$

where $Z_{\text{coupling}} = Z_{12} = Z_{21}$ is the coupling impedance and Z_0 is the characteristic impedance of each signal path if the other signal path was not present.

There are online calculators[7] that determine the differential characteristic impedances given the geometry.

[7] https://resources.altium.com/p/differential-microstrip-impedance-calc

111 Odd and even characteristic impedance

When a purely differential signal propagates along the differential transmission line, $I_2 = -I_1$, then

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2, && (17.8)\\ &= Z_{11}I_1 - Z_{12}I_1, && (17.9)\\ &= (Z_{11} - Z_{12})\, I_1, && (17.10)\\ &= Z_{\text{odd}}I_1. && (17.11) \end{aligned}$$

Thus the characteristic impedance for the differential signal is

$$Z_{\text{odd}} = (Z_{11} - Z_{12}). \quad (17.12)$$

This is called Z_{odd} due to the odd symmetry. Each of the two signals needs to be terminated with this impedance.

The differential termination impedance is defined as

$$\begin{aligned} Z_{0d} &= 2Z_{\text{odd}}, && (17.13)\\ &= 2(Z_{11} - Z_{12}). && (17.14) \end{aligned}$$

When a purely common-mode signal propagates along the differential transmission line, $I_2 = I_1$, then

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2, && (17.15)\\ &= Z_{11}I_1 + Z_{12}I_1, && (17.16)\\ &= (Z_{11} + Z_{12})\, I_1, && (17.17)\\ &= Z_{\text{even}}I_1. && (17.18) \end{aligned}$$

The characteristic impedance for this even symmetry case is

$$Z_{\text{even}} = (Z_{11} + Z_{12}). \quad (17.19)$$

The common-mode termination impedance is defined as

$$\begin{aligned} Z_{0c} &= \frac{1}{2}Z_{\text{even}}, && (17.20)\\ &= 2(Z_{11} - Z_{12}). && (17.21) \end{aligned}$$

For an example, consider USB 2.0. This specifies a differential characteristic impedance Z_{0c} =80–100 ohm and a common-mode characteristic impedance Z_{0d} =21–39 ohm. Let's assume $Z_{0c} = 90$ ohm and $Z_{0d} = 30$ ohm. Thus the even and odd characteristic impedances from (17.21) and (17.14) are

$$Z_{\text{odd}} = \frac{Z_{0d}}{2} = 45\,\text{ohm}, \tag{17.22}$$

$$Z_{\text{even}} = 2Z_{0c} = 60\,\text{ohm}. \tag{17.23}$$

Combining (17.11) and (17.18) gives

$$Z_{11} = \frac{Z_{\text{even}} + Z_{\text{odd}}}{2} = 52.5\,\text{ohm}, \tag{17.24}$$

$$Z_{12} = \frac{Z_{\text{even}} - Z_{\text{odd}}}{2} = 7.5\,\text{ohm}. \tag{17.25}$$

What this means is that each microstrip of the pair (when considered in isolation) should have a characteristic impedance of 52.2 ohm.

1000 Summary

1. Differential signalling is useful for low voltage and high speed systems.
2. Differential signalling has better EMC provided the signals are closely spaced, with a uniform geometry and equal length.
3. An extra connection is required (this takes up more PCB area).
4. The resistance required to terminate the differential-mode for each signal is lower than for an equivalent single-ended system due to coupling of the fields.

18
Clock generation

Microcontrollers require a clock to regulate their operation and to provide timing references.

1 Oscillators

An oscillator can be constructed from an amplifier with positive feedback, provided there is 360 degrees of phase shift at a frequency where the gain is greater than unity. To obtain an accurate frequency requires narrow bandwidth; this is difficult to achieve using inductors and capacitors due to the winding resistance of the inductors. However, a quartz crystal has little vibrational losses and thus the resonance has a very high Q.

1.1 Crystals

Crystals are constructed from quartz, a piezoelectric material. A voltage applied across the device produces mechanical deformation and the resulting stress wave propagates across the crystal and is reflected from boundaries. The reflected waves change the deformation and thus the voltage across the crystal. When the driving voltage has a frequency equal to the mechanical resonant frequency of the crystal, the electrical impedance is a minimum.

A quartz crystal can be modelled as a number of series R-L-C circuits[1] one for each mechanical resonance, in parallel with the electrical capacitance of the crystal, C_0, see Figure 18.1. The lowest frequency mode is the *fundamental mode*. There are also resonances at odd multiples of the fundamental frequency, called *overtones*, and other spurious resonances due to unwanted modes of vibration.

For microcontroller clocks, usually the lowest resonance frequency (the fundamental mode) is used[2].

[1] The inductance is proportional to the crystal mass, the capacitance is proportional to the crystal stiffness (elasticity), and the mass is proportional to the mechanical damping. The scale factor is the piezoelectric constant, newtons per volt.

[2] It is difficult to have a fundamental mode above 30 MHz, since the crystal has to be thin and is thus fragile. For high frequency applications, the oscillator is tuned to an overtone frequency of the crystal.

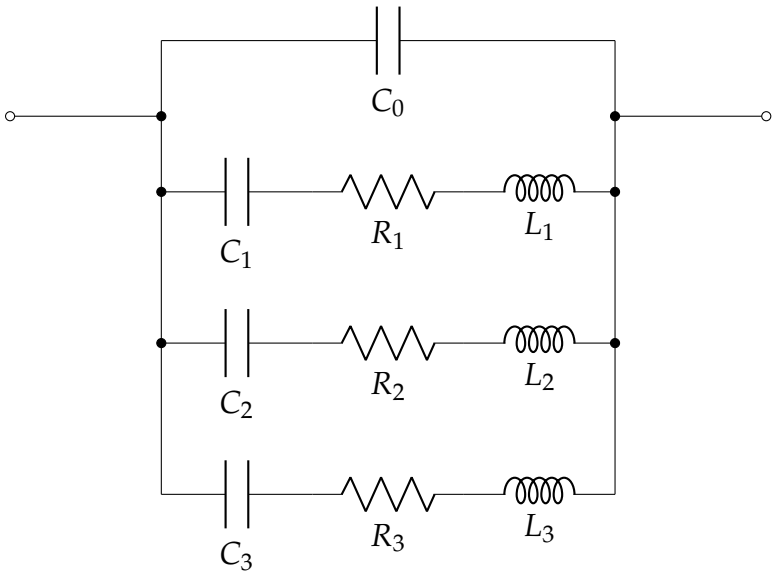


Figure 18.1: Crystal electrical model. C_n, L_n, and R_n model the motional elements of the crystal for the n^{th} resonance.

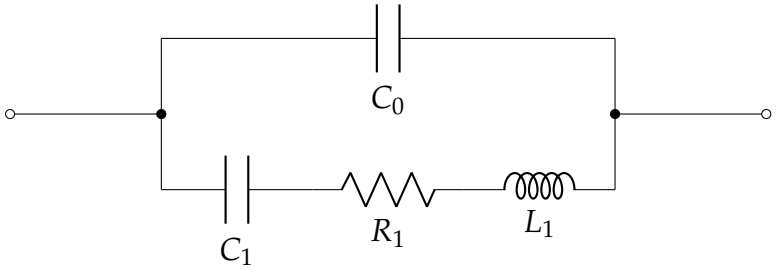


Figure 18.2: Simplified crystal electrical model. C_1, L_1, and R_1 model the motional elements of the crystal for one resonant mode, example, the fundamental mode.

$$Z(s) = \left(\frac{1}{sC_1} + sL_1 + R_1\right) \parallel \frac{1}{sC_0}, \tag{18.1}$$

$$= \frac{1}{sC_0}\frac{s^2 + s\frac{R_1}{L_1} + \omega_s^2}{s^2 + s\frac{R_1}{L_1} + \omega_p^2}, \tag{18.2}$$

where the series resonant angular frequency is

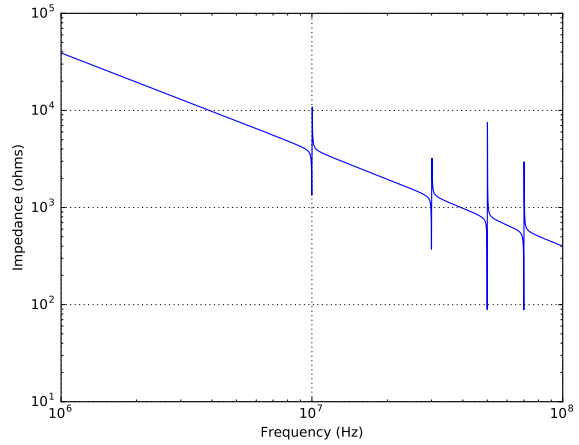
$$\omega_s = \frac{1}{\sqrt{L_1 C_1}}, \tag{18.3}$$

and the parallel resonant[3] angular frequency is

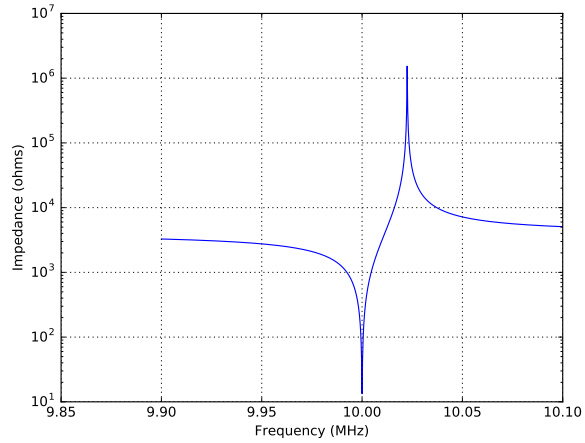
$$\omega_p = \sqrt{\frac{C_1 + C_0}{L_1 C_1 C_0}} = \omega_s \sqrt{1 + \frac{C_1}{C_0}}. \tag{18.4}$$

[3] Also called the anti-resonant frequency.

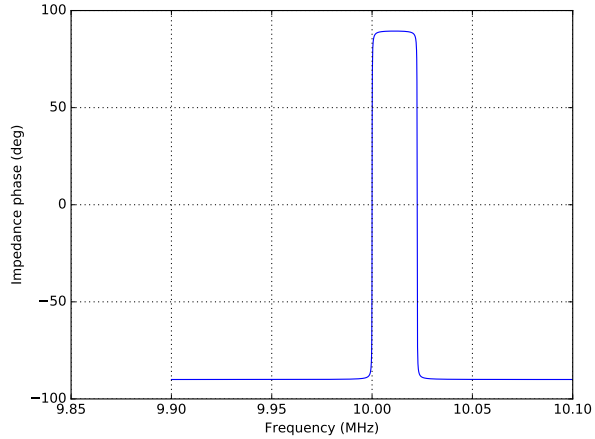
The series resonance is a few kilohertz below the parallel resonance. The impedance of the crystal is a local minimum at the series resonance and a local maximum at the parallel resonance.



(a)



(b)



(c)

Figure 18.3: Crystal impedance (magnitude and phase). $C_0 = 4\,\text{pF}$, $C_1 = 0.018\,\text{pF}$, $R_1 = 20\,\Omega$, $L_1 = 14.1\,\text{mH}$. Series resonant frequency $f_s = 10.0\,\text{MHz}$, parallel resonant frequency $f_p = 10.022\,\text{MHz}$. (a) showing overtones but ignoring spurious modes, (b) magnitude of impedance around fundamental mode, (c) phase of impedance around fundamental mode.

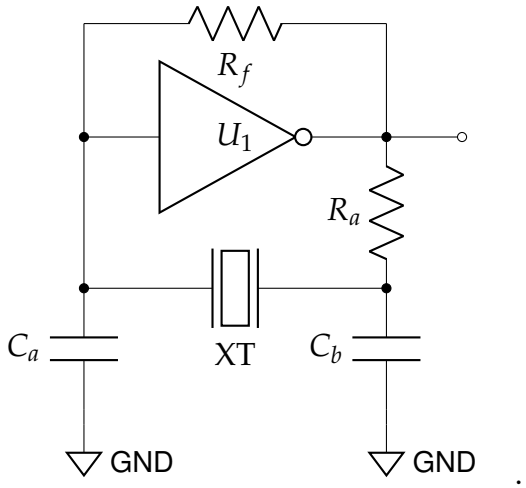


Figure 18.4: The gated Pierce oscillator circuit. R_f is a feedback resistor to ensure the CMOS inverter is biased in the linear region. R_a is a current limiting resistor to prevent over-heating of the crystal.

10 The Pierce oscillator

Microcontroller clocks are usually generated using a Pierce oscillator with a fundamental mode crystal[4]. in the feedback loop, as shown in Figure 18.4. The gain is provided using a CMOS inverter biased in its linear region by connecting a feedback resistor[5] between the input and output. The inverter provides 180 degrees of phase shift so the feedback network needs to provide another 180 degrees. There are two means of obtaining the 180 degrees phase shift with the feedback circuit:

Series resonance Here the crystal is operated at its series resonance where it acts like a resistor. The lowpass filter comprised of the crystal and the load capacitor C_b provides 90 degrees phase shift[6]. The other 90 degrees is provided by the output resistance of the inverter, the output capacitance of the inverter, R_a, and C_a.

Parallel resonance Here the crystal is operated between its series resonance but below its parallel resonance[7] where it acts like an inductor. The lowpass filter comprised of the crystal and the load capacitor C_b provides 180 degrees phase shift. The other components of the feedback circuit provide little phase shift (and little attenuation).

Crystal manufacturers specify the frequency of their crystals for a given load capacitance[8]. This is given by

$$C_L = \frac{(C_a + C_i) \times (C_b + C_o)}{C_a + C_i + C_b + C_o} + C_s, \qquad (18.5)$$

where C_i is the inverter input capacitance, C_o is the inverter output capacitance, and C_s is stray capacitance[9]

[4] To use an overtone crystal requires addition inductors and capacitors to select the desired overtone.

[5] Typically of the order of 1–10 MΩ.

[6] Due to propagation delays the inverter provides more than 180 degrees of phase shift and so the crystal does not operate at the series resonant frequency (where it is purely resistive) but slightly above (where it is inductive).

[7] The operating region above series resonance is called the the *area of usual parallel resonance*.

[8] For a series resonance crystal this is zero. For a parallel resonance crystal the specified frequency is above series resonance.

[9] The stray capacitance is typically 2–5 pF and depends on the PCB layout.

across the crystal pins. Typically the values of C_a and C_b are equal. Larger values make the operating frequency approach the series resonant frequency and provide increased frequency stability. However, the loop gain is reduced and so the oscillator may not start up[10].

Crystals have a maximum operating power so resistor R_a is sometimes needed for low frequency crystals to limit the current.

[10] The oscillator is started by voltage transients or thermal noise within the oscillator passband. If the gain is small it may take a long time for the oscillator to start.

10.1 Ceramic resonators

Ceramic resonators are similar to crystals but are made from a piezoelectric ceramic material[11]. They can be modelled as a voltage dependent capacitor. Quartz crystals have a 0.001% frequency tolerance, while PZT has a 0.5% frequency tolerance.

[11] PZT

10.10 RC oscillator

An RC oscillator circuit is shown Figure 18.5. This is often incorporated into a microcontroller but does not have the frequency precision of a crystal. It is also more temperature dependent.

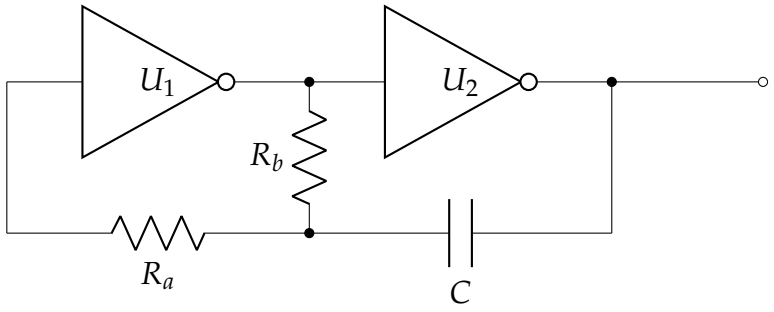


Figure 18.5: An RC oscillator circuit.

11 PLL frequency conversion

A phase-locked loop consists of a voltage controlled oscillator (VCO), a phase comparator, and a filter. The filtered output of the phase comparator is used to set the frequency of the VCO.

$$f_o = f_i \frac{M}{N} \frac{1}{P} \tag{18.6}$$

where M is the PLL multiplier, N is the PLL divider, and P is the post PLL divider; all these are integers. There is a caveat; the VCO frequency Mf_i must be within a specific range.

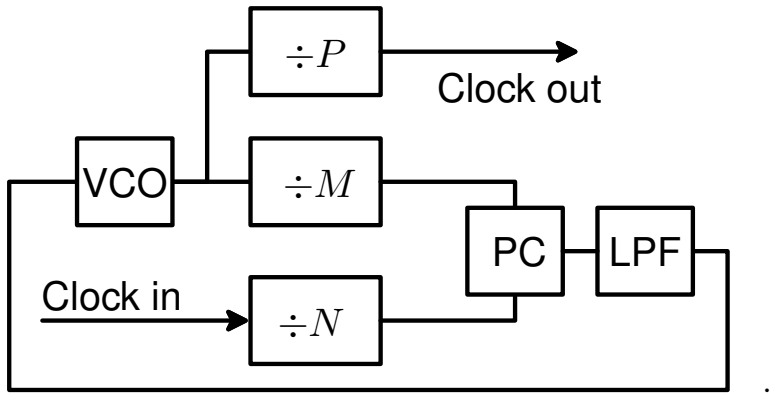


Figure 18.6: Frequency conversion with a phase locked loop (PLL). VCO is a a voltage controlled oscillator, PC is a phase comparator.

11.1 AT91SAM4S example

The master clock frequency is given by

$$f_{\text{MCK}} = \frac{f_{\text{PLL}}}{P} \tag{18.7}$$

where P is a power of two prescaler integer and f_{PLL} is the PLL frequency given by

$$f_{\text{PLL}} = f_{\text{XTAL}} \times \frac{\text{PLL\_MUL}}{\text{PLL\_DIV}}. \tag{18.8}$$

Here f_{XTAL} is the crystal frequency, PLL_MUL is the PLL multiplier, and PLL_DIV is the PLL divisor.

For example, if the desired master clock frequency is 96 MHz and the crystal frequency is 12 MHz, then

$$\frac{f_{\text{MCK}}}{f_{\text{XTAL}}} = \frac{\text{PLL\_MUL}}{P \times \text{PLL\_DIV}} = 8. \tag{18.9}$$

One choice[12] is $P = 2$, PLL_MUL $= 16$, and PLL_DIV $= 1$.

[12] This satisfies the SAM4S constraint 80 MHz $< f_{\text{PLL}} <$ 240 MHz.

100 Spread spectrum operation

Some clock sources will jitter their frequency. The purpose is to spread interference across the spectrum so that it is under the regulatory requirements for EMC.

101 Clock generator chips

Special ICs can generate multiple clocks all referenced to a single oscillator. This is useful for multimedia applications where the video and audio are sampled at different rates. The outputs are usually differential to reduce jitter caused by common-mode noise.

110 Crystal analysis

The impedance of a crystal is

$$Z(\omega) = \frac{1}{\mathrm{j}\omega C_0} \frac{\omega_s^2 - \omega^2 + \mathrm{j}\omega\frac{R_1}{L_1}}{\omega_p^2 - \omega^2 + \mathrm{j}\omega\frac{R_1}{L_1}}. \tag{18.10}$$

When $f \gg f_s$,

$$Z(\omega) \approx \frac{1}{\mathrm{j}\omega C_0}. \tag{18.11}$$

When $f \ll f_s$,

$$Z(\omega) \approx \frac{1}{\mathrm{j}\omega\left(C_0 + C_1\right)}, \tag{18.12}$$

but since $C_1 \ll C_0$, then

$$Z(\omega) \approx \frac{1}{\mathrm{j}\omega C_0}. \tag{18.13}$$

At the series resonance the reactance of L_1 and C_1 cancel and so the impedance of the crystal is

$$\begin{aligned} Z(\omega_s) &= \frac{1}{\mathrm{j}\omega_s C_0} \parallel R_1, & (18.14)\\ &= \frac{R_1}{1 + \left(\omega_s C_0 R_1\right)^2} - \mathrm{j}\frac{\omega_s C_0 R_1}{1 + \left(\omega_s C_0 R_1\right)^2}, & (18.15)\\ &= \frac{R_1}{1 + \frac{\left(C_0 R_1\right)^2}{L_1 C_1}} - \mathrm{j}\frac{\frac{C_0 R_1}{\sqrt{L_1 C_1}}}{1 + \frac{\left(C_0 R_1\right)^2}{L_1 C_1}}. & (18.16) \end{aligned}$$

Since the reactance is negative then the impedance at series resonance is slightly capacitive. However, the real component is much larger than the imaginary component and so the phase is close to zero. The real part of the impedance is the equivalent series resistance (ESR) (not to be confused with the motional resistance R_1),

$$\mathrm{ESR} = \frac{R_1}{1 + \frac{R_1^2 C_0^2}{L_1 C_1}}. \tag{18.17}$$

From this,

$$R_1^2 - \frac{L_1 C_1}{C_0^2 \mathrm{ESR}} + \frac{L_1 C_1}{C_0^2} = 0, \tag{18.18}$$

with a solution

$$R_1 = \frac{L_1 C_1}{2 C_0^2 \mathrm{ESR}} - \sqrt{\left(\frac{L_1 C_1}{2 C_0^2 \mathrm{ESR}}\right)^2 - \frac{L_1 C_1}{C_0^2}}. \tag{18.19}$$

However, since $R_1^2 C_0^2 \ll L_1 C_1$ so ESR $\approx R_1$.

The Effective Resistance is given by

$$R_e = R_1 \left(\frac{C_L + C_0}{C_L} \right)^2. \tag{18.20}$$

The Q[13] of a crystal is given by

$$Q = \frac{X_{L_1}}{R_1} = \frac{1}{X_{C_1} R_1}, \tag{18.21}$$

where X_{L_1} is the reactance of L_1 and where X_{C_1} is the reactance of C_1 at the operating frequency.

[13] This ranges from 20,000 to 200,000 for a standard crystal. In comparison, the Q of a good L C tuned circuit is in the order of 200. This high Q makes crystal oscillators stable.

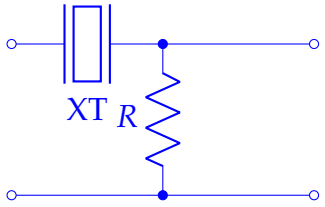


Figure 18.7: A crystal configured as a bandpass filter.

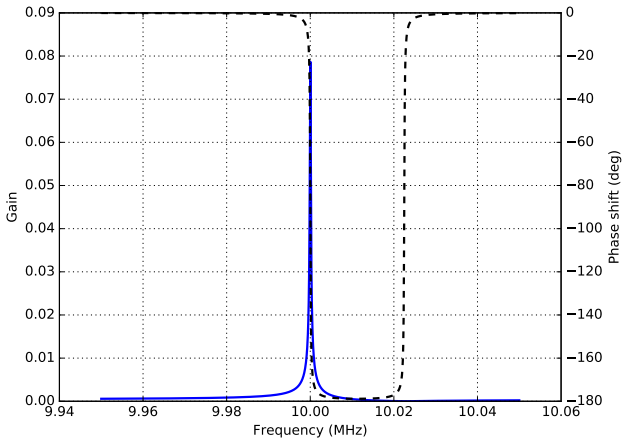


Figure 18.8: Frequency response of crystal bandpass filter. $C_0 = 4\,\text{pF}$, $C_1 = 0.018\,\text{pF}$, $R_1 = 20\,\Omega$, $L_1 = 14.1\,\text{mH}$, $C_b = 10\,\text{nF}$.

111 *Further reading*

http://services.eng.uts.edu.au/pmcl/de/Downloads/Lecture04.pdf has an excellent design example and explains the theory well.

http://ww1.microchip.com/downloads/en/appnotes/00826a.pdf Microchip application note AN826. This assumes that there is a phase shift due to load capacitor C_a.

1000 Summary

- Crystals are made from quartz; a piezoelectric material.
- A crystal oscillator uses a crystal in the feedback circuit to obtain a very high Q bandpass response.
- Crystals are resonant for different vibration modes. Each mode is also resonant at overtones (odd harmonics of the fundamental).
- Microcontroller clocks use the Pierce oscillator; this uses the fundamental mode of the crystal and so is limited to about 30 MHz.
- The Pierce oscillator can be configured for series or parallel resonance.
- The correct load capacitance must be applied to the crystal to obtain the correct frequency and ensure that the circuit oscillates.
- The frequency of a crystal oscillator drifts with temperature and age.
- Ceramic resonators are similar to crystals but use a piezoelectric ceramic. They are cheaper but have a lower Q.
- Phase locked loops (PLL) can be configured to generate a higher frequency clock.
- The jitter of a PLL output is greater than the jitter of the reference clock.

1001 Exercises

1. The SAM4S PLL has a VCO with a frequency range of 80–240 MHz; a multiplier, M, with a range 1–81, a divider, N, with a range 1–255, and a prescaler $P = 1, 2, 4, 8, 16, 32, 64$. If the input clock is 12 MHz and the desired output from the prescaler is 96 MHz, determine suitable values for M, N, and P.
2. Repeat the above exercise for an input clock of 16 MHz.
3. Repeat the above exercise for an input clock of 32 kHz.
4. What is the advantage of a crystal oscillator over an RC oscillator?

5. Most MCUs have an internal oscillator that requires no external parts. What type is it?
6. Why is the jitter of a PLL output greater than the jitter of the input clock.
7. How can the stability of a crystal oscillator be improved?

19
IV curves supplement

An electrical one-port network can be described by current voltage (IV) characteristic curves. There are two variants depending on whether the network acts as a source or load, see Figure 19.1, defined by the current direction.

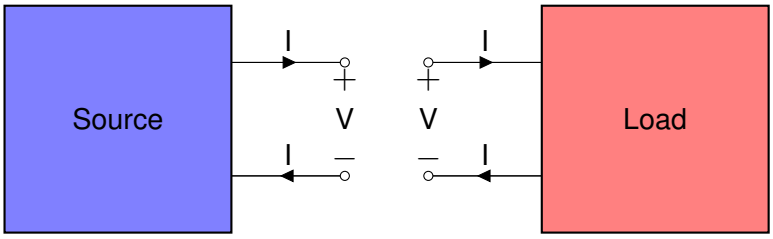


Figure 19.1: Source and load one-port networks. Note, the direction of current.

1 Source characteristic curves

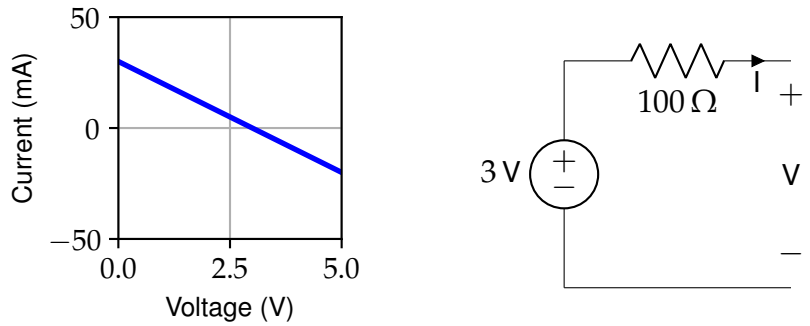


Figure 19.2: Thévenin equivalent source circuit and its IV characteristic curve.

Consider the Thévenin equivalent circuit shown in Figure 19.2 and its IV curve. The IV curve is defined by the open-circuit voltage ($V_{oc} = 3\,\text{V}$) and the short-circuit current ($I_{sc} = 3/100 = 30\,\text{mA}$). The slope of the line is the inverse of resistance. This IV curve indicates that if the output voltage, V, is set to 1 V, then the current, I, is

20 mA. Note if a current, I, is -20 mA (i.e., it flows into the voltage source) then the voltage, V, is 5 V.

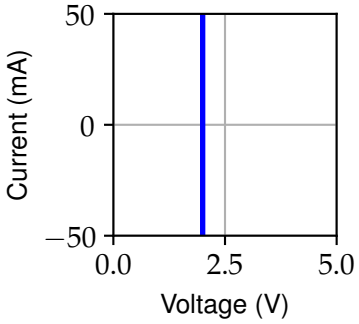
Figure 19.3 shows some example characteristic curves for a variety of sources.

10 Load characteristic curves (load lines)

Figure 19.4 shows some example characteristic curves for a variety of loads.

11 Characteristic curve superposition

Consider a Thévenin equivalent source circuit connected to a resistor load as shown in Figure 19.5. The load provides a constraint between V and I and the resulting values can be found from the intersection of the characteristic curves. This is a graphical approach to solving a simultaneous equation.



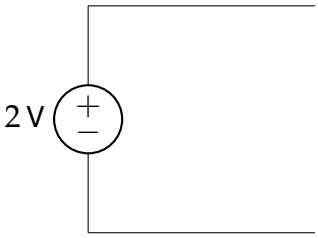
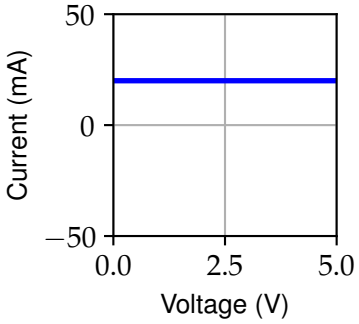
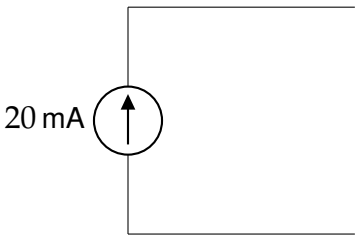
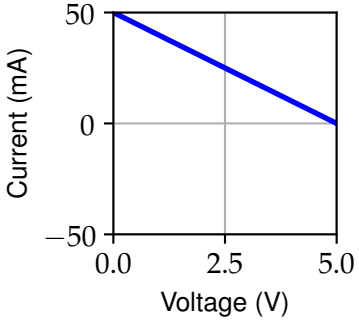
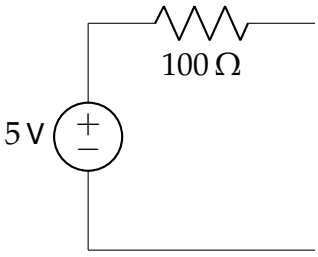


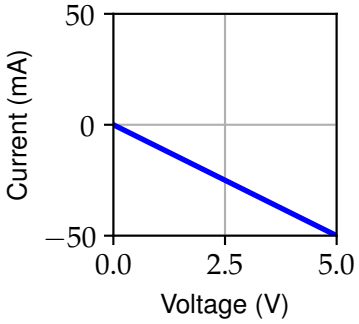
Figure 19.3: IV curves for ideal sources. The weird circuit with the ideal diode is a current-limiting voltage source equivalent to a voltage-limiting current source.

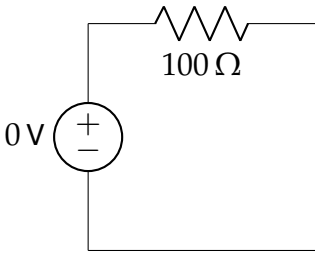


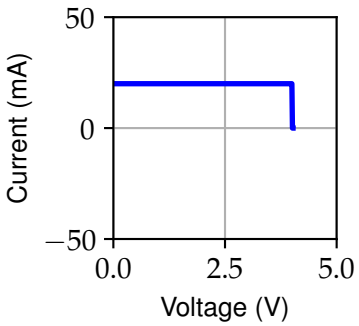


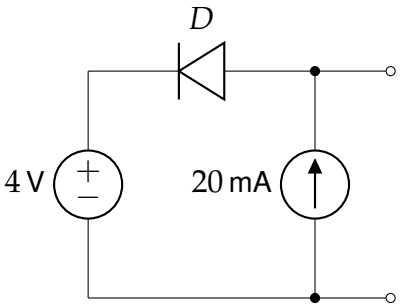












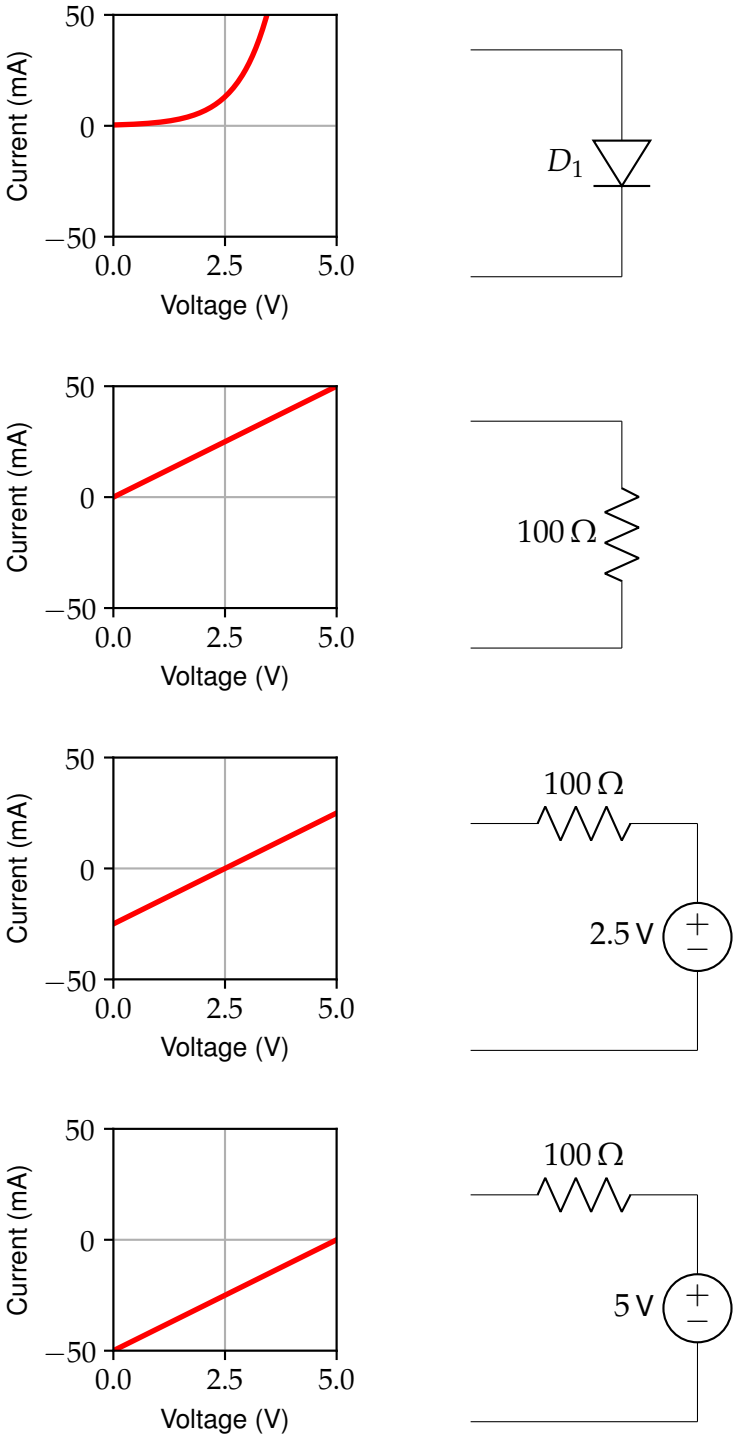
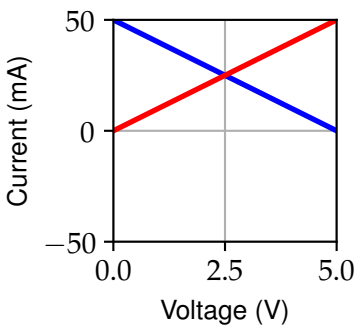


Figure 19.4: IV curves for loads.



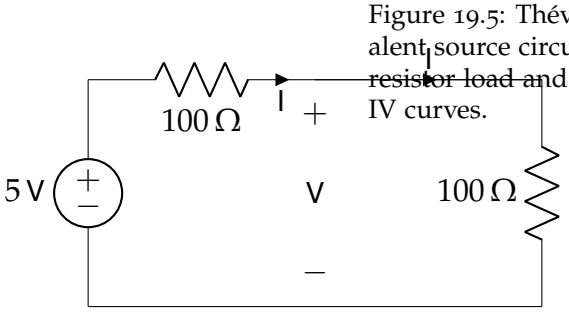
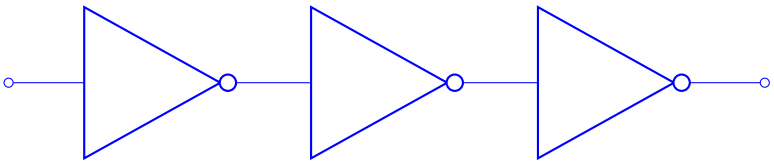


Figure 19.5: Thévenin equivalent source circuit with a resistor load and the associated IV curves.

20
CMOS supplement

1 Buffered CMOS



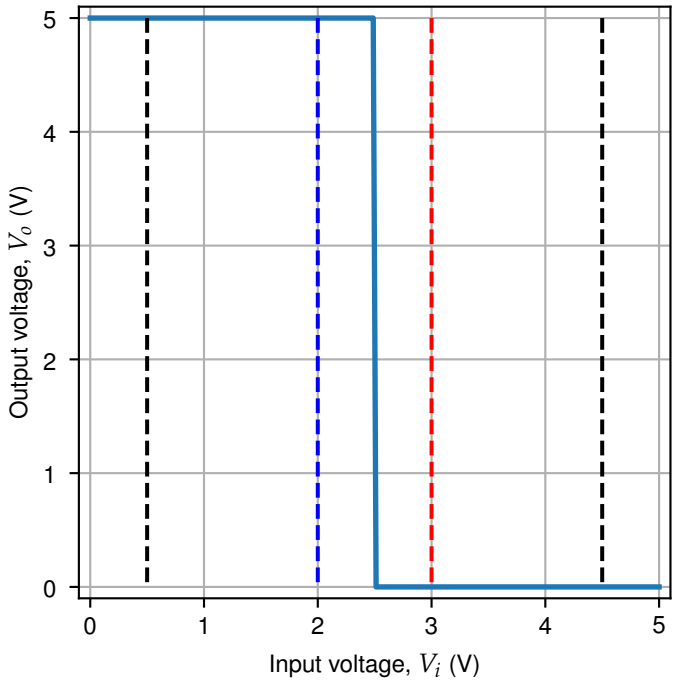


Figure 20.1: Buffered CMOS inverter. In practice, the input inverters can use smaller MOSFETs than the output inverter since they have less capacitive loading and thus require less current drive.
Figure 20.2: Transfer characteristic for three CMOS inverters in series.

Chips can be purchased with one or more CMOS inverters. They come in two flavours: unbuffered and buffered. A buffered CMOS inverter has a faster transition to its voltage transfer function; it is a bit like chaining three unbuffered inverters together (Figure 20.1). The overall transfer function is closer to the transfer function of an ideal inverter, see Figure 20.2.

Characteristics	Buffered	Unbuffered
Propagation delay	Slow†	Fast
Noise immunity/margin	Excellent	Good
Output impedance	Constant	Variable
Output transition time	Constant	Variable
AC gain	High	Low
Output oscillation for slow inputs	Yes	No
Input capacitance	Low	High

Table 20.1: Comparison between buffered and unbuffered logic. † I have seen specifications where the buffered is faster than unbuffered.

10 CMOS inverter analysis

The following is for those with a mathematical bent who are curious as to how a CMOS inverter can be modelled.

10.1 MOSFETs

The drain current, I_d, for an NMOS transistor is

$$I_d = \begin{cases} 0 & V_{gs} < V_{tn} \quad \text{(cutoff region)} \\ K\left(2\left(V_{gs} - V_{tn}\right)V_{ds} - V_{ds}^2\right) & V_{gs} \geq V_{tn} \text{ and } V_{ds} < V_{gs} - V_{tn} \quad \text{(linear region)} \\ K\left(V_{gs} - V_{tn}\right)^2 & V_{gs} \geq V_{tn} \text{ and } V_{ds} \geq V_{gs} - V_{tn} \quad \text{(saturation region)} \end{cases} \tag{20.1}$$

and for a PMOS transistor it is

$$I_d = \begin{cases} 0 & V_{gs} > V_{tp} \quad \text{(cutoff region)} \\ K\left(2\left(V_{gs} - V_{tp}\right)V_{ds} - V_{ds}^2\right) & V_{gs} \leq V_{tp} \text{ and } V_{ds} > V_{gs} - V_{tp} \quad \text{(linear region)} \\ K\left(V_{gs} - V_{tp}\right)^2 & V_{gs} \leq V_{tp} \text{ and } V_{ds} \leq V_{gs} - V_{tp} \quad \text{(saturation region)} \end{cases} \tag{20.2}$$

Note, for a PMOS transistor the threshold voltage V_{tp} is negative and V_{ds} and V_{gs} are negative for positive I_d.

10.10 Voltage transfer function

For a CMOS inverter the parameters of the NMOS and PMOS transistors are typically chosen so that $K = K_N \approx K_P$ and $V_t = V_{tn} \approx -V_{tp}$. With this assumption and equating the drain currents for the NMOS and PMOS transistors, (20.1) and (20.2), yields the transfer function:

$$V_o = \begin{cases} V_{DD} & 0 < V_t \leq V_t \\ V_i + V_t + \sqrt{V_{DD}^2 - 2V_{DD}V_i - 2V_{DD}V_t + 4V_iV_t} & V_t < V_i \leq \frac{V_{DD}}{2} \\ V_i - V_t - \sqrt{-V_{DD}^2 + 2V_{DD}V_i + 2V_{DD}V_t - 4V_iV_t} & \frac{V_{DD}}{2} < V_i \leq V_{DD} - V_t \\ 0 & 0 < V_{DD} - V_t < V_i \leq V_{DD} \end{cases} \tag{20.3}$$

Things to note:

1. The transfer function shape does not depend on K.
2. Input buffers uses small MOSFETs (with low K) whereas output buffers use grunty MOSFETs (with large K). Thus the short-circuit current for an input buffer is much less than that of an output buffer.
3. Have static power consumption if $V_i > V_t$ or $V_i < V_{\text{DD}} - V_t$ since both transistors are turned on[1].
4. Have dynamic power consumption when input voltage transitioning (both transistors on) and for charging/discharging capacitive loads.

[1] Sub-threshold current.

10.11 Output current versus voltage curve

Consider the case when the CMOS input voltage is low, i.e., $V_i = 0$. For the PMOS transistor, $V_s = V_{\text{DD}}$, $V_d = V_o$, and $V_g = 0$. Thus $V_{\text{gs}} = -V_{\text{DD}}$ and $V_{\text{ds}} = V_o - V_{\text{DD}}$, so the transistor is in either the saturation or linear regions. For the NMOS transistor, $V_s = 0$, $V_d = V_o$, and $V_g = 0$. Thus $V_{\text{gs}} = 0$ and $V_{\text{ds}} = V_o$ and so the transistor is in the cut-off region and thus passes no current.

When the CMOS input voltage is high, i.e., $V_i = V_{\text{DD}}$. Thus for the PMOS transistor $V_{\text{gs}} = 0$ and $V_{\text{ds}} = V_o - V_{\text{DD}}$, so the transistor is in the cutoff region and passes no current. For the NMOS transistor, $V_{\text{gs}} = V_{\text{DD}}$ and $V_{\text{ds}} = V_o$ and so the transistor is in either the saturation or linear regions.

10.100 Logic levels

Assuming symmetry of the CMOS inverter transfer function, the logic levels are chosen using

$$V_{\text{il}} = V_M - \Delta V, \tag{20.4}$$
$$V_{\text{ih}} = V_M + \Delta V, \tag{20.5}$$

where ΔV is selected to maximise the noise margins,

$$\Delta V_l = V_{\text{il}} - V_{\text{ol}}, \tag{20.6}$$
$$\Delta V_h = V_{\text{oh}} - V_{\text{ih}}. \tag{20.7}$$

For example, the logic low noise margin is

$$\Delta V_l = V_M - \Delta V - V_o(V_M + \Delta V). \tag{20.8}$$

Differentiating with respect to ΔV, gives

$$\frac{\mathrm{d}\Delta V_l}{\mathrm{d}\Delta V} = -1 - \frac{\mathrm{d}V_o(V_M + \Delta V)}{\mathrm{d}\Delta V}. \tag{20.9}$$

Thus the optimum value for ΔV occurs when the gain is -1. This occurs when

$$V_{\text{il}} = \frac{1}{8}(3V_{\text{DD}} + 2V_t), \tag{20.10}$$

$$V_{\text{ih}} = \frac{1}{8}(5V_{\text{DD}} - 2V_t). \tag{20.11}$$

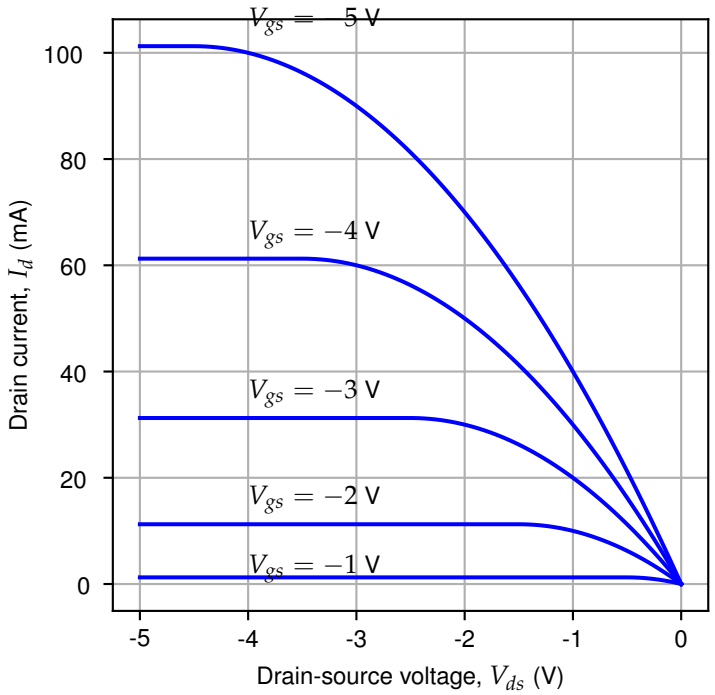


Figure 20.3: PMOS transistor characteristic curves.

11 Logic level errors

Ignoring propagation delays, the receiver input voltage, $v_r(t)$, is a scaled version of the source voltage, $v_s(t)$, corrupted by additive noise and crosstalk, $n(t)$,

$$v_r(t) = av_s(t) + n(t). \tag{20.12}$$

The receiver is usually modelled as a simple threshold detector. If the input voltage is greater than V_{ih} it is considered logic-high; the input voltage is less than V_{il} it is considered logic-low; otherwise the result is unreliable.

Let's assume that the mean receiver voltage μ corresponds to a logic-high, i.e., $\mu > V_{\text{ih}}$. The probability that the receiver does not reliably detect a logic-high can be determined by the probability that the receiver voltage is below the logic-high threshold voltage, V_{ih}, (the red area in Figure 20.6 and Figure 20.7):

$$P(V < V_{\text{ih}}) = F_V(V_{\text{ih}}) = \int_{-\infty}^{V_{\text{ih}}} f_V(v)\mathrm{d}v, \tag{20.13}$$

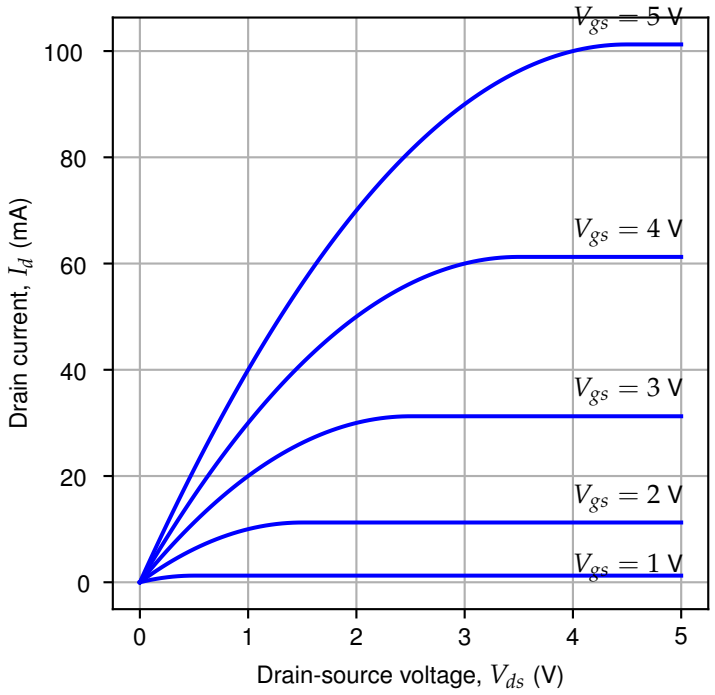


Figure 20.4: NMOS transistor characteristic curves.

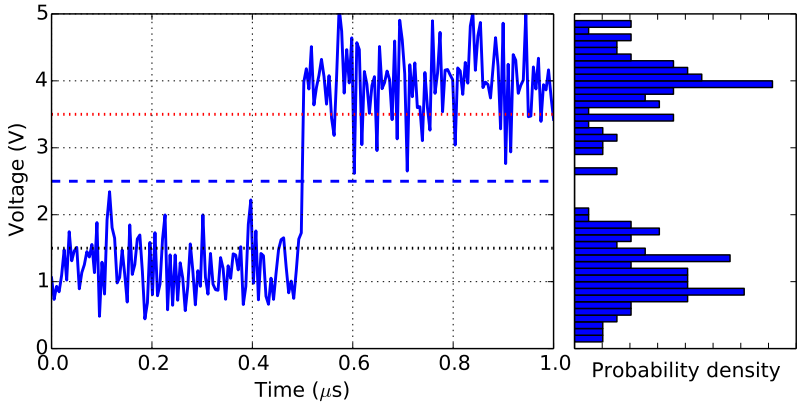


Figure 20.5: Digital waveform corrupted by noise.

where $F_V(V_{\text{ih}})$ is the cumulative distribution function (CDF) evaluated at $v = V_{\text{ih}}$ and $f_V(v)$ is the probability distribution function (PDF).

For simplicity, the noise and crosstalk is modelled with a zero-mean Gaussian distribution of variance σ^2. Thus for a mean receiver voltage μ, the PDF is

$$f_V(v) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(v-\mu)^2}{2\sigma^2}\right), \tag{20.14}$$

and the CDF is

$$F_V(v) = 0.5 - 0.5\,\text{erf}\left(\frac{v-\mu}{\sigma\sqrt{2}}\right), \tag{20.15}$$

where $\text{erf}(z)$ is the error function, given by

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp\left(-x^2\right) \mathrm{d}x. \tag{20.16}$$

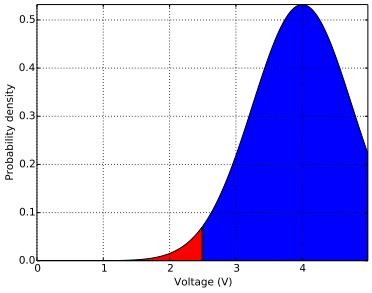


Figure 20.6: Switching voltage $V_T = 2.5\,\text{V}$, mean receiver voltage $\mu = 4.0\,\text{V}$, Gaussian noise with $\sigma = 0.75\,\text{V}$. Blue represents correct logic decisions. The noise margin is $\epsilon = 1.5\,\text{V}$ or two-sigma.

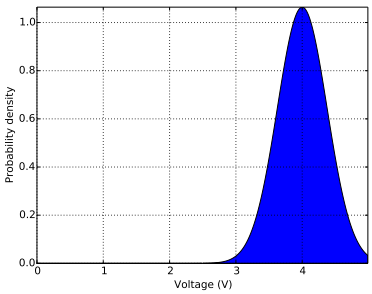


Figure 20.7: Switching voltage $V_T = 2.5\,\text{V}$, mean receiver voltage $\mu = 4.0\,\text{V}$, Gaussian noise with $\sigma = 0.375\,\text{V}$. Blue represents correct logic decisions. The noise margin is $\epsilon = 1.5\,\text{V}$ or four-sigma.

[2] This is equivalent to the bit error rate (BER) when an infinite number of bits are transmitted.

11.1 *Noise margin and signal to noise ratio*

Denoting the mean signal level above the receiver threshold by $\epsilon = \mu - V_{\text{ih}}$ (the noise margin) then the signal to noise ratio is

$$\text{SNR} = \frac{\epsilon^2}{\sigma^2} = \frac{(\mu - V_{\text{ih}})^2}{\sigma^2}, \tag{20.17}$$

and, using (20.13), the probability of an error[2], $P(\text{SNR})$, for a given SNR is

$$P(\text{SNR}) = 0.5 - 0.5\,\text{erf}\left(\sqrt{\text{SNR}}/\sqrt{2}\right) \tag{20.18}$$

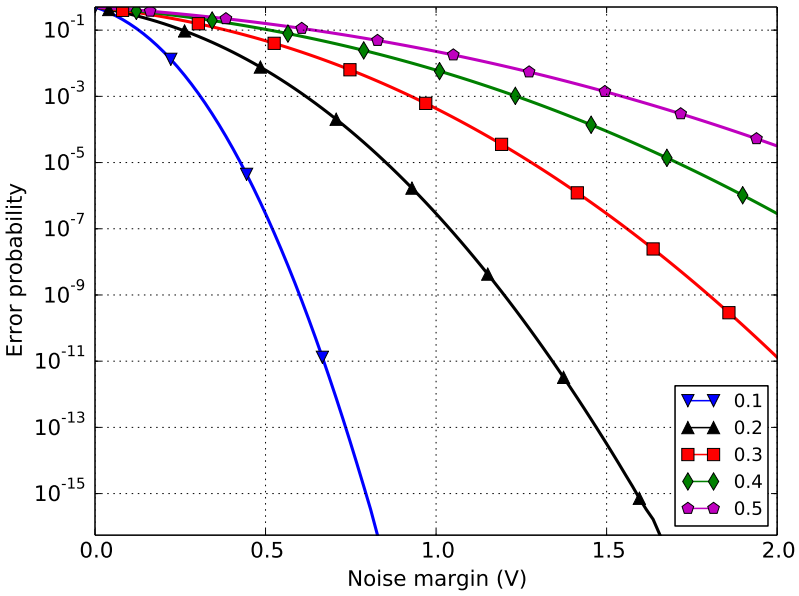


Figure 20.8: Probability of incorrect logic decision as a function of the noise margin ϵ and rms noise σ.

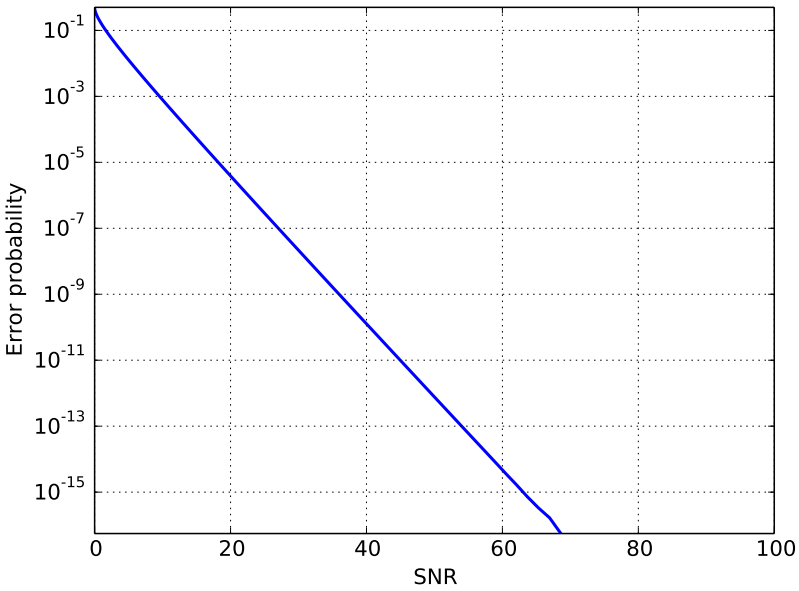


Figure 20.9: Probability of incorrect logic decision as a function of the signal to noise ratio SNR $= \epsilon^2/\sigma^2$, where ϵ is the noise margin and σ is the noise standard deviation.

21
Transmission line supplement

This is for those who like understanding how transmission lines can be modelled with mathematics.

1 *Transmission line analysis*

Transmission lines can be analysed in either the time or frequency domains. Generally, transmission lines for digital systems are analysed in the time domain. This is because the transmission lines have little loss and the source and load impedances are primarily resistive over most frequencies of interest. When the transmission line is lossy or the load and source impedances are not resistive, the differential equations describing the circuit become too complicated to solve in the time domain. Then we need to use Laplace or Fourier transforms to analyse the circuit in the frequency domain[1] and transform the results back into the time domain. Unfortunately, it is hard to achieve a closed-form result and usually we have to resort to numerical methods.

[1] This is why transmission lines in electronics courses are analysed in the frequency domain using Smith's charts, etc.

1.1 *Lossless tx. line time-domain analysis*

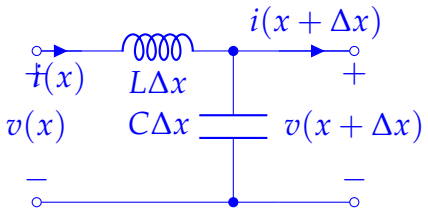


Figure 21.1: Lossless transmission line incremental section of length Δx.

Let's start with a lossless L-C transmission line of length l having a series inductance per unit length L and a shunt capacitance per unit length C. If we analyse an incremental element of length Δx (as shown in Figure 21.1), the incremental series inductance is $L\Delta x$

and the incremental shunt capacitance is $C\Delta x$. The voltage across the series inductance is

$$\Delta v(x,t) = v(x+\Delta x,t) - v(x,t) = -L\Delta x \frac{\partial}{\partial t} i(x,t). \quad (21.1)$$

Similarly, the current flowing[2] through the shunt capacitance element is

$$\Delta i(x+\Delta x,t) = i(x+\Delta x,t) - i(x,t) = -C\Delta x \frac{\partial}{\partial t} v(x+\Delta x,t). \quad (21.2)$$

[2] The negative signs result from the definitions of the incremental currents and voltages.

If we now consider the limit as Δx tends to zero (i.e., we consider an element of infinitesimal length), the voltage and current gradients are:

$$\frac{\partial v}{\partial x} = -L\frac{\partial i}{\partial t}, \quad (21.3)$$

$$\frac{\partial i}{\partial x} = -C\frac{\partial v}{\partial t}. \quad (21.4)$$

Differentiating (21.3) by x, differentiating (21.4) by t, and combining gives a 1-D wave equation:

$$\frac{\partial^2 v}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 v}{\partial t^2} = 0, \quad (21.5)$$

where the speed of propagation is

$$v = \frac{1}{\sqrt{LC}}. \quad (21.6)$$

The general solution[3] to (21.5) is

[3] de Lambert's solution.

$$v(x,t) = V_+\, p\left(t - \frac{x}{v}\right) + V_-\, p\left(t + \frac{x}{v}\right), \quad (21.7)$$

where V_+ and V_- are arbitrary constants constrained by the boundary conditions and $p(t)$ is an arbitrary twice differentiable function. Taking derivatives, (21.7) can be seen to be a solution to (21.5) since:

$$\frac{\partial v}{\partial t} = V_+\, p'\left(t - \frac{x}{v}\right) + V_-\, p'\left(t + \frac{x}{v}\right), \quad (21.8)$$

$$\frac{\partial v}{\partial x} = -\frac{V_+}{v} p'\left(t - \frac{x}{v}\right) - \frac{V_-}{v} p'\left(t + \frac{x}{v}\right), \quad (21.9)$$

$$\frac{\partial^2 v}{\partial t^2} = V_+\, p''\left(t - \frac{x}{v}\right) + V_-\, p''\left(t + \frac{x}{v}\right), \quad (21.10)$$

$$\frac{\partial^2 v}{\partial x^2} = \frac{V_+}{v^2} p''\left(t - \frac{x}{v}\right) + \frac{V_-}{v^2} p''\left(t + \frac{x}{v}\right). \quad (21.11)$$

The current along the transmission line obeys a similar wave equation:

$$\frac{\partial^2 i}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 i}{\partial t^2} = 0, \quad (21.12)$$

with a general solution

$$i(x,t) = I_+\, p\left(t - \frac{x}{v}\right) + I_-\, p\left(t + \frac{x}{v}\right). \tag{21.13}$$

Now,

$$\frac{\partial i}{\partial t} = I_+\, p'\left(t - \frac{x}{v}\right) + I_-\, p'\left(t + \frac{x}{v}\right), \tag{21.14}$$

and equating this to (21.3) yields[4]

$$I_+ = \frac{l}{Lv} V_+ = \frac{1}{Z_0} V_+, \tag{21.15}$$

$$I_- = -\frac{l}{Lv} V_- = -\frac{1}{Z_0} V_-, \tag{21.16}$$

where the characteristic impedance Z_0 of a lossless line is a real constant,

$$Z_0 = \sqrt{\frac{L}{C}}. \tag{21.17}$$

[4] Here I_+ and I_- are defined in the same direction. This results in the current reflection coefficient being the negative of the voltage reflection coefficient and $V_- = -I_- Z_0$.

1.10 *Lossy tx. line time-domain analysis*

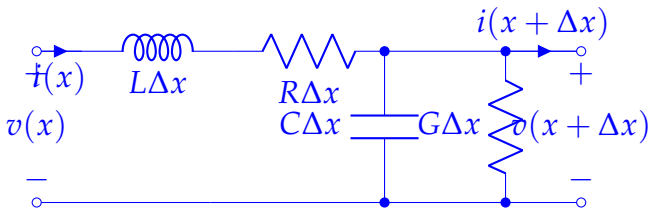


Figure 21.2: Lossy transmission line incremental section.

Let's now introduce some series resistance per unit length R and some shunt inductance per unit length G so the voltage and current gradients are:

$$\frac{\partial v}{\partial x} = -L\frac{\partial i}{\partial t} - Ri, \tag{21.18}$$

$$\frac{\partial i}{\partial x} = -C\frac{\partial v}{\partial t} - Gv, \tag{21.19}$$

Differentiating (21.18) by x, differentiating (21.19) by t, and combining yields the lossy wave equation:

$$\frac{\partial^2 v}{\partial x^2} - (RC + LG)\frac{\partial v}{\partial t} - LC\frac{\partial^2 v}{\partial t^2} = 0. \tag{21.20}$$

In practice this can be simplified[5] since $G \approx 0$ but even so at this point most people (including me) resort to frequency domain techniques to find a solution.

[5] Except at microwave frequencies.

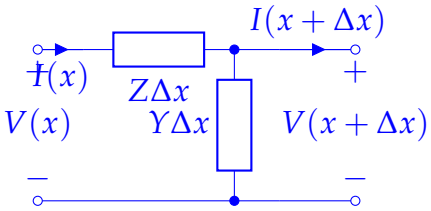


Figure 21.3: Lossy transmission line generalised incremental section in Laplace domain.

1.11 Lossy tx. line frequency-domain analysis

Transforming the model shown in Figure 21.2 into the Laplace domain, the series impedances and shunt admittances can be combined to yield a simpler, more general form, as shown by Figure 21.3 where

$$Z(s) = R + sL, \tag{21.21}$$
$$Y(s) = G + sC. \tag{21.22}$$

The voltage and current gradients (the *telegraphers' equations*) are:

$$\frac{dV}{dx} = -ZI, \tag{21.23}$$
$$\frac{dI}{dx} = -YV, \tag{21.24}$$

where $V = V(x,s)$ and $I = I(x,s)$. Differentiation of (21.23) and substitution in (21.24) yields the one-dimensional Helmholtz equation[6]

$$\frac{d^2V}{dx^2} - \gamma^2 V = 0, \tag{21.25}$$

[6] This is the frequency domain equivalent of the wave equation.

where γ is called the *propagation constant* and is given by

$$\gamma = \alpha + j\beta = \sqrt{ZY}. \tag{21.26}$$

The general solution to (21.25) is

$$V = V_+ \exp(-\gamma x) + V_- \exp(+\gamma x), \tag{21.27}$$

and from (21.23), the corresponding solution for the current is

$$I = \frac{1}{Z_0}\left(V_+ \exp(-\gamma x) - V_- \exp(+\gamma x)\right), \tag{21.28}$$

where Z_0 is the *characteristic impedance* defined by

$$Z_0 = \frac{V}{I} = \frac{\gamma}{Z} = \sqrt{\frac{Z}{Y}}. \tag{21.29}$$

If we consider steady state sinusoids where $s = \mathrm{j}2\pi f$, the speed of propagation (phase speed), c, and attenuation per metre, α, at a frequency f can be found using

$$c = \frac{2\pi f}{\mathrm{Im}\{\gamma\}}, \tag{21.30}$$

$$\alpha = \mathrm{Re}\{\gamma\}. \tag{21.31}$$

In other words, for a sinusoid of frequency f

$$\exp(-\gamma x) = \exp(-\alpha x)\exp(-\mathrm{j}2\pi(f/c)x). \tag{21.32}$$

The first factor describes the attenuation and the second factor describes the phase delay. Note that both c and α are functions of frequency—dispersion is always accompanied with attenuation. For example, the transmission line model shown in Figure 21.2 has a propagation constant

$$\gamma(f) = \sqrt{(R + \mathrm{j}2\pi f L)(G + \mathrm{j}2\pi f C)}. \tag{21.33}$$

and thus a dispersive phase-speed of propagation

$$c(f) = \frac{\mathrm{j}2\pi f}{\sqrt{(R + \mathrm{j}2\pi f L)(G + \mathrm{j}2\pi f C)}}. \tag{21.34}$$

Moreover, with a lossy transmission line, the characteristic impedance is no longer real,

$$Z_0(f) = \sqrt{\frac{R + \mathrm{j}2\pi f L}{G + \mathrm{j}2\pi f C}}. \tag{21.35}$$

10 *Characteristic impedance transfer function*

The characteristic impedance relates the amplitude of the voltage pulse V_p to the amplitude of the current pulse I_p in the *frequency domain*. Thus it can be considered a transfer function, i.e.,

$$V_p(s) = Z_0(s) I_p(s). \tag{21.36}$$

So what is the relationship between the voltage and current in the time domain? Unfortunately, this can get gnarly. Remember the convolution theorem? Applying the convolution theorem to (21.36) yields,

$$v_p(t) = \int_{-\infty}^{\infty} z_0(u) i_p(t-u)\mathrm{d}u, \tag{21.37}$$

where $z_0(t)$ is an impulse response given by the inverse Laplace transform of the characteristic impedance $Z_0(s)$,

$$z_0(t) = \mathcal{L}^{-1} Z_0(s). \tag{21.38}$$

Things become much simpler if the transmission line is lossless implying that $Z_0(s)$ is constant. This results in an impulse response proportional to a Dirac delta,

$$z_0(t) = Z_0 \delta(t), \tag{21.39}$$

and thus the convolution integral in (21.37) collapses[7] to

$$v_p(t) = Z_0 i_p(t). \tag{21.40}$$

[7] This is why most texts assume lossless transmission lines.

11 *An intuitive look at transmission lines*

Consider an uncharged pair of parallel wires. Uncharged means that the electrons on the conductors' surface have spread themselves out uniformly without bunching[8].

Now let's consider a battery. The electrochemical processes inside a battery cause a big bunching of electrons at one electrode (cathode, the negative electrode) and an equivalent depletion of electrons at the other electrode (anode, the positive electrode.). Due to the imbalance of charges, across the battery terminals there is voltage proportional to how tightly the electrons are forced together. The strength of the electric field is proportional to the voltage and inversely proportional to the distance between the terminals. Since no charges are moving there is no magnetic field.

Let's now connect the battery to the ends of a pair of parallel wires. The electrons at the cathode now have an opportunity to move away from each other by flowing into one of the wires. At the end of this wire, there will now be an excess of electrons (compared to the number of protons) and thus there will be a net negative charge. Similarly, electrons will be sucked to the anode of the battery from the end of the other wire, producing a deficit of electrons (a net positive charge). Thus between the ends of the wire, an electric field is created. The strength of the electric field depends on the net charge imbalance, the separation of the wires, and the dielectric constant of the insulator between the wires.

The movement of the electrons into and out of the wires creates a magnetic field around each wire. The strength of this field depends on the current[9]. The magnetic field resists chaanges in the current[10].

[8] In practice, the electrons in the cable conductors will be moving around slightly due to thermal energy. Due to resistance they cannot move around instantaneously and so this will cause the electric field to vary slightly since the electrons may momentarily bunch up (the magnetic field will also vary slightly due to the motion of the electrons). The net effect is that there will be a noise voltage generated across the cable. We can ignore this except when dealing with very small signals.

[9] Current is proportional to the number of electrons entering/leaving the wires per second.

[10] We model this as inductance.

The imbalance of charge at the battery end of the wires has no immediate effect on the other end of the wires otherwise causality will be violated. However, the bunched electrons at the battery end of the negative wire want to spread out; they produce a force on their neighbours and cause them to move (but not immediately). Meanwhile the battery is 'pushing' more electrons into the end of the negative wire (and 'pulling' them out of the positive wire). As a result a moving electric and magnetic field is produced as the disturbance propagates along the wires.

If we used an AC source rather than a DC battery, and if the cable is long enough, parts of the coaxial cable will be charged positively while other parts will be charged negatively. The separation between these two regions is half the wavelength. The wavelength is the ratio of the speed of propagation to the frequency. So higher frequencies have shorter wavelengths.

100 Transmission line reflection analysis

The following section is for those who like understanding things with mathematics.

100.1 Transmission line reflection coefficients

In general, the reflection coefficients Γ_s and Γ_l are transfer functions in the frequency domain,

$$V'_p(s) = \Gamma_l(s)V_p(s). \tag{21.41}$$

To find the time domain response we need to use an inverse Laplace transform

$$v'_p(t) = \mathcal{L}^{-1}\left\{\Gamma_l(s)V_p(s)\right\}. \tag{21.42}$$

Using the convolution theorem, this can also be written as

$$v'_p(t) = v_p(t) * \gamma_l(t), \tag{21.43}$$

where $*$ denotes convolution and $\gamma_l(t)$ is an impulse response, related to the transfer function $\Gamma_l(s)$ by a Laplace transform,

$$\gamma_l(t) = \mathcal{L}^{-1}\left\{\Gamma_l(s)\right\}. \tag{21.44}$$

When the reflection coefficient is a real constant Γ_l (say when there is a resistive load on a lossless[11] transmission line) then $\gamma_l(t) = \Gamma_l\delta(t)$ and

$$v'_p(t) = v_p(t)\Gamma_l. \tag{21.45}$$

[11] Most texts on this subject make this assumption to keep the maths simple.

100.10 Two-port model

A transmission line is a two-port network and the voltages and currents at the ends can be related in the Laplace (or frequency) domain using a transmission matrix

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \frac{H^{-1}+H}{2} & Z_0\frac{H^{-1}-H}{2} \\ \frac{1}{Z_0}\frac{H^{-1}-H}{2} & \frac{H^{-1}+H}{2} \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}, \quad (21.46)$$

where

$$H = \exp(-\gamma l). \quad (21.47)$$

H is the one-way transfer function for a line of length l and $\gamma = \gamma(s)$ is the propagation constant. Substituting for H in (21.46) gives

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \cosh\gamma l & Z_0 \sinh\gamma l \\ \frac{1}{Z_0}\sinh\gamma l & \cosh\gamma l \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}. \quad (21.48)$$

If we terminate the line with an impedance Z_l so that $V_2 = Z_l I_2$ then the (steady-state) impedance looking into the line is

$$Z_{\text{in}} = \frac{V_1}{I_1} = Z_0 \frac{Z_l \cosh\gamma l + Z_0 \sinh\gamma l}{Z_0 \cosh\gamma l + Z_l \sinh\gamma l}. \quad (21.49)$$

Note that the input impedance depends on the load impedance and thus is affected by the reflected wave. Also note that in the frequency domain, a time delay is equivalent to a frequency dependent phase shift.

From (21.49), the open circuit input impedance is

$$Z_{\text{oc}} = \left.\frac{V_1}{I_1}\right|_{Z_l=\infty} = Z_0 \frac{1}{\tanh\gamma l}, \quad (21.50)$$

and the short circuit input impedance is

$$Z_{\text{sc}} = \left.\frac{V_1}{I_1}\right|_{Z_l=0} = Z_0 \tanh\gamma l. \quad (21.51)$$

From these two measurements, the characteristic impedance can be determined,

$$Z_0 = \sqrt{Z_{\text{oc}} Z_{\text{sc}}}. \quad (21.52)$$

If we drive a transmission line with a voltage source $V_d = V_d(s)$ through an impedance Z_d, the voltage at any point on the line is

$$V(x,s) = \frac{V_d}{2}\frac{Z_{\text{in}} + Z_0}{Z_{\text{in}} + Z_d}\exp(-\gamma x) + \frac{V_d}{2}\frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_d}\exp(\gamma x). \quad (21.53)$$

When Z_d, Z_l, and Z_0 are purely resistive then for $x = 0$ (21.53) remarkably transforms into the time response we would calculate heuristically,

$$v_1(t) = \frac{1-\Gamma_s}{2}\left(v_s(t) + \sum_{m=1}^{\infty} \Gamma_l^m \Gamma_s^{m-1}(1+\Gamma_s)v_s(t-2mT)\right). \tag{21.54}$$

In the frequency domain it is common to consider a transmission line in terms of the input acceptance function:

$$A(s) = \frac{Z_0(s)}{Z_d(s)+Z_0(s)}, \tag{21.55}$$

the propagation function:

$$H(x,s) = \exp\left(-\gamma(s)x\right), \tag{21.56}$$

and the output transmission function:

$$T(s) = \frac{2Z_l(s)}{Z_l(s)+Z_0(s)}. \tag{21.57}$$

100.11 Steady-state

When all the reflections have died down the final voltage on the line is the steady-state voltage. If an AC voltage is applied this will be spatially dependent[12] due to the superposition of the incident and reflected waves. If a DC voltage v_d is applied through a source resistance R_s, then all points on a lossless line will charge to a final voltage v_f,

[12] The solution will consist of hyperbolic functions with distance x along the line as a parameter, see (21.53).

$$v_f = \lim_{t\longrightarrow\infty} v_l(t) = \lim_{t\longrightarrow\infty} v_s(t), \tag{21.58}$$

$$= \frac{Z_0 V_d}{R_s+Z_0}\left(1+\Gamma_l+\Gamma_l\Gamma_s+\Gamma_l^2\Gamma_s+\Gamma_l^2\Gamma_s^2+\cdots\right). \tag{21.59}$$

This can be split into two terms, one for waves travelling away from the source and the other for waves travelling toward the source:

$$v_f = \frac{Z_0 v_d}{R_s+Z_0}\left(\sum_{n=0}^{\infty}(\Gamma_l\Gamma_s)^n + \Gamma_l\sum_{n=0}^{\infty}(\Gamma_l\Gamma_s)^n\right), \tag{21.60}$$

$$= \frac{Z_0 v_d}{R_s+Z_0}(1+\Gamma_l)\sum_{n=0}^{\infty}(\Gamma_l\Gamma_s)^n. \tag{21.61}$$

Using the identity for the summation of a geometric series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \tag{21.62}$$

then

$$v_f = \frac{Z_0}{R_s + Z_0} \frac{1 + \Gamma_l}{1 - \Gamma_l \Gamma_s} v_d. \tag{21.63}$$

Finally, substituting for Γ_l and Γ_s in terms of R_l and R_s, then

$$v_f = \frac{R_l}{R_s + R_l} v_d. \tag{21.64}$$

This is the result that would be calculated using simple circuit theory assuming a lossless transmission line with R_s and R_l both resistive.

101 Microstrip trace impedance

A PCB trace running above a ground plane is called a microstrip. A trace is usually modelled with a trapezoidal cross section with 45 degree slopes, width W, and thickness T (see Figure 21.4).

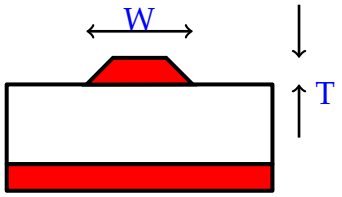


Figure 21.4: Trapezoidal model of microstrip trace cross-section. This approximates the shape of a trace after etching but typically underestimates the resistance by about 10%.

101.1 Capacitance

The capacitance per unit length of a microstrip is

$$C = \frac{26.4 \times 10^{-12} (\epsilon_r + 1.41)}{\ln\left(\frac{5.98H}{0.8W+T}\right)} \quad \text{(F/m)}, \tag{21.65}$$

where W is the trace width, H is the separation between the trace and ground plane, T is the trace thickness, and ϵ_r is the relative permittivity of the dielectric. This is valid for $0.1 < W/H < 3.0$ and $1 < \epsilon_r < 15$.

101.10 Inductance

The inductance per unit length of a microstrip is

$$L = 0.2 \times 10^{-6} \ln\left(\frac{5.98H}{0.8\,W + T}\right) \quad \text{(H/m)}. \tag{21.66}$$

This is valid for $0.1 < W/H < 3.0$ and $1 < \epsilon_r < 15$.

101.11 Characteristic impedance

Assuming no loss, the characteristic impedance of a microstrip is

$$\begin{aligned} Z_0 &= \sqrt{\frac{L}{C}}, \\ &= \frac{87}{\sqrt{\epsilon_r + 1.41}} \ln\left(\frac{5.98H}{0.8\,W + T}\right) \quad \text{(ohms)} \end{aligned} \tag{21.67}$$

and the speed of propagation is

$$\begin{aligned} v &= \frac{1}{\sqrt{LC}}, \\ &= \frac{1}{2.3 \times 10^{-9}\sqrt{\epsilon_r + 1.41}}, \\ &= \frac{4.35 \times 10^8}{\sqrt{\epsilon_r + 1.41}} \quad \text{(m/s)}. \end{aligned} \tag{21.68}$$

This corresponds to a propagation delay per unit metre of

$$\begin{aligned} \tau &= \frac{1}{v}, \\ &= \sqrt{LC}, \\ &= 2.3 \times 10^{-9}\sqrt{\epsilon_r + 1.41} \quad \text{(s/m)}. \end{aligned} \tag{21.69}$$

101.100 DC resistance

The resistance of a copper trace is given by

$$R = \rho\frac{L}{A}, \tag{21.70}$$

where ρ is the resistivity of copper (ohm-m), L is the length of the trace (m), and A is the cross sectional area of the trace (m^2), given by

$$A = (W - T)T + \frac{1}{2}T^2. \tag{21.71}$$

The PCB copper thickness T is often specified by a plating weight in ounces (oz)[13]. Assuming a uniform thickness T, the mass m of copper spread over a surface area A is

$$m = \rho A T, \tag{21.72}$$

where $\rho = 8930\,\text{kg/m}^3$ is the (mass) density of copper. Defining

$$\mu = \frac{m}{A}, \tag{21.73}$$

[13] More accurately it should be ounces per square foot (oz/ft^2).

then

$$T = \frac{\mu}{\rho}. \tag{21.74}$$

Now since 1 oz is equivalent to 28.35×10^{-3} kg and 1 ft^2 is equivalent to 0.645×10^{-3} m^2, then 1 oz/ft^2 is equivalent to 0.305 kg/m^2 and thus a thickness of 34.8 μm. However, the real value is closer to 25 μm before plating since during processing the thickness of the copper is reduced as shown in Table 21.1.

Copper has a positive temperature coefficient so its resistivity increases with temperature. The temperature dependence of the resistivity of annealed copper can be modelled by

$$\rho(T) = \rho_0 \left(1 + \alpha(T - T_0)\right), \tag{21.75}$$

where $T_0 = 20°$ is a reference temperature, $\rho_0 = 1.724 \times 10^{-8}$ ohm-m is the resistivity at the reference temperature, and $\alpha = 0.00393°\text{C}^{-1}$ is the temperature coefficient at the reference temperature.

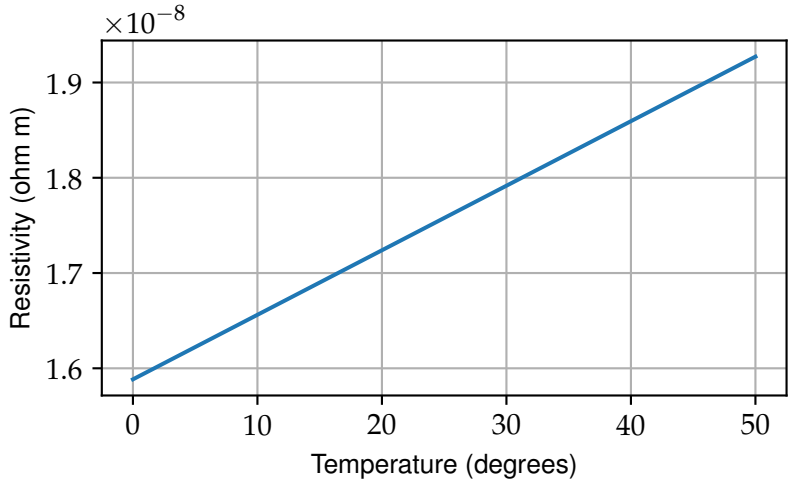


Figure 21.5: Temperature dependence of annealed copper resistivity.

Weight (oz/sq ft)	Internal trace thickness	External trace thickness
0.5	0.5 mil (13 μm)	1.3 mil (33 μm)
1.0	1.0 mil (25 μm)	1.8 mil (46 μm)
2.0	2.2 mil (56 μm)	3.0 mil (76 μm)
3.0	3.6 mil (91 μm)	4.2 mil (107 μm)

Table 21.1: Minimum copper thickness on printed circuit board after fabrication. The external trace thickness includes plating. Source: IPC-6012 Tables 3-8 and 3-9. Note, mils are thousands of an inch; 1 mil = 25.4 μm.

101.101 AC resistance

The AC resistance of a copper trace is greater than the DC resistance due to the skin effect, where the current is constrained to the outer surfaces of the trace and thus the effective cross sectional area is smaller.

101.110 Example

A PCB manufacturer produces 2 and 4 layer boards with FR4 glass epoxy composite 0.062 in (1.57 mm) thick with a copper plating weight of 1 oz/square foot (35 μm). The manufacturing tolerance on the thickness is 33–51 μm. Let's use the worst case giving a trace thickness of $T = 33\,\mu$m. Thus a 6 mil (0.152 mm) trace[14] has a cross sectional area of

[14] The dimensions of PCBs are still mostly specified in inches (25.4 mm) and mils (thousands of inches). 1 mil is equivalent to 25.4 μm (or 25.4 micron).

$$\begin{aligned} A &= (W-T)T + \frac{1}{2}T^2, \\ &= 3.9\times10^{-9} + 0.5\times10^{-9}, \\ &= 4.4\times10^{-9}\,\mathrm{m}^2, \end{aligned} \tag{21.76}$$

and a DC resistance per metre of

$$\begin{aligned} \frac{R}{L} &= \frac{\rho}{A}, \\ &= \frac{1.724\times10^{-8}}{4.4\times10^{-9}}, \\ &= 3.9\,\mathrm{ohm/m}. \end{aligned} \tag{21.77}$$

The capacitance per unit length is

$$\begin{aligned} C &= \frac{26.4\times10^{-12}\,(\epsilon_r+1.41)}{\ln\left(\frac{5.98H}{0.8\,W+T}\right)}, \\ &= \frac{26.4\times10^{-12}\times(4.5+1.41)}{\ln\left(\frac{5.98\times1.5\times10^{-3}}{0.8\times152\times10^{-6}+33\times10^{-6}}\right)}, \\ &= 38\,\mathrm{pF/m}, \end{aligned} \tag{21.78}$$

and the inductance per unit length is

$$\begin{aligned} L &= 0.2\times10^{-6}\ln\left(\frac{5.98H}{0.8\,W+T}\right), \\ &= 0.2\times10^{-6}\ln\left(\frac{5.98\times1.5\times10^{-3}}{0.8\times152\times10^{-6}+33\times10^{-6}}\right), \\ &= 0.81\,\mu\mathrm{H/m}. \end{aligned} \tag{21.79}$$

This gives a characteristic impedance of $Z_0 = 148$ ohms. This high value is due to the skinny trace size and the large separation from the ground plane.

Using a four layer PCB, the separation of the trace from the ground plane can be reduced to $H = 0.23\,\text{mm}$. This lowers the characteristic impedance for a 6 mil trace to $Z_0 = 78\,\text{ohms}$. On an 8 layer board with $H = 0.13\,\text{mm}$ the characteristic impedance for a 6 mil trace can be further reduced to 57 ohms.

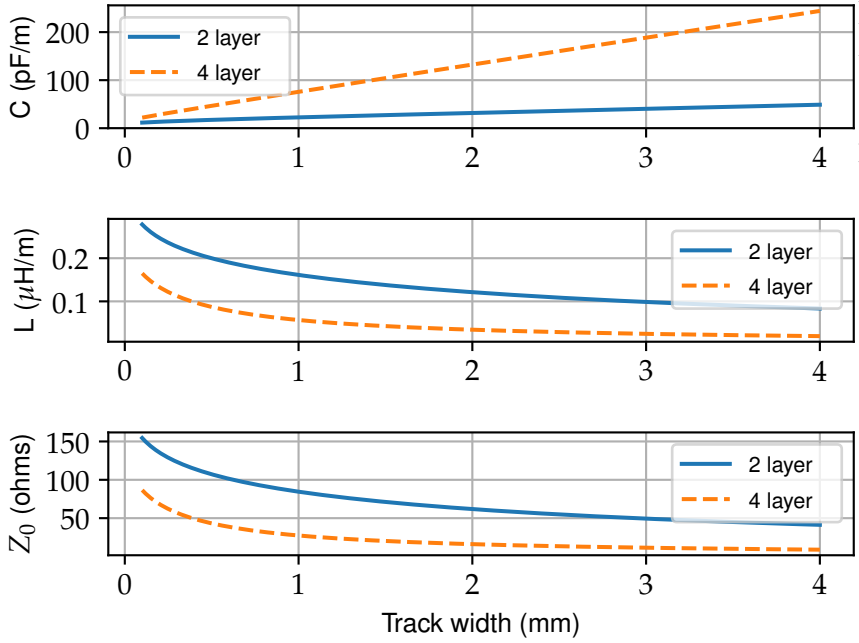


Figure 21.6: Characteristic impedance for microstrip trace above a ground plane on typical 2 layer ($T = 33\,\mu\text{m}$, $H = 1.6\,\text{mm}$) and 4 layer ($T = 33\,\mu\text{m}$, $H = 0.23\,\text{mm}$) PCBs.

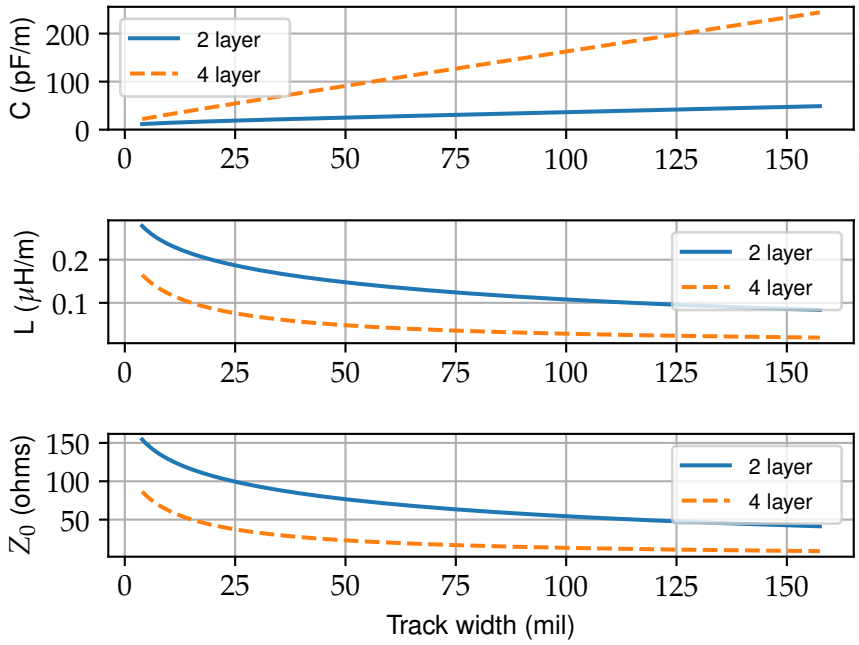


Figure 21.7: Characteristic impedance for microstrip trace above a ground plane on typical 2 layer ($T = 33\,\mu$m, $H = 1.6\,$mm) and 4 layer ($T = 33\,\mu$m, $H = 0.23\,$mm) PCBs.