

---

# **Beyond CMOS Notebook**

**Review Edition**

---

Luka Bostick

LukaBostick@my.unt.edu

Beginning 15 November 2023





L   B   T   M

Research and Development

# Contents

0.1 Acknowledgments . . . . .	1
<b>March 6, 2024</b>	<b>3</b>
<b>Sustainable Computing A Quantitative Perspective</b>	<b>4</b>
0.2 Sustainable Development . . . . .	4
0.3 Pillars of Sustainability . . . . .	4
0.4 Sustainable Computing . . . . .	4
0.5 Achieving computational sustainability . . . . .	6
0.6 CO <sub>2</sub> emissions per capita and global warming trend . . . . .	8
0.7 Legally binding international treaty on climate change . . . . .	10
0.8 The cost and efficiency of computing . . . . .	11
0.9 What are the practices . . . . .	12
0.10 Greenhouse Gases based on the GHG Protocol . . . . .	12
0.11 Foundational, First, and Second Order Trends . . . . .	14
0.12 Phases in Hardware Lifecycle . . . . .	16
0.13 Employs Complex, Often Dirty, and Energy-Hungry Processes . . . . .	17
0.14 The global chip industry has a colossal problem with carbon emissions! . . . . .	18
0.15 Keeping energy use steady despite large growth in demand . . . . .	19
0.16 Carbon emission based on region and in comparison with fossil fuel usage . . . . .	24
0.17 Energy usage and efficiency on personal and small business levels . . . . .	25
0.18 Apple's Environmental Progress Report, 2022 . . . . .	28
0.19 Response to the Climate Crisis . . . . .	29
0.20 Absorbing Carbon by Regenerating Nature — Removing Carbon Through Direct Capture . . . . .	30
0.21 How you can improve the efficiency of our computing infrastructure . . . . .	31
0.22 Using AI to Improve Buildings . . . . .	33
0.23 Using machine learning and AI for landscape connectivity and wildlife corridors . . . . .	35
0.24 How computing can address the challenges of poverty . . . . .	36
0.25 Switching energy and switching speed . . . . .	37
0.26 By shifting to alternative technologies such as superconductor electronics for specific applications, the energy cost can be significantly reduced . . . . .	37
0.27 Effect of IT infrastructure on equity and equality . . . . .	39
0.28 Transformation to a sustainable and resilient society that ensures the safety and security of its people . . . . .	41
0.29 Educating the new generation and workforce development/retraining . . . . .	42
<b>Introduction to Quantum Computing: Qubits, Gates, and Algorithms</b>	<b>44</b>
0.30 Introduction . . . . .	45
0.31 Computing Development Timeline . . . . .	46
0.32 How is a Quantum Computer Different? . . . . .	55
0.33 Classical and Quantum Bits . . . . .	58
0.34 Quantum Parallelism . . . . .	61
0.35 Quantum Interference . . . . .	68
0.36 Classical Boolean Logic Gates . . . . .	74

0.37 Single-Qubit Quantum Gates . . . . .	76
0.38 Two-qubit Quantum Gates . . . . .	77
0.39 Gate-Based Approach:Single-Qubit Operation . . . . .	78
0.40 Gate-Based Approach: Two-Qubit Controlled-NOT . . . . .	79
0.41 Quantum Algorithm (Universal QC) . . . . .	81
0.42 Coherence Gate Time . . . . .	85
0.43 Qubit Modalities (Extensible platforms, benchmarked, ca 2019) . . . . .	88
0.44 Natural and Artificial Atoms . . . . .	91
0.45 Superconducting Artificial Atoms . . . . .	96
0.46 Cryogenic Engineering . . . . .	101
0.47 Fabrication Engineering . . . . .	102
0.48 Microwave Engineering and control . . . . .	106
0.49 Engineering Improved Coherence . . . . .	109
0.50 Gate Fidelities . . . . .	110
0.51 Nascent Commercial Quantum Processors . . . . .	111
0.52 Architectural Layers of a QIP . . . . .	113
0.53 Lacrosse in the Presence of Noise . . . . .	114
0.54 Superconducting Qubit Arrays . . . . .	115
0.55 3D Integration for Quantum Processors IARPA Quantum Enhanced Optimization	118
<b>Compare and contrast: CMOS, NMOS, Josephson Logic</b>	<b>120</b>
0.56 Overview . . . . .	121
0.57 Static CMOS inverter . . . . .	122
0.58 Voltages and Logic levels . . . . .	123
0.59 Voltage waveforms and logic levels . . . . .	127
0.60 Logic level compatibility . . . . .	129
0.61 Saturating inverter model . . . . .	130
0.62 Gain and signal levels . . . . .	131
0.63 Gain and delay example . . . . .	132
0.64 Transistor-level view of-slow rising inputs . . . . .	134
0.65 Step vs. Ramp input . . . . .	136
0.66 Signal integrity . . . . .	136
0.67 Transmissions lines . . . . .	137
0.68 RC transmission lines . . . . .	137
0.69 Pulse propagation . . . . .	137
0.70 RC transmission lines cont . . . . .	139
0.71 RC section pulse response . . . . .	139
0.72 Logic gate reliability . . . . .	139
0.73 Error rates and energy . . . . .	140
0.74 Metastability in practice . . . . .	140
0.75 The vulnerable interval . . . . .	140
0.76 Example: metastability failure . . . . .	141
0.77 Metastability countermeasures . . . . .	141
0.78 Compare and contrast . . . . .	141
0.79 Lessons from history . . . . .	141
0.80 Generators . . . . .	142
0.81 MOS vs Superconducting generators . . . . .	143
0.82 Why generators for superconducting logic? . . . . .	143
0.83 Possible generators . . . . .	143
0.84 nMOS logic and PLA . . . . .	144

## Contents

0.85 PLA as FSM Mead and Conway . . . . .	145
0.86 Summary . . . . .	147
<b>Superconducting Quantum-classical Information Processing Systems</b>	<b>148</b>
0.87 SEEQC global footprint . . . . .	149
0.88 Operating: SEEQC System Red, full-stack reference quantum computer . . . . .	150
0.89 SEEQC System Red compared to state-of-the-art quantum platforms: public cloud data . . . . .	152
0.90 SEEQC System Red . . . . .	158
0.91 Revaluation of The State of Computing . . . . .	159
0.92 Manufactured by SEEQC Foundry . . . . .	161
0.93 Unit Fabrication Process . . . . .	161
0.94 Fabrication Process Specifications . . . . .	162
0.95 Standard Fabrication Processes . . . . .	162
0.96 What else is available . . . . .	162
0.97 How to go from present densely wired to chip-based system . . . . .	163
0.98 Rent's Rule . . . . .	163
0.99 Rent's Rule should be used for Quantum Circuits . . . . .	163
0.100 Present-Day Qubit Control: Analog, Microwave-based . . . . .	163
0.101 Present-Day Qubit Readout: Heterodyne Technique . . . . .	164
0.102 Increasing Difficulties to Scale . . . . .	165
0.103 SEEQC digital SFQ chip-based quantum computing architecture . . . . .	166
0.104 Resonant Excitation of Qubits by SFQ pulses . . . . .	170
0.105 Using superconducting SFQuClass circuits . . . . .	171
0.106 Flip-chip packaging of fluxonium qubits . . . . .	174
0.107 SEEQC digital SFQ controller chip integrated on multi-chip module with qubit chip (@ 20mK) . . . . .	175
0.108 Digital SFQ control . . . . .	177
0.109 Digital SFQ Control of Qubit . . . . .	178
0.110 Digital SFQ multiplexing of 1:4 signals . . . . .	179
0.111 Digital Readout: quantum-to-digital converter (QDC) . . . . .	180
0.112 JD PD – Josephson Digital Phase Detector *Patent Pending) . . . . .	180
0.113 Digital SFQ readout Josephson Digital Phase Detection (JD PD) . . . . .	181
0.114 Rent's Rule for Quantum Circuits cont . . . . .	182
0.115 Qubit Readout/Control: Analog vs Digital . . . . .	182
0.116 Scaling qubit control readout for quantum . . . . .	183
0.117 Advantages of SEEQC digital SFQ chips vs analog room temperature quantum control, readout multiplexing Confidential: Do not distribute without permission . . . . .	184
0.118 SEEQC chip compatibility across quantum technologies Confidential: Do not distribute without permission . . . . .	184
0.119 Competitor 4000 qubit system <sup>1</sup> ; cryo only, no room temp electronics SEEQC Confidential 2023: Do not distribute . . . . .	184
0.120 Competitors 1M qubit quantum data center Confidential: Do not distribute without permission . . . . .	185
0.121 SEEQC scaled quantum computing system: 100k-1m qubits . . . . .	185
0.122 Thank You! . . . . .	185
<b>I.A.A.S: Discover Expedition Education and Outreach</b>	<b>186</b>
0.123 Abstract . . . . .	187

## **0.1 Acknowledgments**

*Thanks to Dr. Timothy M. Pinkston, Dr. Murali Annavaram, and Dr. Sasan Razmkhah for providing me with a voice and a potential platform to share my input and thoughts on superconducting single flux quantum technology.*

**Special thanks to the review committee:**

*Contents*

Name	Institution	Qualifications	Email
Jonas Wagner	UT Dallas, UW-Platteville	Mechanical Engineering PhD Student — Engineering Physics and Electrical Engineering Bs	jonas.wagner.2826@gmail.com
Storm Mc.Cauley	UNT	Computer Science Bs student	storm.mccauley@gmail.com
Pavlo Bondarenko	UNT, Odessa Nat'l Polytechnic Univ.	Computer Science Bs student— Computer Engineering Bs	pavlobondarenko@my.unt.edu
Ethan Jensen	UNT, George Fox Univ.	Mathematics PhD Student — Mathematics and Computer Science Bs	ethanjensen@my.unt.edu
Dr. Jacob Hochstetler	UNT, Fidelity Investments	Computer Science and Engineering PhD — VP, Cloud Engineering at Fidelity	Jacob.Hochstetler@unt.edu
Dr. Austen Moss	UNT	Organic Chemistry PhD	austen.moss@unt.edu
Willy Vasquez	UT Austin, MIT	Cryptography PhD Student — Computer Science Ms	wrv@utexas.edu
Hannah Pil	NC State Univ.	Genetics PhD s	hdpil@ncsu.edu

Table 1: Review Committee Members

**March 6, 2024**

**1**

Text \_\_\_\_\_  
\_\_\_\_\_

# Sustainable Computing A Quantitative Perspective

Written by Massoud Pedram

©2023 Massoud Pedram All rights reserved.

University of Southern California

DISCoVER Expedition Seminar Series

Thanks to Dr. Mehdi Kamal and Dr. Sasan Razmkhah for their valuable inputs

Sustainability, discussion of the environment &development as one issue

## 0.2 Sustainable Development

There are many different definitions of sustainability; if we focus on sustainable development, it is defined as the kind of development that would allow us to meet our needs at present without compromising the ability of future generations to meet their own needs

## 0.3 Pillars of Sustainability

The essential pillars of sustainability are social equity, economic progress and prosperity, mental health support, and cultural sustainability.

## 0.4 Sustainable Computing

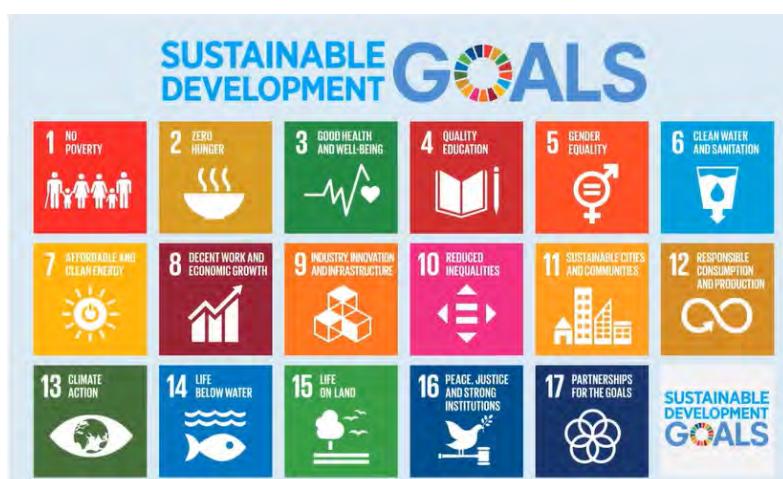


Figure 1:

UN Sustainable Development Goals<sup>1</sup>

- Our holistic life-cycle management approach stretches from power generation to waste removal to manufacturing to education and deploying IT across an organization or nation.

Sustainable computing embodies a comprehensive life-cycle management paradigm, from power generation to manufacturing diverse products, including equipment, instruments, and servers. It encompasses not only the utilization of these resources but also addresses the critical aspect of end-of-life disposal. Adopting a full life cycle asset-time view of the process is imperative to facilitate the deployment of information technology resources in our societal and digital ecosystems in an inherently sustainable manner.

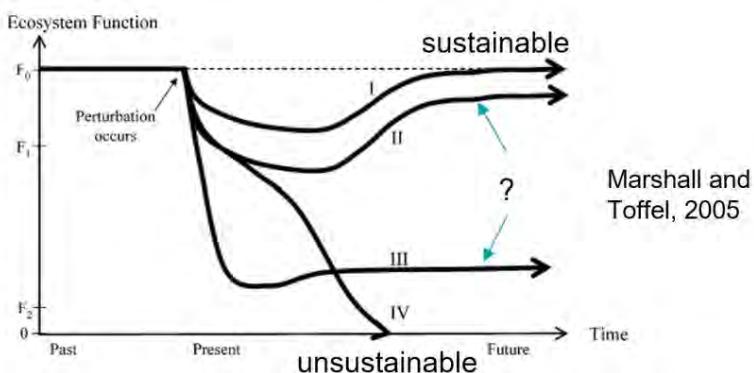


Figure 2:

Changes over time in ecosystem function from initial level  $F_0$ , in response to a perturbation. For line I, the ecosystem has fully recovered from the perturbation. Lines II and III dip below  $F_1$ , and ecosystem function has partially recovered. Line IV falls below  $F_2$ , and ecosystem function does not recover. The system will typically be considered sustainable if the response follows line I and not sustainable if the response follows line IV. If the response is projected to follow lines II or III, the system may not be considered sustainable.<sup>2</sup>

Above, you see the graph from the well-known Marchal and Toffel paper. You know what happens when we are not sustainable? I always like to look at what we are not so that we can understand what we are. An example of a non-sustainable behavior is shown in plot IV. In this graph, the ecosystem is functioning at some city-state level, and then something happens to the system, such as COVID-19, a shock that we all experienced in early 2020 (and even now to some extent). And if the system is resilient enough to recover and return to its normal functions, such as plot I or plot II, we are doing fine; if you end up with plot III or plot IV, you are probably not good. Plot IV is not good, and plot III means that there is a significant loss of productivity or capability to produce functions and services, and that would not be an example of a sustainable environment, be it economic, social, cultural, or what have you. So, there are many different goals for sustainability, some of which have a robust socioeconomic factor and some of which have to do with the environment. I don't need to go through the laundry list here, but you can read and see what we refer to No Poverty, zero hunger, good health, quality education, and so on.

<sup>1</sup>“Communications Materials - United Nations Sustainable Development.” United Nations, United Nations, [www.un.org/sustainabledevelopment/news/communications-material/](http://www.un.org/sustainabledevelopment/news/communications-material/). Accessed 30 Jan. 2024.

<sup>2</sup>Marshall, Julian D., and Michael W. Toffel. “Framing the elusive concept of sustainability: a sustainability hierarchy.” Environmental Science & Technology, vol. 39, no. 3, 22 Dec. 2004, pp. 673–682, <https://doi.org/10.1021/es040394k>.

## 0.5 Achieving computational sustainability

There are many different approaches one could use to achieve computational sustainability. Below is a short list of techniques one can use to ensure sustainability within a computing infrastructure.

**Reduce the energy consumption footprint of IT hardware** Of course, first and foremost, it would be to reduce the energy consumption footprint of the IT hardware itself. This, as I will explain later on, has to do not only with the footprint of the equipment, servers, and devices that are being used but also with when they are being manufactured and when they are being disposed of at the end of life. **Use less GHG-intensive sources of electricity** Try to use less greenhouse gas-intensive sources of electricity. Avoid coal, for example, or other types of energy sources that emit a lot of carbon **Increase the lifetime of IT hardware** Increase the lifetime of IT hardware; by doing so, of course, we prioritize the cost in manufacturing and end-of-life disposal over a longer period, and there, the overall energy efficiency of the IT hardware looked at on a full life-cycle basis increases. **Full life cycle and obsolescence management** Obsolescence management is another critical path. Can we recycle? If we dispose of something, can we do it safely where the environmental impact is minimal, and so on? **Avoid over-provisioning and over-design for performance** Avoid over-provisioning and over-design, which is a tendency that people have because we want to get the maximum performance out of our resources. We tend to over-provision when the system does not meet the higher performance. It was over-provisioned and over-designed, for it will create a lot of waste in terms of energy consumption. **Run-time adaptation and load management** Then, finally, run-time adaptation and load management, being able to respond dynamically to the actual conditions in which the systems are being used and adjust the system's behavior, in terms of energy consumption vs performance level, dynamically.

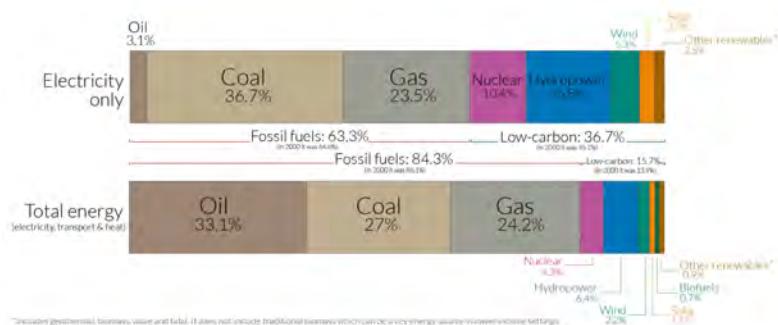


Figure 3:

More than one-third of global electricity comes from low-carbon sources, but a lot less of total energy does Global primary energy consumption by source

\*Includes geothermal, biomass, wave, and tidal. It does not include traditional biomass, which can be a key energy source in lower-income settings.<sup>3</sup>

So I thought to make things into a better perspective, we could look at the whole process of energy generation, consumption, trends, and so on before we can get into more details about IT

<sup>3</sup>Ritchie, Hannah. "Decarbonizing Electricity Is Only One Step towards a Low-Carbon Energy System." Our World in Data, 5 Nov. 2021, [ourworldindata.org/l-low-carbon-electricity](http://ourworldindata.org/l-low-carbon-electricity).

infrastructure and what we can do to make it more sustainable and energy efficient. So this is (above) a plot that shows the sources of generating energy, all kinds of energy, electricity, transport, heat, and electricity only, which we use in our data centers, to charge our batteries, and so forth. So they consistently show that dirty energy sources are coal, gas (of course, gas is better than coal), and oil. These are the dominant ways we generate energy today despite our recent trends to try to switch to renewable energy sources. Renewable energy sources, including hydropower, comprise about 25% of the total electricity generation sources today. **Global energy consumption and CO<sub>2</sub> emission trend by source and country**

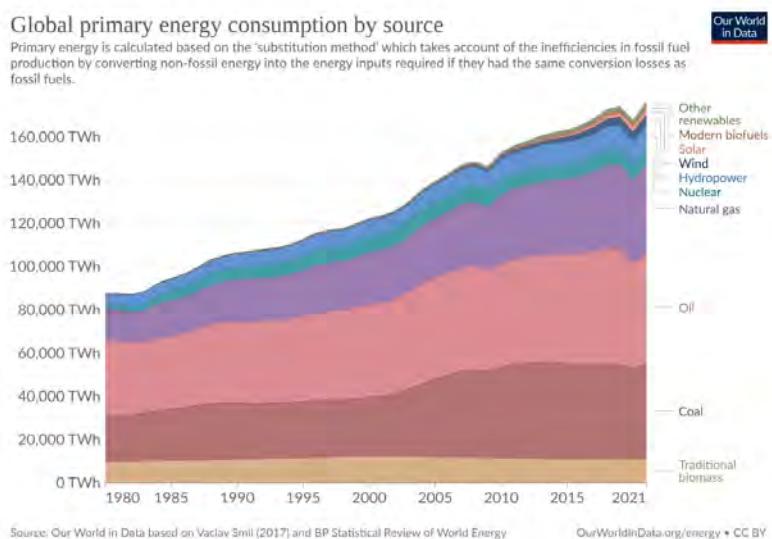


Figure 4:

#### Energy consumption by source<sup>4</sup>

This is the energy consumption trend. Of course, energy consumption is increasing. Generally, it dips when economic recessions or big shocks to the system, such as COVID-19, are captured above, but the trend usually moves upward. You can see again by the source what percentage of dirty energy is being consumed vs renewable ones.

<sup>4</sup>"File: Energy Consumption by Source, OWID.Svg." Wikipedia, Wikimedia Foundation, en.m.wikipedia.org/wiki/File:Energy\_consumption\_by\_source,\_OWID.svg. Accessed 30 Jan. 2024.

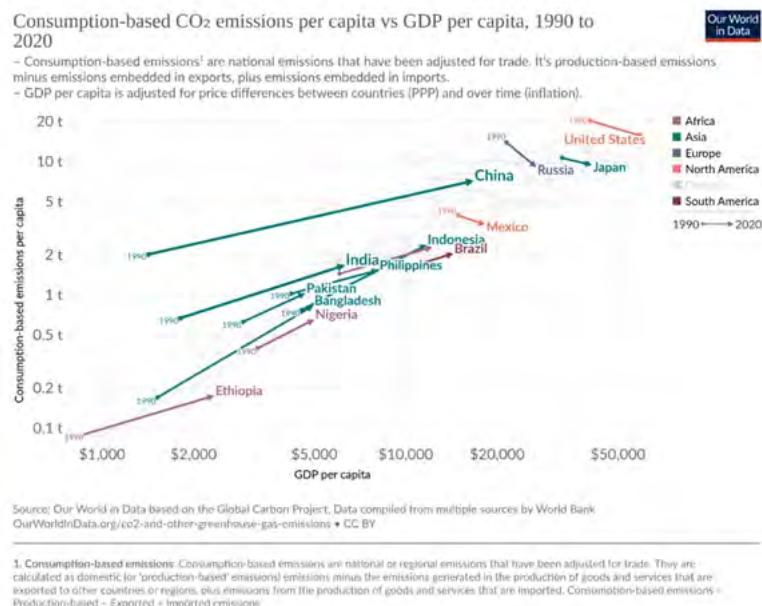


Figure 5:

### Consumption Based CO<sub>2</sub> emissions per capita vs GDP per capita, 1990 to 2020

The plot on the right shows Consumption-based CO<sub>2</sub> emissions per capita vs GDP per capita. The United States has the top honor there, the united states has the highest GDP per capita, but also the highest consumption based emission. Things are improving; you can see from 1990 to about 2020 or so, the US has reduced its consumption-based emissions, same with Japan, Russia, and Mexico. China has an increasing trend, and many other developing countries are also in the same situation, including Brazil. So that's a worry some trend.

## 0.6 CO<sub>2</sub> emissions per capita and global warming trend

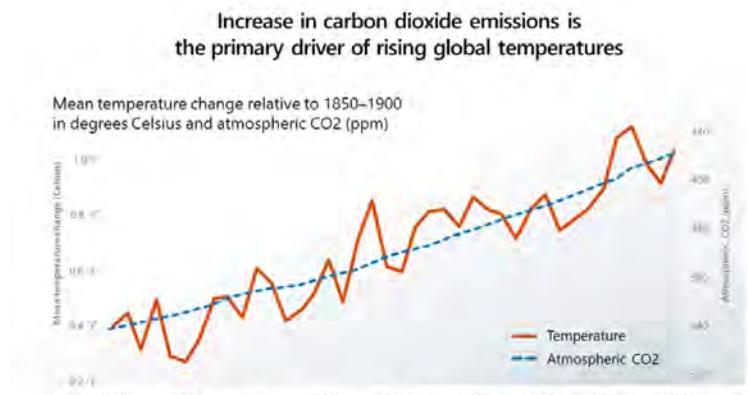


Figure 6:

## 0.6 $CO_2$ emissions per capita and global warming trend

Increase in carbon dioxide emissions is the primary driver of rising global temperatures <sup>5</sup>

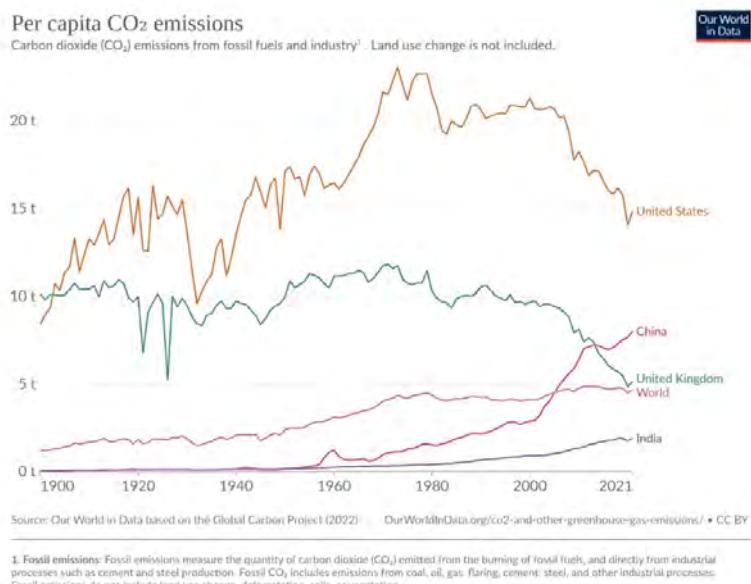


Figure 7:

Per capita CO<sub>2</sub> emissions by country<sup>6</sup> Suppose global warming is well documented and known. For example, the above figure shows the increase in temperature in degrees Celsius from the 1980 baseline to about 2020, and you can see this in the order of .6 – .7 degrees Celsius, which tracks very well with the atmospheric  $CO_2$  that is present. So, it is well documented and the source of concern for us.

<sup>5</sup>“It’s Called Frenchs Forest for a Reason and Why It Is Important to See the Wood for the Trees.” The University of Sydney, [www.sydney.edu.au/business/news-and-events/news/2023/03/08/it-s-called-frenchs-forest-for-a-reason-and-why-it-is-important-.html](http://www.sydney.edu.au/business/news-and-events/news/2023/03/08/it-s-called-frenchs-forest-for-a-reason-and-why-it-is-important-.html). Accessed 30 Jan. 2024.

<sup>6</sup>Smith, Brad. “Microsoft Will Be Carbon Negative by 2030.” The Official Microsoft Blog, 23 July 2020, [blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/](https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/)

## 0.7 Legally binding international treaty on climate change

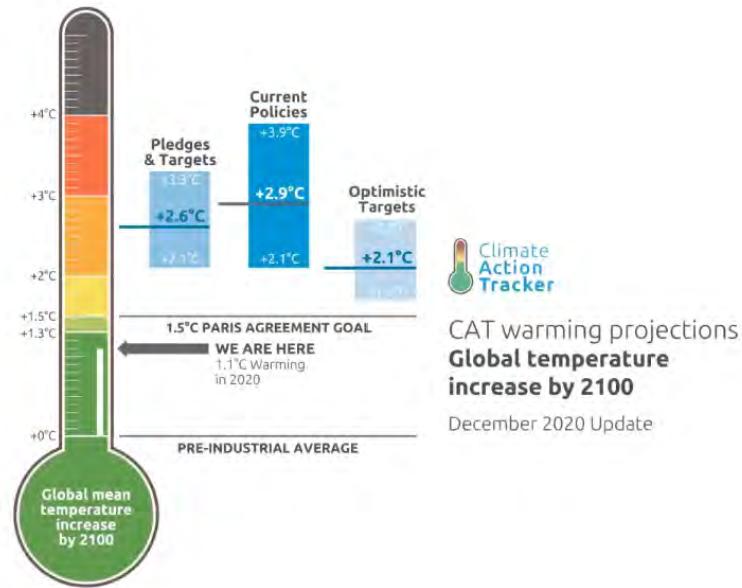


Figure 8:

Global update: Paris Agreement Turning Point<sup>7</sup>

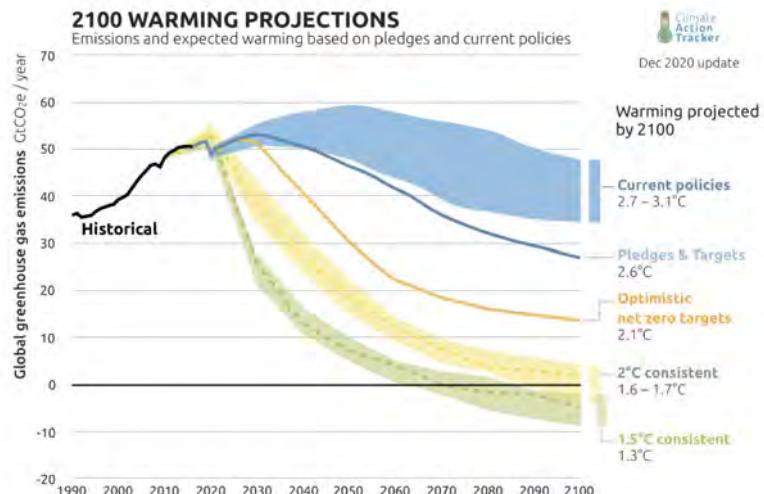


Figure 9:

Global Update: Climate Summit Momentum<sup>8</sup>

This recent meeting in Egypt was called the Sharm El-Sheikh Climate Agreement. Witch the US came back into the fold, and some new agreements were made, but the fact is that if we

<sup>7</sup>"Global Update: Paris Agreement Turning Point." Climate Action Tracker, [climateactiontracker.org/press/global-update-paris-agreement-turning-point/](https://climateactiontracker.org/press/global-update-paris-agreement-turning-point/). Accessed 30 Jan. 2024.

<sup>8</sup>"Global Update: Climate Summit Momentum." Climate Action Tracker, [climateactiontracker.org/publications/global-update-climate-summit-momentum/](https://climateactiontracker.org/publications/global-update-climate-summit-momentum/). Accessed 30 Jan. 2024.

don't make any policy changes now, we expect by the  $22^{\text{nd}}$  century (2100), we will have an increased temperature of about 2.7 to 2.3 Celsius. This isn't good. It would not be good if this happened. So with the pledges and the targets (2.6 deg ), Optimistically (2.1 deg). So we have to be concerned about and care a lot about the climate conditions in the future.

## 0.8 The cost and efficiency of computing

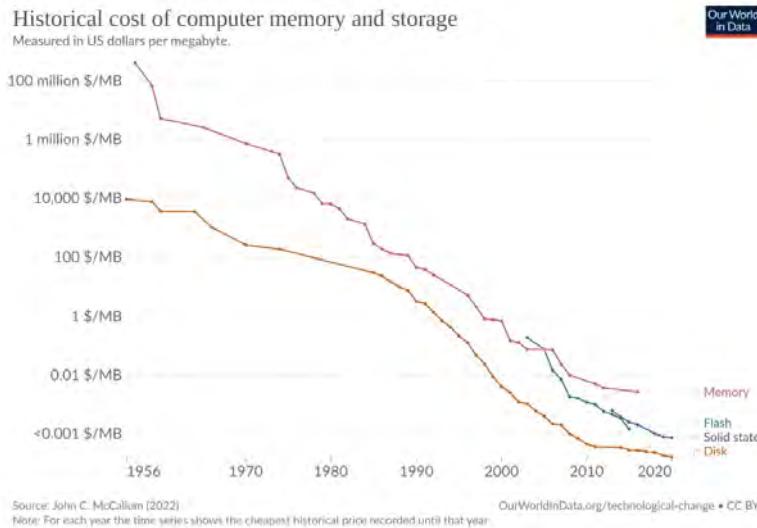


Figure 10:

Historical cost of computer memory and storage<sup>9</sup>

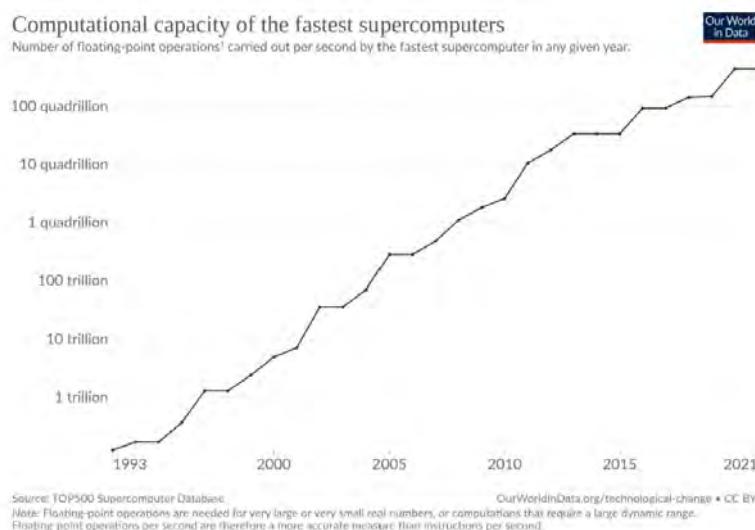


Figure 11:

<sup>9</sup>“Historical Cost of Computer Memory and Storage.” Our World in Data, [ourworldindata.org/grapher/historical-cost-of-computer-memory-and-storage](https://ourworldindata.org/grapher/historical-cost-of-computer-memory-and-storage). Accessed 30 Jan. 2024.

## Computational capacity of the fastest supercomputers<sup>10</sup>

So, let us now look at the computing trends and the historical cost of computer memory and storage. This is another version of Moore's law in which the cost of building the transistor or a storage element drops exponentially over time. Also, we can see that there's been a significant, sustained improvement of our basic computer memory storage capabilities, and that has also given us the computational capacity, which is increasingly similar to the better-known version of Moore's law on the right-hand side. Currently, we are on the level where we can do (in 2021 or so) in excess of 100 quadrillion operations per second in our fastest supercomputers

## 0.9 What are the practices



Figure 12:

## World Word Map

Energy efficiency is important for several different reasons depending on the target platforms, e.g., electric bill savings in hosting data centers, hotspot avoidance in high-end computing systems, error/failure minimization in integrated circuits and systems, and battery service life maximization in laptops and smartphones

## 0.10 Greenhouse Gases based on the GHG Protocol

### 1. Scope 1. Direct GHG emissions

Owned/controlled boilers/furnaces/turbines, backup combustion power generators, vehicles,

<sup>10</sup>“Computational Capacity of the Fastest Supercomputers.” Our World in Data, [ourworldindata.org/grapher/supercomputer-power-flops](https://ourworldindata.org/grapher/supercomputer-power-flops). Accessed 30 Jan. 2024.

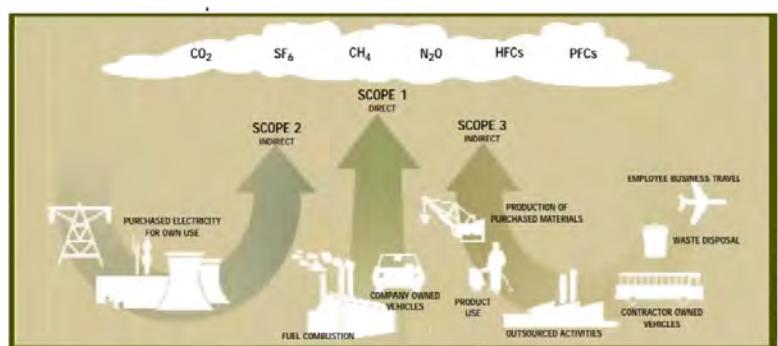
coolants/refrigerants, chemical/gas produced in owned/controlled equipment, transportation of materials, products, waste, and employees

## 2. Scope 2. Electricity indirect GHG emissions from purchased or acquired energy

Try to use less greenhouse gas-intensive sources of electricity. Avoid coal, for example, or other types of energy sources that emit a lot of carbon

## 3. Scope 3. Other indirect GHG emissions

A consequence of the activities of the company that occur from sources now owned or controlled by the company e.g., exploration, extraction, and production of purchased materials and fuels, use of sold products and services, waste disposal, transportation of sold products and waste



<https://ghgprotocol.org/>

Figure 13:

Greenhouse Gas Protocol<sup>11</sup>

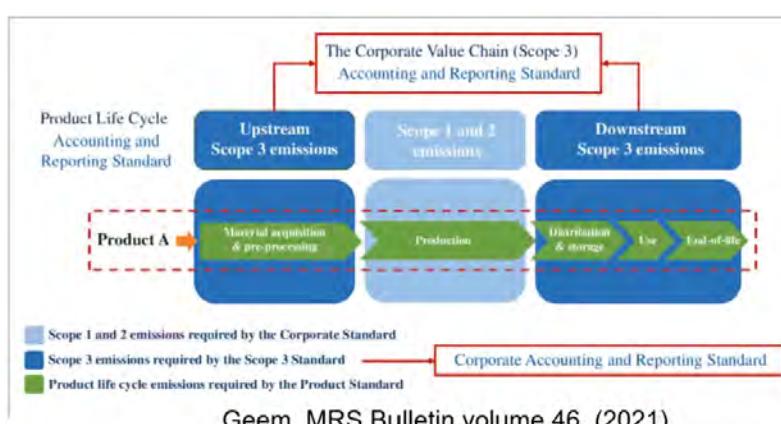


Figure 14:

Greenhouse gas protocol and illustration of Scope 1, Scope 2, and Scope 3 emissions.<sup>12</sup> This is a well-known standard concerning how to look at the emission footprints of anything, including

<sup>11</sup>The Greenhouse Gas Protocol,[ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf](http://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf). Accessed 31 Jan. 2024.

<sup>12</sup>Greenhouse Gas Protocol and Illustration of Scope 1, Scope 2, [www.researchgate.net/figure/Greenhouse-gas-protocol-and-illustration-of-Scope-1-Scope-2-and-Scope-3-emissions\\_fig3\\_357503745](http://www.researchgate.net/figure/Greenhouse-gas-protocol-and-illustration-of-Scope-1-Scope-2-and-Scope-3-emissions_fig3_357503745). Accessed 31 Jan. 2024.

computing. There are three different scopes defined. Scope one: Direct greenhouse gas emissions are due to generators, boils, vehicles, and equipment owned by the entity, for example, Apple, Microsoft, or even the US government entity. Scope two is the indirect Greenhouse gas emissions from purchased or acquired energy from other sources. Scope three is from other indirect greenhouse gas (GHG) emissions because of the activities from OEMs and suppliers to the company that provides the various materials, products, and components that go into the company's final product that is offered for sale. Of course, the company has direct control over scope one and two, but scope three is less controlled by the company or entity. That was one way of characterizing greenhouse gas emissions, and I will use this when I get to the analysis of different companies in manufacturing and software services. What have you deal with these scopes of energy consumption and therefor GHG emissions

## 0.11 Foundational, First, and Second Order Trends



Figure 15:

The three-wheel framwork of value creation<sup>13</sup>

<sup>13</sup>“Where Do You Fit in the New Digital Ecosystem?” Deloitte Insights, [www2.deloitte.com/us/en/insights/topics/emerging-technologies/new-digital-ecosystem-technology-media-telecom-industry.html](http://www2.deloitte.com/us/en/insights/topics/emerging-technologies/new-digital-ecosystem-technology-media-telecom-industry.html). Accessed 30 Jan. 2024.

## 0.11 Foundational, First, and Second Order Trends

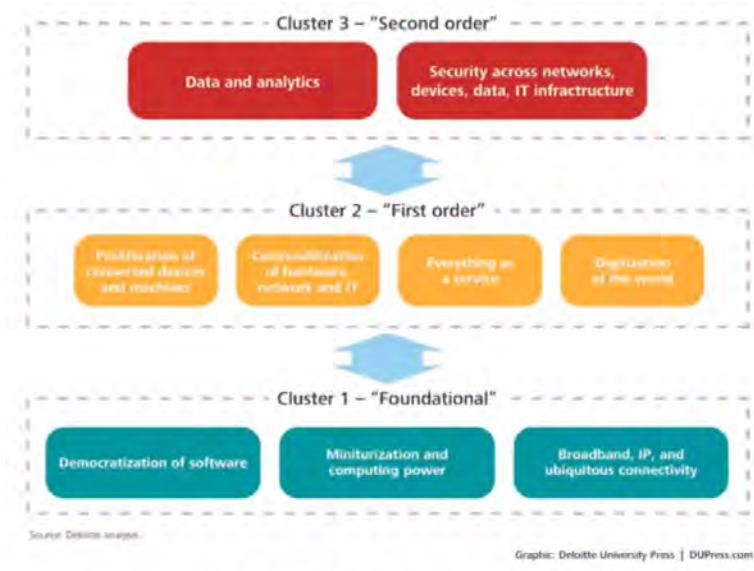


Figure 16:

Dominant trends within the digital ecosystem.<sup>14</sup>

So, suppose you look at our digital Ecosystem, this is trying to summarize graphically the well-documented and generally used terminology that is foundational trends. In that case, there are other first-order effects and second-order effects. The Foundational trends are the democratization of software, meaning access to software services and products regardless of their class, gender, and country of origin, and the democratization of Moore's law (Miniaturization) itself beyond that, in terms of beyond CMOS Miniaturization that is taking place, trends in computing power that are generally becoming more power efficient than before, and this broadband connectivity that we have for example 5G, 6G and so on. Cluster two – “First order” effects is the proliferation of connected devices and machines, hardware being commoditized and used in many different products, we have seen this effect with covid -19 and IC shorted and how it impacted everything, right you don't expect how it is going to impact electronics and it did. The services come in, and the world is really being digitized. CluserTree refers to the punch note capabilities that are being provided, such as advanced data analysis, analytics, machine learning technologies that help with decision-making, prediction forecasting, and more secure access to various types of data, whether through encryption standards or through some other protocols, such as secure multicore computation and so on.

<sup>14</sup>“Asset Management.” Top Software Consulting Company Fairfax, VA — MD — DC - CT Solutions, Inc, [www.ctsols.com/service/asset-management](http://www.ctsols.com/service/asset-management). Accessed 30 Jan. 2024.

## Sustainable Computing A Quantitative Perspective

Type of Release and Other Waste Management Activity	Typical EPCRA Section 313 Chemical
Stack Air, Fugitive Air	N-methyl-2-pyrrolidone, Xylene, glycol ethers, Ethylbenzen, methanol, hydrofluoricacid, sulfuric acid aerosols; hydrochloric acid aerosols MEK
Off-Site Transfer, On-Site Treatment	N-methyl-2-pyrrolidone, xylene, glycol ethers, ethylbenzene, methanol, hydrofluoricacid, MEK
Stack Air, Fugitive Air	Ammonia, 1,2-dichloroethylene
Stack Air, Fugitive Air, On-Site Treatment	Hydrochloricacidaerosols, sulfuricacidaerosols, nitricacid
Stack Air, Fugitive Air, On-Site Transfer, On-Site Treatment	Hydrochloricacid, xylene, Methanol, toluene, hydrochloricacidaerosols, sulfuricacidaerosols
Off-Site Transfer, POTW	Arsenic, antimony, Trace amounts of metals

## 0.12 Phases in Hardware Lifecycle



Figure 17:

### How Can IoT Contribute to a More Sustainable Product Lifecycle?

let us examine the two competing ways to organize a Computing Hardware's Lifecycle. There are many different ways to describe this process, but generally, the first-order effect is to plan the product and compile all the parts and components you need to build it and put it Into operation. The second order is to operate it, of course during this operation you have to maintain it If you service requires upgrades that also must be taken care of. The third effect is to decommission and replace the part. Therefore, if you look at the phases in the hardware's lifecycle, there are as follows.

**Before use** Planning and procurement

**In-use** Deployment and operations/maintenance

**After-use** Recycling and disposal

### CapEX Vs. OpEx

#### CapEX:

Single lump-sum investment covers manufacturing, transportation, and end-of-life disposal/recycling mainly associated with before-use and, to some extent, After-use

#### OpEx:

It covers the cost of operation and maintenance, which refers to the in-use operations.

## 0.13 Employs Complex, Often Dirty, and Energy-Hungry Processes

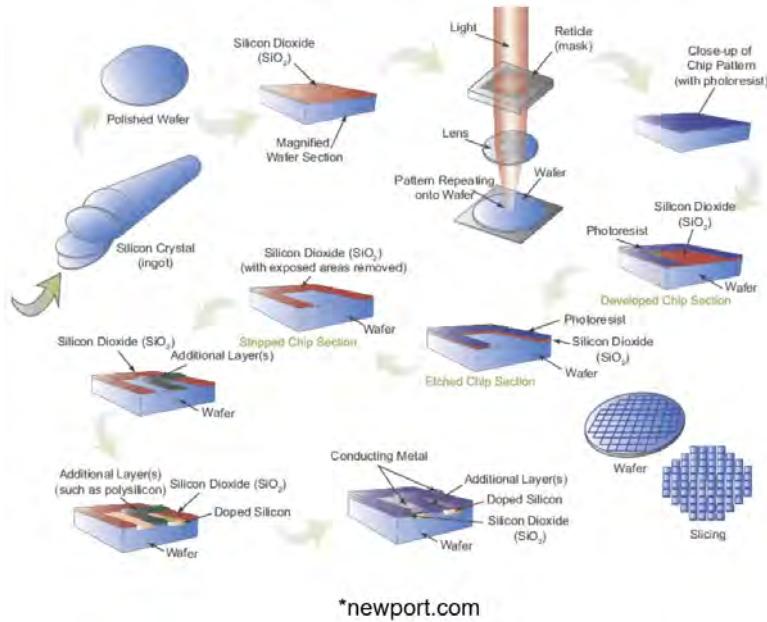


Figure 18:

Overview of the semiconductor manufacturing process.<sup>15</sup>

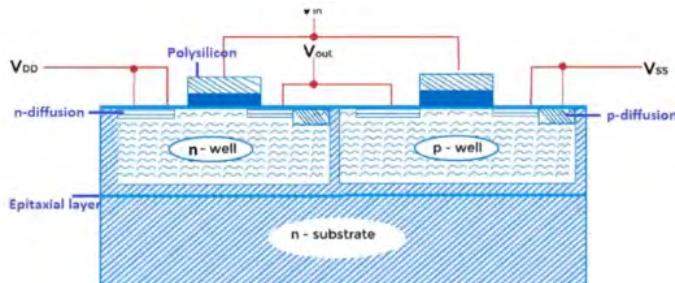


Figure 19:

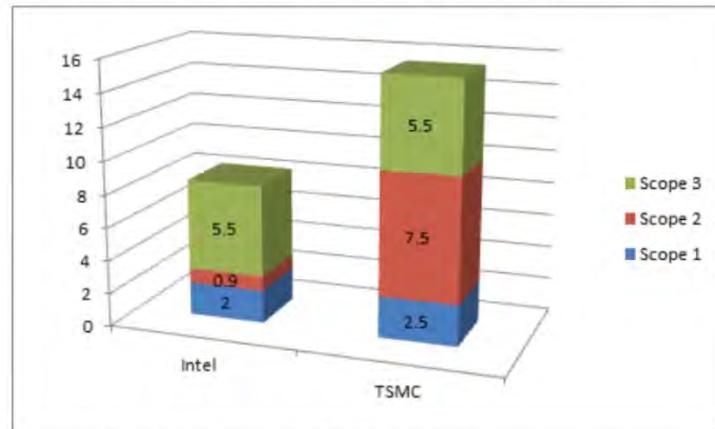
Twin-TUB Fabrication<sup>16</sup> Now, look at some processes used to manufacture our devices → circuits → systems. We have very complex, often very dirty, and power-hungry processes. For example, chemical mechanical polishing, etching, and oxidation are all processes that require many chemicals that (in most cases) are much worse for the environment than standard carbon emission. So generally, chip fabrication is essential, we must do it, but it is a very GHG-intensive process: due to all of the material and processing steps involved in producing state-of-the-art manufacturing. EUV (extreme ultraviolet) lithography is much more energy-hungry and dirty compared to 16nm, 22nm processes, etc .

<sup>15</sup>“Semiconductor Manufacturing.” Semiconductor Manufacturing, [www.newport.com/n/semiciconductor-manufacturing](http://www.newport.com/n/semiciconductor-manufacturing). Accessed 30 Jan. 2024.

<sup>16</sup>“CMOS Fabrication - Javatpoint.” Www.Javatpoint.Com, [www.javatpoint.com/cmos-fabrication](http://www.javatpoint.com/cmos-fabrication). Accessed 30 Jan. 2024.

## 0.14 The global chip industry has a colossal problem with carbon emissions!

**TSMC and Intel: Carbon emissions (million metric tons)**



Source: Companies, compiled by DIGITIMES

Semiconductors firms - big power users, 15 June 2022

Figure 20:

A contrast of speech act verbs' distribution in the two corporations <sup>17</sup>

1. TSMC's greenhouse gas emissions have overtaken those of automotive giant GM 15.5 MtCO<sub>2</sub>eq ("Million Tonnes of carbon dioxide equivalent") in 2020
  - TSMC uses more electricity each year than Taiwan's capital, Taipei
  - TSMC announced that it wants to reach net-zero emissions by 2050
2. Intel's greenhouse gas emissions in 2021 amounted to 8.3 MtCO<sub>2</sub>eq
  - Intel Corporation pledges to emit zero emissions by 2040

The big difference in Scope 2 emissions is mainly due to the lower percentage of green power in Taiwan's electricity mix and TSMC's use of higher power-consuming processes (In TSMC's 7nm and 5nm processing, EUV equipment is indispensable, which also causes TSMC to emit much more greenhouse gases in Scope 2 than intel.)

So, regarding the chip industry and carbon emissions, data compares TSMC vs intel in three separate scopes (emission types). As you can see when you look at Intel, its total million metric ton equivalent carbon emissions are much lower than TSMC's, and it's not because Intel's production capacity, number of employees, etc, is smaller than TSMC's.

Essentially, the difference is in the sources/energy production in the respective countries. In the United States, the energy sources produced are generally cleaner than in Taiwan, which creates a big difference between the two. The other reason is that TSMC relies heavily on advancements in our equipment, but they emit much more greenhouse gasses than Intel, at least on a production basis.

<sup>17</sup>Phrasal Verb Frequencies in the Five Corpora\* — Download Table, [www.researchgate.net/figure/Phrasal-verb-frequencies-in-the-five-corpora\\_tbl1\\_263252963](http://www.researchgate.net/figure/Phrasal-verb-frequencies-in-the-five-corpora_tbl1_263252963). Accessed 31 Jan. 2024.

## 0.15 Keeping energy use steady despite large growth in demand

Photos courtesy of ABB

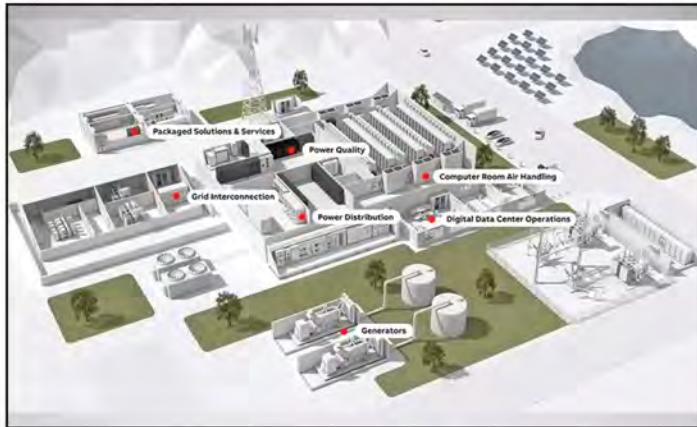


Figure 21:

ABB data center solutions <sup>18</sup>

Now, we will look at data centers. Data centers typically house thousands to hundreds of thousands of powerful servers composed of CPUs, GPUs, specialized accelerators, and so on. This complex setup requires air conditioning, power distribution, power quality control, local generators, lots of batteries, and so on.

## 0.15 Keeping energy use steady despite large growth in demand



Figure 22:

Once opened, data centers don't need much staff to keep running 24 hours a day

<sup>18</sup>ABB Data Center Generators Reliable Solutions for Data Center, [library.e.abb.com/public/87c892ea1ec34ac0ba33b04c368dfd1a/Data\\_center\\_generator\\_9AKK108003\\_RevB\\_EN\\_lowres.pdf?x-sign=RxZtu7tnNpCwdCpvAiqEkqxBcXU2HX23Mp6tcFL97KEJwpWtmE1D8/dlaYGaG9nV](https://library.e.abb.com/public/87c892ea1ec34ac0ba33b04c368dfd1a/Data_center_generator_9AKK108003_RevB_EN_lowres.pdf?x-sign=RxZtu7tnNpCwdCpvAiqEkqxBcXU2HX23Mp6tcFL97KEJwpWtmE1D8/dlaYGaG9nV). Accessed 31 Jan. 2024.

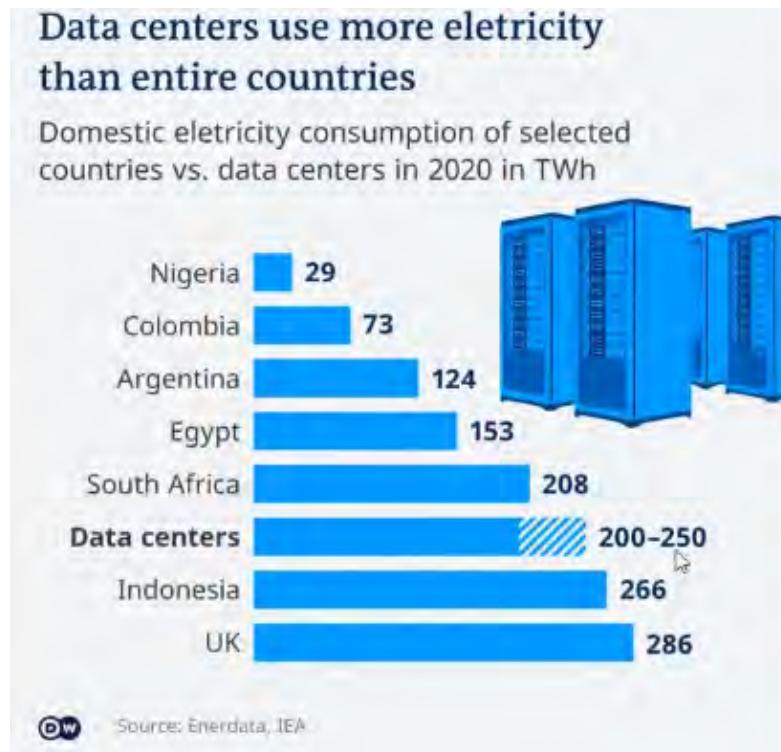


Figure 23:

Data centers use more electricity than entire countries <sup>19</sup> Next, we will look at the global data center electricity consumption trend. Data centers are 200 to 250 TW (terawatt) hours, which is more than the energy consumption of Egypt, Argentina, or South Africa, getting close to Indonesia and the United Kingdom annually. Through this perspective, we must give credit to ourselves as the developers of these technologies and solutions that go into data centers. I remember I started working on a similar talk in 2010 for a keynote speech. Back then, projections were that we expected electricity consumption to keep going up and that it would be a massive problem for us, not in terms of a ten percent increase but in the order of eighty percent or so.

<sup>19</sup>Rooks, Timothy. "Data Centers Keep Energy Use Steady despite Big Growth – DW – 01/24/2022." Dw.Com, Deutsche Welle, 24 Jan. 2022, [www.dw.com/en/data-centers-energy-consumption-steady-despite-big-growth-because-of-increasing-efficiency/a-60444548](http://www.dw.com/en/data-centers-energy-consumption-steady-despite-big-growth-because-of-increasing-efficiency/a-60444548).

## 0.15 Keeping energy use steady despite large growth in demand

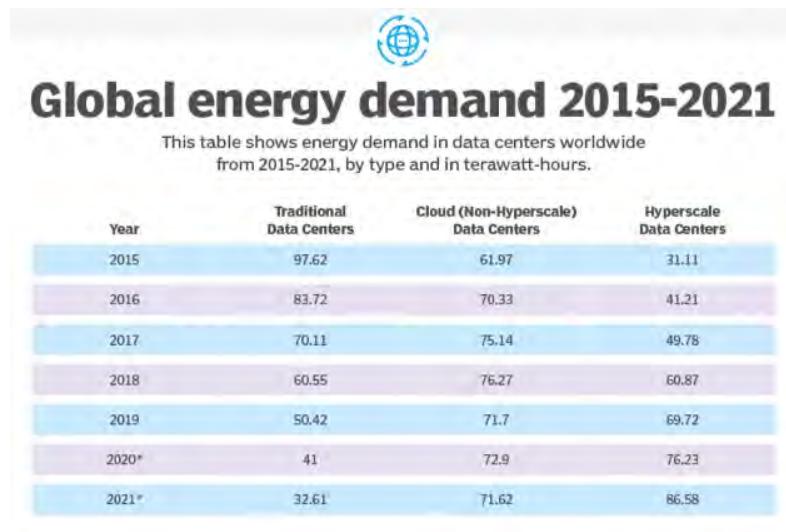


Figure 24:

Global energy demand 2015-2021<sup>20</sup>

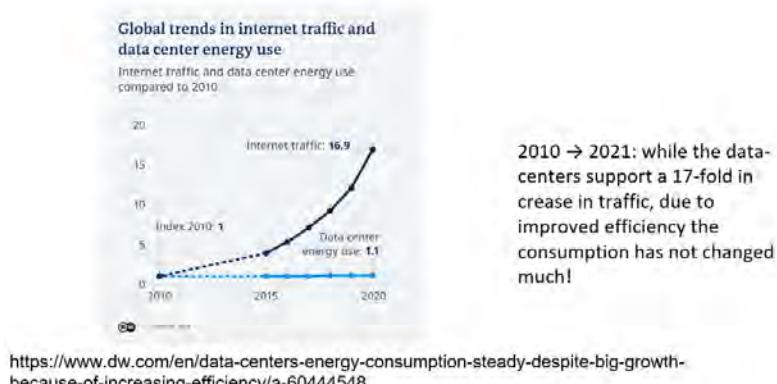


Figure 25:

Global trends in internet traffic and data center energy use<sup>21</sup> Despite the increase in internet traffic, which is a factor of seventeen, we have provided the required services with only maybe a ten percent increase in the total energy consumption since 2010, and that's remarkable! This is only due to many people working very hard to do so, including introducing and deploying so-called energy-proportional computing hardware.

<sup>20</sup>Mok, Edwin. "Orv3, Sustainability, Cooling, Modularity & Integration: Observations from the OCP Global Summit 202." ORv3, Sustainability, Cooling, Modularity & Integration: Observations from the OCP Global Summit 202 — Advanced Energy, [www.advancedenergy.com/de-de/about/news/blog/orv3,-sustainability,-cooling,-modularity-integration-observations-from-the-ocp-global/](http://www.advancedenergy.com/de-de/about/news/blog/orv3,-sustainability,-cooling,-modularity-integration-observations-from-the-ocp-global/). Accessed 30 Jan. 2024.

<sup>21</sup>Rooks, Timothy. "Data Centers Keep Energy Use Steady despite Big Growth — DW — 01/24/2022." Dw.Com, Deutsche Welle, 24 Jan. 2022, [www.dw.com/en/data-centers-energy-consumption-steady-despite-big-growth-because-of-increasing-efficiency/a-60444548](https://www.dw.com/en/data-centers-energy-consumption-steady-despite-big-growth-because-of-increasing-efficiency/a-60444548).

## Sustainable Computing A Quantitative Perspective

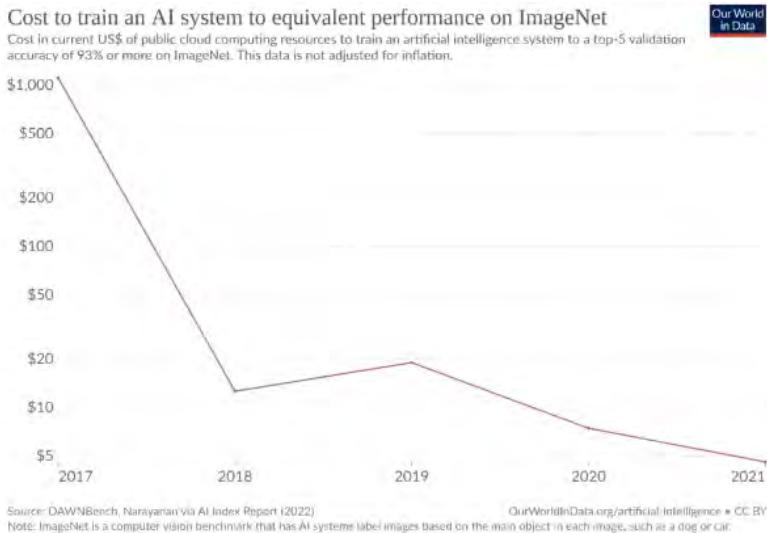


Figure 26:

Cost to train an AI system to equivalent performance on ImageNet <sup>22</sup>

This is another trend that we see these days having to do with artificial intelligence. Of course, this is a very useful technology that everyone is using, thus is becoming indispensable, and is expected only to grow going forward. But because of all the functions, services, and utility it provides to us, fueled by our need to create higher fidelity or higher accuracy solutions for forecasting decision-making control, we are building more and more complex models.

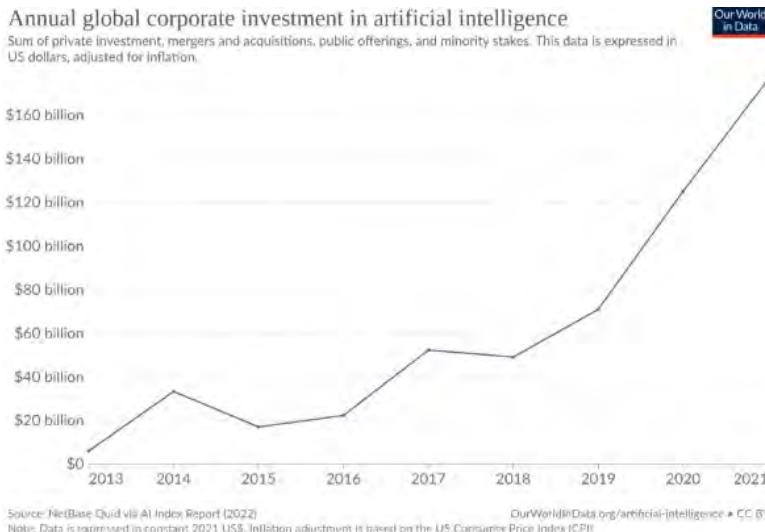


Figure 27:

Annual global corporate investment in artificial intelligence <sup>23</sup>

<sup>22</sup>Data, Our World in. “Our Regularly-Updated Charts on AI Let You Monitor Different Trends. A Few Examples:• Annual Global Corporate Investment in AI• Chess Ability of the Best Computers • Computation Used to Train Notable Systems• Cost to Train a System to Equivalent Performance on ImageNet Pic.Twitter.Com/4j585uz1UU.” Twitter, Twitter, 6 Dec. 2022, [twitter.com/OurWorldInData/status/1600096040442535936](https://twitter.com/OurWorldInData/status/1600096040442535936).

<sup>23</sup>“Annual Global Corporate Investment in Artificial Intelligence, by Type.” Our World in Data, [ourworldindata.org/artificial-intelligence](https://ourworldindata.org/artificial-intelligence).

## 0.15 Keeping energy use steady despite large growth in demand

Some of the models we use today have on the order of 100s of millions or even billions of hidden parameters. So when you start training such models so that you can deploy them and use them for inference, the cost of training could be pretty high.

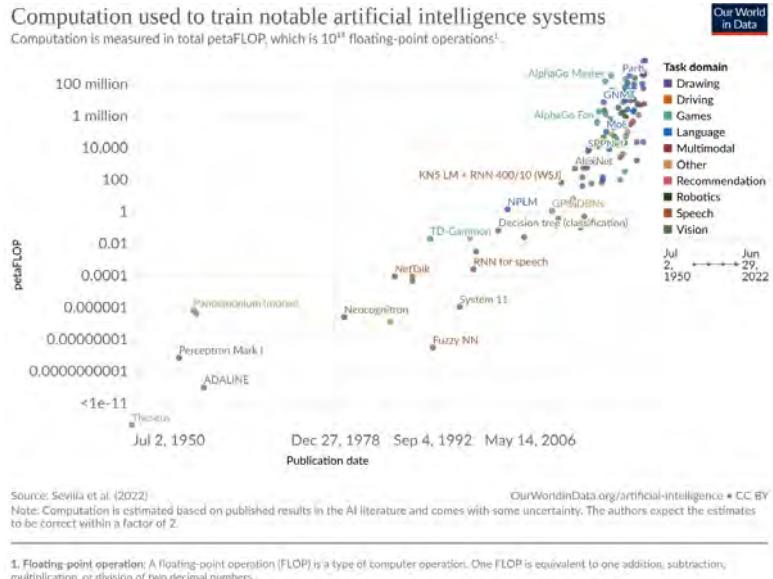


Figure 28:

Estimated compute used for training notable AI systems over the last 70 years. Since 2010, the compute used for training has been doubling every six months<sup>24</sup>. The Figure titled “Computation used to train notable artificial intelligence systems” shows the requirements in terms of petaFLOPS to train some of our notable AI systems, but of course, there are a lot more complicated models these days, right? Some of the BERT models are quite large and require a lot of computing power.

[org/grapher/corporate-investment-in-artificial-intelligence-by-type](https://ourworldindata.org/grapher/corporate-investment-in-artificial-intelligence-by-type). Accessed 30 Jan. 2024.

<sup>24</sup>DeepLearning.AI. “Cutting the Carbon Cost of Training a New Tool Helps NLP Models Lower Their Gas Emissions.” A New Tool Helps NLP Models Lower Their Gas Emissions; A New Tool Helps NLP Models Lower Their Gas Emissions, 28 Oct. 2022, [www.deeplearning.ai/the-batch/cutting-the-carbon-cost-of-training/](http://www.deeplearning.ai/the-batch/cutting-the-carbon-cost-of-training/).

## 0.16 Carbon emission based on region and in comparison with fossil fuel usage

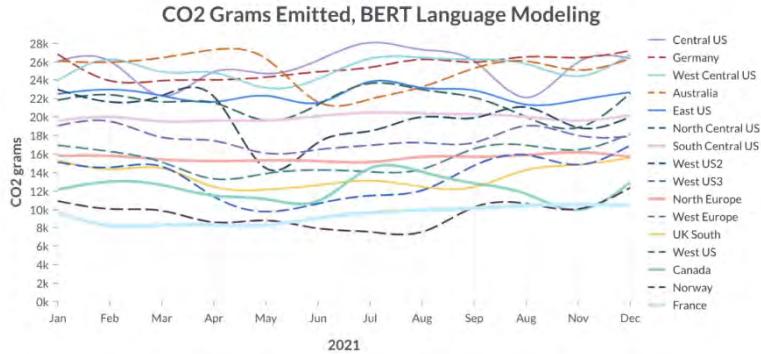


Figure 29:

Carbon emissions accrued training BERT model on eight Nvidia V100s for 36 hours in different regions throughout the year <sup>25</sup> Now, we will bring it back to CO<sub>2</sub> emissions. The Figure Titled “CO<sub>2</sub> Grams Emitted, BERT Language Modeling” shows the cost of training a BERT language model in terms of carbon emission in different countries, and the reason for this is that different countries have a different carbon mix for the generation sources.

Model	BERT finetune	BERT pretrain	6B Transf.	Dense 121	Dense 169	Dense 201	ViT Tiny	ViT Small	ViT Base	ViT Large	ViT Huge
GPU	4-V100	8-V100	256-A100	1-P40	1-P40	1-P40	1-V100	1-V100	1-V100	4-V100	4-V100
Hours	6	36	192	0.3	0.3	0.4	19	19	21	90	216
kWh	3.1	37.3	13,812.4	0.02	0.03	0.04	1.7	2.2	4.7	93.3	237.6

\*Dodge et al. (2022)

Figure 30:

For the 11 models in our analysis: the type of GPU, the number of GPUs, the number of hours, and the energy used in kWh. For example, our BERT language modeling (BERT LM) experiment used 8 V100 GPUs for 36 hours and used a total of 37.3 kWh. We note our training run of the 6 billion parameter transformer only trained for approximately 13% of the time it would take to train to completion; we estimate a full training run would consume approximately 103,593 kWh. <sup>26</sup> So again, that’s a big problem, and we expect it to worsen. There is a trend, fortunately, in the recent past to try to control the size of the Neural Network models that we use to get high-fidelity solutions through advanced architecture optimizations that one performs on neural networks or through aggressive quantization, pruning, and so on. Many folks in academia and industry are focusing on those things, so hopefully, it will come down.

<sup>25</sup>EPA, Environmental Protection Agency, [www.epa.gov/energy/greenhouse-gas-equivalencies-calculator](http://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator). Accessed 31 Jan. 2024.

<sup>26</sup>Dodge, Jesse, et al. “Measuring the Carbon Intensity of AI in Cloud Instances.” arXiv.Org, 10 June 2022, [arxiv.org/abs/2206.05229](https://arxiv.org/abs/2206.05229).

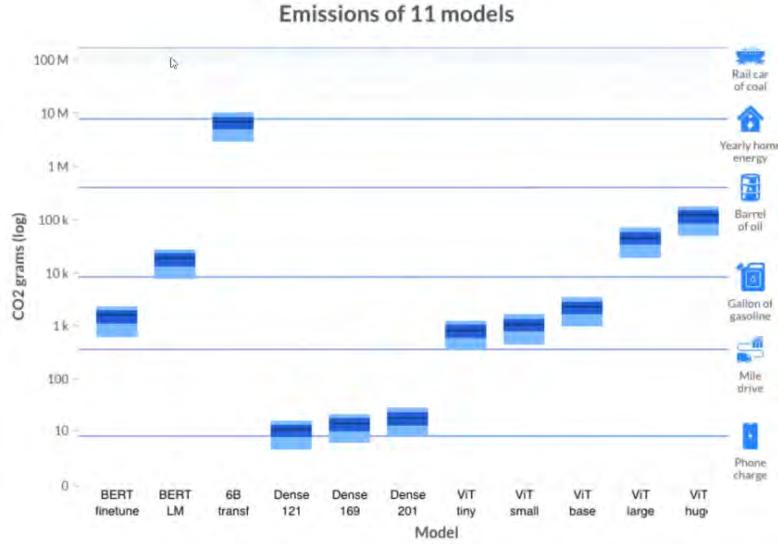


Figure 31:

Emissions for our 11 experiments described in §4. For each model, we show a vertical blue bar, where the top of the bar is the max, the bottom is the min, and the black line represents the average emissions (across regions and time of year). The first and fourth quartiles are represented by the light blue at the top and bottom of each vertical blue bar. The largest training runs (e.g., 6 billion parameter LM) release a significant amount of emissions, no matter the region (and recall, the 6 billion parameter LM is only trained for 13% of a full run, so a full run would emit about an order of magnitude more emissions than reported here). The smallest experiments emit very little. Presented on a log scale, with references on the right indicating equivalent sources of emissions per the United States Environmental Protection Agency <sup>27</sup>. This trend is going to stop, but whether we can reverse it soon is the question.

## 0.17 Energy usage and efficiency on personal and small business levels

- Why extend the device's lifetime?
- What can be done on the personal, small, and mid-size enterprise levels?
  1. Device lifecycle management
  2. IT asset acquisition with responsible IT asset disposal in mind
  3. Smart charging and modular design increase the mobile device's lifetime

<sup>27</sup>EPA, Environmental Protection Agency, [www.epa.gov/energy/greenhouse-gas-equivalencies-calculator](http://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator). Accessed 31 Jan. 2024.

## Sustainable Computing A Quantitative Perspective

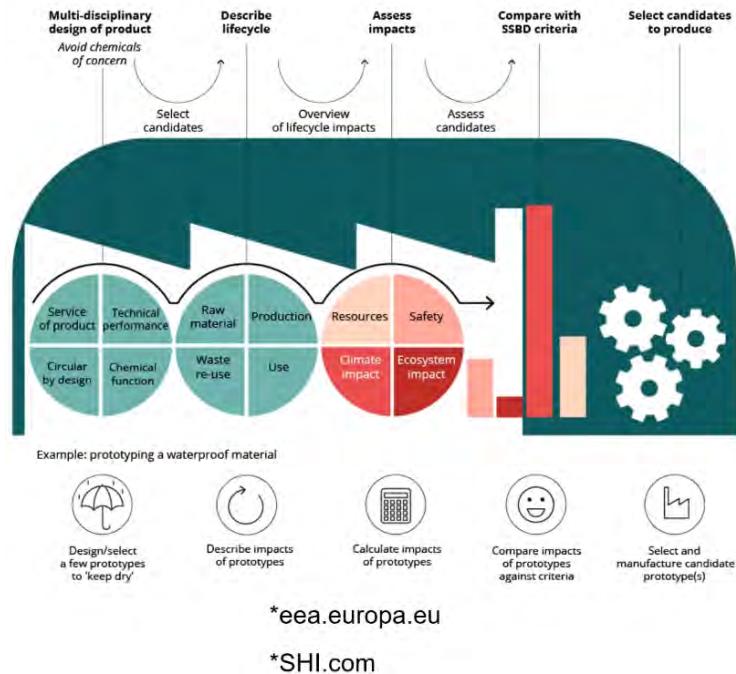
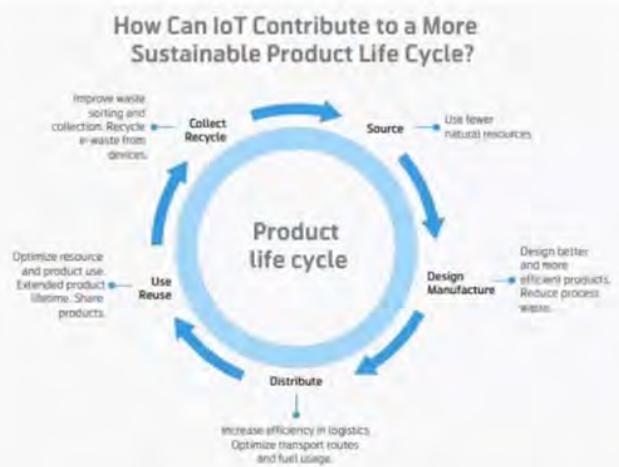


Figure 32:

Implementing safe and sustainable by design approaches <sup>28</sup>



\*Avnet Silica, 2020

Figure 33:

How Can IoT Contribute to a More Sustainable Product Lifecycle? <sup>29</sup> So the other source of energy consumption concerning IT is, of course, all of the mobile and personal devices that we

<sup>28</sup>“Designing Safe and Sustainable Products Requires a New Approach for Chemicals.” European Environment Agency, 7 Feb. 2023, [www.eea.europa.eu/publications/designing-safe-and-sustainable-products-1](http://www.eea.europa.eu/publications/designing-safe-and-sustainable-products-1).

<sup>29</sup>Admin@relevant.software. “IOT and Sustainability: How Business Can Improve Environmental Footprint with the Internet of Things.” Relevant Software, 1 Sept. 2022, [relevant.software/blog/improve-environmental-footprint-with-internet-of-things/](http://relevant.software/blog/improve-environmental-footprint-with-internet-of-things/).

#### *0.17 Energy usage and efficiency on personal and small business levels*

love, use, and depend on every day of our lives. And these devices consume a lot of energy. But more importantly, their manufacturing costs are high, right? let us look at Apple as an example.

## 0.18 Apple's Environmental Progress Report, 2022

Achieve **carbon neutrality** for our entire carbon footprint, including products, by 2030. And reduce related emissions by 75% compared with fiscal year 2015



40% emissions reduction since 2015 across our value chain



Established the \$200M Apple Restore Fund with the aim of removing over 1M metric tons of carbon per year



23M metric tons of emissions avoided in fiscal year 2021 alone due to carbon reduction initiatives across our value chain



Transition our entire manufacturing supply chain to **100% renewable electricity** by 2030



As of March 2022, 213 suppliers have committed to 100% renewable electricity for Apple production, representing the majority of Apple's direct worldwide spend for materials, manufacturing, and assembly of products



Figure 34:

Apple's comprehensive carbon footprint <sup>30</sup>

The Figure above is from Apple's sources and is somewhat of a marketing document, but at the same time, it at least reflects the trends. As we all know, every Fortune 500 company has some carbon footprint or neutrality in mind that gets advertised. For example, Apple says that by 2030, they will achieve carbon neutrality and reduce related emissions by 75 percent compared to 2015. Also, 100 percent renewable electricity is going to be used.

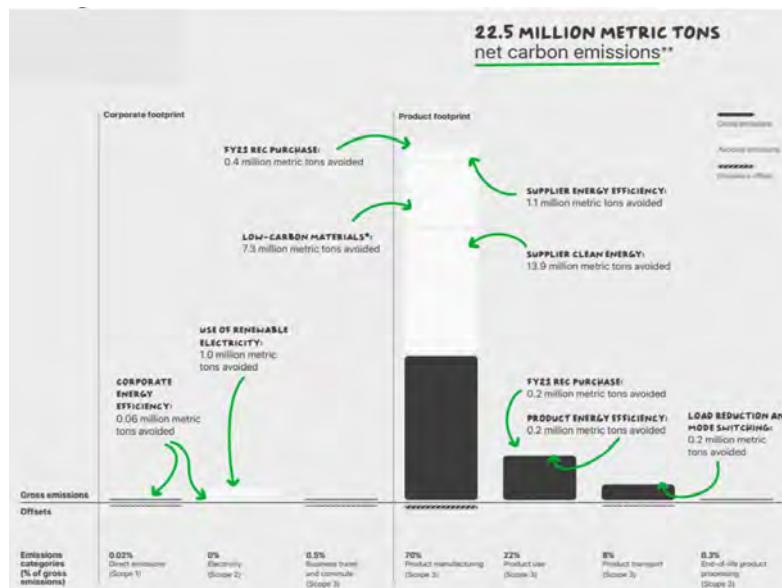


Figure 35:

The graph titled “22.5 Million Metric tons net carbon emissions” shows the three-scope system employed.

- Scope 1: (Direct emissions) is low
- Scope 2: Electricity that they purchase claims that Apple avoided 1.0 million metric tons of electricity use in the year 2022 (Note it was probably the year 2021 because the report came out in early 2022. )

In 2021, they saved 22.5M metric tons of carbon because they used renewable energy sources and worked with their suppliers. So, it is a 70 percent reduction compared to the status quo from 2015 as per the baseline. This is because of product manufacturing, energy efficiencies, and reduced product use. So, the trend is good in the way Apple talks about it. But of course, if you look at our numbers, they are pretty significant in terms of the overall carbon emission.

## 0.19 Response to the Climate Crisis

Microsoft's total emissions in 2021 were approximately 14 MtCO<sub>2</sub>eq

By 2023, Microsoft will be carbon negative, and by 2050, Microsoft will remove from the environment all the carbon the company has emitted directly or by electrical consumption since it was founded in 1975

<sup>30</sup>Environmental Progress Report, [www.apple.com/environment/pdf/Apple\\_Environmental\\_Progress\\_Report\\_2023.pdf](http://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2023.pdf). Accessed 31 Jan. 2024.

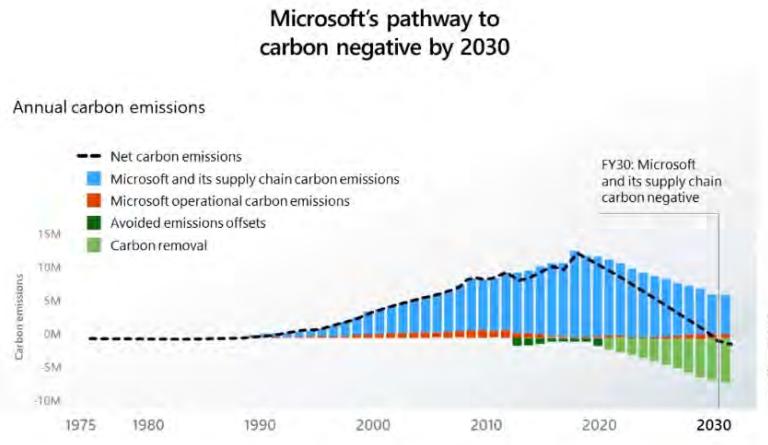


Figure 36:

Microsoft annual carbon emissions <sup>31</sup> Smith, Brad. “Microsoft Will Be Carbon Negative by 2030.” The Official Microsoft Blog, 23 July 2020, blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/.

## 0.20 Absorbing Carbon by Regenerating Nature — Removing Carbon Through Direct Capture

Destruction of rainforests that hold a significant amount of the carbon stored in terrestrial ecosystems contributes significantly to rising atmospheric CO<sub>2</sub> levels, while reductions in SOM levels from soil disturbance from mining can impact the infiltration of rainfall and the storage of soil moisture

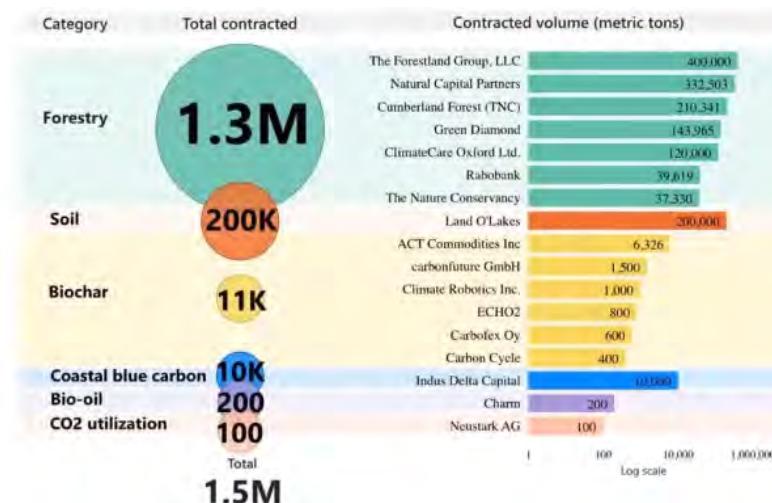


Figure 37:

FY22 additions to Microsoft's carbon removal portfolio (inclusive of volume contracted for current and future years) <sup>32</sup>

<sup>31</sup>Microsoft annual carbon emissions

<sup>32</sup>Bansal, Tima. “How Microsoft Is Leading the Response to the Climate Crisis.” Forbes, Forbes Magazine, 8 Nov. 2022, www.forbes.com/sites/timabansal/2022/06/22/how-microsoft-is-leading-the-response-to-the-

So, this is Microsoft's carbon footprint. The Figure titled Microsoft's Pathway to Carbon Negative by 2030 shows that Microsoft's and its supply chain's carbon emissions peaked around 2019. Microsoft lowers this to some extent by avoiding emissions and carbon removal. They are huge on carbon removal in terms of forestry, soil enrichment, and other natural processes that are used to either capture carbon or make sure that there will be a compensating effect because of the plantation and the trees that inhabit it. So they claim that their total emission in 2021 was 14 million metric tons of CO<sub>2</sub> equivalent and that they will be carbon negative (represented by the green portion of the graph). Microsoft claims that by 2050, they will not only be carbon negative but also remove any carbon emissions contribution they have had since their inception. Ambitious goals that are good to see if it comes to term

Reduce the energy consumption of computing tasks and hardware

## 0.21 How you can improve the efficiency of our computing infrastructure

There are many different approaches one could use to achieve computational sustainability. Below is a short list of techniques one can use to ensure sustainability within a computing infrastructure.

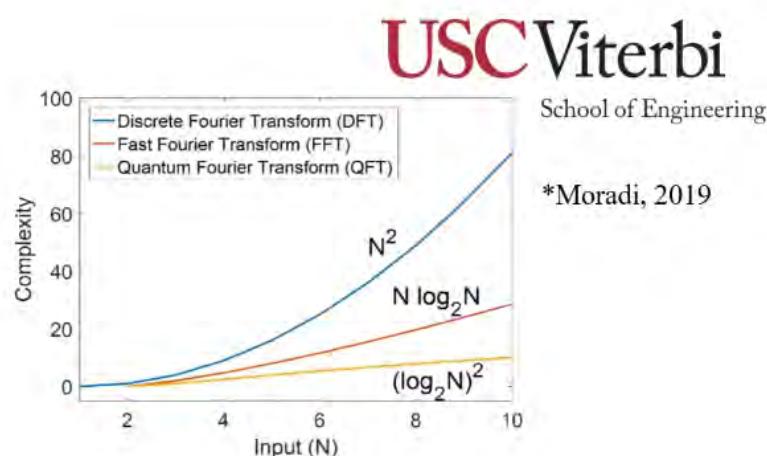


Figure 38:

Comparison of the computational complexities of DFT, FFT and QFT algorithms. <sup>3334</sup>

1. Developing new efficient algorithms at the application level

Fourier transform: DTF( $\mathcal{O}(N^2)$ ) –> FFT( $\mathcal{O}(N \log N)$ ) Now that we know the trends and some of the issues, how do you, as a graduate student, researcher, or competing expert, decide how to improve the efficiency of our computing infrastructure, from devices to circuits and systems, through systems and algorithms that run on it? Of course, there are several different techniques. I give some of them here to develop new efficient algorithms for the application. That is a well-known, straightforward example. I want to do a Fourier transform. I could use the DFT algorithm ( $\mathcal{O}(N^2)$ ), or should I use the FFT algorithm

[climate-crisis/?sh=f86343e50a12](https://climate-crisis/?sh=f86343e50a12).

<sup>33</sup>Comparison of the computational complexities of DFT, FFT, and QFT algorithms.

<sup>34</sup>Comparison of the Computational Complexities of DFT, FFT and QFT..., [www.researchgate.net/figure/Comparison-of-the-computational-complexities-of-DFT-FFT-and-QFT-algorithms\\_fig2\\_338980039](https://www.researchgate.net/figure/Comparison-of-the-computational-complexities-of-DFT-FFT-and-QFT-algorithms_fig2_338980039).

Accessed 31 Jan. 2024.

$(\mathcal{O}(N \log N))$ ? somebody had to develop LFT, but once it's developed, we use it. Now we go from  $(\mathcal{O}(N^2))$  to  $(N \log N))$ , and that's good.

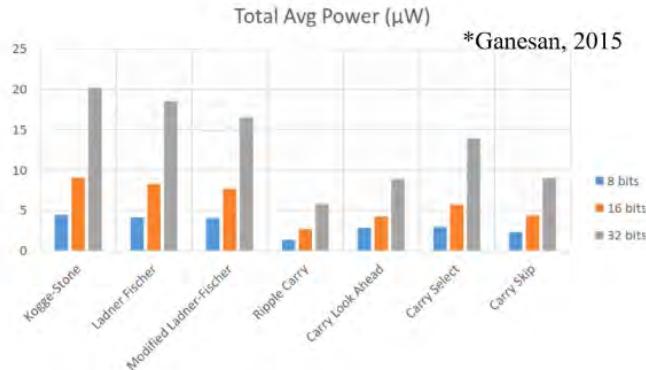


Figure 39:

Total Average Power Comparison for the Adder Topologies for 8, 16 and 32 Bit Sizes<sup>35</sup>

2. Reducing accuracy in favor of a reduction in the computational complexity  
Reduction in the number of input and filter bits to save training and computational resources

Another example would be training on a neural network. But instead of training using 32 floating point numbers, I'll do it using eight-bit floating point numbers. Or maybe I do it with 4-bit fixed point numbers. Right. Of course, there will be a bit of accuracy loss, but maybe one or two percent. But perhaps I can live with that because there are a lot of benefits (increase in energy efficiency and carbon emission reduction.)

3. Developing new models of computation  
Stochastic computing, neuromorphic computing, Ising machines, etc.

It is developing new models of computing stochastic computing, approximate computing, neuromorphic computing, and using machines. So there are usually beyond CMOS technologies that don't rely on teraflop computational power to do the work but on a completely different mechanism (dynamical system mechanism) that slows you to solve certain types of problems very efficiently. Of course, quantum computing could be the Holy Grail if you can develop a large quantum computing machine to solve complicated problems efficiently. I mean, that's not proven right. The best known is still an  $(N^2)$  Grover's algorithm, but the claims of quantum supremacy abound and have yet to be improved.

---

<sup>35</sup>Ganesan, Sarvesh. "Area, Delay and Power Comparison of Adder Topologies." TSW, 1 Dec. 2015, [repositories.lib.utexas.edu/items/d972b963-969d-487c-9d47-c177caff99af](http://repositories.lib.utexas.edu/items/d972b963-969d-487c-9d47-c177caff99af).

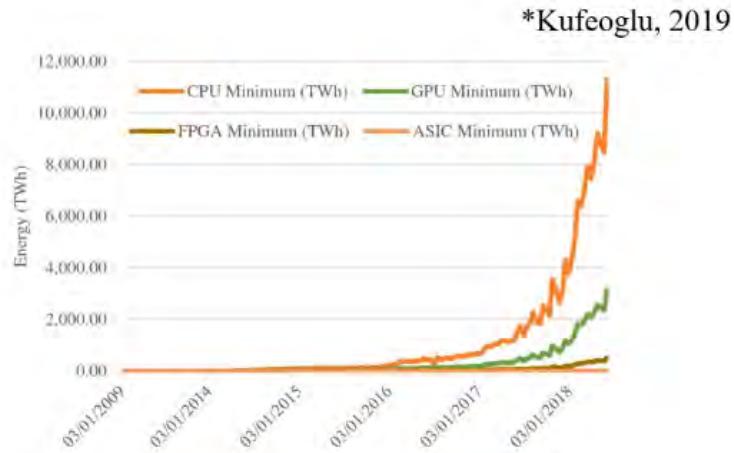


Figure 40:

CPU, GPU, FPGA, and ASIC minimum energy consumption between difficulty recalculation.<sup>36</sup>

#### 4. Application-specific hardware optimized for repetitive tasks

New circuits and custom computing architectures with sustainability in mind

This means finding a standard kernel and making sure you do a good at standard kernel Computations. A lot of the advanced energy efficiency approaches focus on that.

#### 5. New materials, devices, and fabrication techniques (beyond CMOS)

New materials and devices can reduce energy consumption and pollution.

Unfortunately, that's hard to do with the kind of devices we are trying to manufacture these days ( $3nm - 1nm$  CMOS devices). It is getting increasingly more challenging, but we try, and people are doing it.

#### 6. Use AI and ML to solve challenging problems

essentially replacing physical activity with virtual ones. So, using advanced technologies for computing, including decision-making, dynamical control with AI, and machine learning, would help improve energy efficiency.

## 0.22 Using AI to Improve Buildings

Commercial buildings

- 1.3 trillion kWh of electricity annually (1/3 of total US electricity generation)
- Annual energy costs > \$100 billion
- Poorly maintained, degraded, and improperly controlled equipment wastes 15% – 30% of energy in commercial buildings

<sup>36</sup>CPU, GPU, FPGA, and ASIC Minimum Energy Consumption Between ..., [www.researchgate.net/figure/CPU-GPU-FPGA-and-ASIC-minimum-energy-consumption-between-difficulty-recalulation\\_fig5\\_337886683](http://www.researchgate.net/figure/CPU-GPU-FPGA-and-ASIC-minimum-energy-consumption-between-difficulty-recalulation_fig5_337886683). Accessed 31 Jan. 2024.

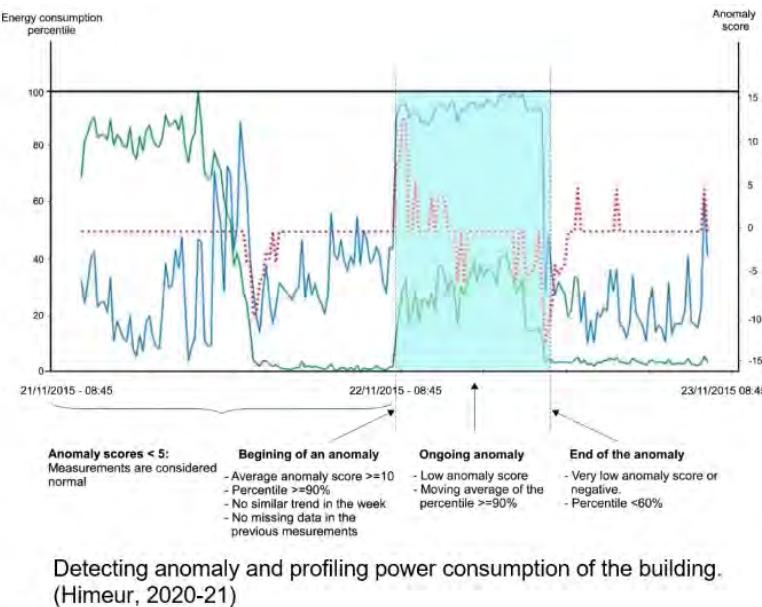


Figure 41:

Anomaly detection using AI and micro-moment analysis <sup>37</sup>

So, these are a few examples of the AI currently being used. I will not spend a lot of time in any one of them. I just picked a sample representative of how AI has been used to improve energy efficiency. So, this is a paper published by Himeur.

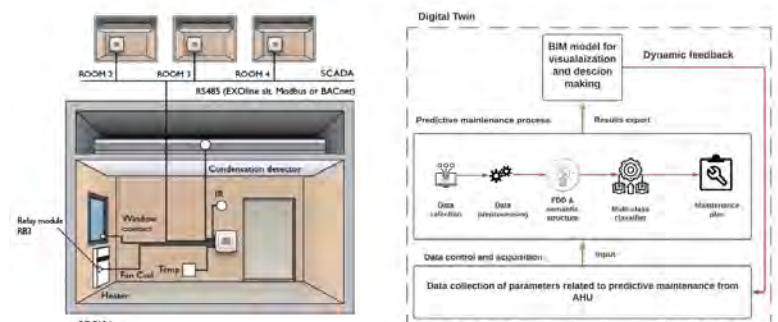


Figure 42:

An example of the application system <sup>38</sup>

The following paper published by Hosamo describes how they have used AI to improve energy efficiency in buildings. I remind you that this is a significant source of electricity consumption in the US. For example, one of the total energy consumption in the US is due to buildings, so being able to control the consumption of energy here is an important one, and usually, it's like

<sup>37</sup>Author links open overlay panelYassine Himeur a, a, b, c, et al. "Artificial Intelligence Based Anomaly Detection of Energy Consumption in Buildings: A Review, Current Trends and New Perspectives." Applied Energy, Elsevier, 9 Feb. 2021, [www.sciencedirect.com/science/article/pii/S0306261921001409](http://www.sciencedirect.com/science/article/pii/S0306261921001409).

<sup>38</sup>(PDF) a Digital Twin Predictive Maintenance Framework of Air ..., [www.researchgate.net/publication/358906612\\_A\\_Digital\\_Twin\\_Predictive\\_Maintenance\\_Framework\\_of\\_Air\\_Handling\\_Units\\_based\\_on\\_Automatic\\_Fault\\_Detection\\_and\\_Diagnostics](http://www.researchgate.net/publication/358906612_A_Digital_Twin_Predictive_Maintenance_Framework_of_Air_Handling_Units_based_on_Automatic_Fault_Detection_and_Diagnostics). Accessed 31 Jan. 2024.

## 0.23 Using machine learning and AI for landscape connectivity and wildlife corridors

when you have a water leak, a lot of water is wasted before you detect there. It's a water leak, and you bring somebody to fix it. So you could use AI to detect anomalies in the building's electrical energy consumption and then know at least there is a problem and go on trying to fix it right. Maybe the fixes turn off the switch, but it could be more severe than that.

## 0.23 Using machine learning and AI for landscape connectivity and wildlife corridors

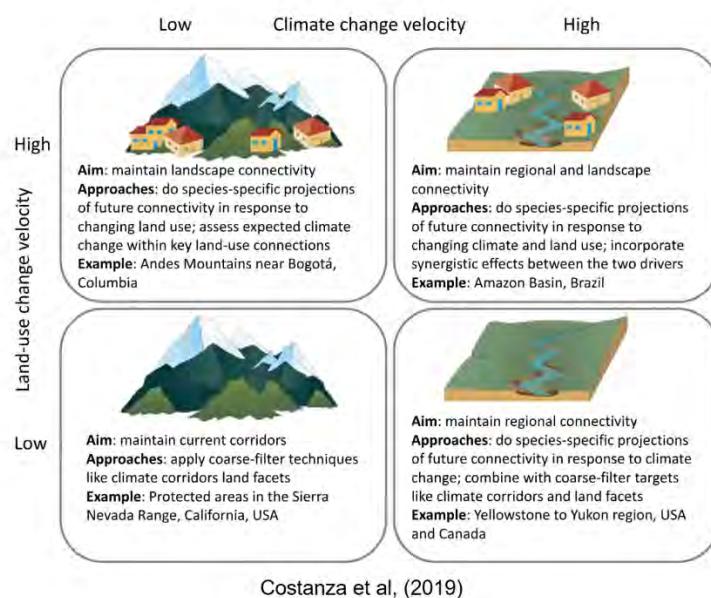
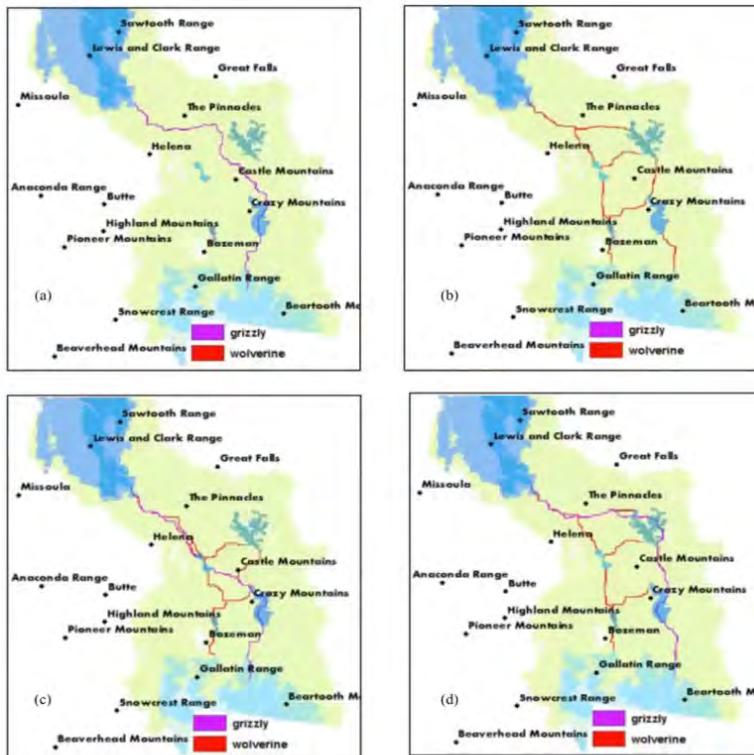


Figure 43:

Suggested approaches for identifying a corridor network that promotes connectivity for both CC and LUC adaptation, given the CC and LUC velocities in a location. See Table 2 for more information on the approaches listed. Images courtesy of the Integration and Application Network, University of Maryland Center for Environmental Science ([ian.umces.edu/symbols/](http://ian.umces.edu/symbols/))<sup>39</sup>

So AI and conservation people are using it to provide landscape connectivity for animals to move in their domain without being hit by cars or being restricted to particular areas that do not have enough resources. So, a study of AI in this domain was conducted by Costanza in 2019.

<sup>39</sup>(PDF) Landscape Connectivity Planning for Adaptation to Future ...[www.researchgate.net/publication/330857374\\_Landscape\\_Connectivity\\_Planning\\_for\\_Adaptation\\_to\\_Future\\_Climate\\_and\\_Land-Use\\_Change](https://www.researchgate.net/publication/330857374_Landscape_Connectivity_Planning_for_Adaptation_to_Future_Climate_and_Land-Use_Change). Accessed 31 Jan. 2024.



Dilkina et al, (2017)

Figure 44:

Designing cost-effective biodiversity corridors <sup>40</sup>

## 0.24 How computing can address the challenges of poverty

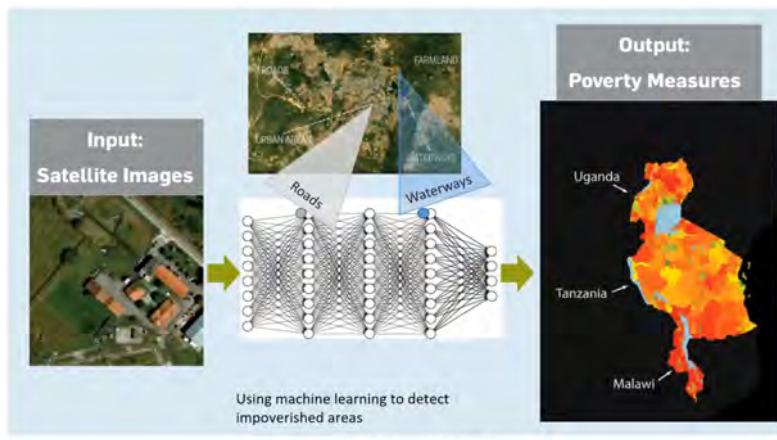


Figure 45:

<sup>40</sup>Reiter, Chuck Rhoades and Erika. "Rocky Mountain Research Station." US Forest Service Research and Development, 12 Jan. 2024, [www.fs.usda.gov/rmrs/science-spotlights/designing-cost-effective-biodiversity-corridors](http://www.fs.usda.gov/rmrs/science-spotlights/designing-cost-effective-biodiversity-corridors).

Transfer learning is an effective model approach and predicts socioeconomic indicators in data-scarce regions that take advantage of globally available satellite images, updated frequently, and becoming increasingly more accurate.<sup>41</sup>

People have used AI to identify from satellite images and some other information, perhaps precisely where the pockets of poverty are. Of course. In a country like ours, where there's a lot of data, maybe that's less important. But in countries that don't have the proper infrastructure or data gathering mechanisms and reporting mechanisms, this kind of AI-based imaging analysis could be beneficial to identify the parts of poverty that require special attention. I'm sure in a case like the recent earthquake in Turkey, This kind of technology could be used effectively and is probably being used.

## 0.25 Switching energy and switching speed

### 0.26 By shifting to alternative technologies such as superconductor electronics for specific applications, the energy cost can be significantly reduced

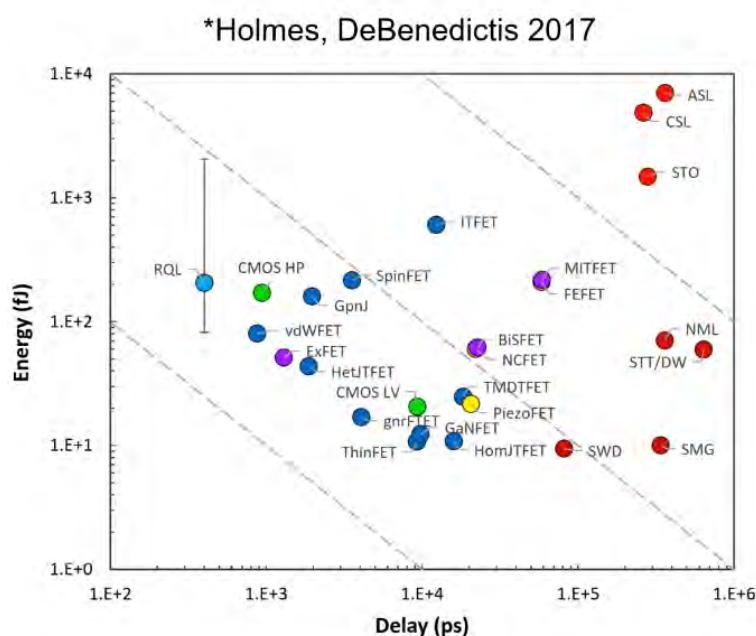


Figure 46:

Switching energy at 300 K versus delay for 32-bit ALUs. Added to Fig. 6 is an RQL superconducting ALU with whiskers showing a range for refrigeration cost from 10,000 to 400  $\frac{w}{\text{K}}$  ( $300\text{K}/4\text{K}$ ) and a symbol at  $1,000 \frac{w}{\text{K}}$ . Dashed lines show constant energy-delay products.<sup>42</sup>  
<sup>43</sup>

<sup>41</sup>(PDF) Computational Sustainability: Computing for a Better World ..., [www.researchgate.net/publication/335373363\\_Computational\\_sustainability\\_Computing\\_for\\_a\\_better\\_world\\_and\\_a\\_sustainable\\_future](http://www.researchgate.net/publication/335373363_Computational_sustainability_Computing_for_a_better_world_and_a_sustainable_future). Accessed 31 Jan. 2024.

<sup>42</sup>DeBenedictis, Erik P. "Quantum Computer Control Using Novel, Hybrid Semiconductor-Superconductor Electronics." arXiv.Org, 15 Dec. 2019, [arxiv.org/abs/1912.11532](https://arxiv.org/abs/1912.11532).

<sup>43</sup>D. E. Nikonorov and I. A. Young, "Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits," IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, vol. 1, pp. 3–11,

Technology	Switching energy	Frequency
CMOS (Charge-transfer based)	$10^6 K_B T \ln 2$	$\sim 6\text{GHz} @ 77\text{K}$ and above
Single Flux Quantum (RSFQ)	$10^3 K_B T \ln 2$	$\sim 50\text{GHz} @ 4.2\text{K}$
Reversible Computing (AQFP)	$\sim K_B T \ln 2$	$\sim 5\text{GHz} @ 4.2$

Table 1:

So, I have briefly mentioned that the end of CMOS law is insight. No one argues against that now, so maybe another 5, 7, or 10 years of life is left in Moore's law, and that's it. So we must find beyond CMOS to have the 'free lunches' we have enjoyed for the last 50 years.

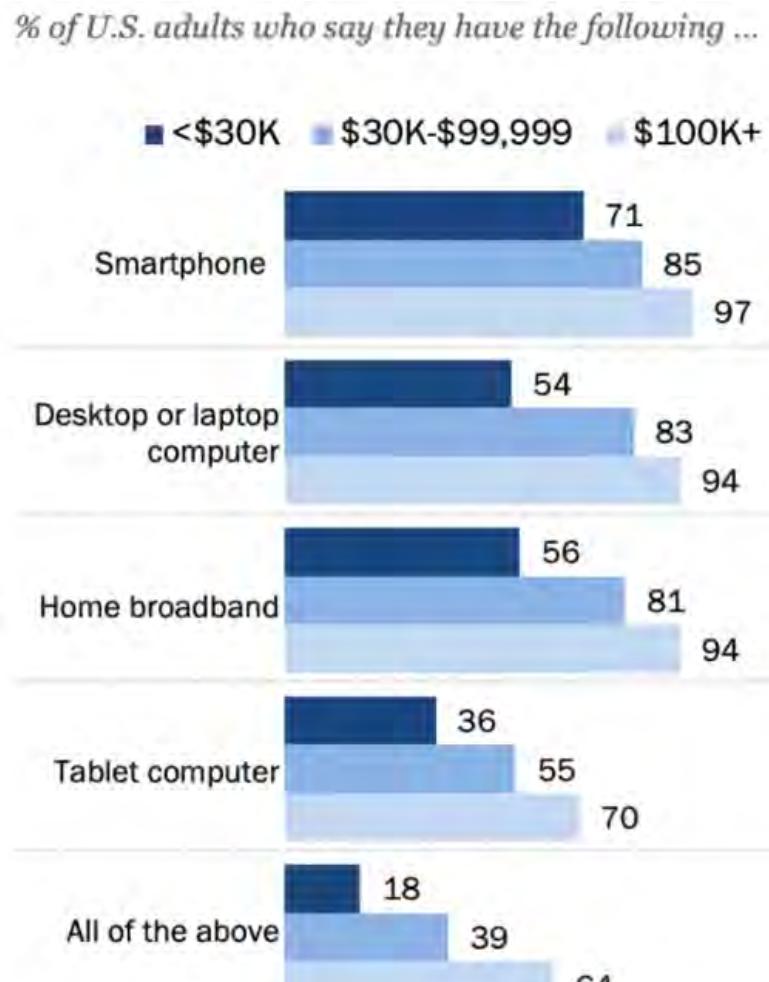
There are many competing technologies, one that we especially like, and the one we have been working on in this expedition project is single flux quantum logic and superconductor electronics.

The main reason for this interest is that if you look at just the switching energy consumption, CMOS versus singular flux quantum versus reversible computing.

(Switching energy) This is the minimum, meaning you have to be at least three times that to have a signal value we can detect. Ok, so  $10^6$  for CMOS; single flux quantum  $10^3$ . Reversible computing using an idiomatic version of sfq logic is the order of the thermal knowns. In the best case, six orders of magnitude improvement in energy efficiency and performance is quite acceptable.

(For AQFP Reversible Computing) we are at 5 GHz at 4.2 kelvin. The big problem is that you have to cool down 4.2 kelvin from room temperature, which may mean a factor of 500 – 1000 loss in energy efficiency itself. The Carnot cycle efficiency is about 100 or so, but effectively, 500 would be pretty good and, more realistically, probably 1000. Despite that, this is still a good win.

## 0.27 Effect of IT infrastructure on equity and equality



\*Ubalde, (LinkedIn 2020)

Figure 47:

Lower-income Americans have lower levels of technology adoption<sup>44</sup>

<sup>44</sup>"Lower-Income Americans Have Lower Levels of Technology Adoption." Pew Research Center, Pew Research Center, 6 May 2019, [www.pewresearch.org/ft\\_19-05-06\\_digitaldivideincome\\_lowerincomeamericanslowertechadoption/](http://www.pewresearch.org/ft_19-05-06_digitaldivideincome_lowerincomeamericanslowertechadoption/).

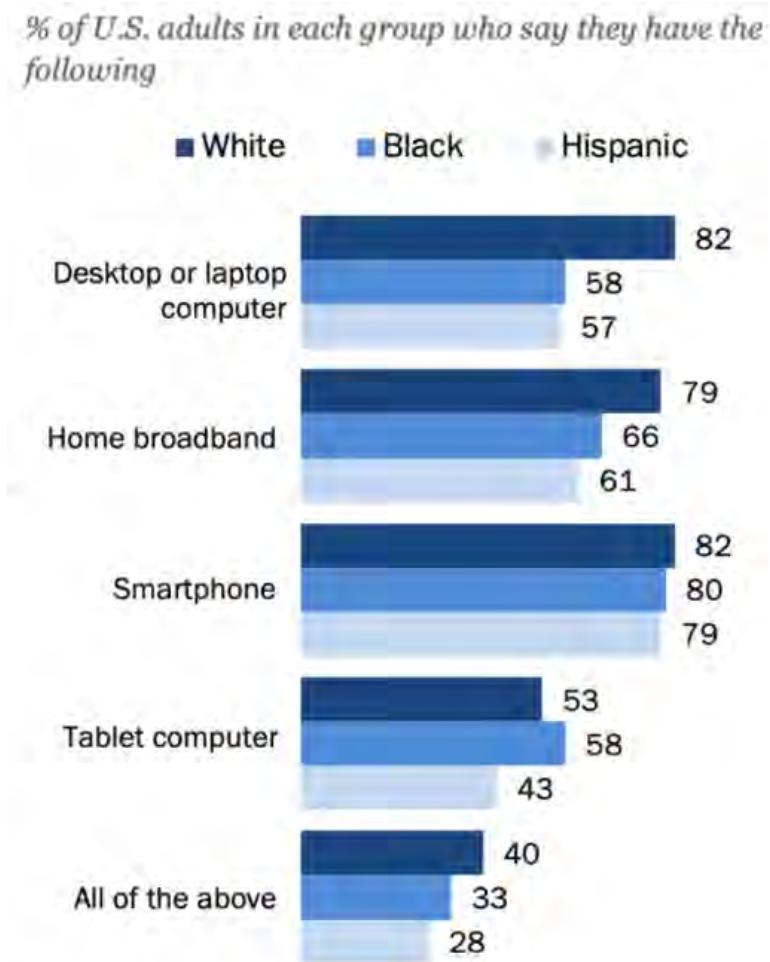


Figure 48:

Blacks and Hispanics own mobile devices at similar rates to whites <sup>45</sup>

- Access to IT is dependent on income and uneven across different populations
- Many women deemphasize their careers when they start families
- Flexible work eliminates the physical headquarters as the focal point of a company and instead leverages technology to give employees control over when and where they work and how they engage and respond
- Reduced transportation and travel leads to less power consumption
- Remote work tends to provide higher equity

<sup>45</sup>"Blacks and Hispanics Own Mobile Devices at Similar Rates to Whites." Pew Research Center, Pew Research Center, 30 Aug. 2017, [www.pewresearch.org/fact-tank/2017/08/30/mobile-devices-devices3/](http://www.pewresearch.org/fact-tank/2017/08/30/mobile-devices-devices3/).

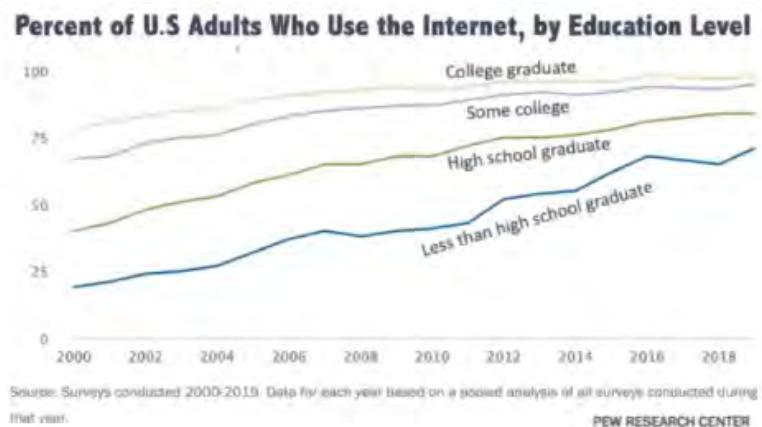


Figure 49:

Internet use by education <sup>46</sup>

Information accessibility is quite essential. Some well-known studies show information is rarely provided to people in different parts of society based on their income, gender, race, etc. Usually, people with higher incomes, white and male, have better access to these resources, which gives you knowledge, power, and agility to respond. So, it's essential to ensure accessibility and equity in distributing these resources and these capabilities across our society for it to prosper.

## 0.28 Transformation to a sustainable and resilient society that ensures the safety and security of its people

- Balances economic advancement with resolving social problems by a system that highly integrates cyberspace and physical space
- Transformation into a sustainable and resilient society through the fusion of cyberspace and physical space
- Creation of “Knowledge” as a source of value creation by designing a new society
- Development of human resources to support a new society

<sup>46</sup>“Internet Use by Education.” Pew Research Center: Internet, Science Tech, Pew Research Center, 11 Jan. 2017, [www.pewresearch.org/internet/chart/internet-use-by-education/](http://www.pewresearch.org/internet/chart/internet-use-by-education/).

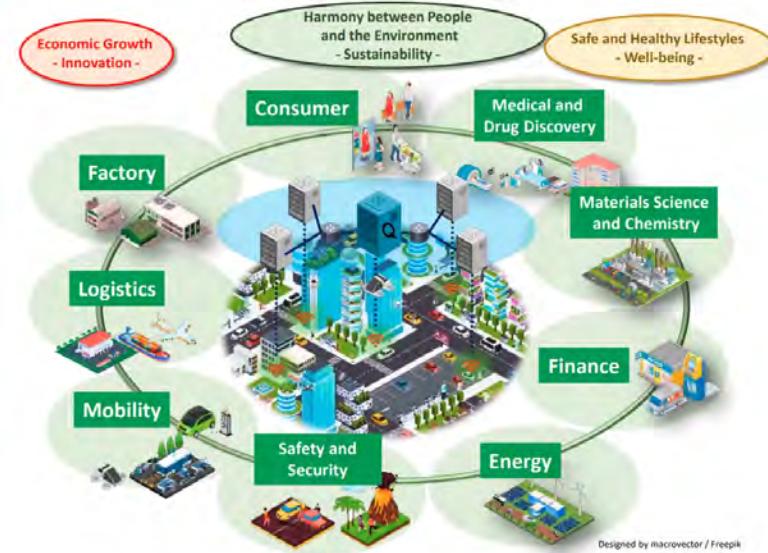


Figure 50:

#### Values created by Quantum Technology in the Future Society <sup>47</sup>

Cultural sustainability, which I have mentioned in the above section (environmental issues), is a combination of environmental issues, economic growth, safety, healthy lifestyles, and individuals' overall well-being. Again, you could use various approaches to analyze this, to try to encourage more culturally sustainable practices, and so on (beyond the scope of this section).

### 0.29 Educating the new generation and workforce development/retraining



<https://www.wastetoenergysystems.com/wp-content/uploads/2015/06/Sustainability-In-Schools.jpg>

Figure 51:

#### Should Kids be Taught Sustainability in School? <sup>48</sup>

<sup>47</sup>Vision of Quantum Future Society, [www8.cao.go.jp/cstp/english/outline\\_vision.pdf](http://www8.cao.go.jp/cstp/english/outline_vision.pdf). Accessed 31 Jan. 2024.

<sup>48</sup>Woods, Alissa. "Should Kids Be Taught Sustainability in School?" Waste to Energy Systems, 24 June 2015, [www.wastetoenergysystems.com/should-kids-be-taught-sustainability-in-school/](http://www.wastetoenergysystems.com/should-kids-be-taught-sustainability-in-school/).



Figure 52:

#### Whole school approach to Sustainable Development <sup>49</sup>

So, finally, being an educator and teacher myself, I recognize that many of you are students (or also faculty); we know that education will have considerable value for an individual but also for a society that has individuals educated in technology and all manners of natural science. So we need to have mechanisms not only for training a new generation of the workforce but also for retraining the existing ones to make them adapt to using the latest and best technology.

#### Let's build sustainability into our digital ecosystem!

So, I'd like to conclude this portion of the notebook here. Let's build sustainability into our digital ecosystem in the future. Develop society 5.0 as some people refer to it. Overcoming many of the issues we face today, improving our quality of life, and ensuring we are not sacrificing or compromising our children's future.

---

<sup>49</sup>“Whole School Approach to Sustainable Development.” Whole School Approach, 10 June 2021, [wholeschoolapproach.lerenvormorgen.org/en/](http://wholeschoolapproach.lerenvormorgen.org/en/).

# **Introduction to Quantum Computing: Qubits, Gates, and Algorithms**

## Outline

1. Introduction to quantum computing
2. Superconducting qubits
3. Engineering quantum systems
4. 3D integration

## 0.30 Introduction

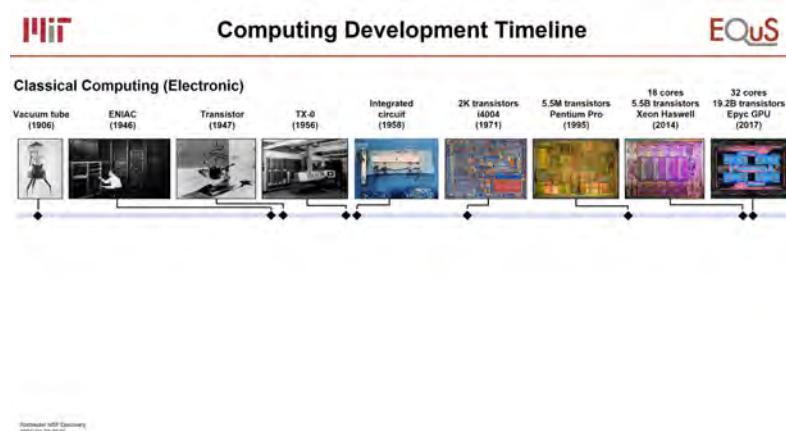


Figure 1:

I am really pleased and honored to have been asked to write an introduction to quantum computing, which will partially focus on superconducting qubits.

I will not be giving a lot of research results from my group (Engineering Quantum Systems), but more of an overview of the field. And I'm hoping to get back to Rochester on another occasion, maybe to give a more in-depth technical talk on my results. But getting into this, I think it's beneficial to take a step back and remember what the last century's development of electronic and classical computing looked like.

### 0.31 Computing Development Timeline

## Vacuum tube (1906)



Figure 2:

In 1906, we invented the vacuum tube, a three-terminal device, and it was used for radio transceivers for decades.

# ENIAC (1946)



Figure 3:

It was 40 years before it was used in a computer that we know as ENIAC—developed at the University of Pennsylvania around 1946.

# Transistor (1947)



Figure 4:

A year later, at Bell Labs, the bipolar junction transistor was invented. Within about ten years, we had an entire transistor computer called TX-0 or tixo developed at Lincoln Lab and MIT. You can see it takes up a whole room, and in fact, it was just transistors soldered together, and it used a magnetic core memory. We needed the integrated circuit, developed in the late 50s, to get closer to today.

# 2K transistors i4004 (1971)

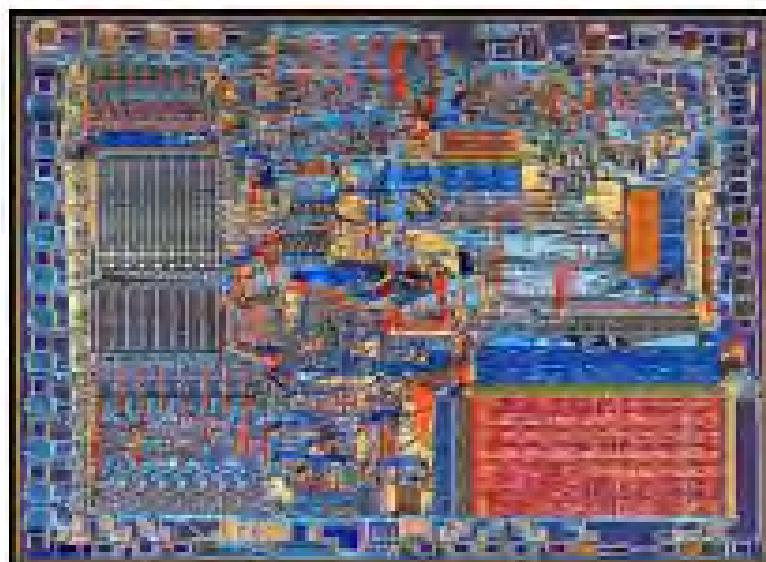


Figure 5:

You would have to wait until the 70s to start to see processors that we would recognize (a few 1000 transistors) like Intel 4004, 8008, etc.

# **5.5M transistors Pentium Pro (1995)**

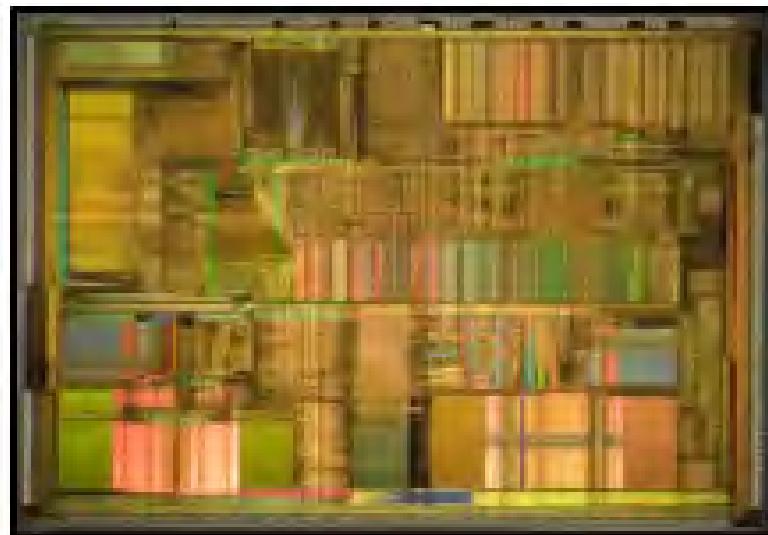


Figure 6:

It took another 25 years to get to the Pentium Pro with millions of transistors. That was the year I graduated from Rochester.

# 18 cores 5.5B transistors Xeon Haswell (2014)

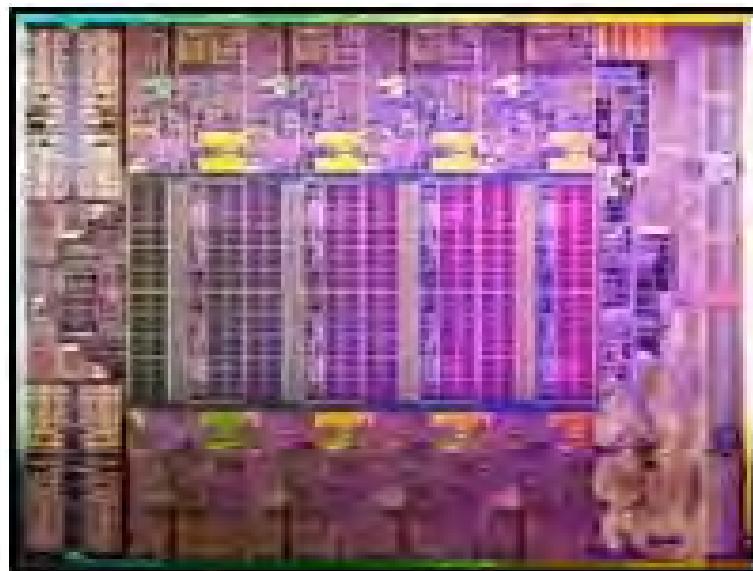


Figure 7:

Then it took another 20-30 years to get to where we are today with multicore processors, billions of transistors, GPUs, and now even approaching 100 billion transistors. So, these wonderful devices we are using today took over 100 years of development. **We can contrast quantum computing with that. Quantum computing is a much younger, newer field.**

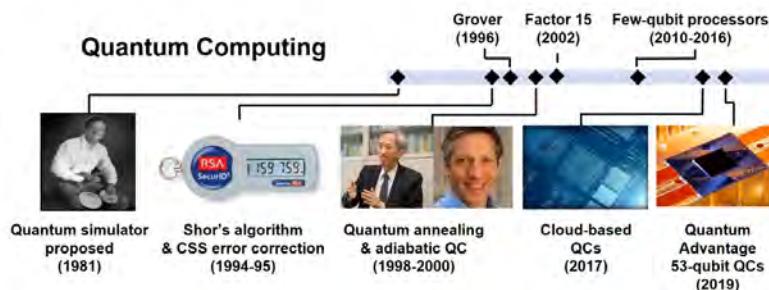


Figure 8:



## **Quantum simulator proposed (1981)**

Figure 9:

In the early 80s. Richard Feynman. Yle Manon and others suggested that if you want to simulate a quantum system, we should use a quantum system to do that, and that's a good idea.



# **Shor's algorithm & CSS error correction (1994-95)**

But it took 15 years to start to come up with the first algorithm that we would identify as applicable, such as Shor's algorithm for crypto analysis, Grover's algorithm for searching an unsorted database, ideas around quantum annealing, adiabatic quantum computing, and importantly, the advent of quantum versions of error correcting codes which would allow us to, at least in principle, make robust machines from faulty devices. Again, this mimics what we went through in the 1950s with electrical and optical communication systems and the computer systems themselves.



## **Quantum annealing & adiabatic QC (1998-2000)**

Figure 10:

Even so, it's another 20 years to get to where we are today, where we do have small-scale quantum computers available in the cloud, maybe 50 –100 qubits or so, but there's still a long way to go.

**Quantum computing is transitioning from scientific curiosity to technical reality.**

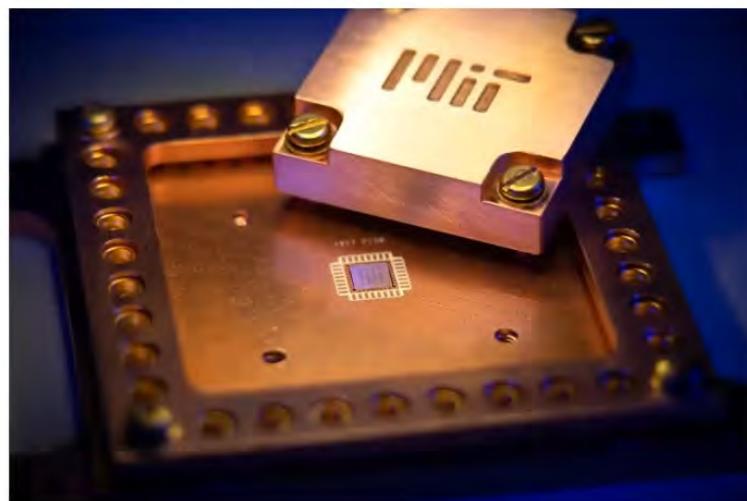
**Advancing from discovery to useful machines takes time & engineering**

**You must be in the game to play**

The takeaway is that quantum computing is natural; there's a lot of promise. It's transitioning from a laboratory or scientific curiosity to a technical reality that's happening. But we know, particularly as engineers, that advancing from discovery to useful machines will take time, and there's still more science to do. Maybe the third takeaway is that you need to be in the game to play. It may not be a surprise that many companies making these fantastic computer chips today have been in this field for decades and decades. They didn't just jump in at the last minute, and we can see that as China tries to develop its base of semiconductor development, it's not easy.

	Classical Computer
<b>Fundamental Logic element</b>	"Bit": classical bit (transistor, spin in magnetic memory, ...)
<b>State</b>	<b>0 or 1</b>
<b>Measurement</b>	Discrete states Deterministic measurement: EX Set as 1, measure as 1

**32-pin Package  
5x5 mm<sup>2</sup> silicon qubit chip**



Y. Sung, ..., WDO, Nature Communications (2019)

Figure 11:

With this, I'd like to introduce quantum computing.

I'll then talk about superconducting qubits at a high level and some high-level views of what it will take to engineer robust quantum systems. Finally, I'll show a few pictures at the end of some 3D integration work we've done here at MIT.

## 0.32 How is a Quantum Computer Different?

So, first off, is how a quantum computer is different? Classical bits are based on digital binary logic, typically a 0 or 1. The transistor has a voltage, or it doesn't, and these transistors rarely fail right; the error rate is probably better than one part in  $10^{-20}$ . There's no noise in this sense. So, states are discrete, and measurement is deterministic. This means that if we set it as 1, we will measure it as 1. 0 is a 0. If this falls apart for some reason and I get knocked off, it's probably because the software crashed, not because of a transistor error.

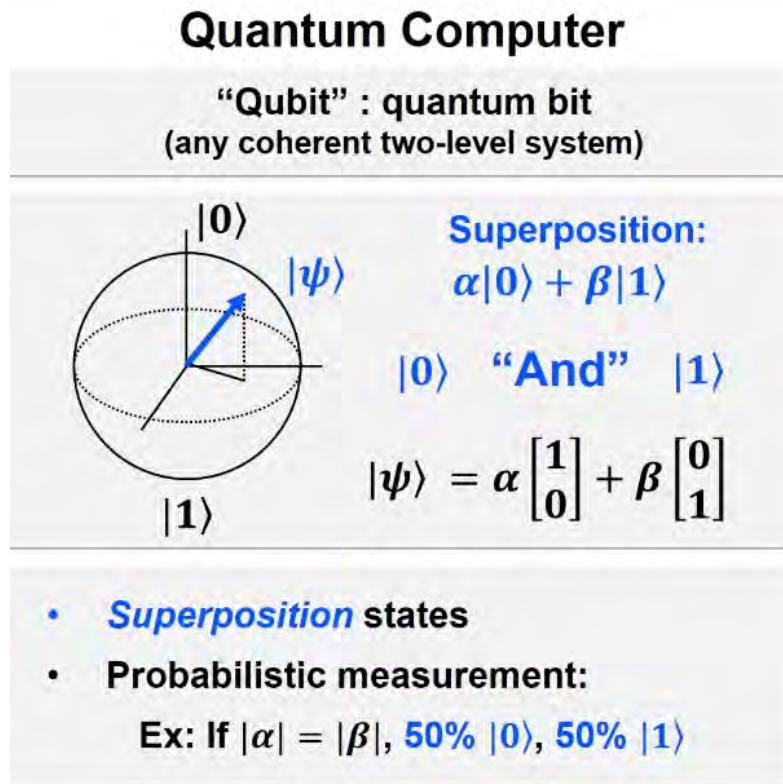


Figure 12:

Quantum computers are quite different in that their quantum systems, so quantum bit or Qubit still has two levels, 0 and 1, but they're quantum mechanical. We think of them on this sphere, which we call the Bloch sphere or planet Earth, and the Qubit can form a superposition of both 0 and 1 simultaneously; there are manifestly quantum mechanical states. One consequence is that measurement is probabilistic, as we'll see in a few minutes. So if we happen to be on the equator or prepare the qubit on the equator with an equal weight of 0 and 1, then half the time when we make a measurement, that qubit vector will snap back to the North Pole zero. Half the time it will go to one, even when we identically prepare it in the same place on the equator repeatedly. So that's very different.

**Quantum computers rely on encoding information in a fundamentally different way than classical computers**

Figure 13:

The question is because it's very different, fundamentally different. How can we leverage this?

Classical Computer	
Fundamental logic element	"Bit" : classical bit (transistor, spin in magnetic memory, ...)
Computing	<ul style="list-style-type: none"> <li>• N bits: <b>One N-bit state</b> <b>000, 001, ..., 111 (N = 3)</b></li> <li>• Change a bit: <b>new calculation</b> (classical parallelism)</li> </ul> <pre> graph LR     A[000] --&gt; B[Computer]     B --&gt; C[f(000)]     D[001] --&gt; B     B --&gt; E[f(001)]   </pre>

Figure 14:

One insight into that is, if we think again about a classical computer with  $N$  bits, there is an exponentially large number of states we can represent ( $2^N$ th), but we can only represent one at a time. So, if  $N$  is three, then  $2^3$  is eight. We have all the combinations 000 to 111. But we only have two choices if we want to process each of those as an input to get some function as an output. One would be taking one computer and feeding one input at a time. // Or take  $2^N$ th computers in parallel and feed all of these input states in parallel. Still, either way, with an exponentially large number of states, it would take either exponential time or exponential hardware to do this.

## Quantum Computer

**“Qubit” : quantum bit  
(any coherent two-level system)**

- **N qubits:  $2^N$  components to one state**

$$\alpha|000\rangle + \beta|001\rangle + \dots + \gamma|111\rangle \quad (N = 3)$$

- **Quantum parallelism & interference**

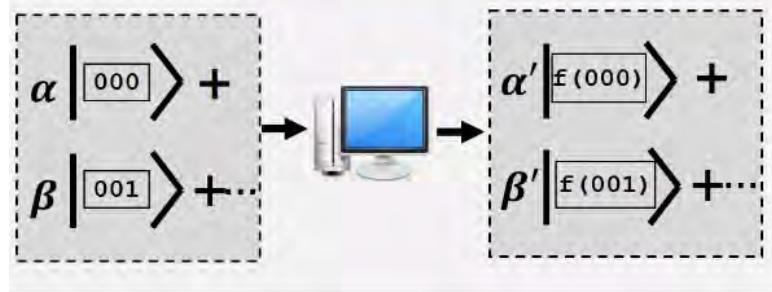


Figure 15:

### 0.33 Classical and Quantum Bits

A quantum computer can take  $N$  qubits, and we have  $2^N$  components (all of those components can be put into a single superposition state, each one with its own weight coefficient here represented as  $\alpha$ ,  $\beta$ , and  $\gamma$ ). So, notionally, we can take that input state (massive superposition state), operate on it with a quantum computer, and have an enormous superposition state at the output. We'll talk about how this collapses down in measurement, but what the critical ingredients are for this to happen and to derive advantage is quantum parallelism or quantum version of parallelism and quantum interference. We'll have some imagery of that in just a moment. So the question is how to take advantage of this?

## Three spins

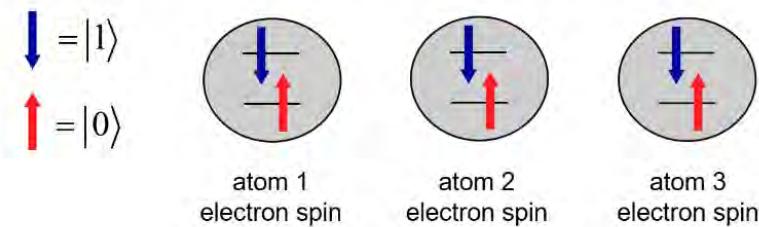


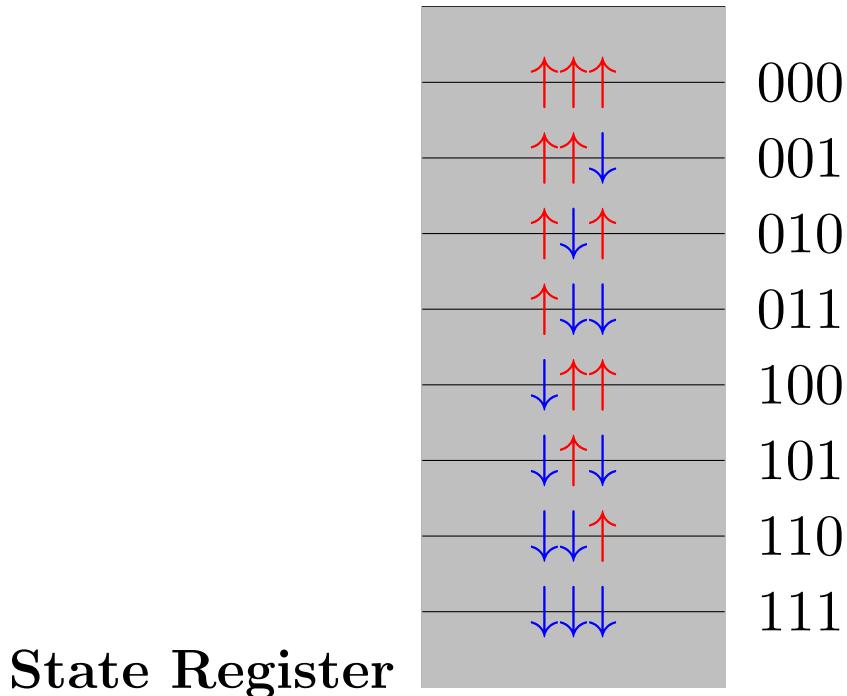
Figure 16:

So, let us take a concrete example of three spins. These could be electrons in an atom and say a

magnetic field. They have a spin-up and a spin-down with different energies. We associate spin up with zero and spin down with state one.

**eight( $2^N$ ) classical states**

(classical parallelism) ↓



Now, we could use these as classical bits, and we would then have  $2^N$  (again Eight) states in our state register from all spins up to all spins down or equivalently to 100 to 111. On the other hand, we could use these in their quantum mechanical nature and make a single quantum state, which we call  $\psi$ .

Comprising all of these components with their weight coefficients  $c_1$  through  $c_8$  properly normalized, it looks more digital if we write the numbers 0 and 1. *Single quantum state*

$$\psi = c_1 |\uparrow\uparrow\uparrow\rangle + c_2 |\uparrow\uparrow\downarrow\rangle + c_3 |\uparrow\downarrow\uparrow\rangle + c_4 |\uparrow\downarrow\downarrow\rangle + c_5 |\downarrow\uparrow\uparrow\rangle + c_6 |\downarrow\uparrow\downarrow\rangle + c_7 |\downarrow\downarrow\uparrow\rangle + c_8 |\downarrow\downarrow\downarrow\rangle \quad (0.1)$$

↓

$$= c_1 |000\rangle c_2 |001\rangle + c_3 |010\rangle + c_4 |011\rangle + c_5 |100\rangle + c_6 |101\rangle + c_7 |110\rangle + c_8 |111\rangle \quad (0.2)$$

**Quantum superposition state: eight complete numbers**

$$2^N \rightarrow 2^3 = 8 \rightarrow \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$$

$c_3$

**State Register**

$C_1$	$\uparrow\uparrow\uparrow$	000
$C_2$	$\uparrow\uparrow\downarrow$	001
$C_3$	$\uparrow\downarrow\uparrow$	010
$C_4$	$\uparrow\downarrow\downarrow$	011
$C_5$	$\downarrow\uparrow\uparrow$	100
$C_6$	$\downarrow\uparrow\downarrow$	101
$C_7$	$\downarrow\downarrow\uparrow$	110
$C_8$	$\downarrow\downarrow\downarrow$	111

We know that this single state comprises itself (single state), but it comprises eight numbers, and in fact, they are complex numbers  $c_1 c_8$ , so 16 real numbers if you want to think of it that way. So, in fact, we can change our state register to a quantum state register by placing these coefficients, or as we call them, probability amplitudes, in front of their corresponding aspect.

**Quantum superposition & gates:**  
**Quantum parallelism**  
**Quantum interference**

Figure 17:

With this Quantum state register, we can look at what it means to have quantum parallelism or quantum interference from the perspective of computation.

## 0.34 Quantum Parallelism

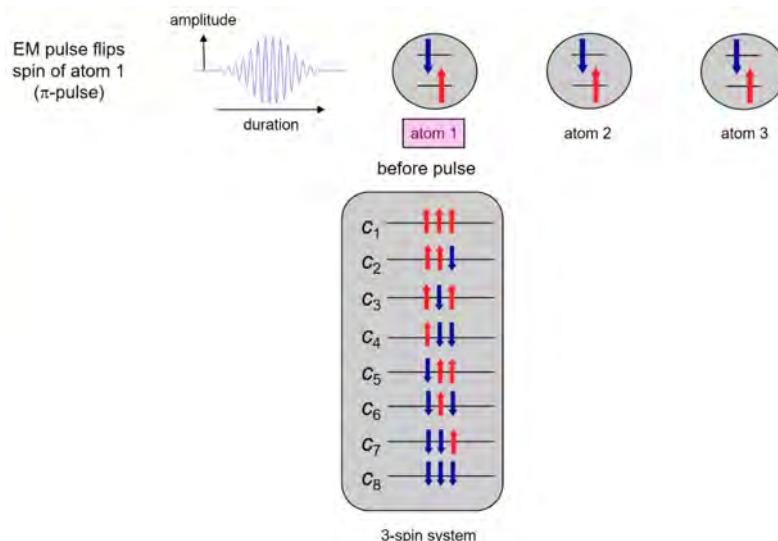


Figure 18:

As a first example, let us imagine that we will perform an operation on this first spin in Atom 1.

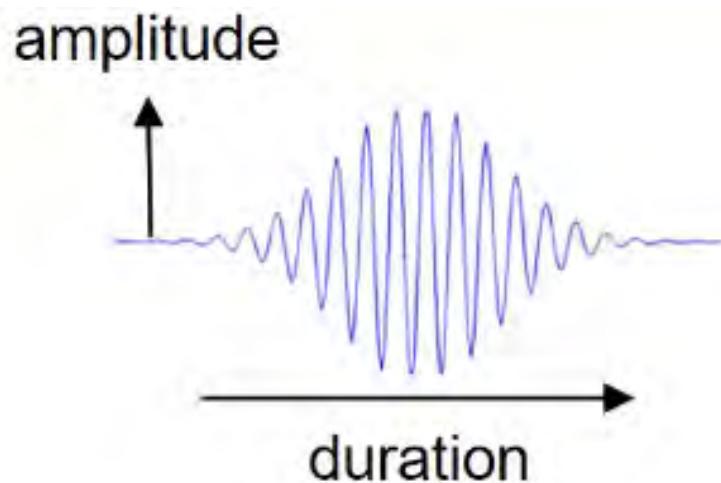


Figure 19:

What you see represents an electromagnetic pulse; this pulse will drive that transition, and it has a certain amplitude and a certain duration chosen so that it will rotate that spin 180 degrees.

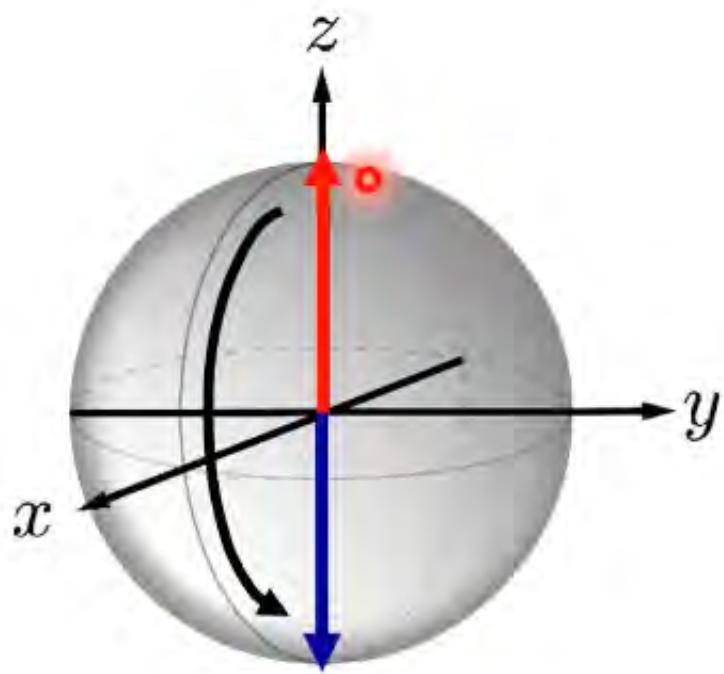


Figure 20:

If we look at our planet Earth or the block sphere, a spin up is pointed towards the North Pole, and this pulse would flip that spin to the South Pole.

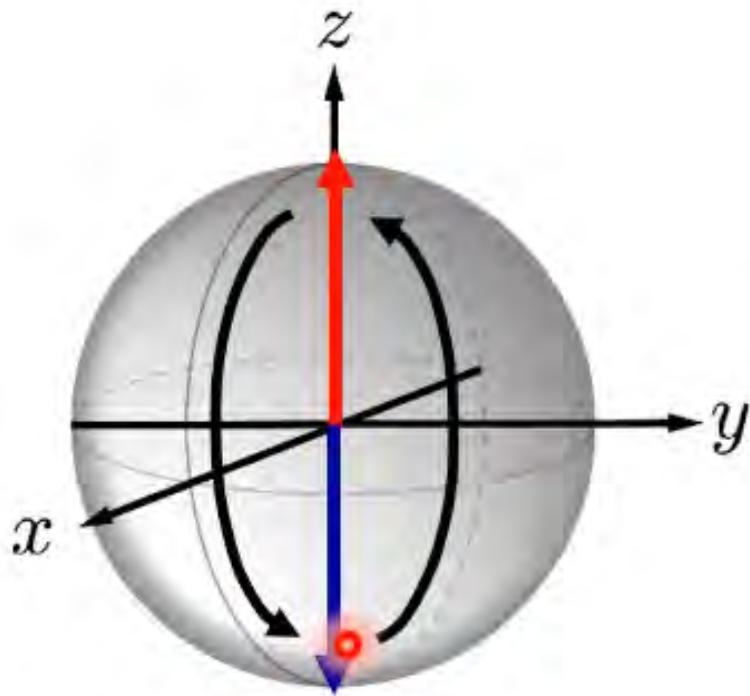


Figure 21:

Now for all of the blue spins at the South Pole, that same pulse would take it and flip it back to the North Pole

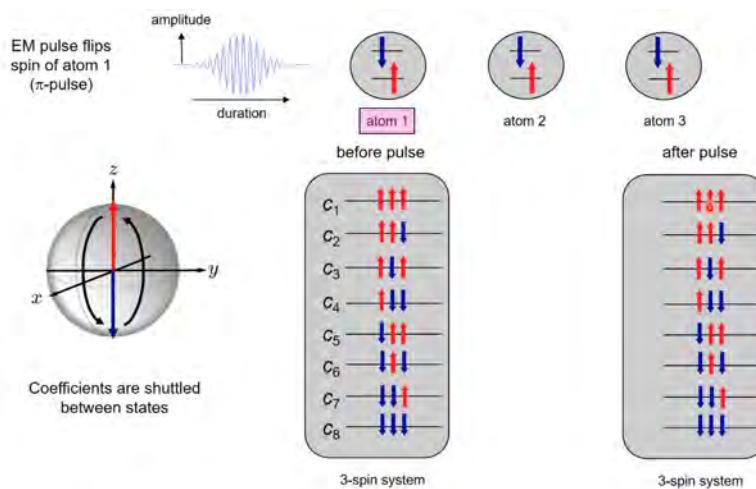


Figure 22:

You can see in our State Register we have eight of these different aspects; four of them the first atom has a spin up and four of them aspects with spin down. Basically, what this pulse will do is it will shuttle coefficients between different states.

So, if we look at before and after, here's what happens after the pulse.

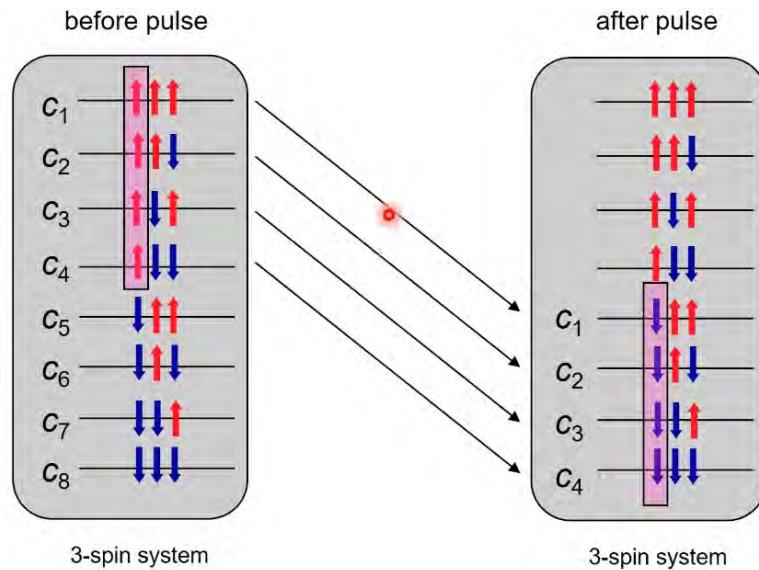
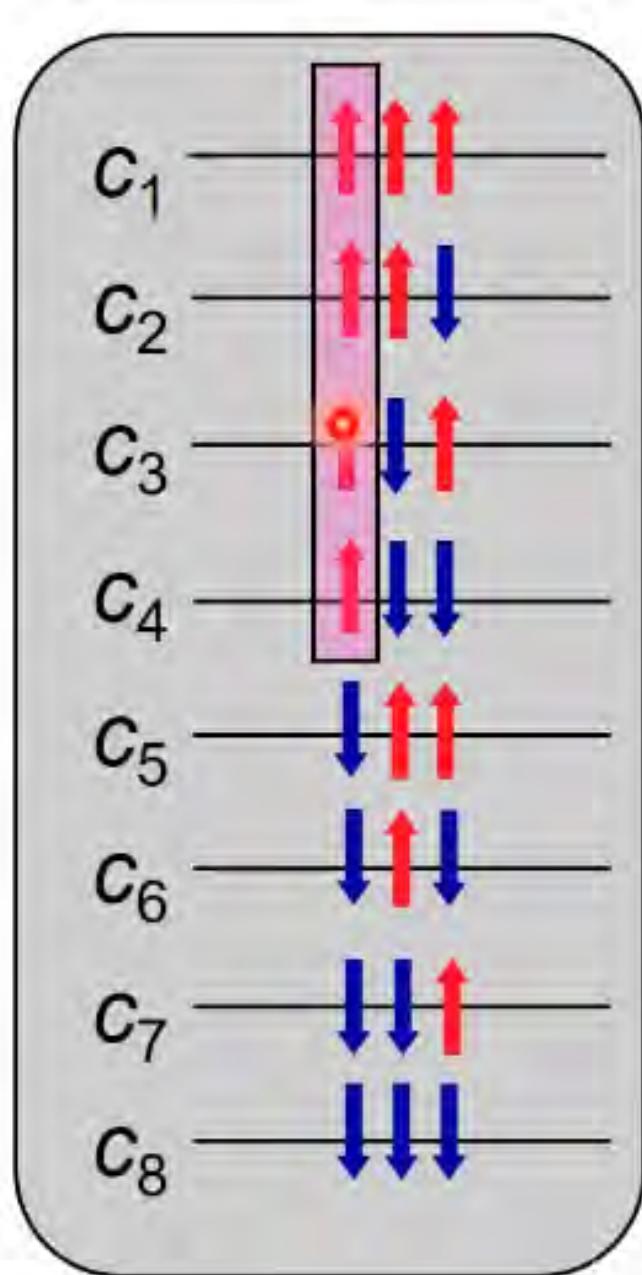


Figure 23:

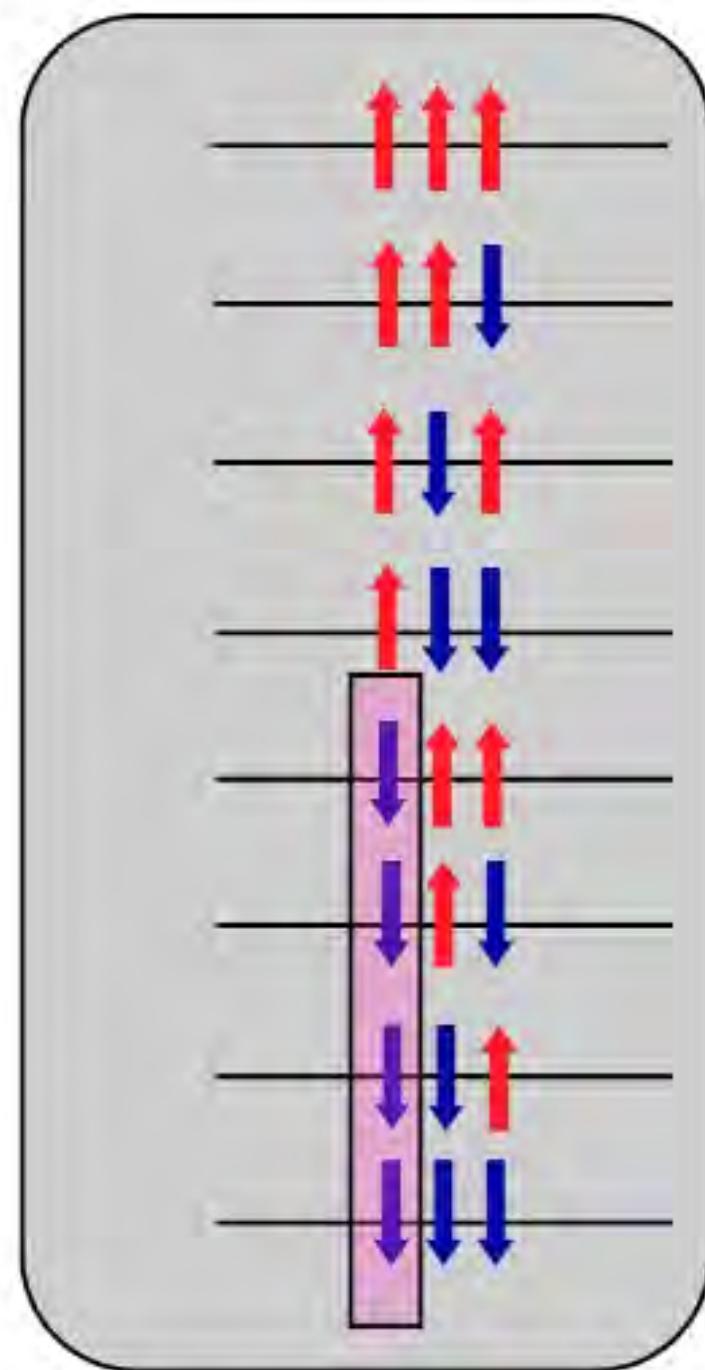
We'll start first with the spin-ups.



3-spin system

Figure 24:

## after pulse



## 3-spin system

Figure 25:

After a  $\pi$  pulse, they will all flip to spin down, and the coefficients are associated with or follow

them. So C1 through C4 now is associated with the downspin, with the other spins remaining unchanged.

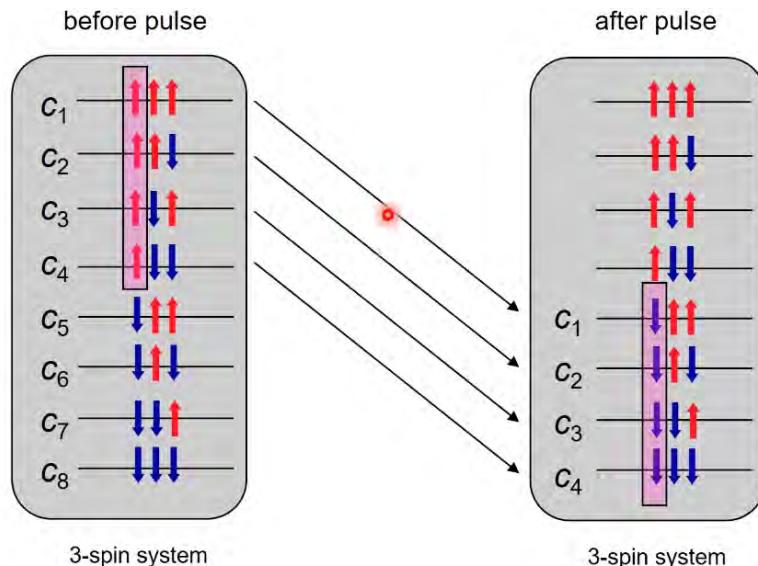


Figure 26:

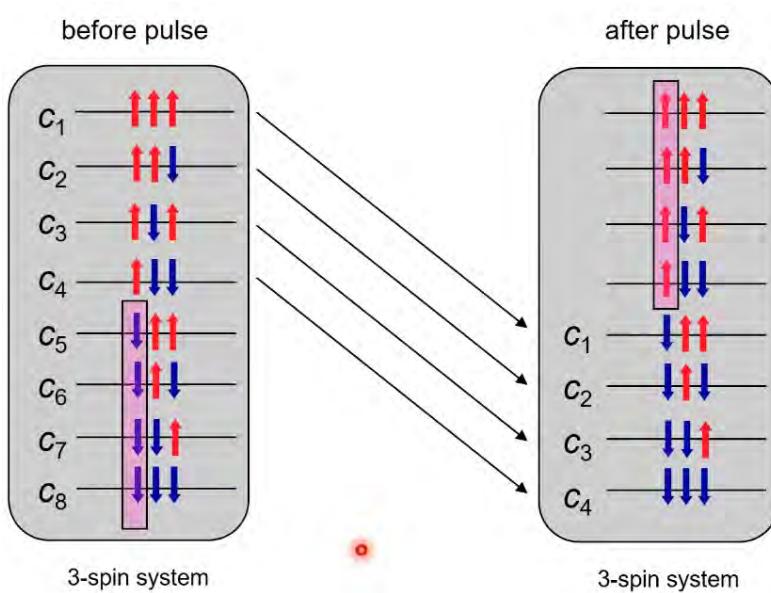


Figure 27:

Similarly, if we had to spin down,

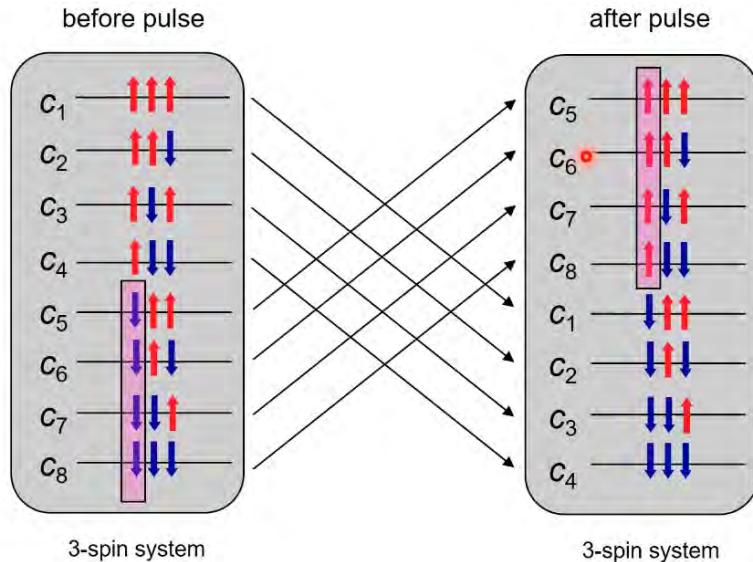


Figure 28:

those coefficients would follow as that spin gets flipped to up  $C_5$  through  $C_8$ .

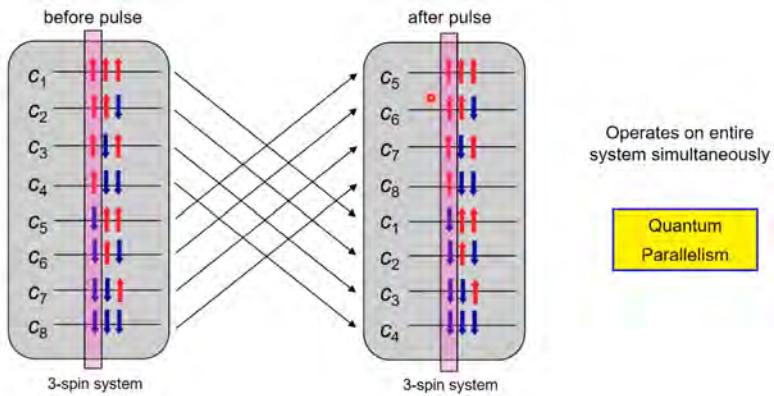


Figure 29:

The point is that we are applying one pulse on one spin, and yet we are shuttling eight coefficients or, in fact, 2 coefficients. This is an example (in a state space) of quantum parallelism.

## 0.35 Quantum Interference

Now the next important ingredient is quantum interference, and to do that, we are going to apply what's called a  $\frac{\pi}{2}$  pulse. If you remember the size of the previous pulse, this one's about half of the area, so the amplitudes smaller may have the same duration.

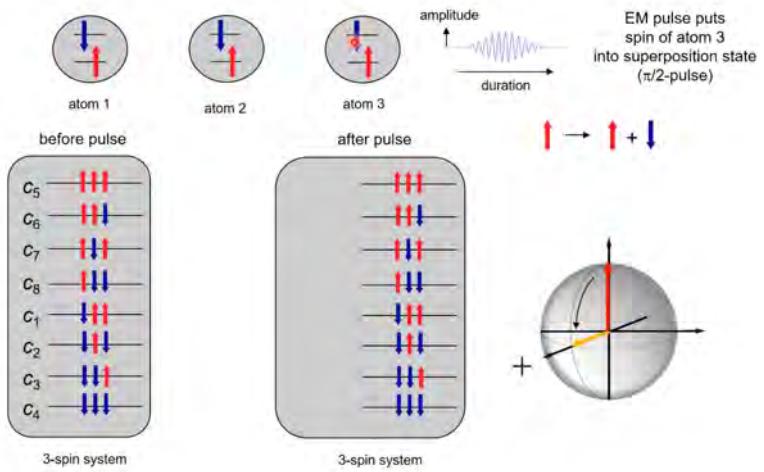


Figure 30:

we are going to look at a before and after situation; since we are applying it to the third atom, we are going to look now at the third spin, and we are going to observe that if a spin is at the North Pole spin up and we rotate it  $\frac{\pi}{2}$  or 90 deg degrees.

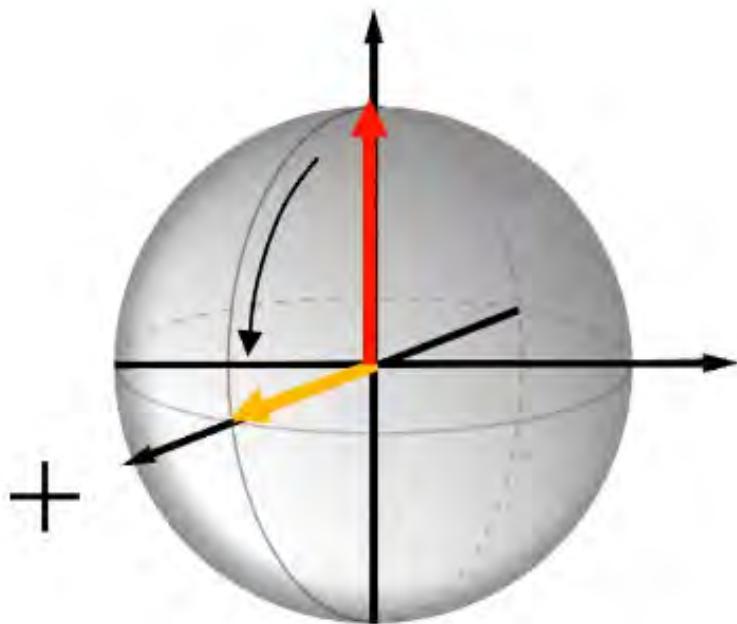


Figure 31:

We'll say that we are rotating it around this axis (it's now pointed on the equator), and we'll associate with this direction a + sign.

So, a spin-up goes to a superposition of

$$\uparrow \rightarrow \uparrow + \downarrow$$

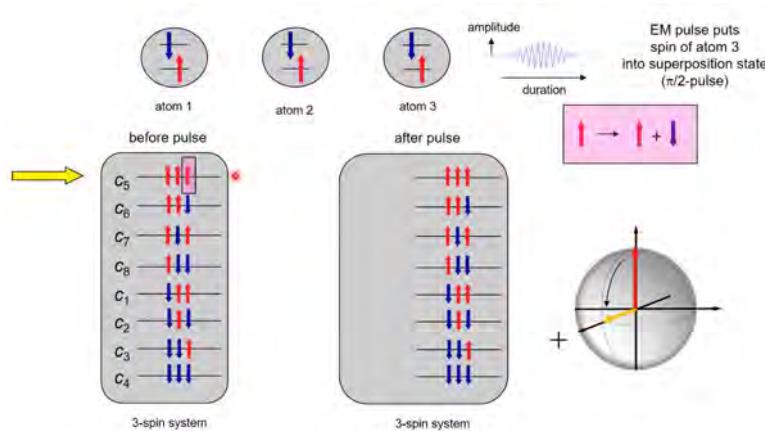


Figure 32:

Okay, so let us just look at this particular aspect corresponding with the coefficient  $c_5$ , so this spin up will go to spin up plus down (I'm not worried about normalization here),

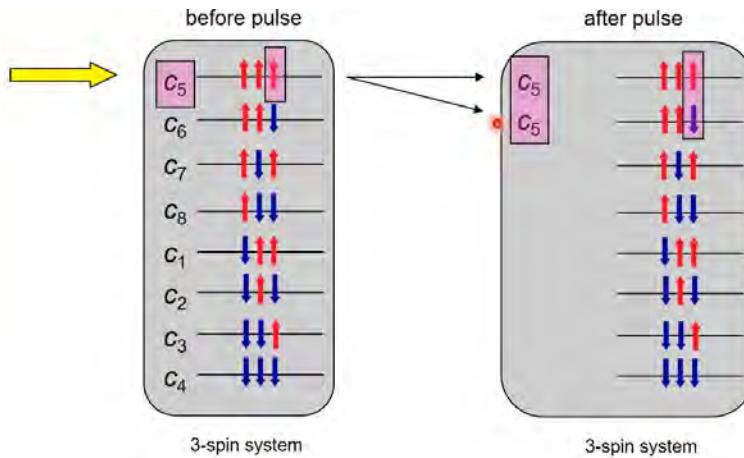


Figure 33:

so that the coefficient will now be shared between these two output states ( $C_5$  is now shared here). Again, properly normalized would have a  $\frac{1}{\sqrt{2}}$  (But let us not worry about that here now.)

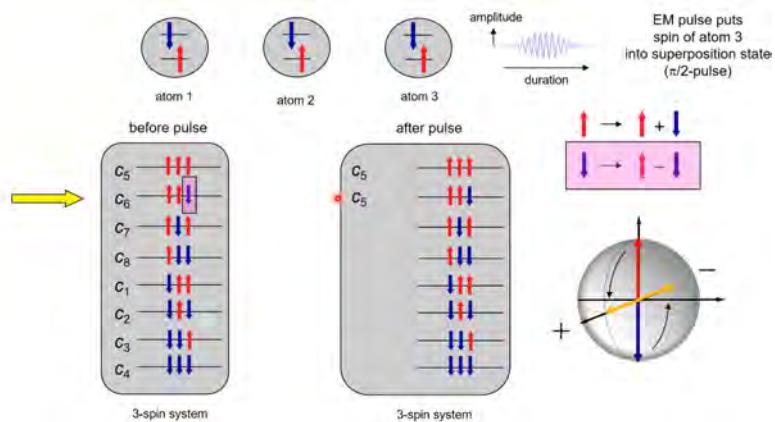


Figure 34:

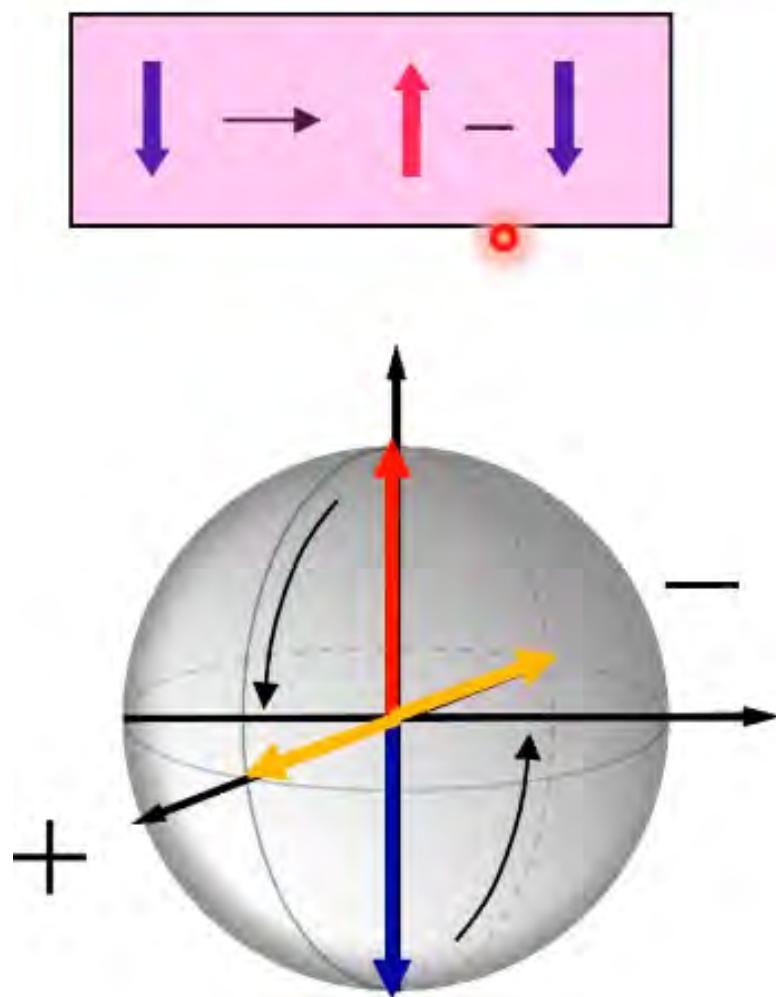


Figure 35:

Now, in the case of this next one where it's a spin down, it rotates in the same direction (equator); of course, it means it's pointing in the opposite direction, which will be associated with a  $-$  sign (spin down goes to up minus down.)

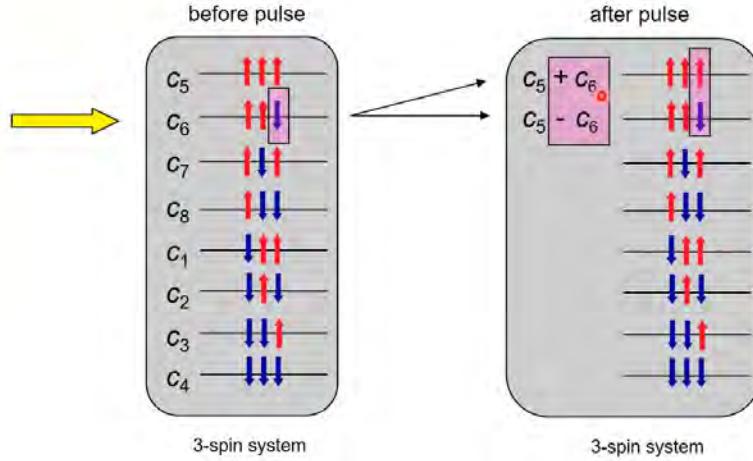


Figure 36:

This coefficient  $C_6$  now appears on both of these aspects here, but one comes in with a  $-$  sign

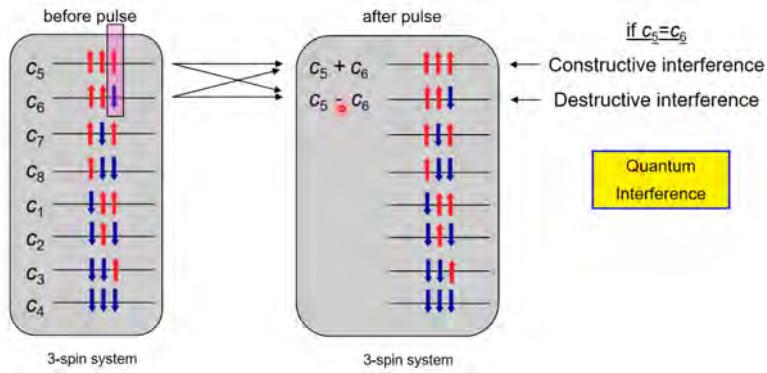


Figure 37:

So let us just say that  $C_5$  equaled  $C_6$ , and if it did, then we see that the waiting on this first  $\uparrow + \uparrow + \uparrow$  (111) state basically doubles that constructive interference, and the  $C_5 - C_6$  completely disappears, and that's complete destructive interference. This is what we mean when we say quantum interference in the context of digital quantum computation; these coefficients are adding and subtracting from one another coherently, and again, they are complex coefficients.

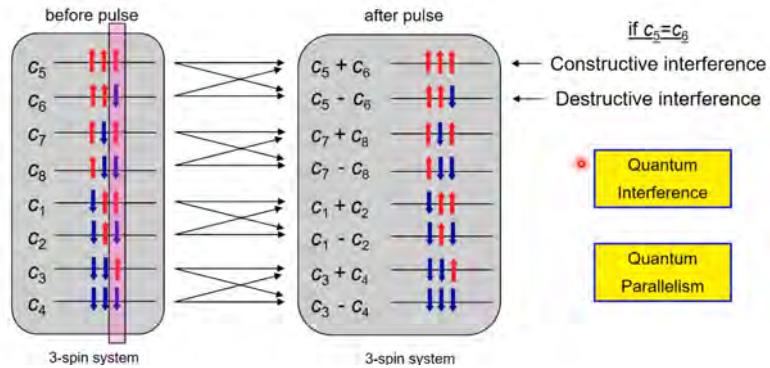


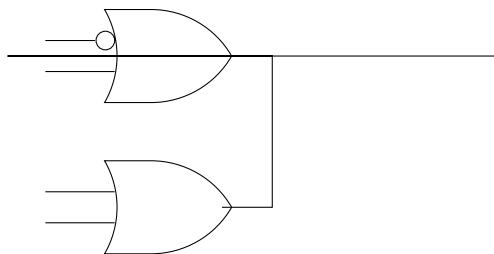
Figure 38:

In fact, this is also an example of quantum parallelism because it's happening on the entire state space at the same time. To emphasize once more, we apply single pulses, and we apply them on one spin at a time. But we are making or affecting change on the entire state space or  $2^n$  coefficients.

## 0.36 Classical Boolean Logic Gates

GATE	CIRCUIT REPRESENTATION	TRUTH TABLE										
<i>NOT</i> The output is 1 when the input is 0 and 0 when the input is 1.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0</td><td>1</td></tr> <tr> <td>1</td><td>0</td></tr> </tbody> </table>	Input	Output	0	1	1	0				
Input	Output											
0	1											
1	0											
<i>AND</i> The output is 1 only when both inputs are 1, otherwise the output is 0.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>0</td></tr> <tr> <td>0 1</td><td>0</td></tr> <tr> <td>1 0</td><td>0</td></tr> <tr> <td>1 1</td><td>1</td></tr> </tbody> </table>	Input	Output	0 0	0	0 1	0	1 0	0	1 1	1
Input	Output											
0 0	0											
0 1	0											
1 0	0											
1 1	1											
<i>OR</i> The output is 0 only when both inputs are 0, otherwise the output is 1.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>0</td></tr> <tr> <td>0 1</td><td>1</td></tr> <tr> <td>1 0</td><td>1</td></tr> <tr> <td>1 1</td><td>1</td></tr> </tbody> </table>	Input	Output	0 0	0	0 1	1	1 0	1	1 1	1
Input	Output											
0 0	0											
0 1	1											
1 0	1											
1 1	1											
<i>NAND</i> The output is 0 only when both inputs are 1, otherwise the output is 1.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>1</td></tr> <tr> <td>0 1</td><td>1</td></tr> <tr> <td>1 0</td><td>1</td></tr> <tr> <td>1 1</td><td>0</td></tr> </tbody> </table>	Input	Output	0 0	1	0 1	1	1 0	1	1 1	0
Input	Output											
0 0	1											
0 1	1											
1 0	1											
1 1	0											
<i>NOR</i> The output is 1 only when both inputs are 0, otherwise the output is 0.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>1</td></tr> <tr> <td>0 1</td><td>0</td></tr> <tr> <td>1 0</td><td>0</td></tr> <tr> <td>1 1</td><td>0</td></tr> </tbody> </table>	Input	Output	0 0	1	0 1	0	1 0	0	1 1	0
Input	Output											
0 0	1											
0 1	0											
1 0	0											
1 1	0											
<i>XOR</i> The output is 1 only when the two inputs have different values, otherwise the output is 0.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>0</td></tr> <tr> <td>0 1</td><td>1</td></tr> <tr> <td>1 0</td><td>1</td></tr> <tr> <td>1 1</td><td>0</td></tr> </tbody> </table>	Input	Output	0 0	0	0 1	1	1 0	1	1 1	0
Input	Output											
0 0	0											
0 1	1											
1 0	1											
1 1	0											
<i>XNOR</i> The output is 1 only when the two inputs have the same value, otherwise the output is 0.		<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>0 0</td><td>1</td></tr> <tr> <td>0 1</td><td>0</td></tr> <tr> <td>1 0</td><td>0</td></tr> <tr> <td>1 1</td><td>1</td></tr> </tbody> </table>	Input	Output	0 0	1	0 1	0	1 0	0	1 1	1
Input	Output											
0 0	1											
0 1	0											
1 0	0											
1 1	1											

Figure 39:



So, of course, we have to develop some language or algorithm to do this, and we can derive a lot of intuition from classical Boolean logic gates, which we list here on the left. I'm not going to talk through all of them, but they are gates like the NOT gate, the AND gate, and the OR gate, and they have input and output truth tables as shown here. And I think we are all pretty familiar with these from electrical engineering.

- **Universal gate sets for Boolean logic**
  - E.g., NOT, AND
  - E.g., NOR
  - And many more (not unique)
  - Requires at least one two-bit gate

Figure 40:

The important part is that we can form universal gate sets that implement arbitrary Boolean logic. We don't need all these gates; we need a handful of them, and many such subsets exist. It could be the NOT gate and the AND gate together. There are many, it could be the NAND gate or the NOR gate. It's not unique. It does require at least one two-bit gate, and with a universal gate set, one can apply or implement any arbitrary Boolean logic.

## 0.37 Single-Qubit Quantum Gates

GATE	CIRCUIT REPRESENTATION	MATRIX REPRESENTATION	TRUTH TABLE	BLOCH SPHERE
$I$ Identity-gate: no rotation is performed.		$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	Input      Output $ 0\rangle$ $ 0\rangle$ $ 1\rangle$ $ 1\rangle$	
$X$ gate: rotates the qubit state by $\pi$ radians ( $180^\circ$ ) about the x-axis.		$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	Input      Output $ 0\rangle$ $ 1\rangle$ $ 1\rangle$ $ 0\rangle$	
$Y$ gate: rotates the qubit state by $\pi$ radians ( $180^\circ$ ) about the y-axis.		$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$	Input      Output $ 0\rangle$ $i 1\rangle$ $ 1\rangle$ $-i 0\rangle$	
$Z$ gate: rotates the qubit state by $\pi$ radians ( $180^\circ$ ) about the z-axis.		$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	Input      Output $ 0\rangle$ $ 0\rangle$ $ 1\rangle$ $- 1\rangle$	
$S$ gate: rotates the qubit state by $\frac{\pi}{2}$ radians ( $90^\circ$ ) about the z-axis.		$S = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{2}} \end{pmatrix}$	Input      Output $ 0\rangle$ $ 0\rangle$ $ 1\rangle$ $e^{i\frac{\pi}{2}} 1\rangle$	
$T$ gate: rotates the qubit state by $\frac{\pi}{4}$ radians ( $45^\circ$ ) about the z-axis.		$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}$	Input      Output $ 0\rangle$ $ 0\rangle$ $ 1\rangle$ $e^{i\frac{\pi}{4}} 1\rangle$	
$H$ gate: rotates the qubit state by $\pi$ radians ( $180^\circ$ ) about an axis diagonal in the x-z plane. This is equivalent to an X-gate followed by a $\frac{\pi}{2}$ rotation about the y-axis.		$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	Input      Output $ 0\rangle$ $\frac{ 0\rangle +  1\rangle}{\sqrt{2}}$ $ 1\rangle$ $\frac{ 0\rangle -  1\rangle}{\sqrt{2}}$	

Figure 41:

So, quantum computing is quite similar. We have single and two qubit gates. I'm showing two-qubit gates (above), and again, I won't talk through them in detail.

$I$ Identity-gate: no rotation is performed.		$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	Input      Output $ 0\rangle$ $ 0\rangle$ $ 1\rangle$ $ 1\rangle$	
--	--	--	---	--

Figure 42:

Above is the identity gate, which means not doing anything. One is the identity gate, which means not doing anything. Then, there are a number of gates, such as an X gate or a Y gate.

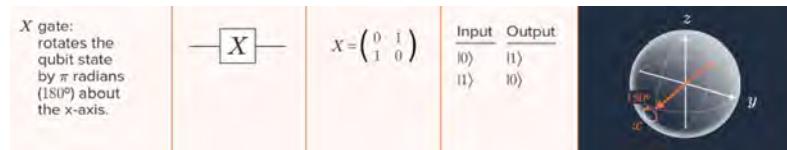


Figure 43:

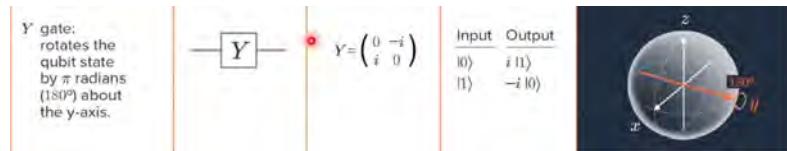


Figure 44:

We'll see an example in a moment, which indicates what axis we are rotating around on the block sphere for some rotation distance. In this case, it's a rotation of  $\pi$ . There are many such single qubit gates.

## 0.38 Two-qubit Quantum Gates

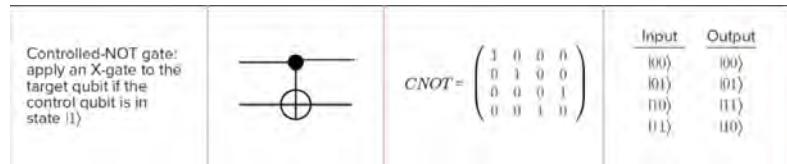


Figure 45:

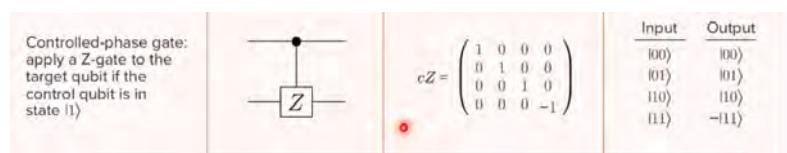


Figure 46:

Two-qubit gates are also necessary here, such as the Controlled-NOT or the Controlled-Z and their truth tables. We'll see an example of a controlled-phase gate in just a moment.

The point is that, just like with classical Boolean logic, we form universal gate sets with quantum logic. With just a handful of these single and two qubit gates, it's not a unique set. There are many, but with any one of these, we can implement arbitrary quantum logic and the other important difference is that the quantum gates and the quantum logic itself are reversible. There are examples of reversible classical logic, but with quantum computing, it's a requirement, and it does require one two-qubit entangling gate. These are basically the ground rules.

## 0.39 Gate-Based Approach: Single-Qubit Operation

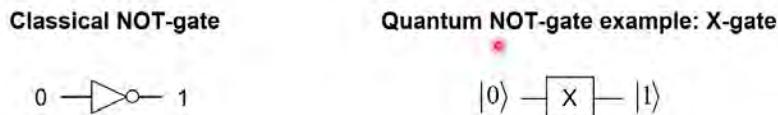


Figure 47:

So let us take a look at a single-qubit operation, some intuition and we'll base it on an analogy with the classical NOT gate. It's a quantum NOT gate, which we call the *X* gate, and it takes quantum state zero to quantum state one or vice versa. You can see it here with this simulation. So we start on the block sphere at the North Pole, and we rotate around the *x*-axis (that's why we call it an *x*-gate) The blue vector is the block vector, and the red vector that comes in and out is basically tracing out the envelope of this resonant pulse that we are applying to drive that spin.

**X-gate applied to qubit along +Z:**  $|0\rangle \rightarrow |1\rangle$

Figure 48:

As it's implemented here, we go from the North Pole to the South Pole, which are classical states.



Figure 49:

So here, it looks fairly classical, but you need not start at the North or South Pole.

**X-gate applied to arbitrary qubit state:**  $\alpha|0\rangle + \beta|1\rangle \rightarrow \beta|0\rangle + \alpha|1\rangle$

Figure 50:

You could start anywhere on the surface of the planet Earth and rotate around the *x*-axis, taking you from one superposition state to another. In fact, we are swapping the coefficients of  $\alpha$  and  $\beta$  on this superposition state. So, this is an example of the *X*-Gate, a single qubit gate.

## 0.40 Gate-Based Approach: Two-Qubit Controlled-NOT

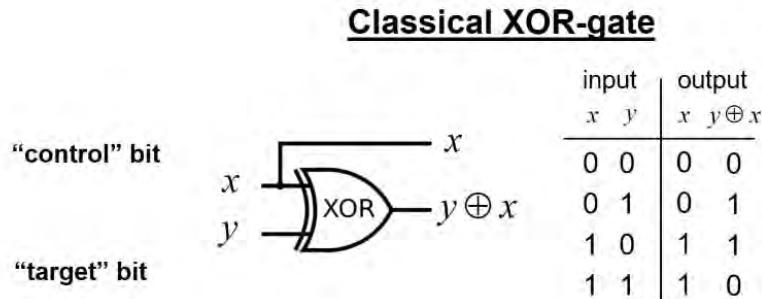


Figure 51:

Now, an example of the two-qubic Controlled-NOT has a classical out analog which is the Exclusive OR gate shown here. one way to describe XOR is to consider one of the input bits the control bit and the other the target bit. When the control bit  $x$  is 0, the target bit-Y is passed to the output as the XOR output. But when the control bit is 1, we take the Y-bit's value and flip it. 0 becomes 1, 1 becomes 0, and that is the output of the XOR.

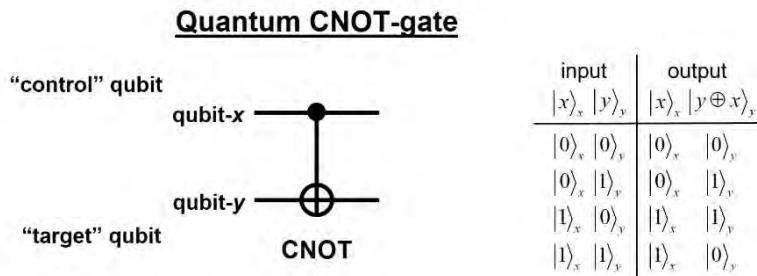


Figure 52:

The Quantum CNOT-gate is the same thing. We have a control qubit-x and a target qubit-y. This is how we diagrammatically represent CNOT, and you see the same thing when the control qubit is in state 0, we leave the target qubit-y alone and pass that to the output. When the control qubit is 1, we will flip the target qubit and pass it to the output. This means that the rotation of qubit-y depends on the state of qubit-x, and it can get very interesting if we start thinking about superposition states.

Rotation of qubit-y depends on the state of qubit-x

Figure 53:

**For example:**  $|\psi_{\text{in}}\rangle \propto (|0\rangle + |1\rangle)_x |0\rangle_y$

Figure 54:

For example, if the input is in a superposition state ( $|0\rangle + |1\rangle$ ) for the control qubit x and 0 for the target qubit y, we can basically apply linear algebra here to see what happens.

**For example:**  $|\psi_{\text{in}}\rangle \propto (|0\rangle + |1\rangle)_x |0\rangle_y$

$$|\psi_{\text{out}}\rangle \propto \underline{|0\rangle_x} |0\rangle_y + \underline{|1\rangle_x} |1\rangle_y$$

Figure 55:

input	output
$ x\rangle_x  y\rangle_y$	$ x\rangle_x  y \oplus x\rangle_y$
$ 0\rangle_x  0\rangle_y$	$ 0\rangle_x  0\rangle_y$
$ 0\rangle_x  1\rangle_y$	$ 0\rangle_x  1\rangle_y$
$ 1\rangle_x  0\rangle_y$	$ 1\rangle_x  1\rangle_y$
$ 1\rangle_x  1\rangle_y$	$ 1\rangle_x  0\rangle_y$

Figure 56:

we can basically apply linear algebra here to see what happens. So, when the control is zero, we leave the target the same and that appears here.

**For example:**

$$|\psi_{\text{in}}\rangle \propto (|0\rangle + |1\rangle)_x |0\rangle_y$$

$$|\psi_{\text{out}}\rangle \propto \underline{|0\rangle_x} \underline{|0\rangle_y} + \underline{|1\rangle_x} \underline{|1\rangle_y}$$

Figure 57:

So, when the control is zero, we leave the target the same and that appears here.

**For example:**

$$|\psi_{\text{in}}\rangle \propto (|0\rangle + |1\rangle)_x |0\rangle_y$$

$$|\psi_{\text{out}}\rangle \propto \underline{|0\rangle_x} \underline{|0\rangle_y} + \underline{|1\rangle_x} \underline{|1\rangle_y} \neq (\dots)_x (\dots)_y$$

Results in an entangled state  
(cannot be factored)

Figure 58:

What we see now is that this is no longer a factorizable state into something that looks solely like an x-qubit crossed with something that looks like a y-qubit right. This is in fact an example of an entangled state. Now it needs to be entangled in this way in all bases, which this one happens to be, but the point is that it cannot be factored out. So the Quantum CNOT-Gate is called an entangling gate because it has this property,

**Universal gate-based quantum computation is achievable with a small set of single and two-qubit gates.**

Figure 59:

and so a universal gate-based quantum computer. Quantum computation is achievable with just a handful of these types of single and two Qubit gates.

## 0.41 Quantum Algorithm (Universal QC)

So, with these, we can build up a quantum algorithm, and I'm going just pictorially to represent what such an algorithm might look like.

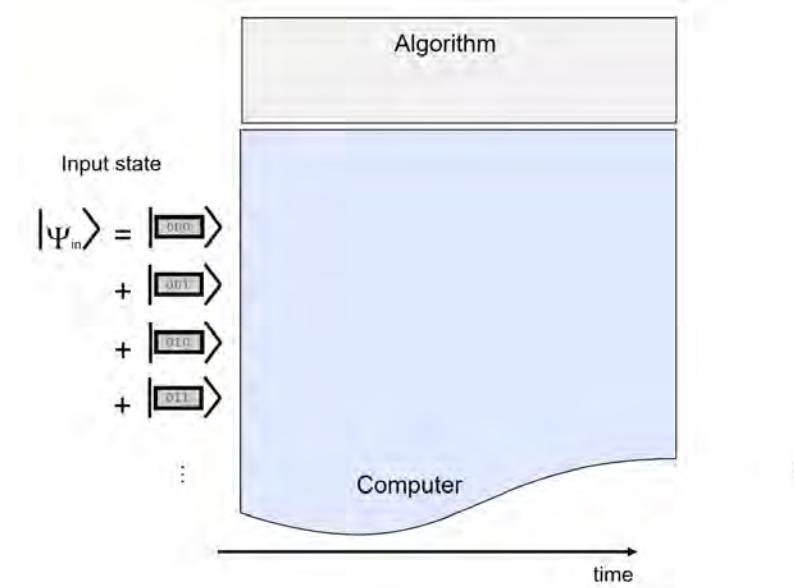


Figure 60:

Typically, we'll start with a massive superposition state with all of the aspects equally weighted, so it's an equal superposition state. Again, I'm not worried about normalization here, but the coefficients in front all have the same magnitude.

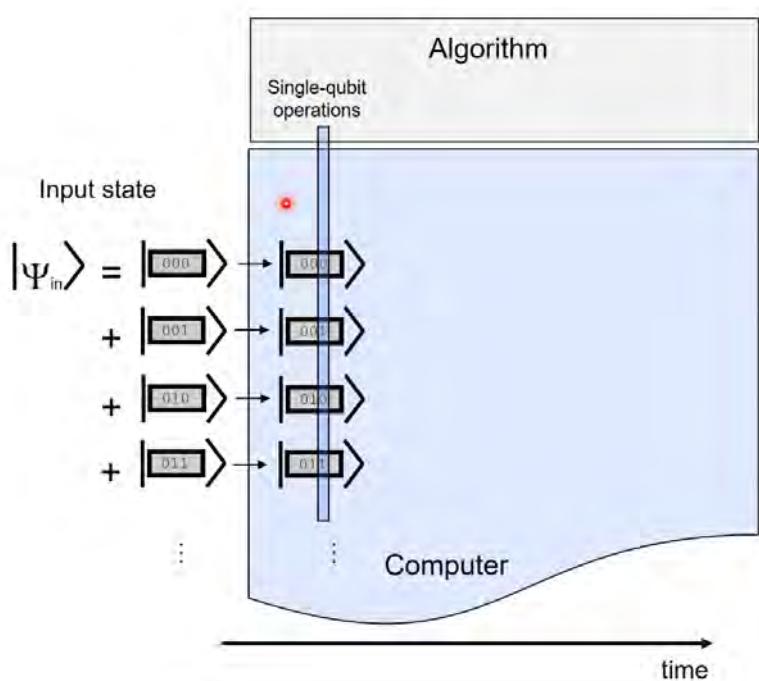


Figure 61:

Then the algorithm proceeds according to some prescription which will include single qubit operations, where again, to emphasize one operation on one qubit, but that impacts all of the aspects in that superposition in the state space followed by quantum interference.

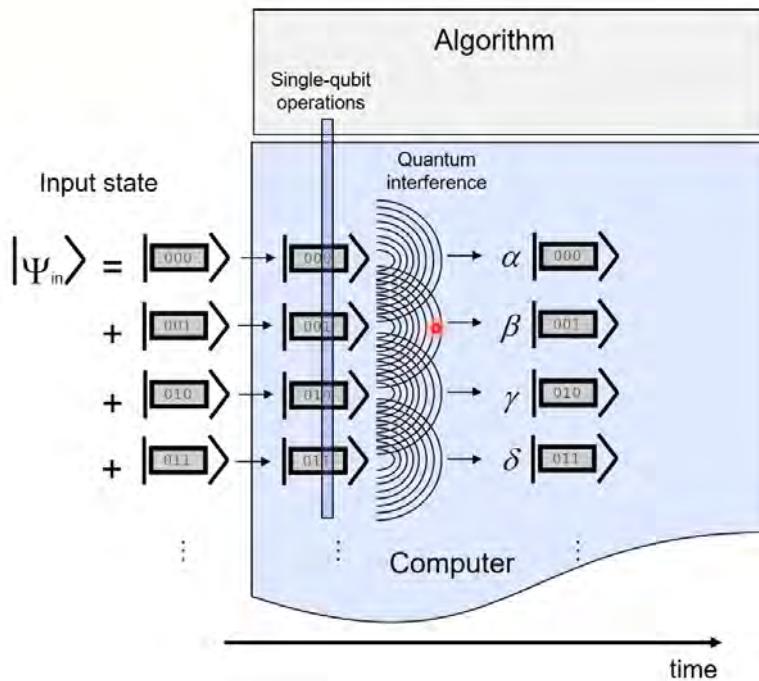


Figure 62:

And the purpose here is that it changes the values of these coefficients. We start to see the values increase or decrease,

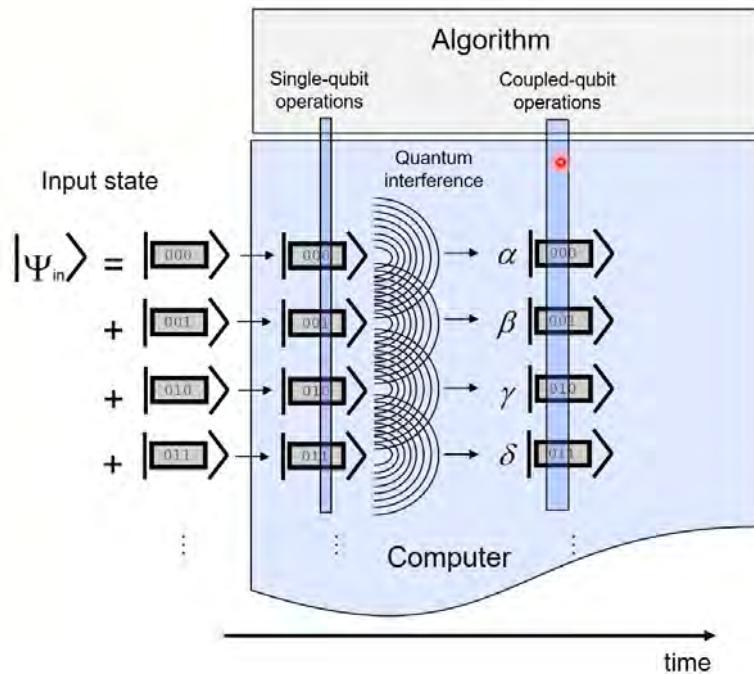


Figure 63:

and of course there will also be coupled qubit gates where we apply one operation in this case on two qubits

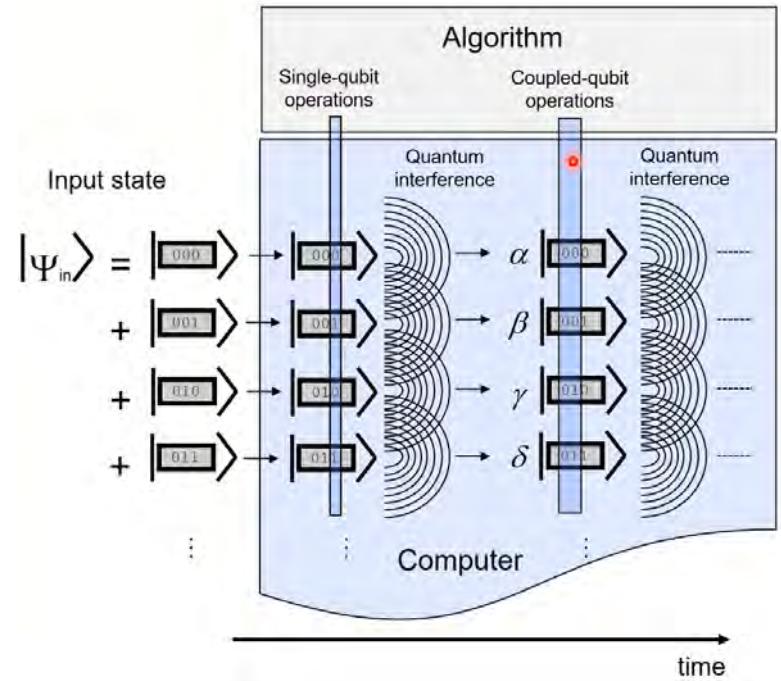


Figure 64:

followed by quantum interference,

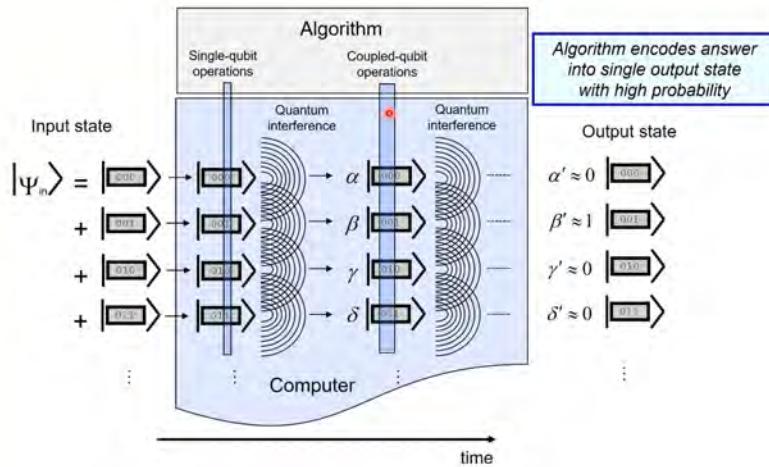


Figure 65:

and at the end of the day the intent or the goal should be that all of the probability amplitude or [q] this weight exists in front of only one of the aspects out of the very large number of components in that superposition state.

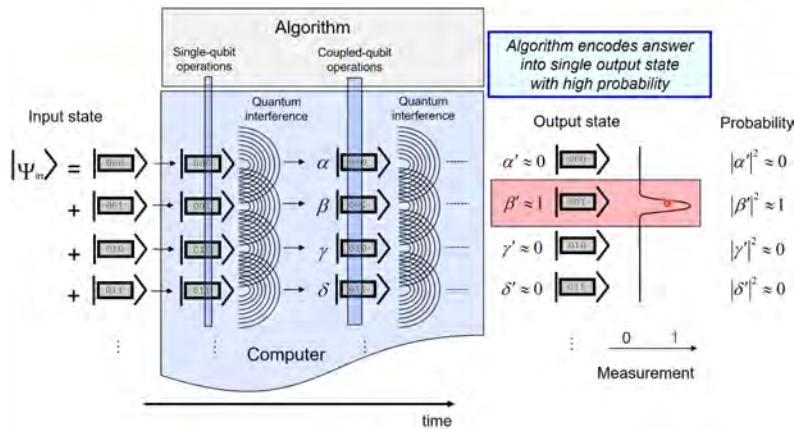


Figure 66:

And that's because we know that when we make a measurement that the classical result we will measure will result from a probability, which goes as the magnitude of that probability amplitude squared.

So, in this case, if beta were equal to beta prime or equal to one, then with unit probability, we would measure the output result ze one [q] on our three qubits, which is critical. So that's the challenge for the algorithm designer is to create the proper sequence of single and two qubit gates to achieve this coalescence of probability amplitude around one and only one or, in some cases, a few, but a select few, of these aspects in a very, very large superposition state. So that when we measure with high probability, we get an answer encoded by the result.

## 0.42 Coherence Gate Time

So then you could ask, that's great, it all works. Why don't we have quantum computers today? It comes down to two times: one is the coherence time, and the other is the gate time.



Figure 67:

The coherence time is the qubit's lifetime. So even if we know exactly a quantum state at time t equals zero.

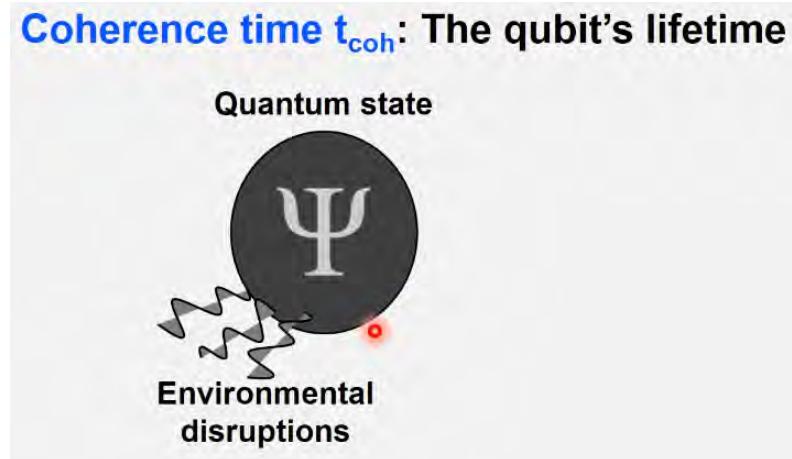


Figure 68:

From that point forward, it starts to interact with its environment in ways that we don't control

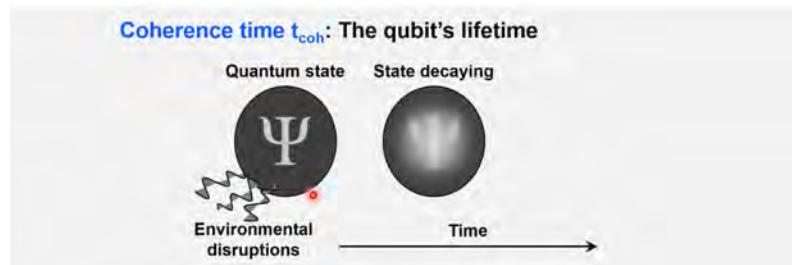


Figure 69:

As we are running our algorithm, you could think of it as the state side blurring and eventually disappearing. Now, the qubits are still there, but their time evolution has taken them to a different place than what we expected with our algorithm because, in addition to our algorithm, there are these environmental disruptions, and so the times scale over which this happens is the coherence time.

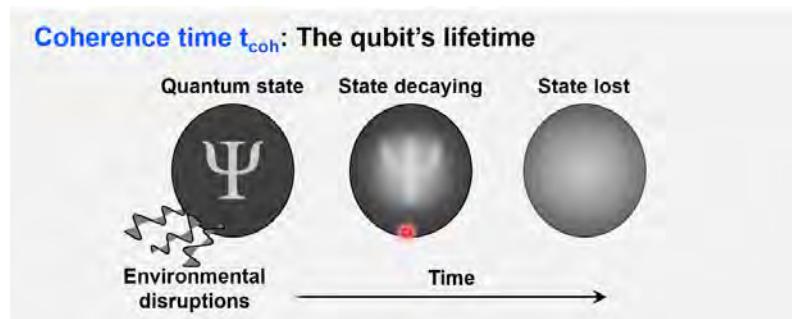


Figure 70:

Then the other time is the gate time, and that's the typical time required for a single gate operation. It could be your slowest operation, which would be a conservative approach

**Gate time  $t_{\text{gate}}$ :** Time required for a single gate operation

Figure 71:

and the figure of merit then is, how many gates can I apply within the coherence time that I have? And this is a figure of merit. For intuition, a rigorous metric would be called fidelity. We'll talk briefly about that.

**Figure of Merit \* :** # of gates per coherence time =  $t_{\text{coh}}/t_{\text{gate}}$

(\* Rigorous metric: gate & readout fidelity)

Figure 72:

One point to make here is that long coherence times in and of themselves are insufficient. It's really the number of gates before an error, and to make an absurd kind of ridiculous example of this, you could say that neutrinos last forever, and so maybe they would make really good qubits. Their coherence time is basically infinite, but they wouldn't, and the reason they wouldn't, even though they have a very long coherence time is because their gate time would be equally long because it's very hard for us even to detect them, let alone control them right.

**Long coherence times are not sufficient, it's the number of gates before an error**

Figure 73:

So of course, that's ridiculous. But it serves to emphasize the point that you need both of these times to know if this is a good qubit.

## 0.43 Qubit Modalities

## (Extensible platforms, benchmarked, ca 2019)

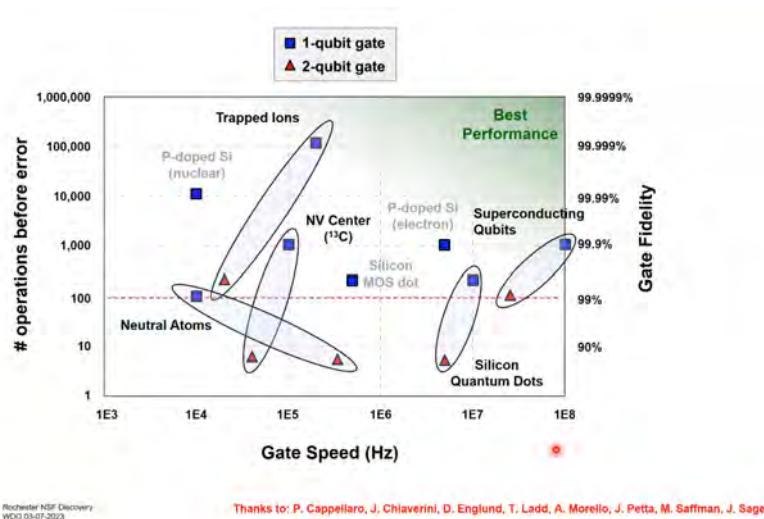


Figure 74:

With that, we can compare different cubit modalities against one another. On the left vertical axis is the number of operations before AND error. This metric we just talked about and on the right axis is a corresponding gate fidelity based on that, so one error in 100 would be 99 percent fidelity one error in 1099.9 percent E CEA [100 operations] [q], and on the horizontal axis is gate speed. How quickly can I apply these gates? Now the best performance is in the upper right corner, and you can see a number of different technologies here such as trapped Dixons superconducting qubits ET, etc.

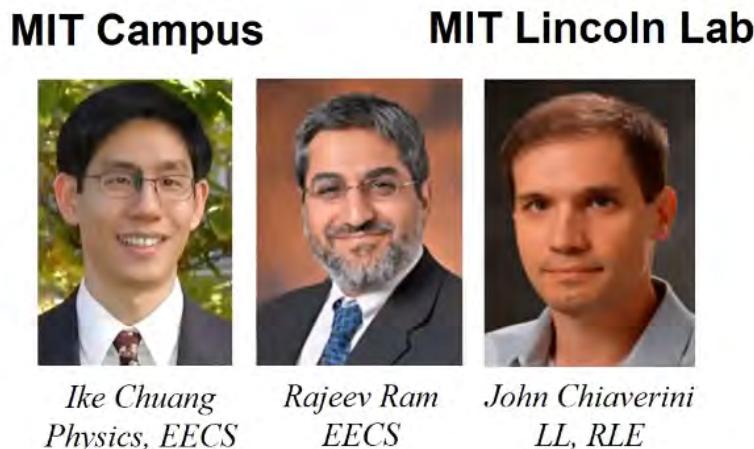


Figure 75:

And here are a few colleagues of mine at MIT Campus and Lincoln Lab working on that are one of the leaders today.

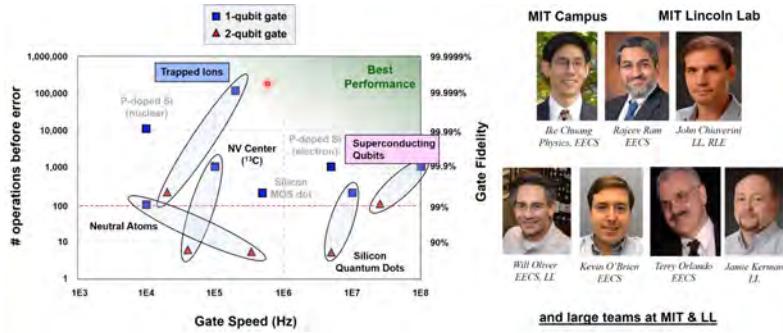


Figure 76:

My own research is in superconducting qubits. Myself and some of my colleagues and you can see that they're both basically equidistant from the upper right corner in some sense, [q] which means we haven't picked really what a winner is, and it's way too early to pick that. Trapped ions are focused on making their gates faster because they're relatively slow about a 1000 times slower than superconducting. Superconducting are pretty fast already and we just want higher fidelity.

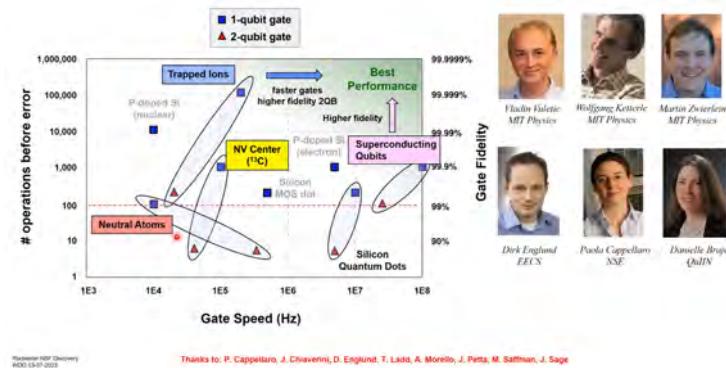


Figure 77:

And of course there are a lot of other efforts that are quickly catching up. For example, neutral atoms have made a lot of progress in the last couple years.

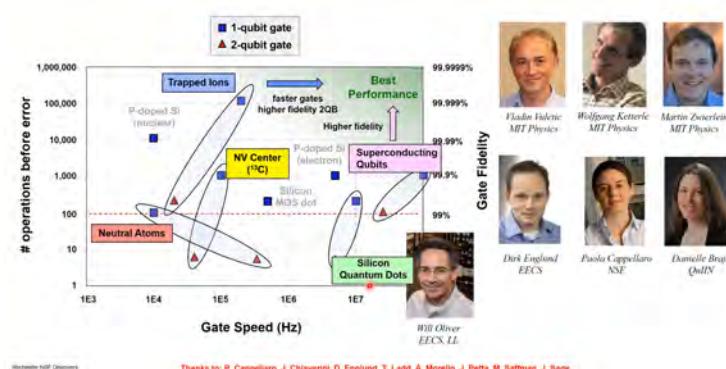


Figure 78:

## *Introduction to Quantum Computing: Qubits, Gates, and Algorithms*

I'm working on silicon quantum dots now in my own lab and you. [q]

### **Many candidate technologies under development to realize the promise of quantum computation**

Figure 79:

The takeaway here is that know there are a lot of candidate technologies. It's very early in this marathon to say who is going to be the winner and very likely different technologies will excel at different periods of time as they mature and wane. Just like we had vacuum tubes and then we had BATs and then ECL logic and then C moths E CE. OKA. [q]

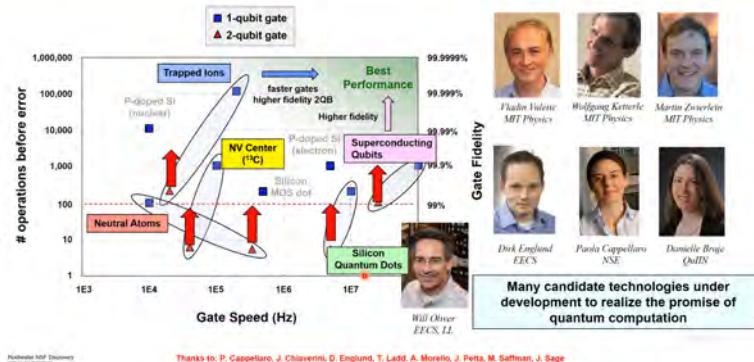


Figure 80:

And also this chart is already a bit dated and so everything has improved by about a factor of 10 since I initially made it, and only getting better.

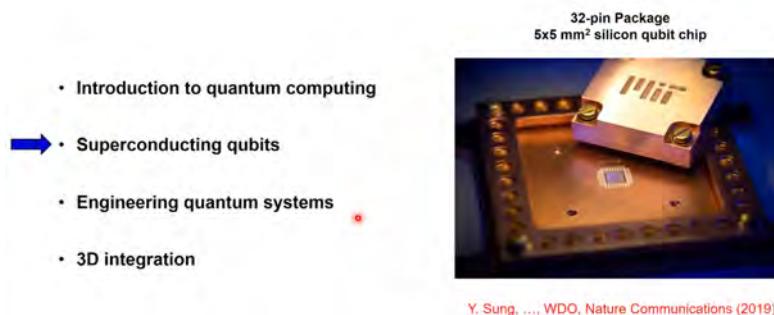


Figure 81:

So with that let me now talk about superconducting qubits.

## 0.44 Natural and Artificial Atoms

### Protons, Neutrons, and Electrons

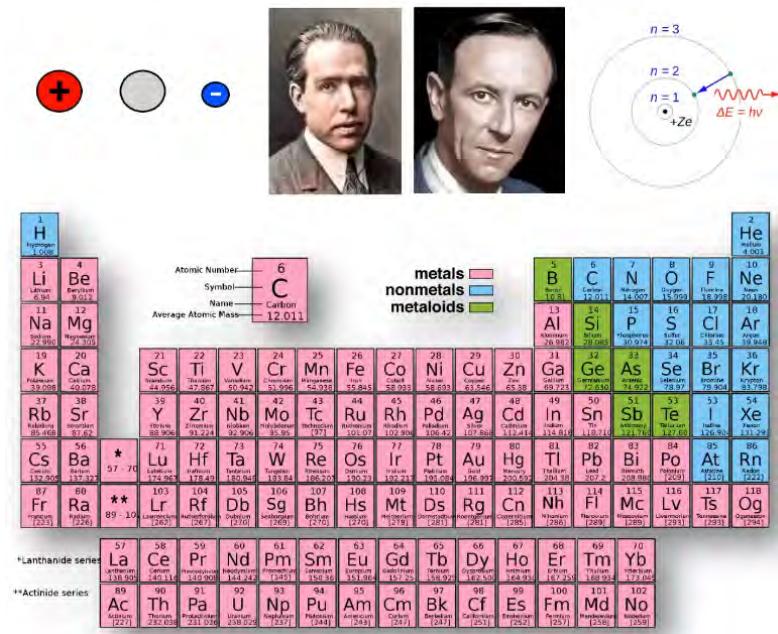


Figure 82:

You know, we sometimes call them [q] artificial atoms by analogy, of course to the natural atoms in our real world. And one way to look at this is natural atoms. They all have different properties based on how many protons, neutrons or electrons that they have.

### Inductors, Capacitors, and Josephson Junctions

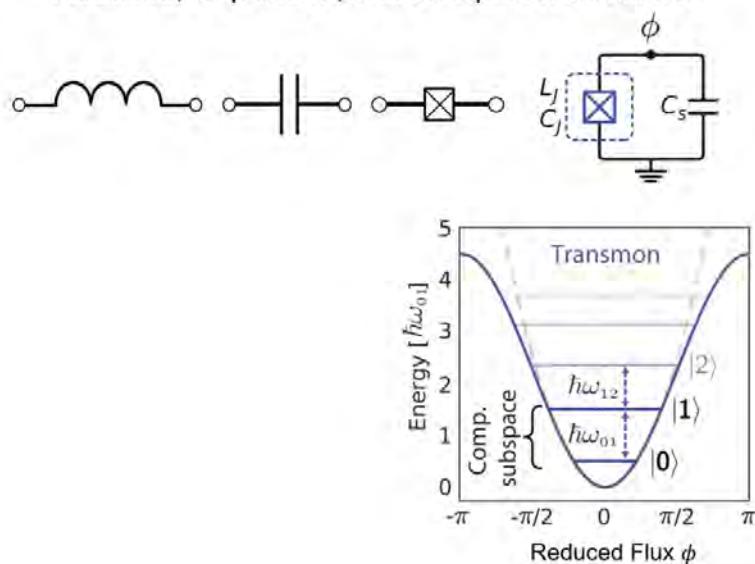


Figure 83:

I don't want to overstress this analogy, but in some sense we build artificial atoms and the ingredients we use are



Figure 84:

inductors,

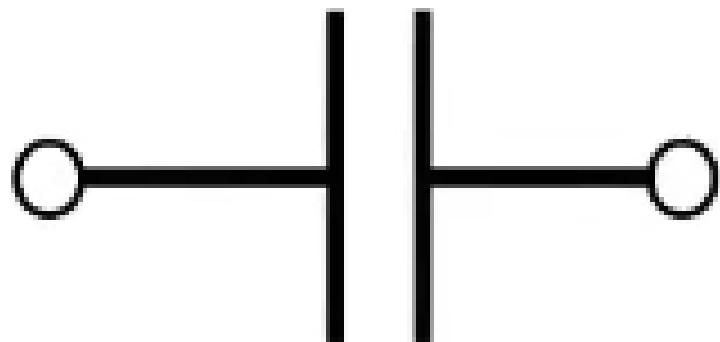


Figure 85:

capacitors

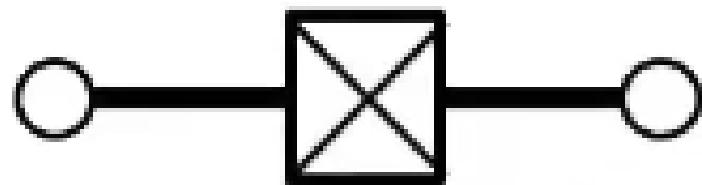


Figure 86:

and joseph's injunctions.

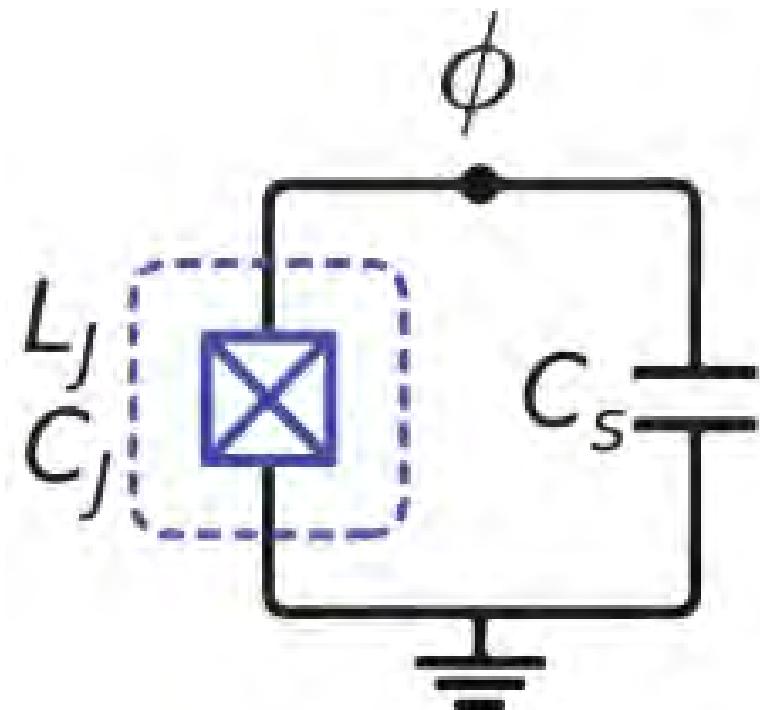


Figure 87:

the joseph's injunction is in our case a nonlinear inductor.

And what happens is you build something like this a resonator, but it's an anharmonic resonator due to the nonlinear Josephs and inductance. So an LC oscillator and [q]

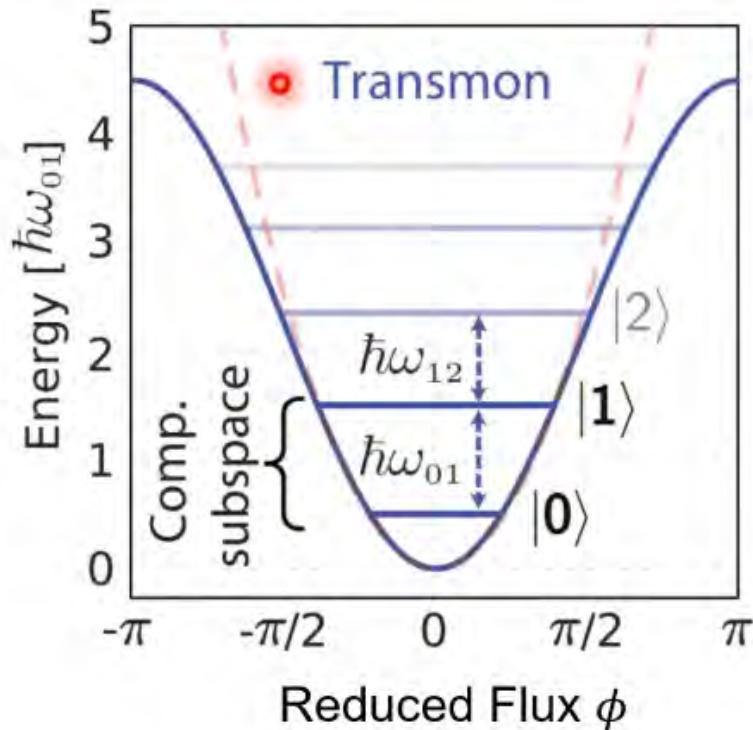
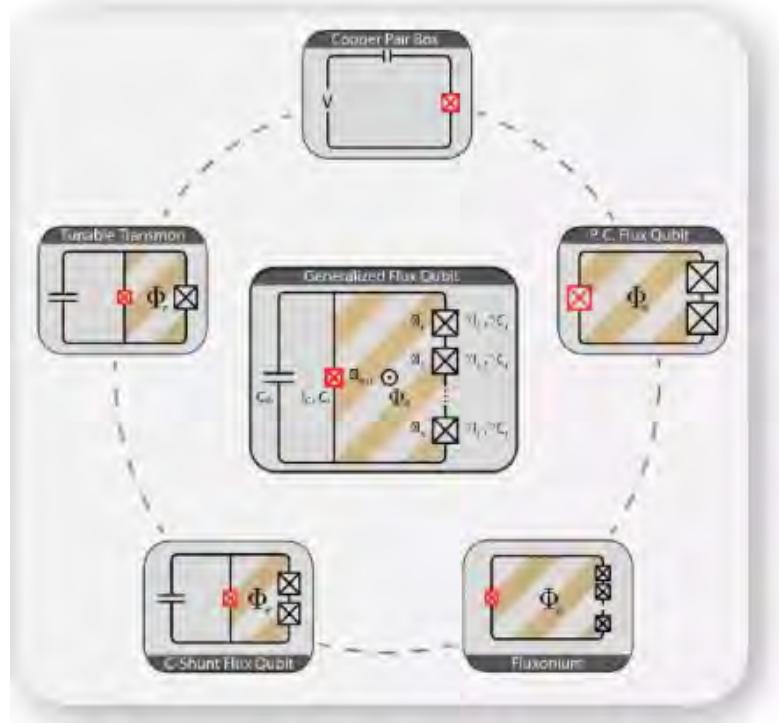


Figure 88:

as a result its potential is not parabolic. But it's a cosine potential for this simple case. And if we cool this down to a low enough temperature, then you know the zero and the one state are discrete. This resonator can have two excitations, but its frequency Omega 1 two is different than Omega one so we can basically look at the zero and one of this resonator and call that qubit. But of course, there are many other higher excited states, just like Adams has many states as well.

As a result, its potential is not parabolic. But it's a cosine potential for this simple case. [q] If we cool this down to a low enough temperature, then you know the 0 and the 1 states are discrete. This resonator can have two excitations, but its frequency Omega 1 two is different than Omega one, so we can basically look at the zero and one of these resonators and call that qubit. But of course, there are many other higher excited states, just like atoms have many states as well.



F. Yan et. al., arXiv:2006.04130 (2020)

Figure 89:

By choosing different numbers and arrangements and parameter values of these junctions, capacitors, linear inductors and Josephson junctions we can make basically a large number of artificial atoms, each with different properties. But the difference is that we can design them because we can choose these parameter values rather than having them given to us by nature.

**Superconducting qubits are artificial atoms with properties we can design**

Figure 90:

So superconducting qubits are artificial atoms with properties that we can design.

## 0.45 Superconducting Artificial Atoms

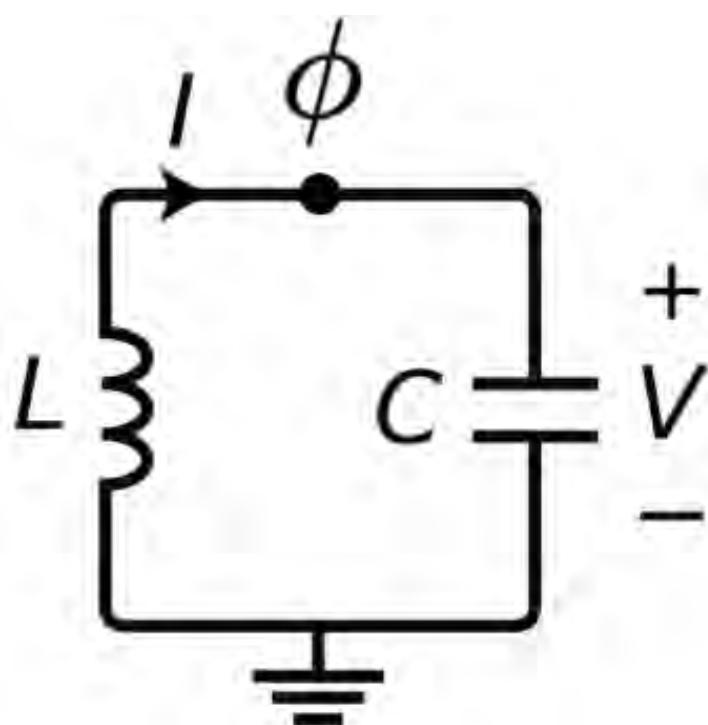


Figure 91:

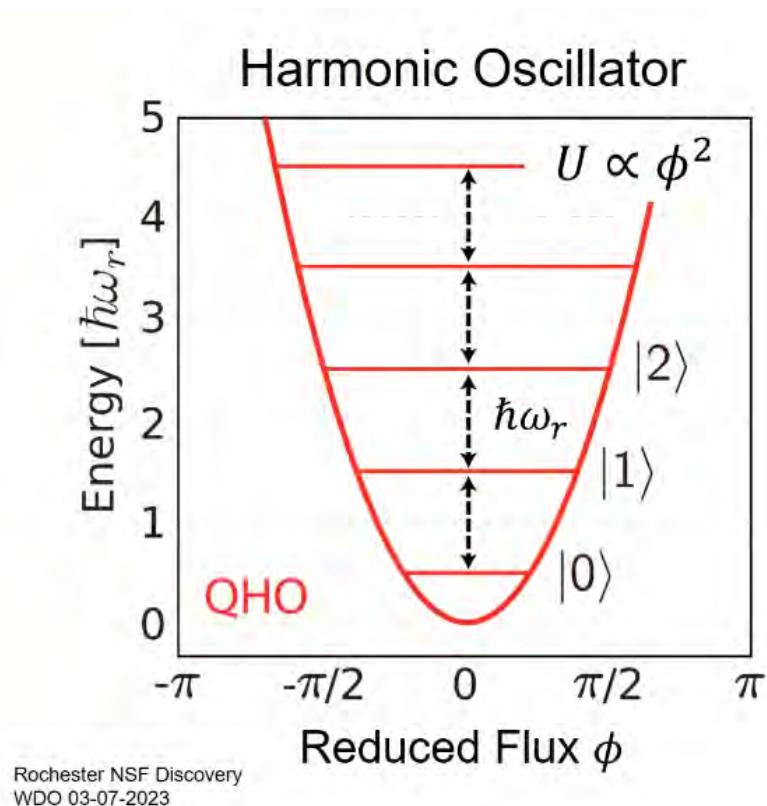


Figure 92:

And they look a lot like this. Let me just emphasize what we looked at, so the LC oscillator, of course, is a harmonic oscillator. It has a parabolic potential, and if you cool this down to low temperatures, it is in fact quantum mechanical. But the challenge is that the energy level spacing between 0 and 1, 1 and 2, and 2 and 3 is equal.

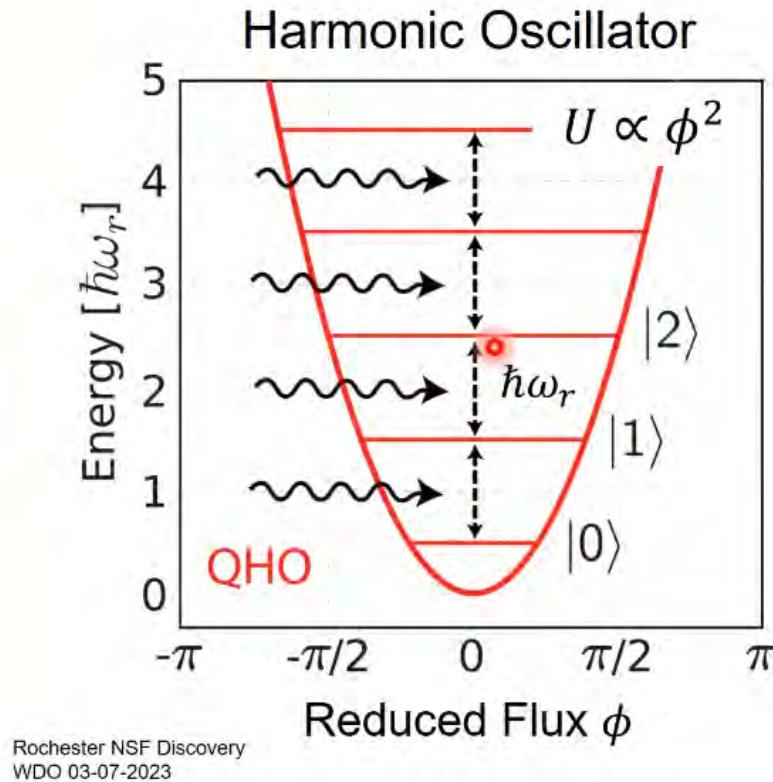


Figure 93:

For each of these, because it's parabolic or harmonic, if I try to drive a one transition, I'll end up driving the whole ladder of states in my harmonic oscillator. So this is not a good qubit, it is quantum, but it's not a good qubit.

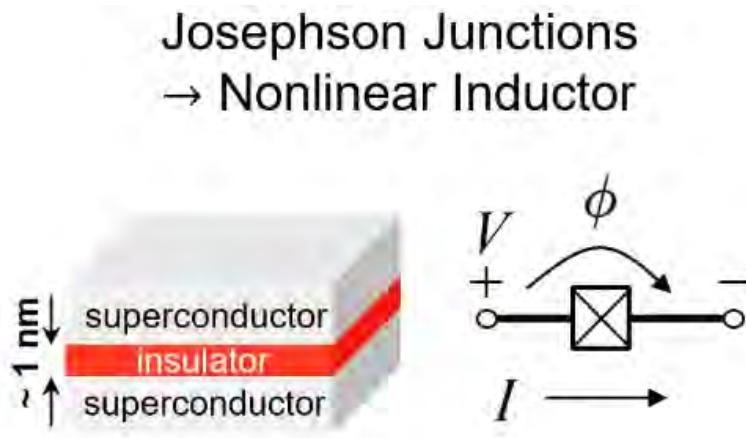


Figure 94:

So to make it anharmonic we need to add a nonlinear element, and for that we use the Josephson junction, which is a non-linear inductor, made up of superconducting leads separated by a thin insulator.

$$I = I_c \sin(\phi) \quad V = \frac{\Phi_0}{2\pi} \frac{d\phi}{dt}$$

Figure 95:

The Josephson Relations from Brian Josephson and tell us what the current and voltage through this junction should be,

$$V = L_J \frac{dI}{dt} \rightarrow L_J = \frac{\Phi_0}{2\pi I_c \cos(\phi)}$$

Figure 96:

for our purposes here, however, we just combine them in the usual way to define a Josephson's inductance [ $q$ ], and its property is that it's non-linear. If you look at this, you can see that it's nonlinear in the current that is going through it. So the inductance changes with current.

$$E = \int dt IV = -\frac{I_c \Phi_0}{2\pi} \cos \phi = -E_J \cos \phi$$

### Anharmonic Potential

Figure 97:

We can calculate an energy and we also see that the energy has this cosine potential and so indeed,

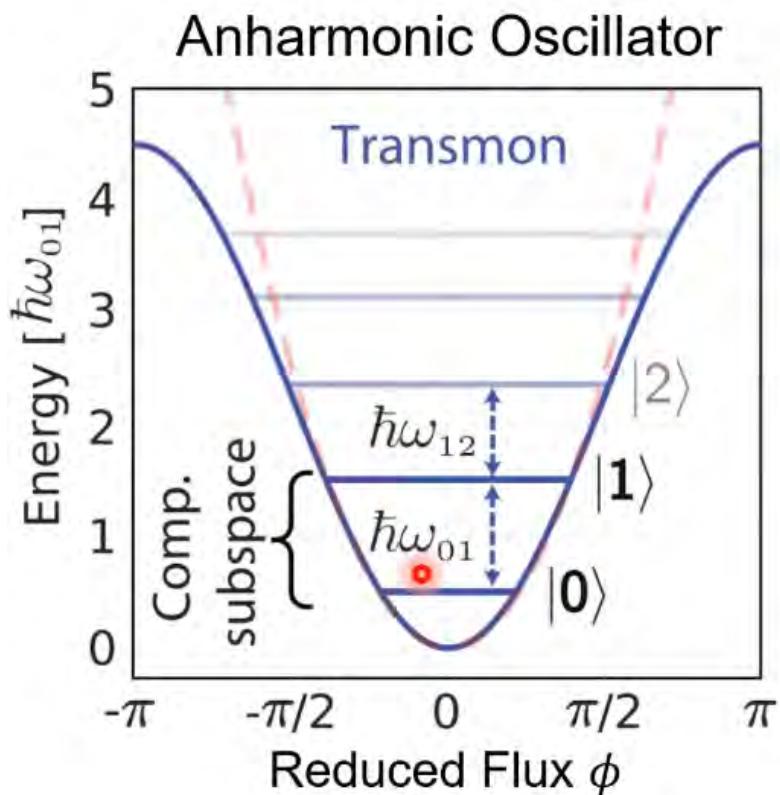


Figure 98:

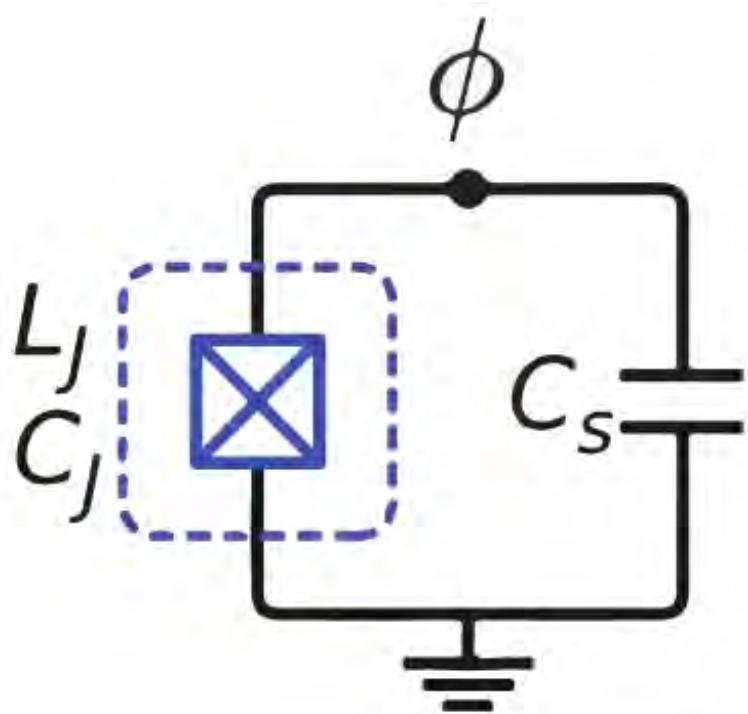


Figure 99:

that's why we drew this cosine potential right here and by just adding a Josephson junction here to make this an anharmonic oscillator, we now have the 0 1 frequency and the 1 2 frequency and the two frequency have different values and that's really what we needed. So, the Josephson junction is the critical element that makes us all possible.

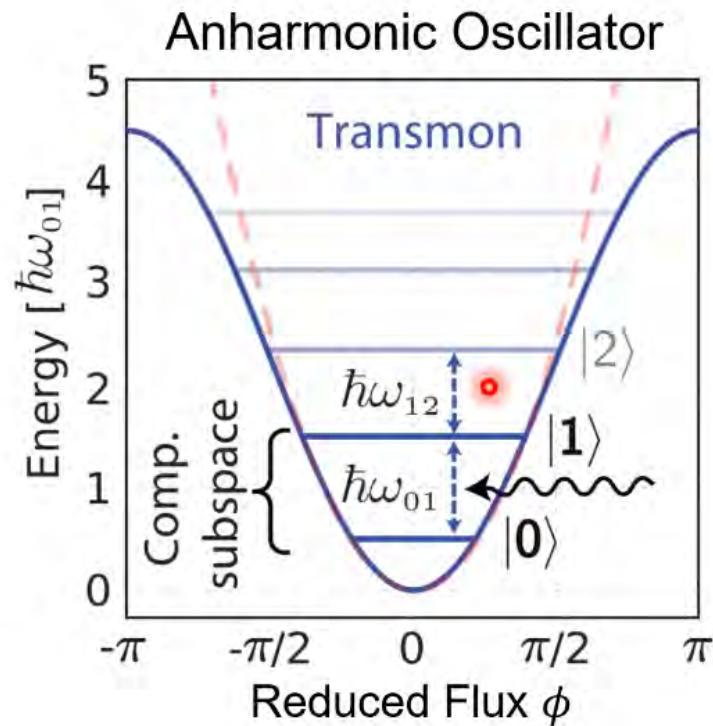


Figure 100:

## 0.46 Cryogenic Engineering

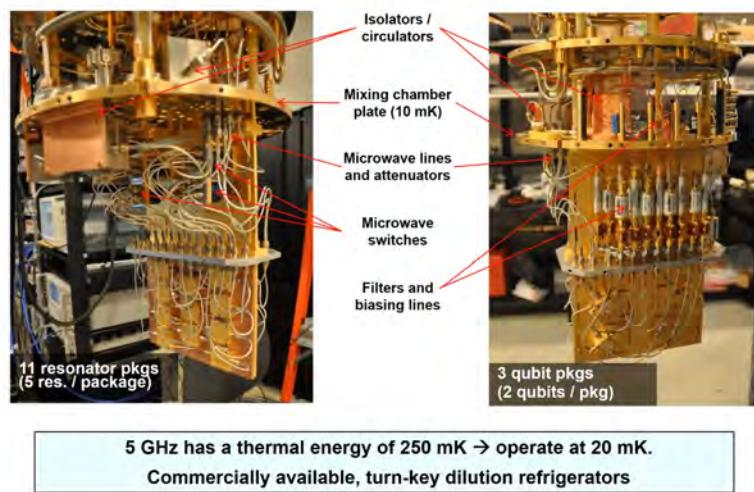


Figure 101:

Now we need to cool them down and of course how cold depends on what resonance frequency we choose. We typically choose the 0 1 frequency to be around 5 GHz and that's a trade-off, but there are a lot of electronics available to us today that will operate at 5 GHz. And it's also a temperature that's achievable in a dilution refrigerator, meaning that 5 GHz has a thermal energy. You equate  $H\bar{R}\omega$  to  $kT$  [q]. The T would be about 200 and 50 millikelvin and you want to be much colder than that, so that there's no thermal population of state one, so ten times colder is typically good enough. This is an Arrhenius function, so we use dilution refrigerators which can operate at 20 millikelvin.

## **0.47 Fabrication Engineering**

1. Manufactured designed qubits
2. Lithographic scalability (Silicon)

High-coherence air-gap cross-overs (Optical microscope and confocal images)

200-mm wafers (49 Reticles X 16 chips)

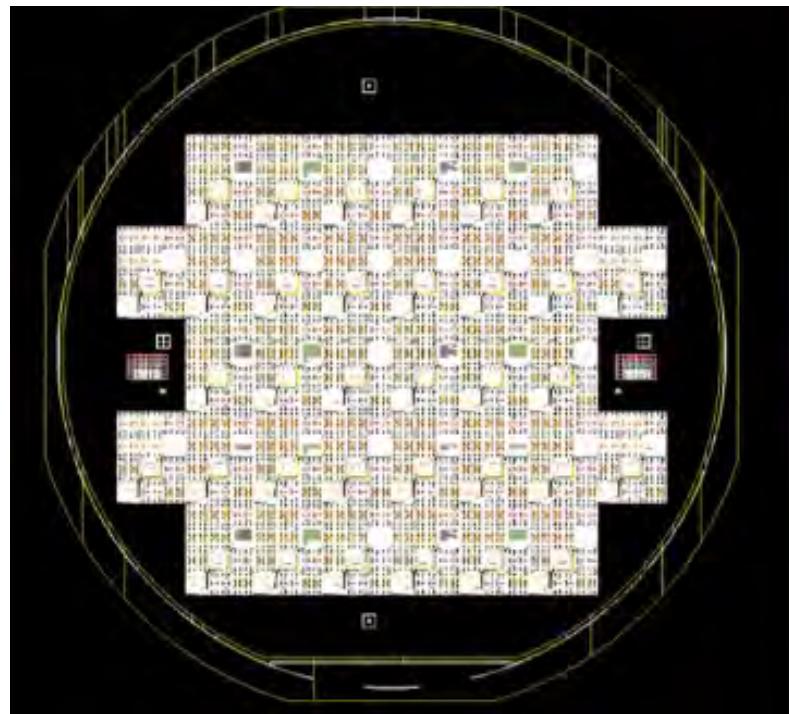


Figure 102:

5-Tranmon chip with readout resonators

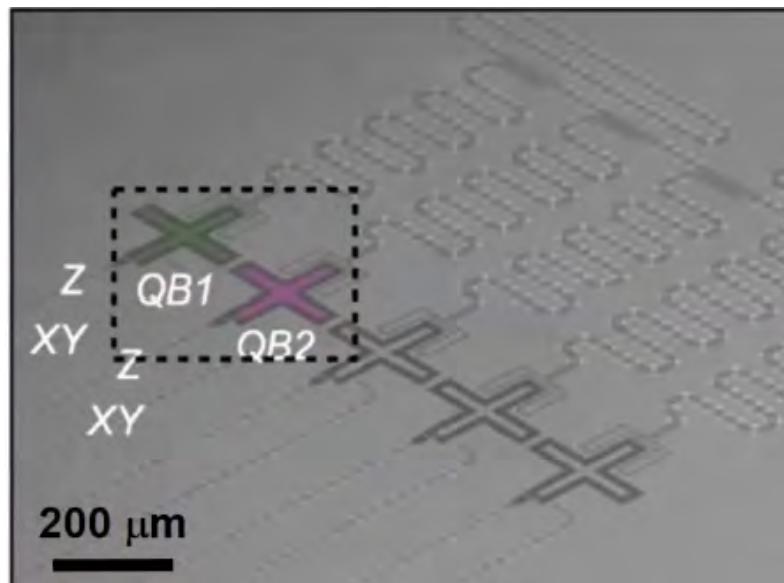


Figure 103:

Transmon capacitor and control lines

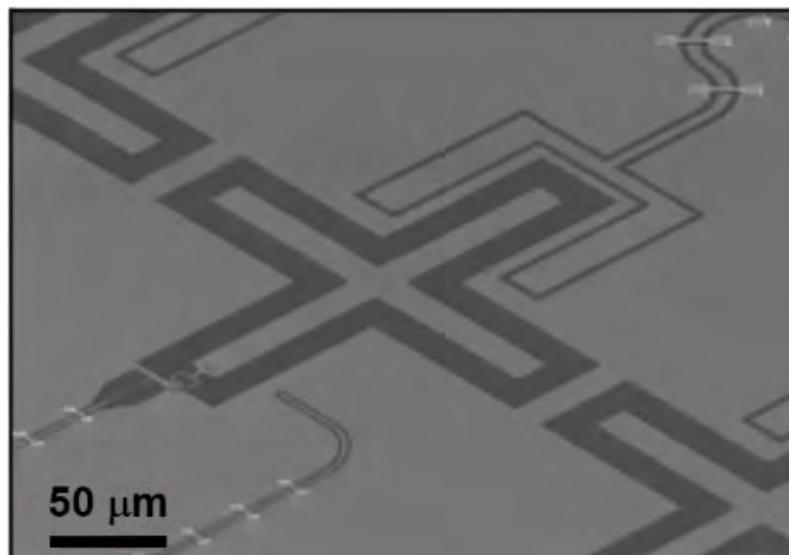


Figure 104:

Tunable transmon qubit loop with junctions

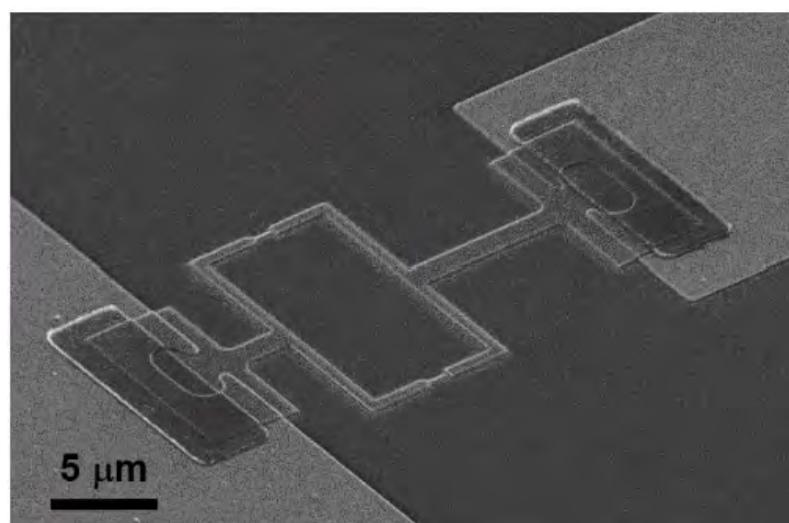


Figure 105:

Josephson Junctions (Aluminum )

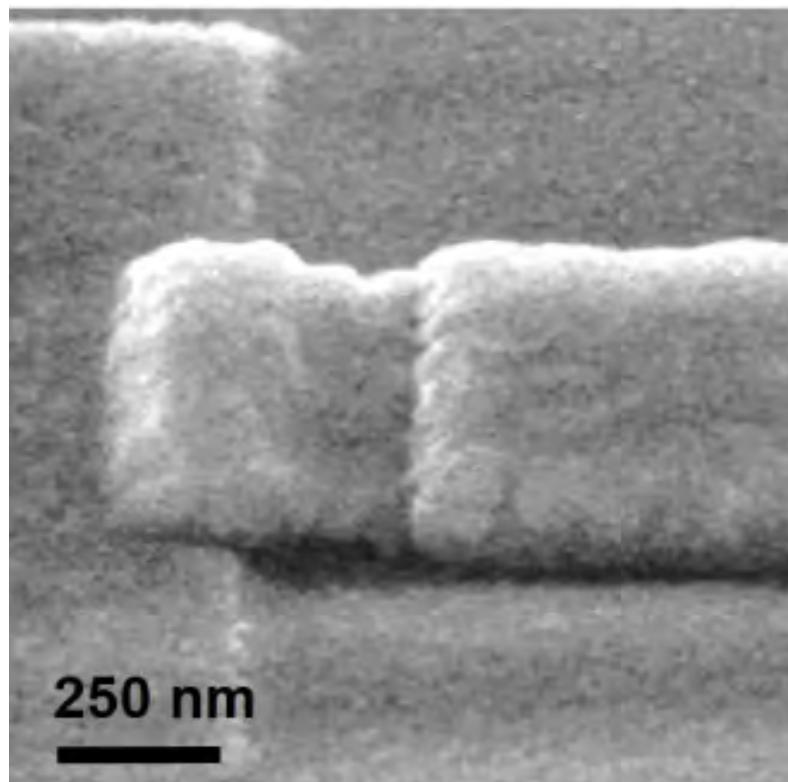


Figure 106:

We fabricate these devices using primarily silicon tool sets, so many of the evaporators and sputtering tools and optical lithography etch tools that we have from the many years of silicon development. We use them here, and in fact, we fabricate these on silicon wafers. Here's a 200-millimeter wafer, and with many recitals and chips, we zoom in on one chip. You can see five plus signs that are actually capacitors zoom in on one of them, and you can see this capacitance. Two ground. The plus to the ground is the capacitance. To the northeast is a readout resonator and to the southwest or control or bias lines. We'll look at 3D integration later of how we can take these out of the plane [q]. But if we zoom in further, this spot right here is where the SQUID [q] is. It's two junctions in a loop. By taking a Josephson junction and putting it in a loop, we can now make the frequency tunable. By putting a magnetic field through this loop, we are basically changing the inductance with the magnetic field and then finally zooming in at the nanoscale. This is a Josephson junction that's about one to 200 nanometers in size, and you're just looking at the top electrode on the bottom electrode.

## 0.48 Microwave Engineering and control

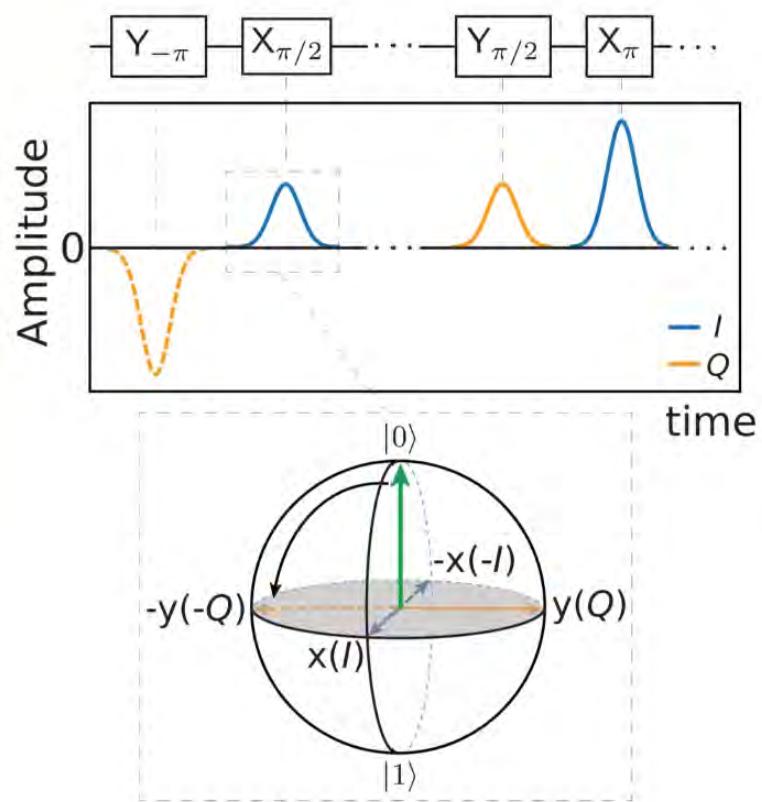


Figure 107:

**Gate Sequence** So the way we control our qubits is through microwave I-Q mixing. We have a gate sequence that we want to apply to the block vector on the block sphere to rotate it around different axes, say the X or the Y, and we can do that by associating X within phase and Y with the quadrature ports of an I-Q mixer.

## I-Q Mixing

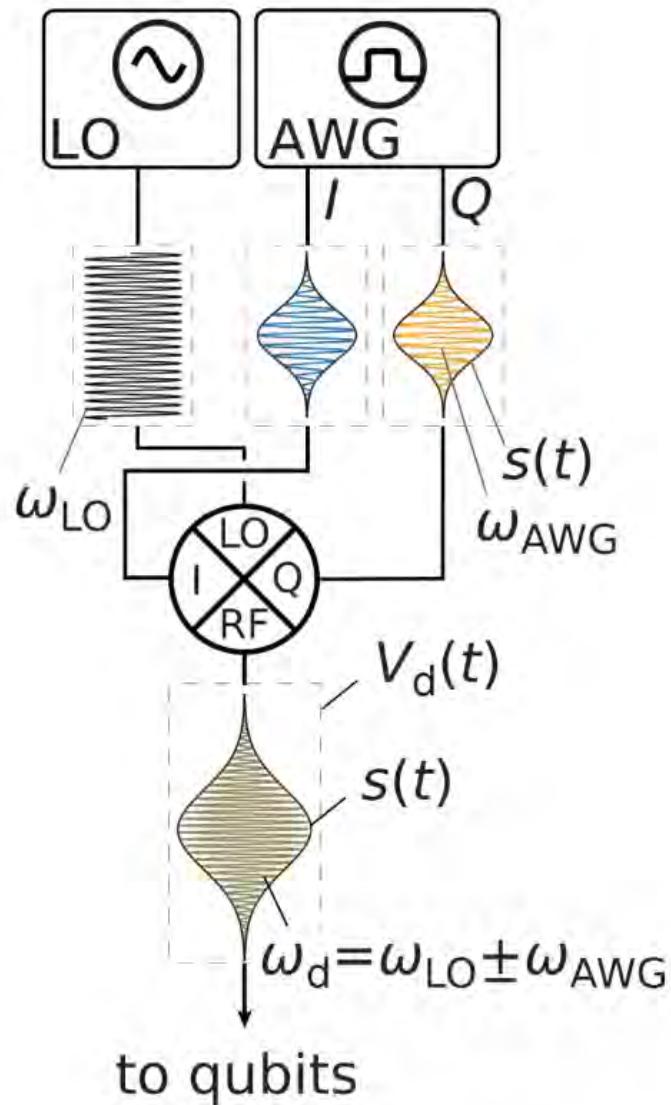
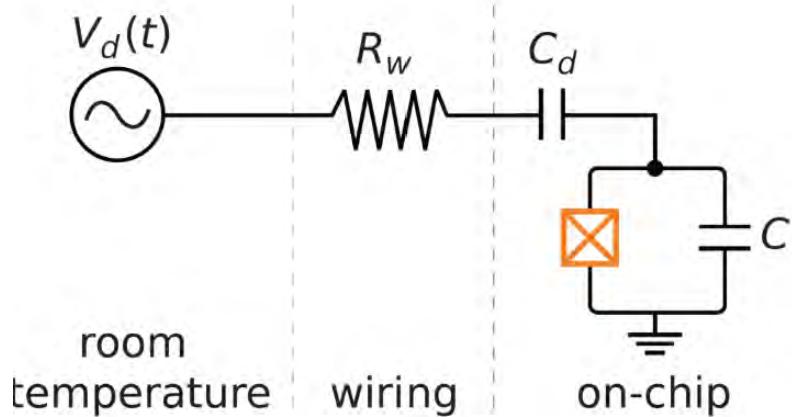


Figure 108:

**I-Q Mixing** So we'll take a local oscillator which is close to resonance with our qubit but not exactly the same as the local oscillator, and then a lower frequency arbitrary waveform generator which outputs two tones from an I in AQ port and puts them an I and queue in the mixer [q] and the intermediate frequency such that adding these two frequencies together makes it resonant with the qubit.

## Application



**Control applied via  
capacitive or inductive coupling  
of a microwave pulse to the qubit**

Figure 109:

We can use single sideband mixing to get rid of the other sideband, and typically, this pulse now will be sent down to the qubit and it will either be capacitive coupled [q] directly to the qubit or if this is a SQUID, we could magnetically couple it through an antenna, so control is applied via capacitive or inductive coupling of a microwave pulse to AQUBI.

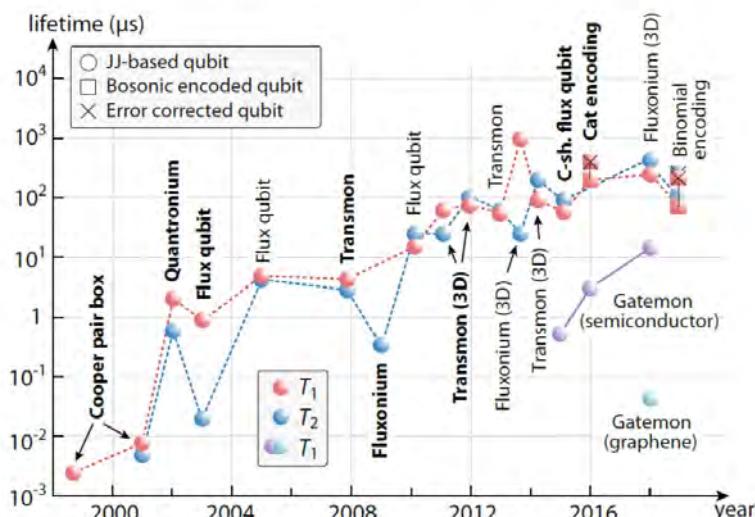
It looks like this on the block sphere, so that was an in-phase around  $X$ . This is minus  $X$ . We'll now move to the quadra port . If we put the pulse in the quadrature port, we now rotate around the y-axis, and with half the area, that's a  $\pi$  pulse, which brings us back to the equator, and another  $\pi$  pulse brings us back to the North Pole

## Major qubit types at MIT + LL

<b>Flux qubit:</b>	T2 = 23 us
<b>c-shunt flux qubit:</b>	T2 = 100 us
<b>2D transmon:</b>	T2 = 100-200 us
<b>3D transmon:</b>	T2=150 us
<b>Fluxonium</b>	T2 = 250 us

## 0.49 Engineering Improved Coherence

“Moore’s Law” for T<sub>2</sub>



M. Kjaergaard, WDO, et al., arXiv:1905.13641

P. Krantz, WDO, et al., Appl. Phys. Rev. 6, 021318 (2019); arXiv:1905.13641

WDO & Welander, MRS Bulletin (2013)

Figure 110:

Finally, one reason the superconducting cubit is doing well today is because the past twenty years have brought about a remarkable improvement in its coherence times. The first superconducting qubit was around a nanosecond coherence very short, but we've improved over five orders of magnitude in the intervening time. Here are a few references if you'd like to read more about superconducting qubits or the materials behind them, but just to give a few numbers that we measure in my own group, we are around 100 microseconds to 350 microseconds for the coherence time. T two [q]. And this has been due to remarkable improvements by groups worldwide improving the materials, the fabrication, and of course the design of these qubits know choosing different inductors, capacitors, inductor values and numbers of those elements to create a new type of qubit, a new artificial atom that may be insensitive to certain noise that we have in the environment.

## Single-Qubit Gate Fidelity > 0.999

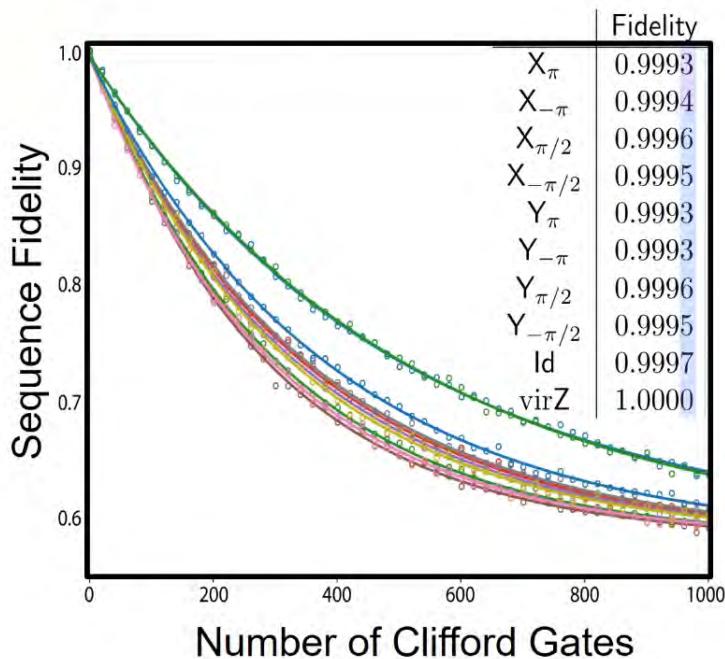


Figure 111:

## 0.50 Gate Fidelities

<sup>1</sup><sup>2</sup> If we ask what the fidelities are of our gates today, our single qubit fidelities are greater than 99.9% percent or three nines, and our two qubit fidelities, I just list them here for two gates, one is called IWGP, the other is Controlled-Z, and you can see that these are almost three nines. And at a March meeting, one of my students, in fact presented a result where we now have three nine fidelity in a two qubit gate and four nines fidelity [q] in a single qubit gate. So I won't go into more detail on this, but if you're interested, here are two references about the theory and the experiment.

<sup>1</sup>F. Yan et al. PRApplied 10, 054062 (2018): Theory

<sup>2</sup>Y. Sung et al. PRX 11, 021058 (2021): Experiment

## 0.51 Nascent Commercial Quantum Processors

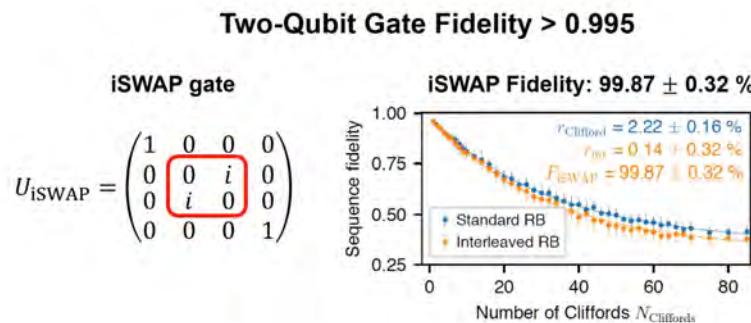


Figure 112:

This is what quantum computers look like today. Above is one of IBM's. I think this is the Eagle processor with 100 and 27 qubits.

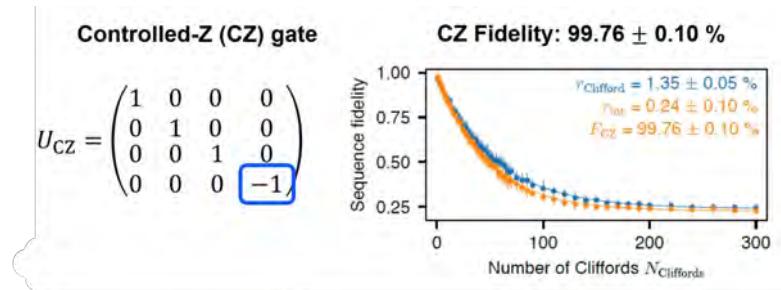


Figure 113:

This is Google's Sycamore processor with 54 qubits. This is their Quantum supremacy demonstration chip.

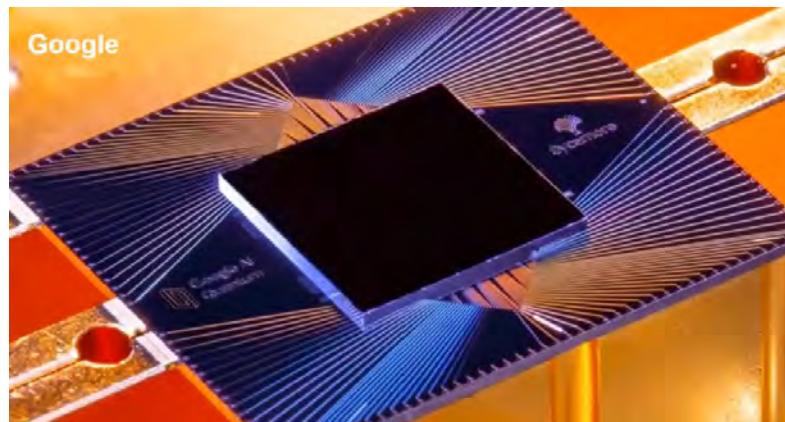


Figure 114:

Rigetti is a publicly traded company. Ion does trapped ions, [q] so the first three were superconducting. This is a trapped ION quantum computer.

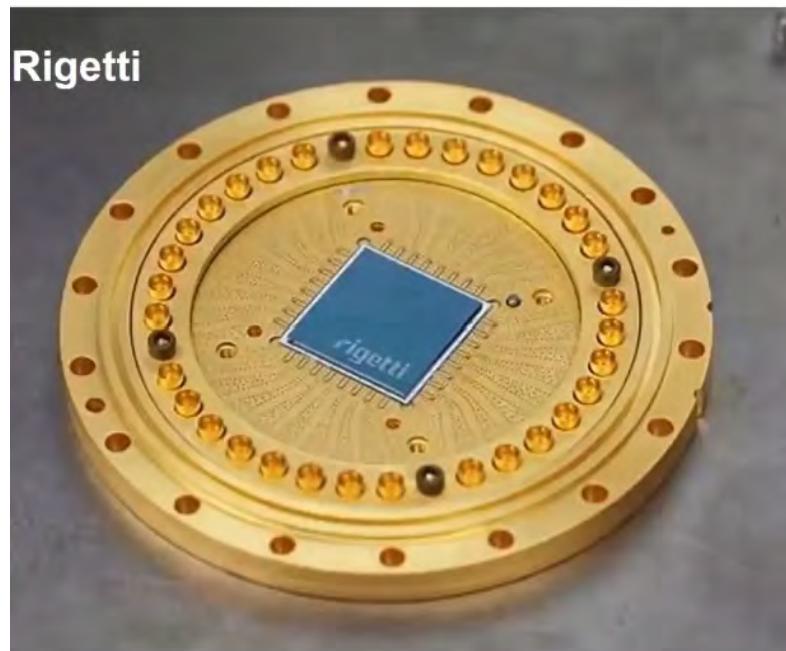


Figure 115:



Figure 116:

D-Wave is based on superconductors but does a different type of quantum computing called annealing, and this, of course, you will recognize as not a quantum computer. This is the Wright Flyer, and we know and remember it as the dawn of aviation because it was the first plane that was manned and operated under its own propulsion. We remember that when we saw this flight we knew that flight was possible and I include this because I think this is the stage where we are at today. That we know that quantum computing is possible, but it's going to be a long time before we have commercial airlines and we go on vacations in a jet.

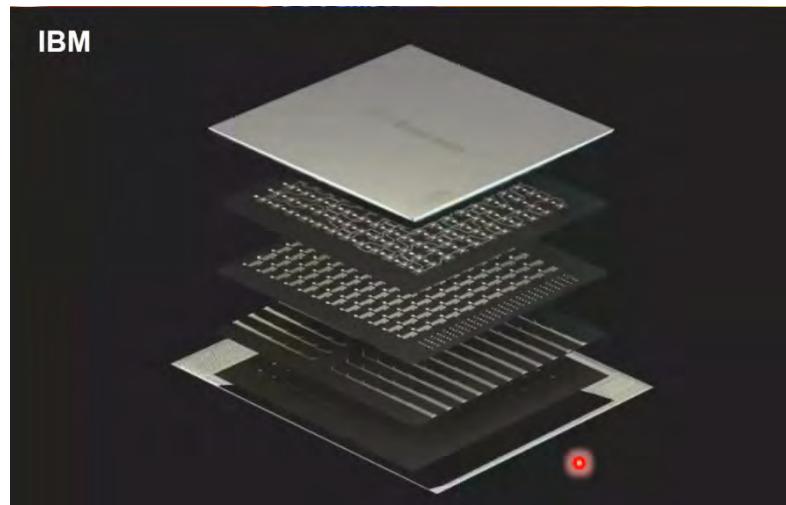


Figure 117:

There's a lot of promise. But to realize the promise, we need to engineer these systems such that they are robust, reproducible, and extensible or scalable, even.

## 0.52 Architectural Layers of a QIP



Figure 118:

So, with that, let me talk briefly about scaling quantum systems and what we need to do. Of course, a system needs architecture, and the one that I'm showing here is an abstracted one from Cody Jones from ten years ago, but I think it still serves a good purpose. At the top is the application or the human-computer interface. This is where we would write and type in our algorithms into a classical computer.

At the bottom are the physical qubits, say in a dilution refrigerator or depending on the modality of the hardware [q], and these are lossy and faulty. Then, as we walk down the stack from our application, we need a logical controller, something that interprets the computer program we wrote and determines all the pulses we need to apply to our physical qubits in order to implement the algorithm. But you'll see there are two more layers here and on every clock cycle. Because the qubits are faulty, we need to do two things. One is we need to suppress the errors as much as possible through open-loop type error cancellation techniques, and I'll show an example in a moment. Then, whatever errors are left, we need to actively measure them or measure syndromes that indicate that an error has occurred and then keep track of those errors and, at certain points in the algorithm, even correct them. All of these things, including the algorithm, need to be implemented by the logical controller.

## 0.53 Lacrosse in the Presence of Noise

So I like to think of this in the context of lacrosse, which I think everyone's familiar with since it was developed in the Northeast by Native Americans. Lacrosse, as you know, is a game with a stick with a basket and the ball is in the basket. You want to carry it down the field and you eventually score goal on your opponent.

Think of somebody who's playing lacrosse the wrong way and what you can see is this person's running and just holding the stick and clearly the ground is bumpy and the act of the running is also bumpy. So not only is the environment noisy, but our control is noisy and the ball falls out quickly. These are the physical qubits that we have today with error rates. What or fidelities of 3 nine s. 49 [q]. Something like that. So, of course,

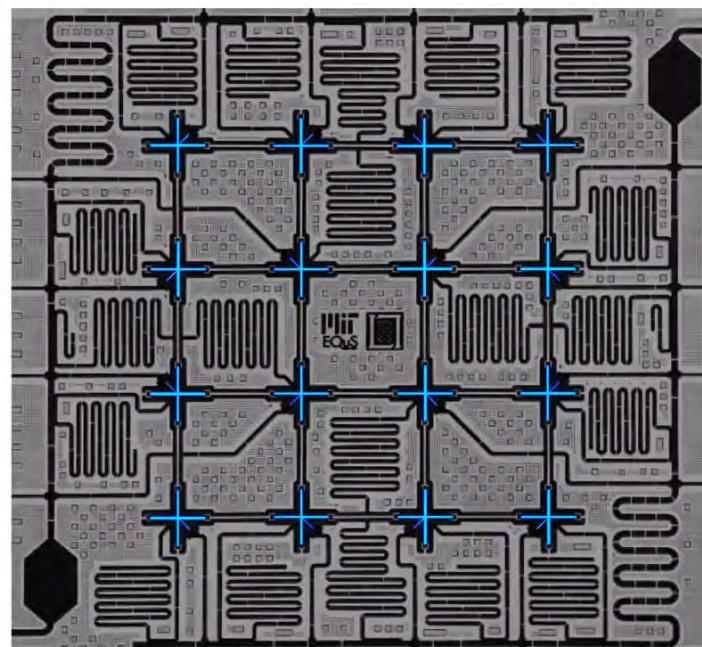
So of course we don't play lacrosse this way. We play it like this. We do what's called cradling. Basically 180 degrees back and forth and back and forth and we just do this in an open loop. Since, this player is not thinking about it, he just does it because he knows it's a good thing to do. And that decouples the noise of the ground and of his running, and the ball stays in his basket and eventually he scores. This is an example of passive error suppression.

And despite that, we've reduced our error rates, but it's not perfect because occasionally the environment or this defender will come up and still knock the ball out. When that happens, two things have to occur. One is you need to recognize an error has occurred, and then you need to go over and do something about it, which is pick it up. You will notice that as soon as this player picks up and corrects the error, he or she just starts cradling once more. So he'll pick it up here and then he starts cradling, it's back to the passive errors suppression right away, so that's what we need to do. And in the interest of time, I will not be able to go into more detail today on how we do that.

## 0.54 Superconducting Qubit Arrays

### 2D Arrays of Qubits

Lattices, Error Propagation, Coherent Errors, ...



Y. Yanay, ..., WDO, C. Tahan, arXiv:1910.00933  
npj Quantum Information 6, 58 (2020)

Figure 119:

# Coherence & Gate Fidelity

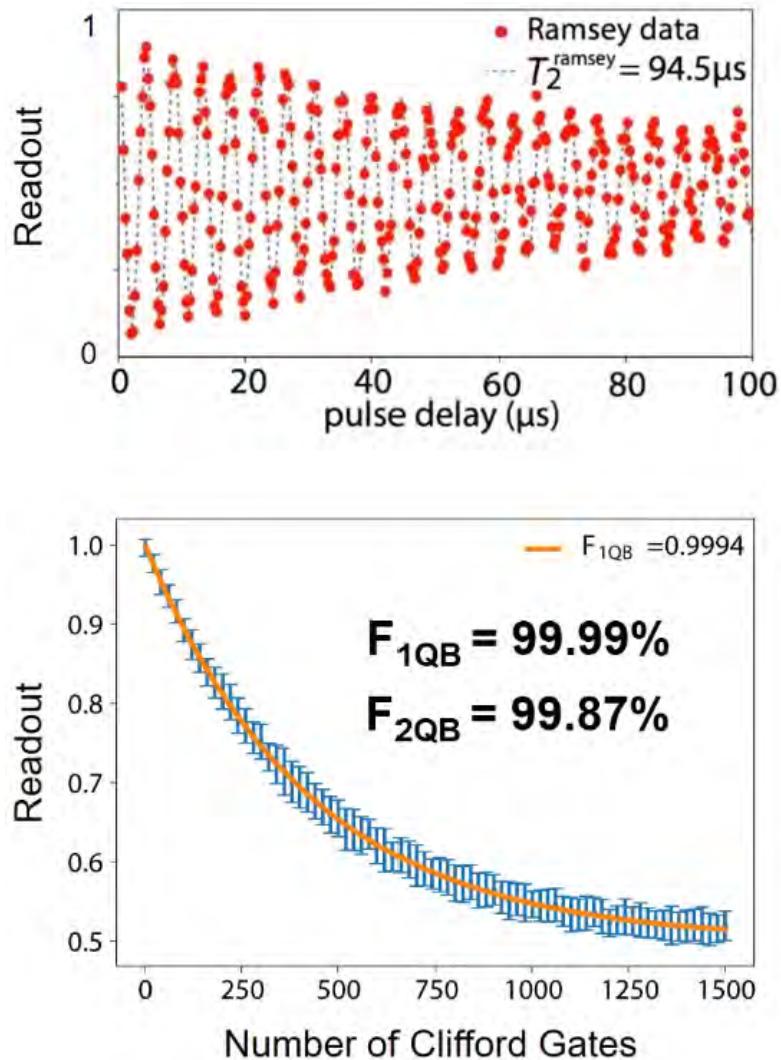


Figure 120:

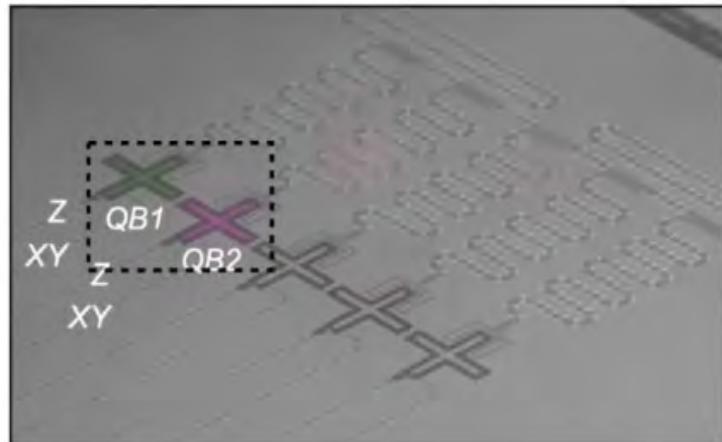


Figure 121:

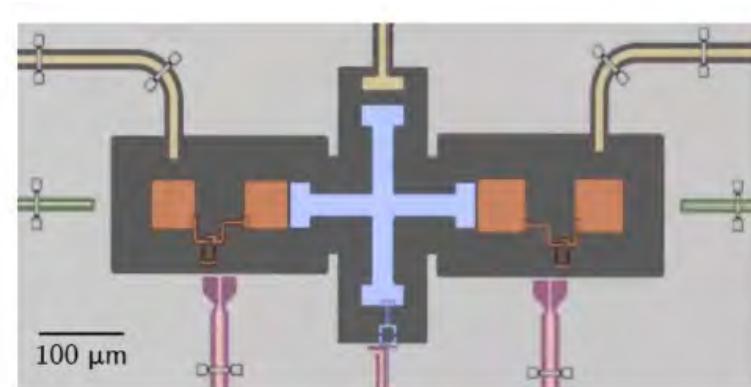


Figure 122:

But let me now just talk in a bit higher level of what we need to do going forward in order to think about implementing error correction. We've talked about the qubits. We've got superconducting qubits with decent coherence time, say on the order of 100 microseconds, or this two cubic gates with single cubit fidelities, four nines and almost three nines for two cubits [q]. And we've done a number of experiments. Our largest chip, or at least with the largest number of qubits, has 16 qubits, so a 4x4 array. But already you can see that it's getting quite crowded coming in from the edge with all the wiring that we need. And of course we know this that this is not an extensible solution. So we need to do 3D integration.

I should mention that we are currently installing a refrigerator that will have a large number of wires capable of measuring up to 64 cubits through a generous donation from Key Site Technologies, and we've been building that in 2022 and 2023 getting the lab ready.

## 0.55 3D Integration for Quantum Processors IARPA Quantum Enhanced Optimization

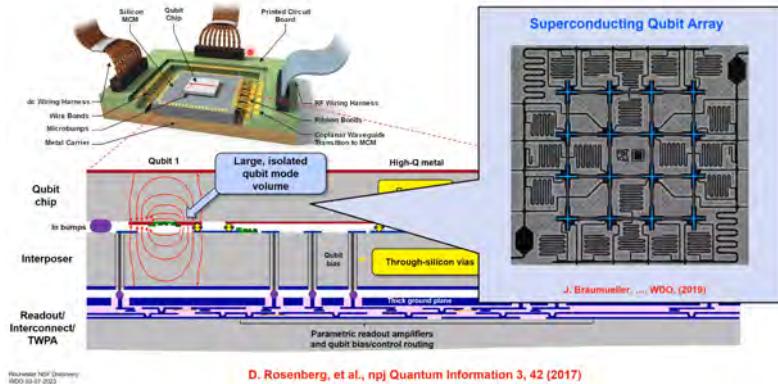


Figure 123:

If we want to go beyond the 16 qubits, we need to have a 3D integration process. And we've developed this over the last many years at MIT and Lincoln Lab. So the idea is that this chip would be flipped upside down, so face down, into this stack that we call a three stack. So the top chip is the qubit chip faced down. We have an interposer with through-silicon vias that are super conducting and then we have a readout/interconnect. We can think of it as a multi-chip module which has multiple wiring layers to allow us to route the signals where they need to go. When the signal is underneath the qubit we want to control the TSV brings that signal up to the qubit and then the signals come back down for readout.

Now one reason we do this you might ask: "Why not just put the qubit chip right on this bottom layer?" This is because the qubits are very sensitive to their environment, and any defects would be viewed as noise and reduce coherence. We use intrinsic silicon which has very high quality factor and the capacitors for example. I just drawn in arbitrarily some electric field lines get sucked into the silicon and they would get sucked in down here as well and they might see some lossy dielectrics [q]. So by putting an interposer layer inbetween also using intrinsic silicon, we are basically protecting the qubit and its environment. So it doesn't look very different or any different at all from when it was just a single chip, and yet we can bring these signals in.

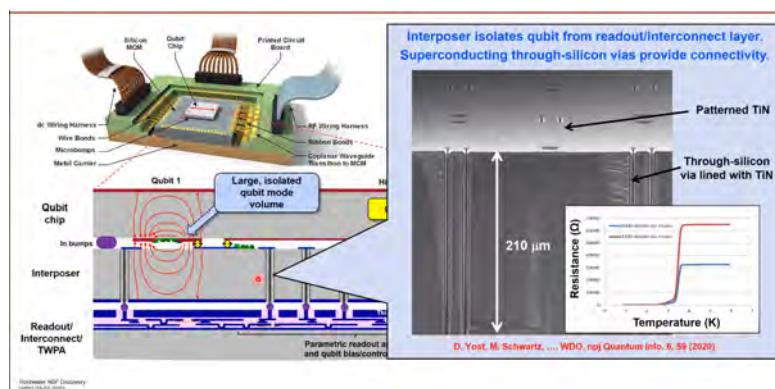


Figure 124:

This is what the through silicon villa look like. These are 200 microns deep holes basically

that are etched in the silicon and then lined in this case with titanium nitride. The aspect ratio is about 10 or 20 to 1, so 10 or 20 micron diameter or oval-shaped vias. You can see here that they are superconducted around 2 to 3 Kelvin with this titanium nitride and what I'm representing. Here are a series array of 1600 and 3200 tvs where the resistance doubles in the normal state because I have twice as many TV tvs, but when I reduce the temperature then they all superconductor [q].

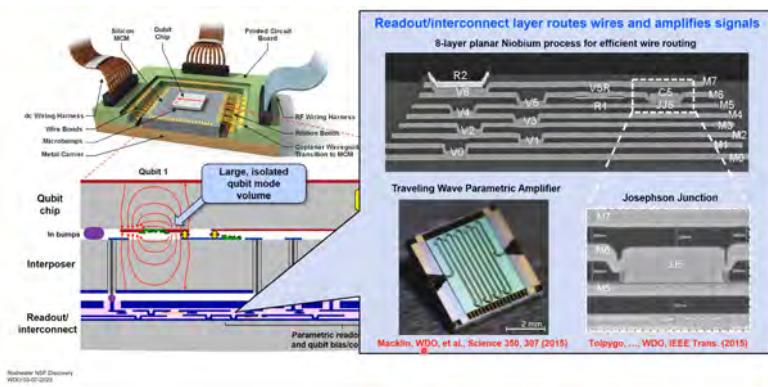


Figure 125:

And then finally this bottom layer in fact is what we've developed over many years, originally for niobium qubits and then of course for niobium single flux quantum logic is this eight wiring layer stack up with active elements Josephson junctions [q]. But for this purpose, basically the wiring layers M0 through M7 allow us to route signals and then vs connect different layers [q]. And this way we can bring in signals from the very edge of the multi-chip module into where we need them to be to address the qubits. Now with junctions, of course, we can make active elements like traveling wave, parametric amplifiers or of course SFQ electronics or its derivatives like adiabatic quantum-flux-parametrons and the like. So there's a lot of promise here, but right now we are not doing that. Right now we are just using this for routing.



Figure 126:

And so with that let me conclude thank you for your attention. I hope this was useful. I want to acknowledge my team down at the MIT campus, in the Engineering Quantum Systems Group and colleagues out at MIT Lincoln Laboratory, and of course, many sponsors at the bottom who make this work possible and thank you for your attention.

## **Compare and contrast: CMOS. NMOS, Josephson Logic**

## 0.56 Overview

I will use this opportunity to compare different technologies and see what lessons from the past we can apply to the future of superconducting supercomputers you folks are building. Some of this material is from my book *The Physics of Computing*, which I wrote for a new course in Gerga tech.

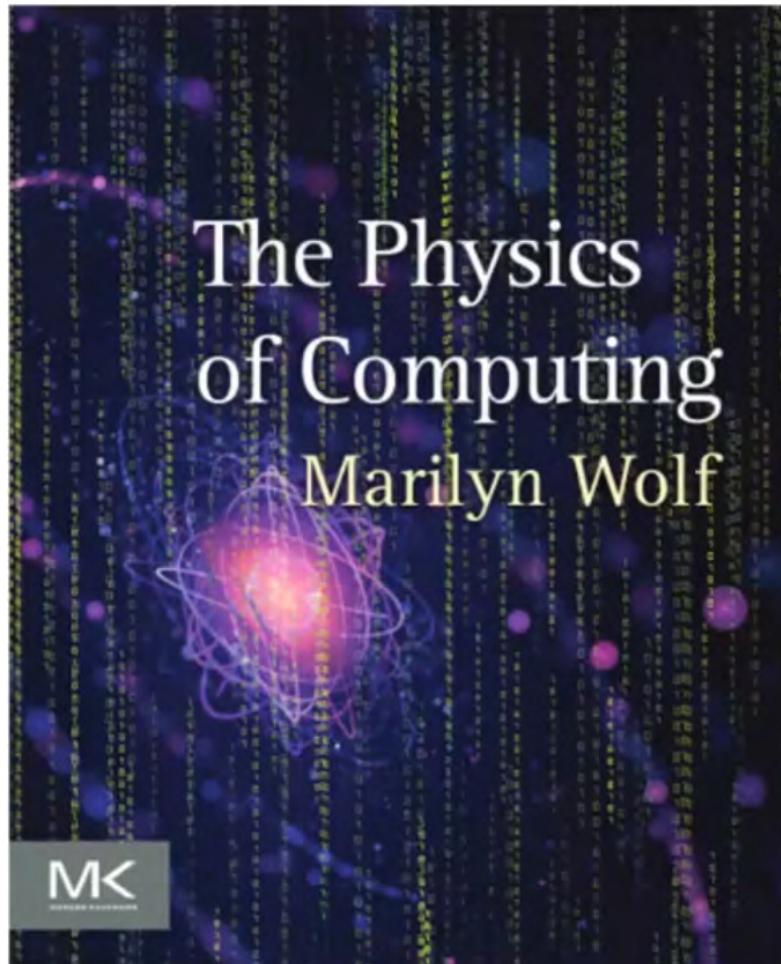


Figure 1:

1. Classic CMOS
2. Lessons from the VLSI era

In particular, I will talk about classic CMOS properties; then, we will look back at the VLSI era to see what approaches we used (back then) to try to get a handle on the challenges posed by VLSI. We will see how that might provide some lessons for the superconducting supercomputer era.

## 0.57 Static CMOS inverter

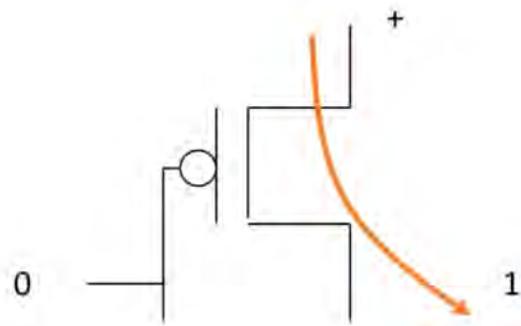


Figure 2:

Let us remind ourselves of how a CMOS inverter works. When the input is

0

, our pull-up transistor (*P*-type transistor) is on, and that connects the output to the + power supply.

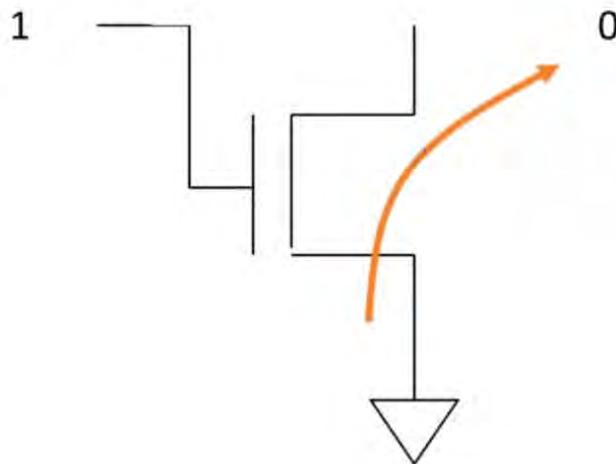


Figure 3:

When the input is

1

, our pull-up transistor comes off, and the pull-down transistor (*N*-type) comes on. Then, we connect the output to the negative power supply (Ground).

## 0.58 Voltages and Logic levels

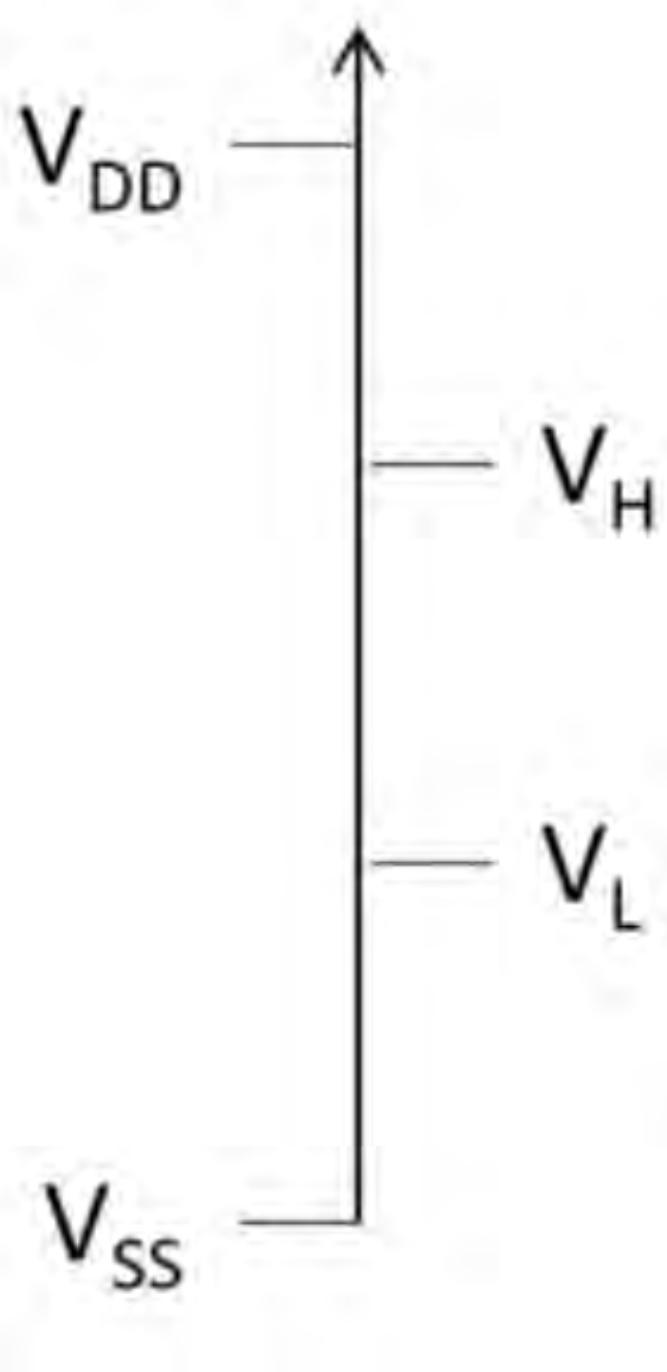


Figure 4:

Figure 4 refers to a scale of voltages in which our inputs and outputs can reside.  $V_{DD}$  at the top and  $V_{SS}$  at the back. So, in steady state, as you saw, the logic levels go to one of the power supply levels ( $V_{DD}$  or  $V_{SS}$ ), but setting a single voltage to be a logic level is unwise and entity Impractical.

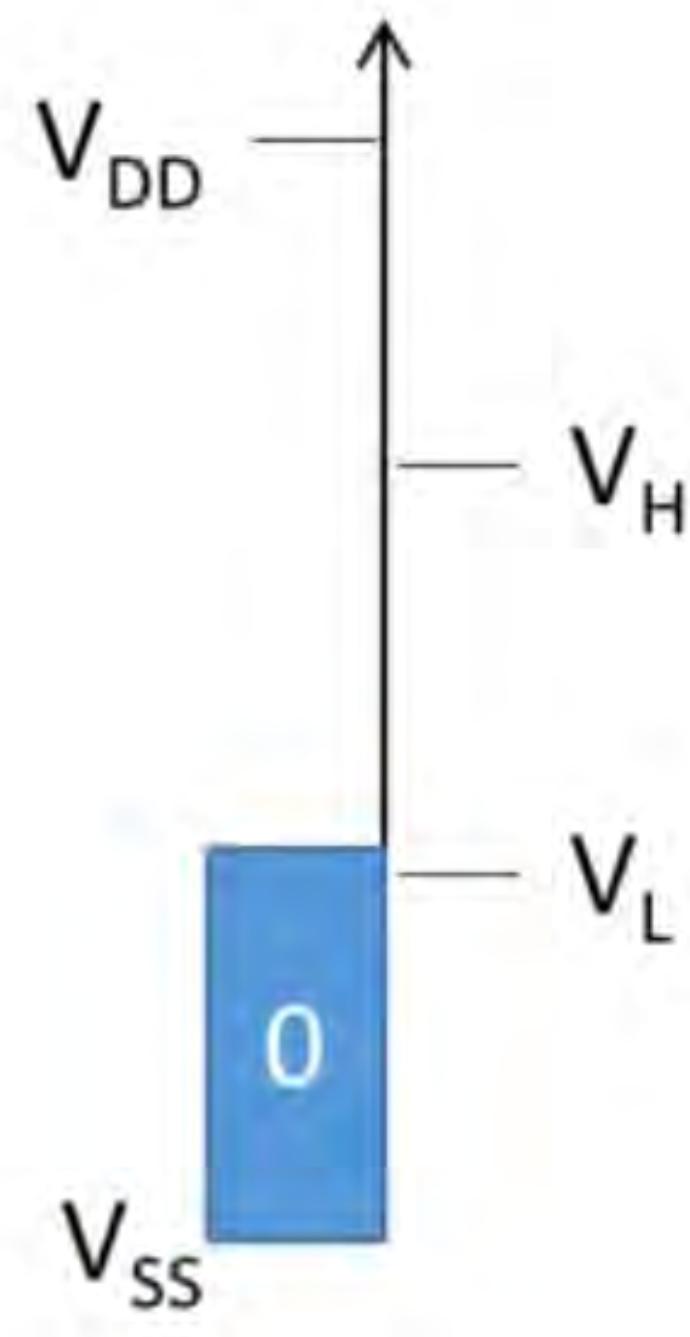


Figure 5:

To remedy this, let us develop some ranges for these voltages to represent a logic 0. Anywhere from  $V_{SS}$  (Ground) up to this  $V_L$  level. That's a logic 0.

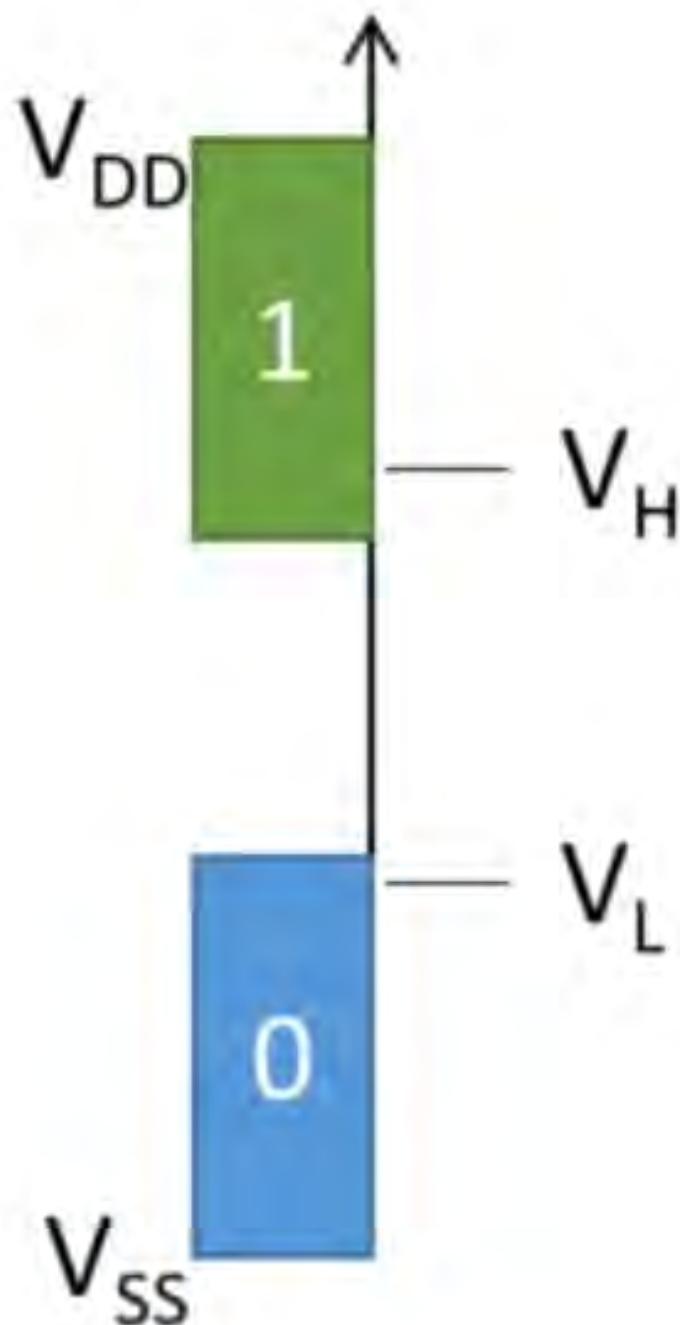


Figure 6:

Next is our range for a logic 1; from  $V_{DD}$  down to  $V_H$ . You will notice that there is a gap. So what do we do with this gap?

We could define  $V_H$  and  $V_L$  such as those levels touch, But then we will have some possible confusion when a signal is right around that transition level.

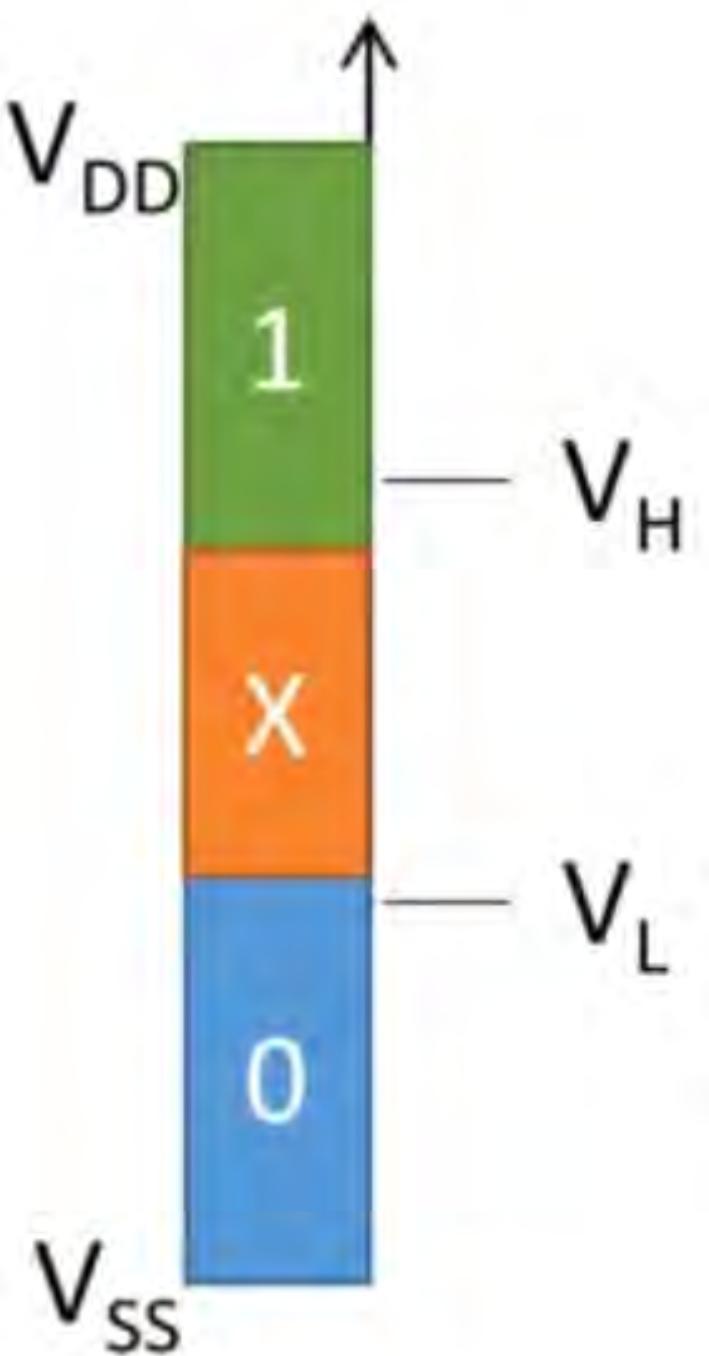


Figure 7:

It is better to define a range of unknown values called

$X$

. So, as a logic single transition, it will go from 0 through the  $X$  range up to 1. Or from 1 through the  $x$  range down to 0. We do not want it to end up in the  $x$  range because then we

do not know what value the gate wants us to compute, but we do want to have a separation zone between a logic 1 and a logic 0; we don't want a tiny amount of random noise to send us directly from 0 to 1

## 0.59 Voltage waveforms and logic levels

We can start to think about how those logic levels play out with Time.

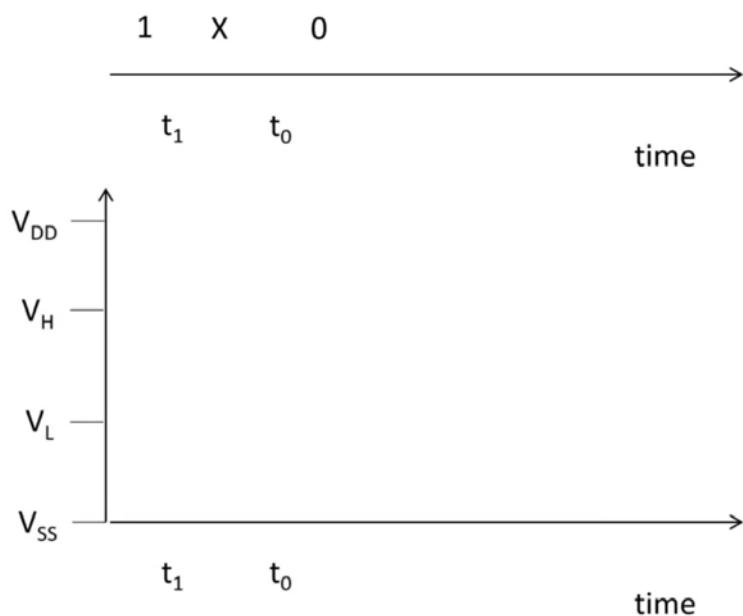


Figure 8:

Okay, so we have a time axis here and we can look at this in the discrete space in which we start at a 1 go through x and end at 0.

Compare and contrast: CMOS, NMOS, Josephson Logic

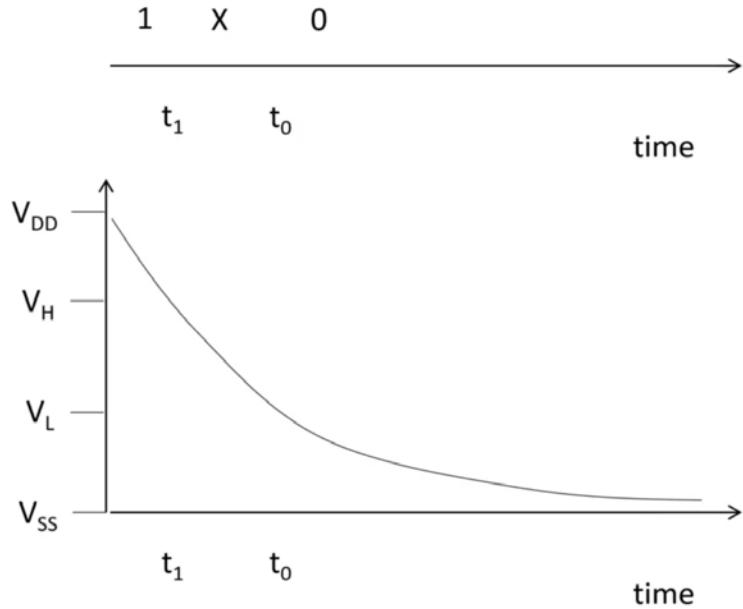


Figure 9:

We can then look at the continuous singles that are actually what logic singles constitute on a chip.

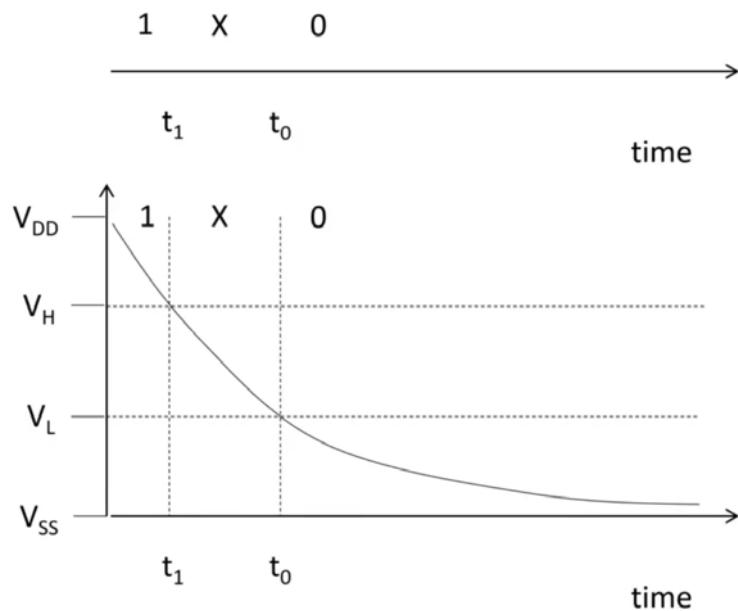


Figure 10:

So, in this case, we go down from the high power supply down to  $V_{SS}$  and we will cross the boundary between different zone at differing times. You can see that we cross the boundary at  $t_1$  we cross the boundary from logic 1 to logic  $x$ . Then at time  $t_0$  we cross the boundary from the  $x$  region to the logic 0. So, while taking this continuous signal and extracting it into discrete values from 0, 1, and  $x$ , we can assign real value times for these events (when I go from 1 to  $x$  or

from x to 0). So far, we have only been talking about one gate; what about two gates? One that talk to another.

## 0.60 Logic level compatibility

**Noise margins measure compatibility:**

$$NM_H = V_{OH} - V_{IH} \quad (0.3)$$

$$NM_L = V_{IL} - V_{OL} \quad (0.4)$$

$$(0.5)$$

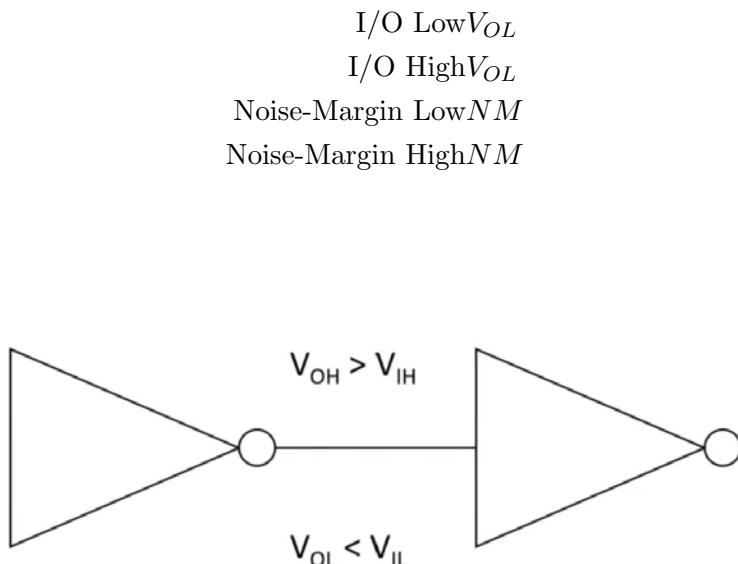


Figure 11:

Well, have  $V$  output low

$$V_{OL}$$

the output voltage level for the logic 0 from the left-hand gate.

We have  $V$  output high

$$V_{OH}$$

the output voltage level for the logic 1 for the left-hand gate. We want the right-hand gate to see a logic 0 when its preceding gate sends a 0 and a logic 1 when its preceding gate sends a 1. The input gate has its own levels. It has a voltage level  $V$  input low  $V_{IL}$  that is the maximum at which that gate will consider that input a 0. We have a  $V$  input high  $V_{IH}$  that is the maximum at which that gate will consider that input a 1. We want  $V_{OL}$  to be below  $V_{IL}$ ; otherwise, the left-hand gate could think it is sending a 0, but the right-hand gate sees an x. Similarly, We want  $V_{OH}$  to be larger  $V_{IH}$ ; otherwise, if the left-hand gate sends a 1, we don't want the right-hand gate to see an x. So, the difference between output and input levels is called noise-margin.

$$NM$$

We have noise-margin Low

$$NM_L$$

Compare and contrast: CMOS, NMOS, Josephson Logic

and noise-margin High.

$$NM_H$$

In some of the older logic technologies, high-low noise margins can be very different; they are more symmetric in CMOS.

Moving on to more dynamic behavior

## 0.61 Saturating inverter model

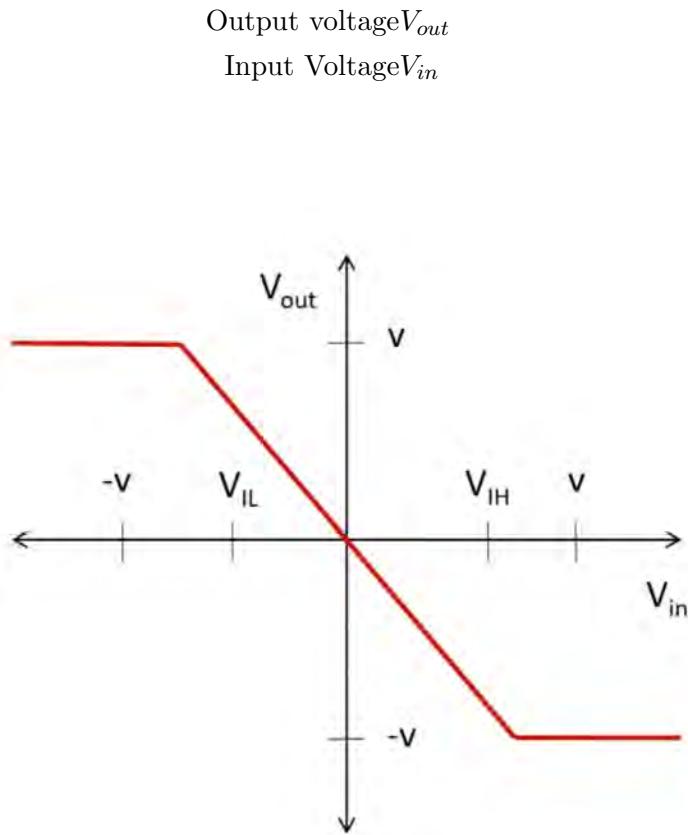


Figure 12:

Above is a simple model of an inverter. The two axis are the Input Voltage (x-axis).

$$V_{in}$$

and the Output Voltage on (y-axis)

$$V_{out}$$

At very low voltages, the output is high, and at high output voltages, the output is zero. Essentially we have taken the axis and changed the origin to be in the middle these logic ranges. This will help us to see what is going on. At either end, the inverter saturates; that is, we change the input to be higher or we change it to be lower, and that does not change the output. In between, this simple model says that there is a linear relationship; we know that gates are not that easy to deal with, but this model helps us to understand the basic principles

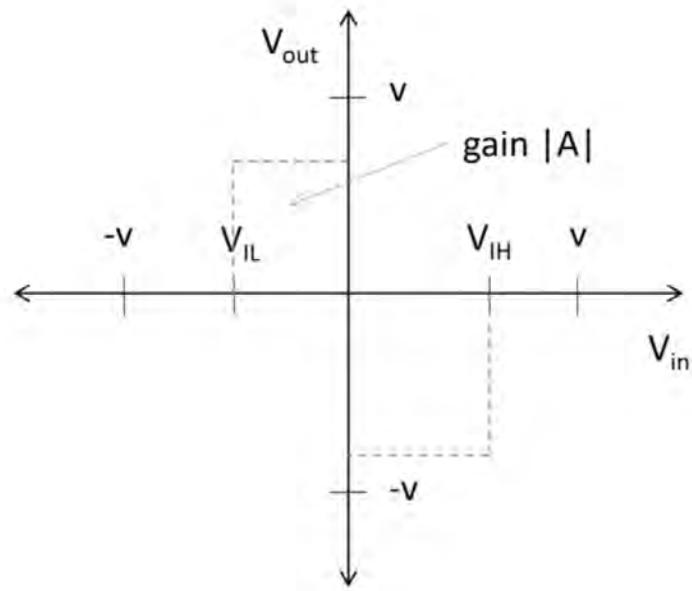


Figure 13:

The slope above is the gain:

## 0.62 Gain and signal levels

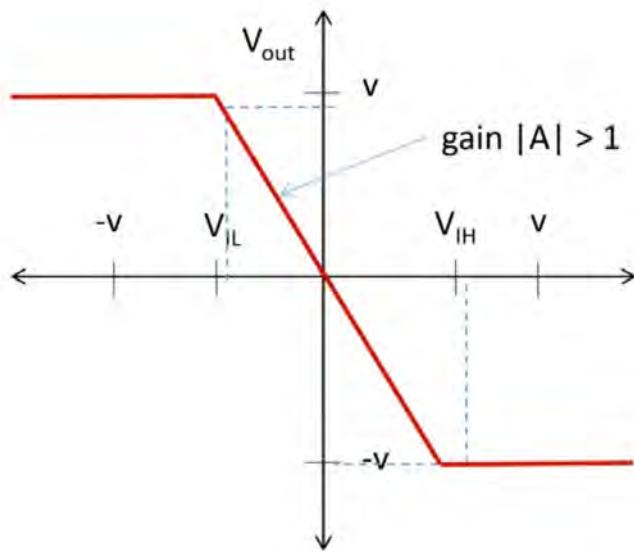


Figure 14:

Logic gates are amplifiers, so the gain is the ratio between the  $V_{out}$  and  $V_{in}$ . For example, if we were building an audio amplifier, we would want gain to be about  $10^{100}$  range. Clearly, if we are designing a logic gate, we want that number to be at least  $gain|A| < 1$ . If it is  $gain|A| > 1$ , that

means that we have put in a voltage that we think is solid on the input side but, in actuality, will give us a less than solid on the output side.

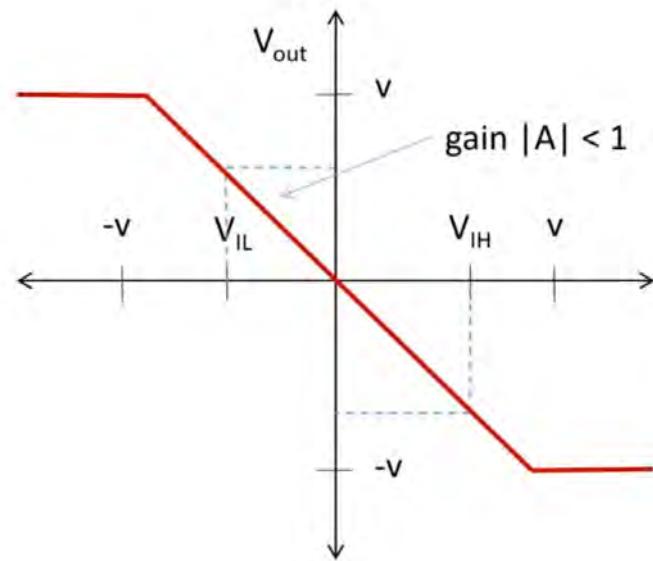


Figure 15:

Now, let us look at a lower slope. That means that we are moving the relationship between the input and output to a less-than-ideal location. We can not make the slope completely vertical, but we would want it to stand up quite a bit.

## 0.63 Gain and delay example

X-axis: Time:

$t$

Y-axis: Voltage

Voltage for Logical 1:

$v$

Voltage for Logical 0:

$-V$

Now, we can use this notion of gain to think about singles over time instead of a transfer curve. Below, we have centered the voltage access in the middle of the logic levels. The red line represents the output for

$$|A| = 1$$

So that the output is (inverted) the mirror image of the input.

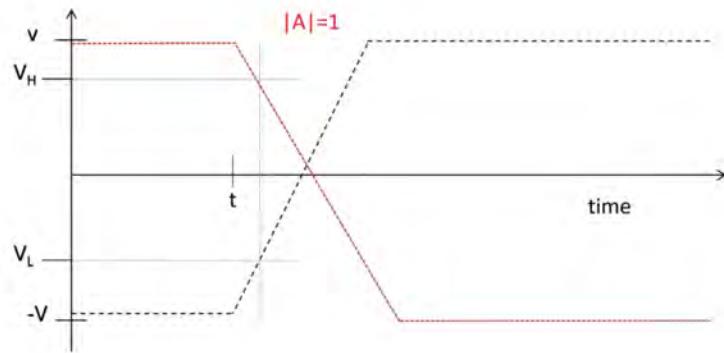


Figure 16:

If we move to.

$$|A| > 1$$

The output transitions faster than the input.

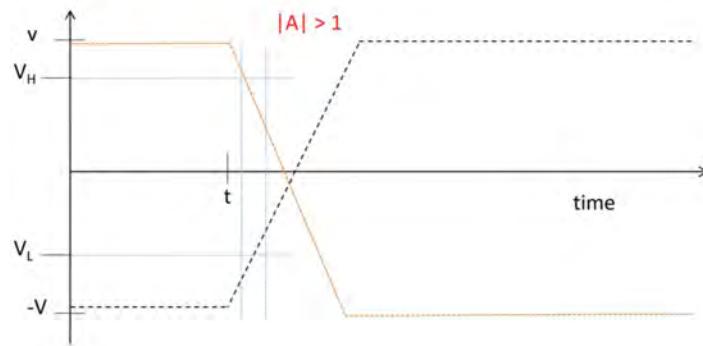


Figure 17:

Then, if we have

$$|A| < 1$$

Then, the output transitions are slower.

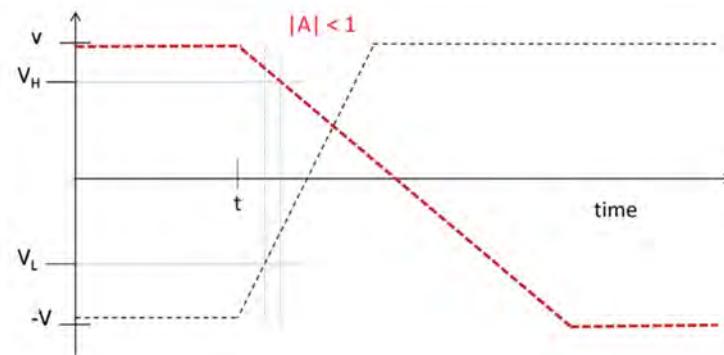


Figure 18:

## 0.64 Transistor-level view of-slow rising inputs

Now, we have not talked much about the Transistor curves yet, so let us move on.

X-axis: voltage across the transistor/across the channel  
 $V_{ds}$   
Y-axis: the current through the channel  
 $I_d$

The red line is the current curve for the power supply inputs.

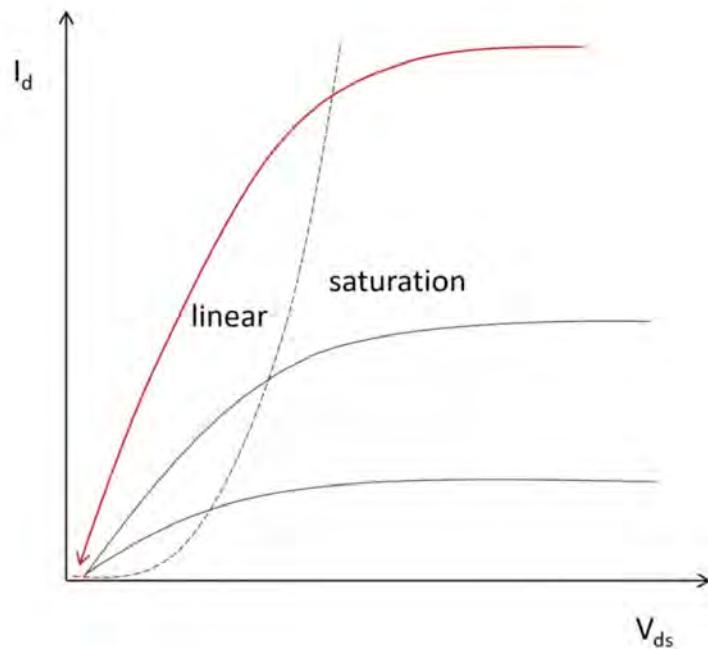


Figure 19:

If we have a fast-rising input, we will jump very quickly to one of these curves and ride it down to zero current as we take the gate down to zero.

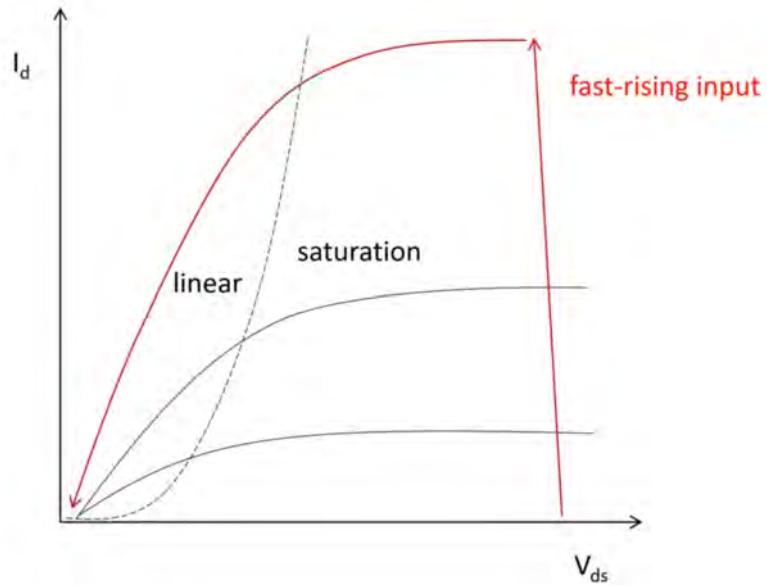


Figure 20:

What happens if the input is slow-rising? First, the trajectory would not be as vertical, and it would take longer to reach the maximum current it could drive. That means that for a big part of the transition time it will not draw as much current as it could be. Compare the red line and the blue line. There is a big difference in area between the red line and the blue line. That area is a lost opportunity to get the output of the gate to where we want it to go.

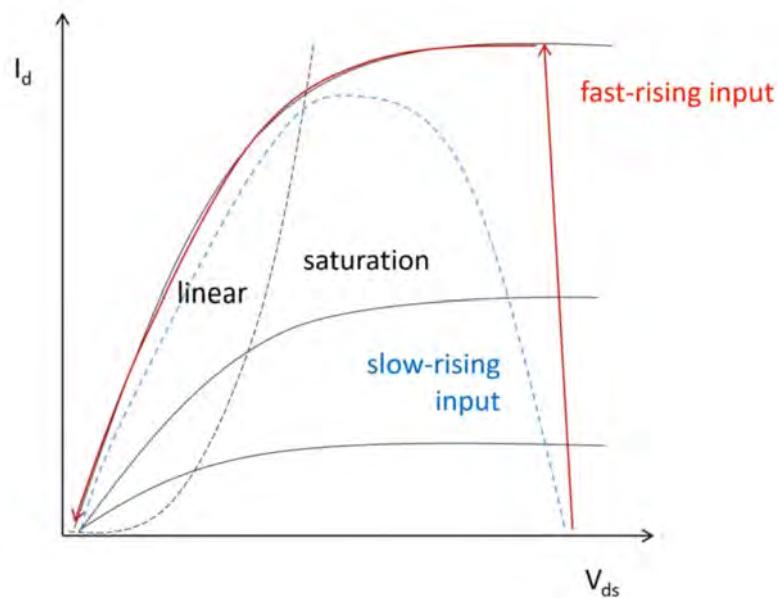


Figure 21:

## 0.65 Step vs. Ramp input

One way to look at this is to compare a step input, which is a perfect transition to a ramp input, as you can see in the circuit simulation below. You can see that the step circuit goes down relatively fast, unlike the ramp input that lowers slower. We can also see that it takes more time for the ramp to get started. For any given voltage that we pick the response to the ramp input is considerably slower than the response to the step input. The gain of CMOS gates is not as high as the gain in some earlier technologies, bi-polar technologies, for example. But at least we have some gain.

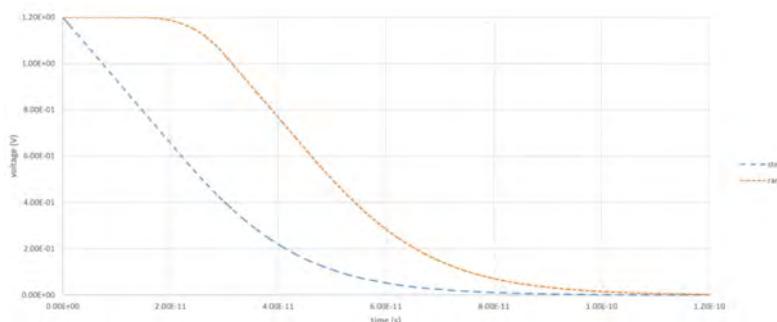


Figure 22:

## 0.66 Signal integrity

let us think next about single integrity and noise

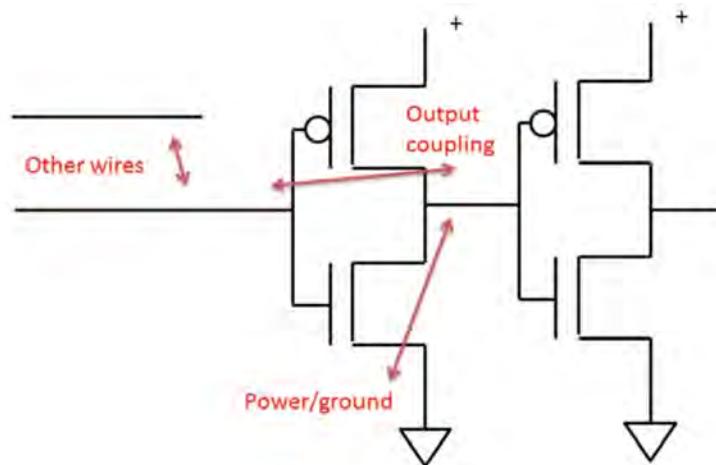


Figure 23:

Here are two gates. Where can we have noise or unwanted values coming in? Well, we could have problems stemming from either the power or the ground singles. Power and ground are not perfect singles; they will be disturbed. Those can be transmitted to the outputs. There can be coupling within gates from the input to the output, and there can be coupling between the input to one gate or the output of another gate to other wires that are completely different singles.

## 0.67 Transmissions lines

Now, let us think for just a moment about transition lines. For longer writes we need to consider the time it takes for a signal to go from one end of the wire to another. The wire is not an ideal conductor, and the time it takes to send from the left-hand side to the right-hand side is noticeable, and for large amounts of wire, it is a large part of its delay. So, the transmission line can have resistive components, capacitive components, and inductive components



Figure 24:

## 0.68 RC transmission lines

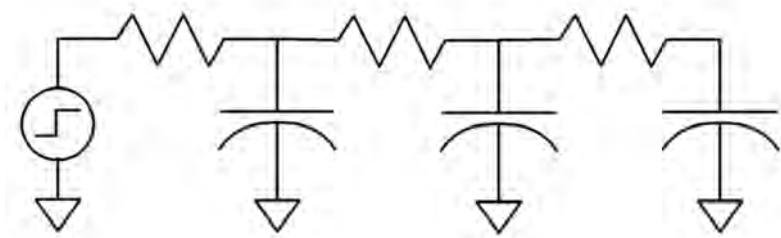


Figure 25:

## 0.69 Pulse propagation



Figure 26:

*Compare and contrast: CMOS. NMOS, Josephson Logic*



Figure 27:

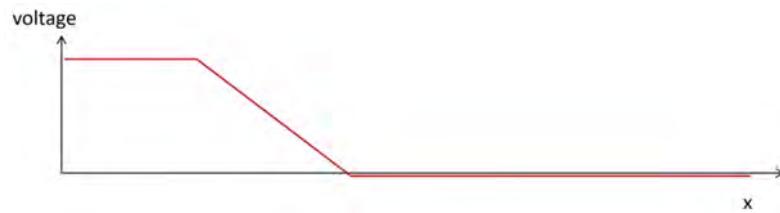


Figure 28:

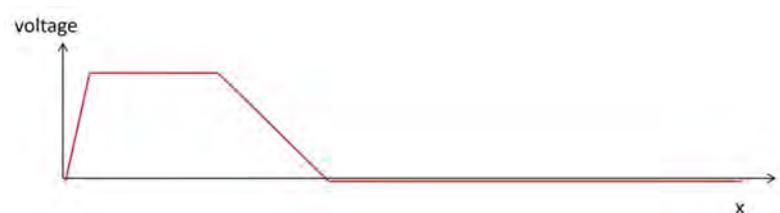


Figure 29:

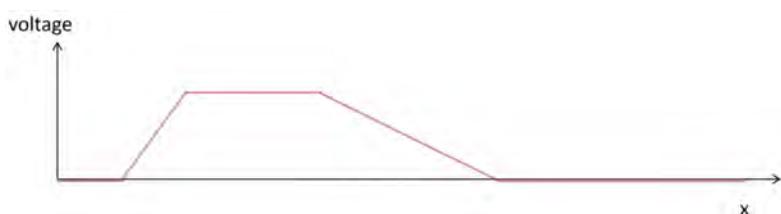


Figure 30:

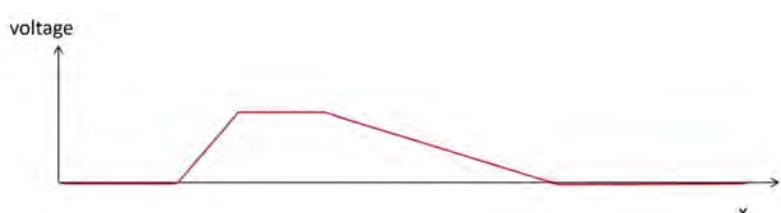


Figure 31:

## 0.70 RC transmission lines cont

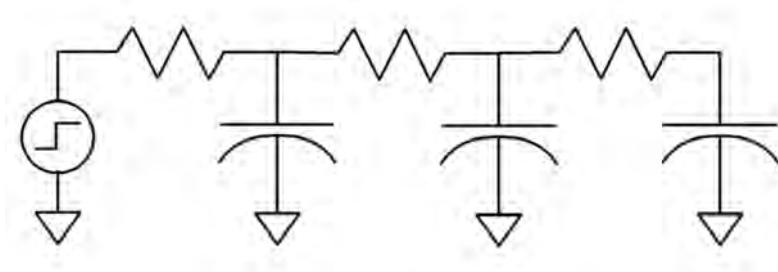


Figure 32:

## 0.71 RC section pulse response

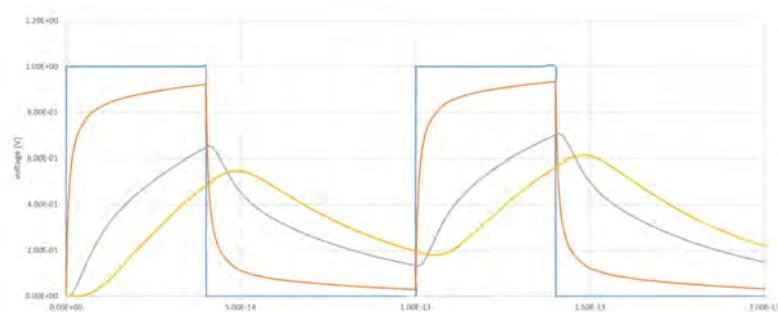


Figure 33:

## 0.72 Logic gate reliability

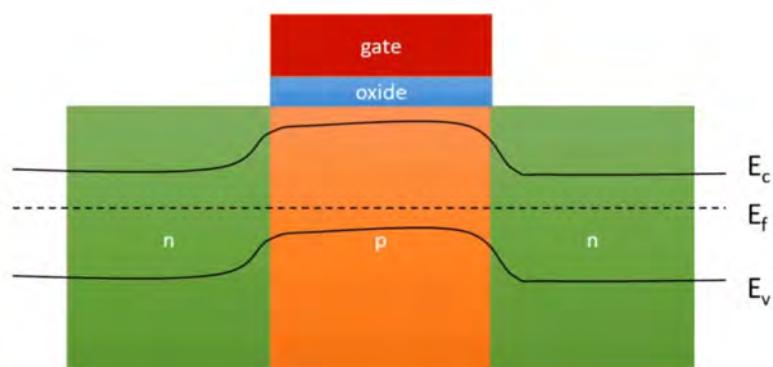


Figure 34:

Compare and contrast: CMOS, NMOS, Josephson Logic

### 0.73 Error rates and energy

### 0.74 Metastability in practice

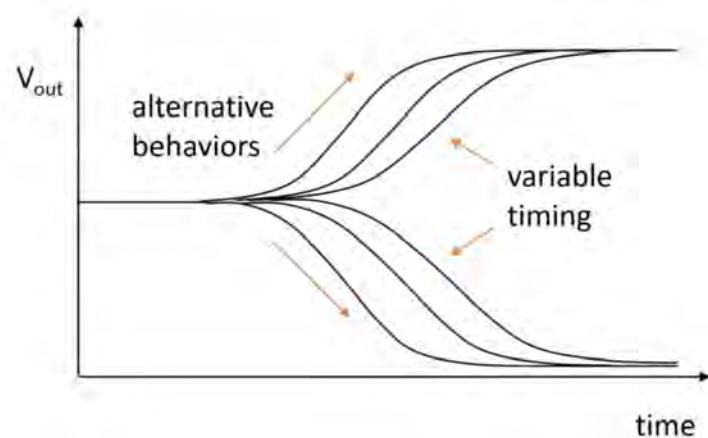


Figure 35:

### 0.75 The vulnerable interval

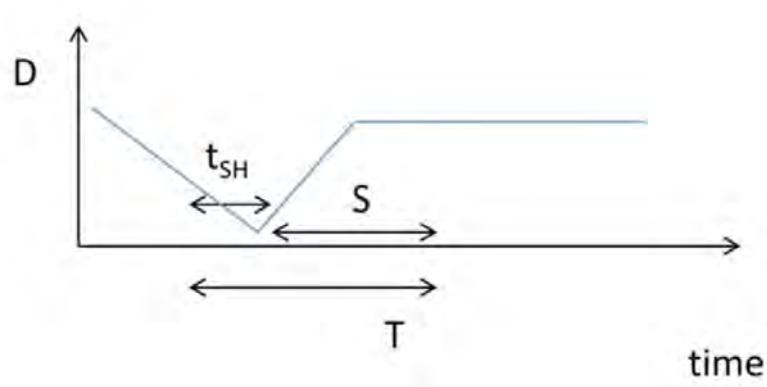


Figure 36:

## 0.76 Example: metastability failure

## 0.77 Metastability countermeasures

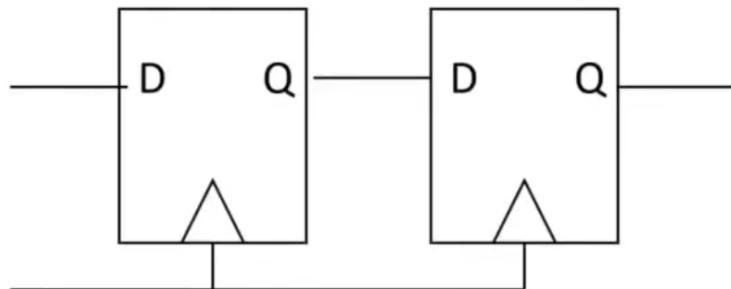


Figure 37:

## 0.78 Compare and contrast

1. MOS:

- Energy consumed for each transition (adiabatic may help).
- Ranges for logic values. Gain helps with noise immunity.
- Dissipation adds delay.

2. Josephson:

- Quantized logic values at small energy levels.
- No gain.
- No dissipation.

## 0.79 Lessons from history

1. The VLSI era is the inflection point between manual design and CAD for silicon.

- Early 1970s: rectangles drawn on butcher paper.
- Early 1980s: layout synthesis, logic synthesis

2. CAD tool flow requires many engineering decisions.

- CAD tools don't pop up overnight.

*Compare and contrast: CMOS. NMOS, Josephson Logic*

## 0.80 Generators

1. Generators are an intermediate step in automation
2. Generates a class of logic.
  - Parameterized for size, function, etc.
  - Tailored algorithms for pieces and their connections.

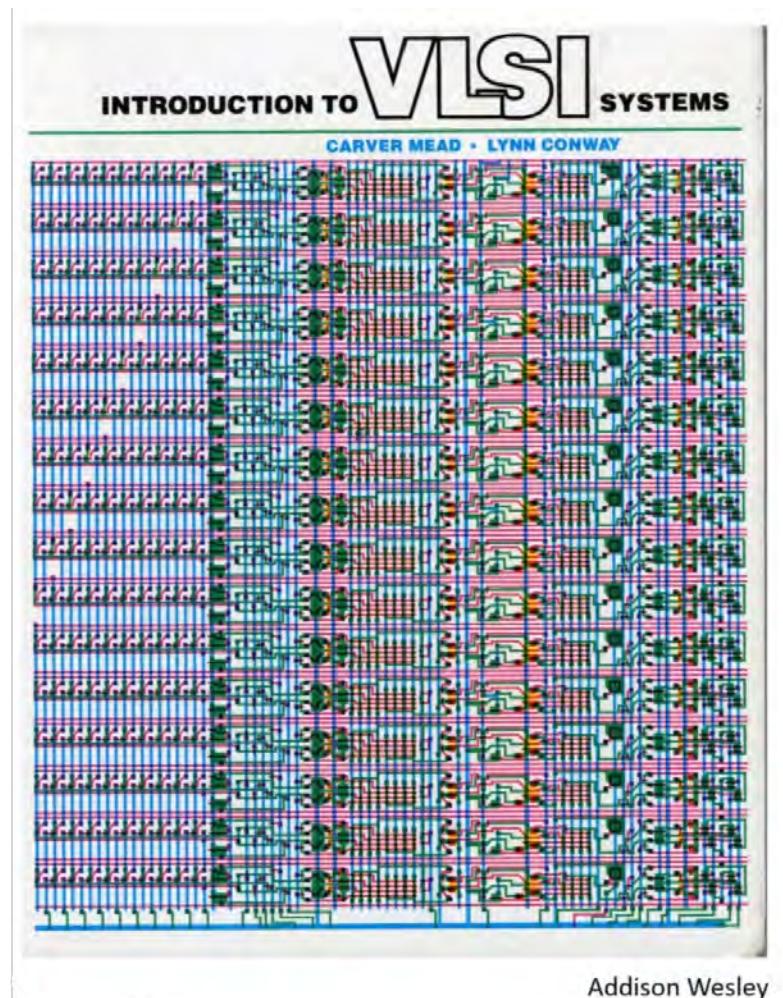


Figure 38:

## 0.81 MOS vs Superconducting generators

1. Some generators still used—memory compilers.

2. Early nMOS generators were simple:

- Tiled structures.

- Simple logic design.

- Structured wiring.

3. Superconducting generators will require:

- In-the-loop wiring modeling.

- In-the-loop circuit simulation.

## 0.82 Why generators for superconducting logic?

1. Allow us to build large modules in a reasonable amount of time.

2. Use generator design to explore the module design space.

3. Encapsulate knowledge about circuit, layout design.

## 0.83 Possible generators

1. Multiplier.

- ALU

- Datapath

- Programmable logic array (PLA).

Compare and contrast: CMOS, NMOS, Josephson Logic

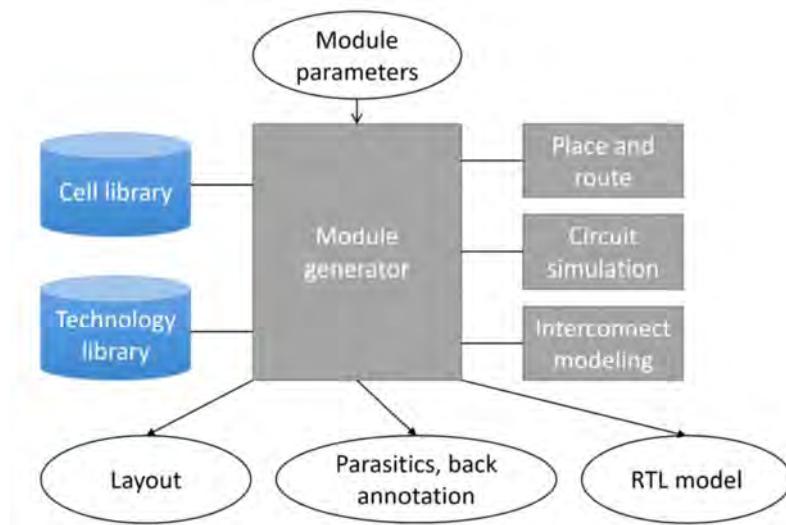


Figure 39:

## 0.84 nMOS logic and PLA

1. NOR gate: small pullup, parallel pulldowns
2. Ideal for PLA

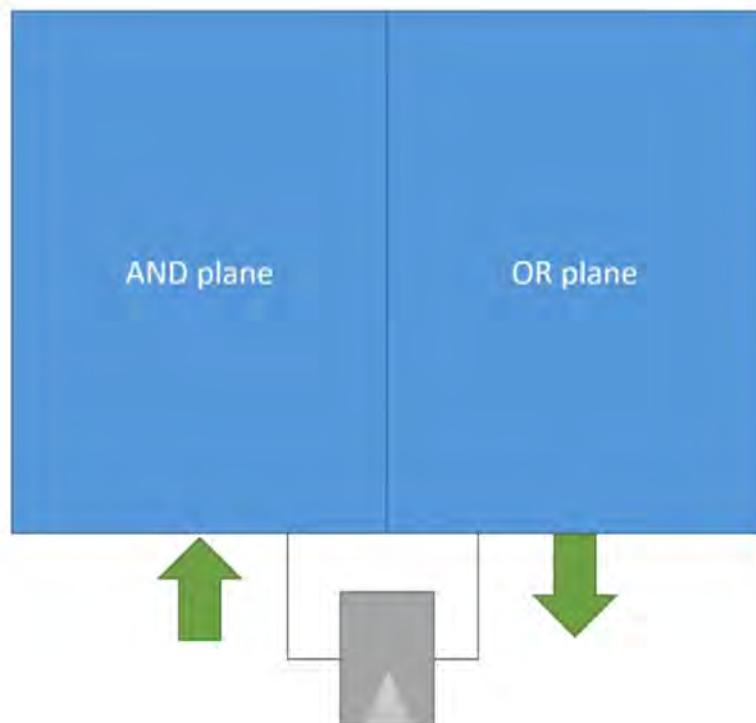


Figure 40:

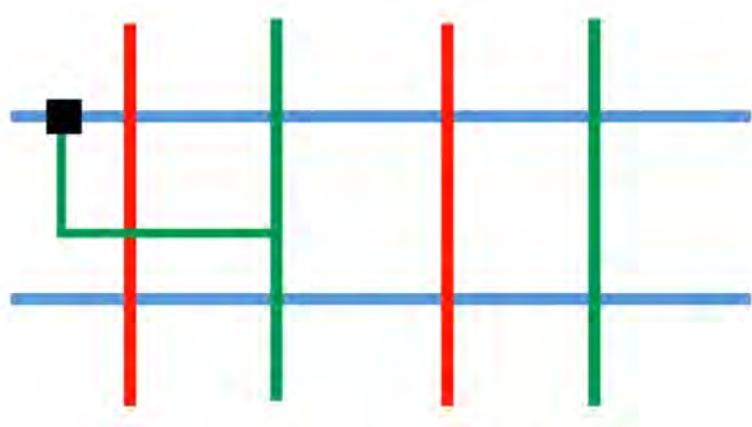


Figure 41:

## 0.85 PLA as FSM Mead and Conway

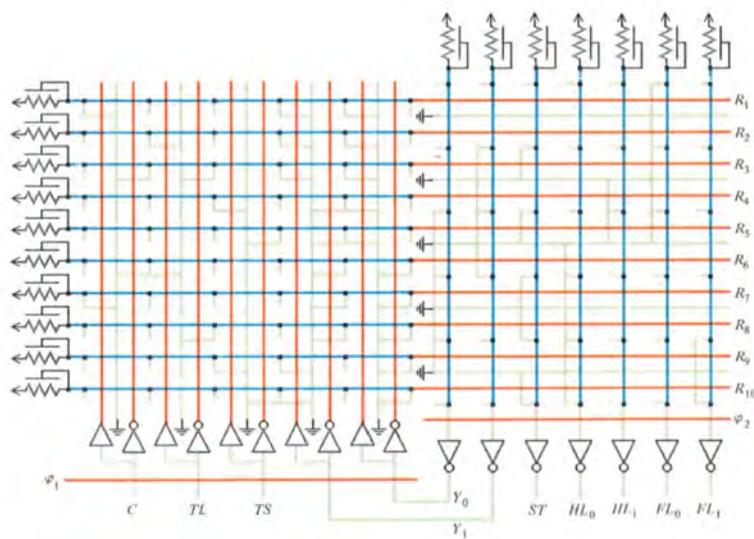


Figure 42:

Compare and contrast: CMOS, NMOS, Josephson Logic

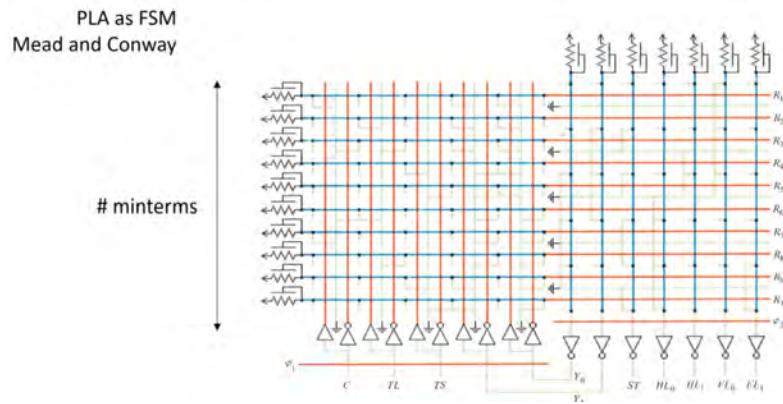


Figure 43:

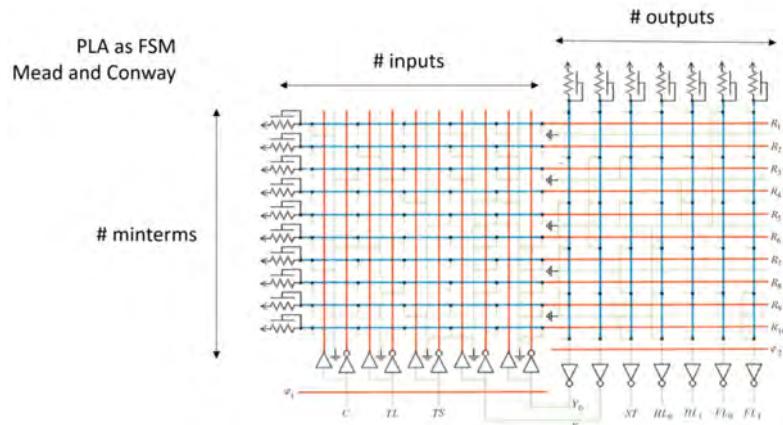


Figure 44:

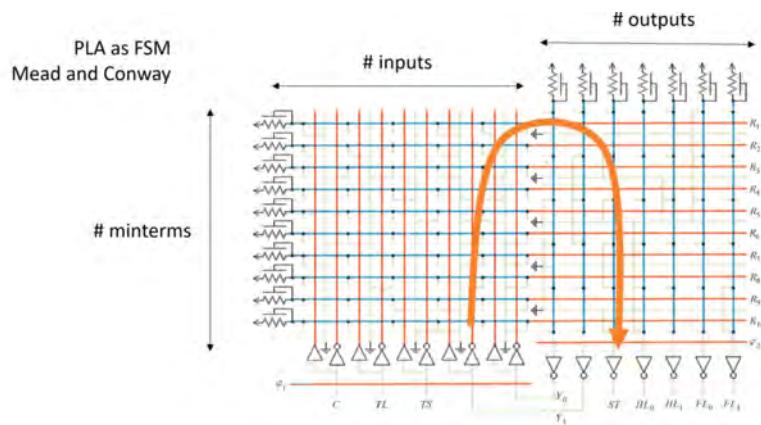


Figure 45:

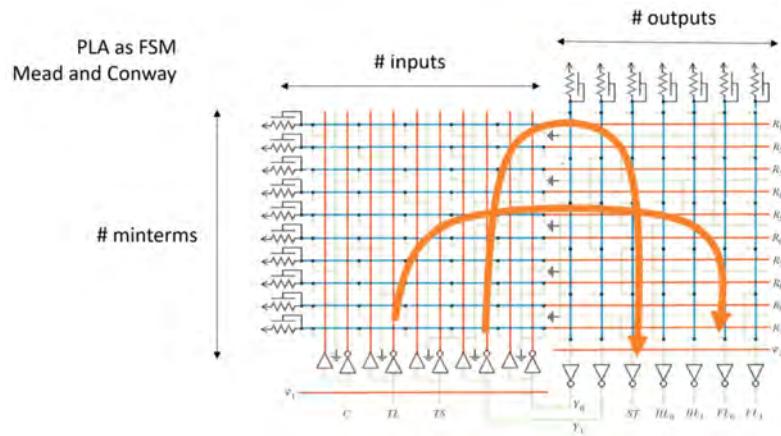


Figure 46:

## 0.86 Summary

- Josephson junction era bears some comparison to early VLSI era.
- VLSI used generators to get passed some early problems.

# **Superconducting Quantum-classical Information Processing Systems**

Thank you for inviting me to give this presentation. Just to clear the air, I am not a professor, and I am not affiliated with any university. Today, I will talk about superconducting quantum - Classical Information Processing Systems. This means I will talk about a hybrid quantum-classical computer made entirely out of superconducting circuits. This does not always have to be the case, but It will be my focus today.

## 0.87 SEEQC global footprint

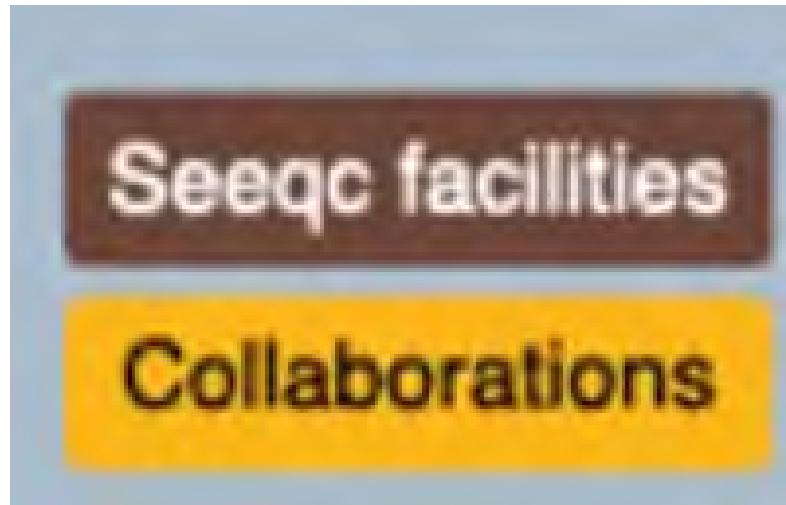


Figure 1:



Figure 2:

Just before jumping into the subject, I would like to tell you a little bit about our company. Seeqc is a recent spinoff of Hypres, which has been a known player in the superconducting electronics space for quite some time and a previous employer of mine. Four years ago, we spun off from Hypres; our main facility sits adjacent to Hypres, including former Hypers facilities that now belong to Seeqc. We have other locations in London, and also Naples, Italy.

## 0.88 Operating: SEEQC System Red, full-stack reference quantum computer

Just before describing the system, I would like tell you we built a quantum computing company because we have a quantum computer. We recently unveiled our "System Red." Red in a spectrum is the first color, and this is our first computer. It consisted of only 5 Qubits.

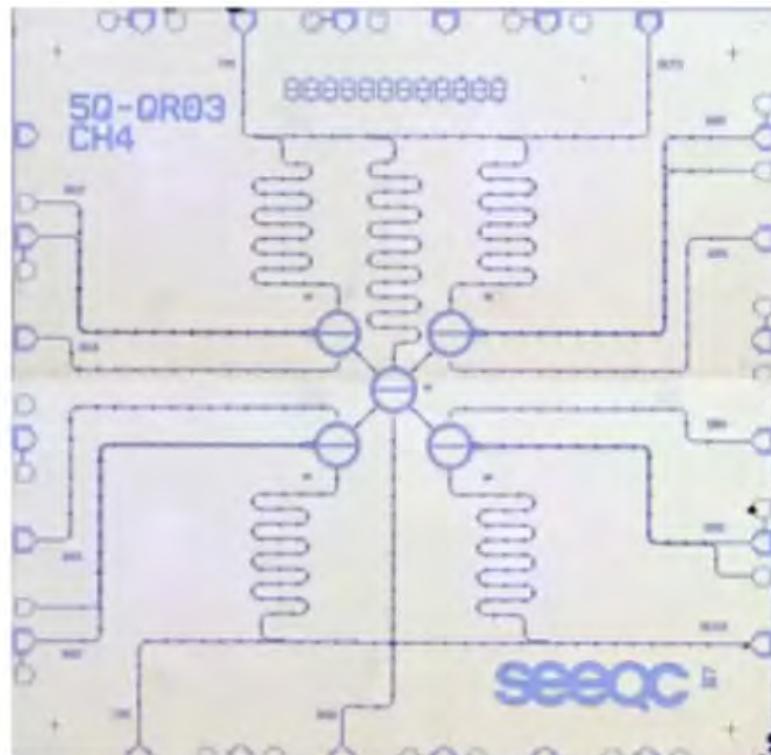


Figure 3:

Above, you see a quantum chip that we designed and was fabricated elsewhere, but it is our design. In the middle, you see 5 round transmons connected in a starlight fashion. Now, It is not just about the Qubit chip itself; we put together the entire system, which means the controller electronics, Firmware, Software, API, etc, meaning that it can actually run algorithms.

Completing this for the first time is quite an achievement for us. It is a reference design, meaning that it is not implemented to our exact specifications and standards, which I will describe in more detail later; it is similar to what other companies in this space configure. Using microwave room temperature controller electronics, in a way, pretty much replicates the conventional approach. To complete the feasibility requirements for the software and control side, meaning what kind of interfaces are needed, it also allows us to hit the ground running. In terms of Adaptation algorithms and quantum algorithms, to see if we can operate with them or if we need to use third-party algorithms.

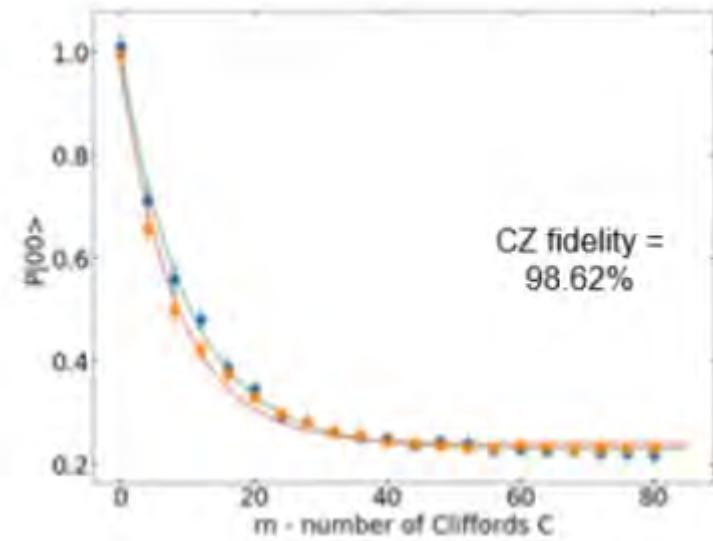


Figure 4:

So the parameters of the gates that we implemented are pretty good there are on pare with the best gates in the industry.

Dispite it only being a 5 Qubit machine the fidelities of the qubit gates are pretty high

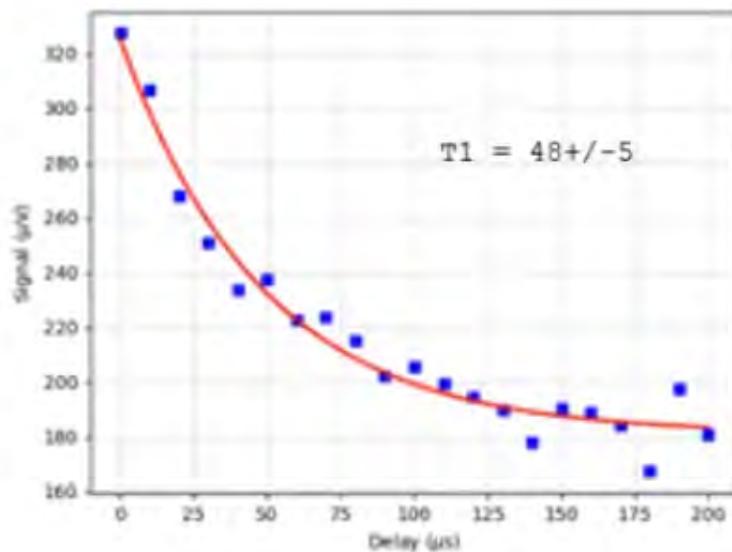


Figure 5:

1. **Algorithmic benchmarking:** Run 2 qASIC error correction algorithms +3<sup>rd</sup> party algos
2. **Firmware and Software:** Build and optimize SEEQC full-stack firmware and software layers
3. **Component Benchmarking:** State-of-the-art infrastructure to test SEEQC proprietary quantum and SFQ chips

4. **R/D:** Confirm SFQ chip spec requirements for next-gen chip-scale SEEQC Orange quantum computer

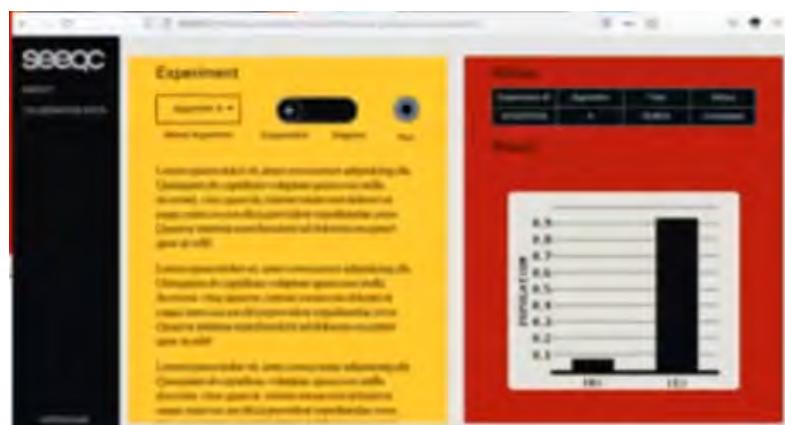


Figure 6:

### **0.89 SEEQC System Red compared to state-of-the-art quantum platforms: public cloud data**

This compares our parameters vs the best parameters in the industry. We have:

SEEQC Red

5 qubits

5 Usable qubits\*

$$\frac{T_1}{T_2} Time = \frac{20\mu s}{16\mu s}$$

$$2-Q gate speed = 39ns$$

$$2-Q gate fidelity = 98.4\%$$

Table 1: Caption

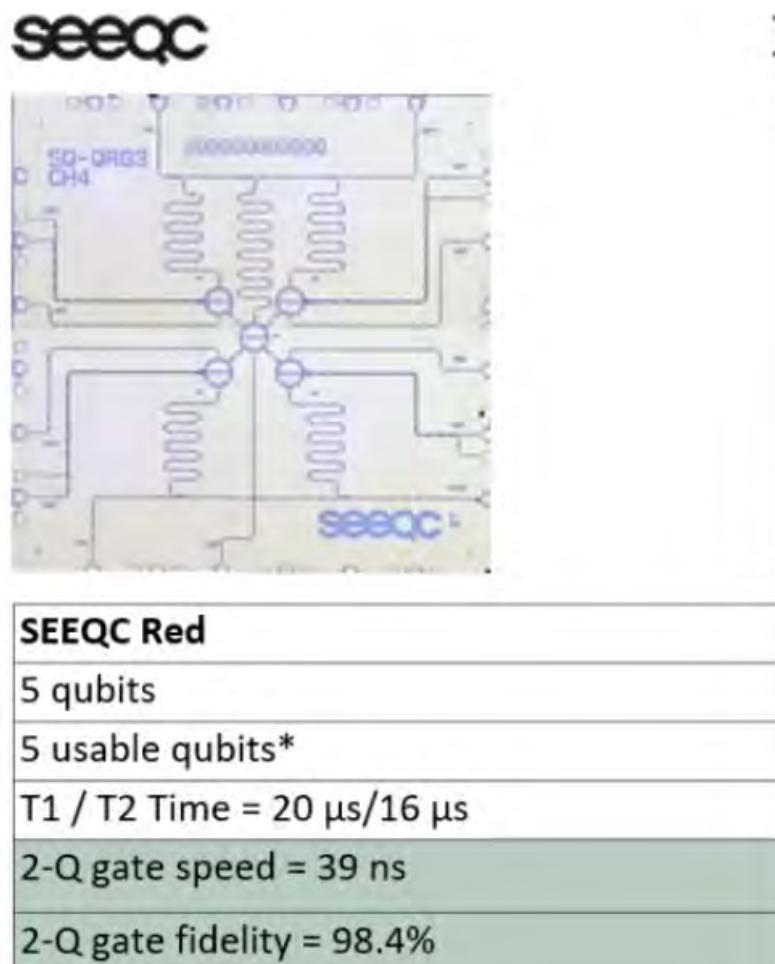
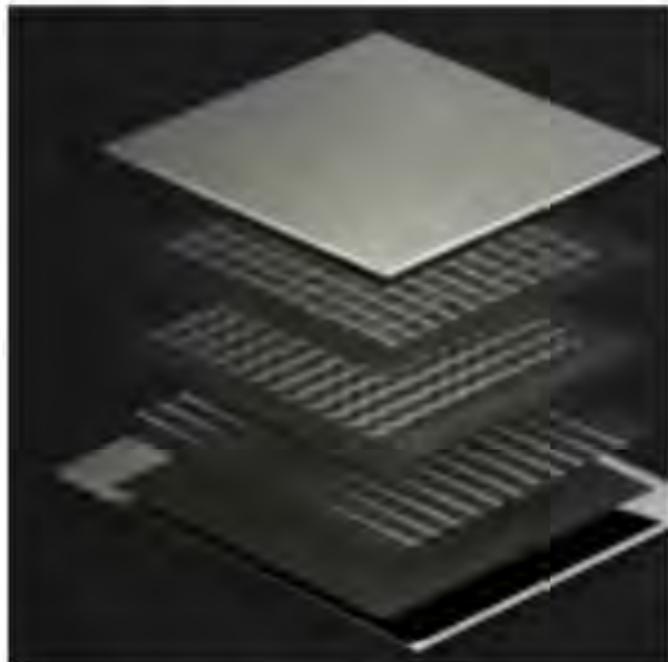


Figure 7:

**IBMQ**



<b>Washington</b>
127 qubits
6 usable qubits*
T1 / T2 Time = 90 $\mu$ s / 94 $\mu$ s
2-Q gate speed = 555 ns
2-Q gate fidelity = 96%

Figure 8:

Washington

127 qubits

6 Usable qubits\*

$$\frac{T_1}{T_2} Time = \frac{90\mu s}{94\mu s}$$

$$2 - Qgate speed = 555ns$$
$$2 - Qgate fidelity = 96\%$$

Table 2: Caption

	
	Aspen M
	80 qubits
	4 usable qubits*
	T1 / T2 Time = 29 μs / 22 μs
	2-Q gate speed = 180 ns
	2-Q gate fidelity = 91.3% <sup>1</sup>

Aspen M

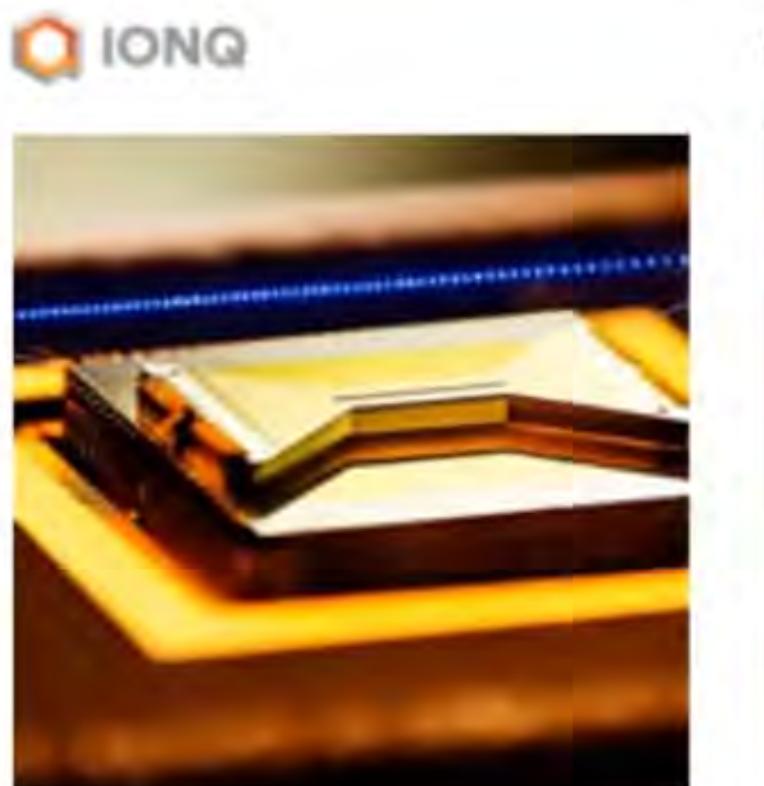
80 qubits

4 Usable qubits\*

$$\frac{T1}{T2} Time = \frac{29\mu s}{22\mu s}$$

$$2 - Q \text{gates speed} = 180ns$$
$$2 - Q \text{gate fidelity} = 91.3^{\textcolor{red}{1}}\%$$

Table 3: Caption



---

### **IonQ Device**

---

**11 qubits**

---

**6 usable qubits\***

---

**T1 / T2 Time = seconds**

---

**2-Q gate speed = 300  $\mu$ s**

---

**2-Q gate fidelity = 98%<sup>2</sup>**

---

IonQ Device

11 qubits

6 Usable qubits\*

$\frac{T_1}{T_2} Time = \text{seconds}$

$2 - Q\text{gatespeed} = 300\mu s$   
 $2 - Q\text{gatefidelity} = 98^{2}\%$

Table 4: Caption

OQC



**Lucy**

8 qubits

4 usable qubits\*

T1 / T2 Time= 41  $\mu s$  / 40  $\mu s$

2-Q gate speed = 187 ns<sup>3</sup>

2-Q gate fidelity = 91.9%<sup>2</sup>

Figure 11:

Lucy

8 qubits

4 Usable qubits\*

$$\frac{T_1}{T_2} Time = \frac{41\mu s}{40\mu s}$$

$$2 - Qgatespeed = 187ns^3$$

$$2 - Qgatefidelity = 91.9\%$$

Table 5: Caption

Of the above, the following implement superconducting Qubits

1. SEEQC

2. IBM Q

3. rigetti

4. OQC

With IONQ using trapped ions. Now, let us look at the specifications of each machine. If we compare the 127 Qubit Washington chip, we see the most physical qubits integrated into an array, but at any one time, algorithms can only operate using 6 qubits. In our case, all qubits can be used for programming.

## 0.90 SEEQC System Red



Figure 12:

## 0.91 Revaluation of The State of Computing

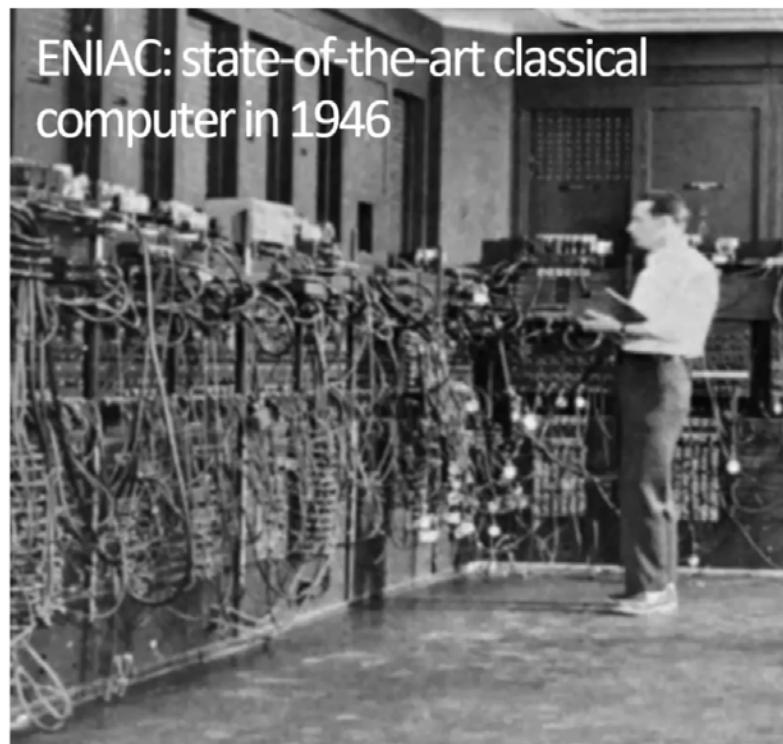


Figure 13:

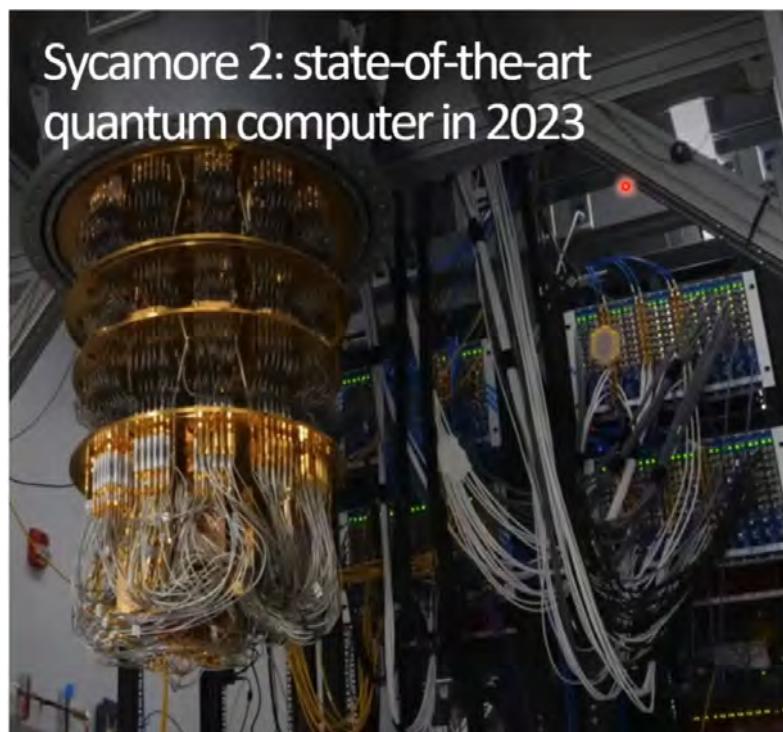


Figure 14:

## State-of-the-art CMOS chip

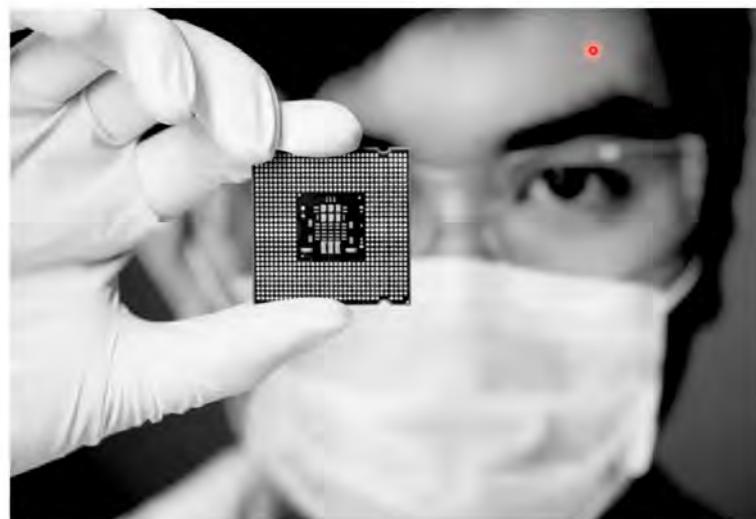


Figure 15:



Figure 16:

## State-of-the-art Single Flux Quantum (SFQ) chip



Figure 17:

## 0.92 Manufactured by SEEQC Foundry



Figure 18:

## 0.93 Unit Fabrication Process

Deposition -Metals -Dielectric

Photolithography - 5 x reduction stepper, 0,25 um resolution - 1x laser writer, 0.6 um resolution  
1x aligners 1.0 um resolution

Etching -RIE -ICP -Ion Beam etch -Wet etch

CMP

Wafer cleaning, inspection and probing - Profiler -Water prober -SEM

## 0.94 Fabrication Process Specifications

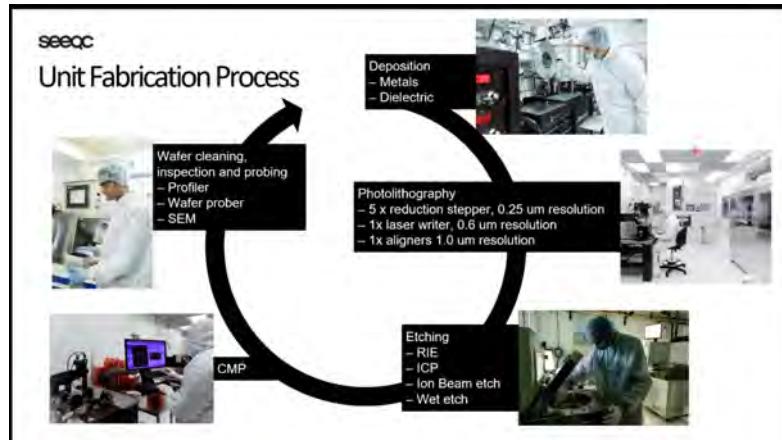


Figure 19:

## 0.95 Standard Fabrication Processes

SEEQC-Q2SL Multi-layer niobium fabrication process for superconducting quantum bits  
SEEQC-C2SL MCM carrier, detectors for CMB-relevant detectors

SEEQC-C4SL Legacy process with 4 superconducting layers

SEEQC-C5SL Process for SQUID microscopy

SEEQC-Q5SL Process for SFQuClass chip

SEEQC-C9SL SFQ co-processor

## 0.96 What else is available

<b>seeqc</b>	<b>Superconductors</b> • Nb, NbNx, Ta, Al, AlMn, Ti, Ha
<b>Fabrication Process Specifications</b>	<b>Critical current density:</b> • Nb/Al/AlOx/Nb Josephson junctions • Available critical current densities for 100, 1000, 4500, 10000 A/cm <sup>2</sup>
	<b>Resistor layer material:</b> • Sheet resistance: Mo: 1.0 Ω/□, 2.1 Ω/□, PdAu: 2.1 Ω/□, MoNx: 4.0 Ω/□
	<b>High Kinetic inductance:</b> • HKIL – 8.5 pH/□
	<b>Inter-layer dielectric:</b> • Low temp PECVD SiO <sub>2</sub> • Low temp PECVD low-loss SiN <sub>x</sub>

Figure 20:

Standard Fabrication Runs - Comprehensive Design Rules - Limited PDK -Functional and DC parametric testing -CAD Support and verification services - Mask-set composition, verification and generation -Cryogenic functional and high-speed testing

Custom Fabrication Runs that included: Photolithography, deposition, etching, dicing, wire bonding (normal metal, superconducting Nb), file-chip bonding with custom specifications. For specialty applications that include: MEMS, GaN, CMOS backend processes

## 0.97 How to go from present densely wired to chip-based system

### 0.98 Rent's Rule

In the 1960s, E.F. Rent of IBM found a remarkable trend between the number of pins (terminals, T) at the boundaries of integrated circuit designs at IBM and the number of internal components (g), such as logic gates or standard cells. On a log-log plot, these datapoints were on a straight line, implying a power-law relation  $T = tg^p$  where t and p are constants ( $p < 1.0$ , and generally  $0.5 < p < 0.8$ ). t can be viewed as an average number of terminals required by a single logic block, since  $T = t$ , when  $g = 1$

### 0.99 Rent's Rule should be used for Quantum Circuits

Qubit Control Wiring Overhead \*table Just Qubit Ctrl; not including I/O for readout and tunable coupler \*Google quantum supremacy chip: c. Neill et. Al., Science 360,6385, 195-199 (2018) Additional system integration considerations: -Cost: current cost of a cable channel from room temperature to milliKelvin stage is ~5,000 (per Oxford Instruments). – Typically 2 – 3.5 cables on average per superconducting qubit (transmon) – >10-11M cable cost for a 1000 qubit system. Most expensive cables are between 4K to 20 mK stages

Reliability of a few hundred-channel cabling farm of a dilution refrigerator system is likely be not sustainable

Large I/O Count = scalability problem. Long Wires = latency problem

### 0.100 Present-Day Qubit Control: Analog, Microwave-based

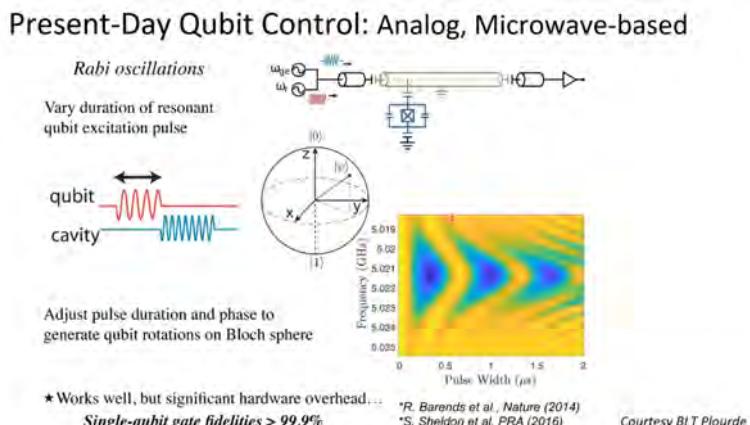


Figure 21:

### 0.101 Present-Day Qubit Readout: Heterodyne Technique

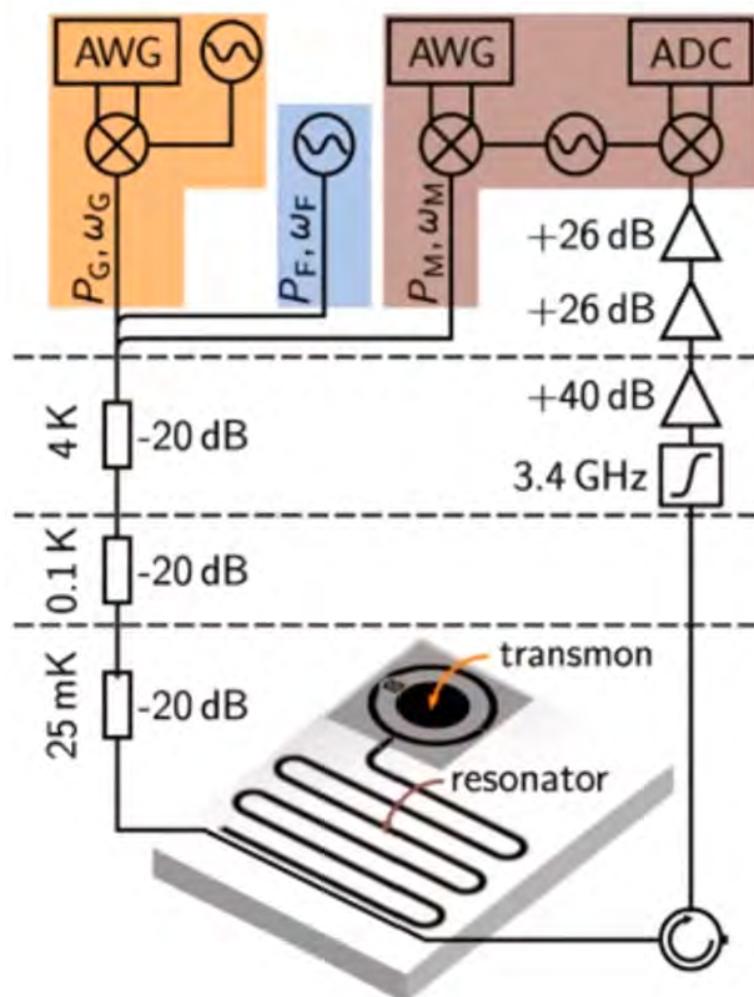
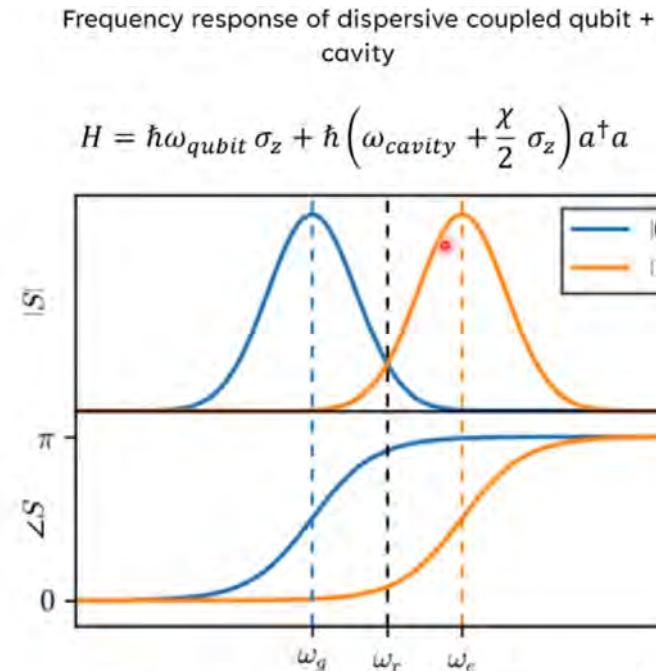


Figure 22:



The information on qubit's state is encoded both in amplitude and phase !

Figure 23:

## 0.102 Increasing Difficulties to Scale

Noise from higher temperatures stages

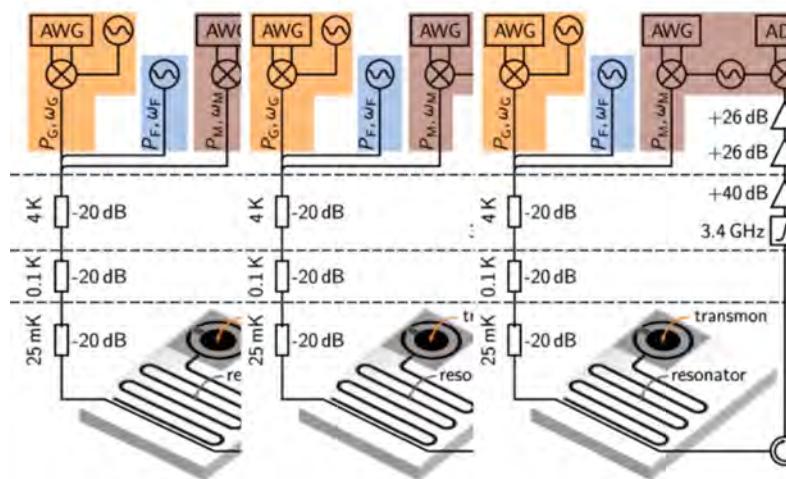


Figure 24:

increased heat load from cable and amplifiers

Crosstalk in multi-qubit system

Increasing complexity of qubit calibration in multiparameter space

Low latency (cycle time delay)

### **0.103 SEEQC digital SFQ chip-based quantum computing architecture**

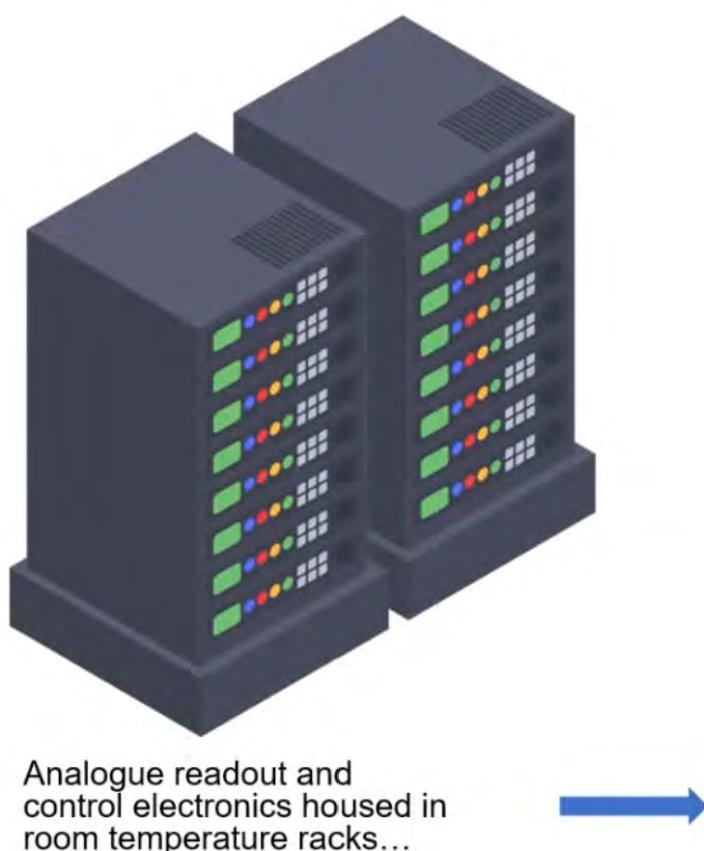
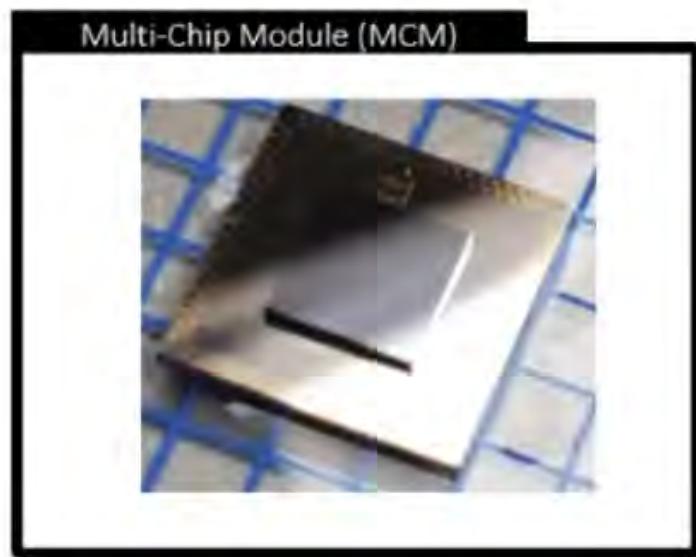
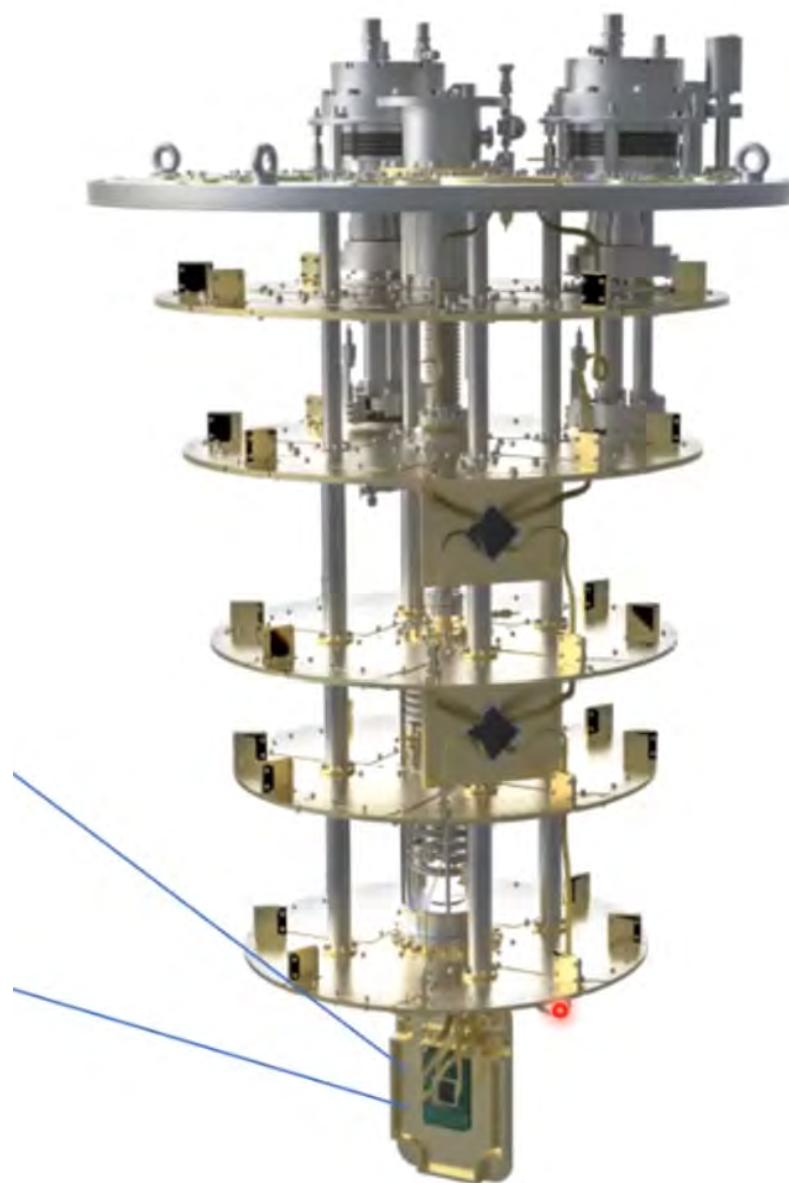


Figure 25:



...are miniaturized and  
engineered as digital SFQ  
chips...

Figure 26:



**...and fit onto a multi-chip module on top of qubits in the DR**

Figure 27:

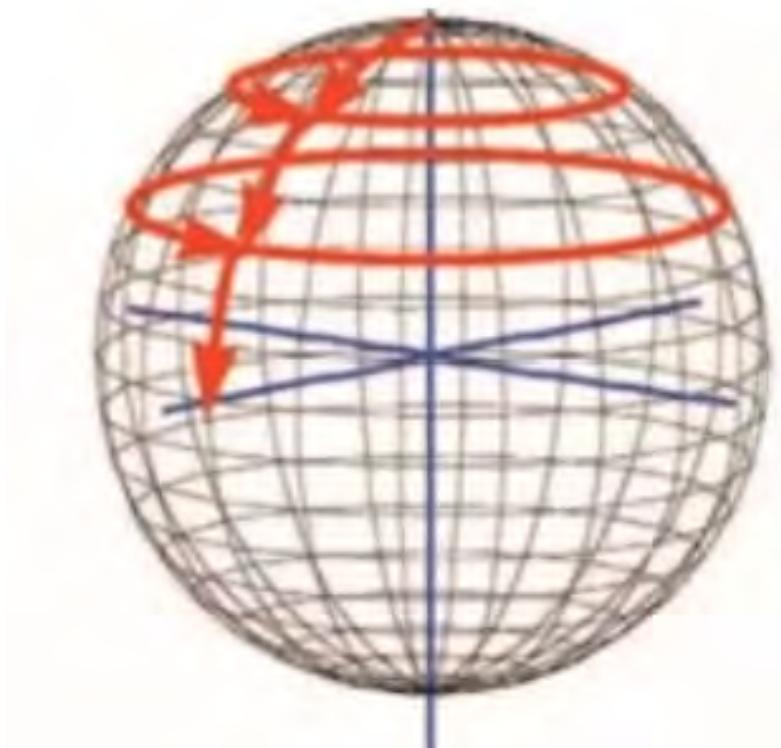


Figure 28:

Analogue readout and control electronics housed in room temperature racks..

... are miniaturized and – engineered as digital SFQ chips...

... and fit onto a multi-chip module on top of qubits in the DR Scaling enabled by chip modularity Scaling enabled by low energy and physical profile Classical co-processing for hybrid quantum-classical algorithms Low cost, low fidelity, low heat leak digital superconducting cables Low latency quantum Error Correction (Pre-decoders) Massive reduction of I/O count 10 nanosecond latency All control signals are low energy SFQ pulses All-digital system All control signals are generated in-cryostat Easily reconfigurable chip based system No need for attenuators, filters, circulators, HEMT amplifiers. Minimized interference, crosstalk No energy dissipated/heat generated in cabling, attenuators compared to conventional QC systems 5 ps signal transfer latency in MCM Ultra-low power superconducting digital control (DQM) circuits

Orders of magnitude lower power than in cryoCMOS

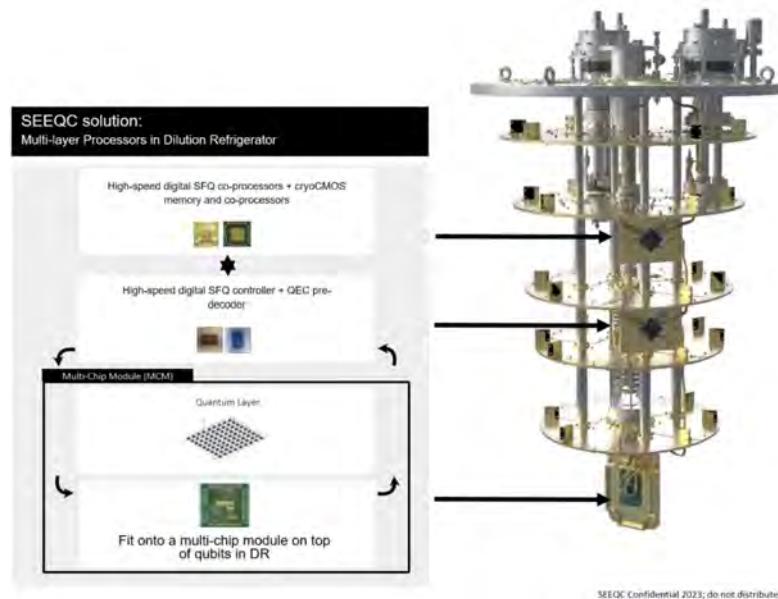


Figure 29:

### 0.104 Resonant Excitation of Qubits by SFQ pulses

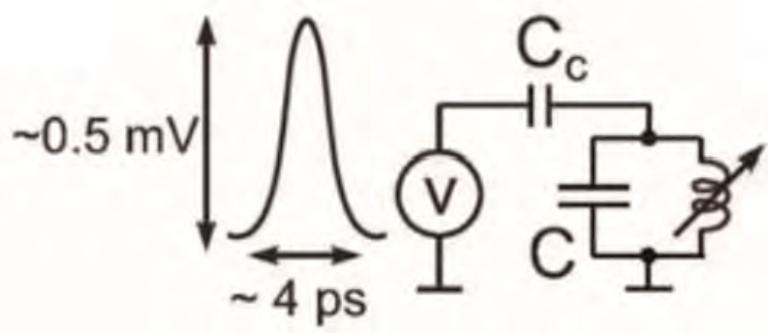


Figure 30:

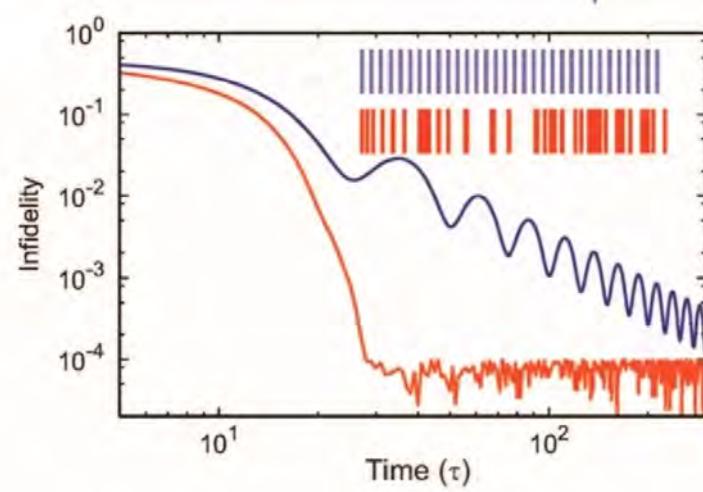


Figure 31:

Replacing microwave with digital SFQ pulse train

Feasibility proven by Wisconsin and Syracuse University groups

By varying distance between SFQ pulses in the train using control theory, one can achieve higher fidelities as shown by Saa rland Univ. team

## 0.105 Using superconducting SFQuClass circuits

Single Flux Quantum Classical circuits fit for a proximal location to qubits

### SFQuClass™ circuit technology:

- Energy-efficient ERSFQ or eSFQ varieties of RSFQ logic
- rescaling JJ critical currents to  $<10\mu\text{A}$  ( $\sim 10^{-20} \text{ J}$  per switching).
- use of high-kinetic inductors.
- qubit-aware SFQ circuit layout

Figure 32:

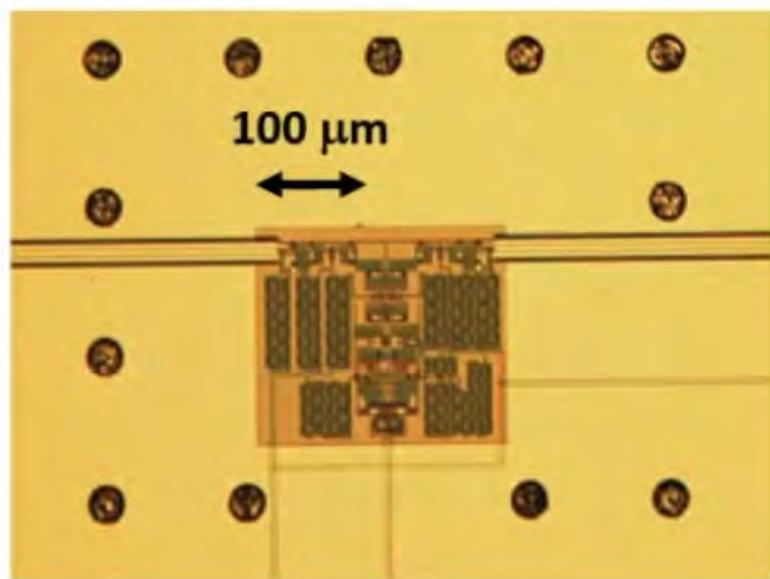


Figure 33:

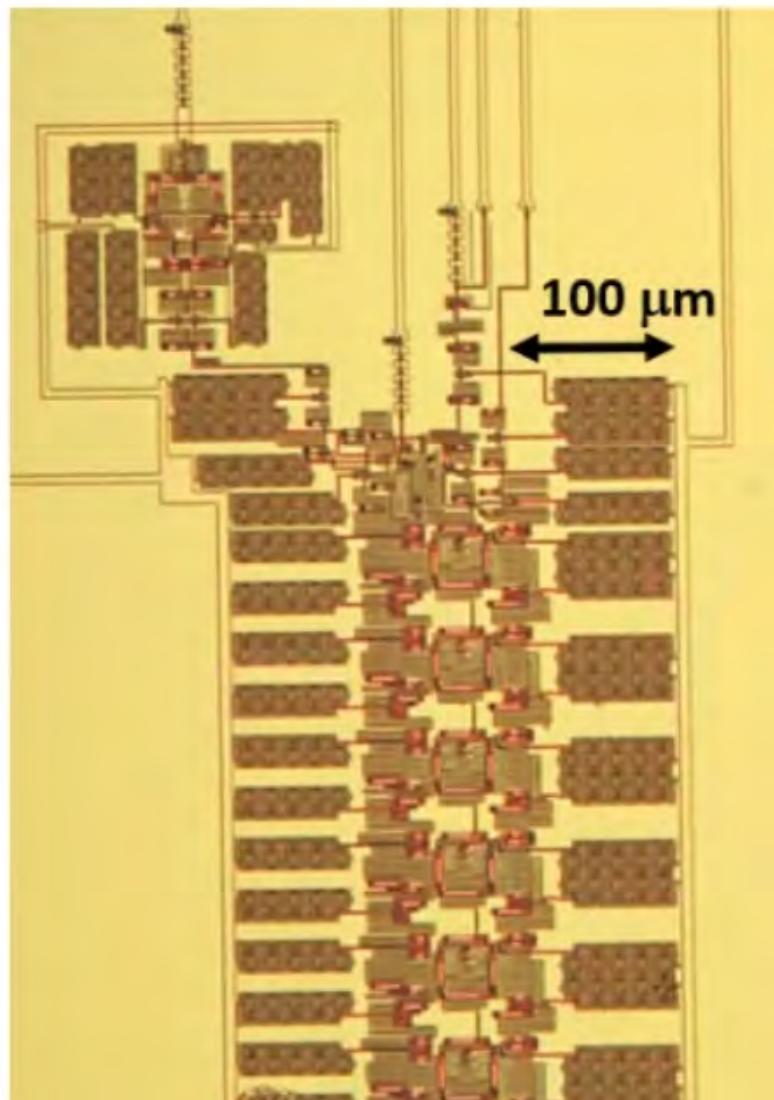


Figure 34:

SFQuClass<sup>TM</sup> circuit technology :

Energy-efficient ERSFQ or eSFQ varieties of RSFQ logic

recalling JJ critical currents to  $\sim 10\text{ }\mu\text{A}$  ( $10^{-20} J$  per switching).

use of high-kinetic inductors.

qubit-aware SFQ circite layout

SFQ Pulser generating SFQ pulse at each cycle of input sinewave clock

Fragment of PGU for storing and generating the optimal SFQ pulse pattern at 25 GHz clock

Fabricated with SEEQC 1000 A/cm<sup>2</sup> critical current density process and a unit Josephson junction critical currents

## 0.106 Flip-chip packaging of fluxonium qubits

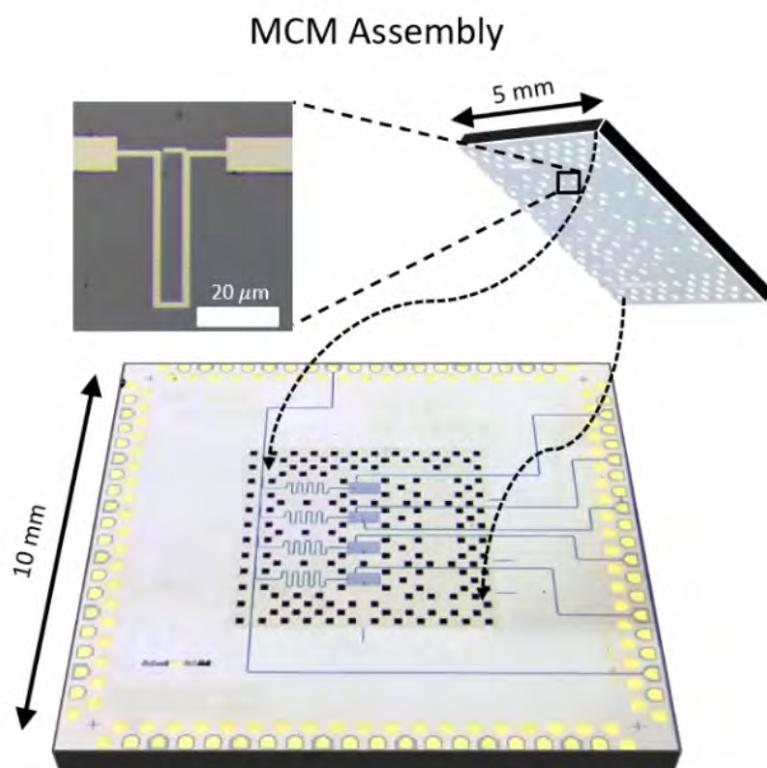


Figure 35:

Fluxonium is packaged in a Multichip Module (MCM) for the first time

Data shows no degradation in qubit performance in this new architecture

paves the way for scaling fluxonium-based quantum processors

0.107 SEEQC digital SFQ controller chip integrated on multi-chip module with qubit chip (@ 20mK)

### 0.107 SEEQC digital SFQ controller chip integrated on multi-chip module with qubit chip (@ 20mK)

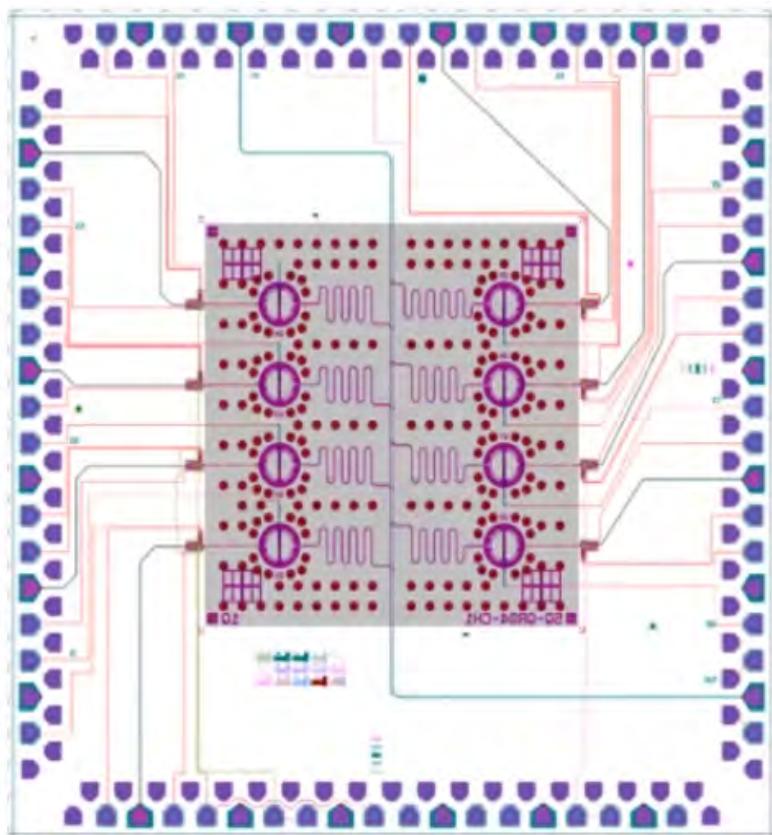


Figure 36:

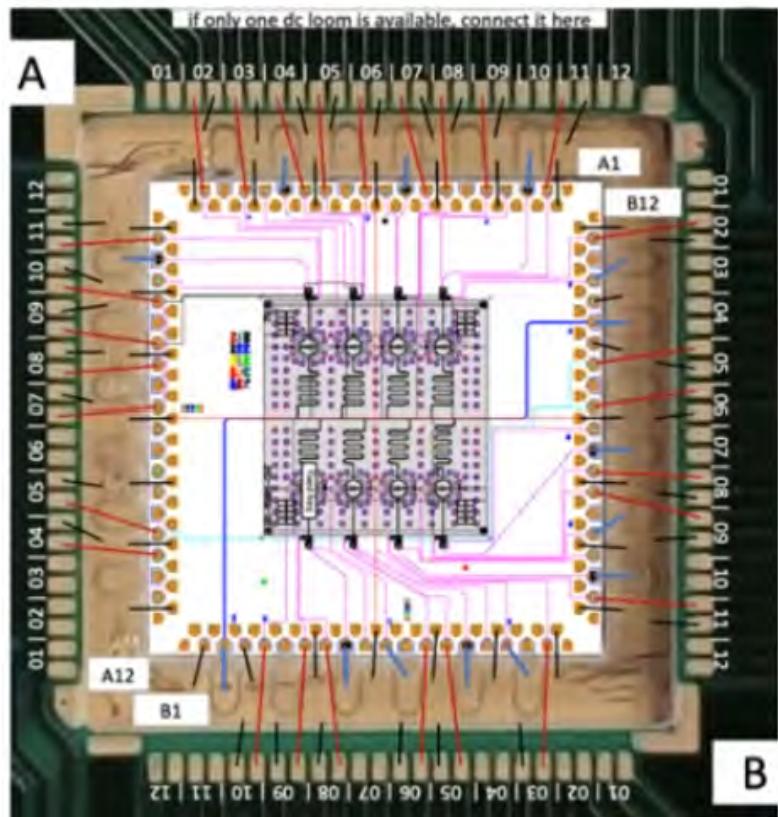


Figure 37:

Design (Virtual) Manufactured and tested (actual)

## 0.108 Digital SFQ control

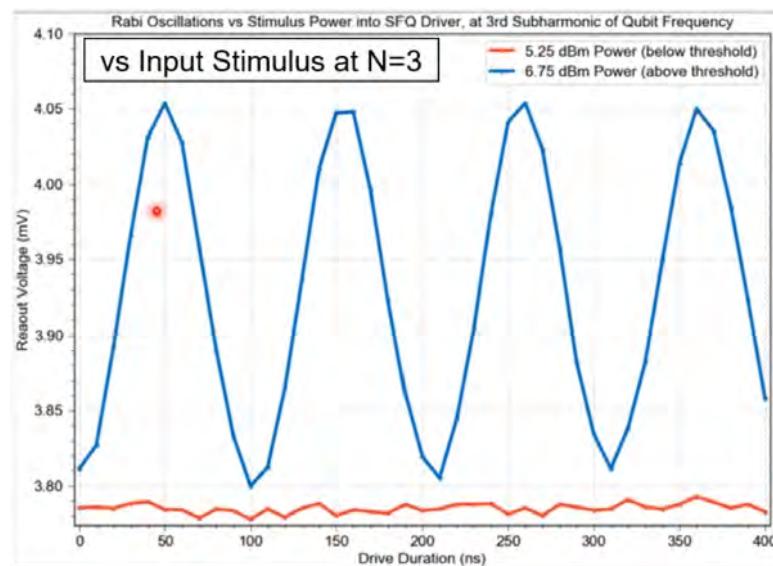


Figure 38:

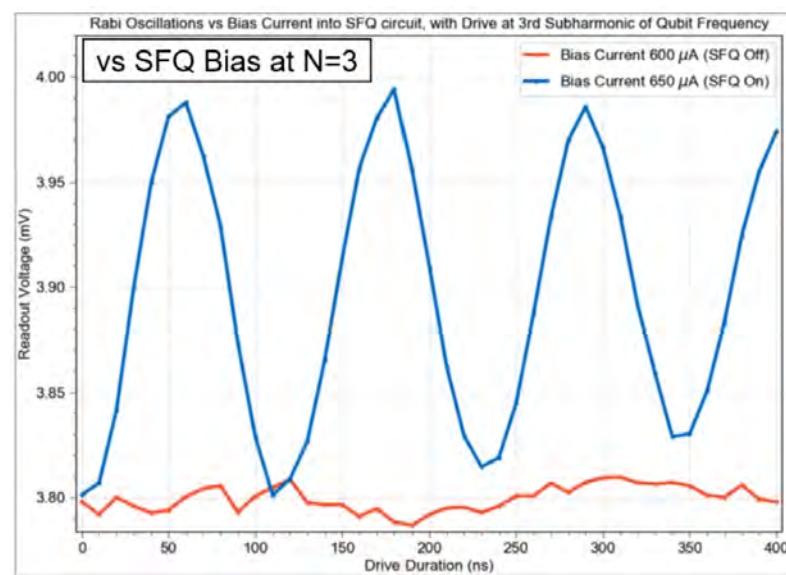


Figure 39:

Rabi oscillation behavior turns on or off depending on...

... whether SFQ circuit is powered or not (current bias... whether input \clock signal into SFQ circuit exceeds operation threshold

### 0.109 Digital SFQ Control of Qubit

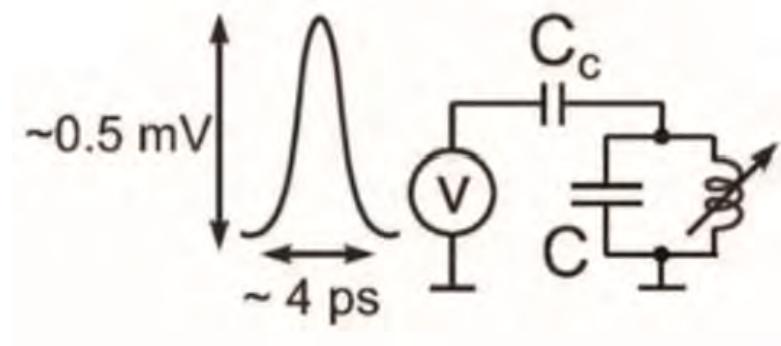


Figure 40:

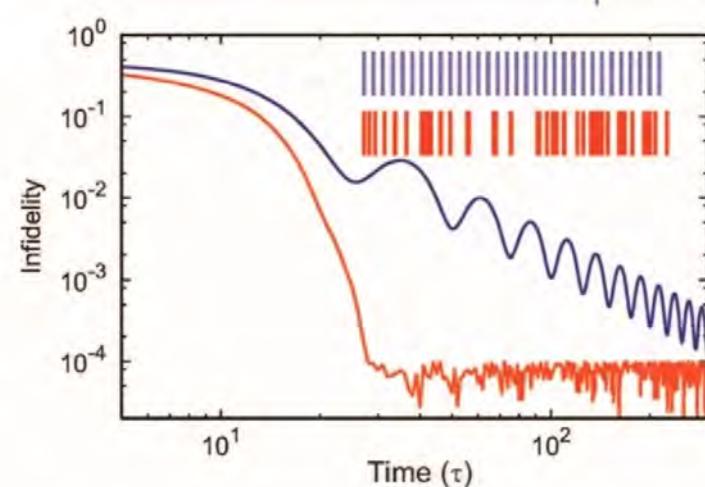


Figure 41:

Qubit shows clear thresholding behavior in Rabi oscillation experiments when the SFQ circuit dc current bias and trigger power are varied proving the SFQ qubit control action

## 0.110 Digital SFQ multiplexing of 1:4 signals

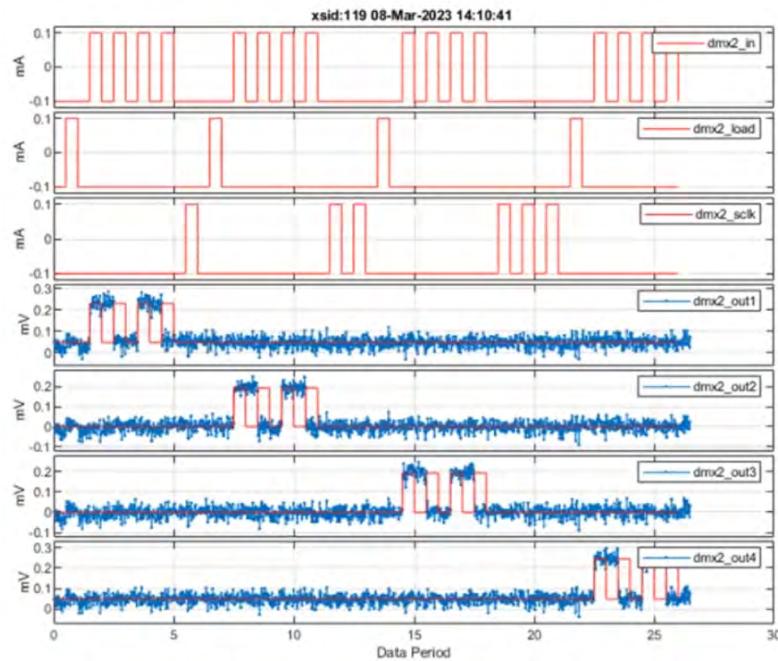


Figure 42:

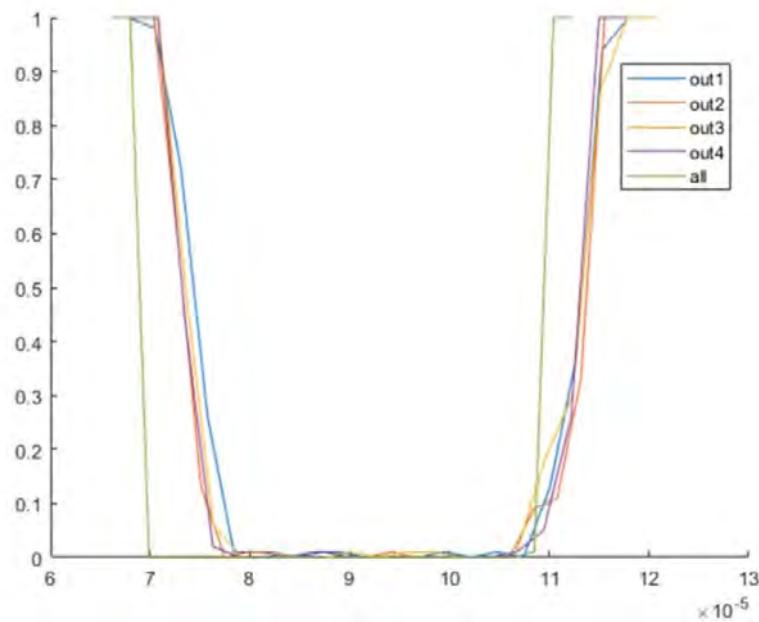


Figure 43:

Testing of MCM with demuced SFQ control is underway  
1:8, 1:16 digital demixes are in fab

### 0.111 Digital Readout: quantum-to-digital converter (QDC)

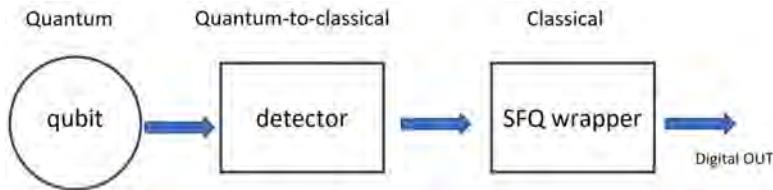


Figure 44:

information Flow

Achieve high fidelity prevent backaction

### 0.112 JD PD – Josephson Digital Phase Detector \*Patent Pending)

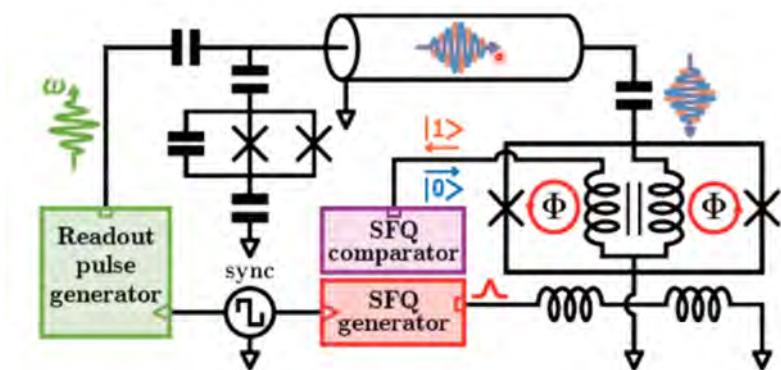


Figure 45:

L. Di Palma, et al. “Discriminating the phase of a coherent tone with a flux switchable superconducting circuit,” Phys. Rev. Applied, accepted for publication.

### 0.113 Digital SFQ readout Josephson Digital Phase Detection (JDPD)

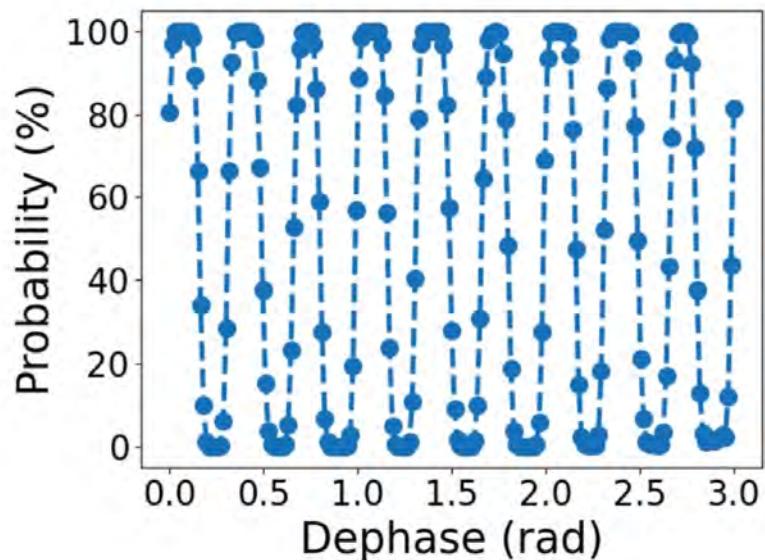


Figure 46:

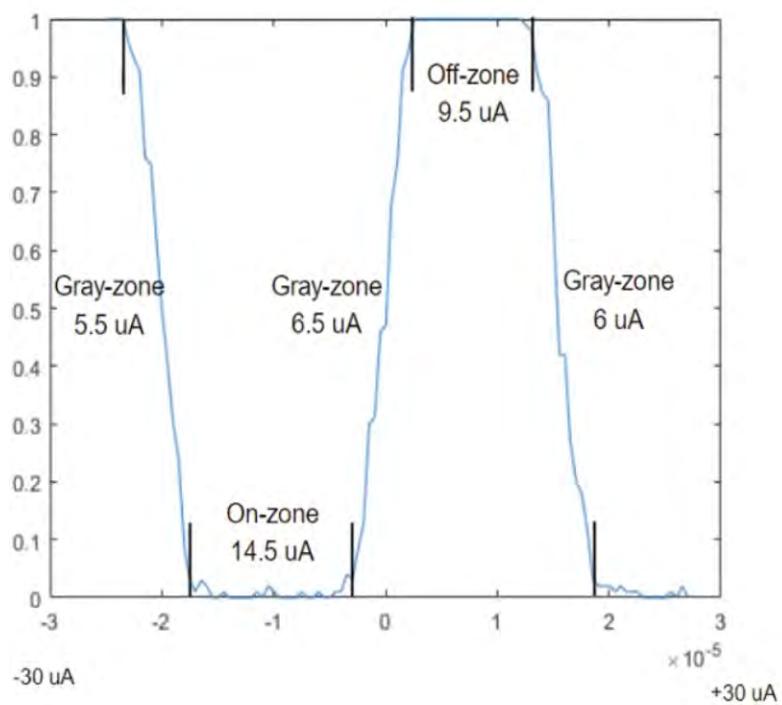


Figure 47:

## 0.114 Rent's Rule for Quantum Circuits cont

*Getting close to microprocessor-like Rent's rule exponent*

Qubit count	Uwaves*	SFQ**
2	4	3
10	20	3
100	200	31
1000	2000	76

\*\*Wires from 4K to 20mK classical SFQ chip  
(classical-to-classical non-coaxial cabling only)

No wires to Quantum Chip (capacitive and magnetic coupling only)

Figure 48:

Qubit count Uwaves\* SFQ\*\* 2 4 3 10 20 3 100 200 31 1000 2000 76 Not including I/O for readout and tunable coupler-only qubit control

\*Google quantum supremacy chip: C. Neil et. Al., Science 360, 6385, 195-199 (2018)

\*\*Wires from 4K to 20nK classical SFQ chip (classical-to-classical non-coaxial cabling only)

No wirers to Quantum Chip (capacitive and magnetic coupling only)

## 0.115 Qubit Readout/Control: Analog vs Digital

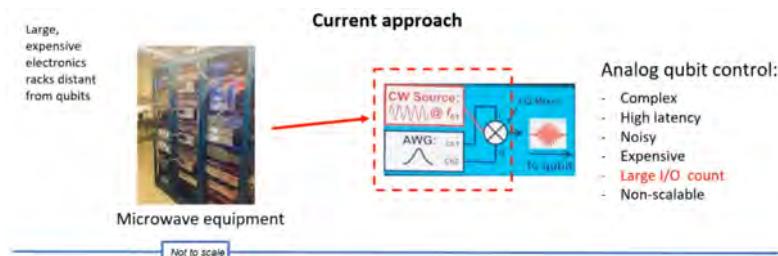


Figure 49:

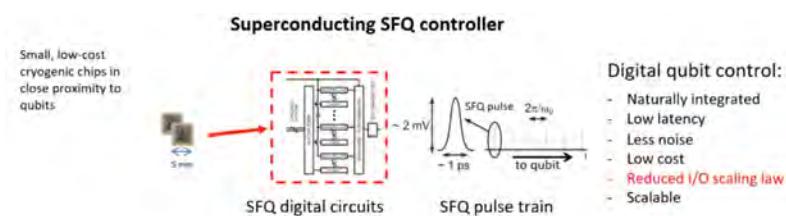


Figure 50:

Current approach Large, expensive electronics racks distant from qubits pic  
Microwave equipment

pic Analog qubit control: -Complex -High latency -Noisy -Expensive -Large I/O count -Non-scalable

Superconducting SFQ controller Small, low-cost cryogenic chips in close proximity to qubits

pic pic pic Digital qubit control: -Naturally integrated -Low latency Less noise -Low cost -Reduced I/O scaling law -Scalable

O.Mukhanov, A Kirichenko, C Howington, J Walkter, M. Hutchings, I. Vermik, D. Yahnmes, K Didge, A Ballard, B. LT, Plourde, A. Opermcak, C.-H. Liu, R. McDermatt, "Scalable Quantum Computing infrastructure Based on Superconducting Electronics," 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019, pp. 31.2.1-31.2.4

## 0.116 Scaling qubit control readout for quantum

	Analogue control/readout	Digital control/readout
Energy Efficiency / Heat dissipation	<ul style="list-style-type: none"> <li>2mW per qubit achieved<sup>1</sup></li> <li>0.2mW possible<sup>2</sup>, fidelity trade-offs required and clock speed</li> </ul>	<ul style="list-style-type: none"> <li>0.01mW per qubit achieved<sup>3</sup></li> <li>0.0002mW possible with high fidelity compatibility</li> </ul>
Speed & Latency	<ul style="list-style-type: none"> <li>10GHz standard clock speed, 10MHz for cryoCMOS to limit head load which limits mo, qubit operations per second and scale 1us qubit reset state-of-the-art analogue latency</li> <li>&gt;100ns cable latency introduced by separate qubit controllers</li> </ul>	<ul style="list-style-type: none"> <li>40GHz clock standard achieves at 4K</li> <li>10ns readout speed, &gt;100ns reset latency achievable</li> <li>&gt;100ns cable latency possible with integrated circuitry</li> </ul>
Multiplexing	<ul style="list-style-type: none"> <li>1:4 RF switching<sup>4</sup>, 1:10 frequency MUX achieved in readout</li> <li>Compromises gate operations &amp; constrains qubit freq.</li> <li>Limits bandwidth and data rates</li> <li>Scale limited by crtoelectronic power dissipation: up to 1000 Q</li> </ul>	<ul style="list-style-type: none"> <li>1:64 digital MUX achievable, Stable, energy efficient cryogenic 1:16 MUX demonstrated<sup>5</sup></li> <li>No compromise to gate operations.</li> <li>High bandwidth and data rates</li> </ul>
Computational complexity	<ul style="list-style-type: none"> <li>14-bit resolution needed</li> <li>16,384 bits of information per qubit control pulse</li> </ul>	<ul style="list-style-type: none"> <li>7-bit resolution needed</li> <li>128 bits of information per qubit control pulse</li> </ul>
Manufacturability	<ul style="list-style-type: none"> <li>Can utilize existing and well established CMOS manufacturing process for components</li> </ul>	<ul style="list-style-type: none"> <li>Utilizes superconducting manufacturing process which is less well established compared to CMOS, but commercial scale</li> </ul>
Cost	<ul style="list-style-type: none"> <li>Expensive RF electronics and high fidelity cables needed</li> <li>RF electronics cost &gt;\$10,000 per qubit today</li> <li>NbTi RF cables cost \$1,000 per channel today</li> <li>Multiplexing maintains need for high fidelity cables</li> </ul>	<ul style="list-style-type: none"> <li>Inexpensive digital chips and low fidelity cables needed.</li> <li>Digital chips cost \$100 per qubit today</li> <li>DC ribbon cables cost \$100 per channel today</li> <li>Multiplexing can utilize low cost low fidelity cables</li> </ul>

Figure 51:

Analogue control/readout Digital control/readout Energy Efficiency/ 2mW per qubit achieved<sup>1</sup> 0.2mW possible<sup>2</sup>, fidelity trade-off required and clock speed 0.01mW per qubit achieved<sup>4</sup>

0.0002mW possible with high fidelity compatibility Speed Latency 1GHz standard clock speed, 10MHz for cryoCMOS to limit head load which limits mo, qubit operations per second and scale 1us qubit reset state-of-the-art analogue latency

>100ns cable latency introduced by separate qubit/controller 40GHz clock speed achieved at 4K 10ns readout speed and >100ns reset latency achievable

>100ns cable latency possible with integrated circuitry Multiplexing 1:4 RF switching<sup>4</sup>, 1 : 10 frequency MUX Achieved in readout

Compromises gate operations constrains qubit freq.

Limits bandwidth and data rates

Scale speed limited by crtoelectronic power dissipation: up to 1000 Q 1:64 digital MUX achievable, Stable, energy efficient cryogenic 1:16 MUX demonstrated<sup>5</sup>

No compromise to gate operations

High bandwidth and data rates Computational complexity 14-bit resolution needed

16,384 bits of information per qubit control pulse 7-bit resolution needed

128 bits of information per qubit control pulse Manufacturability Can utilize existing and well established CMOS manudacting process for components Utilizes cuperconducting manufacturing process which is less well established compared to CMOS, but commercial scale Cost Expensive RF electronics and high fidelity cables needed

RF electronics cost >10,000 per qubit today

NbTi RF cables cost >1,000 per channel today

Multiplexing maintains need for high fidelity cables Inexpensive digital chips and low fidelity cables needed

Digital chips cost 100 per qubit today

DC riddon cables cost 100 per channel today

Multiplexing can utilize low-cost low fidelity cables 1 Intel Horse Ridge 2 and J.C Bardin et al, (ISSCC), 2019

2 Assume 2nm process utilized with associated decreased power consumption

3 CryoCMOS and HEMT based approaches, R Acharyam et. Al, arXiv 2209. 13060 (2022), arCiv, 1409.2202v1 (2014) 4 E Leonard Jr, et al, Phys , Rev, Applied 11, 014008 (2019) 5 Steven B. Japlan and Oleg A. Mukhanov, IEEE Trans, On Appl Spcd, Vol 5, No 2 (1995)

### 0.117 Advantages of SEEQC digital SFQ chips vs analog room temperature quantum control, readout multiplexing

**Confidential: Do not distribute without permission.**

1,000x Lower energy and heat dissipation  $\downarrow$  10x Faster clock speeds + lower latency  $\downarrow$  8-16x Higher multiplexing 128x Lower control pulse complexity Dev. Needed Superconducting manufacture commercial ready, but req, scaling 400x Cheaper system components

### 0.118 SEEQC ship compatibility across quantum technologies

**Confidential: Do not distribute without permission.**

	Control	Readout	Reset	Error decoding & co-processing
Superconducting (Google, IBM, Amazon etc)	✓	✓	✓	✓
Ion Traps (Quantinuum, IonQ etc)	—	✓*	—	✓
Spin (Intel, Quantum Motion, Hitachi etc)	—	✓	✓*	✓
Photonics (PsiQ, ORCA, etc)	—	✓*	—	✓
Cold Atom (Cold Quanta, etc)	—	✓*	—	✓
Topological (Microsoft)	Potentially	✓	Potentially	Potentially

Confidential: Do not distribute without permission

+ SNSPD with SFQ backend

\* SFQ resonator coupled with CryoCMOS circuit

Figure 52:

### 0.119 Competitor 4000 qubit system<sup>1</sup>; cryo only, no room temp electronics SEEQC Confidential 2023: Do not distribute

	Control	Readout	Reset	Error decoding & co-processing
Superconducting (Google, IBM, Amazon etc)	✓	✓	✓	✓
Ion Traps (Quantinuum, IonQ etc)	—	✓*	—	✓
Spin (Intel, Quantum Motion, Hitachi etc)	—	✓	✓*	✓
Photonics (PsiQ, ORCA, etc)	—	✓*	—	✓
Cold Atom (Cold Quanta, etc)	—	✓*	—	✓
Topological (Microsoft)	Potentially	✓	Potentially	Potentially

Confidential: Do not distribute without permission

+ SNSPD with SFQ backend

\* SFQ resonator coupled with CryoCMOS circuit

Figure 53:

Estimated system cost = \$34.2M Estimated energy cost = 130 kW

## 0.120 Competitors 1M qubit quantum data center **Confidential: Do not distribute without permission.**

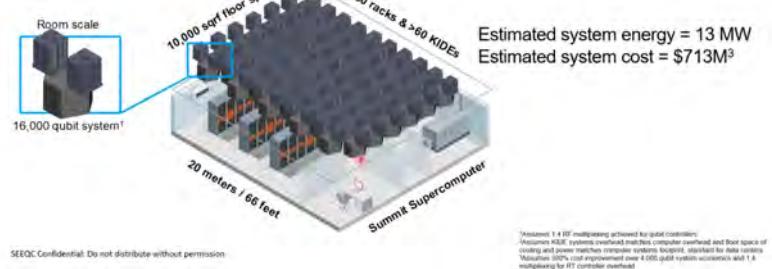


Figure 54:

Estimated system energy = 13MW Estimated system cost = 713M<sup>3</sup> SEEQC scaled quantum computing system: 100k -1m qubits

Estimated total system energy = 63 kW

Estimated cost = 3M

## 0.121 SEEQC scaled quantum computing system: 100k-1m qubits

Estimated total system energy = 63 kW

Estimated cost = 3M

## 0.122 Thank You!

# I.A.A.S: Discover Expedition Education and Outreach

Written by Luka Bostick

©2024 LBTM All rights reserved.

University of North Texas

DISCoVER Expedition Seminar Series

Thanks to Dr. Timothy M. Pinkston, Dr. Murali Annavaram and Dr. Sasan Razmkhah for providing me with a voice and a potential platform to share my input and thoughts on superconducting single flux quantum technology.

Special thanks to the review committee:

Name	Institution	Qualifications	Email
Jonas Wagner	The University of Texas at Dallas — University of Wisconsin-Platteville —	Mechanical Engineering PhD Student — Engineering Physics and Electrical Engineering Bs	jonas.wagner.2826@gmail.com
Storm McCauley	The University of North Texas	Computer Science Bs student	storm.mccauley@gmail.com
Pavlo Bondarenko	The University of North Texas — Odessa National Polytechnic University	Computer Science Bs student— Computer Engineering Bs	pavlobondarenko@my.unt.edu
Ethan Jensen	The University of North Texas — George Fox University	Mathematics PhD Student — Mathematics and Computer Science Bs	ethanjensen@my.unt.edu
Dr. Jacob Hochstetler	The University of North Texas — Fidelity Investments	Computer Science and Engineering PhD — VP, Cloud Engineering at Fidelity Investments	Jacob.Hochstetler@unt.edu
Dr. Austen Moss	The University of North Texas	Organic Chemistry PhD	austen.moss@unt.edu
Willy Vasquez	The University of Texas at Austin —Massachusetts Institute of Technology	Cryptography PhD Student — Computer Science Ms	wrv@utexas.edu
Hannah Pil	North Carolina State University	B.S. in Genetics PhD in Genetics	hdpl@ncsu.edu

Table 1:

## 0.123 Abstract

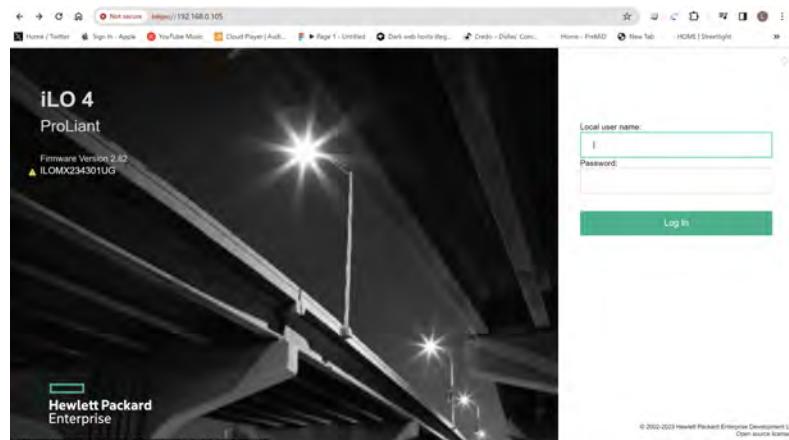


Figure 1:

Figure 2:

Figure 3:

## I.A.A.S: Discover Expedition Education and Outreach

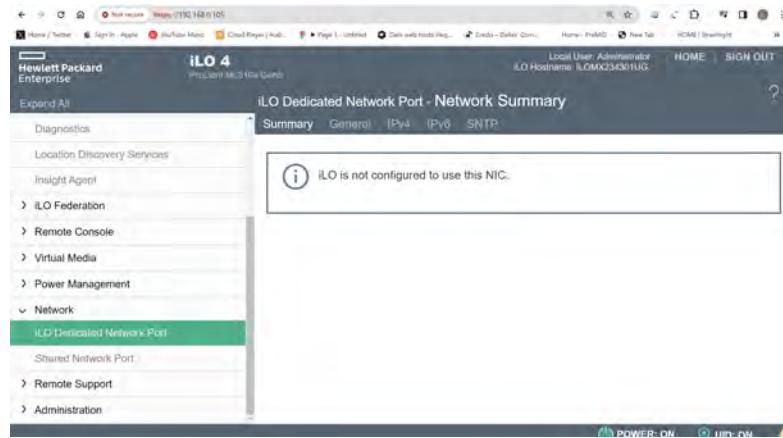


Figure 4:

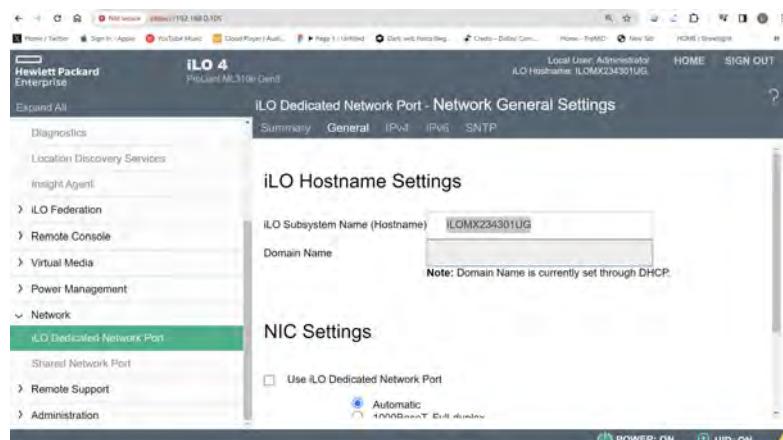


Figure 5:

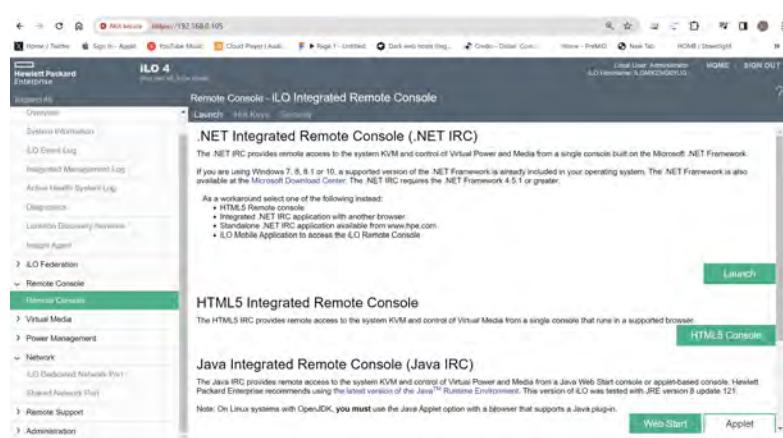


Figure 6:

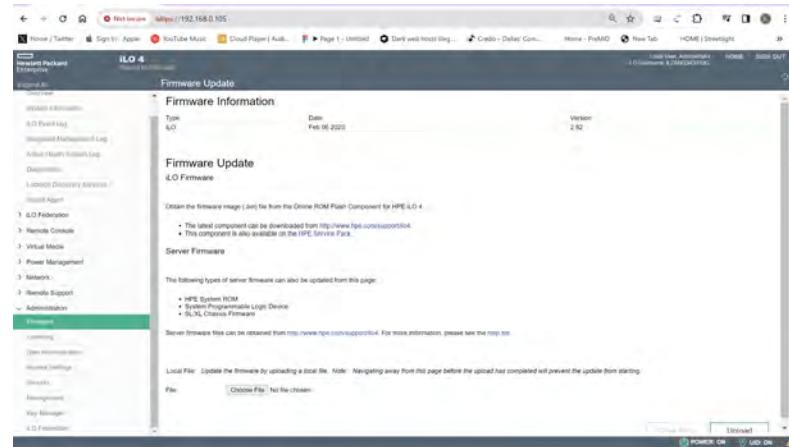


Figure 7:



Figure 8:

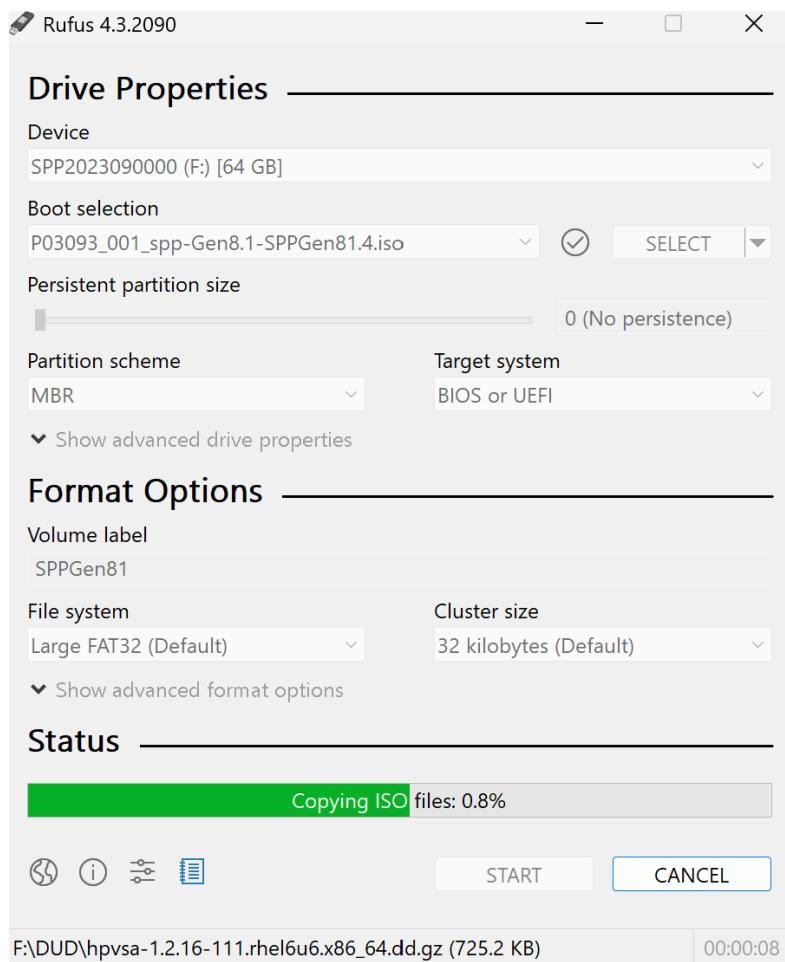


Figure 9:

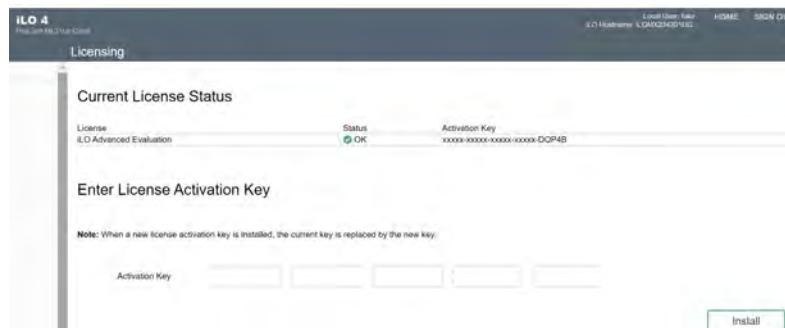


Figure 10:

## Status

System Health	<span style="color: yellow;">⚠️</span> Degraded
iLO Health	<span style="color: green;">✓</span> OK
Server Power	<span style="color: green;">●</span> ON
UID Indicator	<span style="color: green;">💡</span> UID OFF
TPM Status	Not Present
SD-Card Status	Not Present
iLO Date/Time	Sat Jan 13 10:56:55 2024

## Connection to HPE

⚠️ Not registered

Figure 11:

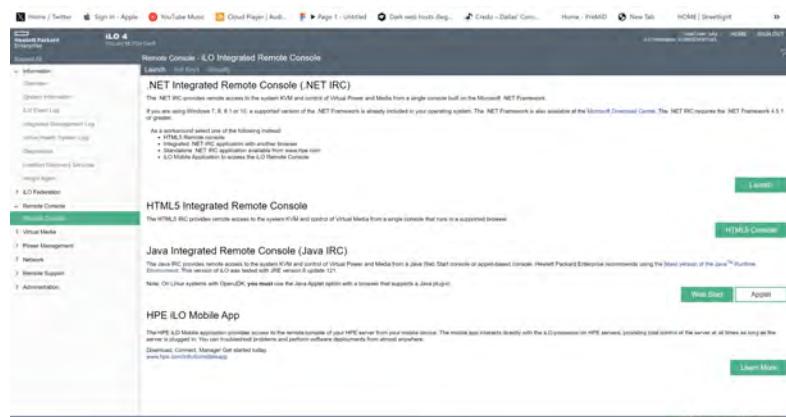


Figure 12:

## I.A.A.S: Discover Expedition Education and Outreach

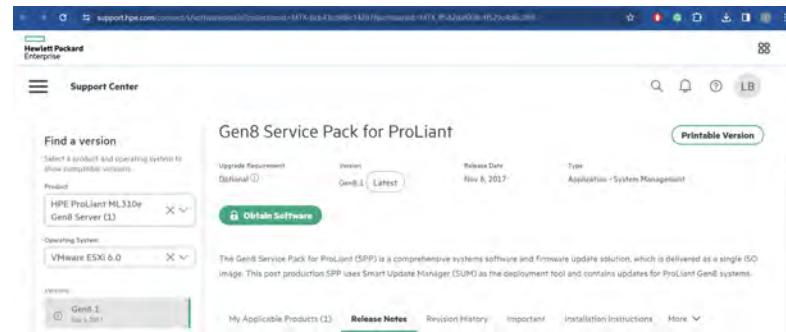


Figure 13:

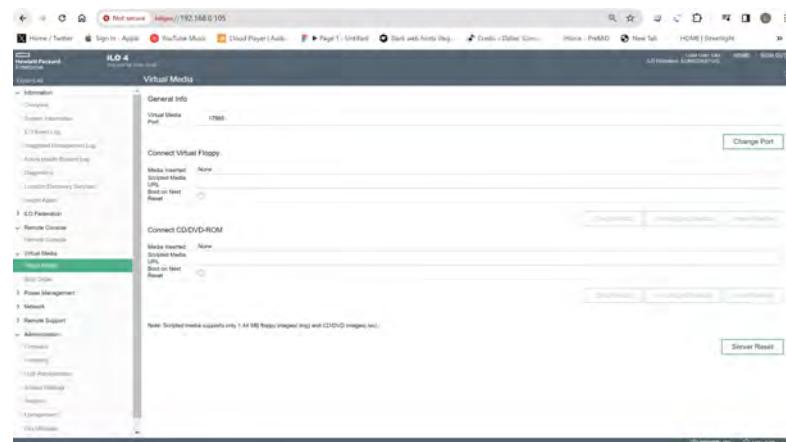


Figure 14:

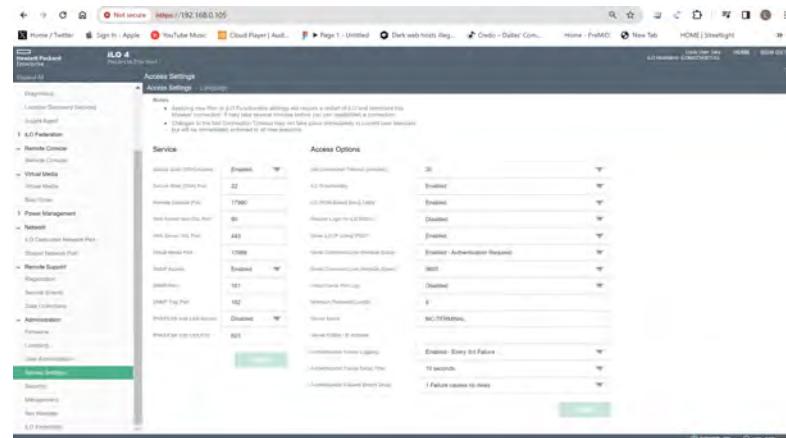


Figure 15:



Figure 16:

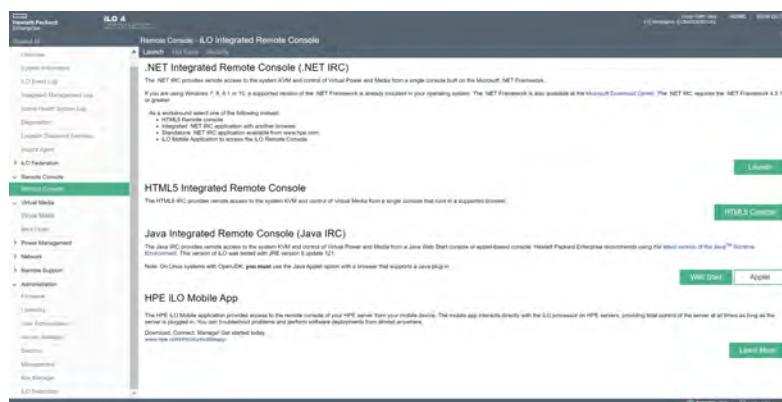


Figure 17:

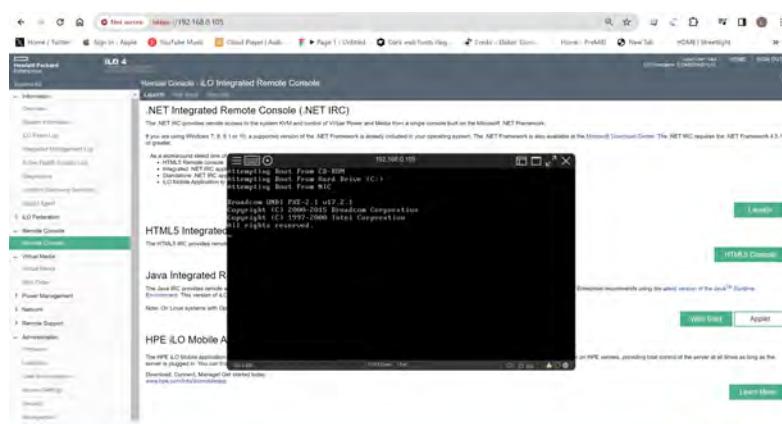


Figure 18:

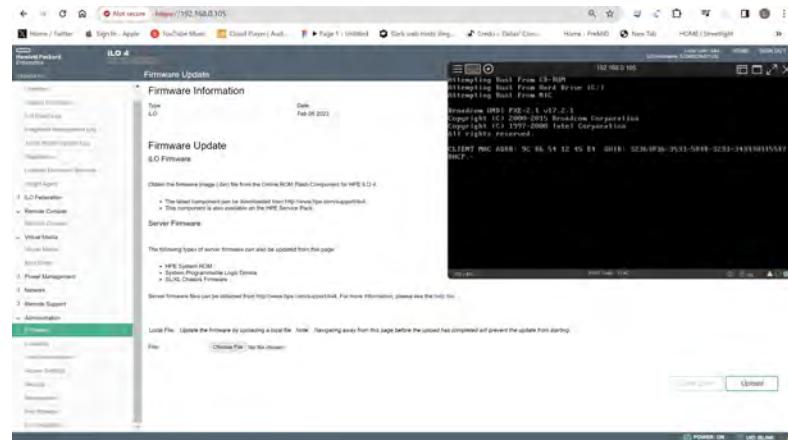


Figure 19:



Figure 20:



Figure 21:

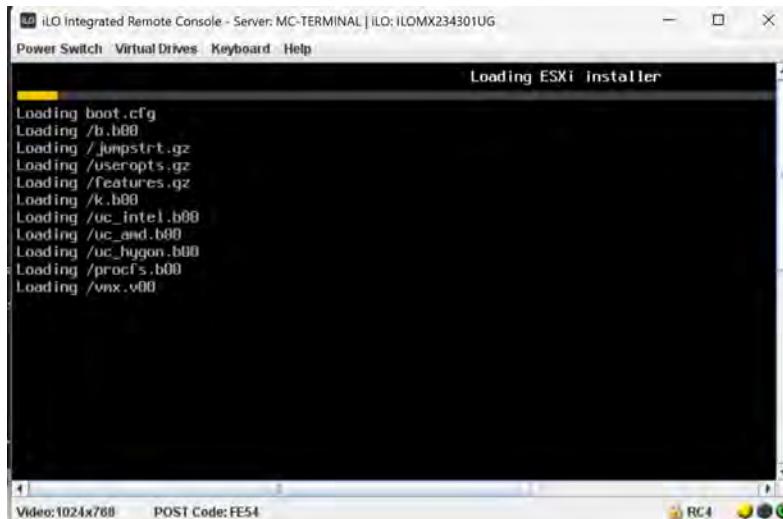


Figure 22:

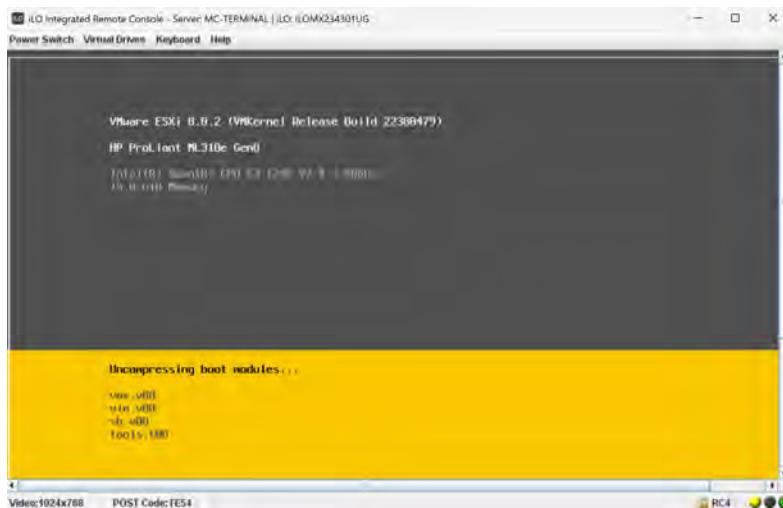


Figure 23:



Figure 24:



Figure 25:



Figure 26:



Figure 27:

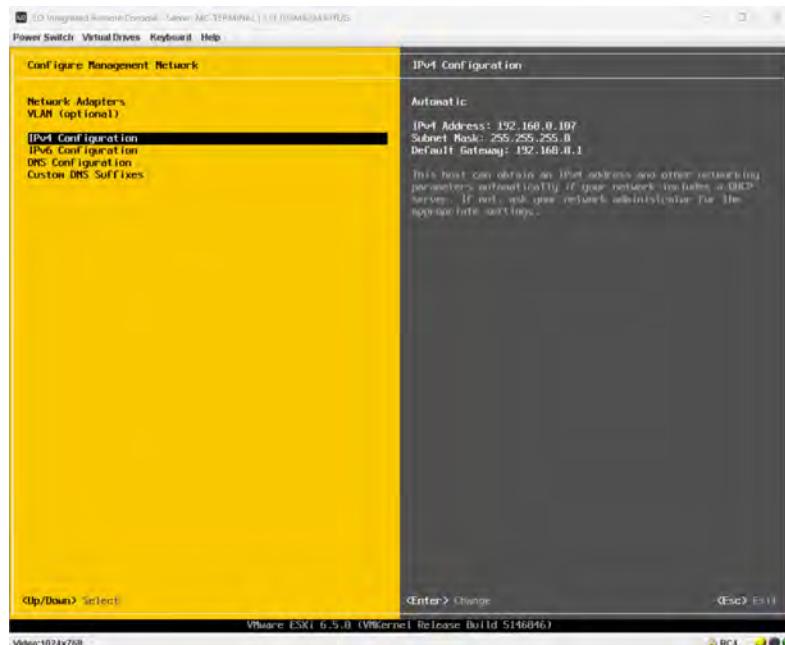


Figure 28:

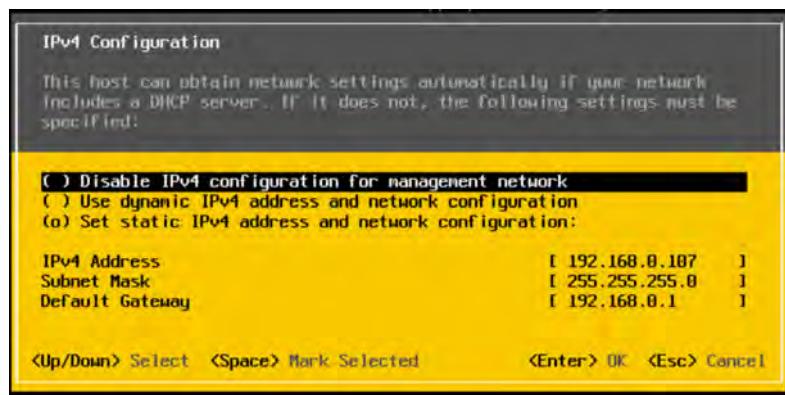


Figure 29:

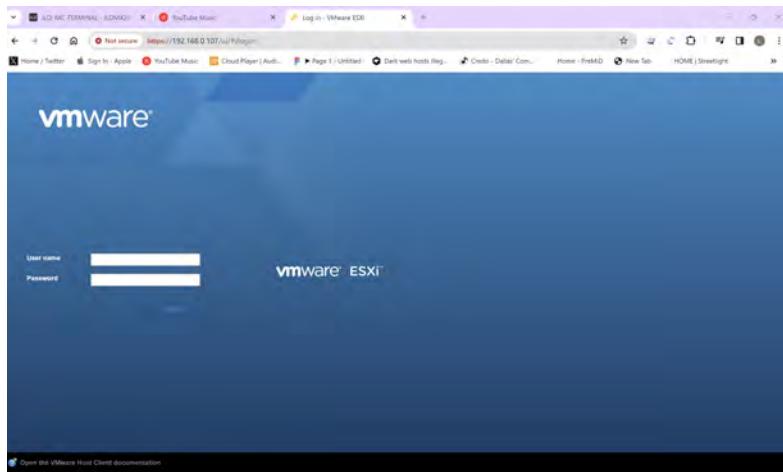


Figure 30:

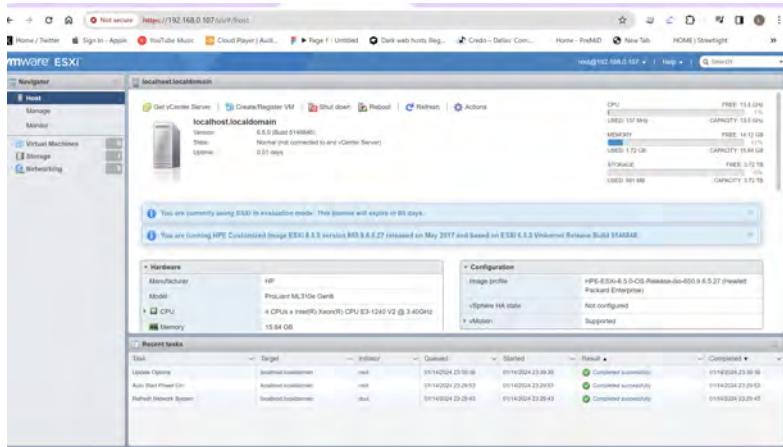


Figure 31:

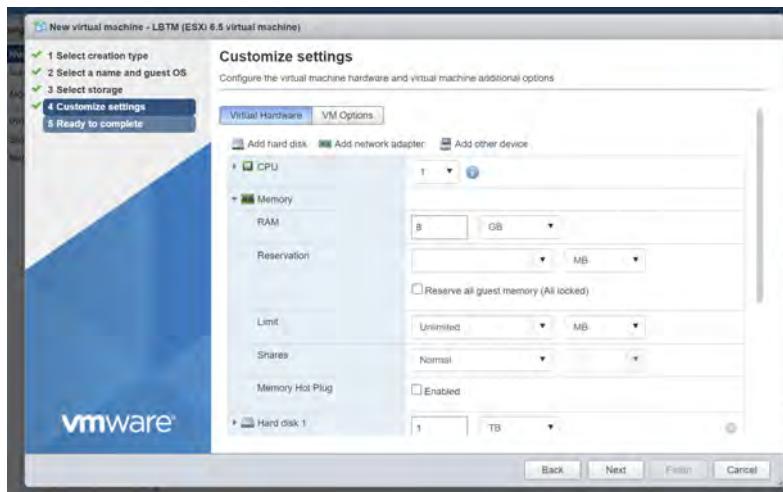


Figure 32:

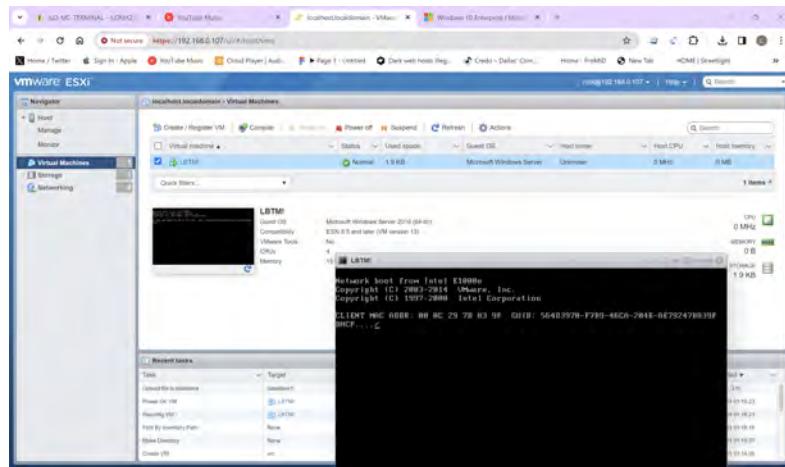


Figure 33:

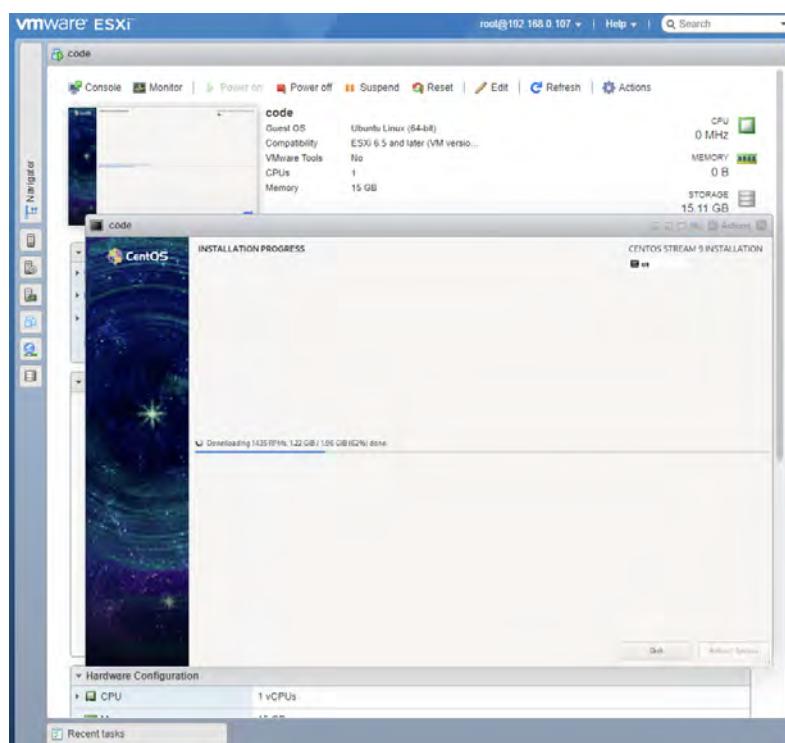


Figure 34:

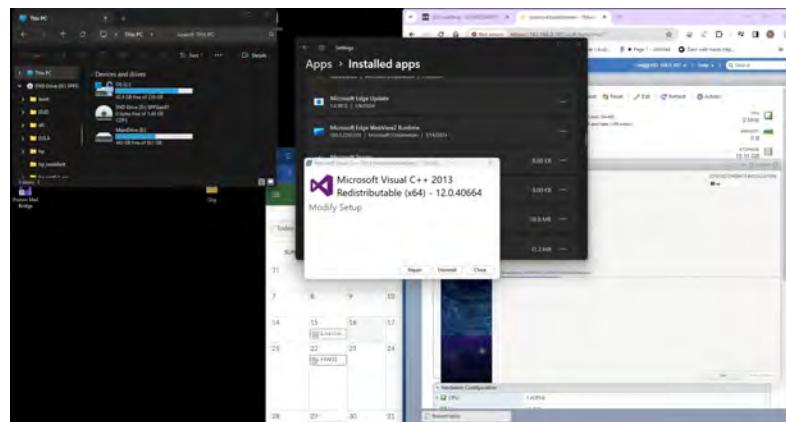


Figure 35:

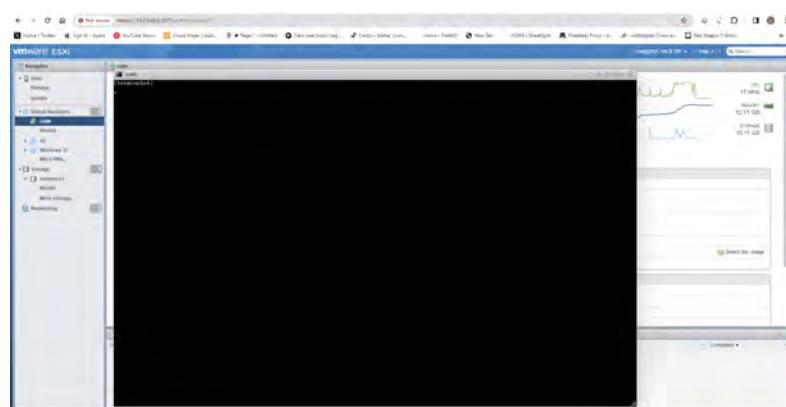


Figure 36:

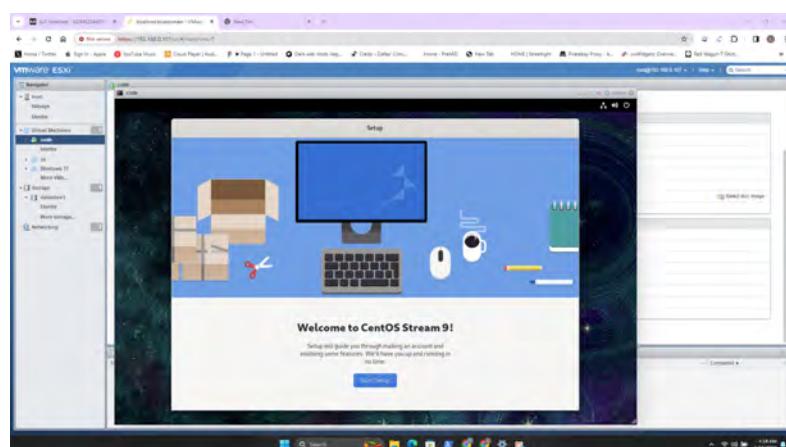


Figure 37:

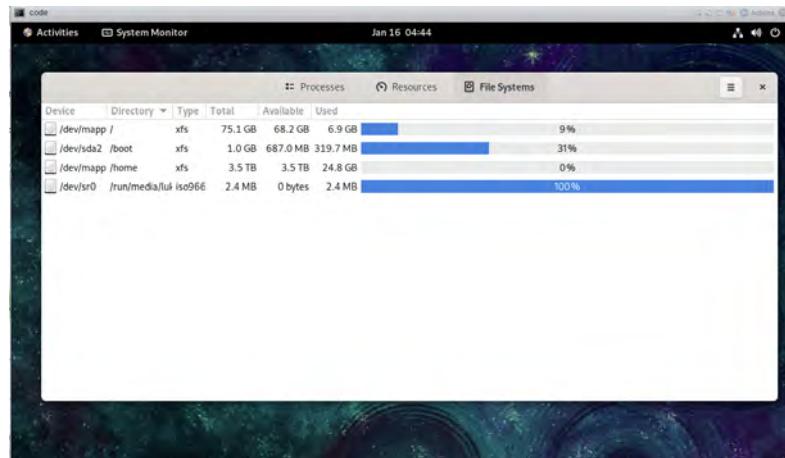


Figure 38:

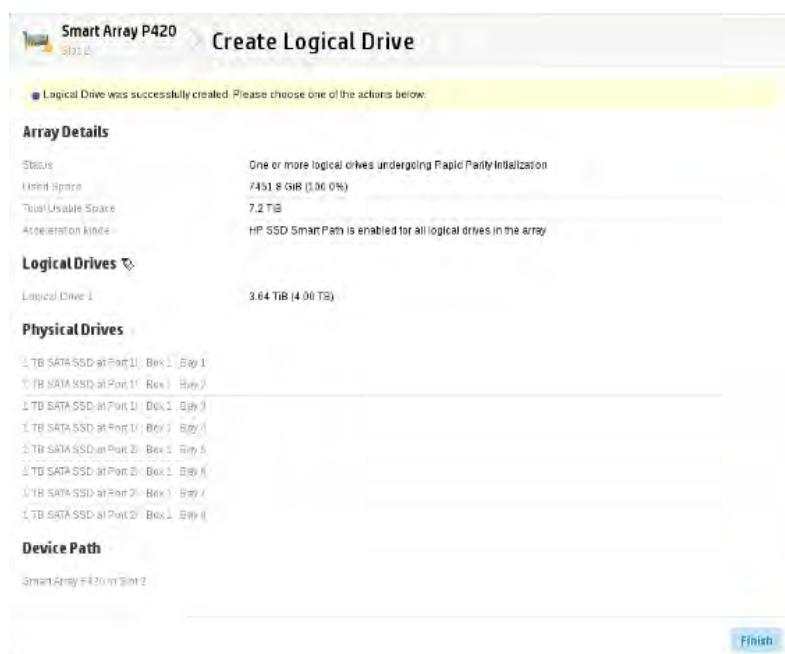


Figure 39:

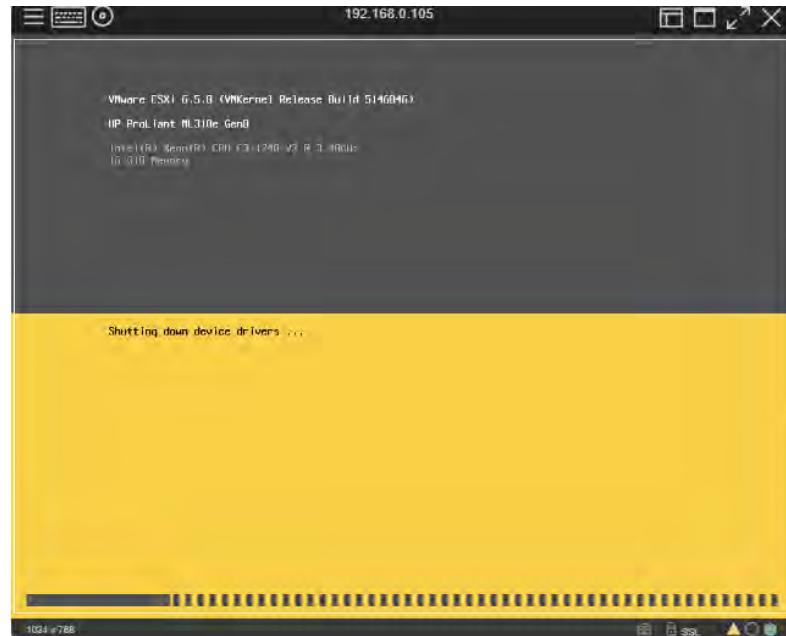


Figure 40:

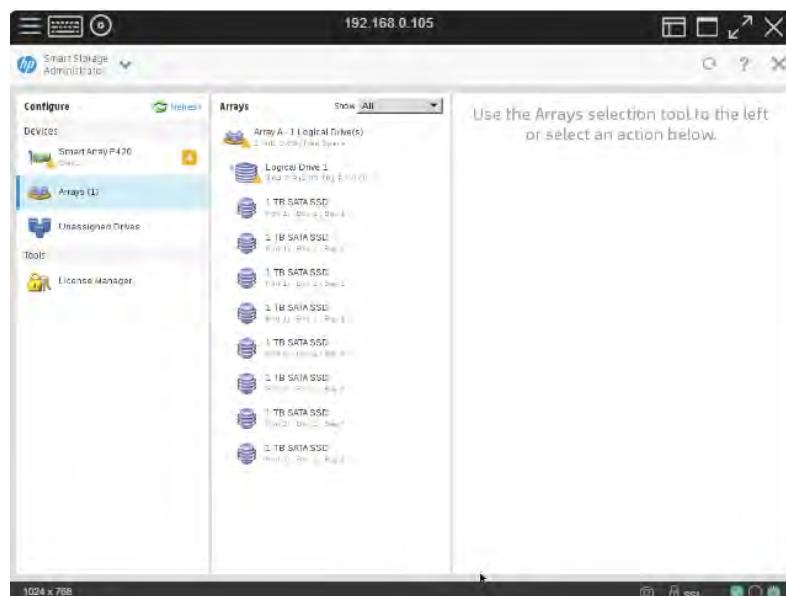


Figure 41:

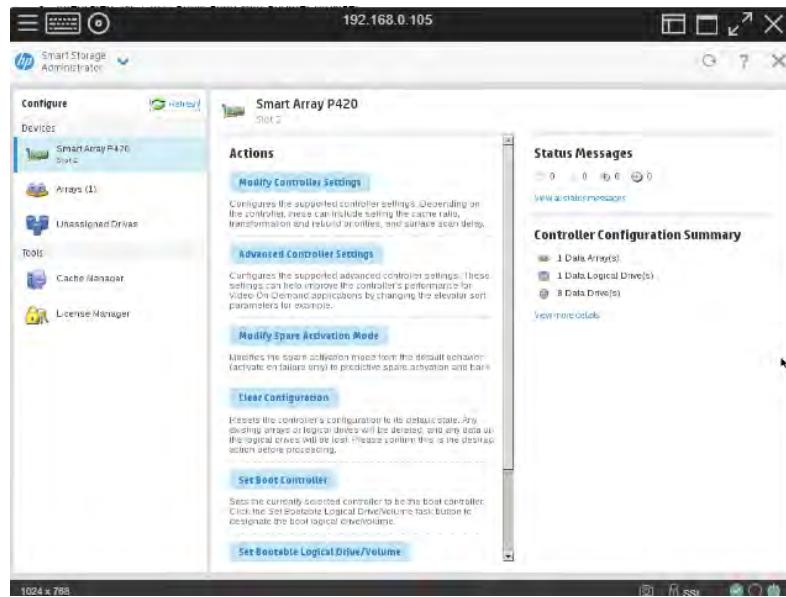


Figure 42:

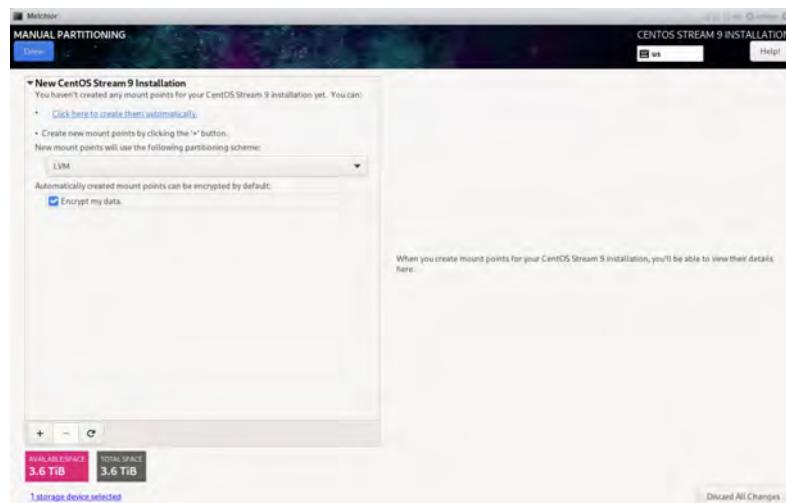


Figure 43: