Advanced Integrated Circuits

Lecture Notes 2025

ASICedu.com

Contents

Background | 1

In the spring of 2024 I lectured Advanced Integrated Circuits for the third time. I have an inherent need to make things better, and the course is no different.

In the first round I noticed that little of what I had on slides, or said in lectures, made it into the student brain. That annoyed me, and I realized that probably a few things needed to change.

I think the lectures have gotten better, but I don't have any specific proof. There were 19 students that took the exam in 2024. An indication of lecture quality could be attendance. I don't have all the dates, but an average attendance of 76 % I think is pretty OK.

Date	Attendance
2024-02-02	19
2024-02-09	17
2024-02-16	16
2024-03-01	14
2024-03-07	14
2024-03-15	12
2024-03-22	13
2024-04-12	16
2024-04-19	10

For the third semester I finally felt I achieved a balance. I spent Thursday's preparing for the lecture, writing these notes, making a YouTube video (so I'll remember next year what I wanted to talk about). I passed 1k subscribers. Friday's I had the lecture and the group work.

For the group work I forced students into groups, and I forced that they for the first 5-10 minutes do a check-in. That I need to do next year too.

For the check in, they had go around in the group and answer one of the following questions:

- ► What is one thing that is going on in your life (personal or professional)?
- ► What is one thing that you're grateful for right now?
- ► What is something funny that happened?

The check-in led to excellent team work for those students that showed up.

Thanks to Jonathan for helping out in the exercise hours.

I love programming and automation. Not much makes me more happy than using the same source (the slide markdowns), to generate the lecture notes, to translate into the book your looking at right now.

If you find an error in what I've made, then fork analogIC, fix , commit, push and create a pull request. That way, we use the global brain power most efficiently, and avoid multiple humans spending time on discovering the same error.

2.1 Who

My name is

Carsten Wulff carstenw@ntnu.no
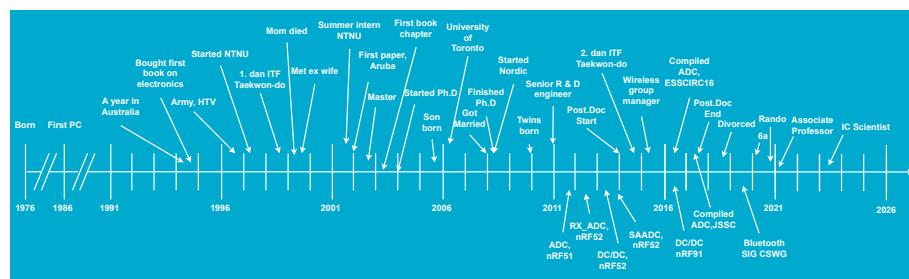
I finished my Masters in 2002, and did a Ph.D on analog-to-digital converters finished in 2008.

Since that time, I've had a three axis in my work/hobby life.

I work at Nordic Semiconductor where I've been since 2008. The first 7 years I did analog design (ADCs, DC/DCs, GPIO). The next 7 years I was the Wireless Group Manager. The Wireless group make most of the analog and RF designs for Nordic's short-range products. Now I'm the IC Scientist, and focus on technical issues with our integrated circuits that occur before we go into volume production.

I work at NTNU where I did a part time postdoc from 2014 - 2017. From 2020 I've been working on and teaching Advanced Integrated Circuits

I have a hobby trying to figure out how to make a new analog circuit design paradigm. The one we have today with schematic/simulation/layout/verification/simulation is too slow

2.2 I want you to learn the skills necessary to make your own ICs

In 2020 the global integrated circuit market was 437.7 billion dollars! The market is expected to grow to 1136 billion in 2028.

Integrated circuits enable pretty much all technologies.

I will be dead in approximately 50 years, and will retire in approximately 30 years. Everything I know will be gone (except for the small pieces I've left behind in videos or written word)

Someone must take over, and to do that, they need to know most of what I know, and hopefully a bit more.

That's were some of you come in. Some of you will find integrated circuits interesting to make, and in addition, you have the stamina, patience, and brain necessary to learn some of the hardest topics in the world.

> Making integrated circuits (that work reliably) is not rocket science, it's much harder.

In this course we'll focus on analog ICs, because the real world is analog, and all ICs must have some analog components, otherwise they won't work.

Insights · Tech The Future

The World Is Analog

10/28/2014

Written by Peter Kinget

The world we live in is analog. We are analog. Any inputs we can perceive are analog. For example, sounds are analog signals; they are continuous time and continuous value. Our ears listen to analog signals and we speak with analog signals. Images, pictures, and video are all analog at the source and our eyes are analog sensors. Measuring our heartbeat, tracking our activity, all requires processing analog sensor information.

https://circuitcellar.com/insights/tech-the-future/kinget-the-world-is-analog/
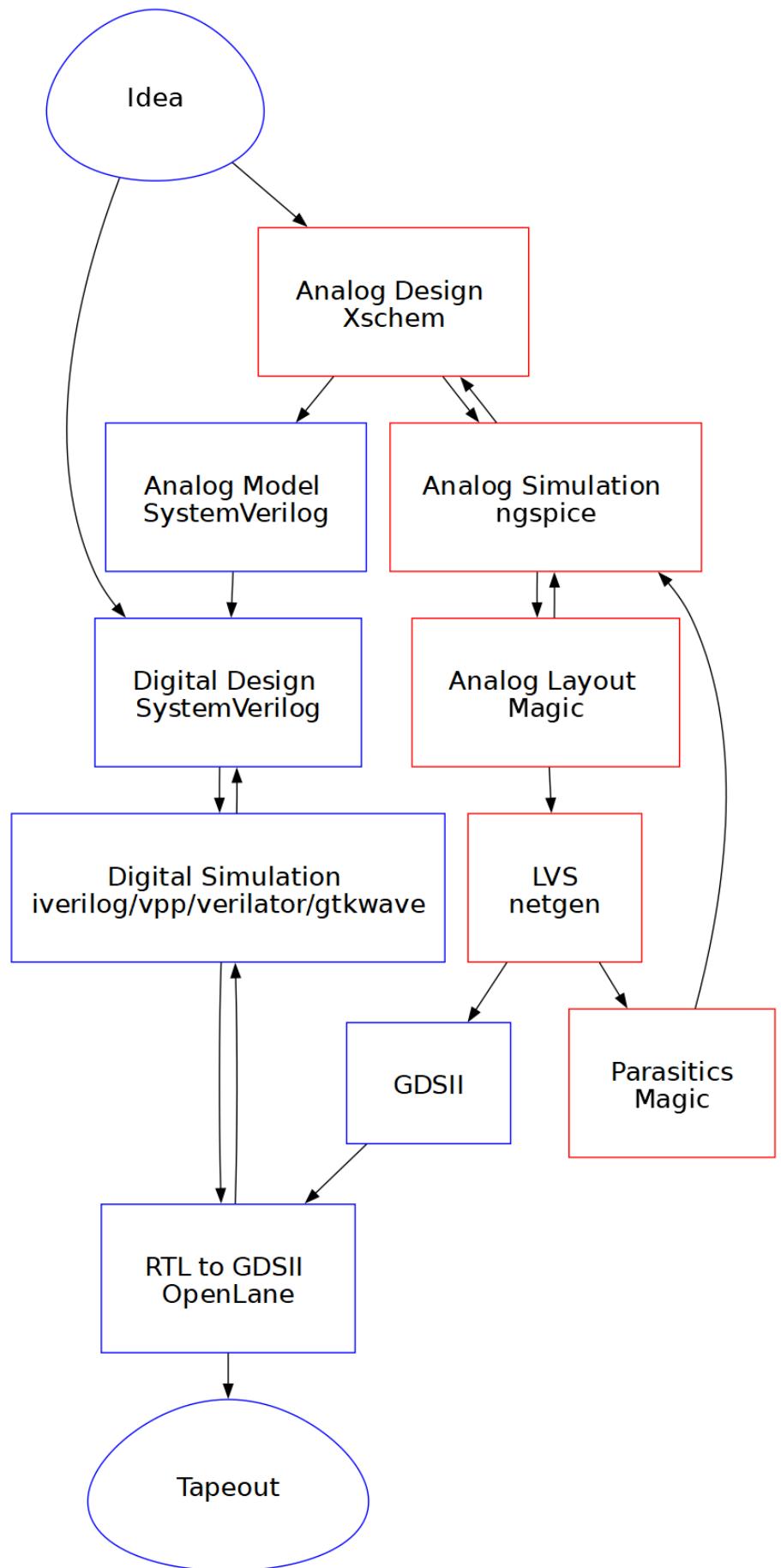
The steps to make integrated circuits is split in two. We have a analog flow, and a digital flow.

It's rare to find a single human that do both flows well. Usually people choose, and I think it's based on what they like and their personality.

If you like the world to be ordered, with definite answers, then it's likely that you'll find the digital flow interesting.

If you're comfortable with not knowing, and an insatiable desire to understand how the world *really* works at a fundamental level, then it's likely that you'll find analog flow interesting.

2.2.1 Will you tape-out an IC?

Something that would make me really happy is if someone is able to tapeout an IC after this course.

It's now possible without signing an NDA or buying expensive software licenses.

In 2020 Google and Skywater joined forces to release a 130 nm process design kit to the public. In addition, they have fueled a renaissance of open source software tools.

Together with Efabless there are cheap alternatives, like tinytapeout, which makes it possible for a private citizen to tape-out their own integrated circuit.

Google just sponsored a GlobalFoundries 180 nm tapeout where you could tape out your circuit for free.

2.2.2 What the team needs to know to design ICs

There are a multitude of tools and skills needed to design professional ICs. It's not likely that you'll find all the skills in one human, and even if you could, one human does not have sufficient bandwidth to design ICs with all it's aspects in a reasonable timeline

That is, unless we can find a way to make ICs easier.

The skills needed are

- ▶ *Project flow support*: **Confluence**, JIRA, risk management (DFMEA), failure analysis (8D)
- ▶ *Language*: **English**, **Writing English (Latex, Word, Email)**
- ▶ *Psychology*: Personalities, convincing people, presentations (Powerpoint, Deckset), **stress management (what makes your brain turn off?)**
- ▶ *DevOps*: **Linux**, bulid systems (CMake, make, ninja), continuous integration (bamboo, jenkins), **version control (git)**, containers (docker), container orchestration (swarm, kubernetes)
- ▶ *Programming*: Python, Go, C, C++, Matlab Since 1999 I've programmed in Python, Go, Visual BASIC, PHP, Ruby, Perl, C#, SKILL, Ocean, Verilog-A, C++, BASH, AWK, VHDL, SPICE, MATLAB, ASP, Java, C, SystemC, Verilog, and probably a few I've forgotten.
- ▶ *Firmware*: signal processing, algorithms
- ▶ *Infrastructure*: **Power management**, **reset**, **bias**, **clocks**
- ▶ *Domains*: CPUs, peripherals, memories, bus systems

> ▶ *Sub-systems*: **Radio's**, **analog-to-digital converters**, **comparators**
> ▶ *Blocks*: **Analog Radio**, Digital radio baseband
> ▶ *Modules*: Transmitter, **receiver**, de-modulator, timing recovery, state machines
> ▶ *Designs*: **Opamps**, **amplifiers**, **current-mirrors**, adders, random access memory blocks, standard cells
> ▶ *Tools*: **schematic**, **layout**, **parasitic extraction**, synthesis, place-and-route, **simulation**, (System)Verilog, **netlist**
> ▶ *Physics*: transistor, pn junctions, quantum mechanics

2.2.3 Zen of IC design (stolen from Zen of Python)

When you learn something new, it's good to listen to someone that has done whatever it is before.

Here is some guiding principles that you'll likely forget.

> ▶ Beautiful is better than ugly.
> ▶ Explicit is better than implicit.
> ▶ Simple is better than complex.
> ▶ Complex is better than complicated.
> ▶ Readability counts (especially schematics).
> ▶ Special cases aren't special enough to break the rules.
> ▶ Although practicality beats purity.
> ▶ In the face of ambiguity, refuse the temptation to guess.
> ▶ There should be one **and preferably only one** obvious way to do it.
> ▶ Now is better than never.
> ▶ Although never is often better than *right* now.
> ▶ If the implementation is hard to explain, it's a bad idea.
> ▶ If the implementation is easy to explain, it may be a good idea.

2.2.4 IC design mantra

To copy an old mantra I have on learning programming

> Find a problem that you really want to solve, and learn programming to solve it. There is no point in saying "I want to learn programming", then sit down with a book to read about programming, and expect that you will learn programming that way. It will not happen. The only way to learn programming is to do it, a lot. – Carsten Wulff

And run the perl program

`s/programming/``analog design``/ig`

2.3 My Goal

Don't expect that I'll magically take information and put it inside your head, and you'll suddenly understand everything about making ICs.

You are the one that must teach yourself everything.

I consider my role as a guide, similar to a mountain guide. I can't carry you up the mountain, you need to walk up the mountain , but I know the safe path to take and increase the likelihood that you'll come back alive.

I want to:

- ▶ Enable you to read the books on integrated circuits
- ▶ Enable you to read papers (latest research)
- ▶ Correct misunderstandings on the topic
- ▶ Answer any questions you have on the chapters

I'm not a mind reader, I can't see inside your head. That means, you must ask questions, only by your questions can I start to understand what pieces of information is missing from your head, or maybe somehow to correct your understanding.
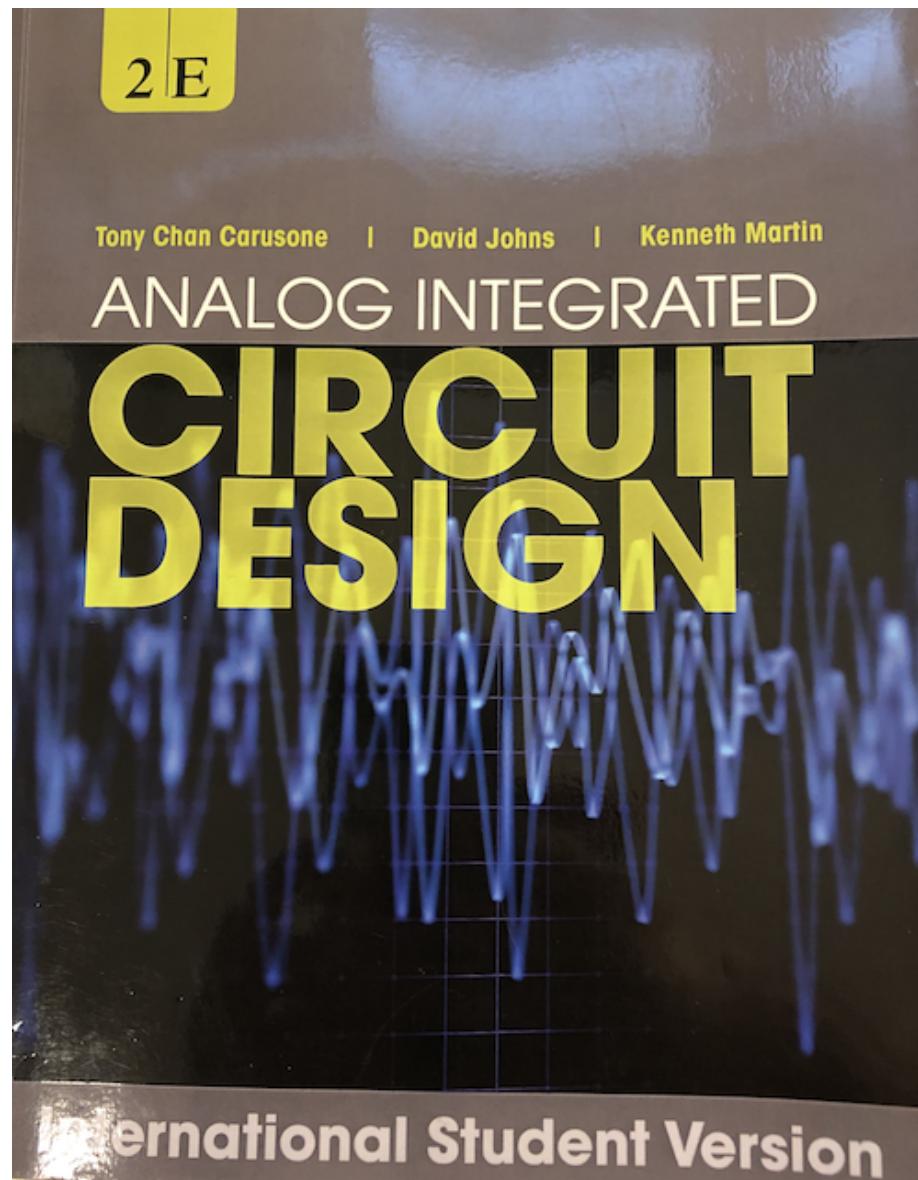
At the same time, and similar to a mountain guide, you should not assume I'm always right. I'm human, and I will make mistakes. And maybe you can correct my understanding of something. All I care about is to *really* understand how the world works, so if you think my understanding is wrong, then I'll happily discuss.

2.4 Syllabus

The syllabus will be from Analog Integrated Circuit Design (CJM) and Circuits for all seasons.

These lecture notes are a supplement to the book. I try to give some background, and how to think about electronics. It's not my goal to repeat information that you can find in the book.

Buy a hard-copy of the book if you don't have that. Don't expect to understand the book by reading the PDF.

2.5 CNR (2024)

"In an insane world, it was the sanest choice." - Sarah Connor, Terminator 2: Judgment Day

The project for 2024 is to

Design a integrated temperature sensor with digital read-out

An outline of the plan is shown below. There will be five milestones in all.

At the end of the project you will have a function that converts temperature to a digital value.

$$D = f_0(T)$$

I've broken down the challenge into three steps, first convert Temperature into a current

$$I = f_1(T)$$
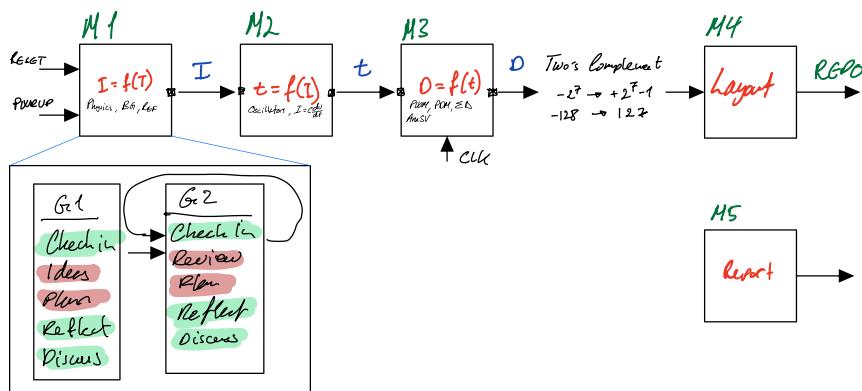
Then convert current into a time

$$t = f_2(I)$$

then time to digital

$$D = f_3(t) = f_3(f_2(f_1(T))) = f_0(T)$$

The fourth milestone is the layout, while the fifth milestone is the report.

2.5.1 Group dynamics

How you work together is important. No-one can do everything by them self. I know from experience it can be magical when bright brains come together. The collective brain can be smarter, better, faster, than anyone in the group.

That's why I think it's important not to just work in groups, but also focus on how we work in groups.

A group shall be maximum 4 members. There must be at least 3 that don't know each-other that well.

The group will meet once per week, and shall have a discussion according to the outline below.

If there is time left at the end of the group session it will be used for Q & A.

2.5.1.1 First session of milestone

During the first group session of a milestone, you will

Check-in (10 minutes)

Some example questions could be

- ▸ Share one thing that is going on in your life (personal or professional.)
- ▸ What is one thing that you are grateful for right now?
- ▸ What is something funny that happened?

Some examples answers could be: - My dog died yesterday, so I'm not feeling great today. - I woke up early, had an omelet, and went running, so I feel motivated and fantastic. - I feel *blaaah* today, motivation is lacking. - I went running yesterday and did not discover before I got home that I'd forgotten to put my pants on, even though it was -10 C.

The point of this exercise is to get to know each other a bit, and attempt to create psychological safety in the group.

Ideas (35 minutes)

Come up with ideas for how the milestone could be implemented. What circuit ideas could work?

Break (15 minutes)

Plan (20 minutes)

Sketch out who does what the next week. What's the goal for the week.

Reflect (5 minutes)

In silence, think about the group dynamics. How did it go today? What was good? What could be improved? Write down one word.

Discuss (10 minutes)

Each group member talks about their one word.

2.5.1.2 Other sessions of a milestone

You shall always Check-in, Reflect and Discuss. Although some may consider it a waste of time, it's important to improve the group dynamics.

Review (35 minutes)

Go through the plan from last week, what worked, what did not work, what should be done differently. Discuss.

Plan (20 minutes)

Sketch out who does what the next week. What's the goal for the
week.

2.6 Software

We'll use professional Open source software (xschem, ngspice,
sky130B PDK, Magic VLSI, netgen)

I've made a rather detailed (at least I think so myself) tutorial on
how to make a current mirror with the open source tools. I strongly
recommend you start with that first.

Skywater 130 nm Tutorial

I've also made some more complex examples, that can be found at
the link below. There are digital logic cells, standard transistors,
and few other blocks.

aicex

How to write a project report | 3

3.1 Why

> Them who has a Why? in life can tolerate almost any How?

You're writing the report on the project for me to be able to see inside your head, and grade how much of the project you have understood.

- ▶ Have you learned what is to be expected?
- ▶ Do you understand what you're trying to explain?

You will work on the project in groups, however, on the report, you will write on your own.

That means, that there will be X projects reports that describe the same circuit. You shall not copy someone elses report text.

It's fine to share figures between reports, and also references.

I'm also forcing you to use a report format that matches well with what would be expected if you were to publish a paper.

Should you make a fantastic temperature sensor, and maybe even reach close to a tapeout I would strongly suggest you submit a paper to NorCas. The deadline is August 15 2024.

3.2 On writing English

Writing well is important. I would recommend that you read On writing Well.

Most of you won't buy the book, as such, a few tips.

3.2.1 Shorter is better

I can write the section title idea in many words:

> A shorter text will more elequently describe the intricacies of your thoughts than a long, distinguished, tirade of carefully, wonderfully, choosen words.

or

Shorter is better

Describe an idea with as few words as possible. The text will be better, and more readable.

3.2.2 Be careful with adjectives

Words like "very, extremely, easily, simply, . . . " don't belong in a readable text. They serve no purpose. Delete them.

3.2.3 Use paragraphs

You write a text to place ideas into anothers head. Ideas and thoughts are best communicated in chunks. I can write a dense set of text, or I can split a dense set of text into multiple paragraphs. The more I try to cram into a paragraph, for example, how magical the weather has been the last weeks, with lots of snow, and good skiing, the more difficult the paragraph is to read.

One paragraph, one thought. For example:

You write a text to place ideas into anothers head. Ideas and thoughts are best communicated in chunks.

I can write a dense set of text, or I can split a dense set of text into multiple paragraphs.

The more I try to cram into a paragraph, for example, how magical the weather has been the last weeks, with lots of snow, and good skiing, the more difficult the paragraph is to read.

3.2.4 Don't be afraid of I

If you did something, then say "I" in the text. If there were more people, then use "we".

3.2.5 Transitions are important

Sentences within a paragraph are sometimes linked. Use

- ► As a result,
- ► As such,
- ► Accordingly,
- ► Consequently,

And mix them up.

3.2.6 However, is not a start of a sentence

If you have to use "However" it should come in the middle of the sentence.

I want to go skiing, however, I cannot today due to work.

3.3 Report Structure

The sections below go through the expected structure of a report, and what the sections should contain.

3.3.1 Introduction

The purpose of the introduction is to put the reader into the right frame of mind. Introduce the problem statement, key references, the key contribution of your work, and an outline of the work presented. Think of the introduction as explaining the "Why" of the work.

Although everyone has the same assignment for the project, you have chosen to solve the problem in different ways. Explain what you consider the problem statement, and tailor the problem statement to what the reader will read.

Key references are introduced. Don't copy the paper text, write why they designed the circuit, how they chose to implement it, and what they achieved. The reason we reference other papers in the introduction is to show that we understand the current state-of-the-art. Provide a summary where state-of-the-art has moved since the original paper.

The outline should be included towards the end of the introduction. The purpose of the outline is to make this document easy to read. A reader should never be surprised by the text. All concepts should be eased into. We don't want the reader to feel like they been thrown in at the end of a long story. As such, if you chosen to solve the problem statement in a way not previously solved in a key references, then you should explain that.

A checklist for all chapters can be seen in table below.

3.3.2 Theory

It is safe to assume that all readers have read the key references, if they have not, then expect them to do so.

The purpose of the theory section is not to demonstrate that you have read the references, but rather, highlight theory that the reader probably does not know.

The theory section should give sufficient explanation to bridge the gap between references, and what you apply in this text.

3.3.3 Implementation

The purpose of the implementation is to explain what you did. How have you chosen to architect the solution, how did you split it up in analog and digital parts? Use one subsection per circuit.

For the analog, explain the design decisions you made, how did you pick the transistor sizes, and the currents. Did you make other choices than in the references? How does the circuit work?

For the digital, how did you divide up the digital? What were the design choices you made? How did you implement readout of the data? Explain what you did, and how it works. Use state diagrams and block diagrams.

Use clear figures (i.e. circuitikz), don't use pictures from schematic editors.

3.3.4 Result

The purpose of the results is to convince the reader that what you made actually works. To do that, explain testbenches and simulation results. The key to good results is to be critical of your own work. Do not try to oversell the results. Your result should speak for themself.

For analog circuits, show results from each block. Highlight key parameters, like current and delay of comparator. Demonstrate that the full analog system works.

Show simulations that demonstrate that the digital works.

3.3.5 Discussion

Explain what the circuit and results show. Be critical.

3.3.6 Future work

Give some insight into what is missing in the work. What should be the next steps?

3.3.7 Conclusion

Summarize why, how, what and what the results show.

3.3.8 Appendix

Include in appendix the necessary files to reproduce the work. One good way to do it is to make a github repository with the files, and give a link here.

3.4 Checklist

Item	Description	OK
Is the problem description clearly defined?	Describe which parts of the problem you chose to focus on. The problem description should match the results you've achieved.	
Is there a clear explanation why the problem is worth solving?	The reader might need help to understand why the problem is interesting	
Is status of state-of-the-art clearly explained?	You should make sure that you know what others have done for the same problem. Check IEEEXplore. Provide summary and references. Explain how your problem or solution is different	
Is the key contribution clearly explained?	Highlight what you've achieved. What was your contribution?	
Is there an outline of the report?	Give a short summary of what the reader is about to read	

Item	Description	OK
Is it possible for a reader skilled in the art to understand the work?	Have you included references to relevant papers	
Is the theory section too long	The theory section should be less than 10 % of the work	
Are all circuits explained?	Have you explained how every single block works?	
Are figures clear?	Remember to explain all colors, and all symbols. Explain what the reader should understand from the figure. All figures must be referenced in the text.	
Is it clear how you verified the circuit?	It's a good idea to explain what type of testbenches you used. For example, did you use dc, ac or transient to verify your circuit?	
Are key parameters simulated?	You at least need current from VDD. Think through what you would need to simulate to prove that the circuit works.	
Have you tried to make the circuit fail?	Knowing how circuits fail will increase confidence that it will work under normal conditions.	
Have you been critical of your own results?	Try to look at the verification from different perspectives. Play devil's advocate, try to think through what could go wrong, then explain how your verification proves that the circuit does work.	
Have you explained the next steps?	Imagine that someone reads your work. Maybe they want to reproduce it, and take one step further. What should that step be?	
No new information in conclusion.	Never put new information into conclusion. It's a summary of what's been done	
Story	Does the work tell a story, is it readable? Don't surprise the reader by introducing new topics without background information.	

Item	Description	OK
Chronology	Don't let the report follow the timeline of the work done. What I mean by that is don't write "first I did this, then I spent huge amount of time on this, then I did that". No one cares what the timeline was. The report does not need to follow the same timeline as the actual work.	
Too much time	How much time you spent on something should not be correlated to how much text there is in the report. No one cares how much time you spent on something. The report is about why, how, what and does it work.	
Length	A report should be concise. Only include what is necessary, but no more. Shorter is almost always better than longer.	
Template	Use IEEEtran.cls. Example can be seen from an old version of this document at https://github.com/wulffern/dic2021/tree/main/2021-10-19_project_report. Write in LaTeX. You will need LaTeX for your project and master thesis. Use http://overleaf.com if you're uncomfortable with local text editors and LaTeX.	
Spellcheck	Always use a spellchecker. Misspelled words are annoying, and may change content and context (peaked versus piqued)	

Refresher | 4
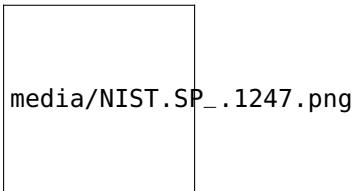
4.1 There are standard units of measurement

All known physical quantities are derived from 7 base units (SI units)

- ► second (s) : time
- ► meter (m) : space
- ► kg (kilogram) : weight
- ► ampere (A) : current
- ► kelvin (K) : temperature
- ► candela (cd) : luminous intensity

All other units (for example volts), are derived from the base units.

I don't go around remembering all of them, they are easily available online. When you forget the equation for charge (Q), voltage (V) and capacitance (C), look at the units below, and you can see it's $Q = CV$ *

media/NIST.SP_.1247.png
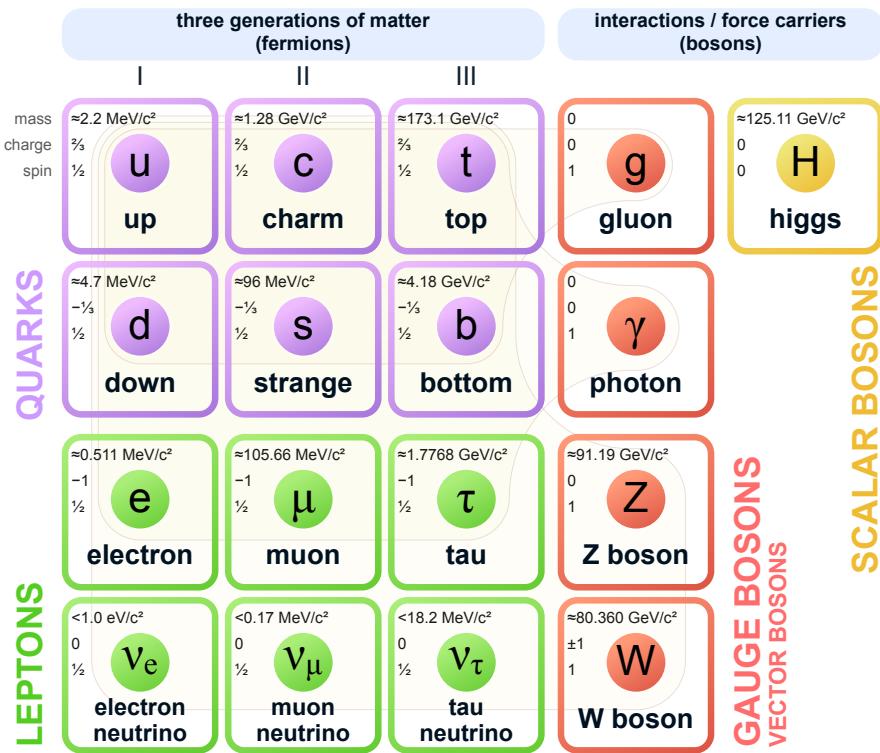
4.2 Electrons

Electrons are fundamental, they cannot (as far as we know), be divided into smaller parts. Explained further in the standard model of particle physics

---

* Although you do have to keep your symbols straight. We use "C" for Capacitance, but C can also mean Columbs. Context matters.

Standard Model of Elementary Particles

|	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
mass charge spin	≈2.2 MeV/c² ⅔ ½ **u** up	≈1.28 GeV/c² ⅔ ½ **c** charm	≈173.1 GeV/c² ⅔ ½ **t** top	0 0 1 **g** gluon	≈125.11 GeV/c² 0 0 **H** higgs
	≈4.7 MeV/c² −⅓ ½ **d** down	≈96 MeV/c² −⅓ ½ **s** strange	≈4.18 GeV/c² −⅓ ½ **b** bottom	0 0 1 **γ** photon	
	≈0.511 MeV/c² −1 ½ **e** electron	≈105.66 MeV/c² −1 ½ **μ** muon	≈1.7768 GeV/c² −1 ½ **τ** tau	≈91.19 GeV/c² 0 1 **Z** Z boson	
	<1.0 eV/c² 0 ½ **νe** electron neutrino	<0.17 MeV/c² 0 ½ **νμ** muon neutrino	<18.2 MeV/c² 0 ½ **ντ** tau neutrino	≈80.360 GeV/c² ±1 1 **W** W boson	

QUARKS · LEPTONS · GAUGE BOSONS (VECTOR BOSONS) · SCALAR BOSONS

Electrons have a negative charge of $q \approx 1.602 \times 10^{-19}$. The proton a positive charge. The two charges balance exactly! If you have a trillion electrons and a trillion protons inside a volume, the net external charge will be 0 (assuming we measure from some distance away). I find this fact absolutely incredible. There must be a fundamental connection between the charge of the proton and electron. It's insane that the charges balance out so exactly.

All electrons are the same, although the quantum state can be different.

An electron cannot occupy the same quantum state as another. This rule that applies to all Fermions (particles with spin of 1/2)

The quantum state of an electron is fully described by it's spin, momentum (p) and position in space (r).

4.3 Probability

The probability of finding an electron in a state as a function of space and time is

$$P = |\psi(r,t)|^2$$

, where ψ is named the probability amplitude, and is a complex function of space and time. In some special cases, it's

$$\psi(r,t) = Ae^{i(kr-\omega t)}$$

, where A is complex number, k is the wave number, r is the position vector from some origin, ω is the frequency and t is time.

The energy is $E = \hbar\omega$, where $\hbar = h/2\pi$ and h is Planck Constant and the momentum is $p = \hbar k$

4.4 Uncertainty principle

We cannot, with ultimate precision, determine both the position and the momentum of a particle, the precision is

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

From the uncertainty (Unschärfe) principle we can actually estimate the size of the atom

4.5 States as a function of time and space

The time-evolution of the probability amplitude is

$$i\hbar\frac{d}{dt}\psi(r,t) = H\psi(r,t)$$

, where H is named the Hamiltonian matrix, or the energy matrix or (if I understand correctly) the amplitude matrix of the probability amplitude to change from one state to another.

For example, if we have a system with two states, a simplified version of two electrons shared between two atoms, as in H_2, or hydrogen gas, or co-valent bonds, then the Hamiltonian is a 2 x 2 matrix. And the ψ is a vector of $[\psi_1, \psi_2]$

Computing the solution to the Schrodinger Equation can be tricky, because you must know the number of relevant states to know the vector size of ψ and the matrix size of H. In addition, the H can be a function of time and space (I think).

Compared to the equations of electric fields, however, Schrodinger is easy, it's a set of linear differential equations.

4.6 Allowed energy levels in atoms

Solutions to Schrodinger result in quantized energy levels for an electron bound to an atom.

Take hydrogen, the electron bound to the proton can only exists in quantized energy levels. The lowest energy state can have two electrons, one with spin up, and one with spin down.

From Schrodinger you can compute the energy levels, which most of us did at some-point, although now, I can't remember how it was done. That's not important. The important is to internalize that the energy levels in bound electrons are discrete.

Electrons can transition from one energy level to another by external influence, i.e temperature, light, or other.

The probability of a state transition (change in energy) can be determined from the probability amplitude and Schrodinger.

4.7 Allowed energy levels in solids

If I have two silicon atoms spaced far apart, then the electrons can have the same spin and same momentum around their respective nuclei. As I bring the atoms closer, however, the probability amplitudes start to interact (or the dimensions of the Hamiltonian matrix grow), and there can be state transitions between the two electrons.

The allowed energy levels will split. If I only had two states interacting, the Hamiltonian could be

$$H = \begin{bmatrix} A & 0 \\ 0 & -A \end{bmatrix}$$

and the new energy levels could be

$$E_1 = E_0 + A$$
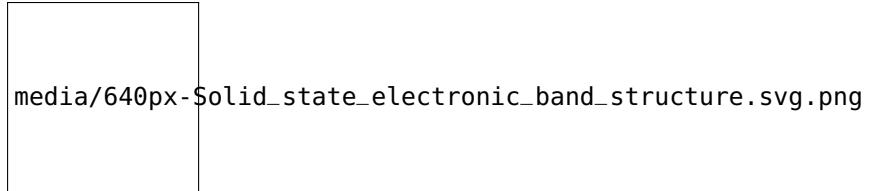
and

$$E_2 = E_0 - A$$

In a silicon crystal we can have trillions of atoms, and those that are close, have states that interact. **That's why crystals stay solids**. All chemical bonds are states of electrons interacting! Some are

strong (co-valent bonds), some are weaker (ionic bonds), but it's all quantum states interacting.
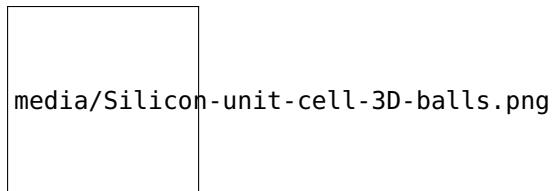
The discrete energy levels of the electron transition into bands of allowed energy states.

media/640px-Solid_state_electronic_band_structure.svg.png

For a crystal, the allowed energy bands is captured in the band structure
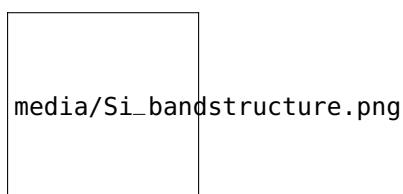
4.8 Silicon Unit Cell

A silicon crystal unit cell is a diamond faced cubic with 8 atoms in the corners spaced at 0.543 nm, 6 at the center of the faces, and 4 atoms inside the unit cell at a nearest neighbor distance of 0.235 nm.

media/Silicon-unit-cell-3D-balls.png

4.9 Band structure

The full band structure of a silicon unit cell is complicated, it's a 3 dimensional concept

media/Si_bandstructure.png

4.10 Valence band and Conduction band

For bulk silicon we simplify, and we think of two bands, the conduction band, and valence band

In the conduction band (E_C) is the lowest energy where electrons are free (not bound to atoms). The valence band (E_V) is the highest band where electrons are bound to silicon atoms.

The difference between E_C and E_V is a property of the material we've named the band gap.

$$E_G = E_C - E_V$$

4.11 Fermi level

From Wikipedia's Fermi level

> In band structure theory, used in solid state physics to analyze the energy levels in a solid, the Fermi level can be considered to be a hypothetical energy level of an electron, such that at thermodynamic equilibrium this energy level would have a 50% probability of being occupied at any given time

The Fermi level is closely linked to the Fermi-Dirac distribution
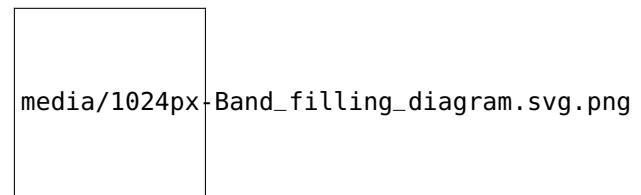
$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}$$

If the energy of the state is more than a few kT away from the Fermi-level, then

$$f(E) \approx e^{(E_F-E)/kT}$$

4.12 Metals

In metals, the band splitting of the energy levels causes the valence band and conduction band to overlap.

media/1024px-Band_filling_diagram.svg.png

As such, electrons can easily transition between bound state and free state. As such, electrons in metals are shared over large distances, and there are many electrons readily available to move under an applied field, or difference in electron density. That's why metals conduct well.

4.13 Insulators

In insulating materials the difference between the conduction band and the valence band is large. As a result, it takes a large energy to excite electrons to a state where they can freely move.

That's why glass is transparent to optical frequencies. Visible light does not have sufficient energy to excite electrons from a bound state.

That's also why glass is opaque to ultra-violet, which has enough energy to excite electrons out of a bound state.

Based on these two pieces of information you could estimate the bandgap of glass.

4.14 Semiconductors
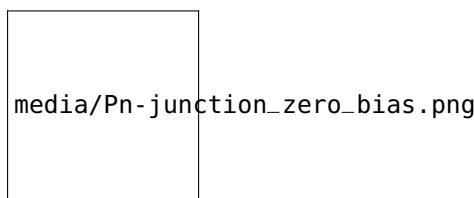
In a silicon the bandgap is lower than an insulator, approximately

$$E_G = 1.12 \, eV$$

At room temperature, that allows a small number of electrons to be excited into the conduction band, leaving behind a "hole" in the valence band.

4.15 Band diagrams

A band diagram or energy level diagrams shows the conduction band energy and valence band energy as a function of distance in the material.

media/Pn-junction_zero_bias.png

The horizontal axis is the distance, the vertical axis is the energy.

The figure shows a PN-junction

4.16 Density of electrons/holes

There are two components needed to determine how many electrons are in the conduction band. The density of available states, and the probability of an electron to be in that quantum state.

The probability is the Fermi-Dirac distribution. The density of available states is a complicated calculation from the band-structure of silicon. for details.

$$n_e = \int_{E_C}^{\infty} N(E)f(E)dE$$

The Fermi level is assumed to be independent of energy level, so we can write

$$n_e = e^{E_F/kT} \int_{E_C}^{\infty} N(E)e^{-E/kT}dE$$

for the density of electrons in the conduction band.

4.17 Fields

There are equations that relate electric field, magnetic field, charge density and current density to each-other.

$$\oint_{\partial \Omega} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \iiint_V \rho \cdot dV$$

,relates net electric flux to net enclosed electric charge

$$\oint_{\partial \Omega} \mathbf{B} \cdot d\mathbf{S} = 0$$

,relates net magnetic flux to net enclosed magnetic charge

$$\oint_{\partial \Sigma} \mathbf{E} \cdot d\ell = -\frac{d}{dt} \iint_{\Sigma} \mathbf{B} \cdot d\mathbf{S}$$

,relates induced electric field to changing magnetic flux

$$\oint_{\partial \Sigma} \mathbf{B} \cdot d\ell = \mu_0 \left(\iint_{\Sigma} \mathbf{J} \cdot d\mathbf{S} + \epsilon_0 \frac{d}{dt} \iint_{\Sigma} \mathbf{E} \cdot d\mathbf{S} \right)$$

,relates induced magnetic field to changing electric flux and to current

These are the Maxwell Equations, and are non-linear time dependent differential equations.

Under the best of circumstances they are fantastically hard to solve! But it's how the real world works.

The permittivity of free space is defined as

$$\epsilon_0 = \frac{1}{\mu_0 c^2}$$

, where c is the speed of light, and μ_0 is the vacuum permeability, which, in SI units, is now

$$\mu_0 = \frac{2\alpha}{q^2}\frac{h}{c}$$

, where α is the fine structure constant.

4.18 Voltage

The electric field has units voltage per meter, so the electric field is the derivative of the voltage as a function of space.

$$E = \frac{dV}{dx}$$

4.19 Current

Current has unit A and charge C has unit As, so the current is the number of charges passing through a volume per second.

The current density J has units A/m^2 and is often used, since we can multiply by the surface area of a conductor, if the current density is uniform.

$$I = A \times J$$

4.20 Drift current

Charges in an electric field will give rise to a drift current.

We know from Newtons laws that force equals mass times acceleration

$$\vec{F} = m\vec{a}$$

If we assume a zero, or constant magnetic field, the force on a particle is

$$\vec{F} = q\vec{E}$$

The current density is then

$$\vec{J} = q\vec{E} \times n \times \mu$$

where n is the charge density, and μ is the mobility (how easily the charges move) and has units $[m^2/Vs]$

Assuming

$$E = V/m$$

, we could write

$$J = \frac{C}{m^3}\frac{V}{m}\frac{m^2}{Vs} = \frac{C}{s}m^{-2}$$

So multiplying by an area

$$A = Bm^2$$

$$I = qn\mu BV$$

and we can see that the conductance

$$G = qn\mu B$$

, and since

$$G = 1/R$$

, where R is the resistance, we have

$$I = GV \Rightarrow V = RI$$

Or Ohms law

4.21 Diffusion current

A difference in charge density will give rise to a diffusion current, and the current density is

$$J = -qD_n \frac{d\rho}{dx}$$

,where D_n is a diffusion constant, and ρ is the charge density.

4.22 Why are there two currents?

I struggled with the concepts diffusion current and drift current for a long time. Why are there two types of current? It was when I read The Schrödinger Equation in a Classical Context: A Seminar on Superconductivity I realised that the two types of current come directly from the Schrodinger equation, there is one component related to the electric field (potential energy) and a component related to the momentum (kinetic energy).

In the absence of an electric field electrons will still jump from state to state set by the probabilities of the Hamiltonian. If there are more electrons in an area, then it will seem like there is an average movement of charges away from that area. That's how I think about the equation above. We can kinda see it from the Schrödinger equation below.

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial^2 x}\psi(x,t) + V(x)\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t)$$

4.23 Currents in a semiconductor

Both electrons, and holes will contribute to current.

Electrons move in the conduction band, and holes move in the valence band.

Both holes and electrons can only move if there are available quantum states.

For example, if the valence band is completely filled (all states filled), then there can be no current.

To compute the total current in a semiconductor one must compute

$$I = I_{n_{drift}} + I_{n_{diffusion}} + I_{p_{drift}} + I_{p_{diffusion}}$$

where n denotes electrons, and p denote holes.

4.24 Resistors

We can make resistors with metal and silicon (a semiconductor)

In metal the dominant carrier depends on the metal, but it's usually electrons. As such, one can often ignore the hole current.

In a semiconductor the dominant carrier depends on the Fermi level in relation to the conduction band and valence band. If the Fermi level is close to the valence band the dominant carrier will be holes. If the Fermi level is close to the conduction band, the dominant carrier will be electrons.

That's why we often talk about "majority carriers" and "minority carriers", both are important in semiconductors.

4.25 Capacitors

A capacitor resists a change in voltage.

$$I = C\frac{dV}{dt}$$

and store energy in an electric field between two conductors with an insulator between.

4.26 Inductors

An inductor resist a change in current.

$$V = L\frac{dI}{dt}$$

and store energy in the magnetic fields in a loop of a conductor.

Diodes | 5

5.1 Why

Diodes are a magical * semiconductor device that conduct current in one direction. It's one of the fundamental electronics components, and it's a good idea to understand how they work.

If you don't understand diodes, then you won't understand transistors, neither bipolar, or field effect transistors.

A useful feature of the diode is the exponential relationship between the forward current, and the voltage across the device.

To understand why a diode works it's necessary to understand the physics behind semiconductors.

This paper attempts to explain in the simplest possible terms how a diode works †

5.2 Silicon

Integrated circuits use single crystalline silicon. The silicon crystal is grown with the Czochralski method which forms a ingot that is cut into wafers. The wafer is a regular silicon crystal, although, it is not perfect.

A silicon crystal unit cell, as seen in Figure 1 is a diamond faced cubic with 8 atoms in the corners spaced at 0.543 nm, 6 at the center of the faces, and 4 atoms inside the unit cell at a nearest neighbor distance of 0.235 nm.

---

* It doesn't stop being magic just because you know how it works. Terry Pratchett, The Wee Free Men

† Simplify as much as possible, but no more. Einstein
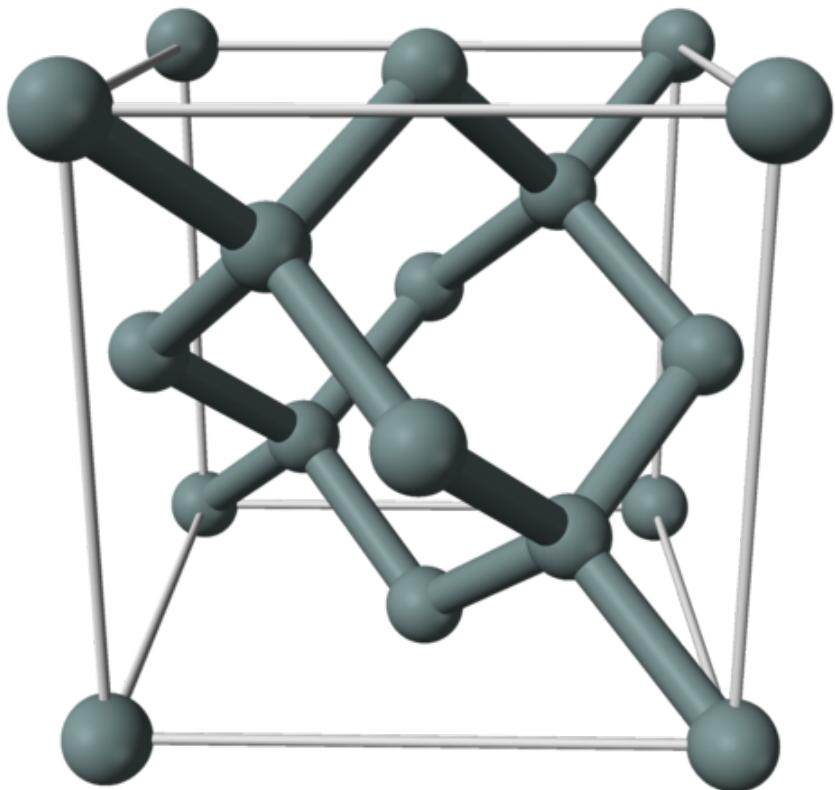
Figure 1: Silicon crystal unit cell

As you hopefully know, the energy levels of an electron around a positive nucleus are quantized, and we call them orbitals (or shells). For an atom far away from any others, these orbitals, and energy levels are distinct. As we bring atoms closer together, the orbitals start to interact, and in a crystal, the distinct orbital energies split into bands of allowed energy states. No two electrons, or any Fermion (spin of 1/2), can occupy the same quantum state. We call the outermost "shared" orbitial, or band, in a crystal the valence band. Hence covalent bonds.

If we assume the crystal is perfect, then at 0 Kelvin all electrons will be part of covalent bonds. Each silicon atom share 4 electrons with its neighbors. I think what we really mean when we say "share 4 electrons" is that the wave-functions of the outer orbitals interact, and we can no longer think of the orbitals as belonging to either of the silicon nuclei. All the neighbors atoms "share" electrons, and nowhere is there an vacant state, or a hole, in the valence band. If such a crystal were to exist, it would not conduct any current, as the charges cannot move.

In a atom, or a crystal, there are also higher energy states where the carriers are "free" to move. We call these energy levels, or bands of energy levels, conduction bands. In singular form "conduction

band", refers to the lowest available energy level where the electrons are free to move.

Due to imperfectness of the silicon crystal, and non-zero temperature, there will be some electrons that achieve sufficient energy to jump to the conduction band. The electrons in the conduction band leave vacant states, or holes, in the valence band.

Electrons can move both in the conduction band, as free electrons, and in the valence band, as a positive particle, or hole.

5.3 Intrinsic carrier concentration

The intrinsic carrier concentration of silicon, or how many free electrons and holes at a given temperature, is given by

$$n_i = \sqrt{N_c N_v} e^{-\frac{E_g}{2kT}} \qquad (1)$$

where E_g is the bandgap energy of silicon (approx 1.12 eV), k is Boltzmann's constant, T is the temperature in Kelvin, N_c is the density of states in conduction band, and N_v is the density of states in the valence band.

The density of states are

$$N_c = 2 \left[\frac{2\pi kT m_n^*}{h^2} \right]^{3/2} \quad N_v = 2 \left[\frac{2\pi kT m_p^*}{h^2} \right]^{3/2}$$

where h is Planck's constant, m_n^* is the effective mass of electrons, and m_p^* is the effective mass of holes.

In [1] they claim the intrinsic carrier concentration is a constant, although they do mention n_i doubles every 11 degrees Kelvin.

In BSIM 4.8 [2] the intrinsic carrier concentration is

$$n_i = 1.45e10 \frac{TNOM}{300.15} \sqrt{\frac{T}{300.15}} \exp^{21.5565981 - \frac{E_g}{2kT}}$$

Comparing the three models in Figure 2, we see the shape of BSIM and the full equation is almost the same, while the "doubling every 11 degrees" is just wrong.
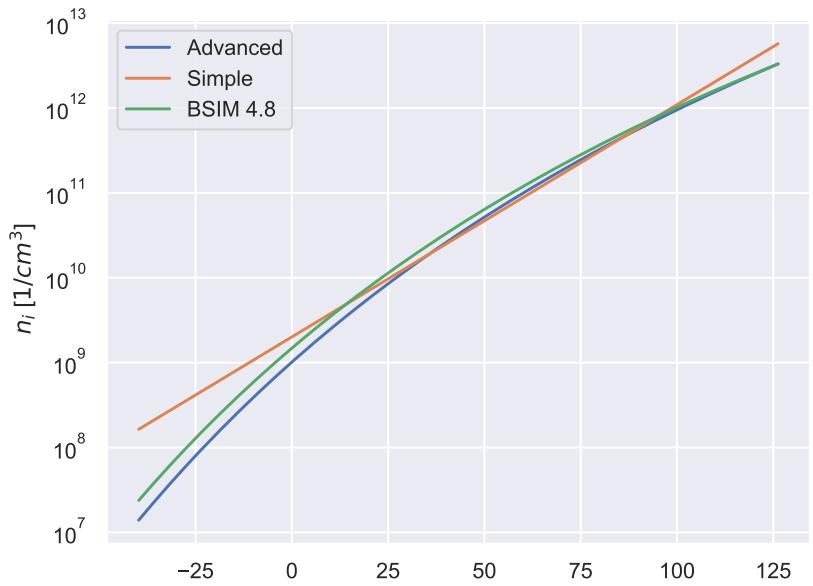
Figure 2: Intrinsic carrier concentration versus temperature

At room temperature the intrinsic carrier consentration is approximately $n_i = 1 \times 10^{16}$ carriers/m^3.

That may sound like a big number, however, if we calculate the electrons per um^3 it's $n_i = \frac{1 \times 10^{16}}{(1 \times 10^6)^3}$ carriers/μm$^3 < 1$, so there are really not that many free carriers in intrinsic silicon.

But where does Eq (1) come from? I find it unsatisfying if I don't understand where things come from. I like to understand why there is an exponential, or effective mass, or Planck's constant. If you're like me, then read the next section. If you don't care, and just want to memorize the equations, or indeed the number of intrinsic carrier concentration number at room temperature, then skip the next section.

5.4 It's all quantum

There are two components needed to determine how many electrons are in the conduction band. The density of available states, and the probability of an electron to be in that quantum state.

For the density of states we must turn to quantum mechanics. The probability amplitude of a particle can be described as

$$\psi = Ae^{i(k\mathbf{r} - \omega t)}$$

where k is the wave number, and ω is the angular frequency, and $\mathbf{r}$ is a spatial vector.

In one dimension we could write $\psi(x,t) = Ae^{i(kx-\omega t)}$

In classical physics we described the Energy of the system as

$$\frac{1}{2m}p^2 + V = E$$

where $p = mv$, m is the mass, v is the velocity and V is the potential.

In the quantum realm we must use the Schrodinger equation to compute the time evolution of the Energy, in one space dimension

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial^2 x}\psi(x,t) + V(x)\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t)$$

where m is the mass, V is the potential, $\hbar = h/2\pi$.

We could rewrite the equation above as

$$\widehat{H}\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t) = \widehat{E}\psi(x,t)$$

where $\widehat{H}$ is sometimes called the *Hamiltonian* and is an operator, or something that act on the wave-function. I recently read Feynman's Lectures on Physics, and Feynman called the Hamiltonian the *Energy Matrix* of a system. I like that better. The $\widehat{E}$ is the energy operator, something that operates on the wave-function to give the Energy.

We could re-arrange

$$[\widehat{H} - \widehat{E}]\psi(r,t) = 0$$

This is an equation with at least 5 unknowns, the space vector in three dimensions, time, and the energy matrix $\widehat{H}$. It turns out, that the energy matrix depends on the system. The energy matrix further up is for one free electron. For an atom, the energy matrix will have more dimensions to describe the possible quantum states.

I was watching Quantum computing in the 21st Century and David Jamison mentioned that the largest system we could today compute would be a system with about 30 electrons. So although we know exactly how the equations of quantum mechanics appear to be, and they've proven extremely successful, we must make simplifications before we can predict how electrons behave in complicated systems

like the silicon lattice with approximately 0.7 trillion electrons per cube micro meter. You can check the calculation

$$\left[\frac{1\ \mu m}{0.543\ nm}\right]^3 \times 8 \text{ atoms per unit cell} \times 14 \text{ electrons per atom}$$

5.4.1 Density of states

To compute "how many Energy states are there per unit volume in the conduction band", or the "density of states", we start with the three dimensional Schrodinger equation for a free electron

$$-\frac{\hbar^2}{2m}\nabla^2\psi = E\psi$$

I'm not going to repeat the computation here, but rather paraphrase the steps. You can find the full derivation in Solid State Electronic Devices.

The derivation starts by computing the density of states in the k-space, or momentum space,

$$N(dk) = \frac{2}{(2\pi)^p}dk$$

Where p is the number of dimensions (in our case 3).

Then uses the band structure $E(k)$ to convert to the density of states as a function of energy $N(E)$. The simplest band structure, and a approxmiation of the lowest conduction band is

$$E(k) = \frac{\hbar^2 k^2}{2m^*}$$

where m^* is the effective mass of the particle. It is within this effective mass that we "hide" the complexity of the actual three-dimensional crystal structure of silicon.

The effective mass when we compute the density of states is

$$m^* = \frac{\hbar^2}{\frac{d^2E}{dk^2}}$$

as such, the effective mass depends on the localized band structure of the silicon unit cell, and depends on direction of movement, strain of the silicon lattice, and probably other things.

In 3D, once we use the above equations, one can compute that the density of states per unit energy is

$$N(E)dE = \frac{2}{\pi^2}\frac{m^{*3/2}}{\hbar^2}E^{1/2}dE$$

In order to find the number of electrons, we need the probability of an electron being in a quantum state, which is given by the Fermi-Dirac distribution

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \tag{2}$$

where E is the energy of the electron, E_F is the Fermi level or checmical potential, k is Boltzmann's constant, and T is the temperature in Kelvin.

Fun fact, the Fermi level difference between two points is what you measure with a voltmeter.

If the $E - E_F > kT$, then we can start to ignore the +1 and the probability reduces to

$$f(E) = \frac{1}{e^{(E-E_F)/kT}} = e^{(E_F-E)/kT}$$

A few observation on the Fermi-Dirac distribution. If the Energy of a particle is at the Fermi level, then $f(E) = \frac{1}{2}$, or a 50 % probability.

In a metal, the Fermi level lies within a band, as the conduction band and valence band overlap. As a result, there are a bunch of free electrons that can move around. Metal does not have the same type of covalent bonds as silicon, but electrons are shared between a large part of the metal structure. I would also assume that the location of the Fermi level within the band structure explains the difference in conductivity of metals, as it would determined how many electrons are free to move.

In an insulator, the Fermi level lies in the bandgap between valence band and conduction band, and usually, the bandgap is large, so there is a low probability of finding electrons in the conduction band.

In a semiconductor we also have a bandgap, but much lower energy than an insulator. If we have thermal equilibrium, no external forces, and we have an un-doped (intrinsic) silicon semiconductor, then the fermi level E_F lies half way between the conduction band edge E_C and the valence band edge E_V.

The bandgap is defined as the $E_C - E_V = E_g$, and we can use that to get $E_F - E_C = E_C - E_g/2 - E_C = -E_g/2$. This is why the bandgap of silicon keeps showing up in our diode equations.

The number of electrons per delta energy will then be given by

$$N_e dE = N(E) f(E) dE$$

, which can be integrated to get

$$n_e = 2 \left(\frac{2\pi m^* kT}{h^2} \right)^{3/2} e^{(E_F - E_C)/kT}$$

For intrinsic silicon at thermal equlibrium, we could write

$$n_0 = 2 \left(\frac{2\pi m^* kT}{h^2} \right)^{3/2} e^{-E_g/(2kT)} \tag{3}$$

As we can see, Equation (3) has the same coefficients and form as the computation in Equation (1). The difference is that we also have to account for holes. At thermal equilibrium and intrinsic silicon $n_i^2 = n_0 p_0$.

5.4.2 How to think about electrons (and holes)

I've come to the realization that to imagine electrons as balls moving around in the silicon crystal is a bad mental image.

For example, for a metal-oxide-semiconductor field effect transistor (MOSFET) it is not the case that the electrons that form the inversion layer under strong inversion come from somewhere else. They are already at the silicon surface, but they are bound in covalent bonds (there are literaly trillions of bound electrons in a typical transistor).

What happens is that the applied voltage at the gate shifts the energy bands close to the surface (or bends the bands in relation to the Fermi level), and the density of carriers in the conduction band in that location changes, according to the type of derivations above.

Once the electrons are in the conduction band, then they follow the same equations as diffusion of a gas, Fick's law of diffusion. Any charge concentration difference will give rise to a diffusion current given by

$$J_{\text{diffusion}} = -q D_n \frac{\partial \rho}{\partial x} \tag{4}$$

where J is the current density, q is the charge, ρ is the charge density, and D is a diffusion coefficient that through the Einstein relation can be expressed as $D = \mu kT$, where mobility $\mu = v_d/F$ is the ratio of drift velocity v_d to an applied force F.

To make matters more complicated, an inversion layer of a MOSFET is not in three dimensions, but rather a two dimensional electron gas, as the density of states is confined to the silicon surface. As such, we should not expect the mobility of bulk silicon to be the same as the mobility of a MOSFET transistor.

5.5 Doping

We can change the property of silicon by introducing other elements, something we've called doping. Phosphor has one more electron than silicon, Boron has one less electron. Injecting these elements into the silicon crystal lattice changes the number of free electron/holes.

These days, we usually dope with ion implantation, while in the olden days, most doping was done by diffusion. You'd paint something containing Boron on the silicon, and then heat it in a furnace to "diffuse" the Boron atoms into the silicon.

If we have an element with more electrons we call it a donor, and the donor concentration N_D.

The main effect of doping is that it changes the location of the Fermi level at thermal equilibirum. For donors, the Fermi level will shift closer to the conduction band, and increase the probabilty of free electrons, as determined by Equation (2).

Since the crystal now has an abundance of free electrons, which have negative charge, we call it n-type.

If the element has less electrons we call it an acceptor, and the acceptor concentration N_A. Since the crystal now has an abundance of free holes, we call it p-type.

The doped material does not have a net charge, however, as it's the same number of electrons and protons, so even though we dope silicon, it does remain neutral.

The doping concentrations are larger than the intrinsic carrier concentration, from maybe 10^{21} to 10^{27} carriers/m^3. To separate between these concentrations we use $p-, p, p+$ or $n-, n, n+$.

The number of electrons and holes in a n-type material is

$$n_n = N_D \,,\, p_n = \frac{n_i^2}{N_D}$$

and in a p-type material

$$p_p = N_A \,,\, n_p = \frac{n_i^2}{N_A}$$

In a p-type crystal there is a majority of holes, and a minority of electrons. Thus we name holes majority carriers, and electrons minority carriers. For n-type it's opposite.

5.6 PN junctions

Imagine an n-type material, and a p-type material, both are neutral in charge, because they have the same number of electrons and protons. Within both materials there are free electrons, and free holes which move around constantly.

Now imagine we bring the two materials together, and we call where they meet the junction. Some of the electrons in the n-type will wander across the junction to the p-type material, and visa versa. On the opposite side of the junction they might find an opposite charge, and might get locked in place. They will become stuck.

After a while, the diffusion of charges across the junction creates a depletion region with immobile charges. Where as the two materials used to be neutrally charged, there will now be a build up of negative charge on the p-side, and positive charge on the n-side.

5.6.1 Built-in voltage

The charge difference will create a field, and a built-in voltage will develop across the depletion region.

The density of free electrons in the conduction band is

$$n = \int_{E_C}^{\infty} N(E)f(E)dE$$

, where $N(E)$ is the density of states, and $f(E)$ is a probability of a electron being in that state (Equation (2)).

We could write the density of electrons on the n-side as

$$n_n = e^{E_{Fn}/kT} \int_{E_C}^{\infty} N_n(E)e^{-E/kT} dE$$

since the Fermi level is independent of the energy state of the electrons (I think).

The density of electrons on the p-side could be written as

$$n_p = e^{E_{Fp}/kT} \int_{E_C}^{\infty} N_p(E)e^{-E/kT} dE$$

If we assume that the density of states, $N_n(E)$ and $N_p(E)$ are the same, and the temperature is the same, then

$$\frac{n_n}{n_p} = \frac{e^{E_{Fn}/kT}}{e^{E_{Fp}/kT}} = e^{(E_{Fn}-E_{Fp})/kT}$$

The difference in Fermi levels is the built-in voltage multiplied by the unit charge.

$$E_{F_n} - E_{F_p} = q\Phi$$

and by substituting for the minority carrier concentration on the p-side we get

$$\frac{N_A N_D}{n_i^2} = e^{q\Phi_0/kT}$$

or rearranged to

$$\Phi_0 = \frac{kT}{q} ln \left(\frac{N_A N_D}{n_i^2} \right)$$

5.6.2 Current

The derivation of current is a bit involved, but let's try.

The hole concentration on the p-side and n-side could be written as

$$\frac{p_p}{p_n} = e^{-q\Phi_0/kT}$$

The negative sign is because the built in voltage is positive on the n-type side

Asssume that $-x_{p0}$ is the start of the junction on the p-side, and x_{n0} is the start of the junction on the n-side.

Assume that we lift the p-side by a voltage qV

Then the hole concentration would change to

$$\frac{p(-x_{p0})}{p(x_{n0})} = e^{q(V-\Phi_0)/kT}$$

while on the n-side the hole concentration would be

$$\frac{p(x_{n0})}{p_n} = e^{qV/kT}$$

So the excess hole concentration on the n-side due to an increase of V would be

$$\Delta p_n = p(x_{n0}) - p_n = p_n \left(e^{qV/kT} - 1 \right)$$

The diffusion current density, given by Equation (4) states

$$J(x_n) = -qD_p \frac{\partial \rho}{\partial x}$$

Thus we need to know the charge density as a function of x. I'm not sure why, but apparently it's

$$\partial \rho(x_n) = \Delta p_n e^{-x_n/L_p}$$

where L_p is a diffusion length. This equation smells to me like a simplified model of reality, I'm not sure how much it's based on fundamental physics.

Anyhow, we can now compute the current density, and need only compute it for $x_n = 0$, so you can show it's

$$J(0) = q \frac{D_p}{L_p} p_n \left(e^{qV/kT} - 1 \right)$$

which start's to look like the normal diode equation. The p_n is the minority concentration of holes on the n-side, which we've before estimated as $p_n = \frac{n_i^2}{N_D}$

We've only computed for holes, but there will be electron transport from the p-side to the n-side also.

We also need to multiply by the area of the diode to get current from current density. The full equation thus becomes

$$I = qAn_i^2 \left(\frac{1}{N_A} \frac{D_n}{L_n} + \frac{1}{N_D} \frac{D_p}{L_p} \right) \left[e^{qV/kT} - 1 \right]$$

where A is the area of the diode, D_n, D_p is the diffusion coefficient of electrons and holes and L_n, L_p is the diffusion length of electrons and holes.

Which we usually write as

$$I_D = I_S(e^{\frac{V_D}{V_T}} - 1), \text{ where } V_T = kT/q$$

5.6.3 Forward voltage temperature dependence

We can rearrange I_D equation to get

$$V_D = V_T \ln \left(\frac{I_D}{I_S} \right)$$

and at first glance, it appears like V_D has a positive temperature coefficient. That is, however, wrong.

First rewrite

$$V_D = V_T \ln I_D - V_T \ln I_S$$

$$\ln I_S = 2 \ln n_i + \ln Aq \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right)$$

Assume that diffusion coefficient ‡, and diffusion lengths are independent of temperature.

That leaves n_i that varies with temperature.

$$n_i = \sqrt{B_c B_v} T^{3/2} e^{\frac{-E_g}{2kT}}$$

where

$$B_c = 2 \left[\frac{2\pi k m_n^*}{h^2} \right]^{3/2} \quad B_v = 2 \left[\frac{2\pi k m_p^*}{h^2} \right]^{3/2}$$

---

‡ From the Einstein relation $D = \mu kT$ it does appear that the diffusion coefficient increases with temperature, however, the mobility decreases with temperature. I'm unsure of whether the mobility decreases with the same rate though.

$$2 \ln n_i = 2 \ln \sqrt{B_c B_v} + 3 \ln T - \frac{V_G}{V_T}$$

with $V_G = E_G/q$ and inserting back into equation for V_D

$$V_D = \frac{kT}{q}(\ell - 3 \ln T) + V_G$$

Where ℓ is temperature independent, and given by

$$\ell = \ln I_D - \ln \left(Aq \frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right) - 2 \ln \sqrt{B_c B_v}$$

From equations above we can see that at 0 K, we expect the diode voltage to be equal to the bandgap of silicon. Diodes don't work at 0 K though.

Although it's not trivial to see that the diode voltage has a negative temperature coefficient, if you do compute it as in vd.py, then you'll see it decreases.

The slope of the diode voltage can be seen to depend on the area, the current, doping, diffusion constant, diffusion length and the effective masses.

Figure 3 shows the V_D and the deviation of V_D from a straight line. The non-linear component of V_D is only a few mV. If we could combine V_D with a voltage that increased with temperature, then we could get a stable voltage across temperature to within a few mV.
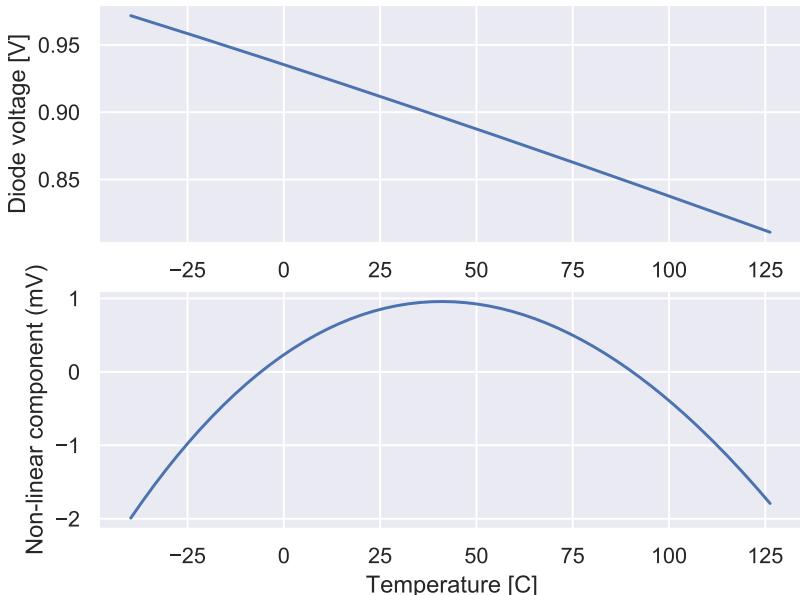
Figure 3: Diode forward voltage as a function of temperature

5.6.4 Current proportional to temperature

Assume we have a circuit like Figure 4.

Here we have two diodes, biased at different current densities. The voltage on the left diode V_{D1} is equal to the sum of the voltage on the right diode V_{D2} and voltage across the resistor R_1. The current in the two diodes are the same due to the current mirror. A such, we have that

$$I_S e^{\frac{qV_{D1}}{kT}} = NI_S e^{\frac{qV_{D2}}{kT}}$$

Taking logarithm of both sides, and rearranging, we see that

$$V_{D1} - V_{D2} = \frac{kT}{q} \ln N$$

Or that the difference between two diode voltages biased at different current densities is proportional to absolute temperature.

In the circuit above, this ΔV_D is across the resistor R_1, as such, the $I_D = \Delta V_D / R_1$. We have a current that is proportional to temperature.

If we copied the current, and sent it into a series combination of a resistor R_2 and a diode, we could scale the R_2 value to give us the exactly right slope to compensate for the negative slope of the V_D voltage.

The voltage across the resistor and diode would be constant over temperature, with the small exception of the non-linear component of V_D.
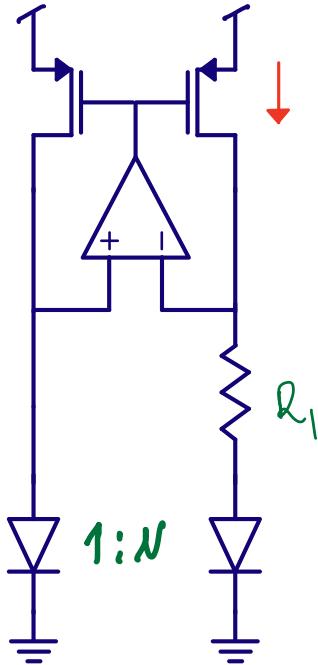
Figure 4: Circuit to generate a current proportional to kT

5.7 Equations aren't real

Nature does not care about equations. It just is.

We know, at the fundamental level, nature appears to obey the mathematics on quantum mechanics, however, due to the complexity of nature, it's not possible today (which is not the same as impossible), to compute exactly how the current in a diode works. We can get close, by measuring a diode we know well, and hope that the next time we make the same diode, the behavior will be the same.

As such, I want to warn you about the "lies" or "simplifications" we tell you. Take the diode equation above, some parts, like the intrinsic carrier concentration n_i has roots directly from quantum mechanics, with few simplifications, which means it's likely solid truth, at least for a single unit cell.

But there is no reason nature should make all unit cells the same, and infact, we know they are not the same, we put in dopants. As we scale down to a few nano-meter transistors the simplification that "all unit cells of silicon are the same, and extend to infinity" is

no longer true, and must be taken into account in how we describe reality.

Other parts, like the exact value of the bandgap E_g, the diffusion constant D_p or diffusion length L_p are macroscopic phenomena, we can't expect them to be 100 true. The values would be based on measurement, but not always exact, and maybe, if you rotate your diode, they would be different.

You should realize that the consequence of our imperfection is that the equations in electronics should always be taken with a grain of salt.

Nature does not care about your equations. Nature will easily have the superposition of trillions of electrons, and they don't have to agree with your equations.

But most of the time, the behavior is similar.

References

[1]   T. C. Carusone, D. Johns, and K. Martin, *Analog integrated circuit design*. Wiley, 2011 [Online]. Available: https://books.google.no/books?id=1OIJZzLvVhcC

[2]   Berkeley, "Berkeley short-channel IGFET model." [Online]. Available: http://bsim.berkeley.edu/models/bsim4/

Mosfets | 6

ESD and IC $\Big|$ 8

References and bias | 9

Analog frontend and filter 10

Switched capacitor circuits **11**

Oversampling and Sigma-Delta ADCs 12

Voltage Regulation 13

Clocks and PLLs 14

Oscillators | 15

Low Power Radio | 16

Analog SystemVerilog | **17**

Energy Sources 18