

Fundamentals of Technological Processes

Martina Paola Calvi

December 2, 2025

Contents

1 Part 1	6
1.1 Wafer Production and Monocrystals Growth	6
1.1.1 Monocrystals growth for miniaturised devices technology	6
1.1.2 Monocrystals growth	7
1.1.3 Electronic-Grade Silicon (EGS) Production	8
1.1.4 Czochralski Process (CZ)	9
1.1.5 Ingot Characteristics and Parts	9
1.1.6 Doping from the Melt	10
1.1.7 III-V Monocrystals Growth: Liquid Encapsulated CZ (LEC)	10
1.1.8 From Boule to Wafers	11
1.1.9 Wafer Evolution and Advantages of Larger Diameters	12
1.2 Crystallography and crystalline defects	14
1.2.1 3 types of solids	14
1.2.2 Crystal structure	14
1.2.3 Silicon lattice (Diamond Structure)	16
1.2.4 Miller indices	17
1.2.5 Crystalline defects	18
1.3 Cleanroom Technology	20
1.3.1 Introduction and Rationale for Cleanroom Technology	20
1.3.2 Contamination Induced Problems	20
1.3.3 Sources of Contamination	20
1.3.4 Classification Standards and Measurement	21
1.3.5 Environmental Control and Airflow	22
1.3.6 Cleanroom Air Filters	23
1.3.7 Cleanroom Clothing (Gowning)	23
1.3.8 Deionized Water (DI Water)	24
1.4 Introduction to Lithography	25
1.4.1 Micro/Nano Technologies Overview	25
1.4.2 Photolithography Function	25
1.4.3 Masks and Alignment	25
1.4.4 The Photolithography Process Flow	26
1.4.5 Photoresist Components and Properties	27
1.4.6 Back to lithographic process: Exposure Sources and Tools	29
1.4.7 Post-Exposure Processing and Pattern Transfer	32

1.4.8	Developing	33
1.4.9	Hard Bake	33
1.4.10	Pattern Transfer Methods	34
1.4.11	Models and Simulation	34
1.5	Silicon Oxidation	35
1.5.1	Properties of Silicon Dioxide	35
1.5.2	How does Silicon Oxide?	36
1.5.3	Substrate Orientation + Mono/Poli comparison	36
1.5.4	Key variables in Oxidation	36
1.5.5	Modelin Thermal Oxidation of Si	37
1.5.6	Oxidation Graphs	40
1.5.7	Growth rate regimes	41
1.5.8	Rapid thermal Oxidation (RTO)	42
1.6	Epitaxy, CVD, ALD, PVD, electroplating	43
1.6.1	Conditions and Types	43
1.6.2	Classification by Mother Phase	43
1.6.3	Why is Epitaxy important?	43
1.6.4	Vapor Phase Epitaxy (VPE) / Chemical Vapor Deposition (CVD)	44
1.6.5	Liquid Phase Epitaxy (LPE)	45
1.6.6	Molecular Beam Epitaxy (MBE)	46
1.6.7	Chemical Vapor Deposition (CVD)	47
1.6.8	Types of Chemical Vapor Deposition (CVD) Reactors	47
1.6.9	PECVD process control parameters	50
1.6.10	Atomic Layer Deposition(ALD)	50
1.6.11	Silicon Dioxide (CVD)	51
1.6.12	Polycrystalline Silicon ("Poly")	53
1.6.13	PVD (Physical Vapor Deposition)	53
1.6.14	PVD-Thermal Evaporation	53
1.6.15	PVD-Thermal Evaporation	54
1.6.16	PVD- Sputtering	56
1.6.17	Electroplating – Electrodeposition-Galvanic Deposition	59
1.7	Implantation	61
1.7.1	Silicon Doping	61
1.7.2	Dopant Species: P-Type, N-Type	61
1.7.3	Piezoresistive Pressure Sensor	62
1.7.4	Doping Parameters	62
1.7.5	Ion Implantation	62
1.7.6	Gaussian approximation	64
1.7.7	Implantation Control Parameters	65
1.7.8	Junction Depth (As Implanted)	66
1.7.9	Annealing - Activation	66
1.7.10	Channeling Ions Through Lattice	67
1.7.11	Sheet (or Square) Resistance	68
1.7.12	Implanting Through a Mask	68
1.7.13	Masked Implantation	69

1.7.14	Screen Oxides	69
1.7.15	Silicon on insulators (SOI) wafers	70
1.7.16	SIMOX	70
1.7.17	BESOI	71
1.7.18	Smart Cut	71
1.8	Wet & Dry Etching	73
1.8.1	Etching	73
1.8.2	Wet Etching	73
1.8.3	Etchant properties	74
1.8.4	Wet etching silicon ("HNA")	75
1.8.5	Anisotropically wet etching Silicon	76
1.8.6	HYDROFLUORIC ACID (SiO_2)	77
1.8.7	What is a Dry Etch?	78
1.8.8	Vapor etching	78
1.8.9	RF plasma-based dry etching	79
1.8.10	TODO:Types of Dry Etching	80
1.8.11	DRY ETCH CHEMISTRIES	81
1.8.12	TODO:Selectivity	82
1.8.13	TODO:Anisotropy in Dry Etching	82
1.8.14	Deep reactive ion etching (DRIE)→Time multiplexed deep etching	83
1.9	Packaging	84
1.9.1	Packaging Process Flow	84
1.9.2	Steps in Packaging	84
1.9.3	Functions of Packaging	85
1.9.4	Package Types	85
1.9.5	Advanced Interconnection and Integration	86
1.10	Complementary Technologies	89
1.10.1	Bulk micromachining	89
1.10.2	SPM Machining	90
2	Part 2: Cryogenic technology and Superconductivity	92
2.1	Introduction	92
2.1.1	Elements of classical thermodynamics	92
2.1.2	Thermodynamic systems	93
2.1.3	Thermodynamic transformations	94
2.1.4	The Laws of Thermodynamics	94
2.1.5	Thermal equilibrium	94
2.1.6	Zeroth Law – Temperature	94
2.1.7	First Law of Thermodynamics	95
2.1.8	Second Law of Thermodynamics	96
2.1.9	Third Law of Thermodynamics	96
2.1.10	Entropy	96
2.1.11	Principle of Increase of Entropy	98
2.1.12	Enthalpy	100
2.2	The Thermodynamics of Refrigeration	103

2.2.1	Cryogenics	103
2.2.2	Historical development of refrigeration	104
2.2.3	Isentropic cooling	105
2.2.4	Use of a throttling process	105
2.2.5	Refrigerants	107
2.2.6	The throttling process	108
2.2.7	Enthalpy and throttling	109
2.2.8	Joule-Thomson expansion	109
2.2.9	Liquefaction of gases by the Joule-Thomson expansion . .	111
2.2.10	Coolers using turbo-expanders	113
2.2.11	Countercurrent heat exchangers – calculating desirable dimensions	114
2.3	Refrigeration machines	117
2.3.1	Coolers using regenerative heat exchangers	117
2.3.2	Stirling refrigerator cycle	117
2.3.3	Stirling cryocoolers	117
2.3.4	Gifford-McMahon cryocoolers	120
2.3.5	Pulse tube refrigerators	120
2.4	Properties of cryogenic fluids	122
2.4.1	Properties of Cryoliquids	122
2.4.2	Liquid Air, Liquid Oxygen, Liquid Nitrogen	122
2.4.3	Liquid Helium: <i>Transport Properties of Liquid ^4He</i>	127
2.5	Thermal Contact and Thermal Isolation	129
2.5.1	Heat transfer	129
2.5.2	Radiation	132
2.5.3	Other causes of heat transfer	133
2.5.4	Selection of the Material with the Appropriate Cryogenic Thermal Conductivity	136
2.5.5	Heat Switches	137
2.5.6	Thermal Boundary Resistance	138
2.6	^4He Cryostats and Cryogenic Technology	141
2.6.1	Key Technological Requirements and Challenges	141
2.6.2	Use of Liquid ^4He in Low-Temperature Equipment	141
2.6.3	Cryostat Types	142
2.6.4	Cryostats for Variable Low Temperatures ($1.3 \text{ K} \leq T \leq 4.2 \text{ K}$)	146
2.6.5	Auxiliary Equipment	147
2.7	^3He cryostats: refrigeration below 1K	150
2.7.1	Advantages and Disadvantages of ^3He Refrigeration	150
2.7.2	Types of ^3He Cryostats	150
2.8	Dilution Refrigerator	155
2.8.1	The $^3\text{He} - ^4\text{He}$ Dilution Refrigerator	155
2.8.2	Finite Solubility of ^3He in ^4He	157
2.8.3	Cooling Power of the Dilution Process	158
2.8.4	Osmotic Pressure	159
2.8.5	Realization of a $^3\text{He} - ^4\text{He}$ Dilution Refrigerator	161

2.8.6	Properties of the Main Components of a $^3\text{He} \sim ^4\text{He}$ Dilution Refrigerator:	163
2.8.7	Examples of $^3\text{He} \sim ^4\text{He}$ Dilution Refrigerators	167
2.8.8	Developments	169
2.8.9	$^3\text{He} \sim ^4\text{He}$ Dilution Refrigerator: Addendum: Example	169
2.9	^3He solidification - Demagnetization - Others	175
2.9.1	Outline	175
2.9.2	TODO:Refrigeration by Solidification of Liquid ^3He : Pomeranchuk Cooling	175
2.9.3	TODO:Refrigeration by Adiabatic Demagnetization of a Paramagnetic Salt	176
2.9.4	TODO:Nuclear Adiabatic Demagnetization	179
2.9.5	TODO:Other techniques for refrigeration below 1 K	182
2.9.6	TODO:Summary	182
2.10	Superconducting materials and technologies	184
2.10.1	TODO:Introduction to the phenomenology of superconductivity	184
2.10.2	TODO:Vortex dynamics, vortex pinning	185
2.10.3	TODO:Review on superconducting materials (relevant for applications) and preparation methods	186
2.10.4	Iron-based superconductors (IBS)	189
2.10.5	Hints on Superconducting Film Deposition	191
2.10.6	Cryoelectronic devices	192
2.10.7	TODO:state of the art	193
2.10.8	TODO:Material requirements	193
2.10.9	TODO:Substrate Requirements	193
2.10.10	Brief overview of applications of superconductivity	193
2.11	The Technology of Josephson Junctions	196
2.11.1	Josephsons Junctions:main concepts and phenomenology	196
2.11.2	Current-Phase Relation and Junction Types	198
2.11.3	Revisiting the Current-Phase Relation (CPR)	202
2.11.4	SFS Josephson Junctions: The π Junction	203
2.11.5	More on the SFS Junctions	205
2.11.6	The π -ring Tricrystal Experiment with d-wave Superconductors	206
2.11.7	TODO:Applications of π -JJs: π -SQUIDs and Beyond	207
2.11.8	Serial Junctions	209
2.11.9	Hybrid Josephson Junctions (Hybrid JJs)	209
2.11.10	The Resistively Shunted Junction (RSJ) Model	210
2.11.11	The Evolution of Josephson Junction Technology	213
2.11.12	High-Temperature Superconducting Junctions (HTS JJs)	214
2.11.13	Applications of Josephson Junctions	214

Chapter 1

Part 1

1.1 Wafer Production and Monocrystals Growth

To create micro-things we need a special technology: semiconductive materials. They're useful because typically they can modify electrical and optical properties.

1.1.1 Monocrystals growth for miniaturised devices technology

. Semiconductor materials used include:

- **Silicon:** the most used, even if the 90% of what we'll see is valid as semiconductor.
- **III-V compounds:** *GaAs* (*gallium arsenide*) (mainly for high frequency applications) and *InP* (mainly optoelectronic applications).
- *SiC* (→ high T & harsh environments applications).

The aims of monocrystals growth process are:

- Obtaining **extremely pure silicon** to be able to control doping. → To control electrical properties. The usual accepted level of impurities is < 1 over 10^9 atoms of silicon.
- Obtaining **very regular silicon crystals** (defects in the crystallographic building worsen conduction characteristics of the semiconductor). Note: Perfection does not exist, we try to minimize the defect. In quantum technologies we voluntarily introduce defect.
- Obtain a **monocrystalline structure** rather than a **polycrystalline one** (a set of different crystallographic buildings with different orientations → polysilicon). In a polycrystalline structure, the carriers mobility is reduced because of discontinuities between different crystals.

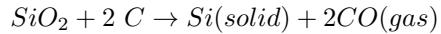
The fabrication process moves sequentially from raw materials to finished wafers:

Raw materials	(Sand, coke, coal, wood chips)
Metallurgical grade silicon	(used in making steel)
Electronic grade silicon	(polysilicon)
Single crystal boule	
Single crystal wafers	
Finished wafers	

1.1.2 Monocrystals growth

The starting material for silicon is relatively pure quartz based sand (SiO_2 or silica).

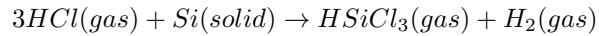
Silicon Refining Sand is reduced in an arc furnace with coal, coke, and wood chips at $2000^\circ C$ for ~ 8 days. The reaction is:



This forms **metallurgical grade silicon (MGS)**. Its purity is $\sim 98\%$. Major impurities (metals) include: Fe 0.8%, Al 0.3%, Cr 0.04%, Ti 0.03%, Mn & V 0.02%. **MGS** is used in Steel, Semiconductor, Silicones, and Aluminium alloys.

The furnace structure includes quartzite, coal, coke, wood chips, and a submerged electrode. Intermediate reactions occur, such as: $SiO + CO \rightarrow SiO_2 + C$ @ $1600^\circ C$. $SiO + 2SiC + CO$ from SiO and C . $SiO_2 + C \rightarrow CO$ at $1780^\circ C$ (to melt SiO_2). $SiC + SiO_2 \rightarrow Si + SiO + CO$. Total production is $\sim 900,000$ tonnes/p.a.

Trichlorosilane (TCS) Production MGS is crushed and reacted with HCl to obtain trichlorosilane (TCS). The reaction is:



This occurs at $325^\circ C$. The reaction is exothermic and must be cooled to keep at $325^\circ C$ to minimize byproduct formation.

TCS Purification $HSiCl_3$ is a liquid at room temperature (boiling point at $32^\circ C$). B and P impurities are present in $HSiCl_3$. Impurities are eliminated through **fractioned distillation** of the liquid.

Fractioned distillation

Fractioned distillation is the separation of a mixture into its component parts, or fractions. Chemical compounds are separated by heating them to a temperature at which one or more fractions of the mixture will vaporize. It uses distillation to fractionate.

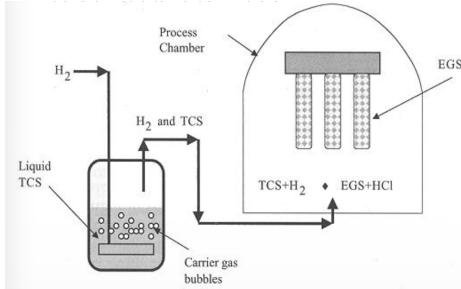


Figure 1.1: Siemens process

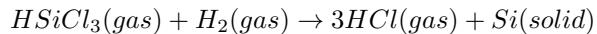
In the distillation process, the crude oil mixture is heated in a boiler into a super-hot mixture of liquid and vapour called the feed. The mixture is fed into a distillation tower. Compounds with a lower boiling point rise up as vapours, while compounds with a higher boiling point fall downwards as liquids. The tower contains trays that allow the vapour to bubble upward through the liquid, helping to exchange heat and resulting in more effective separation. The distilled products (fractions or distillates) are piped off from the different levels. This may take place along multiple distillation towers.

Petrochemical outputs resulting from distillation can include Propane & Butane, Petrol (Gasoline), Kerosene / Jet Fuel, Diesel Fuel, Heavy Fuel Oil, Waxes & Lubricants, and Bitumen (Asphalt). Light Ends have a Boiling point $< 0^{\circ}C$ and Residuals have a Boiling point $> 500^{\circ}C$.

After purification, TCS is 99.9999999% pure.

1.1.3 Electronic-Grade Silicon (EGS) Production

Purified TCS is reduced with H_2 to obtain **electronic-grade silicon (EGS)**. The **Siemens process** (Figure 1.1) is used:



This reaction (Chemical Vapor Deposition, **CVD**) occurs in a cold chamber of a reactor containing an heated Si rod which acts as a deposition seed at $1100^{\circ}C$.

After a few deposition days, the result is an **extremely pure** (~ 1 impurities over 10^9 atoms) **polycrystalline silicon**, which will be the raw material for the production of the monocrystalline one with the Czochralski (CZ) or with the Float-Zone (FZ) process.

The fabrication of slices from pure silicon consists first of fabricating a crystal, precisely an **ingot**, which is then cut into slices and transformed into wafers.

A comparison of impurity levels (ppm) between MGS and EGS is provided:

Impurity	MGS (ppm)	EGS (ppm)
A1	1570	-
B	44	<0,001
Fe	2070	4
P	28	<0,002
Sb	-	0,001
Au	-	0,00007

1.1.4 Czochralski Process (CZ)

The **Czochralski Process** was discovered, more than invented, in 1916.

Process Conditions

- Hyper-pure *Si* (EGS *Si*) is placed in a quartz crucible.
- Temperature: $\geq 1415^{\circ}C$ (Si m.p.).
- Pressure: 20 Torr (Ar or He).
- Ar Flow Rate: 20 to 50 *liters/min.*
- Time: 18 to 24 hours (for initial setup/melting).
- Typical growth speed: some *mm/min.*
- A monocrystalline *Si* seed is introduced.

The seed is rotated about its axis to mix the melt and produce a circular cross-section crystal. The rotation inhibits the natural tendency of the crystal to grow along certain orientations to produce a faceted crystal. The pulling time takes 3 days.

1.1.5 Ingot Characteristics and Parts

- Length: Up to 2 meters.
- Diameter: Up to 300mm (12").
- Weight: Up to 250 kg.

The parts of the ingot definition are:

- **Seed:** determines crystal orientation.
- **Neck:** removes the dislocations (like Edge Dislocation or "extra" plane of atoms) generated by thermal shock when the seed is dipped into the melt.
- **Shoulder or taper:** transitions to required diameter.
- **Body:** Part of the crystal actually used for wafers.
- **Endcone or bottom:** prevents shockback into the body when the crystal is removed from the melt.

1.1.6 Doping from the Melt

Dopants are added to the melt to pull controlled doped crystals. Dopants are typically added in the form of doped polysilicon for concentration control.

The concentration of dopant in the crystal (C_S) is generally not the same as that in the melt (C_L). This difference is governed by the **segregation coefficient** (k):

$$k = \frac{C_S}{C_L}$$

Where C_S = Concentration of dopant in solid (wt dopant/wt solid) and C_L = Concentration of dopant in melt (wt dopant/wt melt).

Since k in general is less than 1, the dopant becomes increasingly more concentrated in the melt. Dopant concentration thus changes along the length of the crystal.

Dopant Segregation Coefficients (k)

Dopant	P	As	B
k	0.35	0.3	0.8
atomic weight	30.97	74.92	10.81

Since k in general is less than 1, the dopant becomes increasingly more concentrated in the melt.

Boron produces a relatively flat profile because its k is close to 1. Dopants with $k \ll 1$ (like Antimony, $k = 0.023$) produce much more doping variation along the crystal.

1.1.7 III-V Monocrystals Growth: Liquid Encapsulated CZ (LEC)

The **Liquid Encapsulated CZ (LEC)** is a CZ technique modification used for the growth of $GaAs$. $GaAs$ has a low boiling point, and evaporation could result in toxic fumes and non-uniform crystal growth. Furthermore, the components Ga and As have differing vapor pressures, which complicates a stoichiometric precipitation over a large crystal volume.

Solution: The solution is to pressurize the chamber and melt B_2O_3 (about 1 cm thick) on top to seal and suppress evaporation.

- B_2O_3 does not react with $GaAs$ during growth.
- Ga melts at $30^\circ C$, B_2O_3 melts at $500^\circ C$ and seals the crucible, while Ga and As start reacting at $800^\circ C$ to produce $GaAs$.
- A high pressure As atmosphere is maintained in the reactor to avoid As evaporation (a very volatile element).
- Process pressure is approximately $200\ atm$ $Ar + AsH_3$.
- Pyrolytic BN crucibles are used, which do not react with $GaAs$.

After crystalline seed introduction and monocrystal growth, slow cooling down ($30\text{--}80^{\circ}\text{C}/h$) is performed. The melt GaAs required temperature is $T_F\text{ GaAs} = 1238^{\circ}\text{C}$.

1.1.8 From Boule to Wafers

The process steps after monocrystal growth are:

1. Ingot cropping This consists of cutting, or cropping, the extremities of the ingot, which are regions with high defect concentration and variable diameter.

2. Ingot inspection This involves:

- Control of the size of the ingot: if undersized, is trashed.
- Control of the resistivities on the top and bottom faces of the ingot: due to the variation of the doping concentration during the pulling, the final resistivity varies as a function of location. A check of resistivity and an agreement with specifications are needed. A four probe method equipment makes these measurements.
- Check of the crystallographic orientation of the ingot (XRD).

Approximately 50% of the material can be rejected at this stage.

3. Shaping of ingot (Outer Diameter Grinding) Make it round and of the correct diameter: during the pulling, due to the very large set of physical parameters to control, the diameter of the ingot slightly varies. That creates some waves at the ingot surface. To get slices of calibrated diameter suitable for automatic equipment, a cylindrical polishing is needed.

4. Flat(s) grinding One or two flat zones on the edge of the ingot are processed to get a crystallographic orientation reference for the wafer fabrication. This reference will be used during the wafer process (orientation of conducting zones, crystallographic axes for the die cutting...). Orientation flat grinding references: 45° for (111) type n, 180° for (111) type p, 190° for (100) type n, and a primary flat for (100) type p.

5. Ingot sawing (Wafer Slicing) This sawing is proceeded with a diamond tooth saw. The creation of a thickness of powder equivalent to the saw thickness leads to loss. Typical thickness range is $300\text{ }\mu\text{m} - 1\text{ mm}$.

The saw blade itself is about $400\text{ }\mu\text{m}$ thick. Together with the loss at the seed and tail end of the crystal, only 50% of the ingot ends up in wafer form. The dominant state of the art slicing technology is **Multi-Wire Sawing (MWS)**. In MWS, a thin wire is arranged over cylindrical spools so that hundreds of

parallel wire segments simultaneously travel through the ingot. The sawing effect is achieved by SiC or other grinding agents that run along the rotating wire.

6. Edge grinding After sawing, "peaks of matter" remain on the peripheral zones of the slices, which must be removed. A circle shaped edge is created to make manipulation easier during IC fabrication, avoiding degradation of transfer equipment and preventing cracks or dislocations that lead to definitive breaking. This step can improve cleaning results and reduce breakage up to 400%.

7. Wafer lapping and grinding The thicknesses and section shapes after sawing can be significantly different. To decrease the cost, the quantity of matter to grind has to be minimized. Usually, the wafers are sorted by thickness range of ten micrometers ($10 \mu m$). In order to improve their surface quality, the slices are polished using a mixture that contains alumina or diamond grains for which the size is about several micrometers (final roughness $< 2 \mu m$).

8. Wafer polishing This polishing is **chemico-mechanical based**. The equipment is similar to the lapper, but the solution is less aggressive, containing smaller grains of alumina or diamond (grain diameters can be as low as $0.1 \mu m$) and acid or basic chemical agents. Chemical agents used can include $NaClO$, $Br_2 + CH_3OH$, and $NaOH + SiO_2$ gel.

Mirror finishing is necessary for lithographic steps. The polishing process creates a very smooth surface with a roughness of $1 nm$ or better. Polishing results in regular reflection from a smooth surface, unlike diffuse reflection from a rough surface. Wafers can be **SSP** (Single side Polished) or **DSP** (Double side Polished).

9. Wafer cleaning This step removes abrasive species and contaminants by ultra pure deionized water rinsing.

10. Laser marking for ID The writing of the lots mentioning ingot number, date, etc., is performed through a laser beam scanning. These indications allow controlling each wafer during the fabrication process of circuits and devices. The identification code includes VENDOR IDENTIFICATION, RESISTIVITY IDENTIFICATION, DOPANT SPECIES, CRYSTAL GROWTH ORIENTATION, and CHECK CHARACTERS.

1.1.9 Wafer Evolution and Advantages of Larger Diameters

Wafer Evolution The evolution of standard wafer diameters is:

- 1964: 25 mm.
- 1969: 50 mm.

- 1974: 75 mm.
- 1978: 100 mm.
- 1982: 125 mm.
- 1985: 150 mm.
- 1990: 200 mm. (Current MEMS standard).
- 1998: 300 mm. (Current ICs standard).
- 450 mm (18") diameter wafer was expected in production by 2018-20 through an alliance between Intel, Samsung, and TSMC, but was abandoned.

Note that 1 inch = 2,54 cm, and 1 mil = 0,001 in = 25,4 μ m.

Advantages of Larger Diameter Wafers Larger Diameter Wafer Advantages are summarized below, comparing surface area relative to a 50mm wafer, considering a 3mm exclusion zone (black ring):

	100mm	150mm	200mm
Gross surface area advantage over 50mm	4X	9X	16X
Surface area % lost to 3mm exclusion zone	22%	12%	8%
Net surface area advantage over 50mm	4.4X	10.3X	18.8X

Further comparison of surface area and chip count for various configurations is available:

	56x50mm (2")	8x150mm (6")	5x200mm (8")
Net surface area	88,274 mm^2	130,288 mm^2	147,796 mm^2
Surface area gain vs. 50mm		48%	67%
Count of 45x45mil LED chips	62,944	97,600	111,280
LED chip count gain vs. 50mm		55%	77%

1.2 Crystallography and crystalline defects

1.2.1 3 types of solids

- **Crystalline:** All atoms arranged on a common lattice.
- **Polycrystalline:** Different lattice orientation for each grain.
- **Amorphous:** Atoms are disordered (no lattice).

Microstructure of electronic materials Examples of microstructure components include:

- Polycrystalline materials: Lines show lattice orientation, Grain boundary.
- Amorphous materials: SiO_2 , nitride spacer, silicide, gate oxide.
- Single-Crystal Material: substrate Si .
- Features size example: 50nm.

1.2.2 Crystal structure

Atoms are arranged on a regular grid called a **lattice**. A Bravais lattice is an infinite array of points that appears the same from any lattice point, and there are **only 14 unique lattices possible in three dimensions**.(Figure 1.2)

The 14 Bravais Lattices

1. Simple Cubic
2. Body Centered Cubic (bcc)
3. Face Centered Cubic (fcc)
4. Simple Tetragonal
5. Centered Tetragonal
6. Simple Orthorhombic
7. Base Centered Orthorhombic
8. Body Centered Orthorhombic
9. Face Centered Orthorhombic
10. Simple Monoclinic
11. Centered Monoclinic
12. Triclinic

Crystal Family	Lattice System	Schönlies	14 Bravais Lattices			
			Primitive (P)	Base-centered (C)	Body-centered (I)	Face-centered (F)
Triclinic		C _i				
Monoclinic		C _{2h}	$\beta \neq 90^\circ$ $a \neq c$ 	$\beta \neq 90^\circ$ $a \neq c$ 		
Orthorhombic		D _{2h}	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$
Tetragonal		D _{4h}	$a = c$ 		$a \neq c$ 	
Hexagonal	Rhombohedral	D _{3d}	$\alpha \neq 90^\circ$ 			
	Hexagonal	D _{6h}	$\gamma = 120^\circ$ 			
Cubic		O _h				

Figure 1.2: 14 Bravais Lattices

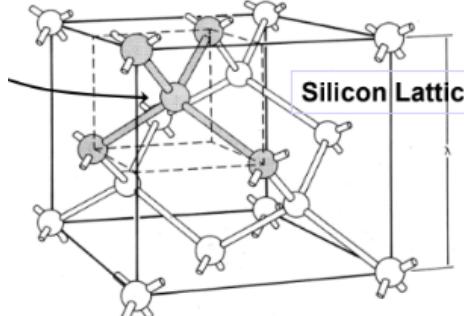


Figure 1.3: Silicon lattice. Gray spheres denote the corners of a tetrahedron

13. Rhombohedral

14. Hexagonal

These lattices correspond to 7 Crystal families/Lattice systems: triclinic, monoclinic ($\beta \neq 90^\circ$), orthorhombic ($a \neq b \neq c$), tetragonal, rhombohedral, hexagonal ($\gamma = 120^\circ$), and cubic ($a = b = c, \alpha = \beta = \gamma = 90^\circ$).

Cubic Crystal Structure

- Simple Cubic (only Polonium).
- Body Centered Cubic (bcc) (Na, W, Cr, Fe, K, Mo ...).
- Face Centered Cubic (fcc) (Al, Au, Cu, Pt, ...).

The **lattice parameter** ($= a_0$) is the length of a cube side.

1.2.3 Silicon lattice (Diamond Structure)

The crystalline structure of *Si* and *Ge* (also *Sn*) belongs to the family of cubic crystals.

It is formed by the **interpenetration of two fcc** sub-lattices where one sub-lattice is moved away from the other along the diagonal of the cube by an amount corresponding to $\frac{1}{4}$ of its length (a shift $= a_0 \frac{\sqrt{3}}{4}$).

In diamond structure:

- All atoms are equal.
- Every atom is surrounded by 4 atoms placed at the same distance at the corners of a tetrahedron.(Figure 1.3)

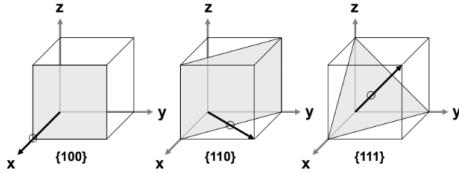


Figure 1.4: Common cubic crystal planes illustrated are $\{100\}$, $\{110\}$, and $\{111\}$.

Zincoblenz Structure This structure is typical of **III-V compounds** (GaAs, InP, ...).

It is **similar to diamond**, but one fcc lattice is made by III atoms and the other by V atoms.

1.2.4 Miller indices

Miller indices are used to describe surfaces (or planes) cut in crystals. Different cuts result in different physical/chemical properties.

General Procedure

1. **Cut** crystal along some plane.
2. **Determine x, y, z, intercepts** (x_o, y_o, z_o). If a plane does not intersect an axis, the intercept is infinity.
3. **Take reciprocals** ($\frac{1}{x_o}, \frac{1}{y_o}, \frac{1}{z_o}$) and **reduce to smallest integer**.

A bar is placed over negative intercepts.

Anisotropic Wet Etching and Surface Properties

Different surfaces have different arrangements of atoms, leading to different properties (mechanical, chemical, electrical, ...).

- **(100) Surface Orientation (for Silicon):**
 - Gives a lesser number of dangling bonds on the surface.
 - Results in less impurity accumulation, leading to better control.
 - This characteristic is **required for a MOSFET**.
- **(111) Surface Orientation (for Silicon):**
 - Gives an advantage of **fast oxidation** as more atoms are available on the surface to react with oxygen.
 - This is a characteristic **desired for BJTs**.

The angle between the (100) and (111) surfaces in silicon is 54.74° . (Figure 1.4)

1.2.5 Crystalline defects

A real crystal is different from ideality due to:

- **Finite dimensions** (surface atoms present "dangling bonds").
- **Presence of defects.**

Defects have a **strong influence** on electrical, chemical, mechanical, and optical properties of the semiconductor.

Defects Classification

- Point defects.
- Line defects.
- Surface defects.
- Volume defects.

Point Defects

Point defects often arise from too fast deposition or low substrate temperatures, leaving no time for atoms to move to crystal lattice sites.

- **Vacancy** (missing atom).
- **Self interstitial** (extra atom).
- **Substitutional impurity** (impurity atom in lattice).
- **Interstitial impurity** (impurity atom not in regular lattice site).
- **Antisite**.

These defects cause distortion of planes.

Line Defects: Edge Dislocation

An edge dislocation is the **insertion of an extra half-plane of atoms** which distorts the lattice. This defect is called a line defect because the locus of defective points produced in the lattice lies along a line, which runs along the edge of the extra half-plane. Inter-atomic bonds are significantly distorted only in the immediate vicinity of the dislocation line.

- The dislocation creates stresses (compression and tension).
- Typical densities are often $10^{10} - 10^{12}$ dislocations/cm² in films.
- Dislocations form from the film growth process, dislocations in the substrate continuing into the film, or contamination on the substrate.

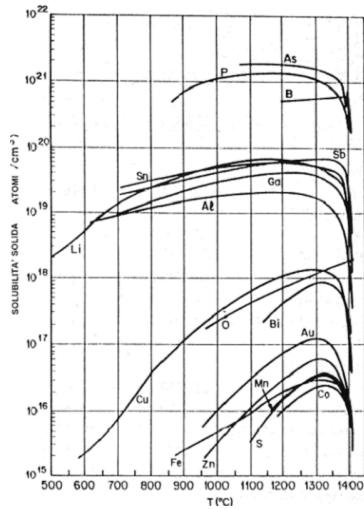


Figure 1.5: A graph detailing Solid Solubility (ATOMI/cm³) versus Temperature (T(°C)) shows solubility profiles for various elements (As, P, B, Sb, Sn, Ga, Al, Li, Bi, Au, Mn, Co, Cu, S, Fe, Zn).

Surface Defects

Grain Boundaries are interfaces between two single crystal regions of different orientation. Atoms at grain boundaries tend to be **loosely bound**. That implies that those boundaries are **more reactive** (leading to corrosion/etching/oxidation) and allow **accelerated diffusion** along boundaries. The number of grain boundaries in a film (grain size) depends on deposition rate and substrate temperature. Generally, lower T yields smaller grains and many boundaries, while higher T yields larger grains and fewer boundaries. Grain size is often proportional to film thickness (thinner films tend to have smaller grains).

Gemination is an abrupt variation of crystal orientation.

Volume Defects

Impurities have an intrinsic **solid solubility** in the host crystal, which is variable with temperature. Usually, solid solubility decreases with decreasing temperature.

If impurities are introduced at maximum concentration for a certain temperature (T) and then T is lowered, the excess impurities (respect to the solid solubility at that temperature) can **precipitate in local concentration**. This local concentration change can drastically change the volume of the crystal, leading to volume dislocations.

1.3 Cleanroom Technology

1.3.1 Introduction and Rationale for Cleanroom Technology

A cleanroom is defined as **an environment where particulate concentration is reduced and kept under a specified level**. Typically, temperature and humidity are also controlled. Cleanroom Technology is crucial because contamination can cause several significant problems:

- **Ruin devices:** A single ruined device in a complex circuit can cause the whole chip to fail. This leads to a lower yield of good chips per wafer, resulting in higher costs and lower profits per chip.
- **"Poison" equipment:** Equipment must be removed from the manufacturing line, which reduces production throughput and revenue.
- **Pose a health risk:** Contamination can endanger employees, customers, and the environment, greatly increasing costs, and potentially leading to litigation, insurance issues, and compliance with safety issues.

1.3.2 Contamination Induced Problems

Particles are defined as undesired objects (size between $\sim 0.06 \mu\text{m}$ and a few microns) whose composition depends on the contamination source. Particles adhere to surfaces by van der Waals forces, condensation of vapour, or the formation of chemical bonds.

Contamination leads to various issues in semiconductor manufacturing:

- **Mobile ions in oxides** (e.g., Gate (+5V), Source (OV), Drain (OV)) can change electric fields and voltages at the semiconductor/dielectric interface.
- **Unintentional films between layers** can create open circuits or short circuits between layers. They can also impede adhesion between films.
- **Effect on photolithography:** Dust particles can create defects (cuts or protrusions) through shadowing.
- **Effect on movable microstructures:** Contamination can cause jamming.

1.3.3 Sources of Contamination

Humans cause most of the contamination. People continuously exfoliate skin and replace hair. Sources associated with people include dirt and oils tracked into labs on shoe soles, widespread use of make-up, perfume, and hair gels.

Particle shedding rates from people are substantial:

- Sitting quietly: 100,000 particles shed per minute.
- Moving: 1 million particles shed per minute.
- Walking: 5 million particles shed per minute.

Other sources of contamination include:

- Ventilation and Room Structure.
- Equipment and Machines.
- Abrasion during automated wafer handling.
- Mechanical mechanism wear and lubrication.
- Aging plastic and rubber parts (valves, pumps, filters, etc.).
- Materials (chemicals, water, gas, targets).
- Processes (residues, depositions, slurry).

Examples of Common Air Contaminants and Size (in microns)

Contaminant	Size (μm)
Human Hair	70-100
Human Sneeze	10-100
Pollen	10-100
Pet Dander	5-100
Spores from Plants	6-100
Household Dust	0.05-100
Dust Mite Debris	0.5-50
Mold	2-20
Skin Flakes	0.4-10
Bacteria	0.35-10
Smoke	0.01-1

1.3.4 Classification Standards and Measurement

The 1966 publication of Fed. Std. 209 defined a standard for clean rooms divided by classes. A clean room belongs to a certain class if its distribution of concentration/dimension of particles is under the line corresponding to the class. For example, **Class 100** must have less than 100 particles greater than $0.5 \mu\text{m}$ per cubic foot. The Normal environment value is 10,000 particles per cubic foot.

Particulate concentration can be directly measured with methods of filtering and counting (usually too slow) or with a suitable apparatus exploiting **optical scattering of a laser beam** (just like looking at particulate through a sun beam).

The ISO standard (ISO 14644-1) requires results to be shown in cubic meters (1 cubic meter = 35.2 cubic feet).

Class limits (maximum allowable particles cubic foot)						
ISO	SI	CLASS	[0.1	[0.2	[0.3	[0.5
CLASS 3	M 1	1	100 35		35 1	
	M 1.5	1	35	7	3	1
	M 2	10	99.1			2.83
CLASS 4	M 2.5	10	10,000 345	75	30	562 10
	M 3	100	991			28.1
CLASS 5	M 3.5	100	100,000 3,460	750	300	2620 100
CLASS 6	M 4.5	1,000	1,000,000 34,500	N/A	N/A	96,200 1,000
CLASS 7	M 5.5	10,000	945,000	N/A	N/A	962,000 10,000
CLASS 8	M 6.5	100,000	9,450,000	N/A	N/A	9,520,000 100,000

Figure 1.6: ISO 14644-1 Fed. Std. 209 E(SI) International ISO standard requires results to be shown in cubic meters (1 cubic meter=35.2 cubic feet)

Classification Details (Fed. Std. 209 E)

The classification table provides maximum particles per cubic foot and associated design requirements (Air changes per hour, ACH, and Capital Cost per ft²):

- **Class 100,000:** 100,000 particles per ft³ at 0.5 μm. Air changes per hour: 12-18. Approx. capital cost: \$10 per ft².
- **Class 10,000:** 10,000 particles per ft³ at 0.5 μm. Air changes per hour: 18-30. Approx. capital cost: \$50 per ft².
- **Class 1,000:** 1,000 particles per ft³ at 0.5 μm. Air changes per hour: 150-300. Approx. capital cost: \$350 – 400 per ft².
- **Class 100:** 100 particles per ft³ at 0.5 μm. Air changes per hour: 400-540. Approx. capital cost: ~ \$1200 per ft².
- **Class 10:** 10 particles per ft³ at 0.5 μm. Air changes per hour: 400-540. Approx. capital cost: ~ \$3500 per ft².
- **Class 1:** 1 particle per ft³ at 0.5 μm. Air changes per hour: 540-600. Approx. capital cost: ~ \$10,000+ per ft².

1.3.5 Environmental Control and Airflow

Controlled Parameters

- **Temperature** is controlled, typically 20 to 22°C. Tolerances range from ±3.0°F (Class 10,000) to ±0.1°F (Class 5).
- **Humidity (RH)** is controlled, usually around 50% RH, or 40 ÷ 60%. Tolerances range from ±5% (Class 10,000/1,000/100) down to ±1% (Class 5).
- **Pressure:** The room is held at **positive pressure** to blow dust OUT.
 - Pressure is typically ~ 25 Pa for Class 100, 1000, and 10,000.

- Pressure is typically 75 to 100 Pa for Class 1 and 10.
- Usually, doors open inward, so room pressure closes them shut.
- Note: Biohazard rooms operate at negative pressure to keep bugs IN.

Airflow and Design

Laminar flow is a key factor to reduce environmental particle contamination. Factors affecting laminar flow include the number of operators, cleanroom design, equipment layout, and operator behavior.

Design elements include:

- Supply air often flows through HEPA Filters located in the ceiling.
- Return air can utilize low-wall returns or return air plenums below a raised floor.
- A **raised floor** assures optimal performance for Class 5. Raised floors are very common, though low wall returns also work.
- Class 1 requires Gel/Flush grid ceiling systems with raised floors. It also mandates 540 to 600+ air changes per hour and 98 + % ceiling coverage.
- Class 4 requires 540 to 600 air changes per hour and 85 – 90% ceiling coverage.

1.3.6 Cleanroom Air Filters

- **High Efficiency Particulate Air (HEPA) Filters:** This is the most common type of cleanroom air filter. They offer high efficiency, low pressure drop, and good loading characteristics. They use glass fibers in a paper-like medium. By definition, a true HEPA-rated filter will retain **99.97%** of incident particles with a diameter of 0.3 μm or larger.
- **ULPA Filters:** Class of air filters which meet a minimum performance level of **99.999%** on 0.3 μm efficiency. In the cleanroom market, ULPA can also be rated to 99.9995% on 0.12 – 0.16 μm . Class 1 requires ULPA filters.

1.3.7 Cleanroom Clothing (Gowning)

Cleanroom clothing requirements escalate with the cleanliness class.

- **Highest Classes (Class 1, Class 10):** Require Hood, Hair Cover, Coverall, Intersuit, Boots, Facial Cover, and Gloves. The frequency of change is **Each Entry**.
- **Class 100:** Requires Hood, Hair Cover, Coverall, Boots, Facial Cover, and Gloves. The Intersuit is Optional. Frequency of change is Daily.

- **Lower Classes (Class 1,000, Class 10,000, Class 100,000):** Require less coverage, typically Hood/Cap/Hair Cover and Coverall/Frock, Boots/Footwear, and optional facial cover/gloves. Frequency of change ranges from 3 Times/Week to 2 Times/Week.

Gowning procedures include specific steps, such as putting on safety glasses, ensuring all hair is contained by the cap, checking for a snug fit and good face seal for the hood, bending the nose piece of the face mask for a snug fit, and pulling the hem of the glove over the coverall sleeve. Shoe Covers may be needed before entering the gowning area.

1.3.8 Deionized Water (DI Water)

Deionized water must be particulate- (and contaminant-) free. A large amount of water is used in washing steps.

- **Salt Removal:** Salts are removed through ion-exchange resins, which are both cationic (steal metals substituted by H^+) and anionic (steal anions substituted by OH^-). These resins finally capture salts, producing pure water.
- **Organic Removal:** Organic compounds (non-ionic) are adsorbed by active carbon. Water is then filtered through an absolute filter ($< 0.2 \mu m$).
- **Quality Control:** Electric conductance is a good parameter for evaluating water quality. Resistance must always be in the $10 - 18 M\Omega \cdot cm$ range.
- **Sterilization:** Water is sterilized with UV light to avoid bacteria and algae proliferation.
- **Drying:** After washing, substrates are dried with a **filtered hyperpure nitrogen flow under pressure** to avoid the slow natural drying (evaporation) of water, which would leave a small amount of contamination.

1.4 Introduction to Lithography

Lithography derives from the Greek words *λίθος* (lithos, 'stone') and *γράφειν* (graphein, 'to write'). It is the fundamental process used to **transfer circuit or device patterns from a layout to substrates**.

1.4.1 Micro/Nano Technologies Overview

Micro and Nano Technology often employs **Batch Fabrication**. Because selective processing (directly growing, depositing, or etching films or introducing dopants only in desired areas) is typically impossible, processes involve blanket deposition followed by removal from areas where the material is not needed. Technologies are categorized as:

- **SELECTION:** Lithography.
- **ADDITIVE:** Deposition (Epitaxy, CVD, PVD, ALD, Electroplating, ...) and Thermal Oxidation.
- **SUBTRACTIVE:** Etching.
- **MODIFICATION:** Doping, Annealing.

1.4.2 Photolithography Function

Photolithography is the specific technique used to **select**, transferring a desired pattern from a master copy (or mask) into a photo-sensitive polymer layer called **photoresist** on the wafer surface. The patterned polymer layer is then used to protect areas of the wafer during subsequent processes, such as etching or ion implantation. For the realization of a single device, **many photolithographic steps are needed** (many mask layers required), and alignment between these steps is crucial. During every step, similar regions of different devices are realized in parallel across the wafer surface.

1.4.3 Masks and Alignment

Mask Generation

Masks are fabricated using high-resolution lithographic techniques.

- **Substrates:** Pure and amorphous **SiO₂** slabs that are perfectly planar (rectified).
- **Absorbing Layer:** A thin absorbing metal layer is deposited, typically **Cr** (Chromium), chosen for optimal adhesion to the substrate.
- **Materials Transparency:** Quartz glass maintains a transmission of approximately 90%, even at wavelengths lower than soda lime glass.
- **Features:** Obtained through laser or e-beam lithography, depending on the required resolution.

Alignment

Alignment (or Overlay) ensures that circuit elements from the current step are precisely positioned relative to features already on the wafer. It is accomplished by superimposing previously printed alignment marks on the wafer with corresponding marks on the mask. Verification is achieved through a split-field microscope that views opposite sides of the wafer simultaneously. Sources of misalignment include **X – Y** displacement and θ rotation. For alignment using front-backside exposure tools:

- **IR transparent samples:** Infrared (IR) alignment provides resolution of $1 - 5 \mu\text{m}$.
- **Not IR transparent samples:** Alignment marks on the front side of the wafer are used, achieving a resolution of $5 - 20 \mu\text{m}$.

1.4.4 The Photolithography Process Flow

A typical Photolithography process recipe flow involves several sequential steps:

1. Priming (Wafer Preparation)
2. Resist Coating
3. Resist SoftBake
4. Resist Exposure
5. Resist Post Exposure Bake (PEB)
6. Resist Image Developing
7. Resist HardBake

Wafer Preparation and Priming

Wafer surfaces are prepared prior to resist coating.

- **Dehydration Bake:** Wafer surfaces are baked to drive off adsorbed water, typically at **130°C** to **200°C** in a convection oven.
- **Adhesion Promoter (Primer):** An adhesion promoter is applied. **HMDS** (hexamethyl disilazane, $[(CH_3)_3Si]_2NH$) is commonly used and can be applied as a vapor in an oven or spun on.

Resist Spin-On Coating

The resist is deposited onto the wafer by the spin-on method.

1. **Deposition:** An excess of the liquid polymer (photoresist) is delivered onto the substrate.

2. **Spin-up:** Under centrifugal force generated by the spinning disk, the liquid flows radially.
3. **Spin-off:** The liquid flies off the edge, and a uniform thick film tends to form and thins slowly to an equilibrium thickness.
4. **Evaporation:** The liquid solidifies due to solvent evaporation.

Resist thickness (t) is dependent on variables like viscosity (ν) and angular velocity (ω):

$$t = K \cdot S \cdot \left(\frac{\nu}{\omega^2 \cdot R^2} \right)^{1/3}$$

Where K is a constant, S is the fraction of solids, and R is the radius. The thickness decreases as the spin speed (rpm) increases. Quality measures during this step include thickness, uniformity, and the absence of particles and defects.

Soft Bake

The Soft Bake (or Pre-Bake) step is typically performed on a hot plate at **90°C** to **115°C** for 60 up to 90 seconds. **Purpose of Soft Bake:**

- Partial evaporation of photoresist solvents.
- Improves adhesion, uniformity, etch resistance, and linewidth control.
- Optimizes the light absorbance characteristics of the photoresist.

To leave the resist film, a solvent molecule must diffuse on the resist surface, evaporate, diffuse through the diffusion boundary layer above the surface, and ultimately be carried away by the air stream. The time and temperature of the soft bake significantly impact the average remaining solvent content (e.g., PGMEA).

1.4.5 Photoresist Components and Properties

Components

Photoresists (resists) are photosensitive organic mixtures that contain:

- **Inactive polymer resins:** A binder that provides mechanical and chemical properties, such as adhesion, chemical resistance, rigidity, and thermal stability.
- **PhotoActive Compounds (PAC).**
- **Solvent:** Controls the viscosity of the base, keeping it liquid for spinning (affecting film thickness).
- Other components: Surfactants, leveling agents, sensitizers, and dyes (to absorb λ).

Properties: Polarity (Positive vs. Negative)

- Positive Resist:

- Light (photons) breaks the polymer chains (CHAIN SCISSION), making them more soluble in the developing solution.
- Areas exposed to light are dissolved.
- Resulting pattern replicates the mask pattern (holes in the resist correspond to holes in the mask).
- Positive resist generally has **better resolution** due to the smaller size of the polymer.

- Negative Resist:

- Light induces **cross-linking** of the polymer chains, making them less soluble in the developing solution.
- Areas exposed to light become cross-linked and resist the developer chemical.
- Resulting pattern replicates the inverse (negative) pattern of the mask (resist remains where the mask had holes).

Properties: Resolution and Thickness

Resolution is the smallest opening or space that can be produced in the photoresist layer. It is related to the exposure source and developing process. A **thinner layer has better resolution**. However, thicker layers may be required to serve as effective etch and implantation barriers or to ensure a pinhole-free film.

Properties: Sensitivity and Exposure Energy

Sensitivity is also known as the **Dose-to-Clear** (D_0 or E_i), which is the amount of exposure energy required to just clear (or consolidate) the resist in a large area for a given process.

- Typical exposure energies (Dose) range from $50 - 150 \text{ mJ/cm}^2$ for Positive (UV) resist and $20 - 30 \text{ mJ/cm}^2$ for Negative (UV) resist.
- Energy (E) is related to intensity (I) by $E := I \cdot t$.

Properties: Contrast (γ)

The photoresist contrast (γ) is the **maximum slope of the development rate curve** (H-D curve).

$$\gamma := \frac{1}{\log_{10}(E_f) - \log_{10}(E_i)} = \frac{1}{\log_{10} \frac{|E_f|}{|E_i|}}$$

Typically, γ is ~ 2 to 4 for g- and i-line resists, and ~ 5 to 10 for DUV resists.

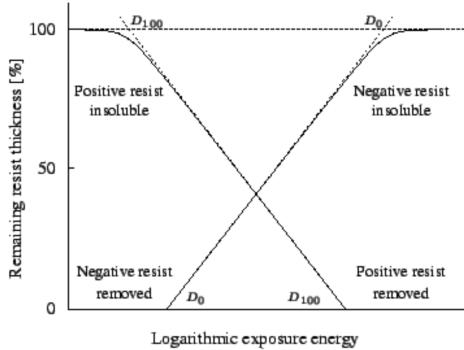


Figure 1.7: This figure presents the relation between Sensitivity and Contrast.

- **Good Resist Contrast** yields sharp walls and no swelling.
- **Poor Resist Contrast** yields sloped walls and swelling.

Sensitivity and contrast are not purely intrinsic properties; they also depend on process conditions (developer, development time, baking time, wavelength, substrate).

1.4.6 Back to lithographic process: Exposure Sources and Tools

The exposure step transfers the mask image to the resist-coated wafer and activates the photo-sensitive components. Quality measures include linewidth resolution, overlay accuracy, particles, and defects. Note: it's always a fight against diffraction.

Exposure Sources

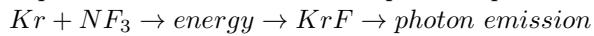
- **Mercury (Hg) Lamps:** Traditionally used, has some spectral lines from a high-intensity plasma. Key lines include **G-line** (436 nm), **H-Line** (405 nm), and **I-Line** (365 nm). Optical filters can limit exposure wavelengths. The use of filters depends on the photoresistor: if we're using a positive one, we should use a filter, because its peak absorption frequency is the same as **I-line**(we'll limit diffraction); instead for a negative we don't use a filter.
- **UV LEDs:** An attractive alternative to energy-consuming mercury lamps due to technological advancements. Benefits include long lifespan, low power consumption, limited heating, instant-on functionality (no mechanical shutter), and no daily calibration or maintenance costs. High-power LEDs can cover the UV spectrum between 360 and 410 nm, mimicking Hg lines.

- **DUV Lasers (Excimer Lasers):** Used to increase resolution for **deep ultraviolet (DUV)** lithography systems. **Excimer pulsed lasers** are typically used. These are powerful, but extremely expensive to purchase and maintain.

In excimer lasers, two elements, e.g. a noble gas and a halogen (from a halogen containing compound), which can react and “bind” together only in the excited state but not in their ground states, are present.

Providing energy will therefore drive the reaction, creating the excimer. When the excitation energy is removed, the excimer dissociates and releases the energy at the characteristic wavelength.

A pulsed excitation is used to repeat the process.



- *KrF* (Krypton Fluoride) at **248 nm** is used for $0.25 \mu\text{m}$ resolution.(lower wavelength than Hg, brighter).
- *ArF* (Argon Fluoride) at **193 nm** is used for $0.12 \mu\text{m}$ resolution.
- *F₂* at **157nm** was theorized but never used due to technical issues: finding suitable resists and transparent optical components at these wavelengths.

- **EUV (Extreme Ultraviolet):** Generates light at **13.5nm** by firing a high-energy IR pulsed laser on a droplet of molten tin. It’s hit two times: First time is melt and the second vaporized →emits radiation. EUV lithography must operate in a **vacuum** and exclusively use nearly 11 **reflective optics** (including reflective masks) because EUV is absorbed by glass, oxygen, and water. Mirrors typically use about 50 pairs of $\lambda/4$ multilayer layers. It’s important to remember that those mirrors reflect only 70% of total light: this means that the wafer receives only 2% of the initial light. ASML is the leader company in this industry.

- **Electron-Beam Lithography (EBL):** Uses a focused beam of electrons instead of light, allowing for **direct writing** of nm-scale features without a mask. Electrons have wave-like properties with wavelengths on the order of $0.2 - 0.5\text{\AA}$, making diffraction negligible. E-beam lithography is primarily used to produce photomasks or in Nanotech/Quantum Research. It requires a vacuum environment.

- **PMMA:** Poly(methylmethacrylate) is a positive photoresist used for short-wavelength lithography (deep UV, EUV, e-beam, X-ray). It is not sensitive to $\lambda > 240 \text{ nm}$ and has low sensitivity, often requiring sensitizers to improve speed. This ensures and high resolution.

Exposure Tools

The three main optical exposure systems are Contact, Proximity, and Projection.

- **Contact Exposure:**

- *Advantages:* Fast (you expose the whole wafer at once), simple, inexpensive.
- *Disadvantages:* The mask **contacts** the wafer, causing mask wear and contamination, leading to unacceptable defect densities and a shorter lifetime for the mask. No magnification is possible.
- Max resolution is given by **Minimum Line/Gap Period** ($2 \cdot b_{min}$): $2 \cdot b_{min} = 3\sqrt{\lambda \cdot \frac{z}{2}}$; being λ the wavelength of the radiation and z the thickness of the photoresist. The definition is about $0.5 - 1 \mu m$.

- **Proximity Exposure:**

- *Advantages:* Fast (wafer exposed at once) and the mask **does NOT contact** the wafer (gap $s = 10 - 25 \mu m$), preventing mask wear or contamination.
- *Disadvantages:* The separation gap causes greater diffraction, leading to **less resolution**. Cannot easily print features below a few μm (except for X-ray systems). It's also more expensive because the mask is usually the same size of the wafer.
- Minimum Line/Gap Period ($2 \cdot b_{min}$): $2 \cdot b_{min} = \sqrt{\lambda \cdot (s + \frac{z}{2})}$; s being the distance between the Mask and the Photoresist.

Note: Contact and Proximity are performed by the same machine, that costs about half a million, for Projection we need another one that costs about 50 millions.

- **Projection Exposure (Steppers/Scanners):**

- *Characteristics:* Projection printing **dominates today** due to high resolution and low defect densities. It uses a complex projection lens (e.g., 25-30 glass elements, 500 Kg).
- *Advantages:* Mask **does NOT contact** the wafer. Offers de-magnification (usually 4X or 5X reduction), which makes masks cheaper to create because they have larger features. High resolution.
- *Disadvantages:* An high de-magnification means each die exposed separately, this means it takes longer to expose the entire wafer. It's also very complex and expensive, requires precision stepper motor. This all implies a lower throughput.

Steppers and Scanners

Note: now the mask is called **reticle**.

- **Stepper:** The step and repeat method exposes one die (or part of one die) at a time. Only the wafers move to expose different dice. Steppers rely on a precision system to automatically expose all dice after an initial

alignment. Extra lenses allow reduction of the mask dimensions onto the wafer (4:1 to 10:1): it's easier to create defect-free masks at larger geometries.

- **Scanner:** All state-of-the-art tools use this approach today. **Both wafers and reticle move** to expose different dyes (reticle's speed depends on magnification). By using a fixed exposure slit, scanners achieve the required exposure field size with a **much smaller lens**, providing better control for aberrations. Nowadays it's used with immersion optical tool.

Front-Backside Alignment & Exposure Tool

For device realization, many photolithographic steps and corresponding masks are required, necessitating precise alignment between layers.

Alignment, also referred to as Overlay, is achieved by superimposing two previously printed alignment marks on the wafer with corresponding alignment marks on the mask. Alignment of these two pairs of marks is verified through a **split – field microscope** that views opposite sides of the wafer simultaneously. Sources of misalignment that must be accounted for include X-Y displacement and θ rotation.

The resolution capabilities of the tool depend on whether the sample is transparent to infrared (IR) light:

- **IR transparent samples such ad Si:** Alignment is achieved using Infrared (IR) technology.
 - Resolution: $1 - 5 \mu m$.
- **Mask Alignment:** Is performed when the sample is not IR transparent (Si covered in photoresist). Alignment relies on standard mask alignment techniques.
 - Resolution: $5 - 20 \mu m$.

1.4.7 Post-Exposure Processing and Pattern Transfer

Reflections, Standing Waves, and PEB

Standing Waves

Standing waves occur due to **interference between the incident and reflected light** waves within the photoresist film. This creates a sinusoidal intensity variation along the thickness of the photoresist, leading to high and low exposed regions separated by $\lambda/4n$ (n =refractive index). This non-uniform exposure causes faster and slower development rates.

It'll show photoresistor's areas that are Over and Under exposed, creating a knurling.

Fixing: Post Exposure Bake (PEB)

PEB is a crucial (and more used) step used to **remove standing waves**; though is not always necessary.

- **Mechanism:** PEB causes thermal re-distribution (diffusion) of the exposure products, smoothing the concentration pattern created by the sinusoidal intensity distribution.
- **Benefits:** It increases focus latitude and resolution, and improves thermal stability.
- **Conditions:** Hot plate bake 110°C to 125°C for 60 up to 90 seconds.

Prevention: Antireflective Coatings (ARC)

The use of ARCs, dyes, and filters can also help prevent interference. ARC materials common for 365 nm are Titanium Nitride, while Silicon nitride and silicon oxynitride are used for 365 – 248 nm.

1.4.8 Developing

Developing involves dissolving the soluble areas of the photoresist using a developer chemical. The final visible patterns (windows or islands) appear on the wafer.

- **Developer Solution:** Typically an aqueous-alkaline solution (normality: 0.1N – 0.35N). NaOH, KOH, or TMAH based solutions are used for PR, while xylene is used for NR.
- **Methods:** Puddle development (100-200cc dispensed), Immersion Develop, or Spray-On Develop.
- **Timing:** The timing is critical. **Overdevelopment** (too long) or **underdevelopment** (too short) both negatively affect linewidth. Underdeveloped resist can prevent access to the underlying layer.

1.4.9 Hard Bake

The Hard Bake step involves evaporating all solvents remaining in the PR. It is not always mandatory but is needed for acid wet etching.

- **Purpose:** Improves etch and implantation resistance, improves PR adhesion, polymerizes and stabilizes photoresist, and allows PR flow to fill pinholes.
- **Conditions:** Hot plate bake 110°C to 125°C for 60 up to 120 seconds.
- **Caution:** Over-baking can cause excessive PR flow, negatively affecting photolithography resolution.

1.4.10 Pattern Transfer Methods

- **Etching Process:** The PR is spun, patterned, and used as a mask while the underlying layer is etched, before the PR is removed. Will always introduce impurities, and the process is too long and difficult in the case we have multiple metals.
- **Lift-Off Process:** The PR is spun and patterned. Metal (or other material) is deposited over the patterned PR, and then the excess metal is lifted off with the PR layer.
 - It's used a lot in the Quantum world.
 - This process allows features to be obtained without chemical wet etching: a solvent is used (often the same of the photoresistor) instead of acid.
 - It requires relatively high thickness resist ($> 2 \mu\text{m}$).
 - The deposited film thickness must be less than the resist thickness (typically a maximum of 1/3).
 - **Image Reversal (IR) resists** are beneficial as they have a particular behavior: they act like a positive resistor under exposure; thanks to a **reversal bake** the soluble area becomes cross-linked, creating a negative profile slope (undercut) which facilitates the lift-off of deposited layers right after a **flood exposure**.
In this way we have strong adhesion, a negative slope with resolution of a positive resistor.

1.4.11 Models and Simulation

Lithography simulation relies on models from two fields: Optics (to model aerial image formation) and Chemistry (to model latent image formation in the resist).

- **Aerial Image Simulation:** Commercial tools (PROLITH, DEPICT, ATHENA) calculate the aerial image based on mask design and optical system characteristics. These simulators are powerful, as the math is well understood and fast algorithms are implemented. Simulation can show how variables like numerical aperture (NA), feature size, and wavelength (λ) affect image quality.
- **Latent Image Simulation:** This step calculates the concentration of the PhotoActive Compound (PAC) after exposure. Simulation can demonstrate how PEB "smears out" standing wave effects, producing a more uniform PAC distribution.
- **Development Simulation:** A number of models exist for resist developing. The simplest is purely empirical (Dill et al.), which uses three optical parameters (A, B, C) and a rate relationship $R(M)$ (rate of removal vs. degree of exposure) to describe positive photoresist exposure and development.

1.5 Silicon Oxidation

Thermal Oxidation of Silicon

Thermal Oxidation of Silicon is the formation of Silicon Dioxide (SiO_2) on the silicon surface exposed to an oxidant ambient ($\text{O}_2 \text{ H}_2\text{O}$) at high temperature ($700 - 1200^\circ\text{C}$) or starting from a polysilicon film.

By the Thermal Oxidation process, SiO_2 layers of 30 \AA to $2 \mu\text{m}$ thickness with controlled interface properties can be grown.

SiO_2 growth is a key process step in manufacturing all Si devices

- Electrical Insulation
- Masking Layer: to protect selected regions during doping or etching
- Thermal Insulation: thermally almost insulating
- Sacrificial Layer: in Surface Micromachining Technology

SiO_2 can be also obtained by other techniques, like CVD (Chemical Vapor Deposition), but with different and typically worse electrical characteristics.

The stability and ease of formation of SiO_2 was one of the reasons that Si replaced Ge as the semiconductor of choice and GaAs is not so widely applied though its excellent electrical properties (thermal oxidation of GaAs produces non-stoichiometric films, with consequent bad electrical insulation and surface protection).

Silicon is naturally covered by a thin ($\sim 1\text{nm}$) layer of silicon dioxide ("native oxide")

1.5.1 Properties of Silicon Dioxide

Thermal Silicon Oxide \rightarrow Glass or Amorphous state (fused silica), with a short range structure of tetrahedrons with oxygen ions (O^{2-}) at each corner centered around a silicon atom (Si^{4+}).

- It is amorphous.
- Stable and reproducible SiO_2 growth
- Melting point: 1700°C
- Density: 2.21 g/cm^3 (almost the same as Si that is 2.33 g/cm^3)
- Crystalline SiO_2 [Quartz] = 2.65 gm/cm^3
- Atomic density: $2.3 \times 10^{22} \text{ molecules/cm}^3$ (For Si, it is $5 \times 10^{22} \text{ atoms/cm}^3$)
- Refractive index: $n = 1.46$
- Dielectric constant: $\epsilon = 3.9$

- Excellent electrical insulator: resistivity $\rho > 10^{20} \Omega\text{cm}$, energy gap $E_g = 8 - 9\text{eV}$.
- High breakdown electric field: $> 10^7 \text{V/cm}$

SiO_2 can also be crystalline \rightarrow quartz

- Crystalline quartz density: 2.65 g/cm^3
- Thermal SiO_2 (amorphous structure) density: $2.21 \text{ g/cm}^3 \rightarrow$ more open structure (only 43% of available volume is occupied by silicon dioxide molecules) \rightarrow interstitial diffusion of impurities (e.g. Na) through the network

1.5.2 How does Silicon Oxide?

Heat wafers in an atmosphere containing an oxidant, usually O_2 or steam.

- Dry Oxidation: $\text{Si} + \text{O}_2 \rightarrow \text{SiO}_2$
- Wet Oxidation: $\text{Si} + 2\text{H}_2\text{O} \rightarrow \text{SiO}_2 + 2\text{H}_2$

Note: Silicon is **consumed**. Oxidation occurs at the $\text{Si} - \text{SiO}_2$ interface NOT on top of the oxide!

Since oxidation occurs at the $\text{Si} - \text{SiO}_2$ interface:

- O_2 or H_2O must diffuse through the previously grown oxide film .
- oxidation (growth) rate will fall off with increasing time and oxide thickness.

$$\frac{\text{molecular density of Oxide}}{\text{molecular density of Silicon}} = \frac{2.2 \times 10^{22} \cdot \frac{\text{molecules}}{\text{cm}^3}}{5.0 \times 10^{22} \cdot \frac{\text{atoms}}{\text{cm}^3}} = 0.44$$

1.5.3 Substrate Orientation + Mono/Poli comparison

Silicon dioxide grows ~ 1.7 times faster on (111) surfaces than on (100) surfaces (because of the greater silicon atoms density of the (111) planes).

1.5.4 Key variables in Oxidation

- Time
- Temperature
 - reaction rate
 - solid state diffusion
- Oxidizing species

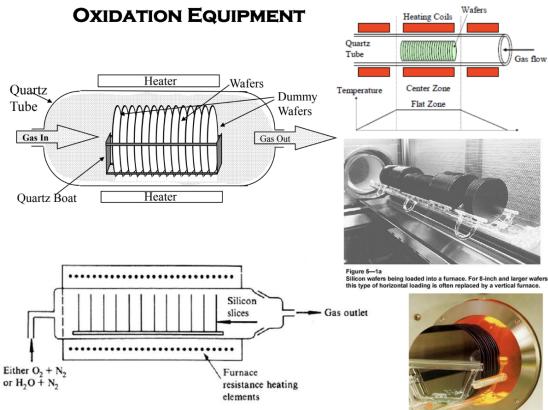


Figure 1.8: Oxidation Equipment

- wet oxidation is much faster than dry oxidation
- Crystallographic orientation
 - Monocrystalline Si orientation / Polycrystalline Si
- 2nd order variables:
 - Surface cleanliness metallic contamination can catalyze reaction
 - Furnace pressure (high P increases oxidation rate)
 - Silicon doping (different behavior among p and n dopants)

1.5.5 Modelin Thermal Oxidation of Si

Grove worked at Fairchild Semiconductor before becoming the 4th employee at the nascent Intel Corporation. He became Intel's president in 1979, CEO in 1987.

Aim of the model → provide a simplified tool to predict and calculate:

- Oxidation rate $\frac{\partial x_{\text{ox}}(t)}{\partial t}$
- Final oxide thickness at time $t \rightarrow x_{\text{ox}}(t)$

Most successful model - published by B.E. Deal and A. S. Grove in 1965. "General Relationship for the Thermal Oxidation of Silicon , Journal of Applied Physics, Vol. 16, no. 12, pg. 3770, 1965.

Grove worked at Fairchild Semiconductor before becoming the 4th employee at the nascent Intel Corporation. He became Intel's president in 1979, CEO in 1987.

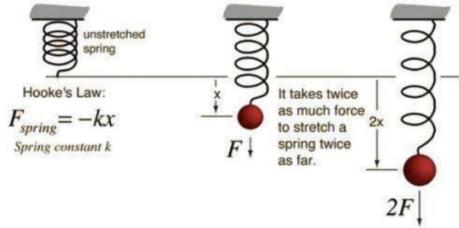


Figure 1.9:

Diffusion Limited Process Fluxes:

- F_1 = gas phase oxidant (O_2 or H_2O) diffuses to the surface of growing film
- F_2 = through oxide \rightarrow oxidant molecules diffuse through growing film to SiO_2/Si interface
- F_3 = reaction flux \rightarrow oxidant molecules react with Si

Steady State Requires:

$$F_1 = F_2 = F_3$$

N_0, N_i = concentration of oxidant species $\frac{\text{number of particles}}{\text{cm}^3}$

Fick's first law of diffusion:

$$J = F_2 = -D \cdot \frac{dN}{dx} = D \cdot \left(\frac{N_0 - N_i}{x_o} \right) \frac{\text{number of particles}}{\text{cm}^2 \text{ sec}}$$

N_0, N_i = concentration of oxidant species $\frac{\text{number of particles}}{[\text{cm}^3]}$

$$\begin{cases} N_0 = 5.2 \times 10^6 \frac{\text{molecule}}{\text{cm}^3} & \text{for dry oxidation (at } 1000^\circ\text{C and 1 atm)} \\ N_0 = 2 \times 10^{19} \frac{\text{molecule}}{\text{cm}^3} & \text{for wet oxidation (at } 1000^\circ\text{C and 1 atm)} \end{cases}$$

D = coefficient of diffusivity $\left[\frac{\text{cm}^2}{\text{sec}} \right]$ combine with

$$J = F_3 = k_s \cdot N_i$$

k_s = surface reaction speed for oxidation
and $J = F_2 = F_3$ steady-state condition yields (by eliminating N_i)

$$J = \frac{D \cdot N_0}{\left(x_0 + \frac{D}{k_s} \right)}$$

Which spring characteristics are inside k ?

- Spring diameter
- Number of coil & spring length
- Spring wire diameter
- Spring material
- Metallurgical process for spring fabrication

Similarly which characteristics are inside k_s (surface reaction speed for oxidation)?

- Temperature
- Silicon solid state (poly/monocrystalline)
- Crystallographic orientation
- Oxidant molecules (since different oxidation activation energies for O₂ and H₂O)
- Silicon surface contamination
- Silicon doping (dopants affect oxidation)

but no characteristics related to Diffusion phenomena!

Reaction of oxidant molecules with silicon generates silicon dioxide

- there are 2.2×10^{22} molecules/cm³ of silicon dioxide in the oxide
- it is necessary one O₂ molecule to produce one SiO₂ molecule
- and two H₂O molecules to produce one SiO₂ molecule
- so, assuming N_{ox} as the concentration of oxidant molecules in the oxide to form the oxide

→

$$\begin{cases} N_{ox} = 2.2 \times 10^{22} \text{ molecules/cm}^3 & \text{for dry oxidation,} \\ N_{ox} = 4.4 \times 10^{22} \text{ molecules/cm}^3 & \text{for wet oxidation.} \end{cases}$$

As a consequence the growth rate of the oxide layer can be expressed as:

$$\frac{dx_{ox}}{dt} = \frac{J}{N_{ox}} = \frac{\left(\frac{D \cdot N_0}{N_{ox}} \right)}{\left(x_{ox} + \frac{D}{k_s} \right)}$$

The differential equation can be solved by assuming

$$x_{ox}(t = 0) = x_i$$

where x_i is the starting thickness of silicon dioxide, or the thickness of the silicon dioxide layer grown in a previous step

$$\frac{dx_{\text{ox}}}{dt} = \frac{J}{N_{\text{ox}}}$$

Having J = Flow that is to say what is available (oxidant molecules/cm²/s) to perform the oxidation, and

N_{ox} = Concentration of oxidant molecules in the oxide to form the oxide that is to say what is necessary (molecules/cm³) to generate the oxide.

So, the general relationship for the silicon oxidation is:

$$\frac{x_{\text{ox}}^2}{B} + \frac{x_{\text{ox}}}{\left(\frac{B}{A}\right)} - (\tau + t) = 0$$

where:

$$x_{\text{ox}}(t = 0) = x_i$$

$$\tau = \frac{x_i^2}{B} + \frac{x_i}{\left(\frac{B}{A}\right)}$$

; τ is a time shift to take in account the starting thickness x_i

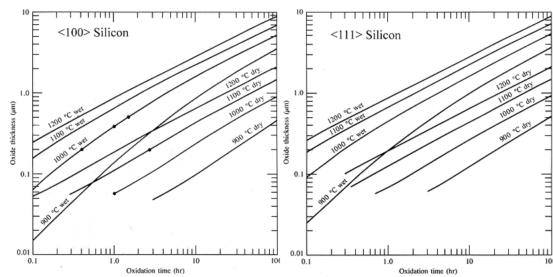
$$A = \frac{2 \cdot D}{k_s}$$

$$B = \frac{2 \cdot D \cdot N_0}{N_{\text{ox}}}$$

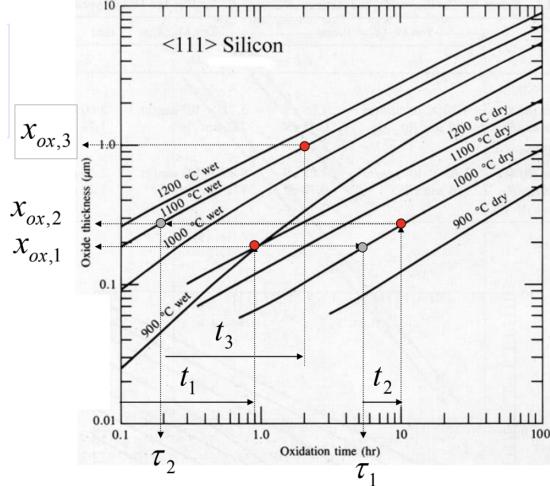
Finally, the solution is given by:

$$x_{\text{ox}}(t) = \frac{1}{2} A \left[\left(\sqrt{1 + \frac{4 \cdot B}{A^2} (t + \tau)} \right) - 1 \right]$$

1.5.6 Oxidation Graphs



Multiple oxidations (tracking τ)



$$t = \frac{x_{\text{ar}}^2}{B} + \frac{x_{\text{ar}}}{(\frac{B}{A})} - \tau$$

$$\tau = \frac{x_i^2}{B} + \frac{x_i}{(\frac{B}{A})}$$

Step	Time	Temp	Type	Overall Thickness
0	0.0 hr			0 nm
1	0.9 hr	900°C	Wet	190 nm
2	4.5 hr	1000°C	Dry	280 nm
3	1.8 hr	1100°C	Wet	

The question is: what is the time required to obtain the current thickness, but with the new oxidation conditions (T & Type of oxidation)?

If you can answer the current oxidation step may be dealt with as a simple time extension of the previous oxidation step → shift of the time origin.

1.5.7 Growth rate regimes

$$x_{\text{ox}}(t) = \frac{1}{2} A \left[\sqrt{1 + \frac{4 \cdot B}{A^2} (t + \tau)} - 1 \right]$$

Short Times: $(t + \tau) \ll \frac{A^2}{4 \cdot B}$

Simplify with Taylor Series Expansion:

$$\chi_{\text{ox}}(t) = x_{\text{ox}}(t_0) + \frac{x'_{\text{ox}}(t_0)}{1}(t - t_0) + \frac{x''_{\text{ox}}(t_0)}{1 \cdot 2} (t - t_0)^2 \dots$$

reduces to:

$$x_{\text{ox}}(t) = \left(\frac{B}{A}\right) \cdot (t + \tau)$$

At short time, the growth rate is limited by the rate of reaction of oxidant with silicon.

Growth rate is linear. B/A = linear growth parameter

Long Times: $(t + \tau) \gg \frac{A^2}{4 \cdot B}$

$$x_{\text{ox}}(t) = \sqrt{B \cdot (t + \tau)}$$

At long time, the growth rate is limited by the rate of oxidant diffusion through the growing layer Growth rate is parabolic. B = parabolic growth parameter

$$\frac{B}{A} = \frac{\frac{2DN_0}{N_{\text{ox}}}}{\frac{2D}{k_s}} = k_s \frac{N'_0}{N_{\text{ox}}}.$$

$$B = \frac{2DN_0}{N_{\text{ox}}}$$

1.5.8 Rapid thermal Oxidation (RTO)

Rapid thermal oxidation (RTO) is increasingly used in the growth of thin, high-quality dielectric layers and to reduce the thermal budget.

The primary issues that differentiate RTO from conventional thermal oxidation are the more complex chamber design, radiation source, as well as temperature monitoring. From the point of view of oxide-growth kinetics, RTO may be influenced by both thermally activated processes and a non-thermal, photon-induced process involving monatomic O atoms generated by UV (0° radicals) and creating a parallel oxidation reaction that dominates at lower temperature.

1.6 Epitaxy, CVD, ALD, PVD, electroplating

Epitaxy is a growth technique involving monocrystalline layers grown on a monocrystalline substrate, ensuring structural continuity. The substrate acts as a crystal seed, and the film is grown by progressively adding atoms.

1.6.1 Conditions and Types

For successful epitaxial growth, **the crystalline structure and lattice parameter should be similar**, generally having a maximum mismatch of approximately 2%. Epitaxial techniques are categorized based on the substrate (A) and the grown layer (S):

- **Homoepitaxy:** Layer S and Substrate A are the same material ($S = A$).
- **Heteroepitaxy:** Layer S and Substrate A are different materials ($S \neq A$), for example, 3C-SiC on Si.

Epitaxy is used to control the doping profile, such as adding a lightly doped or undoped layer on a heavily doped substrate to increase the breakdown voltage of the junction, or creating isolation junctions with opposing doping. Epi-layers often have a smoother surface than the substrate and are generally oxygen and carbon-free, unlike semiconductive substrates. Epi-layers can also be deposited over heavily doped areas called **buried layers**, which serve as low-resistance contacts.

1.6.2 Classification by Mother Phase

Epitaxial techniques are divided according to the characteristic of the Mother Phase (This is the source of the atoms that will form the new crystal layer):

1. **Vapour Phase Epitaxy (VPE) or Chemical Vapor Deposition (CVD):** The mother phase is a mix of gases or vapors at room or reduced pressure.
2. **Liquid Phase Epitaxy (LPE):** The mother phase is a solution of the active material with a suitable diluent.
3. **Molecular Beam Epitaxy (MBE):** The mother phase is a gas at a very reduced pressure, where collisions between molecules are practically absent.

1.6.3 Why is Epitaxy important?

- Add a lightly doped (or totally undoped) layer on a heavily doped substrate → increases breakdown voltage of the junction
- Substrates and epi-layer with opposing doping → isolation (reverse polarization junction)

- Deposited over patterned heavily doped areas called **buried layers** (which serves as a low-resistance contact)
- Epi-layer → smoother surface than the substrate
- One (or more) epi-layers on a substrate offer the device designer a further means of controlling the doping profile in a device structure (beyond that available with doping by diffusion or ion implantation)
- Physical properties of epi-layers differ from those of bulk material (i.e. epi-layers are generally oxygen and carbon-free, situation not obtained with semiconductive substrates)

1.6.4 Vapor Phase Epitaxy (VPE) / Chemical Vapor Deposition (CVD)

VPE Details

VPE consists of growing the crystal from **gaseous precursors** (gas sources) including the material and its associated doping. It is the *most used* epitaxial growth technique, especially for Si.

For VPE, **gases are dissociated** in the chamber to produce atoms (e.g., silicon atoms) that move towards and bond with the pre-existing lattice. Substrates must be heated, typically at **high temperatures** (e.g., around 1100°C) to ensure epitaxial growth. An example reaction is $\text{SiCl}_4 + 2\text{H}_2 \rightleftharpoons \text{Si} + 4\text{HCl}$.

CVD Hazards

Many gases utilized in CVD systems present significant hazards.

Table 1.1: Properties and Hazards of Common CVD Gases

Hazard	Description
Toxic	Hazardous to humans. E.g., Ar-sine (AsH_3) has an exposure limit of 0.05 ppm.
Corrosive	Causes corrosion to stainless steel and other metals. E.g., Hydrogen chloride (HCl).
Flammable	Burns when exposed to an ignition source and oxygen source. E.g., Hydrogen (H_2) (4-74% vol limits).
Pyrophoric	Spontaneously burns or explodes in air, moisture, or when exposed to oxygen. E.g., Phosphine (PH_3) and Silane (SiH_4).

VPE Reactors

VPE reactors consist of a gas inlet/control system, a deposition chamber, and an exhaust (e.g., scrubber). Wafers are typically loaded onto graphite susceptors that are heated via RF induction. This RF heating warms the susceptor while the chamber walls remain cold, limiting outgassing and contamination. Common reactor shapes include **Horizontal**, **Vertical**, **Barrel**, **Pancake**, and **Rotating Disk (Planetary)** reactors.

1.6.5 Liquid Phase Epitaxy (LPE)

- **Key Concepts:** Oversaturation, epitaxy, substrate, spontaneous nucleation, liquid, metallic homogeneous solution.
- **Conditions:** No wetting of crucible.
- **Growth Mechanism:**
 - Direct precipitation from liquid phase due to cooling down.
 - Oversaturated solution on a crystalline substrate → solute excess precipitates **epitaxially** during cooling.
- **Growth Conditions:**
 - Near equilibrium growth conditions → $\Delta T/T \approx 10^{-3} - 10^{-2}$.
- **Applications:** Particularly used for **III-V** compounds, **SiC**, or **multi-layered epitaxies** with different characteristics (i.e., doping).
- **Solvent:**
 - Low melting point metal component of the compound, for instance:
 - * Ga (m.p. 30°C) for the growth of **GaAs** and **GaAlAs** on GaAs substrates.
 - * In (m.p. 156°C) for the growth of **InP**, **InGaAs** and **InGaAsP** on InP substrates.
- **Growth Temperature (T_{growth}):** Compromise between salt solubility (increases with T) and risk of thermal degradation of substrates.
 - About 800°C for epitaxy on **GaAs**.
 - About 600°C for epitaxy on **InP**.

LPE Reactor

Growth takes place under a flux of purified H₂ (passed through a Pd barrier) at atmospheric pressure to avoid oxidation of the solution and the substrate. The apparatus utilizes a **Graphite crucible** which features fixed wells for the liquid growth solutions and a sliding part that hosts the substrate.

The substrate is pushed under the oversaturated solutions. The temperature (T) is then slowly lowered ($\approx 1^\circ\text{C}/\text{min}$), and the substrate is kept in position for the time necessary to achieve the desired layer thickness.

Growth stop is obtained by removing the substrate from the solution. This is possible because molten metallic solutions do not wet graphite and substrates → **sliding off** of the substrate is enough to remove the solution and immediately stop the growth.

1.6.6 Molecular Beam Epitaxy (MBE)

General Conditions: Low pressure gaseous mother phase in molecular flux conditions. An UHV (Ultra-High Vacuum) deposition chamber where the heated substrate faces a certain number of furnaces. Pressure is one of the most tedious thing to measure and so vacuum in nearly impossible to obtain.

Key Components:

- Furnaces (Knudsen Cells): Thermally controlled and host the elements and dopants to be deposited. They generate fluxes of atoms (for metals) and molecules (for non-metals) necessary for the growth.
- Substrate: Kept in rotation for growth uniformity.
- Sliding Shutter: Located in front of every furnace and used to stop the flux **abruptly** → allows for rapid change of flux composition (to obtain a layered epitaxy surface).
- Liquid N₂ Panels and Pumps: Used to achieve the necessary UHV conditions (along with the pumps).
- Loading Chamber (LOAD-LOCK): Used to load substrates without breaking the UHV conditions. Is not always present, but frequently is: it's a sort of preloading chamber. Used to, for example, introduce a new wafer and lower the pressure.

Importance of UHV:

- UHV is necessary to maintain molecular flux conditions. We can deposit with a very small flux, granting a high resolution (few nanometers). This implies it's terrible in terms of productivity.→ Can be used for very complex compounds.
- UHV reduces the incorporation of unwanted species that could dope the material.

Advantages/Control:

- Optimal technique for the control of **thickness** (atomic layers are feasible).

- Optimal technique for the control of **chemical composition** ($\pm 1\%$ of doping variation).

Reference Atmosphere Conversion: 1 atm = 760 torr (mm Hg) = 101323.2 Pa (N/m²) = 1.013232 bars = 1013.232 mbar = 14.7 psi

1.6.7 Chemical Vapor Deposition (CVD)

CVD is the formation of thin films from vapor phase reactants (gaseous precursors). High temperatures and low pressures are common conditions, but not necessary. CVD films include metals (W, Al, Cu), semiconductors (Si, Ge, GaAsP), insulators (Si₃N₄, SiO₂), and silicides (TiSi₂).

All CVD involves using an energy source to break reactant gases into reactive species for deposition.

CVD and epitaxy are two families of deposition techniques: VPE is both CVD and Epitaxy. CVD is not strictly monocrystalline.

Acceptance Angle and Conformality (Non-Conformal Deposition)

The growth rate is dependent on the flux density of gas molecules incident on the surface, which is a function of the **acceptance angle**.

Not uniform coverage → NOT CONFORMAL. This occurs because the reagent is adsorbed and reacts without significant surface migration.

Deposition Rate and Incidence Angle: The deposition rate is proportional to the incidence angle of molecules:

- 180° for the upper surface (maximum deposition rate).
- 270° for the upper edge of the trench.
- An angle proportional to $\arctan(w/d)$ for the bottom surface of the trench (where w is the width and d is the depth).

This angular dependence leads to a **reduced thickness** on the side walls of the trench and on its bottom. A non-continuous film (due to **auto-shadowing**) is even possible.

This characteristic is typical of **PVD (Physical Vapor Deposition)** methods.

Step coverage is defined using ratios like Conformity = b/c , Sidewall Step Coverage = b/a , and Bottom Step Coverage = d/a , where a, b, c, d are thickness measurements at specific points. High aspect ratio (h/w) structures are more difficult to fill without forming **voids** or **keyholes**.

Sometimes you want your film to be not conformal < 3 .

1.6.8 Types of Chemical Vapor Deposition (CVD) Reactors

Chemical Vapor Deposition (CVD) is a foundational technique in semiconductor fabrication, with different reactor types offering various trade-offs in perfor-

mance, dictated primarily by operating pressure and temperature. The three major types are Atmospheric Pressure CVD (APCVD), Low Pressure CVD (LPCVD), and Plasma Enhanced CVD (PECVD).

Atmospheric Pressure CVD (APCVD)

APCVD operates at high temperatures, typically between 350°C and 1200°C, and near atmospheric pressure (10 to 100 kPa). Its main appeal lies in its operational simplicity, resulting in **very high deposition rates** and high throughput. However, the high pressure environment leads to severe limitations, namely **poor uniformity** and **poor conformality** (step coverage), and the resulting film purity is lower compared to other methods. APCVD is primarily used for less demanding applications, such as depositing low-temperature oxides.

Low Pressure CVD (LPCVD)

LPCVD operates at high temperatures, usually between 550°C and 600°C, but at a much lower pressure of approximately 100 Pa. This low-pressure environment significantly improves the quality of the deposited film. LPCVD is highly valued for its **excellent uniformity**, high film **purity**, and **good conformity** across complex topographies. The primary drawback compared to APCVD is its **lower deposition rate**. It is the workhorse for depositing crucial films like **polysilicon**, **nitride**, and high-quality **oxide** layers.

Plasma Enhanced CVD (PECVD)

PECVD is distinguished by its use of plasma (that is a complex mixture resulting from the ionization and excitation of gas atoms) to supply the energy needed for the chemical reactions, which allows it to operate at much **lower temperatures** (150°C to 450°C). This low-temperature capability is critical for depositing films over substrates that cannot withstand high heat, such as those with pre-deposited metallization. PECVD also offers **good step coverage** (conformality) and a deposition rate that is higher than LPCVD. However, the use of plasma introduces potential drawbacks, including **plasma damage** to the underlying device structures and possible **chemical contamination**. The tool itself is also more complex. PECVD is often used for creating low-temperature oxides and for **passivation** layers.

Collision Processes

:

- **Ionization:** $e^- + Ar \rightarrow 2e^- + Ar^+$ (sustains the plasma).
- **Dissociation:** $e^- + SiH_4 \rightarrow e^- + SiH_3^\bullet + H^\bullet$ (generates highly reactive free radicals).

- **Excitation:** $e^- + Ar \rightarrow e^- + Ar^*$ (transfer of energy to higher level).
- **Relaxation:** $Ar^* \rightarrow Ar + h\nu$
 - As the excited element/molecule falls back to its original energy level, it must release energy. In this case a photon is released. This gives the plasma its “glow”.
 - Color of the glow is characteristic of the gas: $h\nu = \Delta E$ oxygen plasmas are blue, nitrogen purple. A stable colored glow is the proof that everything is stable.

Plasma Generation

Applying voltage to a neutral gas accelerates free electrons. When voltage is increased, energetic electrons ionize or excite gas atoms (cascading collisions), sustaining the reaction. The glow visible in a plasma is emitted when excited gas atoms return to the ground state. The process causes:

- **No Applied Voltage:** gas consists of neutral atoms.
- **Apply Voltage:** free electrons are accelerated, elastically collide with atoms.
- **Increase Voltage & Field:**
 - energetic electrons ionize or excite gas atoms and contribute another electron to the process (cascading collisions) → sustained reaction.
 - excited gas atoms returning to ground state emit visible photons (e.g., the glow that is visible).
 - oxygen plasmas are blue, nitrogen purple
- Chamber is evacuated
- Chamber is filled with gas(es)
- RF energy is applied to a pair of electrodes (typically 13.56 MHz) (The electric field is switched more than thirteen millions times per second)
- Applied energy accelerates electrons increasing kinetic energy
- Electrons collide with neutral gas molecules, forming ions and more electrons → Electrons oscillating in the glow region acquire enough energy to cause ionization (the alternating potential polarity affects the electrons, sending them into an oscillatory path). Gas ions are too massive to respond to the high-frequency field (ions, being much heavier and therefore slower, don't keep up with the alternating field).
- Steady state is reached (plasma); ionization = recombination

Plasmas are used to force reactions that would not be possible at low temperature (plasma is the energy source generating reactive gases).

- **Advantages:** uses lower temperatures ($150 - 450^{\circ}\text{C}$) necessary for near end processing or deposition on T sensitive materials, more degrees of freedom to achieve film properties (good step coverage, density, composition), higher deposition rate than LPCVD.

- **Disadvantages:** plasma damage or chemical contamination typically results.

The following is the content starting from slide 22, presented in LaTeX format, drawing on the provided sources.

1.6.9 PECVD process control parameters

1. Temperature.
2. Deposition time.
3. Gas composition (precursors) and flow rate.
4. System pressure and reactant partial pressure.
5. RF power, frequency, and bias.

PECVD Pros and Cons

Pros:

- Lower temp. than LPCVD (suitable for back-end processing)
- High deposition rates

Cons:

- More complex tool to control and maintain
- Poor quality of grown films (contamination, hydrogenation)
- Plasma generates "soup" of perhaps unsavory chemicals

1.6.10 Atomic Layer Deposition(ALD)

Atomic Layer Deposition (ALD) is a modified CVD process to manufacture conformal thin films. The process uses several gases led into the process chamber alternating. Each gas reacts in a self-limiting reaction at the surface with the first gas, causing the reaction to come to a standstill once the surface between the reactions, the chamber is purged with an inert gas, like nitrogen or argon. A specific example for an ALD process is the deposition of aluminum oxide, which can be realized with trimethylaluminum (TMA, $\text{C}_3\text{H}_9\text{Al}$) and water (H_2O).

ALD Process Steps (Example: Al_2O_3)

1. **First step:** elimination of hydrogen atoms bound to oxygen (e.g., due to environment humidity) at the wafer surface. The methyl groups (CH_3) of TMA react with the hydrogen to form methane (CH_4). The remaining molecules bond with the unsaturated oxygen. If these atoms are saturated, no more TMA molecules can react at the surface.
2. **Second step:** The chamber is purged and subsequent water steam (H_2O) is led into the chamber. Every hydrogen atom of the H_2O molecules reacts with the former deposited surface atoms to form methane, while the hydroxyl anion is bond to the aluminum atoms. This leaves new hydrogen atoms at the surface which can react in a subsequent step with TMA like in the beginning.

ALD Characteristics

- Each cycle, including gas injection and pumping, takes few seconds.
- With ALD, even conformal 3D structures can be deposited very uniformly.
- Insulating films are possible as well as conductive ones, on different substrates (semiconductors, polymers, ...).
- Film thickness can be controlled very precisely by the number of cycles.

1.6.11 Silicon Dioxide (CVD)

CVD silicon dioxide has worst electrical properties with respect to thermally grown oxide, but complementary applications. The reaction is:



- **Undoped oxide:** used for separation and insulation among different levels of metals (IMO: InterMetalOxide; where oxidation is not feasible), as a masking layer for implantation and diffusion, and to increase the thickness of thermal oxides.
- **Phosphorous doped oxides (PSG):** used for separation and insulation among different levels of metals, and as the final passivation layer for devices.
- **P, As or B doped oxides:** occasionally used as diffusion sources.

Silicon Dioxide Deposition Methods Comparison

- **PECVD $SiH_4 + O_2$** (max productivity, lowest quality)
 - Deposition temp: 200°C

- Composition: $\text{SiO}_2(\text{H})$
- Thermal stability: Loses H
- Dielectric Strength: 5×10^6 V/cm
- Etch Rate (100 H₂O: 1 HF): 400 Å/min
- Step coverage: Non-conformal

- **LPCVD SiH₄ + O₂**

- Deposition temp: 450°C
- Composition: $\text{SiO}_2(\text{H})$
- Thermal stability: Densifies
- Dielectric Strength: 8×10^6 V/cm
- Etch Rate (100 H₂O: 1 HF): 60 Å/min
- Step coverage: Non-conformal

- **LPCVD TEOS**

- Deposition temp: 700°C
- Composition: $\text{SiO}_2(\text{C...})$
- Thermal stability: Stable
- Dielectric Strength: 10×10^6 V/cm
- Etch Rate (100 H₂O: 1 HF): 30 Å/min
- Step coverage: Conformal

- **LPCVD SiCl₂H₂ + N₂O**

- Deposition temp: 900°C
- Composition: $\text{SiO}_2(\text{Cl})$
- Thermal stability: Loses Cl
- Dielectric Strength: 10×10^6 V/cm
- Etch Rate (100 H₂O: 1 HF): 30 Å/min
- Step coverage: Conformal

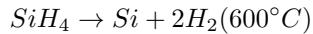
- **Thermal oxidation**(benchmark, used as reference; max quality)

- Deposition temp: 1000°C
- Composition: SiO_2
- Thermal stability: Stable
- Dielectric Strength: 11×10^6 V/cm
- Etch Rate (100 H₂O: 1 HF): 25 Å/min
- Step coverage: Conformal

A lower HF etch rate means better film quality (denser film).

1.6.12 Polycrystalline Silicon ("Poly")

- Used as gate electrode in MOS devices (more affordable than Al, particularly for thin gate oxides). Also used for the fabrication of high resistivity resistors.
- Process: LPCVD process at $600 - 650^{\circ}C \rightarrow$ pyrolysis of silane reaction:



- Can be doped by diffusion, implantation or in-situ doping (doping gases in the reaction mixture). Implantation is the usual choice for its low process temperature.
- **Grain size depends on deposition temperature; hotter deposition leads to larger grain structure.**
- Polysilicon can exhibit columnar or random small grain structures.

1.6.13 PVD (Physical Vapor Deposition)

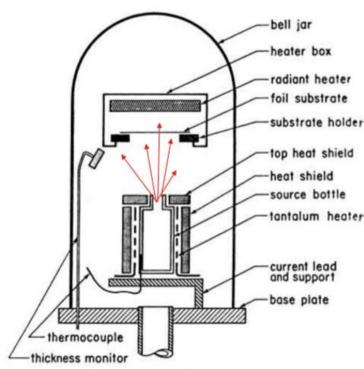
What does it mean physical? Is the type of bonding. Is a phenomena similar to vapor condensation on windows during winter. Physical Vapor Deposition → Film growth is achieved by the accumulation (condensation) of a vapor onto a cooler substrate.

PVD involves two main steps:

1. Evaporation
2. Sputtering

PVD is used mostly for metal layers or alloys.

1.6.14 PVD-Thermal Evaporation



It's done inside a bell jar.

- Requires low working pressure (5×10^{-5} Pa $\approx 5 \times 10^{-7}$ Torr; High vacuum condition, intermediate to epitaxy and low pressure.) to increase the **mean free path**. →Deposition is directional and a line of sight deposition.
- Deposition rate is determined by emitted flux and by geometry of the target and wafer holder.
- It is faster than sputtering.
- It is limited to certain materials, with relatively low melting point.
- Wafers are placed upside down to reduce particulates.
- A shutter is used for better timing.

Conformality

- Because thermal evaporation is line of sight (like a point source), it results in shadowing, meaning steps are often NOT covered, they are not uniform.
- Thickness depends on the angle between the surface and the incoming source flux.
- PVD is characterized by a long mean free path and no surface migration, unlike CVD which exhibits rapid surface migration (thanks to the surface at high temperature, less dependent on the surface morphology).

1.6.15 PVD-Thermal Evaporation

Thermal Evaporation

The two primary evaporation techniques are Thermal and E-beam. In thermal evaporation, the source material is hosted in a **crucible** in which the heating current flows, and produces the vapor. Crucibles are made of different materials (W, Ta, Mo, C, etc.) to avoid contamination. Alternatively, the crucible can be made of a non-conductive refractory material (alumina, BN, etc.) with the shape of a small basket, heated by a spiral filament (W, Ta, etc.).

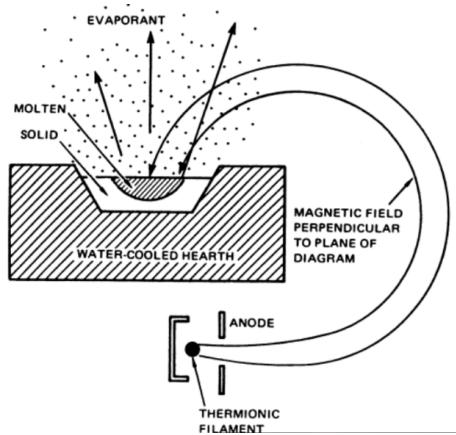
High deposition rate is achievable, but care must be taken for the choice of the crucible. The shape of the crucible allows to use different materials, to achieve heating with Joule effect.

Not every material can be evaporated (W is nearly impossible, Mo and Pt are very difficult for example).

For high throughput with uniform thickness, a rotating planetary wafer holder hosting lots of wafers can be used.

Material	Symbol	MP (°C)	S/D	g/cm ³	Temp.(°C) for Given Vap. Press. (Torr)			Evaporation Techniques				Sputter	Comments		
					10 ⁻⁸	10 ⁻⁶	10 ⁻⁴	E- Beam		Thermal Sources					
					Boat	Coil	Basket	Crucible							
Aluminum	Al	660		2.70	677	821	1010	Ex	TiB ₂ , W	W	W	TiB ₂ -BN, ZrB ₂ -BN	RF, DC	Alloys and wets. Stranded W is best.	
Aluminum Oxide	Al ₂ O ₃	2072		3.97	-	-	1560	Ex	W	-	W	-	RF-R	Sapphire excellent in E-beam; forms smooth, hard films. n = 1.66	
Chromium	Cr	1857	S	7.20	837	977	1157	G	**	W	W	VC	RF, DC	Very adherent. High rates possible.	
Copper	Cu	1083		8.92	727	857	1017	Ex	Mo	W	W	Al ₂ O ₃ , Mo, Ta	DC, RF	Adhesion poor Use interlayer (Cr). Evaporates using any source material	
Silicon	Si	1410		2.32	992	1147	1337	F	W, Ta	-	-	BeO, Ta, VC	DC, RF	Alloys with tungsten; use heavy tungsten boat. SiO produced about 4 x 10 ⁻⁶ Torr. E-beam best.	
Titanium	Ti	1660		4.5	1067	1236	1463	Ex	W	-	-	TiC	DC, RF	Alloys with refractory metals; evolves gas on first heating.	

E-beam Evaporation



Electron beam or Electron gun. A filament is heated up to electron emission; electrons are accelerated and deviated thanks to a magnetic field to hit the material in a **liner** (crucible), causing evaporation.

Liners are water cooled, preventing contamination from the liner (kept at low T) and the filament (magnetic deviation separates contaminants from electrons). Liners can host a greater amount of material, making long evaporation for thick films feasible.

Many materials can be evaporated in sequence without breaking vacuum (liners are hosted in a rotating housing). This allows for multi-layers (with no air or moisture contamination) or alloys.

High deposition rate (up to $0.5 \mu\text{m}/\text{min}$ for example), depending on the distance between the source and substrate.

Materials like W, Mo, Pt, ... can be evaporated using E-beam. Even having high melting point.

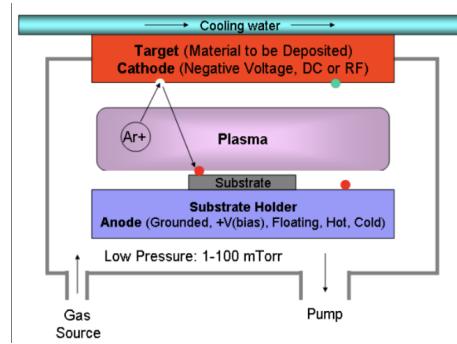
Example: Ti/Al/Au Ohmic Contacts, requiring an adhesion layer (e.g., $h_{Ti} = 20 \text{ nm}$) beneath the metal bulk ($h_{Al} = 500 \text{ nm}$), and a passivation layer

$(h_{Au} = 20nm)$. Alternative adhesion layer materials include Cr and Ta. We do that because we're interested in Al conductivity, but you are forced to put a multilayered substrate to exploit Al weaknesses (like oxidation). We use this tool because we need to put every element under vacuum, otherwise they'll be subject to oxidation (Ti will lose its adhesion properties).

How do we do this process on many wafers? this concept is extendable to every process. For evaporation we'd like the same deposition rate. The solution is a Planetary Substrate Holder, in a sphere, every wafer tangent to that, and geometrically works! (believe me)

We have a double rotation, see a video.

1.6.16 PVD- Sputtering



Sputtering is based on ion bombardment (uses plasma, usually Ar^+ , to sputter the target, dislodging atoms which then travel through the vacuum chamber and condense as a thin film on wafers).

Requires higher pressures than evaporation (1-100 mtorr), which carries possible greater contamination risks.

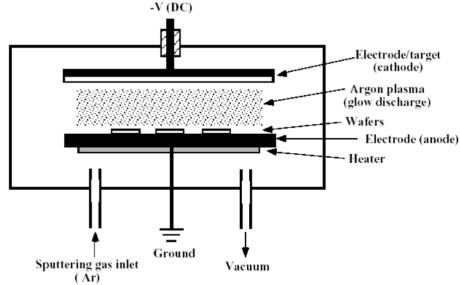
Offers unlimited material selection and better deposition of alloys/compounds than evaporation, resulting in replication of the target composition in the deposited films.

Provides excellent adhesion because adatoms are more energetic than in evaporation.

It is more expensive than evaporation (due to necessary voltage).

Glow discharges for sustainable plasma can be created with DC voltages or AC voltages (RF systems).

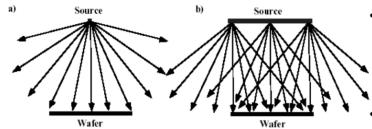
DC Sputtering



This is the simplest sputtering approach. The target material (cathode, negative potential) is eroded and deposits on the substrate (anode, ground or positive potential). Several 100 volts between plates ignite the plasma discharge, and positive ions are accelerated towards the target.

Sputtered atoms arrive at the substrate mostly as neutral atoms. The discharge is maintained as accelerated electrons continuously ionize new ions by collisions with the sputter gas.

Sputtering Conformality

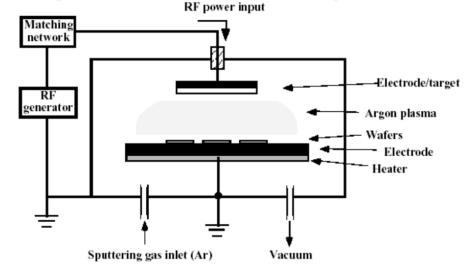


The conformality (step coverage) of standard sputtering is geometrically limited by each surface's line of sight to the target, like evaporation.

The difference is that in evaporation we have only one source, arriving in only one angle; in sputtering targets are generally large, each point is a source, and provide a wide range of arrival angles, (making it **more conformal than evaporation**).

Step coverage (η) is typically poor because the sticking factor of incoming species is ~ 1 (no surface diffusion). Step coverage worsens as deposition progresses because the top grows faster than the bottom. To improve step coverage, one can use CVD or heat the substrate to enhance surface diffusion.

RF Sputtering



For DC sputtering, the target electrode must be conductive; it's necessary to find an alternative (with a dielectric we risk to have, after bombarding with positive ions, after a while we'll have a positive surface, so the positive ions become useless).

The top electrode is smaller, in this way we have a higher electric field on the top electrode. This creates a potential that pushes the ions to the top. → the bottom electrode is connected to the metal box (?).

To sputter dielectric materials (non-conductive), an RF power discharge is used to avoid the accumulation of electric load (positive charge, Ar^+) on the cathode (target). Typically 13.56 MHz is capacitively coupled to the target. When frequencies are above 50 kHz, ions (heavy) can no longer follow the polarity switching, and electrons can neutralize positive charge buildup on each electrode during each half cycle, causing the plasma to bias positively with respect to both electrodes.

RF is better for insulators, DC for conductive.

Magnetron Sputtering

Is an upgrade you can apply to both RF and DC.

Premise: Standard plasma has a low ionization degree (< 1% of the atoms, is enough to be a powerful tool).

To improve the ionization rate (with more ions sputtering rate is increased, even if working at lower pressure (better quality and lower contamination)), magnetic fields are applied (perpendicular to the electric field). These fields force electrons into helical (longer) paths close to the cathode, leading to a much higher ionization probability (increasing the probability of the collision event).

As a result, deposition rates increase by 10-100x. It allows for higher sputter rates at lower Argon pressures (down to 0.5 mTorr). A disadvantage is more inhomogeneous target erosion than a simple planar geometry. Nowadays the magnetic field is put behind the target.

Reactive Sputtering

Reactive sputtering is a solution to deposit dielectric materials using DC sputtering or to adjust stoichiometry of a deposited material. A metallic or semiconductor target (e.g., Si) is used, and a reactive gas (O_2 or NH_3) is employed

in the plasma phase instead of Ar. Dielectric molecules (SiO_x or SiN_x) are produced and deposit on the substrate.

- **Non-Reactive Sputtering:** The sputtering gas is not incorporated into the film. Common choices are Ar, He, Ne, Kr, Xe, with Argon being the most popular.

- **Reactive Sputtering:** The sputtering gas is incorporated into the film to form alloys like Al_2O_3 , Si_xN_y , or SiO_2 . Common choices are O_2 , N_2 , NH_3 .

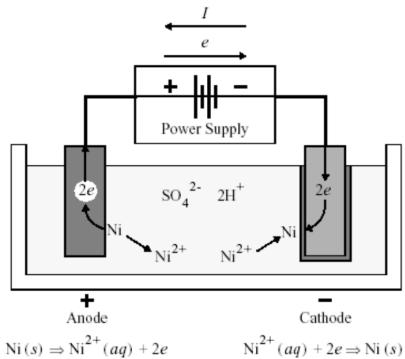
Ionized Sputtering (HDP Sputtering)

Is a modification to sputtering systems.

In these systems, the depositing atoms themselves are ionized (50-85% ionized). Why? you want an absolutely non conformal deposition. To achieve this, you want a vertical arriving angle. How? Transforming the neutral molecule sputtered to a charge.

An RF coil around the plasma induces collisions in the plasma creating the ions. This process provides a narrow distribution of arrival angles which may be useful when filling or coating the bottom of a deep contact hole. Example: VIA is a hole in a dielectric layer that must be filled in with metal that works like a vertical interconnection between two metal layers.(In this way we have the lower electrical resistivity so the higher quality)

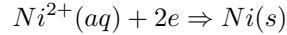
1.6.17 Electroplating – Electrodeposition-Galvanic Deposition



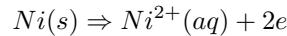
Electrodeposition is the oldest deposition technology, having been around for centuries. The simplest form requires a bath consisting of an electrolyte containing metal-ions, an electrode or substrate on which deposition is desired (cathode), a counter electrode (anode), and an external power supply. When a current flows through the electrolyte, cations move toward the cathode, and anions move towards the anode.

Processes

- **Cathodic Processes:** species are reduced by transfer of electrons from an electrode to an ion.



- **Anodic Processes:** species are oxidized by transfer of electrons to an electrode from an ion.



Characteristics and Applications

- Various metals can be deposited (Au, Ni, Cu, etc.), but not all (may be expansive or dangerous).
- It is **fast** ($\mu m/min$).
- Thicknesses $> 100\mu m$ are possible.
- It needs a seed layer (the entire loop must be electrically conductive).
- Electrodeposit can grow to connect isolated segments.
- 3-D Structures can be formed.
- The process relies on a patterned seed layer, involving the steps: Deposit Seed Layer (e.g., Cu seed layer PVD deposited) \rightarrow Define Seed Layer & Contact \rightarrow Electroplate Material.

The surface of the substrate MUST be **electrically conductive**. It's a mandatory requirement but not sufficient. They also MUST be **electrically connected**.

1.7 Implantation

1.7.1 Silicon Doping

Adding parts/billion to parts/thousand of "dopants" to pure Si can change resistivity by 8 orders of magnitude!

The key to building semiconductor devices lies in the ability to control the local doping and hence local electronic properties of a semiconductor crystal.

Approximate definition of doping levels:

- N⁻ or P⁻: $10^{14} \text{ cm}^{-3} < N_D \text{ or } N_A < 10^{16} \text{ cm}^{-3}$
- N or P: $10^{16} \text{ cm}^{-3} < N_D \text{ or } N_A < 10^{18} \text{ cm}^{-3}$
- N⁺ or P⁺: $10^{18} \text{ cm}^{-3} < N_D \text{ or } N_A < 10^{20} \text{ cm}^{-3}$
- N⁺⁺ or P⁺⁺: $N_D \text{ or } N_A > 10^{20} \text{ cm}^{-3}$
- Si# density: $5 \times 10^{22} \text{ cm}^{-3}$
- Intrinsic Si at RT: $n_i = 1.45 \times 10^{10} \text{ cm}^{-3}$

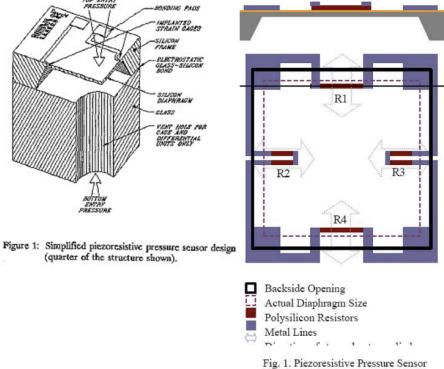
1.7.2 Dopant Species: P-Type, N-Type

Typical Dopants (gives or takes an e⁻) Doping atoms generally must sit on substitutional sites to be electrically active. Doping levels: 10^{15} to $10^{20} \text{ atoms/cm}^3$ (Silicon has $5.2 \times 10^{22} \text{ atoms/cm}^3$).

- Donor Atoms (n-type): P, As, Sb, ...
- Acceptor Atoms (p-type): B, Ga, Al
- P, As and B most used thanks to the greater solubility in Si ($> 5 \cdot 10^{20} \text{ cm}^{-3}$)
- Unwanted Dopants: Au, Fe, Cu, Ni, ...
 - can ruin solid-state electronic devices
 - fast diffusers (high diffusivity)

Doping is done to alter electrical, so physical, properties. But also Chemical!

1.7.3 Piezoresistive Pressure Sensor



In the center there's a membrane that reacts to differential pressure. The translation of pressure to an electrical signal is done thanks to a piezoresistor, it works thanks to mechanical pressure, changing the resistance, having a voltage proportional to the differential of pressure. Doped Si is a piezoresistor. You can alter Electro-mechanical properties.

1.7.4 Doping Parameters

Aim of the doping model characterization parameters is the prediction of a minimum set of 3 doping process

1. **Dopant Distribution:** $N(x)$ (or $C(x)$ depending on literature) [atoms/cm^3]

$$Q(t) = \int_0^\infty N(x, t) \cdot dx$$

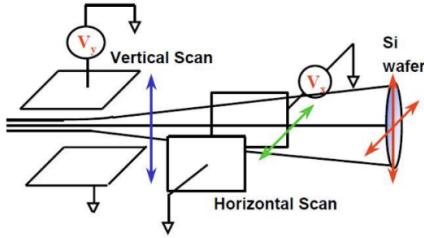
2. **Total Dose:** $Q(t)$ [atoms/cm^2] → is the integral of the Distribution.

3. **Junction Depth** → x_j [nm], [μm] is the intersection of the profile $N(x, t)$ with the background concentration N_B (or C_b), thus where $N(x, t) = N_B$.

1.7.5 Ion Implantation

- The purpose: To introduce doping atoms/impurities in substrates (e.g. silicon, GaAs, InP, SiC, diamond, ...)
- An ion implanter tool generates an ionic beam of **controlled species and energy**.
- In case of doping, the beam is directed toward the substrate scanning its surface to introduce the wanted quantity of element.

You basically spray dopant everywhere, it will act on non-masked areas.

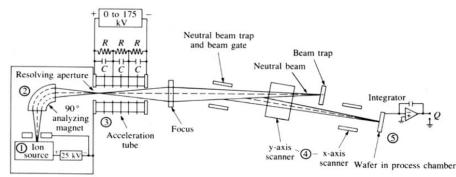


The idea was proposed by Shockley in 1954, but used for mass production only after late 1970s.

Electrostatically (through high potentials 3 kV - 3 MV) accelerate ionized dopant atoms to velocities and energies that can implant dopants below the surface of the substrate:

- process performed at low temperatures (room temperature – 200°C) → can use photoresist as a mask (being cheaper)
- instant-on and instant-off control
- precise control of implanting current and charge allow for better control of the implanted dose (Q_0) that can be monitored in situ
- increased implant energies can penetrate thin films of materials
- the peak of implanted dopant profiles are always below the surface (buried)
- mass separator allows each implanter to run multiple processes (B, P, BF_2 , As...)
- ISSUE: needs to be followed by an anneal
 - heal crystal damage (cause of the damage of the process)
 - activate dopant (put in substitutional sites from interstitial)

Implanter Components



- Ion source: generates dopant ions by electron cascading from tungsten filament, magnet forces e^- into spiral paths.

- Extraction region: pulls ions out of the source and focuses them toward the mass analyzer.
- Mass analysis: used to filter out undesired ions, by tuning magnetic fields to allow a specific m/q ratio to pass through.
- Accelerator: accelerates ions (electric field) toward the wafers ($E = qV$)
- Vacuum: need clean vacuum to reduce loss of ions through neutralization, scattering.
- Deflector plate: bend to remove neutralized species → Being trapped by a Beam Trap.
- Scanning system: x and y axis deflection plates are used to scan the beam across the wafer to produce uniform implantation of desired dose.

Why we want to eliminate any neutral element? Because they're uncountable! Every ion carries one charge, to count the number of ions it's sufficient to measure the current.

$$Q = \int_0^T \frac{I \cdot dt}{n \cdot q \cdot A}$$

dose in ions cm^{-2} → measured by the integrated beam current.

1.7.6 Gaussian approximation

At its heart Implantation is a random process→ Dopants follow a completely Gaussian distribution

- x = depth from wafer surface (μm)
- $N(x)$ or $C(x)$ = dopant distribution = dopant concentration at depth x
- N_p or C_p = peak dopant concentration
- Q = dose, total number of dopant ion implanted ($atoms/\mu m^2$)
- R_p = range of implant (or projected range), depth of highest concentration (μm) or average depth
- Variance (Vertical Straggle (or projected straggle)) = standard deviation of concentration) = $\pm \Delta R_p$
- Lateral Variance (Lateral Straggle (or transverse straggle)) = $\pm \Delta R_{\perp}$ → produces also a lateral implantation
- Gaussian asymmetry (Skewness)

$$N(r) = N_p \exp - \frac{(r - R_p)^2}{2\Delta R_p^2}$$

Dose:

$$Q = \int_0^\infty \frac{I \cdot dt}{n \cdot q \cdot A} = \int_0^\infty N(x) \cdot dx$$

Projected range:

$$R_p = \frac{\sum_i x_i}{N}$$

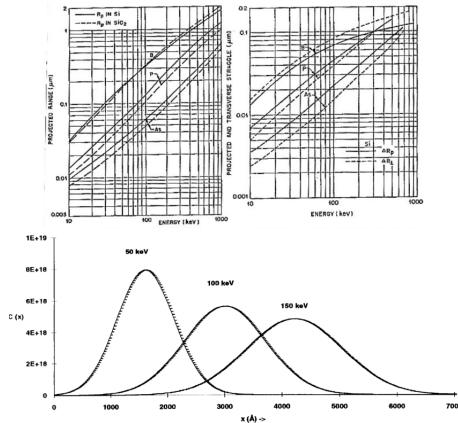
Straggle \rightarrow Order 2 moment:

$$\Delta R_p = \sigma_p = \sqrt{\frac{\sum_i (x_i - R_p)^2}{N}}$$

Gaussian distribution:

$$C(x) = \frac{Q}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{(x - R_p)^2}{2\sigma_p^2}\right) = N(x)$$

Projected Range and Straggle



Heavy ions, shallower and more vertical gaussian. Light ions, deeper and wider distribution.

1.7.7 Implantation Control Parameters

- Ion Energy
 - controls the depth and shape of implant
 - affects amount of crystal damage created
- Dose
 - control concentration \rightarrow many device parameters

- Ion Species
 - controls the type (n or p)
 - impacts profile depth and shape
- Mask Shape (layout) and thickness
 - controls in-plane geometry of the implant
- Annealing technique → time and temperature

1.7.8 Junction Depth (As Implanted)

Junction depth is the depth at which the concentration of implanted dopant atoms equals the substrate background concentration of dopant atoms.

$$C(x_j) = C_b \rightarrow C_b = \text{background doping}$$

$$X_j = R_p \pm \sqrt{2}(\Delta R_p) \sqrt{\ln\left(\frac{C_p}{C_b}\right)}$$

where

$$C_p = \frac{Q_0}{\sqrt{2\pi}(\Delta R_p)}$$

Note: R_p and ΔR_p are tabulated for various dopants and materials

1.7.9 Annealing - Activation

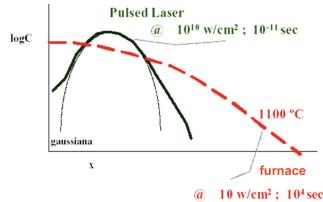
Annealing effects:

- Electric activation: Atoms after implant are distributed casually in the crystalline lattice. To activate their dopant properties they need to be brought to substitutional position
- Reorganize the atomic scale order
- Side effect → Diffusion of dopants and contaminants

The four thermal treatments are classified as:

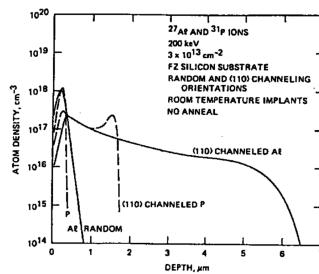
- Isothermal Annealing: the whole substrate maintains thermal equilibrium at every instant (heat transfer is slow with respect to substrate material thermal constants).
 - Furnace based system → Heat transfer mechanisms are dominated by heat conduction through environmental gas. Hours or minutes time scale process. Redistribution of dopant will be wider.
 - Rapid Thermal Process (or Rapid Thermal Annealing) → lamps based system (thermal gradient $\sim 150^{\circ}\text{C/s}$). Heat transfer mechanisms are dominated by electromagnetic radiation. Seconds time scale process. Redistribution of dopant will be limited.

- Adiabatic Annealing: thermal equilibrium is not maintained: heat transfer is faster than capability of substrate to dissipate energy.
 - E-beam anneal → Heat transfer by radiation. Time scale of few seconds or ms.
 - Laser anneal → Heat transfer by radiation. Time scale of ms or ns.



1.7.10 Channeling Ions Through Lattice

- In a monocrystalline material there is a significant probability that the implanted ions trajectory is aligned with crystallographic planes of low Miller indices (larger space). This results in a channel where ion is not subjected to nuclear scattering and the energy loss is mainly due to electronic interaction.
- In this condition, the interatomic potential itself contributes to keep the ion in the direction of the channel → CHANNELING
- The distance before the ion stop can be significantly larger than in the case of amorphous layer → results in $\sim 2\times$ increase in implant depth (no Gaussian distribution or better Gaussian distribution with exponential tail)



Minimizing Channeling

- Implant into a damaged target (Ar^+ pre-implantation to damage surface layers (how? using neutral ions- low energy, high dose))
- Implant through an amorphous layer (Silicon Oxide).
 - Not easy to implement in I.C. process flow

- Implant in direction not aligned with crystal channels (Twist and or Tilt)-is the most used.

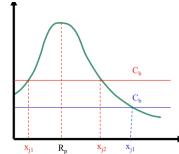
Twist and or Tilt: A tilt angle of $7^\circ - 8^\circ$ was found to be effective for a large variety of implant conditions. **Shadowing** effects must be considered when tilt angle is set. A shifted or a bigger mask is not the solutions, because we may have problems with horizontal vertical lines.

1.7.11 Sheet (or Square) Resistance

$$R = \frac{\rho \cdot L}{A} = \left(\frac{\rho}{t}\right)\left(\frac{L}{W}\right) = R_s\left(\frac{L}{W}\right) \rightarrow R_s = \left(\frac{\rho}{t}\right) = R_{\square}$$

- Resistance R [Ohm]
- Resistivity ρ [Ohm·m]
- Resistor Length L, Width W, Thickness t
- Sheet Resistance R_s [Ohm/square]
 - square is unit less (the aspect ratio of the resistor)

Sheet Resistance of an Implant



$$R_s = \left(\frac{\rho}{t}\right) = \frac{1}{\int_{x_\beta}^{x_j} \sigma(x) \cdot dx}$$

n-type: $\sigma(x) = q \cdot \mu(x) \cdot n(x)$

p-type: $\sigma(x) = q \cdot \mu(x) \cdot p(x)$

$$R_s = \frac{1}{\int_{x_0}^{x_{j/2}} q \cdot \mu(x) \cdot (N_{A,D}(x) - N_B) \cdot dx}$$

Numerical calculation is required since $\mu(x)$ is concentration dependent

1.7.12 Implanting Through a Mask

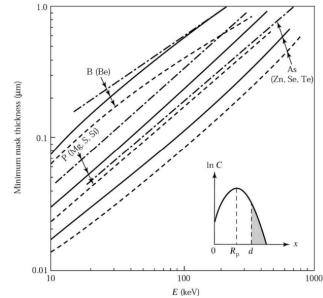
$$\text{Fraction Transmitted} = \frac{\int_d^\infty N(x) \cdot dx}{\int_0^\infty N(x) \cdot dx} \quad N(x) = N_p \cdot e^{-\left(\frac{x-R_p}{\Delta R_p \sqrt{2}}\right)^2}$$

$$\int_z^\infty e^{-a \cdot u^2} \cdot du = \frac{1}{2} \sqrt{\frac{\pi}{a}} \operatorname{erfc}(z\sqrt{a}) \rightarrow \int_z^\infty N_p \cdot e^{-\left(\frac{x-R_p}{\Delta R_p \sqrt{2}}\right)^2} \cdot du = N_p \Delta R_p \sqrt{\frac{\pi}{a}} \operatorname{erfc}\left(\frac{z-R_p}{\Delta R_p \sqrt{2}}\right)$$

Resulting in:

$$\text{Fraction Transmitted} = \frac{1}{2} \operatorname{erfc}\left(\frac{d-R_p}{\Delta R_p \sqrt{2}}\right)$$

1.7.13 Masked Implantation



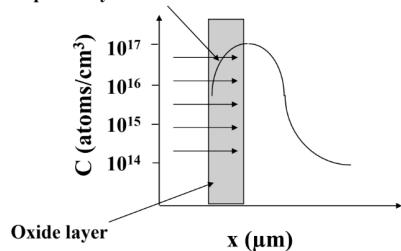
May and Sze, 2004

Minimum thickness of SiO_2 (—), Si_3N_4 (- - -), and a photoresist (——) to produce a masking effectiveness of 99.99%. (....)=dopants in Si.

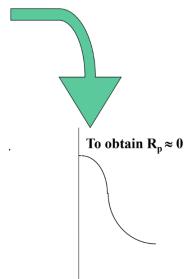
1.7.14 Screen Oxides

IMPLANTATION

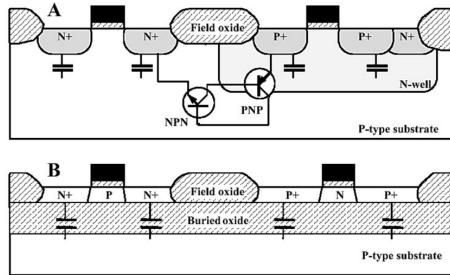
Some dopant stays in the oxide



OXIDE ETCHING



1.7.15 Silicon on insulators (SOI) wafers



Silicon based devices are subjected to problems due to electrical parasitic elements connected to junctions capacitances. In GaAs these problems are minimized by using semi-isolant substrates (for example Cr doped).

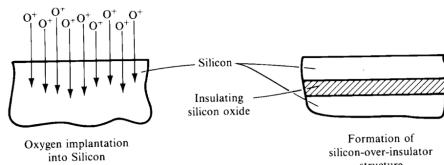
Two options:

- Silicon on Sapphire (SOS) (older technology, seldom used today)
- Silicon on Insulator (SOI), usually an SiO_2 film

SOI (Silicon-On-Insulator)

- Fabricated with at least 3 different process
 - SIMOX, BESOI, Smart-Cut[®]
- Specify the same criteria as with plain wafers
 - Size, Orientation, Doping, Doping Level, Grade, etc.
- Thickness of upper silicon (Device Layer)
 - Typically $\sim 2 \mu m$ for ICs but can be up to $100 \mu m$ for MEMS
- Thickness of oxide
 - Typically ~ 0.5 to $2 \mu m$
- Cost:
 - Expensive: $\sim \$100$ to $\$500$ / wafer

1.7.16 SIMOX

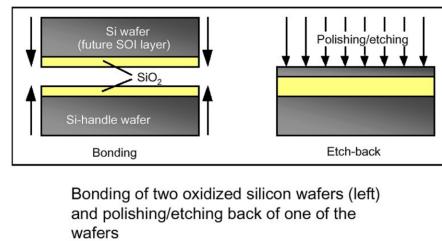


By the early 1990's, a new SOI method called Separation by Implantation of OXYgen (SIMOX) had begun to gain acceptance because it could be made with nearly-standard processing equipment and there was at least one commercial supplier of the wafers.

There are still current users of SIMOX wafers but increased device performance requirements and costs are forcing SIMOX out of the marketplace.

1.7.17 BESOI

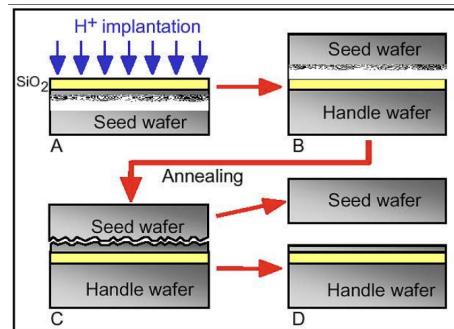
A second method based on wafer bonding and thinning, called Bond and Etch-back Silicon On Insulator (BESOI), was also contending for the SOI market. In this technique, two wafers are bonded together. One or both of the wafers has an insulating oxide on the surface. After bonding, almost all of one wafer is polished away, leaving a structure with a thin top silicon layer, a buried oxide and a full wafer below.



1.7.18 Smart Cut

- Smart Cut is Soitec's proprietary technology, used to manufacture the company's UNIBOND SOI wafers.
- Based on two key techniques—ion implantation and wafer bonding—the process begins with two bulk silicon wafers, one of which is reused to create a SOI wafer later on.
- The UNIBOND wafer is created by growing an oxide layer on one wafer (Wafer A). This oxide layer subsequently forms the bulk-oxide (BOX) layer of the SOI structure.
- Ion implantation of high dose hydrogen ions through the oxide into the underlying silicon forms a damage layer at the end of the ion's range.
- Wafer A is bonded to Wafer B. Wafer A is then cut across the damage plane (heat to break the layer off), and a thin layer of silicon is transferred to Wafer B to form the SOI structure.
- Finally, a CMP polish finishes the SOI surface to the original bulk-silicon wafer specification.

- H^+ ion implantation acts as an atomic scalpel in the Smart Cut process, enabling thin slices of monocrystalline film to be cut from a donor wafer and transferred on top of a receiving wafer. The process provides optimal usage of valuable material by placing the amount of materials needed for electronic functionality on top of a very inexpensive support layer.



1.8 Wet & Dry Etching

1.8.1 Etching

- Etching is the process in which unwanted areas of films/bulk are removed
 - Dissolve them in a wet chemical solution **WET ETCHING**
 - React them with gases (thermal or plasma) to form volatile products
→ **DRY ETCHING**
- Photoresist protects areas which are to remain.
- In some cases a "hard mask", usually patterned layers of SiO_2 or Si_3N_4 , but also metals (Cr, Al, ...), is used
 - When the etch selectivity to photoresist is low
 - Etching environment causes resist to delaminate

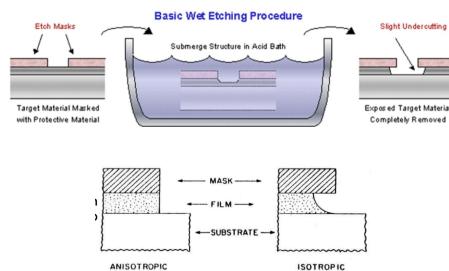
Wet: Usually acid bath.

- Reactions occur at liquid-solid interface.
- Immersion - whole wafer placed in etch bath.
- Spray- chemicals sprayed onto wafer on a spinning chuck.
- Chemical mechanical polishing (CMP)

Dry: Usually a plasma

- Plasma or Reactive Ion Etching (RIE)
- Reactions occur at gas-solid interface.
- Gaseous by-products are required for dry etching.

1.8.2 Wet Etching



Isotropic etchants (e.g. HNA) give rounded profiles

- best to use with large geometries, when sidewall slope does not matter, and to undercut the mask
- quick, easy, cheap

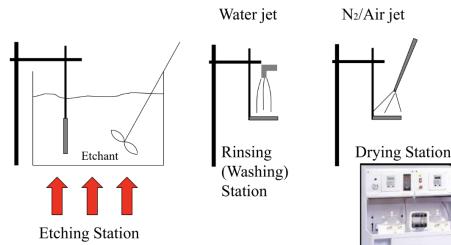
Anisotropic etchants (e.g. KOH, TMAH, EDP) etch preferably $<100>$ planes and slow down markedly on $<111>$ crystal planes of silicon

- best for making small gaps and vertical sidewalls (or sidewalls with a precise angulation)
- typically more costly

You obtain etching with a specific angle, see the image, based on the plane

Steps in Wet Etching

- Etching station
- Rinsing(Washing) Station (Probably the most critical step for some devices)
- Drying Station



Mechanical agitation required to keep fresh etching solution in contact with wafer surface.

Surfactant added to promote wetting and remove/inhibit by-product gasses clinging to wafer (H₂ usually).

1.8.3 Etchant properties

Etch rate Depends on the etchant and the material that is been etched; but also:

- rate of material removal ($\mu\text{m}/\text{min}$)
- function of concentration, mixing, temperature, material, layout, ...
- Arrhenius relationship

Must be controlled and repeatable. Also the movement must be controlled: if too fast, it's more difficult to control.

Selectivity to masking layer(s) and their availability. When you perform etching, you want to etch ONLY the target material (ideally). Because it will also etch the mask and the underlying layer, you must calculate the necessary time to avoid that.

- relative (ratio) of the etch rate of the film to the mask, substrate, or another film
- trade off etch rate and selectivity

Film:Mask Selectivity

$$S_{f:m} = \frac{R_{film}}{R_{mask}}$$

High value needed to minimize linewidth resolution loss

Film:Underlying Film Selectivity

$$S_{f:u} = \frac{R_{film}}{R_{under}}$$

High value needed to minimize overetch material loss

Anisotropy (crystal plane selectivity) Is a parameter you consider only when etching a monocrystalline material. For example KOH has a relative etch rate of 40:30:1 respect to $<110>$, $<100>$ and $<111>$ crystal planes (relative rates can dramatically change with temperature),

Surface roughness

Control of etch parameters (according for example to temperature, agitation conditions,...)

Safety of reactant(s) and product(s)

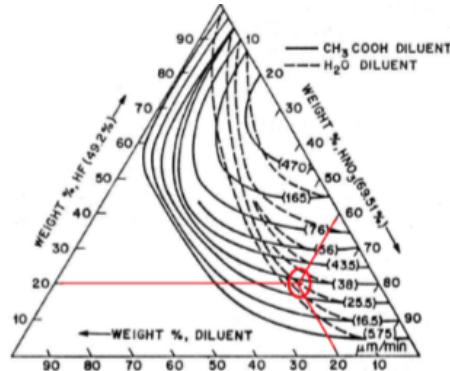
Cost (including disposal and fixed costs) Disposal is usually the higher cost, since it is calculated on volume!

Capacity for etch-stops (simply time controlled or concentration of the solution, doping level, doping dependant or electrochemical etch-stop)

1.8.4 Wet etching silicon ("HNA")

- mixture of nitric (HNO_3) and hydrofluoric (HF) acids
- HNO_3 oxidizes Si, HF removes SiO_2 , repeat...
- high HNO_3 : HF ratio (etch limited by oxide removal)

- low HNO_3 : HF ratio (etch limited by oxide formation)
- dilute with water or acetic acid (CH_3COOH)
- acetic acid is preferred because it prevents HNO_3 disassociation
- **fully isotropic**



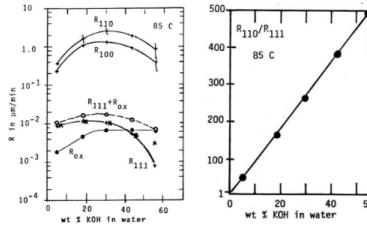
H. Robbins and B. Schwartz, "Chemical Etching of Silicon II, the System HF, HNO_3 , H_2O , and $HC_2H_3O_2$ ", J. Electrochem. Soc., v. 107, p. 108 (1960)

1.8.5 Anisotropically wet etching Silicon

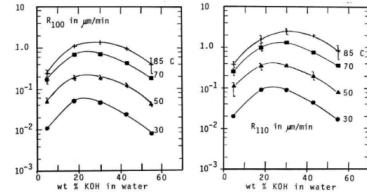
Inorganic Alkaline Solutions (KOH)

- KOH pellets dissolved in water (reaction produces H_2)
- The quickest silicon anisotropic etchant, but also the less controllable
- Not compatible with Microelectronics (CMOS -in particular)
- SiO_2 Si_3N_4 (better) used as masking materials (resist will not survive, oxide is attacked slowly)
- Simple set-up: hot plate, Pyrex beaker and stirrer (improve roughness eliminating H_2)
- High concentrations are needed to limit roughness
- Etch rate $\approx 1 \mu\text{m}/\text{min}$, but strongly dependent on solution concentration and temperature (typically used around $80^\circ C$)

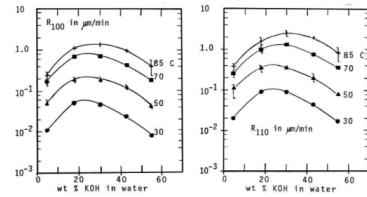
Etch rate of Si in KOH depends on crystallographic plane (anisotropy)



Etch rate of Si in KOH depends on temperature



KOH etch slowly silicon oxide (selectivity)



1.8.6 HYDROFLUORIC ACID (SiO_2)

- Selective (room temperature)
 - etches SiO_2 and not Si
 - will also attack Al
- Rate depends strongly on concentration
 - maximum: 49% HF (“concentrated”) $\sim > 2 \mu m/min$
 - controlled: 5 to 50:1 (“timed”) $\sim < 0.1 \mu m/min$
 - etch rate is not linear with high concentration
- Dangerous!
 - not a strong acid (check the pH of your HF etch)
 - deceptive (looks just like water)
 - penetrates skin (adsorption) and attacks slowly
 - will target bones

- Etch Geometry:
 - completely isotropic (used to undercut / release)
- Buffering with NH_4F keeps pH constant (a buffer agent maintains narrow range of pH during life of etch bath) and thus stabilizes etch rate (even lowering it)

BOE (Buffered Oxide Etch) or BHF (Buffered HF)

1.8.7 What is a Dry Etch?

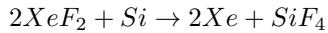
- In wet etchants, the etch reactants come from a liquid source
- In dry etchants, the etch reactants come from a gas or vapor-phase source and are typically ionized (atoms or ions from the gas are the reactive species that etch the exposed film) → 2 kinds of dry etches: plasma and non-plasma based

Why use Plasmas?

- Perform reactions at lower temperatures than thermally driven reactions → PECVD
- Can provide directionality of processing.
- Several more parameters to play with to obtain desired properties → higher versatility

1.8.8 Vapor etching

- A nonplasma, isotropic dry-etch process for silicon is possible using XeF_2 and provides very high selectivity for aluminum, silicon dioxide, PSG, silicon nitride, and photoresist.
- These properties make it an extremely useful etchant, although the etched surfaces produced are quite rough.
- In a bell-jar setup run at 1 torr, XeF_2 can be sublimed from its solid form at room temperature and this etch has excellent selectivity with respect to CMOS process layers.



- The heat generated by this exothermic reaction may adversely affect some microstructures. An important concern is that XeF_2 reacts with water (even moisture in air) to form Xe and HF (the latter can also unintentionally etch silicon dioxide in addition to being a safety hazard).
- Etch rates are generally $1 - 3 \mu\text{m}/\text{min}$ (FAST !!!)

Neutral Molecules at a density of	$10e16/cm^3$
Radicals	$10e14/cm^3$
Electrons	$10e8/cm^3$
Positive ions	$10e8/cm^3$

Table 1.2: Content of typical plasma

HF-vapor phase etching

The HF vapor etcher consists of a reaction cell and heated wafer holder, in addition to an electronic control unit. HF vapor is generated passively from a small liquid reservoir at the bottom of the reaction cell. Only a minimal amount of about 60-80 ml of liquid HF is required. The HF Vapor Etcher Basic is perfectly adapted for R&D work in surface micromachining, SOI-MEMS, for dicing-free release, structure thinning, and many other applications. Safety is an important issue when working with HF. AMMT's engineers have designed this easy-to-use and straight forward etching system with a maximum of security including special.

1.8.9 RF plasma-based dry etching

- Chamber is evacuated
- Chamber is filled with gas(es)
- RF energy is applied to a pair of electrodes
- Applied energy accelerates electrons increasing kinetic energy
- Electrons collide with neutral gas molecules, forming ions and more electrons
- Steady state is reached (plasma); ionization recombination
- RF: Frequencies (typically 13.56 MHz)
 - Electrons oscillating in the glow region acquire enough energy to cause ionization (the alternating potential polarity affects the electrons, sending them into an oscillatory path).
 - Gas ions are too massive to respond to the high-frequency field (ions, being much heavier and therefore slower, don't keep up with the alternating field).
- Similar to PECVD except that etch gas is used instead of precursor gas.
- Etching is isotropic and selective due to the strong chemical component.

1.8.10 TODO:Types of Dry Etching

Neutral

Volatile product

Neutral Ion +

Neutral Ion +

Volatile product

Volatile product

Chemical PLASMA isotropic, chemical, selective

Ion-enhanced energetic RIE directional, physical & chemical, fairly selective

DRIE Like RIE, but high Aspect Ratio is the target

Ion-enhanced inhibitor

Inhibitor

e

TODO:Reactive Ion Etching (RIE)

TOP ELECTRODE

N

PLASMA

R^{\bullet}

Si wafer

BOTTOM ELECTRODE

ion bombardment

|+

- in both plasma etching and RIE feed- gas composition produces the reactive species necessary for etching

chemistry tends to be isotropic

ion bombardment of surfaces generates the anisotropy in plasma assisted pattern transfer

damage

surface damage

Examples of using Electric Fields for Enhanced Sidewall Abruptness (Anisotropy) in a Plasma Etch System

Mask Material

15KU

H10.866

NRC, IMS

InP etched by CAIBE

0000

12.8KX 25X0 WD 00000 200001 201

Low Electric Field

5 μm

High Electric Field

TODO:Inductively Coupled Plasma (ICP-RIE) Reactor

101

00

Inductive coil

~

RF generator

Two RF power generators to control ion energy and ion density separately

Plasma chamber

Dielectric window

Electromagnet

Biased wafer chuck

FIGURE 16.21 Inductively-Coupled Plasma Etch

Dual plasma source:

Bias RF generator

~

Redrawn from Y. Lii, "Etching," ULSI Technology, ed. by C. Chang and S. Sze (New York: McGraw-Hill, 1996), p. 351.

Operating pressure >5 mTorr and <60 mTorr

- Top RF current (typically 13.56 MHz) generates an opposite inductive current in the plasma which sustains the discharge
- Bottom RF power (typically 13.56 MHz) is responsible for ion bombardment
- Advanced etcher for (poly)silicon etching and aluminum etching

+

RIE

Top one (ICP RF power) generates high-density plasma, determines ion density/current.

Bottom one (CCP RF power) generates bias voltage like regular RIE, determines ion energy.

00

ICP

13,56 MHz

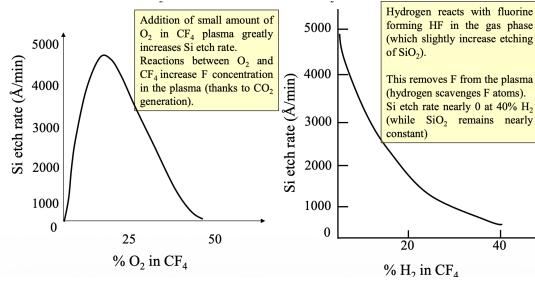
13,56 MHz

13,56 MHz

1.8.11 DRY ETCH CHEMISTRIES

Material	Etch Gases
Si, SiO_2 , Si_3N_4	CF_4 , SF_6 , CHF_3 NF_3
Si	Cl_2 , CCl_2F_2
SiC	CF_4 SF_6 , CHF_3
Al	BCl_3 , CCl_4 , $SiCl_4$, Cl_2
Organics	O_2 , $O_2 + CF_4$
Other: (W, Ta, Mo, ...)	CF_4

1.8.12 TODO:Selectivity



- CF₄ plasmas processes can be used to etch Si selectively to SiO₂ and SiO₂ selectively to Si

Addition of H₂ or O₂ enhances etch rates of one film over the other

- Small additions of O₂ enhances Si etch rate more than SiO₂ etch rate
- Addition of H₂ decreases Si etch rate more than SiO₂ etch rate

Addition of small amount of O₂ in CF₄ plasma greatly increases Si etch rate. Reactions between O₂ and CF₄ increase F concentration in the plasma (thanks to CO₂ generation).

Hydrogen reacts with fluorine forming HF in the gas phase (which slightly increase etching of SiO₂).

This removes F from the plasma (hydrogen scavenges F atoms). Si etch rate nearly 0 at 40% H₂ (while SiO₂ remains nearly constant)

The only etch rate that is being modified is the chemical, no physical component is touched. The only way to change the physical component is by changing the voltage.

1.8.13 TODO:Anisotropy in Dry Etching

$$\frac{x}{z} = \frac{R_x}{R_z}$$

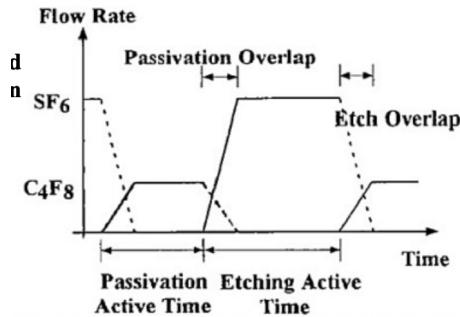
- When H₂ is added to CF₄, the etch rate (isotropic) of Si is decreased(lowering chemical component)
- When bias is applied, the etch rate (anisotropic) is increased(enhance the physical component)
- By combining these two effects, a good anisotropic etch can be achieved

Useful when you have to transfer small portions.

TODo:ASPECT RATIO

$$AspectRatio = \frac{Height}{Width}$$

1.8.14 Deep reactive ion etching (DRIE) → Time multiplexed deep etching



- The etching step uses SF_6 with a substrate bias of 5 to 30 V so that the cations generated in the plasma are accelerated nearly vertically into the substrate being etched.
- After etching for a short time, the polymerization process is started. A mixture of C_4F_8 (high C/F ratio) and argon is used and all exposed surfaces (side walls and horizontal surfaces) are coated with a **Teflonlike** (polymerized CF_2) polymer layer approximately 50 nm thick.
- If ion bombardment, due to a small applied bias voltage, is used during the polymerization step, the formation of polymer on the **horizontal** surfaces can essentially be prevented, and can also be eliminated if formed.
- The etching step is then repeated, and the polymer deposited on the horizontal surfaces is rapidly moved due to the ion bombardment and the presence of reactive fluorine radicals.

Summarizing:

1. Start
2. Isotropic SF_6 Etch With Anisotropic Bombardment
3. Isotropic Polymer Formation with C_4F_8
4. SF_6 Etch Again

Showing the sequential steps during TMDE. In (a) the masking material (photoresist or silicon dioxide) has been patterned. During the etch step in (b), a shallow, isotropic trench is formed. In the subsequent passivation step (c), a protective fluorocarbon film is deposited everywhere. During the next etch step (d), the fluorocarbon film is removed from all horizontal surfaces by directional ion bombardment, and another shallow trench is formed.

1.9 Packaging

Packaging is a critical stage in the manufacturing of integrated circuits. The process starts after the wafer has been fabricated and tested.

1.9.1 Packaging Process Flow

The process generally follows these steps:

1. **Wafer Test:** The finished wafer undergoes testing. Defective dies are electrically identified.
2. **Defective die inked:** The defective dies are marked (inked) to distinguish them from functional dies, the eventually eliminated.
3. **Wafer Saw:** The wafer is cut into individual dies.
4. **Package only "good" die:** Only the dies that passed the test are prepared for packaging.
5. **Final Test:** The packaged integrated circuit undergoes a final test.

1.9.2 Steps in Packaging

The physical packaging process involves several key micro-assembly steps:

Die Separation and Handling

- **Saw wafers into dies:** The process of cutting the wafer into individual chips.
- **Wafers mounted on mylar tape in stainless steel ring:** This is done to hold the dies in place during the sawing and subsequent handling.

Die Bonding and Interconnection

- **Bond good dies into package:** The selected good die is attached to the package's heat sink or base. This is the **Die bond**.
- **Wire bond die pads to package leads:** This process, known as **Wire Bond**, connects the on-chip aluminum or gold bond pads to the package's external leads using Au or Al wires.
- This bonding is typically achieved through Heat or ultrasonic bonding.

Package Sealing (Encapsulation)

- **Seal package:** This step covers the die, providing protection and sealing the package.
- **Ceramic/metal packages** use a metal lid and solder for sealing.
- **Plastic packages** use encapsulant or injection molding, which must be cured in an oven.

Final Inspection

- **Inspect:** The final packaged components are inspected for quality.

1.9.3 Functions of Packaging

Packaging provides several critical functions necessary for a chip to operate reliably within an electronic system:

1. **Power Distribution:** Distributes power from the circuit board to the chip.
2. **Signal Distribution:** Distributes electrical signals to and from the chip.
3. **Dissipates Heat:** Manages thermal load. This involves structural and materials concerns to ensure the chip does not overheat.
4. **Houses and Protects Die:** Provides protection against mechanical stress, chemical attack, and electromagnetic interference.

1.9.4 Package Types

Packages are classified primarily by how they interface with the Printed Circuit Board (PCB).

Types of First-Level Packages

Through Hole Package These packages have leads that are inserted through holes in the PCB.

- **DIP (Dual In-line Package):** Two lines of leads electrically connect the die to the circuit board. Is a low performance package, but is simple and cheap
- **SH-DIP (Shrink DIP)**
- **SK-DIP, SL-DIP (Skinny DIP, Slim DIP)**
- **SIP (Single In-line Package):** The hole and the metal surface are usually attached to a heat dissipator. It is commonly used in power electronics.
- **ZIP (Zig-zag In-line Package)**

Surface Mounted Package These packages are mounted directly onto the surface of the PCB.

- **SO or SOP (Small Out-line Package)**
- **QFP (Quad Flat Package)**
- **LCC (Leadless Chip Carrier)**: Chip carrier with no leads, metal lines run along package sides.
- **PLCC, SOJ** (Plastic Leaded Chip Carrier with Butt Leads)
- **TAB (Tape Automated Bonding)**
- **CSP (Chip Scale Package)**

Pin Array Packages Electrical contact is made through an array of connections on the underside of the package. The advantage is a higher density of connections compared to traditional leaded packages, enabling more compact designs and better electrical performance. The difference between PGA and BGA is the shape.

- **PGA (Pin Grid Array)** or Column Package: Pin array where contacts are pins.
- **BGA (Ball Grid Array)**: Pin array where contacts are solder balls.

1.9.5 Advanced Interconnection and Integration

Flip-Chip Bonding

The flip-chip method is an alternative to wire bonding where the chip is inverted (flipped) and connected directly to the substrate using solder bumps.

In this method, instead of using long gold wires, small pin contacts are utilized to establish electrical connections between the chip and the substrate. These pin contacts are directly bonded to the solder bumps on the chip, providing a compact and efficient interconnection. This approach reduces parasitic inductance and resistance, improving the overall electrical performance of the package.

The Pb : Sn alloy system is critical for many solder processes. The phase diagram (Figure 6.3) shows a eutectic composition at 61.9% Sn melting at 183°C (362°F).

Multi-Chip Modules (MCMs)

The need for high density, high performance, and cost-effectiveness has spurred the development of Multi-Chip Modules (MCMs)(today called SiP). MCMs enable the integration of different chips with varied functions into a mini substrate, instead of inserting individual packages onto a large PCB (also known as second level package).

Figure Placeholder: Flip-Chip Bonding Diagrams and Phase Diagram

Diagrams showing the structure of a flip-chip bond (contact to active region, passivation, solderable layer) and the tin-lead alloy phase diagram (Figure 6.3) with the 183° C eutectic composition.

Figure 1.10: Illustration of Flip-Chip bonding structure and the Pb : Sn phase diagram.

Definitions An MCM can be defined in a couple of ways:

- A structure consisting of two or more integrated circuits electrically connected to a common circuit base and interconnected by conductors in that base.
- A structure in which a **packaging efficiency** of greater than 30% is achieved. **Packaging efficiency** is the ratio of silicon die area to printed circuit board area. This definition implies a particular technology that allows chips to be packed closely together.

Goals The basic idea behind developing MCM technology is to decrease the average spacing between ICs in an electronic system. The goals are:

- Higher performance resulting from reduced signal delays between chips.
- Improved signal quality between chips.
- Reduced overall size.
- Reduced number of external components.

Structure An MCM substrate can be composed of different layers, and their number depends on the MCM technology used. These layers provide all the interconnections between the different mounted ICs.

3D Packaging

As a result, 3D packaging technology has evolved as a natural progression from the 2D packaging technology (MCMs).

3D packaging is achieved by stacking multiple chips vertically, one on top of the other. This stacking is facilitated by Through Silicon Vias (TSVs), which provide vertical electrical connections between the layers. The stacked configuration minimizes the footprint on the PCB while enabling high-density integration

Figure Placeholder: Multi-Chip Module Structures
Diagrams showing a generic MCM package with chips attached via Wire Interconnection and Flip-chip bonding, resting on a multilayer substrate with terminals for Second Level Connection.

Figure 1.11: Structure of a typical Multi-Chip Module (MCM) showing first and second level connections.

and improved performance. This approach is particularly useful in applications requiring compact designs, such as mobile devices and high-performance computing systems. As the requirements for high performance systems continually increase, even MCM technology is often not enough.

Advantages

- Footprint used by the 3D device is much smaller than the footprint used by every chip in a 2D MCM layout.
- **Through Via Connection (TSV)**: Provides short interconnect length, high density interconnect, and a low profile (without a spacer).

Through Silicon Vias (TSV) TSV via diameters typically range from $5 \mu\text{m}$ to several tens of μm . They provide vertical electrical connections through the chip.

Figure Placeholder: 3D Packaging Diagrams
Diagram showing stacked dies (e.g., ASIC and MEMS die) interconnected using Through Via Connection (TSV) for high-density, low-profile packaging.

Figure 1.12: Conceptual diagram of 3D Packaging utilizing Through Silicon Vias (TSV).

1.10 Complementary Technologies

1.10.1 Bulk micromachining

Bulk micromachining is a foundational technique in micro-electro-mechanical systems (MEMS) fabrication.

Definition and Characteristics

Bulk micromachining involves the selective removal of the "bulk" substrate material to release suspended three-dimensional structures.

- It is the most common micromachining technique and was one of the first to be developed, with initial experiments dating back to the 1970s for the fabrication of micro pressure sensors.
- While it is generally less versatile and offers lower resolution compared to surface micromachining, it allows for the fabrication of structures with relatively "big" mass.

Types of Bulk Micromachining

This technique is broadly categorized based on the side of the wafer being etched:

- Front-Side Bulk Micromachining
- Back-Side Bulk Micromachining

Typical resulting structures include square membranes, heating resistors, thermopiles, and piezoresistors. The process often involves layers of materials such as Si_3N_4 , SiO_2 , and the Silicon (Si) substrate, with typical dimensions in the 100 μm to 500 μm range.

Bulk Micromachining Etching Techniques

The removal of the bulk substrate can be achieved through various etching methods.

Wet Etching

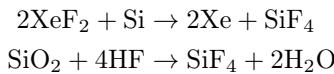
Wet etching involves submerging the substrate in a liquid chemical bath.

Basic Wet Etching Procedure The procedure follows these basic steps:

1. The target material is masked with a protective material (e.g., SiO_2 or Si_3N_4 are commonly used etch masks).
2. The structure is submerged in an acid bath.
3. The exposed target material is completely removed.
4. The process often results in slight undercutting beneath the mask material.

Vapour Etching Vapour etching is a form of chemical etching where the etchant is in the gas phase. A common example is the use of Xenon Difluoride (XeF_2) for silicon etching.

The key chemical reactions include:



Dry (Plasma) Etching

Dry etching, or plasma etching, utilizes chemically reactive species generated in a plasma state. This technique is typically performed inside a vacuum chamber.

Plasma Etching Components The general setup for plasma etching, such as in a Reactive Ion Etching (RIE) system, includes:

- Top and bottom electrodes
- A vacuum pump to maintain the low-pressure environment
- A source for etching gases
- An RF power source to generate and maintain the plasma
- Cooling (e.g., water) to regulate the temperature

Techniques The two most prominent dry etching techniques are:

- **RIE (Reactive Ion Etching):** Provides a balance of chemical and physical etching, resulting in good anisotropy.
- **DRIE (Deep Reactive Ion Etching):** A highly anisotropic process capable of etching deep, vertical structures in silicon, often utilizing the Bosch process.

1.10.2 SPM Machining

Scanning Probe Microscopy (SPM) techniques are highly versatile and offer ultra-high resolution capabilities for both additive and subtractive manufacturing at the nanoscale.

Thermal Scanning Probe Lithography

Thermal Scanning Probe Lithography (t-SPL) is a method that uses a heated AFM tip to locally modify a polymer resist layer, allowing for the patterning of features with nanometer resolution.

Scanning Probe Metal 3D Printing

Scanning Probe Metal 3D Printing, particularly demonstrated by the CERES system, enables true additive manufacturing with submicron resolution.

The CERES System The CERES system, developed by Exaddon AG, performs direct 3D printing of metal under ambient conditions. The system is capable of building complex structures through an additive manufacturing process.

Voxel-Based Printing Process The metal printing process is voxel-based, where voxels are elementary three-dimensional blocks.

1. **Setup:** A small printing nozzle, called an iontip (a type of cantilever), is immersed in a supporting electrolyte bath.
2. **Flow:** A precisely regulated air pressure pushes a liquid containing metal ions through a microchannel inside the iontip. The liquid flow is highly controlled, reaching rates as low as femtoliters per second.
3. **Deposition:** At the end of the microchannel, the ion-containing liquid is released onto the print surface. The dissolved metal ions are then electrodeposited into solid metal atoms.
4. **Structure Building:** Stacked in a layer-by-layer sequence at defined coordinates, these voxels form the desired 2D or 3D geometry.

Design Freedom and Feedback The use of electrodeposition enables significant design freedom, as free-standing structures and 90° overhanging angles are feasible without the need for support structures.

The printing process is automated by real-time feedback of the deflection of the iontip. As a voxel reaches completion, the top side of the voxel interacts with the tip, deflecting the cantilever by a minute, measurable amount. This mechanism operates much like an Atomic Force Microscope (AFM) cantilever operating in contact-mode. Upon detection of the deflection, the tip is quickly retracted to a safe traveling height and moved to the next voxel coordinate.

Chapter 2

Part 2: Cryogenic technology and Superconductivity

2.1 Introduction

2.1.1 Elements of classical thermodynamics

Here, we recall the basic concepts of classical thermodynamics (surely analyzed in detail in previous courses), with the aim to establish a proper ‘language and proper terms useful later to describe the cryogenic processes. Among key concepts: the laws of thermodynamics, entropy and enthalpy functions.

Thermodynamics is the branch of Physics that describes the transformations undergone by a system as a result of a process of **energy exchange** with other systems or with the external environment.

Classical thermodynamics is based on the concept of macroscopic **system**, i.e. a portion of matter physically or conceptually separated from the **surroundings** by a **boundary**.

Microscopic vs. macroscopic points of view

A **microscopic** description of a system involves assumptions about the internal structure of the system and then calculations of system-wide characteristics enormous number of interacting particles, mathematics of probability statistical mechanics. In brief:

- **assumptions** are made concerning the structure of matter, fields, radiation
- **many quantities** must be specified to describe the system, as suggested by our mathematical models rather than by our sensory perceptions
- they cannot be measured directly, they must be **calculated**

A **macroscopic** description involves the specification of a few, fundamental (suggested by our perceptions) and **measurable** (in general, directly measurable) properties, without special assumptions concerning the structure of matter,..

The state of a macroscopic system at equilibrium is specified by quantities called **thermodynamic coordinates** (or variables) such as temperature, pressure, volume, etc.

→ **Equation of state:** a mathematical function relating the coordinates of a system in equilibrium, specific for each system (**time** is not directly involved)

Thermodynamics is the branch of Physics that deals with the macroscopic properties of nature and always includes the macroscopic coordinate of temperature for every system

Examples of intensive and extensive thermodynamic coordinates:

Simple systems	Intensive coordinate	Extensive coordinate
Hydrostatic system	Pressure P	Volume V
Stretched wire	Tension \mathcal{T}	Length L
Surface	Surface tension σ	Area A
Electrochemical cell	Electromotive force (emf) \mathcal{E}	Charge Z
Dielectric slab	Electric field E	Total polarization P
Paramagnetic rod	Magnetic field \mathcal{H}	Total magnetization M

2.1.2 Thermodynamic systems

Among the various types of systems, in relation to the way of exchanging energy with the external environment, we can distinguish:

- **Isolated systems:** they do not exchange heat, matter, work with the outside;
- **Closed systems:** they exchange energy (heat, work), but not matter with the outside. When a system exchanges heat, work or both, it can be classified according to the properties of the boundary:
 - *adiabatic wall*: does not allow heat exchange
 - *diathermic wall*: it allows heat exchange
 - *rigid wall*: does not allow exchange of work
- **Open systems:** allows exchange of energy and matter with the outside. A boundary allowing exchange of matter is called *permeable*.

2.1.3 Thermodynamic transformations

When a system passes from one state of equilibrium to another one, it is said that a thermodynamic transformation takes place.

A distinction is made between reversible and irreversible transformations.

A **reversible process** is the one that is performed in such a way that, at the conclusion of the process, both the system and the local surroundings may be restored to the initial state without producing any changes in the rest of the universe.

A process that does not fulfill these stringent requirements is said to be **irreversible**.

2.1.4 The Laws of Thermodynamics

The laws (or principles) of thermodynamics, enunciated during the XIX century, regulate the thermodynamic transformations and their limits. One can distinguish three basic laws plus a zeroth principle that defines the concept of temperature, and which is implicit in the other three.

2.1.5 Thermal equilibrium

- **Equilibrium state:** a state of a system in which the coordinates have definite values that remain constant as long as the external conditions are unchanged
- **Thermal equilibrium:** the state achieved by two or more systems after they have been in communication with each other through a **diathermic wall**. It is characterized by restricted values of the coordinates

2.1.6 Zeroth Law – Temperature

Two systems in thermal equilibrium with a third are in thermal equilibrium with each other.

In other words: if a body "A" is in thermal equilibrium with a body "B", and "B" is in thermal equilibrium with a body "C", "A" and "C" are in thermal equilibrium with each other.

This law establishes the bases for the concept of temperature and for the use of thermometers.

The property that ensures the system being in thermal equilibrium with one another is called **temperature**. The temperature of a system is a property that determines whether or not a system is in thermal equilibrium with other systems.

This law explains the fact that two bodies initially at different temperatures, between which heat is exchanged, finally reach the same temperature.

In the kinetic formulation of thermodynamics, the zeroth law represents the tendency to reach a common average kinetic energy of the atoms and molecules

of the bodies between which the heat exchange takes place, a tendency that therefore leads to equal temperatures (temperature → mean kinetic energy of particles) .

The efficiency of energy exchange determines the specific heats of the materials involved.

Heat

When a system is placed in contact with another system with a different temperature, a transformation takes place that leads to a state of equilibrium, in which the temperatures of the two systems finally become equal.

To explain this process, XVIII-century scientists hypothesized that a substance, called **caloric**, present in greater quantities in the warmer body, passed into the colder body.

Thermodynamics instead identifies **heat as a form of energy** that can be converted into mechanical work and can be stored, but which is not a material substance.

Work and Heat

- **WORK.** If a system as a whole exerts or experiences a force on/from its surroundings and a **displacement** takes place, a work is done, either by the system or on the system
- **HEAT.** (calorimetric definition) the energy transferred between a system and its surroundings by virtue of a temperature difference only
- It has been experimentally proven that heat, originally measured in calories, and work or energy, measured in joules, are absolutely **equivalent**
- Each calorie equals **4.186 joules**

2.1.7 First Law of Thermodynamics

The first law is a principle of conservation of energy. In every thermal machine a certain amount of energy is transformed into work: there can be no machine that produces work without consuming energy. Such a machine, if it existed, would in fact produce the so-called perpetual motion of the first kind.

The first law is traditionally stated as:

"In a closed system we have that

$$\Delta U = Q - W$$

where U is the internal energy of the system".

Internal Energy is the sum of the kinetic and interaction energies of the different particles of a system. Q is the heat exchanged between environment and system (positive if supplied **to** the system) and W the work done (positive if done **by**

the system on the environment). The convention of signs is due to the study of heat engines, in which heat is transformed (partially) into work.

Differential form: $dU = \delta Q - \delta W$

dU exact differential (U refers to properties of the system)

$\delta Q, \delta W$ inexact differentials (Q, W refer to the way surroundings interact with the system by means of processes of transferring energy)

2.1.8 Second Law of Thermodynamics

The second law states that there are limits in the exchange of work and heat regulated by the first law. There are several statements of the second law, all equivalent, and each of the formulations highlights a particular aspect of it:

- **Kelvin-Planck statement:** it is impossible to construct an engine that, operating in a cycle, will produce no effect other than the extraction of heat from a reservoir and the performance of an equivalent amount of work
- **Clausius statement:** it is impossible to construct a refrigerator that, operating in a cycle, will produce no effects other than the transfer of heat from a lower-temperature reservoir to a higher-temperature reservoir

The K-P statement excludes the possibility of realizing the so-called perpetual motion machine of the second kind. A mathematical formulation of the second law of thermodynamics makes use of the entropy function (see below).

The second law of thermodynamics can be rephrased as follows:

Every time a certain amount of energy is converted from one state to another, there is a penalty that consists in the degradation of a part of the energy itself in the form of heat, in particular this part will no longer be usable to produce work.

2.1.9 Third Law of Thermodynamics

It is closely related to the Second Law, and in some cases it is considered as a consequence of it. It can be stated by saying that

it is impossible to reach absolute zero with a finite number of transformations.

2.1.10 Entropy

In thermodynamics, **entropy** is a state function that is introduced together with the second law of thermodynamics and that can be roughly interpreted as a measure of the "disorder" of a physical system or more generally of the universe. Based on this definition we can say that when a system passes from an ordered state to a disordered one its entropy increases.

In the International System it is measured in joules per kelvin (J/K).
The concept of entropy was introduced in the early nineteenth century, in the context of thermodynamics, to describe a characteristic (whose extreme generality was observed for the first time by Sadi Carnot in 1824) of all systems then known in which it was observed that transformations invariably occurred in only one direction, namely that towards maximum 'disorder'.
The concept of entropy has known great popularity in '800 and '900, thanks to the large number of phenomena that helps to describe, up to leave the purely physical field and be adopted also by the social sciences, in the theory of signals and theoretical computer science.
However, it should be noted that there is a whole class of phenomena, called nonlinear phenomena (for example chaotic phenomena) for which the laws of thermodynamics (and therefore also entropy) must be reviewed and no longer have general validity.

Entropy - Disorder - Time

To further clarify this concept, one can consider some examples:

- Dropping a droplet of ink into a glass of water: what is immediately observed is that, instead of remaining a drop more or less separated from the rest of the environment (which would be an ordered state), the ink begins to diffuse and, in a certain time, a uniform mixture is obtained (completely disordered state).
→ It is common experience that, while this process occurs spontaneously, the reverse process (separating water and ink) would require external energy.
- Imagine a perfume contained in a bottle as a set of point molecules with a certain speed deriving from the temperature of the perfume. As long as the bottle is corked, that is, isolated from the rest of the universe, the molecules will be forced to remain inside and having no space (the bottle is full) they will remain quite orderly in a liquid state. When the bottle is opened, the molecules of the surface of the liquid will begin to detach from the others and randomly colliding with each other and against the walls of the bottle will come out of this, dispersing outside (evaporation). After a certain time, all the molecules will have come out dispersing. → Even if by chance some molecule will fall into the bottle, the overall system is now disordered and the thermal energy that has set in motion the phenomenon is dispersed and therefore no longer recoverable.

If the process is filmed and the video is shown in a reverse mode, you immediately understand that time has been inverted. This is different from the case of a pendulum oscillating without friction (no way to understand if the video is in forward or reverse mode).

→ only when heat comes into play (thermodynamics) time acquires a clear direction.

Entropy

Thus, an increase in entropy of a system can be described as an increase in the disorder of the system.

The concept of disorder is relative to a reference state (e.g., vapour is in a state of disorder, relative to its solid phase), just as entropy is.

It is possible to regard all natural processes from the point of view of orderliness, and, in all cases, the result obtained is that isolated systems or systems plus surroundings experiencing irreversible processes proceed towards a state of greater disorder.

The increase of entropy of the Universe during natural processes is an expression of this tendency → **entropy is the arrow of time**

Before formally introducing these concepts in the entropy principle, let's formally define the entropy function.

Entropy, thermodynamic definition

Entropy S as a state function was introduced in 1864 by Rudolf Clausius in thermodynamics, through its variation, as:

$$dS = \frac{\delta Q_{rev}}{T}$$
$$\Delta S = \int_{rev} \frac{\delta Q}{T}$$

where δQ_{rev} is the amount of heat reversibly absorbed/released by the system at temperature T . Note that while δQ_{rev} is not an exact differential, dividing it by temperature T makes it so.

$\frac{1}{T}$ is therefore the factor of integration.

It should be emphasized that dS is an exact differential only if the second law of thermodynamics is valid.

2.1.11 Principle of Increase of Entropy

The behavior of entropy of the universe (and of any thermally isolated systems) as a result of any kind of process may be represented by:

$$\Delta S \geq 0 \text{ for a thermally isolated system (universe)}$$

where the equality sign refers to reversible processes and the inequality sign to irreversible processes.

This is a mathematical and compact statement of the second law of thermodynamics.

It also implies that not only can energy be neither created nor destroyed, but neither can it be completely transformed from one form into another without a part being dissipated in the form of heat.

Entropy: mathematical definition

Alongside this treatment of entropy, there is another (mathematical) that treats entropy as a function of state of temperature only.

As a continuous and monotonic increasing function of temperature alone, it admits an absolute minimum and maximum (Weierstrass' theorem) to which the universe converges continuously (by the principle of increasing entropy).

Entropy and the Universe

The increase in temperature is a structural factor of the universe. It is impossible at the moment to quantify this maximum temperature, since an analytical dependence of entropy on temperature is not known: in the thermodynamic theory they are represented as independent variables (T-S diagrams).

Therefore, while the initial state of the universe is known at zero entropy, the final state to which it converges (at maximum entropy and temperature) is not known.

The entropy function does not depend on and does not give information on the path that has been and will be followed to get there, that is, it does not tell us the thermodynamic future of the universe. However, it is possible to infer some hints about the ultimate fate of the Universe from thermodynamic considerations.

Assuming that the entire Universe is an isolated system - that is, a system for which it is impossible to exchange matter and energy with the outside - the first and second laws of thermodynamics can be summarized by a single sentence:

The total energy of the universe is constant and the total entropy
is constantly increasing.

The state in which entropy reaches its maximum level and there is no more free energy available to do further work is called the **equilibrium state**.

For the entire universe conceived as an isolated system this implies the progressive conversion of work into heat (by the principle of increasing total entropy). This, in the face of the finite mass of the universe, will eventually lead to a state in which the entire universe will be in conditions of uniform temperature
→ so-called **heat death of the Universe** (the Universe will be unable to sustain further processes that increase entropy).

Entropy: statistical definition

In statistical mechanics, entropy is the means to obtain macroscopic information from microscopic configurations.

It is imagined that a certain macroscopic equilibrium condition of the system corresponds to a multitude of microscopic configurations.

Such microscopic configurations occupy a volume in phase space which is denoted by Γ (Γ is the number of microstates corresponding to a given macrostate of the system)

Then we can define entropy "à la Boltzmann" as

$$S = k_B \ln \Gamma$$

where k_B IS Boltzmann's constant.

It can be shown that entropy thus defined has all the characteristics of thermodynamic entropy and in particular it is shown that it is extensive, that is, it has the property of additivity (and difference: for which the entropy variation is calculable, and the entropy function is differentiable, or it makes sense to speak of entropy in microscopic terms).

The physical difference in meaning between entropy and temperature is that the former measures the state of (physical) disorder of the system and the latter, the state of molecular agitation.

2.1.12 Enthalpy

Enthalpy, H of a macroscopic physical system is a thermodynamic state function defined as

$$H = U + pV$$

where U is the internal energy of the system, p is pressure and V is volume.

Being a state function, the enthalpy variations ΔH that accompany a process depend exclusively on the characteristics of the initial and final state, regardless of the path followed.

H is a convenient function for problems involving heat quantities, such as heat capacities, latent heats, and heats of reaction, when pressure is the variable being controlled.

For example, it is easily shown that the enthalpy variation of a system is equal to the heat exchanged at constant pressure. Differentiating the expression that defines enthalpy, thus considering an infinitesimal transformation of the system, we obtain:

$$dH = dU + pdV + Vdp \quad (1)$$

operating at constant pressure $Vdp = 0$ and then

$$dH = dU + pdV \quad (2)$$

Now, assuming that the only work is that of expansion/compression, and therefore $\delta W = pdV$, we get $dU = \delta Q - \delta W = \delta Q - pdV$, that substituted in (2) gives

$$dH = \delta Q$$

If $dH < 0$ heat is transmitted from the system to the environment, an **exothermic process** takes place. If $dH > 0$, the system absorbs heat → **endothermic process**.

For a hydrostatic system (pure substance, constant mass):

$$H = U + pV$$

$$dH = dU + pdV + Vdp$$

$$\delta Q = dU + pdV \text{ (first law)}$$

$$\rightarrow dH = \delta Q + Vdp$$

for a reversible transformation:

$$\rightarrow dH = TdS + Vdp$$

This is one of the **thermodynamic potential functions**.

All the thermodynamic properties of a system can be calculated by differentiation of them For instance:

$$\left(\frac{\partial H}{\partial S} \right)_p = T \quad ; \quad \left(\frac{\partial H}{\partial p} \right)_S = V$$

$\rightarrow (H,S)$ and (H,p) diagrams are useful to study the properties of a pure substance.

Comparison of properties of the internal energy U and enthalpy H for a hydrostatic system (real and ideal gas)

Internal energy $U(V, S)$	Enthalpy $H(P, S)$
Free expansion (irreversible) $U_i = U_f$	Throttling process (irreversible) $H_i = H_f$
In general $dU = dQ - P dV$ $(\frac{\partial U}{\partial T})_V = C_V$	In general $dH = dQ + Vdp$ $(\frac{\partial H}{\partial T})_P = C_P$
Isochoric process $U_f - U_i = Q_V$	Isobaric process $H_f - H_i = Q_P$
For an ideal gas $U_f - U_i = \int_i^f C_V dT$	For an ideal gas $H_f - H_i = \int_i^f C_P dT$
Adiabatic process $U_f - U_i = - \int_i^f P dV$	Adiabatic process $H_f - H_i = \int_i^f V dP$
Nearby equilibrium states $dU = T dS - P dV$ $(\frac{\partial U}{\partial S})_V = T$ $(\frac{\partial U}{\partial V})_S = -P$	Nearby equilibrium states $dH = T dS + V dp$ $(\frac{\partial H}{\partial S})_P = T$ $(\frac{\partial H}{\partial P})_S = V$

2.2 The Thermodynamics of Refrigeration

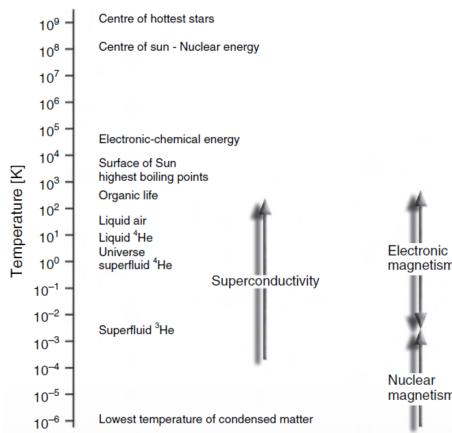
Here we define what Cryogenics is, starting from historical notes, then describing the thermodynamical approach to refrigeration. The cycles at the base of domestic and industrial refrigerators are presented, and the throttling process and Joule-Thomson expansion are described in connection with the enthalpy function and in view of their employment in the liquefaction of gases. Finally, the countercurrent heat exchangers are discussed in detail, and an example calculation desirable dimensions is given.

2.2.1 Cryogenics

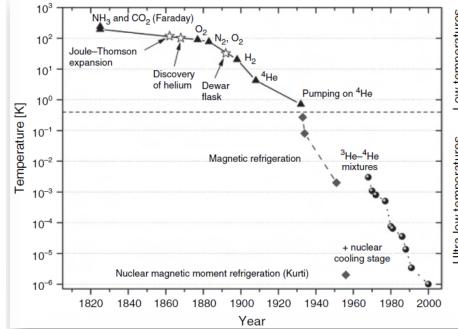
Cryogenics, ‘production of cold’, is a branch of physics which studies the methods to obtain temperatures well below room temperature and the methods to measure them. This technology is fundamental in different areas, from the conservation of food to advanced scientific application such as quantum computing. The range of temperature existing in nature or obtainable in laboratory covers 15 decades.

In nature the lowest temperature is the universe background at 2.7K , due to ‘fossil photons’ from the ‘big bang’.

In laboratory, it is possible to freeze materials down to about 10^{-6} K . Moreover, it is nowadays possible to cool a small number of atoms or molecules ($\sim 10^6$) down to $\sim 500\text{ pK}$.



2.2.2 Historical development of refrigeration



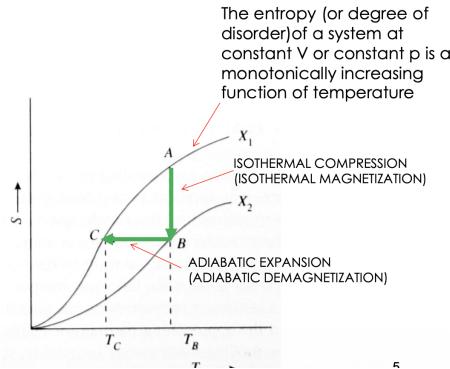
- 1823: Faraday almost "accidentally" manages to obtain liquid chlorine by means of pressure.
A few years later he managed to liquefy SO_2 , CO_2 , NO , NH_3 using similar methods
This allowed the beginning of the development of evaporative cycles based on these liquids:
 - 1850: first air cooler made in Florida to produce ice for hospitals.
 - 1877: successful transport from Argentina to France of meat refrigerated by means of a NH_4 refrigerator.
- For the so-called "permanent gases" (O_2 , H_2 , He) the problem of the right combination of pressure and temperature was more complex and required a few more years and the use (as we will see) of the "throttling" valve.

Table 1.1. Refrigeration techniques. The methods which dominate in the three temperature ranges are in italics

Temperature range	Refrigeration technique	Available since	Typical T_{min}	Record T_{min}
I Kelvin	Universe			2.73 K
II Milli-kelvin	Helium-4 evaporation Helium-3 evaporation 3He - 4He dilution <i>Pomeranchuk cooling</i> <i>Electronic magnetic refrigeration</i>	1908 1950 1965 1965 1934	1.3 K 0.3 K 10 mK 3 mK 3 mK	0.7 K 0.23 K 2 mK 2 mK 1 mK
III Micro-kelvin	<i>Nuclear magnetic refrigeration</i>	1956	$100 \mu K$	$1.5 \mu K^a$

The given minimum temperature for the microkelvin temperature range is the lattice (electronic) equilibrium temperature. Nuclear spin temperatures as low as $0.3nK$ have been reached (Table 10.2)

2.2.3 Isentropic cooling



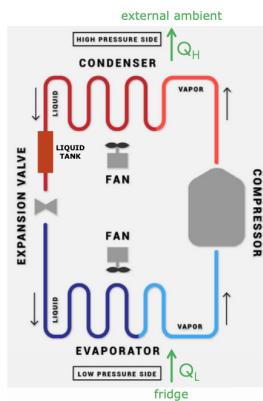
Schematization in an S-T diagram of a possible thermodynamic cooling process:

- $S = S(T, X)$ where the parameter X is a physical property of the system which can be varied within limits so as to change the entropy.
- e.g., $X = \text{pressure } p$ applied to a gas (or magnetic field H applied to an assembly of magnetic dipoles)
- $A \rightarrow B \rightarrow C$: cooling process
- If repeated cyclically it can lead to very low temperatures but always > 0 (for the third principle)

"ISENTROPIC cooling": *cooling \Leftrightarrow ordering \Leftrightarrow entropy reduction*

2.2.4 Use of a throttling process

Example: commercial refrigeration system (electric refrigerator for domestic or industrial use)



$$Q_H > Q_c$$

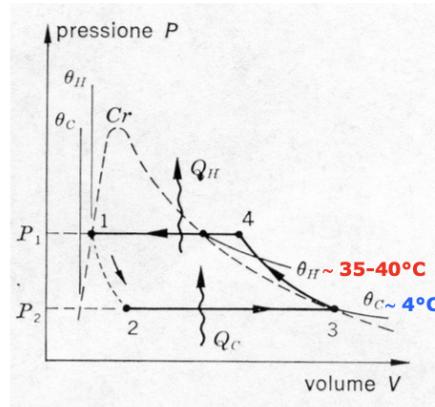
$$T_H > T_c$$

$$W = Q_H - Q_C$$

Work done on the system

Refrigerant fluid = any fluid (liquid, vapor or gas) which can perform the thermodynamic cycle.

- Constant mass of fluid passing from the condenser to the liquid tank and through the valve where it expands and then evaporate
- In the liquid tank there is a suitable liquid at $T_H > T_{amb}$ and $p_H > p_{amb}$ (saturated liquid)
- Expansion valve: (Joule – Thompson expansion, see below) adiabatic expansion of the saturated liquid \rightarrow cooling ($T_C < T_H$) + partial vaporization
- Evaporator: complete evaporation of the fluid \rightarrow the latent heat of evaporation Q_L is provided by the material or chamber to be cooled
- In the compressor the vapor is returned to $T_H > T_{amb}$ and $p_H > p_{amb}$
- Condenser: the vapour condenses by transferring the heat Q_H to air and/or water \rightarrow the substance to this p and T returns to be a saturated liquid
- The minimum temperature obtained in this type of system is of the order of $-18 \div -30^\circ C$



1 \rightarrow 2 Throttled expansion (no equilibrium, see below)

2 \rightarrow 3 Isothermal and isobaric vaporization

3 \rightarrow 4 Adiabatic compression

($T_f > T_H \rightarrow$ over saturated vapor)

4 \rightarrow 1 T_H isobaric cooling and condensation

$COP = \text{coefficient of performance}$
 $= (\text{heat subtracted from } T_C \text{ reservoir}) / W$

$$COP = \frac{Q_c}{|W|} = \frac{Q_c}{Q_H - Q_c}$$

Typical COP of heat pumps ranges from 2 to 7

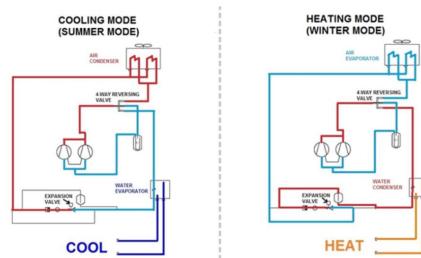
Example: if $COP = 5 \Rightarrow (Q_H - |W|)/|W| = COP$
 $\Rightarrow Q_H/|W| = COP + 1 = 6$

The heat transferred to the thermostat T_H is 6 times the work spent W
 \rightarrow each Joule of work leads to production of 6 J of heat (while with electrical resistance 1 J work = 1 Jheat).

It is convenient to heat a house by cooling the outside!

Heat pump (from an idea of Lord Kelvin, 1852)

Only by rotating a reversing valve the thermal cycle be inverted, realizing both the cooling of the house by heating the outside air, or the heating of the house.



2.2.5 Refrigerants

It is important to choose the right refrigeration fluid to adjust the temperature of the cold body and that of the hot reservoir.

Among the first used fluids there was **ammonia** (NH_3) which has the advantage of having a particularly high latent heat, but it is corrosive and toxic.

For many years the most used were halogenated hydrocarbons, commercially called **freon** (registered trademark), containing carbon (C), hydrogen (H), chlorine (Cl), and fluorine (F), produced as volatile derivatives of methane, ethane, and propane (e.g. chlorofluorocarbons, **CFC**). These are colorless gases, without odor or with a faint odor of ether, non-flammable, chemically stable, without any toxic action.

Being heavier than air, in case of leakage and leaks tend to accumulate in the lower layers of the air and can therefore cause asphyxia due to the depletion of the oxygen content that can take place in the atmosphere.

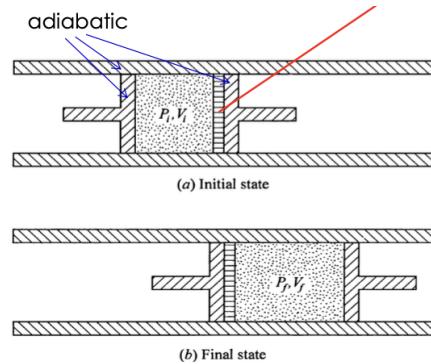
In 1987 (Montreal Protocol) CFCs were banned due to the effect on the formation of the atmospheric ozone layer (ozone depletion).

The most common type of refrigerants used today across the world are hydrofluorocarbons (HFCs).

R-134a ($F_3C - CH_2 - F$) is one of the world's most used refrigerants, widely embedded in automotive, commercial and residential air conditioning systems, across the world.

In any case, these refrigerants are powerful greenhouse gases that remain in the atmosphere for a long time, such as CFC, which persists for 102 years and has a warming capacity, that is, it increases the greenhouse effect 3,800 times more than normal carbon dioxide.

2.2.6 The throttling process

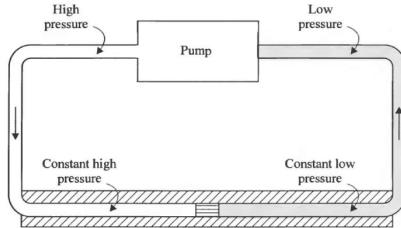


In red: **porous wall** (= porous plug, narrow constriction, series of small holes,...)

It permits mass to flow from one chamber to another while controlling the pressure, unlike a free expansion.

- Both the pistons move simultaneously at different speeds to the right such that a constant higher pressure p_i is maintained on the left-hand side of the porous plug and a constant lower pressure p_f is maintained on the right-hand side.
- A throttling process exhibits internal mechanical irreversibility, due to friction between the gas and the walls of the pores. The gas passes through dissipative nonequilibrium states on its way from the initial equilibrium state to the final equilibrium state.

2.2.7 Enthalpy and throttling



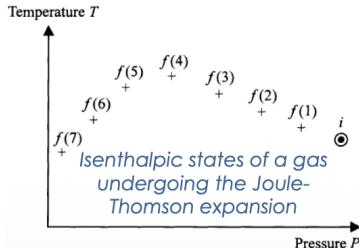
$$\begin{aligned}
 U_f - U_i &= Q - W = -W \\
 W &= \int_{V_i}^0 p_i dV + \int_0^{V_f} p_f dV = p_f V_f - p_i V_i \\
 U_f - U_i &= p_i V_i - p_f V_f \\
 U_i + p_i V_i &= U_f + p_f V_f \\
 H_i &= H_f \text{ (throttling process)}
 \end{aligned}$$

In a throttling process the initial and final enthalpies are equal, but the process is not strictly isenthalpic (the system passes through nonequilibrium states in an irreversible process).

A continuous throttling process may be achieved by a pump that maintains a constant high pressure on one side of the porous wall (or expansion valve) and a constant lower pressure on the other side.

For every fluid of mass m that undergoes the throttling process: $h_i = h_f$
where $h = H/m$ indicates the specific enthalpy.

2.2.8 Joule-Thomson expansion



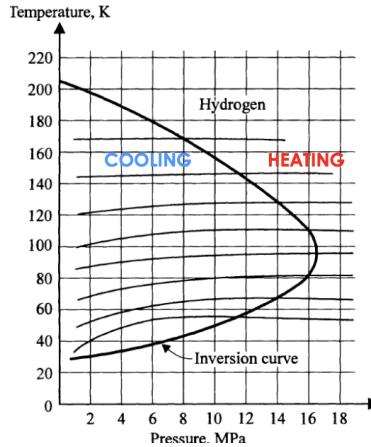
This process is needed to liquefy gases as N , H , and He

In the Joule-Thompson expansion, a gas is made to undergo a continuous throttling process, by means of a pump. In modern apparatus, the gas flows in radial direction through porous walls made of materials able to sustain high pressures. The throttling valve and the surrounding walls must be isolated very carefully. Description of the process: p_i and T_i of the high-pressure side are chosen arbitrarily. P_f on the other side is then set at any value $< p_i$ and T_f is measured → graph (typical of most gases).

Not necessarily a throttled expansion leads to a T decrease, it depends on p_i ,

T_i and p_f . The points on the graph represent equilibrium states of the gas → a smooth curve can be drawn through the discrete points. The curve is called ‘isenthalpic curves’. It is not the graph of a throttling process (that is irreversible).

An isenthalpic curve is the locus of all points representing equilibrium initial and final states of the same molar enthalpy.



By starting from a different $(p_{i'}, T_i)$ point, several isenthalpic curves can be drawn.

The numerical value of the slope of an isenthalpic curve on a TP diagram at any point, is called the Joule-Thomson coefficient:

$$\mu = \left(\frac{\partial T}{\partial p} \right)_h$$

The locus of all points where $\mu = 0$ (maxima of the isenthalpic curves) is the inversion curve

$\mu > 0$ region of cooling

$\mu < 0$ region of heating

Since involves T, p and h, a thermodynamic equation for μ can be obtained starting from the difference in the molar enthalpy between two neighboring equilibrium states.

$$H = U + pV \rightarrow dH = dU + pdV + Vdp = \delta Q + Vdp$$

$$\text{Dividing by } dt: \frac{dH}{dT} = \frac{\delta Q}{dT} + V \frac{dp}{dT}$$

$$\text{At constant } p: \left(\frac{dH}{dT} \right)_p = \left(\frac{\delta Q}{dT} \right)_p = [C_p] \text{ (isobaric heat capacity)}$$

$$dH = \delta Q + Vdp = TdS + Vdp$$

$$\left(\frac{\partial H}{\partial p} \right)_T = T \left(\frac{\partial S}{\partial p} \right)_T + V = -T \left(\frac{\partial V}{\partial T} \right)_p + V$$

Consider $H = H(p, T)$

$$dH = \left(\frac{\partial H}{\partial p} \right)_T dp + \left(\frac{\partial H}{\partial T} \right)_p dT$$

At constant H:

$$0 = \left(\frac{\partial H}{\partial p}\right)_T dp + \left(\frac{\partial H}{\partial T}\right)_p dT$$

Dividing by dp:

$$0 = \left(\frac{\partial H}{\partial p}\right)_T + \left(\frac{\partial H}{\partial T}\right)_p \left(\frac{\partial T}{\partial p}\right)_H$$

$$\left(\frac{\partial T}{\partial p}\right)_H = -\frac{\left(\frac{\partial H}{\partial p}\right)_T}{\left(\frac{\partial H}{\partial T}\right)_p}$$

Gibbs function: $G = H - TS = U + pV - TS$

$$dG = dU + pdV + Vdp - TdS - SdT = Vdp - SdT$$

Condition for an exact differential $dz = Mdx + Ndy$

$$\left(\frac{\partial M}{\partial y}\right)_x = \left(\frac{\partial N}{\partial x}\right)_y \rightarrow \left(\frac{\partial V}{\partial T}\right)_p = -\left(\frac{\partial S}{\partial p}\right)_T \text{ one of the "Maxwell relations")}$$

$$\text{Thus: } \left(\frac{\partial T}{\partial p}\right)_H = \frac{1}{C_p} [T \left(\frac{\partial V}{\partial T}\right)_p - V]$$

$$\text{and for 1 mol: } \mu = \left(\frac{\partial T}{\partial p}\right)_H = \frac{1}{C_p} [T \left(\frac{\partial v}{\partial T}\right)_p - v]$$

For 1 mol of an ideal gas:

$$\mu_{\text{ideal gas}} = \frac{1}{C_p} [T \frac{R}{p} - V] = 0$$

since $pV = RT$

i.e. $T_f = T_i$ under all conditions for an ideal gas (see the curves for a real gas at high T, where it is similar to the ideal one: they are quite flat).

⇒ Cooling by throttling process is only possible for real gases because the temperature change is determined by the change in energy of the gas when the average separation between the gas molecules is increased («internal work» process)

2.2.9 Liquefaction of gases by the Joule-Thomson expansion

To give rise to cooling by the J-T expansion, the initial T must be below the maximum inversion temperature. For many gases (air, nitrogen), room temperature is already below such value, so that no precooling is needed.

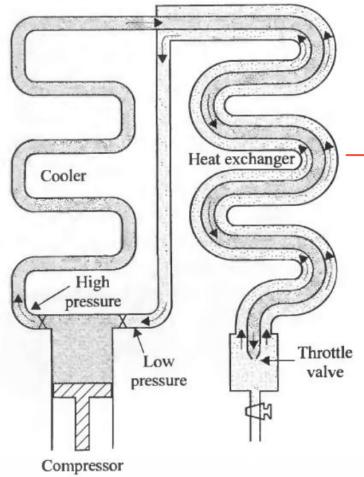
Ex.: if air is compressed to a pressure of 200 atm (about 2×10^7 Pa) and $T_i = 52^\circ C$, then, after throttling to $p_f = 1$ atm, it will be cooled to $23^\circ C$. But if we do the same with He → $T_f = 64^\circ C (> T_i)$.

For hydrogen: $T_i < 200 K$ (precooling with liquid nitrogen required).

For helium: $T_i < 40 K$ (precooling with liquid hydrogen required).

Gas	Maximum inversion temperature, K
Xe	1486
CO_2	1275
Kr	1079
Ar	794
CO	644
N_2	607
Ne	228
H_2	204
4He	43

- Once the gas has been precooled to $T_i < T_{max}$, the optimal pressure from which to start throttling is a point on the inversion curve (ending at atmospheric pressure, the process produces the largest T drop, see below).
- This may not be enough to produce liquefaction → successive cooling processes are usually needed, with the gas that has been cooled by throttling used to cool the incoming gas → this is done by a **countercurrent heat exchanger**.



Double-walled pipe :

For the heat exchanger to be efficient, the temperatures of the opposing streams of gases must differ only slightly

→ heat exchanger long and well isolated, gas with sufficient speed to cause turbulent flow

→ good thermal contact between opposing streams of gas.

When a steady state is established, liquid is formed at a constant rate: for a mass unit of supplied gas, y is the fraction liquefied and $1 - y$ returns to the pump.

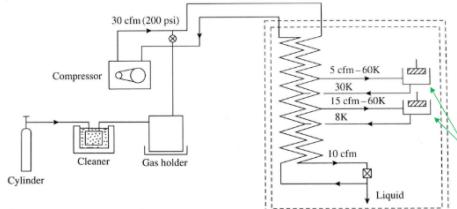


Figure 2.1: A simplified diagram of a Collins He liquefier showing flow rates typical of the original machine and early commercial models

Heat exchanger and throttling valve insulated.
 $h: \text{molar enthalpy } h_i = yh_L + (1 - y)h_f \quad y = \frac{h_f - h_i}{h_f - h_L}$
 $h_L = \text{const.} \rightarrow \text{depends on the pressure of the liquid}$
 $h_f = \text{const.} \rightarrow \text{depends on the pressure drop in the return tube and on } T(C)$
 $\Rightarrow y \text{ depends only on } h_i \rightarrow \text{on } p_i \text{ (once } T_i \text{ has been fixed)} \quad y \text{ max when } h_i \text{ min:}$
 $(\frac{\partial h_i}{\partial p}) = 0 \rightarrow (\frac{\partial h}{\partial T})_p (\frac{\partial T}{\partial p})_h = c_p \mu = 0$
 To maximize y the point (T_i, p_i) must lie on the inversion curve ($\mu = 0$)

Example: Collins He liquefier

($cfm = \text{cubic foot per minute}$; $1 cfm = 28.3 \text{ litres per minute}$) ($psi = \text{pound per square inch}$; $1 psi = 0.0689476 \text{ bar}$)

Advantages of using J-T to liquefy gases:

- no low-T moving parts
- the lower the T , the higher the ΔT (for the same Δp)

Disadvantage:

- strong preliminary cooling required (for H and He)

Adiabatic expansion (about reversible) against a piston or turbine blades

\rightarrow cooling (but ΔT is lower at lower T)

\rightarrow suitable for precooling

The Collins liquefier combines the two methods:

- He expands adiabatically in two stages
- equipped with piston engines

2.2.10 Coolers using turbo-expanders

Turbine refrigeration was proposed by Lord Rayleigh in 1898. Gas expands through a turbine, instead of pistons. The process is generally more efficient, but

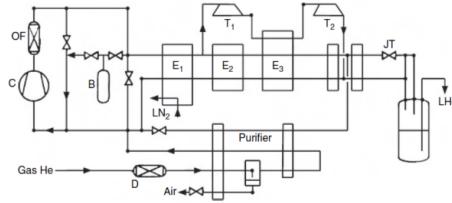


Figure 2.2: Flow diagram of the Linde TCF series of turbine helium liquefiers. C denotes the screw compressor, OF the oil filter, $E_{1,2,3}$ the heat exchangers, $T_{1,2}$ the turbo-expanders and JT the final Joule-Thomson stage. Cylinder B is for pure He gas storage and Gas He is the normal He gas input to the purifier.

the engineering is more sophisticated to implement. Commercial development began in the 1930s. Kapitza (1939) did a careful analysis of turbo-refrigeration and developed radial flow turbines for air refrigeration. Linde models of the TCF series, designed to produce liquid helium at 10 – 100 l/h.

2.2.11 Countercurrent heat exchangers – calculating desirable dimensions

As was discussed above, the heat exchanger is essential in the J-T process for liquefying helium, but it is important in a number of cryogenic applications.

For maximum efficiency, the main goals in HE design are:

- maximum amount of heat should be transferred from the incoming gas stream to the outgoing streams
- small pressure drop for the gas flowing through the exchanger, in comparison with the total entry or exit pressure of the gas
- In low T equipments, $Re \gg 2300 \rightarrow$ turbulent flow (\rightarrow maximum heat transfer)
- For non-circular section tubes \rightarrow hydrodynamic diameter

$$D_h = 4 \times \frac{\text{cross-sectional area}}{\text{total perimeter of surfaces in contact with gas stream}}$$

Calculation of pressure gradient

From classical hydrodynamics: pressure drop

For a circular pipe of diameter D and length L:

$$\Delta p = \frac{1}{2} \psi \frac{LG^2}{D\rho} \text{ where } G = \frac{4m}{\pi D^2}$$

m = mass rate ; ρ = density ; η = viscosity

$\psi = \frac{64\eta}{GD}$ for laminar flow

$\psi = 0.316(\frac{GD}{\eta})^{-0.25}$ for turbulent flow

Calculation of heat transfer

If the gas flowing through the tube is at a temperature different by ΔT from the one of the tube wall, the heat will be transferred at a rate

$$\dot{Q} = h \cdot \Delta T \text{ where } h \text{ is the heat transfer coefficient}$$

For turbulent flow:

$$h = \frac{0.023}{(Pr)^{0.6}} c_p \frac{G^{0.8} \eta^{0.2}}{D_e^{0.2}} [W/m^2 k]$$

where $Pr = \eta c_p / \lambda$ = Prandtl number; η = viscosity; c_p = specific heat [$J/kg \cdot K$]; λ = thermal conductivity [$W/m \cdot K$], D_e = equivalent diameter (ex.: for a simple countercurrent exchanger of two concentric tubes $D_e = D_1 - D_2$)

Representative values for some gases

Gas		N_2	O_2	H_2	4He
c_p [$J/kg \cdot K$]	at s.t.p.	1040	917	14200	5190
ρ [kg/m^3]	at s.t.p.	1.250	1.428	0.090	0.178
η [$\mu Pa \cdot s$]	at 180 K	11.8	13.3	6.3	13.9
	at 50 K			2.5	6.3
k_f [mW/Km]	at 180 K	16.5	16.5	116	107
	at 50 K	4.9	4.5	36	46
$Pr = \eta c_p / k_f$	at 180 K	0.74	0.74	0.77	0.68
	at 50 K			0.84	0.72

Since $Pr \cong 0.75$ and $(\eta/D)^{0.2} \cong 0.25$ (for $D \cong 0.01$ m): $h \cong 0.006 \cdot G^{0.8} c_p$

Required length

By applying considerations of conservation of energy, the exchanger required length can be calculated. With reference to the figure:

$$L = \frac{\alpha}{\gamma} \ln \frac{T_1^H + \beta/\gamma}{T_1^c + \beta/\gamma}$$

where

h_1, h_2 : heat transfer coefficients

m_1, m_2 : mass rates

C_1, C_2 : specific heats

S_1, S_2 heat transfer perimeters

$$\alpha \cong m_1 c_1 \left(\frac{1}{h_1 S_1} + \frac{1}{h_2 S_2} \right)$$

$$\beta = \left(\frac{m_1 c_1}{m_2 c_2} T_1^c - T_2^c \right)$$

$$\gamma = \left(1 - \frac{m_1 c_1}{m_2 c_2} \right)$$

$$\text{If } m_1 c_1 \rightarrow m_2 c_2 \Rightarrow L = \frac{\alpha}{\beta} (T_1^H - T_1^c)$$

Methods of construction

Linde exchanger:

The tube-in-tube or multtube-in-tube heat exchangers are useful in small Linde liquefiers or in the final J-T stage of any liquefier. The performance of Linde-type exchangers is easy to calculate, and their realization is simple. Usually:

- inner tube(s) → high pressure stream
- annular space between the tubes → low pressure stream

Disadvantages: the outer wall contributes seriously to the pressure drop in the gas stream occupying the outer annular space.

Microversion of the Linde heat exchanger: it consists of glass plates which are etched by photolithography to produce fine channels of $\sim 200\mu m$ width and $\sim 30\mu m$ depth. The plates are then bonded together and the gas circulate through these fine channels.

Hampson exchangers:

Shortening the low-pressure return path of the gas stream → the flow resistance (and pressure drop) for the returning stream is smaller than in Linde exchangers.

Disadvantage: more difficult to treat mathematically

Generally, the high-pressure stream is forced to flow around a long spiral path through the shorter, low pressure stream (they are often made by winding a finned capillary tube on a mandrel and enclosing it in the annular space between two cylinders).

2.3 Refrigeration machines

Here we present the Stirling, Gifford-McMahon and pulse-tube cryocoolers, starting from describing the regenerative heat exchangers, then the thermodynamical cycles and way of operation. Specific applications are also mentioned.

2.3.1 Coolers using regenerative heat exchangers

Regenerative heat exchangers differ from the continuous-flow exchangers discussed above as they act as '**cold storage**' systems for the discontinuous or pulsed flow processes used in many cooling cycles and in refrigeration machines.

- They can store the energy from one stream and later transfer the energy to a second stream.
- Regenerative exchangers usually consist of a porous matrix of finely divided material, in the form of wire mesh, plates or small balls (to maximize surface area in order to promote more efficient heat transfer).
- The exchanger should have a heat capacity (in the operating range of temperature) higher than that of the gas stream.
- For operation down to 70 or 80 K, stainless steel, phosphor bronze or nickel are used.
- For cooling stages extending down to 4 K, a combination of lead balls and rare earths is used (e.g. mix of Nd, ErNi, and $HoCu_2$ with about 25% Pb balls: here the heat capacities remain large as a result of low Debye temperatures or magnetic ordering effects).

2.3.2 Stirling refrigerator cycle

1 → 2 Right piston fixed, left piston compresses the gas at constant $T_H \rightarrow Q_H$ released to the hot reservoir

2 → 3 The two pistons move simultaneously → constant gas flow volume → gas through the regenerator → it constant releases $Q_R \rightarrow$ it enters colder in the right volume

3 → 4 Left piston fixed, right piston moves → expansion at constant $T_C \rightarrow$ the gas absorbs Q_C from the cold reservoir

4 → 1 The two pistons move simultaneously → constant volume → gas through the regenerator as in the phase 2-3 → the gas absorbs the same heat (Q_R) that was released during 2 → 3

2.3.3 Stirling cryocoolers

Small portable single-stage version very popular in the 1980s → down to 80 K (widely used for lab air or nitrogen supplies). Two-stage Stirling refrigerators can achieve temperatures of 4K and are used for magnet cooling or for

precooling in J-T liquefiers.

Working gas (helium) is compressed in one chamber and then transferred through a regenerative heat exchanger to a second chamber where it is expanded to cool the 'head' and later transferred back through the exchanger to the lower chamber

1. Compression in space D by the piston B; the heat of compression is removed by water cooling (or air cooling in the mini-versions).
2. Transfer of gas through regenerator G to space E by the displacer C; the gas cools through contact with the cold regenerator.
3. Expansion and further cooling in space E.
4. Transfer of cold gas back through G to space D, thus cooling the regenerator for the next cycle.

Product:

16W @77K

Mass: 3.1 kg

Mean Time to Failure > 200,000 h

Used successfully in applications such as HTS filters, high altitude balloons, refrigeration, germanium detectors, IR detectors, radio telescopes, laser diode cooling.

Split-cycle cooler

Separation of the 'cold head' from the compressor.

- The compression is produced by the piston in the left-hand cylinder which moves the gas towards the right-hand cylinder.
- The combined displacer/regenerator then moves to transfer the compressed gas to the 'head' of the cylinder after which the pistons move apart to allow expansion and cooling.
- The separation of the expansion chamber (with cold finger) from the compressor by a capillary (~30 cm) reduces vibrations.

Typical small split-cycle single-stage units:

- linear drive Stirling cooler, designed for high performance IR-applications (detector temperatures down to 50K) life of over 8000 h
- 1.5 W cooling capacity at 67K cold tip temperature
- cold tip diameter about 12 mm

Special cryocoolers: High-power Stirling cryocoolers

Single stage cryogenerators that provides cooling power in the range of 0,8-11kW @ 40- 160K.

- Often used to produce liquid nitrogen for different cooling purposes (about 1 m³, 500 kg)

Two-stage cryogenerator providing cooling power in the range of 30-175W @ 15-60K.

- Mostly used to remotely cool an application by means of a flow of pressurized cold helium gas
- He gas is first cooled in the coldhead of the cryogenerator to e.g. 20K and will then cool the application while flowing through it. If useful, a separate second loop on the first stage of the cryogenerator can be used at 80 K for e.g. shielding or pre-cooling
- To circulate the He → integrated CryoFans (cryogenic circulators) in the vacuum space of the cold head. In many cases this eliminates the need for a separate cryostat.

Special cryocoolers for space

Applications on satellite orbits:

- X-ray astronomy satellites
- Lunar, Venus probes
- Climate change observation satellite
- Hyper spectrum sensor (equipped on the International Space Station)
- Maintenance free
- Continued efficiency and performance throughout operational life
- 21 years+ continuous operational running on
- demonstrator with no degradation in performance
- Typical cooling power range of 0.5W to 5W at 80K
- **Dual opposed piston** driven by linear motors to compensate vibrations and minimize the noise

FS1ST is a lightweight, single-stage Stirling cryocooler with long service life and specialized for the space environment. This cryocooler is ideal for cooling detectors and heat shields.

2.3.4 Gifford-McMahon cryocoolers

Displacer/regenerator (as in the Stirling cycle), but here the flow from the compressor is controlled by inlet and outlet valves. The compressor can be at some meters from the cold head, to reduce vibrational problems (however: presence of moving parts in the cold area → noise).

The cycle:

- pressurization phase: intake valve is opened with exhaust valve closed. The displacer is at the cold end. High pressure helium gas from the compressor fills the warm end volume W.
- input phase: the input valve remains open and the displacer moves upwards to enhance the cold volume, with gas from the warm end passing through the regenerator .
- de-pressurization phase: the input valve is closed and the exhaust valve is opened slowly so that cooling occurs at the cold end.
- exhaust phase: the displacer is moved downwards to push out the remaining cold gas through the regenerator. The exhaust valve is then shut and the cycle is repeated.

Single-stage GM coolers are mainly used for communication electronics and cryogen free magnets (see LAB).

Two-stage and three-stage versions have been developed for cooling microwave amplifiers and precooling J-T He liquefiers, for driving a dilution refrigerator,...

2.3.5 Pulse tube refrigerators

PTR is a closed-cycle regenerative mechanical cooler, with no moving parts at low temperatures (no displacer) → long lifetime and low vibration level.

Based on the expansion of a gas (usually He) undergoing an oscillating flow with oscillating pressure (avg. pressure 10–25 bar, pressure amplitude 2–7 bar).

To obtain temperatures below 20 K, the systems are usually operated at 1–5 Hz.

Two types of PTRs, depending on the way the pressure oscillations are generated:

- GM-type PTR: a compressor produces continuous high and low pressures and uses a rotary valve to generate pressure oscillations in the pulse tube
- Stirling-type PTR: pressure oscillations are created by the movement of a piston

The theory behind Pulse Tube Coolers is very similar to that of the Stirling Refrigerators, with the volume displacement mechanism of the displacer replaced by the orifice/buffer volume configuration.

The combination of the orifice and the buffer provides a phase difference between the flow of the gas in the tube and the pressure oscillation.

GM-type PTR: compressor (CP), room temperature heat exchanger (E1), rotary valve (RV), regenerator (RG), low- temperature heat exchanger (E2), pulse tube (PT), another room temperature heat exchanger (E3), two orifices (O_1 and O_2) and a buffer volume (BF)

- The compressed gas enters the regenerator at room temperature
- The regenerator stores heat adsorbed from the gas, when it flows from the compressor to the tube, and reject this heat back to the gas, when the gas flows from the tube to the compressor.
- After the regenerator, the gas enters the pulse tube. At the low-temperature end T_L the cold heat exchanger E_2 is used to extract the heat from the cooling object. The hot heat exchanger E_3 rejects this heat to the surroundings. The orifices O_1 and O_2 are flow resistances.
- The buffer volume is a reservoir with a volume typically 10 times larger than the volume of the pulse tube. The pressure in the buffer is almost constant and close to the average pressure in the pulse tube.
- The combination of the orifice O_1 and the buffer provides a phase difference between the flow of the gas in the tube and the pressure oscillation. Phase shift mechanisms are essential: a phase shift, within the periodic cycle, between the source pressure signal and the resultant volumetric flow rate results in a steady temperature variation across the pulse tube
- The function of the double-inlet orifice O_2 is to reduce losses. It allows some gas to bypass the regenerator and to enter the pulse tube via E_3

Regenerator: issues

- Nowadays, PTRs provide cooling powers ~ 15 W at 4.2 K and ~ 40 W at 45 K.
- The material of the regenerator of a PTR must have a high specific heat to provide a good heat storage. Unfortunately, below 20 K, the specific heat of most regenerators rapidly decreases, whereas the heat capacity of helium increases and has a maximum at ~ 10 K
- This drawback has been overcome by using rare- earth magnetic materials in the coldest part of the regenerator. These materials have a magnetic phase transitions below 15 K, with an increase in their heat capacities
- However, to find a material suitable for sub 4K applications is still a challenge

2.4 Properties of cryogenic fluids

2.4.1 Properties of Cryoliquids

Cryoliquids are very important for low-temperature physics because they are the simplest means of achieving low temperatures. In particular, the properties of liquid helium are essential because all refrigeration methods to $T < 10\text{K}$ use liquid helium as a final or intermediate refrigeration stage. Of particular importance for refrigeration are:

- the boiling point T_b (defining the accessible temperature range)
- the latent heat of evaporation L (defining the cooling power)
- the price (defining the availability)

2.4.2 Liquid Air, Liquid Oxygen, Liquid Nitrogen

Today **liquid oxygen** is not in common use as a refrigerant because it is extremely reactive (an explosive oxidation reaction can occur if the oxygen comes into contact with organic liquids, like oil used in pumps, or with solid matter having a large surface area, like metal powder).

Liquid air is not commonly used as well, because nitrogen evaporates first (lower T_b), resulting in an enrichment of dangerous liquid oxygen. For these reasons, today air is liquefied and separated into oxygen and nitrogen in separation plants.

Liquid nitrogen (LN_2) is low-cost and therefore is commonly used for refrigeration down to temperatures higher than about 60 K. (Evaporating nitrogen can cause asphyxia if it displaces much of the usual 20% oxygen in an enclosed volume. Therefore, rooms in which evaporating liquid nitrogen is kept have to be fitted with an appropriate ventilation system).

Liquid Hydrogen

In liquid hydrogen the pair of atoms forming an H_2 molecule are bound by a strong covalent force. The interactions between H_2 molecules, which lead to the liquid and solid states, are the weak *Van Der Waals* forces. These weak dipolar forces, as well as the large zero-point motion of the light H_2 , result in rather low boiling and melting points.

The dangerous feature of hydrogen is its exothermic reaction with oxygen to form water (however, the reaction needs a concentration of more than 4% in air in order to occur). Therefore, liquid hydrogen should be used in a closed system.

It is not often used in laboratories these days because the temperature range between the boiling point of nitrogen (77 K) and the boiling point of helium (4.2 K) is now accessible by means of closed-cycle refrigerators. →HOWEVER its use may become more common in the emerging hydrogen energy technologies....

One of its properties which is important in various low-temperature experiments and which demonstrates some rather important atomic and statistical physics is the **ortho–para conversion** of H₂.

The proton H has a *nuclear spin* $I = 1/2$. The H₂ molecule can thus have a symmetric nuclear state ($I = 1$), so-called **ortho-H₂**, with degeneracy of $2I + 1 = 3$, or an antisymmetric state ($I = 0$), so-called **para-H₂**, with the degeneracy $2I + 1 = 1$.

Para-H₂ has lower energy than ortho-H₂, but the difference is 172 K, if expressed in terms of temperature. It means that at room temperature all the states are equally populated and since three belong to ortho-H₂ and only one to para-H₂, we have 25% para-hydrogen and 75% ortho-hydrogen in thermal equilibrium at room temperature.

However, when we cool hydrogen below 172 K, the lower energy of para-H₂ state starts to be more populated, i.e. a **conversion from ortho- to para-hydrogen occurs upon cooling**.

It is an **exothermic reaction** giving rise to the rather large heat release Problems in experiments at low or ultralow temperatures:

- If the liquid consists mainly of ortho-hydrogen, it will evaporate due to the ortho–para conversion even without any extra external heat being introduced. (To avoid this, one has first to convert the ortho-H₂ to para-H₂ before cooling, by bringing H₂ in contact with a catalyst containing electronic magnetic moments, e.g., ferric hydroxide, iron).
- Many metals such as Cu, Ag, Au, Pt and Rh, may contain traces of hydrogen, in small gas bubbles with a typical diameter of some 0.1 μm (due to their electrolytic production or from purification processes → typical concentration of 10–100 ppm). Then again, the ortho–para conversion in the bubbles gives rise to heat release in these metals, limiting refrigeration.

Liquid Helium: *Some Properties of the Helium Isotopes*

The common stable helium isotope is ⁴He. Its nucleus contains two protons and two neutrons, each with anti-parallel nuclear spin orientation. Therefore, the total nuclear spin of ⁴He is $I = 0$: it is a **boson**.

The nucleus of the (rare) ³He isotope contains two protons, but only one neutron. Therefore, $I = 1/2$, and ³He is a **fermion**.

The different statistics causes substantial differences in their low-temperature behavior.

³He in use today is a byproduct of tritium manufacture in nuclear reactors and is very expensive (more than 200 €/l of gas at STP).

Some important properties are:

- Rather low boiling points and critical temperatures.
- They do not become solid under their own vapour pressure even when $T \rightarrow 0\text{K}$. One has to apply at least about 25.4 or 34.4 bar.

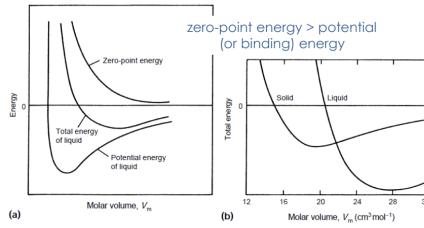


Figure 2.3: (a) Zero-point and potential energies of liquid ${}^4\text{He}$ as a function of molar volume. The total energy is the sum of these two energies. (b) The illustration of why the liquid state is the stable one for helium at saturated vapour pressure even at the $T = 0\text{K}$.

- The helium isotopes have no triple point where gas, liquid and solid phases coexist.
- The two liquids have a rather small density or large molar volume (more than a factor of two larger than for a).

The peculiarities of ${}^4\text{He}$ and ${}^3\text{He}$ are essentially due to:

- *very weak binding forces between the atoms* (van der Waals forces), weak because of the closed electronic s-shell of helium \rightarrow absence of static dipole moments \rightarrow smallest known atomic polarizability: the resulting dielectric constants $\epsilon_4 = 1.0572$ and $\epsilon_3 = 1.0426$.
- due to the small atomic mass m , they have a *large quantum mechanical zero point energy*.

$$E_0 = \hbar^2 / 8ma^2$$

where $a = (V_m/N_A)^{1/3}$ is the radius of the sphere to which the atoms are confined, and N_A is Avogadro's number, 6.022×10^{23} atoms mol^{-1} .

The large E_0 gives rise to a *zero-point vibration* amplitude which is about 1/3 of the mean separation of the atoms in the liquid.

The Figure 2.3 illustrates the influence of the zero-point energy on the total energy as a function of distance between the atoms, and demonstrates why helium – in contrast to all other substances – will remain in the liquid state under its own vapour pressure even when cooled towards absolute zero.

Due to its smaller mass, the influence of the zero-point energy is more pronounced for ${}^3\text{He}$, giving rise to its lower boiling point, smaller density, and larger vapour pressure.

Because of the strong influence of quantum effects on their properties, the helium liquids are called **quantum liquids**.

Liquid Helium: Latent Heat of Evaporation and Vapour Pressure

Latent heat of evaporation L and vapour pressure p_{vap} are important properties that determine whether a liquid is suitable for use as a refrigerant due to

superfluid transition.

Due to the ZPE, for ${}^4\text{He}$ the latent heat of evaporation is only about 1/3 of its value for the corresponding classical liquid.

Due to the small L , liquid helium baths have a rather small cooling power (it is very easy to evaporate them) → they require efficient shielding against introduction of heat from the surroundings.

The vapour pressure can be calculated, at least to a first approximation, from the Clausius–Clapeyron equation

$$\left(\frac{dp}{dT}\right)_{\text{vap}} = \frac{S_{\text{gas}} - S_{\text{liq}}}{V_{m,\text{gas}} - V_{m,\text{liq}}}$$

where S is the entropy and V_m the molar volume.

If we take into account that $\Delta S_{\text{gas,liq}} = L/T$, that $V_{m,\text{gas}} \gg V_{m,\text{liq}}$, and that in a rough approximation (as for the ideal gas) for helium $V_{\text{gas}} \cong RT/p$, then we obtain

$$\left(\frac{dp}{dT}\right)_{\text{vap}} \cong L(T) \cdot \frac{p}{RT^2}$$

and eventually the result

$$p_{\text{vap}} \propto e^{-L/RT}$$

if we make the further approximation that $L \cong \text{const.}$ Thus, the vapour pressure decreases roughly exponentially with decreasing temperature.

One can **pump on the vapour above a liquid**, to obtain temperatures below the normal (1 bar) boiling point.

If one pumps away atoms from the vapour phase, the most energetic atoms will leave the liquid to replenish the vapour → the mean energy of the liquid will decrease → it will cool.

For a pumped-on liquid bath where \dot{n} particles/time are moved to the vapour phase, the cooling power is given by

$$\dot{Q} = \dot{n}L$$

.With a pump with a constant-volume pumping speed \dot{V} , the mass flow \dot{n} across the liquid–vapour boundary is proportional to the vapour pressure $\dot{n} \propto p_{\text{vap}}(T)$, giving a cooling power

$$\dot{Q} \propto Lp_{\text{vap}} \propto e^{-1/T}$$

.The cooling power decreases rapidly with decreasing temperature because the vapour pressure decreases and pumping becomes less and less efficient.

Eventually there is almost no vapour left, resulting in a *practical limit* for the minimum temperature obtainable by pumping on a bath of an evaporating cryoliquid (reached when the refrigeration due to evaporation of atoms is balanced by the external heat flowing to the bath).

Other advantages of this pronounced temperature dependence of the vapour pressure:

- “**cryopumping**” (trapping gases and vapours by condensing them on a cold surface)
- **vapour-pressure thermometry** (the helium-vapour pressure scale represents the low- T part of the

Liquid Helium: *Specific Heat*

- The specific heat of liquid helium is very large compared to the specific heat of other materials at low temperatures.
This remarkable fact is of great cryotechnical importance for low-temperature physics. It means that the thermal behavior (e.g. thermal response time) of a low-temperature apparatus is in most cases determined by the amount and thermal behavior of the liquid helium it contains.

Moreover:

- the latent heat of evaporation of liquid helium – even though it is rather small compared to the latent heat of other materials – is large compared to the specific heat of other materials at low temperatures, enabling us to cool other materials, e.g., a metal, by liquid helium.

Both properties mean that **the temperature of an experiment rapidly follows any temperature change of its refrigerating helium bath** (warning: at very low T problems could arise from the thermal boundary resistance, see below II-A-4). The specific heat of ^4He has a pronounced maximum at about 2.17 K, indicating a phase transition to a new state of this liquid, the **superfluid state** ($^4\text{He-II}$).

The characteristic shape of the maximum has led to the term **λ -transition** for the superfluid transition of ^4He .

$T_\lambda = 2.1768$ K at saturated vapour pressure, while it decreases with increasing pressure to a value of 1.7673 K at the melting line.

Above the λ -transition, ^4He ($^4\text{He-I}$) behaves essentially like a classical fluid (almost like a classical gas, due to its low density).

Below T_λ , due to its (Bose-Einstein like) condensation in momentum space, its entropy and specific heat decrease rapidly with temperature.

Between 1 and about 2K the specific heat has a strong temperature dependence due to so-called roton excitations. Finally, below 0.6 K, the specific heat decreases with T^3 due to phonon-like excitations, as.

The isotope ^3He has a nuclear spin $I = 1/2$. It is therefore a Fermi particle and has to obey Fermi statistics and the Pauli principle. This liquid has many properties in common with the “conducting Fermi liquid” composed of electrons in metals, which are also spin-1/2 particles.

Because ^3He atoms are Fermi particles, single ^3He atoms cannot undergo the analogue of a Bose condensation into a superfluid state as the bosons ^4He do at T_λ .

However, there exists a weak attractive interaction between ^3He atoms in the

liquid which gives rise to pairing of two ^3He atoms. Like the paired conduction electrons in a superconducting metal, the paired ^3He atoms then behave like bosons and can undergo a transition into a superfluid state. Because the pairing forces are rather weak, this transition occurs only at $2.4(0.92)\text{mK}$ for $P = P_{\text{melting}}(P_{\text{svp}})$

2.4.3 Liquid Helium: *Transport Properties of Liquid ^4He Thermal Conductivity*

In its **normal fluid state** above $T = 2.2 \text{ K}$, liquid ^4He shows transport properties almost like a classical fluid:

- low thermal conductivity
- non-uniform T in the liquid/surface while pumping or heating from the bottom
- $^4\text{He-I}$ boils with bubbling.

In the **superfluid state**, under ideal conditions (heat flow $Q \rightarrow 0$) the thermal conductivity of $^4\text{He-II}$ is infinite.

- for realistic conditions it is finite but quite large.
- The very high thermal conductivity of superfluid helium makes it a very efficient medium for establishing temperature homogeneity or for transporting heat.
- For example, at these temperatures the liquid does not boil if heated from the bottom or if pumped, because for creation of bubbles and their transport a temperature gradient is necessary. This is not possible below T_λ if the heat current is not so large that it “destroys” the superfluid state.
- Because of high κ (ideally infinite) temperature waves or entropy waves can propagate in this unusual liquid. “Second sound”, with lower velocity than the usual “first sound” (density waves)

Viscosity and Superfluid Film Flow

In the superfluid state, liquid helium has a **vanishing viscosity** $\eta_s = 0$ for flow through fine capillaries or holes. It is indeed “superfluid” if the flow **velocity** does not exceed a critical value (corresponding to a critical current in a superconductor).

$\eta_s \rightarrow 0$ allows superfluid helium to **flow in a persistent mode** as the persistent supercurrents in a superconductor do.

Very efficient way of **detecting extremely small leaks** in apparatus coming into contact with superfluid helium: it can flow through minute cracks or holes which are impermeable to viscous materials.

The walls of a container partly filled with liquid He are coated with a film of He via the adsorption of atoms from the vapour phase.

Rather strong van der Waals interaction exerted by the wall on helium atoms → the layer is thick, typically, up to 30nm.

Usually, due to its viscosity, such a film is immobile. However, in the superfluid state with $\eta_s = 0$, the film can move→**superfluid film flow**.

Superfluid film flow:

If two containers which are partly filled with liquid He II to different levels are connected, their levels will equalize by means of frictionless flow of the He II film from one container to the other one *driven by the difference in gravitational potential*: the film acts as though being siphoned.

Amazingly, superfluid-helium film will creep up and over the vessel wall, then collect at the bottom as a drop that drips into the reservoir (totally unlike a siphon). This process will continue spontaneously until the vessel is empty.

This superfluid film flow will lead to an enhanced evaporation rate from ^4He baths at $T < T_\lambda$ because the superfluid film flows to hotter places and evaporates there.

2.5 Thermal Contact and Thermal Isolation

In any low-temperature apparatus it is necessary to **couple** some parts **thermally** very well, whereas other parts have to be well **isolated** from each other and, in particular, from ambient temperature.

The transfer of "heat" and the thermal isolation are essential considerations when designing a low-temperature apparatus.

→ a general treatment of the thermal conductivity of materials is needed.

2.5.1 Heat transfer

In general, heat may be transferred by conduction, convection and radiation. In most low-temperature applications, thermal isolation is assisted by partial evacuation of gas from the interior of the cryostat so convection is eliminated. Thus, effective heat transfer is dominated by:

- conduction through the residual low-pressure gas
- conduction through the solids that interconnect the parts of the cryostat
- radiation

Further factors that may contribute are: Joule heating in electrical leads, eddy current heating, mechanical vibration, adsorption/desorption of gases.

Conduction of heat by a low-pressure gas

The residual gas pressure in a cryostat is nearly always reduced to a point where the mean free path becomes comparable with the dimensions of the system.

At such pressures, the average molecule may travel from a hot wall to a cold wall without collisions with another gas molecule → thermal conductivity becomes function of the number of molecules present and their mean velocity.

Within an elementary treatment on the basis of kinetic theory it turns out that for approximately parallel surfaces of equal area A at temperatures T_1 and T_2 the heat transferred by conduction through a gas at low pressure p is

$$\dot{Q} \approx \text{const} \cdot \bar{a} \cdot A \cdot p(T_2 - T_1)$$

where

- const. = 1.2 for air; 4.4 for hydrogen; 2.1 for helium
- \bar{a} is the "accommodation coefficient" (the fraction of gas-surface collisions that result in uptake of the gas by the surface)
 - for He on Pt, W or glass → $\alpha \cong 0.3$ @300K, 0.4 @80K, 0.6 below 20K
 - for air or N on Cu → $a = 0.7$ @243K, 0.8 @77K, 0.6 below 20K
 - for heavier polyatomic molecules → $0.8 < a < 1.0$

Heat transfer through solids

The heat flow through a solid of cross section A under a temperature gradient is given by

$$\dot{Q} = \kappa(T)A \frac{dT}{dx}$$

where κ is the thermal conductivity. If the ends of a solid bar of uniform cross-section A and length l are at temperatures T_1 and T_2

$$Q = \frac{A}{l} \cdot \int_{T_1}^{T_2} \kappa(T) dT$$

The peculiar behavior of dielectric crystals and of metals (with a maximum at low temperature), of the alloys (always decreasing) and of glasses and polymers are to be ascribed to the ***phononic and electronic mechanisms of heat transfer in solids***.

In fact, heat can be carried by conduction electrons or by lattice vibrations; their contributions are additive.

These carriers of heat usually do not fly ballistically from the heated end of the material to the other colder end. They are *scattered* by other electrons or phonons or by defects in the material; therefore, they perform a **diffusion process**.

To calculate thermal transport, we have to apply transport theory, which in its simplest form is a kinetic gas theory. In this simplified version, *electrons and/or phonons are considered as a gas diffusing through the material*.

Within the kinetic gas theory, roughly:

$$\kappa \propto \tilde{c} \cdot v \cdot \lambda$$

\tilde{c} = "what is transported [here the specific heat c]"

v = "the velocity of the carriers performing the transport"

λ = "how far the carriers fly before they are scattered again [λ = mean free path]"

Two scattering processes dominate in different temperature regions and with opposite temperature dependences:

- Low T: scattering from lattice defects and impurities (about constant with T)
- High T: scattering from thermally excited phonons, increasing with T

As a result, thermal conductivity goes through a maximum (the value and position of the **maximum** depend on purity).

Neutral superfluids such as superfluid liquid helium are excellent thermal conductors.

- Unlike charged systems, density fluctuations (second sound, phonons/rotons) do not have an energy cost (i.e. they are gapless) and can carry heat extremely well.

- In fact, superfluid liquid helium is by far the best known thermal conductor of all materials.

At very low temperatures ($T < 1 \text{ K}$) thermal conductivities are about linear in temperature. They range from that very high of pure and crystalline copper to that very low (10^7 times lower) of a particular type of graphite.

κ depends on the purity and crystalline perfection of a material (e.g. in the literature the so-called residual resistivity ratio (RRR) is very often given as a measure of the “purity” of a metal. This is the ratio of the electrical resistivity at room temperature to the electrical resistivity at very low temperatures).

A correct measurement of the low-temperature thermal conductivity of a metal can be rather cumbersome, while a measurement of the electrical conductivity σ is much easier.

In a metal usually both conductivities are determined by the flow of electrons and are mostly limited by the same scattering processes → Wiedemann-Franz law for the ratio of these two conductivities: $\kappa/\sigma = L_0 T$, where the Lorenz number L_0 is a universal constant $= 2.4453 \times 10^{-8} \text{ W} \Omega \text{K}^{-2}$. Thus, from measurements of σ , κ can be easily estimated.

For practical cryogenic calculations, the effective heat conductance of a solid bar with certain end temperatures is needed:

$$\bar{\kappa} = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \kappa(T) dT$$

Commonly encountered pairs of end temperatures are 300-77 K; 300-4.2 K; 77-4.2 K; and 4.2-1 K

	$T_2 = 300 \text{ K}$ $T_1 = 77 \text{ K}$	$T_2 = 300 \text{ K}$ $T_1 = 4 \text{ K}$	$T_2 = 77 \text{ K}$ $T_1 = 4 \text{ K}$	$T_2 = 4 \text{ K}$ $T_1 = 1 \text{ K}$	$T_2 = 1 \text{ K}$ $T_1 = 0.1 \text{ K}$
Nylon	0.31	0.27	0.17	0.006	0.001
Pyrex glass	0.82	0.68	0.25	0.06	0.006
Machineable glass-ceramic	2	1.6	1.3	0.03	0.004
Graphite (AGOT)				0.0025	0.0002
18/8 stainless steel	12.3	10.3	4.5	0.2	0.06
Constantan (60 Cu, 40 Ni)	20	18	14	0.4	0.05
Brass (70 Cu, 30 Zn)	81	67	26	1.7	0.35
Copper (phosphorus deoxidized)	190	160	80	5	(1)
Copper (electrolytic)	410	570	980	200	(40)

Table 2.1: Effective thermal conductivity $\bar{\kappa}$ [$\text{W}/(\text{m} \cdot \text{K})$]

- Alloys of the cupro-nickel family or of the stainless-steel group are frequently chosen for the tubes in a cryostat where low thermal conductivity is a requirement
- The former (Cu-Ni) are often preferred due to the greater ease with which they may be soldered

2.5.2 Radiation

A perfect **black body** may be defined as the one which absorbs all radiation falling upon it. For such body, the absorptivity a and the emissivity ϵ are unity, and so its reflectivity R is zero:

$$R = 1 - \epsilon = 1 - a$$

For a black body at a temperature T the total radiant energy emitted per second and unit area is given by

$$E = \sigma T^4$$

where the Stefan's constant $\sigma = 5.67 \cdot 10^{-8} W/(m^2 K^4)$.

Such radiant energy is distributed over a range of wavelengths λ . $E(\lambda)$ at any temperature T has a maximum value for λ_m , and it may be shown that

$$\lambda_m T = \text{const.} \cong 2900 \mu m K \quad (\text{Wien's constant})$$

- Most non-metallic surfaces such as glass, polymers and varnishes do approximate to black bodies in that their emissivities are $\epsilon \cong 0.9$.
- Metallic surfaces have $0.1 < \epsilon < 1$ depending on the wavelength of the incident radiation and the physical state of the surface, e.g.:
 - Cu oxidized, at room temperature for $\lambda = 14 \mu m \rightarrow \epsilon = 0.6$.
 - Cu polished, at room temperature for $\lambda = 14 \mu m \rightarrow \epsilon = 0.02$.
- A thin layer of oil or Apiezon grease on a low emissivity surface rises ϵ to 0.2-0.3, varnish to 0.87 and Scotch tape to 0.88.

Table 2.2: Values of ϵ at room T and for radiation from 300 K on a 78 K surface

Material	radiation on 78 K surface	room temperature
Al, clean polished foil	0.02	0.04
Al, highly oxidized	-	0.31
Brass, clean polished	0.029	0.03
Brass, highly oxidized	-	0.6
Cu, clean polished	0.015-0.019	0.02
Cu, highly oxidized	-	0.6
Cr, plate	0.08	0.08
Au, foil	0.010-0.023	0.02-0.03
Ni, polished	-	0.045
Ag, plate	0.008	0.02-0.03
Stainless steel	0.048	0.074
Sn, clean foil	0.013	0.06
Soft solder	0.03	-
Glass	-	0.9

For two plane parallel surfaces of area A and emissivities ϵ_1 and ϵ_2 and at temperatures T_1 and T_2 the heat transfer by radiation per unit time is

$$\dot{Q} = \sigma A(T_1^4 - T_2^4) \frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2 - \epsilon_1 \epsilon_2}$$

Hence, if $\epsilon_1 = \epsilon_2 = \epsilon \ll 1$:

$$\dot{Q} = \sigma A(T_1^4 - T_2^4) \frac{\epsilon}{2}$$

Radiation shields are commonly used in cryostats to reduce the heat inflow

- polished metal (copper) cylinder kept at a temperature intermediate between ambient and cryogenic bath)
- multilayers (**aluminized Mylar**) inserted in the vacuum interspace of a cryostat
-

To reduce heat inflow through measurement inserts introduced into the cryostat, circular or semicircular **thermal radiation baffles** of polished metal are soldered and sometimes high-density Styrofoam is inserted between them.

2.5.3 Other causes of heat transfer

Joule heating

Joule heating ($\dot{Q} = I^2 R$) in connecting leads and in resistance thermometers. The opposing demands of low thermal conduction along electrical leads (therefore high R) and low Joule dissipation (therefore low R) often present a problem. At $T < 7$ K this can be solved by wires of a poor thermal conductor such as constantan, and tinning the surface with a thin lead coating: this becomes superconducting below 7 K and the wire remains a poor heat conductor (*). Pure metals and alloys behave in different ways

- copper: about same κ at 300K and 4K, while ρ drops of a factor 100
- constantan: about same ρ at 300K and 4K, while κ drops of a factor 100

(*) *Superconductors are poor thermal conductors*, for the reason that the charge carriers (Cooper pairs) do not carry entropy/heat. It is a charged system, *excitations are gapped* (different from neutral superfluid liquid helium that is an excellent thermal conductor, since excitations do not have an energy cost, they are gapless and can carry heat extremely well)

Gas adsorption/desorption

The thermal energy required to desorb a layer of adsorbed gas from a surface is of the same order of magnitude as the latent heat of vaporization. In precision calorimetry it is undesirable to expose the sample or calorimeter to exchange gas at $T < 10 \text{ K}$, because *adsorption and subsequent gradual desorption of the He layer* when the space is pumped to a high vacuum *can cause temperature drifts*. To avoid the use of exchange gas, in these cases one can use mechanical or superconducting heat switches (see below).

Inadequate anchoring of leads

It is just another case of conduction by solids: any lead wires, control rods, etc. which offer a flow-path for heat from room temperature to the experimental space need to be connected to a heat sink before reaching the experiment (**thermal anchoring**)

He film creep and thermal oscillations

At $T < T_\lambda$ the superfluid He film covering the walls moves towards hotter places and evaporates there → He loss and difficulty to reduce the pressure. The He film flow can be reduced by having a small restriction (e.g. 0.5-mm orifice) in the pumping tube below the evaporation point (or by “He-film burner”, see below). Thermal oscillations in a tube leading down into a liquid helium vessel can increase the evaporation rate.

They can be cured by altering the tube dimensions or inserting some form of damping (e.g. thread of cloth) in the tube. Such oscillations themselves are useful to produce a He-level detector (see below).

Mechanical vibrations

They are relevant only in experiments at ultralow temperatures or in precision calorimetry. Heating at a level of $0.1 \mu\text{W}$ and more can occur in an otherwise thermally isolated system due to being mechanically coupled to a building (vibrations from elevators, nearby traffic,...).

Solutions: rigid suspensions inside the cryostat, e.g. graphite rods with resonant frequency much higher than building or pump vibrations ($\sim 10 \text{ Hz}$), or mounting the cryostat on a pier in sandy soil to decouple it from the building,

...

Other sources of vibrations: bubbling in the cryogenic fluids, pumping flexible tubes, cryocoolers' vibrations.

Electrical losses (induced currents)

Heating due to induced currents (relevant in the presence of intense magnetic fields variable in time) or coupling with radiofrequency sources → use of brass

instead of copper, use of shielded or twisted cables, rf shielded enclosures or rooms,...

Example of heat-transfer calculation

Consider the heat inflow from various sources in the following example.

Setup Description:

- 12 electrical leads (8 of Cu $\phi = 0.1 \text{ mm}$ and 4 of constantan $\phi = 0.2 \text{ mm}$) thermally anchored on a copper bush at 77K... then taken to the inner chamber with a path of 12 cm
- polished copper inner chamber, $A = 0.05 \text{ m}^2$ (4.2 K)
- Stainless steel tube, $\phi = 2 \text{ cm}$, wall thickness = 0.3 mm, length 6 cm
- tarnished brass vacuum chamber, with He gas pressure 10^{-3} Pa (77 K Shield)

We want to calculate:

- a) Radiant heat inflow from the brass to the copper chamber
For tarnished brass $\epsilon_1 \rightarrow 1$, for polished Cu $\epsilon_2 \rightarrow 0.03 \ll \epsilon_1$

$$\dot{Q} = \sigma A (T_1^4 - T_2^4) \frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2 - \epsilon_1 \epsilon_2} \cong \sigma A (T_1^4 - T_2^4) \epsilon_2$$

$$\dot{Q} = 5.67 \cdot 10^{-8} \text{ W}/(\text{m}^2 \text{ K}^4) \times 0.05 \text{ m}^2 \times 0.03 \times (77^4 - 4.2^4) \text{ K}^4 = 0.003 \text{ W}$$

- b) Radiant heat inflow down the stainless steel tube to the inner chamber (no baffle)

Rough approximation: max value for perfect specular reflection from the inner wall of the tube (perfect ‘channeling’) → all radiation reaches the inner space

$$\dot{Q} = \sigma A_t T_a^4 = 5.67 \cdot 10^{-8} \times \pi \times (0.01)^2 \times 295^4 = 0.135 \text{ W}$$

- c) Heat conducted down the stainless steel tube from the outer to the inner chamber

For the heat conducted by the stainless steel tube ($77K \rightarrow 4.2K$: $\bar{\kappa} = 4.5 \text{ W}/(\text{m K})$):

$$\dot{Q} = \bar{\kappa} \frac{A}{l} \Delta T = 4.5 \times \frac{2\pi \times 0.0003 \times 0.01}{0.06} \times 72.8 = 0.103 \text{ W}$$

- d) Heat conducted down the electrical leads from one chamber to the other
Cu $\rightarrow \phi \cong 0.1 \text{ mm}$. $\bar{\kappa} \cong 980 \text{ W}/(\text{m K})$, $l = 0.12 \text{ m.}$, $\Delta T = 72.8 \text{ K}$
Costantana $\rightarrow \phi = 0.2 \text{ mm}$, $\bar{\kappa} = 14 \text{ W}/(\text{m K})$, $l = 0.12 \text{ m.}$, $\Delta T = 72.8 \text{ K}$

$$\dot{Q} = (\bar{\kappa}_1 A_1 + \bar{\kappa}_2 A_2) \frac{\Delta T}{l} = 0.039 \text{ W}$$

- e) Heat conducted through the low-pressure helium gas in inter-chamber space

Low-pressure He gas. Considering parallel surfaces with the same area and accommodation coefficient of 0.6

$$\dot{Q} = \text{const.} \bar{a} A p \Delta T = 2.1 \times 0.6 \times 0.05 \times 1E - 3 \times 72.8 = 4.6 \text{ mW}$$

- f) Examine whether Joule heating due to a current of 5 mA in the electrical leads is serious in view of the other heat inflows

Joule heating due to a current of 5 mA in the electrical leads 8 Cu wires $\phi \cong 0.1 \text{ mm}$, $l = 0.12 \text{ m}$, $R \cong 0.46 \Omega/\text{m}$ at 77 K and $R \cong 0.03 \Omega/\text{m}$ at 4.2 K \rightarrow average resistance $0.25 \Omega/\text{m}$ in the range 77-4K 4 constantan wires $\phi \cong 0.2 \text{ mm}$, $l = 0.12 \text{ m}$, $R \cong 14 \Omega/\text{m}$

$$\dot{Q} = 8 \times (5E - 3)^2 \times 0.25 \times 0.12 + 4 \times (5E - 3)^2 \times 14 \times 0.12 = 0.17 \text{ mW}$$

In summary: $\dot{Q}_{\text{tot}} = 0.28 \text{ W} \rightarrow$ evaporation of about 320 cm^3 of liquid helium per hour ($\dot{Q}V_m/L = 3600$) If the evaporated gas is used for cooling while it goes up, the needed amount of LHe can be reduced significantly, up to a factor 10

2.5.4 Selection of the Material with the Appropriate Cryogenic Thermal Conductivity

Following the descriptions above, one can conclude:

For good thermal conductivity the right choices are

- Cu (but: soft; nuclear specific heat at $T < 0.1\text{K}$)
- Ag (but: soft; expensive)
- Al (but: soft, superconducting below 1K; soldering only possible in an elaborate process).
- The highest practical conductivities of these metals are $\kappa \approx 10 \text{ } T[\text{WK}^{-1}\text{cm}^{-1}]$ if they are very pure; more typical is $\kappa \approx T[\text{WK}^{-1}\text{cm}^{-1}]$

For thermal isolation the right choices are

- plastics (Teflon, Nylon, Vespel, PMMA, etc.)
- graphite (careful, there exists a wide variety)
- Al_2O_3 (alumina)
- thin-walled tubing from stainless steel (but: soldering not easy) or from $\text{Cu}_{0.7}\text{Ni}_{0.3}$ (easier to solder); both can be slightly magnetic at low temperatures

In general, glasses or materials composed of small crystallites (for phonon scattering) and containing a lot of defects and impurities (for electron scattering) are good thermal insulators.

If other properties do not matter too much, aluminium alloys or brass should be used because of their relatively low prices and, above all, because they can be easily machined.

- **For low-current leads** thin **Constantan** ($\rho_{300K} = 52.5 \mu\Omega \text{ cm}$, $\rho_{4K} = 44\mu\Omega \text{ cm}$) or **Manganin** ($\rho_{300K} = 48\mu\Omega \text{ cm}$, $\rho_{4K} = 43\mu\Omega \text{ cm}$) wires should be used because of their low thermal conductivity and the small temperature dependence of their electrical resistivity
- *For large electrical currents* (e.g. for superconducting magnets) the advantage gained by using a good conductor and a large wire diameter to reduce Joule heating, and the disadvantage of the then increased thermal conductivity have to be carefully considered. Often one may end up using **Cu wires with a proper heat sinking** at various places on their way to low temperatures
- *At $T < 1K$* the use of **superconducting wires** with their vanishing thermal conductivity for $T \rightarrow 0$ is the right choice (often it is adequate just to cover a Manganin or Constantan wire with a thin layer of superconducting solder).
- *At $T < 0.1K$* one can use monofilamentary NbTi wires
- Wires for measurements of small signals have to be **twisted pairwise** on their way in the cryostat, **rigidly fixed** and **well shielded** to keep pick-up signals low (coaxial cryogenic cables with proper heat-sinking of leads)

2.5.5 Heat Switches

To thermally connect and disconnect various parts of a low-temperature apparatus

Gaseous Heat Switches

The simplest way is to use a gas for thermal coupling and then remove it by pumping. This method is often employed in precooling the inner parts of a cryostat.

Even low gas pressure of 10^{-4} bar is sufficient for an adequate heat transfer. But usually many hours of pumping are then required to reduce the gas pressure for sufficient thermal isolation.

- Warning: if the exchange gas has not been pumped to a low enough pressure, time-dependent heat leaks due to a continuing desorption and condensation of the remaining gas at the coldest surfaces may result
- For He, the superfluid film contributes to the heat transfer, too.

Mechanical Heat Switches

Thermal contact is made by metallic, usually gold-plated contacts pressed together mechanically. The main disadvantages of these switches are the large forces (typically 100 N) necessary to make adequate thermal contact and the heat generated when the contact is open. They are adequate for calorimetry at $T > 1\text{K}$ and for special applications (e.g. for space missions).

Superconducting Heat Switches

- The thermal conductivity K_S of a metal in the superconducting state can become very small because the number of electrons decreases exponentially with temperature; it can be orders of magnitude smaller than the thermal conductivity K_n of the same material in the normal state.
- Some metals can easily be switched from the superconducting to the normal state by applying a magnetic field → superconducting heat switch.
- Superconducting heat switches are the most common thermal switches at temperatures below about 1K.
- Advantages: the heat flow in the open state is small, they are very easy to switch, the switching ratio κ_n/κ_s can be very large for $T < T_c/10$.
- In fact, $\kappa_n \propto T$ whereas $\kappa_s \propto T \exp(-\Delta E/k_B T)$ for $T > T_c/10$ (from the remaining unpaired electrons) and $\kappa_s \propto (T/\Theta_D)^3$ at $T < T_c/10$ (from the now dominating phonons); in this last case $\kappa_n/\kappa_s \approx 0.05(\theta/T)^2$ for high purity thin Al foils.
- Al is a good candidate because of its high κ_n and large Θ_D (400 K), it is easily available in very high purity, has a convenient critical field (10.5 mT), good durability and it is easy to handle (but problems with contacts due to the tenacious surface oxide on aluminum).
- The behavior of the switch may be deteriorated by frozen-in magnetic flux from the switching field, which may cause parts of the metal to remain in the normal state when the field is removed. This can be avoided by orientation of at least part of the metal perpendicular to the field, so that the normal cores of trapped flux lines will not short-circuit the switch material.

2.5.6 Thermal Boundary Resistance

Boundary Resistance Between Metals

To achieve thermal equilibrium in a system becomes *increasingly more difficult when the temperature is lowered*, not only because the thermal conductivity of materials decreases with decreasing temperature but also because the **thermal boundary resistance at the interface between two materials** becomes

increasingly important.

If we have two different materials in contact and heat \dot{Q} has to flow from one material to the other, for example in a cooling process, there will be a temperature step at the boundary between them. This temperature step is given by

$$\Delta T = R_K \dot{Q}$$

where R_K is the thermal boundary resistance, or **Kapitza resistance** (mechanism not fully understood, at least for very low temperatures).

Between metals, the actual contact area often is only about 10^{-6} of the apparent one, due to the microscopic irregularities of the opposing surfaces; R_K scales (inversely) with the **actual contact area**, that can be increased by pressure.

- The boundary resistance can be kept reasonably small, if the surfaces are **clean**, possibly **gold-plated**, and **pressed together with a high force**. We should then have an overlap of the electronic wave functions of the two metals, giving a good electric and thermal flow between them
- Extremely small contact resistances (from $0.1 \mu\Omega$ down to $10 n\Omega$ at 4.2K) between gold-plated Cu discs bolted together with stainless-steel screws (contact resistance inversely proportional to the tightening torque on the screws)
- The addition of In **foil** or **Apiezon grease** between the contact surfaces of Cu, Al, brass or stainless steel can result in improvements of up to an order of magnitude

The mechanical and electrical contact between two metals often is made by soldering them together. Unfortunately, most solders, in particular **soft solders**, become superconducting at low temperature: good for electrical contact, but eventually behave like a dielectric with regard to thermal conductivity.

Boundary Resistance Between Liquid Helium and Solids

- Between dielectrics, for example a nonmagnetic dielectric in contact with liquid or solid helium, the transfer of energy can only occur via phonon transmission.
- A temperature step ΔT at the interface arises from the acoustic mismatch of the two materials, which can be treated in analogy to optics (Snell's law).
- For helium/solid interfaces the situation is particularly severe because acoustic impedances are 3 orders of magnitude different, $Z = \rho_s v_s \approx 10^6 g(cm^2 s)^{-1}$ for solids but $\rho_h v_h \approx 10^3 g(cm^2 s)^{-1}$ for liquid helium (v is the velocity of phonons).
- For the angles α at which the phonons cross the boundary: $\sin \alpha_h / \sin \alpha_s = v_h / v_s$.

- Because $v_h \approx 238 \text{ ms}^{-1}$ for ${}^4\text{He}$ at $T \leq 1\text{K}$ whereas $v_s \approx 3000 - 5000 \text{ ms}^{-1}$ for metals, the critical angle of incidence at which phonons from helium may enter the solid is very small, $\alpha_{crit} = \arcsin(v_h/v_s) \cong 3^\circ$.
- The fraction of phonons hitting the interface that fall into the critical cone is $f \cong \sin^2(\alpha_{crit})/4 \cong 10^{-3}$.
- However, because of the difference in acoustic impedance $Z = \rho v$, not even all of these phonons are transmitted, but only a fraction $< 10^{-5}$.
- Hence the two bodies are rather well isolated from each other. *The combination of acoustic mismatch and a small critical angle severely limits the energy exchange between helium and other materials, at low temperatures.*

Different approaches to the problem essentially give the result that the boundary resistance scales as

$$R_K \propto (AT^3)^{-1}$$

- This prediction is in reasonable agreement with most experimental data for $0.02\text{K} < T < 0.2\text{K}$ with typical values of $AR_KT^3 \cong 10^{-2} \text{ m}^2 \text{ K}^4 \text{ W}^{-1}$ for liquid and solid helium in contact with metals, but deviates considerably both in the Kelvin temperature range and at $T < 10\text{mK}$ (unexplained)
- *A way to reduce R_K is to increase the actual contact area, usually by means of sintered metal powders*

In summary, even though much remains to be understood, the Kapitza resistance problem can be divided into three distinct temperature regimes:

- *Above 1 K: R_K is essentially the same for liquid and solid ${}^3\text{He}$ and ${}^4\text{He}$; it is an order of magnitude smaller than predicted by the theory (not understood)*
- *For $20\text{mK} \leq T \leq 200\text{mK}$: $R_K \propto T^{-3}$ and behaves as predicted by the acoustic mismatch theory if well characterized, clean, bulk metallic surfaces are used*
- *At $T \leq 10\text{mK}$: $R_K \propto T^{-2}$ or T^{-1} between liquid ${}^3\text{He}$ or helium mixtures and metals, and is again much smaller than predicted by the theory (here probably due to a magnetic dipole coupling between the ${}^3\text{He}$ nuclear moments and electronic moments in the solid together with a coupling of helium phonon modes to soft vibrational modes, if a sintered metal is used)*

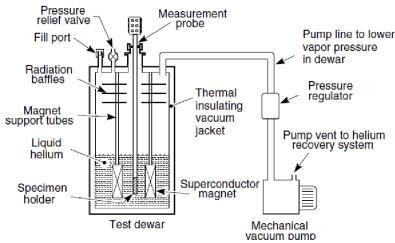


Figure 2.4: Schematic of an immersion measurement probe in a cryogenic liquid with a vacuum pump for lowering the boiling temperature of the cryogen bath.

2.6 ^4He Cryostats and Cryogenic Technology

2.6.1 Key Technological Requirements and Challenges

- **Isolation:** The main isolation of cold parts from warm parts is achieved via vacuum isolation. Consequently, cryotechnology always involves vacuum technology, including designing vacuum-tight equipment, techniques for avoiding leaks, and technologies for soldering, welding, and gluing (or finding and fixing leaks if they occur).
- **Mechanical Stress:** Extreme mechanical stresses arise from large temperature gradients and the differing thermal expansion coefficients of the various materials used.
- **Testing:** Equipment is built and tested at room temperature, but defects may emerge only at, or due to, low temperatures (T), yet they cannot be removed at low temperature.
- **Safety:** Containers holding cryogenic liquid must either be designed to withstand high pressure or, preferably, contain safety features (like safety valves) to ensure pressure cannot rise to dangerous values following an accident. Safety regulations, which are legally enforced in several countries, must be followed.

2.6.2 Use of Liquid ^4He in Low-Temperature Equipment

Every refrigeration process aimed at a temperature of 10 K or lower requires the use of liquid helium in either the final or pre-cooled stage. Because liquid helium is an expensive cryoliquid, minimizing its consumption is important.

Cool-Down Phase Efficiency

The heat of evaporation for helium transforming from liquid to gaseous phase is about 2.6 kJ/l at 4.2 K. This is significantly less than the enthalpy of helium gas between 4.2 K and 300 (77) K, which is 200 (64) kJ/l.

Cooling Al from 300 → 77 K	N ₂ : 1.0 (0.63) l
Cooling Al from 77 → 4.2 K	⁴ He: 3.2 (0.20) l
Cooling Al from 300 → 4.2 K	⁴ He: 66 (1.6) l

Table 2.3: Amount of cryoliquid [l] necessary to refrigerate 1 kg of aluminum (Al) when comparing the use of only latent heat versus latent heat plus the enthalpy of the gas (values in parentheses)

- If 1 W of heat is applied, 1.4 lHe/h will evaporate if only the heat of evaporation is used, but only 0.017 lHe/h evaporates if the enthalpy of the cold gas between 4.2 K and 300 K is utilized.
- It is thus very important to use the *enthalpy of the cold helium gas* after the liquid transitions to the gaseous state when cooling the equipment, and the gas should leave the cryostat at a temperature as close as possible to room temperature.

Precooling

Precooling with liquid nitrogen (LN₂) from 300 K to 77 K saves a large amount of liquid helium. N₂ is almost an order of magnitude cheaper than LHe and has about 60 times the latent heat of evaporation.

Look at *Table 2.3*.

A crucial precaution during precooling is ensuring that really all liquid nitrogen has been removed before liquid helium is transferred, as LN₂ has a large specific heat, and cooling it to the low Kelvin temperature range would consume a sizeable quantity of LHe.

Running Phase

When the equipment has reached the required low temperature, to maintain a stationary state heat transferred from external sources has to be compensated by the cooling power provided by the cryo liquid, which now mainly comes from the heat of evaporation. (The main sources of heat for experiments in the Kelvin temperature range have been considered above).

2.6.3 Cryostat Types

Double-Walled Glass Dewars

These systems consist of nested, double-walled glass dewars.

- **Advantages:** They have a reasonably low price, feature the low thermal conductivity of glass, and allow the level of the cryogenic liquid to be easily seen.
- **Disadvantages/Safety:** Glass breaks easily. A small leak into the vacuum space can allow air to enter and condense onto cold surfaces. Upon

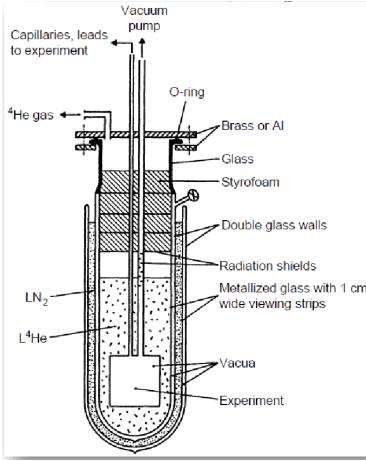


Figure 2.5: A double-walled glass dewar for LN_2 and L^4He . The inner surfaces of the glass are covered with aluminium or silver films to block thermal radiation (except for a vertical 1cm wide viewing strip). Four or five radiation baffles force the exiting cold gas to flow along the dewar walls, thus cooling it as well as the pumping tubes. The effectiveness of the baffles is increased by filling the spaces between them with Styrofoam.

warming, the condensed air evaporates, and if it cannot escape quickly, the pressure buildup may cause the glass dewar to explode, necessitating protective containers.

- **Vacuum Maintenance:** Helium diffuses through glass walls at room temperature (typically $10^{-12} - 10^{-10} \text{ cm}^3\text{s}^{-1}$ through 1 cm^2 of 1 mm thick glass if $\Delta p = 1 \text{ bar}$). To maintain the vacuum, helium gas from the inner volume should be pumped out immediately after warming the system to room temperature.
- **Thermal Shielding:** The inner surfaces of the glass are often covered with aluminum or silver films to block thermal radiation, except for a vertical 1 cm wide viewing strip. Radiation baffles (four or five) reduce thermal radiation and force the exiting cold gas to flow along the dewar walls, cooling the walls and the pumping tubes. Styrofoam can fill the spaces between baffles to increase their effectiveness.

Metal Dewars

- **Advantages:** A metal dewar is more rugged and can withstand higher pressures or stresses than glass. It offers greater flexibility for more complex designs and avoids the helium diffusion problem. Is the most used.

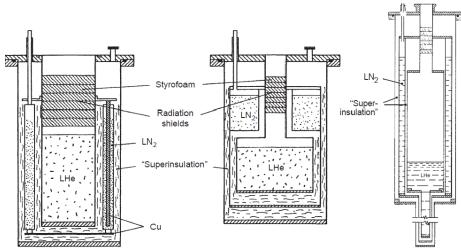


Figure 2.6: Metal Dewar

- **Design/Materials:** A disadvantage is the higher price. They are usually constructed from stainless steel or, increasingly, a combination of aluminum and fibreglass. They generally do not have two vacuum spaces.
- **Shielding/Pre-cooling:** Many modern metal cryostats omit LN_2 vessels for pre-cooling and radiation shielding to avoid vibrations produced by constantly boiling LN_2 . In such cases, superinsulation layers are necessary.
- **Superinsulation:** This consists of thin plastic foil (Mylar) onto which a reflective layer of aluminum (Al) has been evaporated, resulting in an emissivity coefficient of about 0.06. These sheets are wrapped around the LHe vessel, acting as radiation shields at continuously decreasing temperatures from the outer to the inner layers.
- **Further Improvements:** Metallic radiation shields can be placed between the room-temperature vessel and the LHe vessel, with tubing soldered to them to vent the evaporating cold He gas before it leaves the cryostat.
- **Performance:** A good, simple helium dewar should have an evaporation rate of no more than 0.1 l/h (or up to 1 l/h for very large dewars).

He Gas-Flow Cryostats ($T > 4.2 \text{ K}$)

For measurements above the normal boiling point of liquid ${}^4\text{He}$ ($T > 4.2 \text{ K}$), using the main He bath at 4.2 K as the temperature reservoir is uneconomical.

- **Principle:** It is much more efficient to use the cold gas evaporating from liquid helium and leverage its enthalpy for cooling the experiment in a continuous gas-flow cryostat.
- **Configuration:** This design allows the liquid helium storage vessel to be physically separated from the cryostat containing the experiment. The cooling power and temperature are controlled by regulating the flow rate via a needle valve.

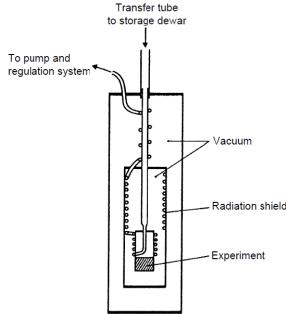


Figure 2.7: Evaporation cryostat for the temperature range $T > 5K$. A mixture of liquid and gaseous 4He is transferred via a transfer tube from the storage dewar to the cryostat. The mixture is pumped through a spiral tube which is first connected to a chamber on which the experiment is mounted and then soldered to a radiation shield before it leaves the cryostat. The temperature of the experiment can be regulated by a heater and/or a valve which regulates the helium flow.

- **Advantages:** Low consumption of cryogenic liquid (especially if $T > 10$ K), a temperature that is variable in a wide range up to room temperature, and very fast cool-down and warm-up times.
- **Operation:** In an evaporation cryostat for $T > 5$ K, a mixture of liquid and gaseous He is transferred via a transfer tube from a storage dewar. The mixture is pumped through a spiral tube connected first to a chamber where the experiment is mounted, and then soldered to a radiation shield before exiting. Temperature regulation can use a heater and/or a valve controlling the helium flow.

Use of the Temperature Gradient Above the LHe Bath (Dipper Probes)

For measurements where $T > 4.2$ K, a cryogenic probe (variable temperature insert) holding a small experiment chamber can be moved within the vapors above the liquid helium bath.

- **Principle:** The sample-holder temperature is varied by changing the probe's height above the liquid surface, utilizing the natural temperature stratification that occurs in the vapor space. The temperature gradient ranges from 4.2 K near the liquid surface up to 200 K or more near the terminal part of the dewar "neck".
- **Stability:** The temperature gradient is generally rather "long" (over 1 m) and very stable, enabling easy stabilization at the desired temperature, particularly at low T.

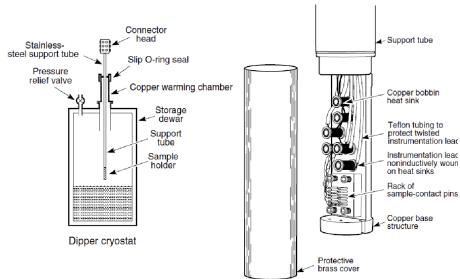


Figure 2.8: Sample-holder section of a dipper probe, showing copper base structure and sample holder for temperature stability, and copper-bobbin heatsinking arrangement for the instrumentation leads

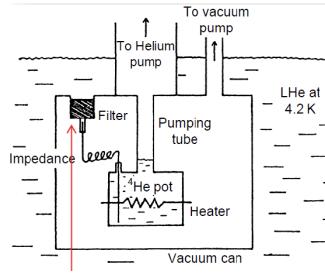


Figure 2.9: Cryostat with Variable Temperature for $1.3\text{ K} \leq T \leq 4.2\text{ K}$

- **Efficiency/Versatility:** This system uses the enthalpy of the gas for cooling the insert, consuming very little helium. The dipper probe is considered the simplest, most common, useful, and versatile type of cryostat. A significant advantage is the ease with which it can often be inserted into a storage dewar, avoiding the need to transfer cryogenic fluid into a separate vessel.

2.6.4 Cryostats for Variable Low Temperatures ($1.3\text{ K} \leq T \leq 4.2\text{ K}$)

This temperature range is bounded by the normal boiling point of ${}^4\text{He}$ (4.2 K) and the temperature at which its vapor pressure becomes very small (1.3 K). This range is accessed using pumped liquid ${}^4\text{He}$.

Pumping on the Main ${}^4\text{He}$ Bath

This approach is very uneconomical because approximately 40% of the liquid He must be evaporated simply to cool the remaining liquid from 4.2 K to 1.3 K, due to the large change in its specific heat over this range.

Continuously Operating ^4He Evaporation Cryostat

Since the specific heat of solids is small in this T range, cooling them from 4.2 K to 1.3 K requires evaporating only a small fraction of liquid He. It is much more efficient to pump on only a small fraction of the liquid contained in a separate container (a He pot).

- **Flow and Expansion:** This small fraction flows from the main 4.2 K bath through a suitable flow impedance (capillary) into a small vessel (several cm^3) located within a vacuum can inside the cryostat. The liquid, initially at 1 bar in the main bath, is isenthalpically expanded through the impedance and arrives at a lower temperature in the evaporation vessel.
- **Impedance:** The required pressure drop (1 bar down to about 1 mbar) can be achieved using a capillary (1 to several meters long with an inner diameter of about 0.05 or 0.1 mm), a shorter capillary (10 – 20 cm) with a tightly fitting wire inserted, or a needle valve controlled by an external motor.
- **Filtration:** Impurities like frozen air in the main LHe bath could block the fine capillary impedance, necessitating a filter (e.g., of Cu powder).
- **Self-Regulation:** The vessel continues to fill until the liquid level in the pumping tube reaches a height (h) where the cooling power available from the latent heat (L) of evaporation is precisely balanced by the total heat transferred (heat from the main bath through the liquid column + heat from the experiment).
 - If the external load is increased, the liquid level in the pumping tube drops, which reduces the heat contribution from the main bath. Therefore, the refrigerator is self-regulating, and its temperature remains fairly constant at about 1.3 K even when the heat load is varied.
 - The refrigerator described can typically operate at a constant temperature up to a heat input of about 3 mW (a value that can be changed by adjusting the impedance). At higher heat inputs, the device enters an overload state until the pumped ^4He pot is empty.
- **Higher Temperatures ($T > 4.2 \text{ K}$ in Pot):** In a stationary state at $T > 4.2 \text{ K}$, the refrigerator operates with an empty ^4He pot at higher temperatures. For temperatures $T > 4.2 \text{ K}$, balance can be maintained using electronic control between the heater power and the flow rate of the pump, regulated through a needle valve.

2.6.5 Auxiliary Equipment

Storage Vessels

These vessels are used for the transport of liquid helium from the supplier, typically having volumes from 30 to 100 l.

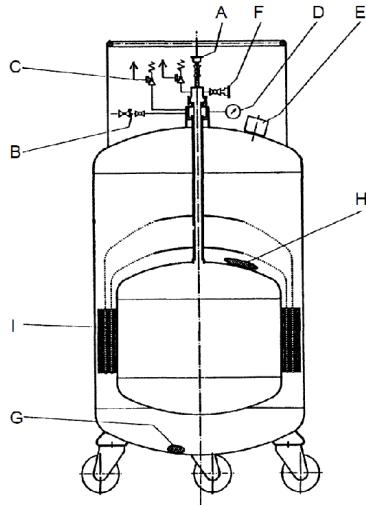


Figure 2.10: Commercial storage vessel for liquid 4He (A: connection for transfer tube, B: overflow valve, C: safety valve, D: manometer, E: vacuum and safety valves, F: gas valve, G: getter material, H: adsorbent material to maintain and improve the vacuum, I: superinsulation only partly shown).

- They are vacuum-isolated and utilize superinsulation instead of LN_2 shielding.
- Materials used include aluminum (to keep weight low) or stainless steel (for more rugged applications).
- Typical evaporation rates are about 1% per day.
- Features include safety features to avoid overpressure, such as a safety valve, overflow valve, and manometer.

Transfer Tube

A double-walled vacuum transfer tube is required for transferring liquid helium from the storage vessel to the experimental cryostat.

Level Detectors

- **Acoustic Level Detection:** A thin stainless steel tube is lowered into the cold gas or cryogenic liquid. Liquid evaporation increases pressure, causing oscillations transferred to a diaphragm (rubber or plastic sheet). High amplitude/high frequency oscillations occur when the tube is in the

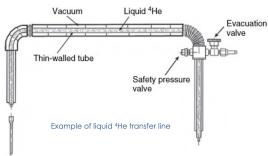


Figure 2.11: Transfer tube

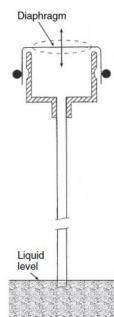


Figure 2.12: Level detector

gas, which are easily distinguishable from the low amplitude/low frequency oscillations when the tube is lowered all the way into the liquid.

- **Resistive Level Detection:** This utilizes the distinct temperature dependence of resistance in conductors.
 - Semiconductor or carbon resistors can be used, but the current may overheat the resistor if it is raised into the gas phase.
 - A commercial superconducting wire (e.g., NbTi or a normal-conducting wire with a superconducting coating like PbSn solder) with a transition temperature between 5 and 10 K is common. In LHe, it exhibits zero resistance; above the liquid, it becomes resistive, making the total resistance a direct measure of how much of its length is outside the liquid.
- **Capacitive-Level Detection:** The liquid level is determined by measuring the capacitance of two concentric tubes partially immersed in the liquid. This works because the dielectric constant of liquid ^4He is 1.0572.

2.7 ^3He cryostats: refrigeration below 1K

The temperature range achievable by standard cryogenic methods can be extended down to about **0.3 K** if the rare isotope ^3He is used. This range places ^3He evaporation among several advanced refrigeration methods operating below 1 K, including dilution refrigeration, nuclear demagnetization, Pomeranchuk cooling, and electron demagnetization.

2.7.1 Advantages and Disadvantages of ^3He Refrigeration

The use of ^3He offers several key advantages over ^4He in the sub-Kelvin range:

- **Vapour Pressure:** ^3He has a substantially larger vapour pressure than ^4He at the same temperature, with a ratio of about 10^4 at 0.5 K.
- **Specific Heat:** The specific heat of liquid ^3He is larger and varies much less between 2 K and 0.5 K compared to liquid ^4He . This results in a larger heat reservoir in this temperature range. Consequently, only about 20% of ^3He must be evaporated to cool the liquid from 1.5 K to 0.3 K using its own heat of evaporation.
- **Superfluidity:** In the relevant temperature range, ^3He is not superfluid, thus avoiding heat transfer problems that arise from superfluid film flow.

However, there are significant disadvantages associated with ^3He :

- **Latent Heat:** The ^3He latent heat of evaporation (L) is substantially smaller than that of liquid ^4He .
- **Cost:** ^3He is much more expensive than ^4He .

Due to its high cost and smaller latent heat, ^3He must be utilized in a closed gas handling and cryogenic system to ensure no gas is lost. ^3He is transformed from gas to liquid by bringing it into contact with a ^4He bath, often maintained at a temperature of about 1.3 K.

For effective refrigeration below 1 K, all parts cooled by ^3He are typically surrounded by shields maintained at the ^4He temperature. Furthermore, tubing and wiring leading to the experiment are thermally heat sunk at the ^4He bath, which has a larger volume and a larger heat of evaporation, allowing it to absorb heat transferred from the outside world.

2.7.2 Types of ^3He Cryostats

^3He cryostats are generally categorized based on their pumping mechanism:

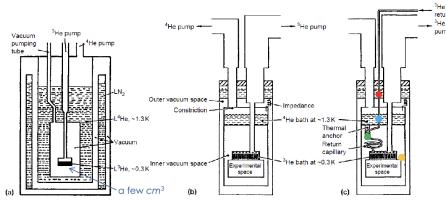


Figure 2.13: Cryostat with external pump: ^3He cryostats of increasing sophistication. (a) Non-recirculating ^3He refrigerator with a pumped main ^4He bath. (b) Non-recirculating ^3He refrigerator with a continuously operating ^4He evaporator. (c) Recirculating ^3He refrigerator with a continuously operating ^4He evaporator

Cryostats with External Pumps

For this purpose: heat exchangers in the 4.2K bath and in the 1.3K bath . The now liquid ^3He will then be isenthalpically expanded through an impedance of order $10^{12}\text{--}10^{13}\text{cm}^{-3}$, which maintains a pressure sufficient for condensation before the ^3He arrives as a low-pressure, low-temperature liquid at the pumped ^3He pot.

When all the ^3He has condensed in the pot, we pump on this liquid to reduce its temperature from about 1.3K to the desired temperature(minimum typically 0.3 K).

In (c) the ^3He refrigerator is run in a continuous mode by the recondensing tube. The ^3He vapour that we pump away will leave the room-temperature pump at a pressure of several 0.1 bar, it is pre-cooled by the ^4He pot and eventually recondensed into the ^3He pot.

The narrow capillaries used as impedances can easily be blocked by impurities (frozen air)→ LN_2 cooled trap after the ^3He room-temperature pump.

- vacuum-tight closed ^3He system with a sealed pump avoiding any loss of this expensive gas.
- careful design of the room-temperature part with the smallest possible volumes, so that not too much of the ^3He gas remains unused in these “dead” volumes.
- suitably dimensioned pumps and pumping tubes, to circulate the required amount of ^3He :
 - for $T_{fin} = 0.5 \text{ K}$: $p_{fin} \cong 0.2 \text{ mbar}$ mechanical pump
 - for $T_{fin} = 0.3 \text{ K}$: $p_{fin} \cong 2 \mu\text{bar}$ oil-diffusion pump + mechanical pump (or turbo).

Often the limitation is not the pump but rather the conductance of the pumping tubes. If we need a cooling power of $\dot{Q} = 1 \text{ mW}$, this requires

an evaporation rate of

$$V = \frac{\dot{Q}}{L} \cong 5 \text{ cm}_{\text{liq}}^3 {}^3\text{He} \text{ h}^{-1} \cong 3 \ell_{\text{gas}} {}^3\text{He} \text{ h}^{-1} \text{ at } p = 1 \text{ bar},$$

but at the mentioned p_{fin} the volume flow rate have to be

$$\begin{aligned} - p_{fin} &\cong 0.2 \text{ mbar} \rightarrow V \cong 15 \text{ m}_{\text{gas}}^3 {}^3\text{He} \text{ h}^{-1} \\ - p_{fin} &\cong 2 \mu\text{bar} \rightarrow V \cong 1500 \text{ m}_{\text{gas}}^3 {}^3\text{He} \text{ h}^{-1} \end{aligned}$$

- These volume rates are not a problem for the pumps, provided that the tubing is adequate. The conductance $L^* [\text{m}^3 \text{ h}^{-1}]$ of a tube for a laminar flow is given by

$$L^* = 486 \frac{p \cdot d^2}{l},$$

where d is the diameter [cm], l is the length [cm], and p is the mean pressure [mbar]. We then find that we need pumping tubes with diameters of 3 and 10 cm, respectively, if the length is several meters (**bulky!**).

- Inside the cryostat, the diameter of the pumping tube can be reduced according to the temperature profile because the density of the evaporating gas increases with decreasing T and the circulation rate is the same everywhere.

Cryostats with Internal Adsorption Pumps

- One can avoid the room-temperature pump as well as the often-bulky pumping tubes by inserting a cold adsorption pump inside the cryostat.
- Working principle: gases adsorb at cold surfaces if their temperature is low enough.
- Maximize surface/volume ratio: various suitable materials, e.g., charcoal, zeolites, or fine metal powder with surface areas of at least several $\text{m}^2 \text{ g}^{-1}$.
- A volume of several cm^3 with such an adsorbent at low temperature: it will very effectively pump the liquid ${}^3\text{He}$ bath.
- When all the ${}^3\text{He}$ has been pumped away, so that the ${}^3\text{He}$ pot is empty, the charcoal pumping system is lifted into a space at higher temperature in the cryostat to desorb the helium, which will then enter the gas phase, condense at the cold surfaces of the cryostat and eventually drip back down into the ${}^3\text{He}$ pot.

The charcoal pump can also be equipped with an heater so that its temperature can be regulated to adsorb or desorb the ${}^3\text{He}$.

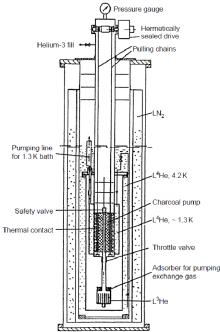


Figure 2.14: cryostat with a hermetically sealed ${}^3\text{He}$ system, reaching 0.25 K. The charcoal pump can be raised or lowered by means of a chain drive operated via a hermetically closed rotating seal

- Cooling power of a ${}^3\text{He}$ refrigerator pumped by 20 g of charcoal. One can maintain temperatures as low as 0.25 K for cooling powers below about 0.01 mW, and 0.4 K at 1 mW.
- “Pumping power” of activated charcoal: adsorption isotherms of ${}^3\text{He}$ on activated charcoal as a function of the helium-gas pressure.
- The pumping speed of charcoal pumps is not only a function of its temperature but also, in practice, a complicated function of geometry, thermal coupling and pressure.
- Cooling to 0.3 K can be achieved e.g. within a few minutes.
- Cold charcoal pumping systems are more efficient than room-temperature mechanical pumps because they are connected via a short, cold pumping tube to the ${}^3\text{He}$ pot, taking advantage of the very high pumping speed of the adsorbent.

Example: ${}^3\text{He}$ Cryostat with Two Internal Adsorption Pumps

A laboratory example utilizes two internal adsorption pumps: a Main sorb (for ${}^3\text{He}$) and a Mini sorb (for ${}^4\text{He}$). The cycle involves several steps (with example temperatures noted in Kelvin) leading to condensation and cooling:

1. Both sorb heaters are switched on (Main sorb at 330.0 K, Mini sorb at 330.0 K) and gas is pumped away.
2. A small amount of liquid ${}^4\text{He}$ is introduced into the VTI (Vacuum Tight Insert).
3. The VTI is pumped, and a small amount of ${}^4\text{He}$ gas enters the cone seal can, causing the ${}^3\text{He}$ pot temperature to begin decreasing ($T_{\text{He}3 \text{ pot}} \approx 280.4 \text{ K}$).

4. The VTI temperature is stabilized at about 2 K while pumping continues (Main sorb \approx 2.000 K, Mini sorb \approx 3.000 K), and the system waits for ^3He condensation, which takes about 30 min.
5. The heating power of the sorbs is switched off, and the system waits about 20 min.
6. Cooldown to 300 mK takes approximately 20 min, reaching a base temperature of **0.282** mK (or 0.298 K in another reading).

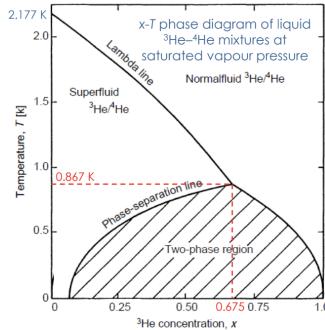


Figure 2.15: x-T phase diagram of liquid ${}^3\text{He}$ - ${}^4\text{He}$ mixtures at saturated vapour pressure (Descriptive labels: Si 2.177 kelvin, 2.0 Temperature, 1.5, Lambda line, Normalfluid ${}^3\text{He}/{}^4\text{He}$, 1.0, Superfluid ${}^3\text{He}/{}^4\text{He}$, 0.5, Si 0.867 kelvin, Phase-separation line, Two-phase region, 0.25, 0.50, 0.675, 0.75, 1.00 ${}^3\text{He}$ concentration, x)

2.8 Dilution Refrigerator

From 1962: in contrast to the helium refrigerators discussed above, where the latent heat of evaporation is used for cooling, it was suggested to use the heat of mixing of the two helium isotopes to obtain low temperatures. Today it is the most important refrigeration technology for the temperature range between about 5 mK (record of 2 mK) and 1 K, and it is the base from which lower temperatures can be reached.

2.8.1 The ${}^3\text{He}-{}^4\text{He}$ Dilution Refrigerator

Liquid ${}^3\text{He}-{}^4\text{He}$ Mixtures

In the liquid ${}^3\text{He}-{}^4\text{He}$ mixture, the respective concentrations of the two helium isotopes are expressed as $x = x_3 = n_3/(n_3 + n_4)$ and $x_4 = n_4/(n_3 + n_4)$ where n_3 (n_4) is the number of He^3 (He^4) atoms or moles.

Phase Diagram and Solubility

The temperature of the superfluid phase transition of liquid ${}^4\text{He}$ is depressed if we dilute the Bose liquid ${}^4\text{He}$ with the Fermi liquid ${}^3\text{He}$. At the concentration $x = 0.675$ and at $T = 0.867$ K the λ -line meets the phase separation line. Below this temperature the two isotopes are only miscible for certain limiting concentrations which depend on the temperature. The shaded region is a non-accessible range of temperatures and concentrations for helium mixtures.

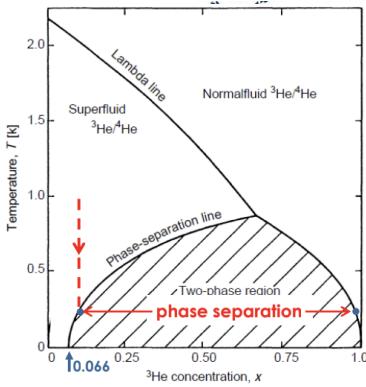


Figure 2.16: Phase diagram showing phase separation (Descriptive labels: 2.0 Temperature, T [K], 1.5, 1.0, 0.5, Lambda line, Normalfluid ${}^3\text{He}/{}^4\text{He}$, Superfluid ${}^3\text{He}/{}^4\text{He}$, Phase-separation line, Two-phase region, 0.066 ${}^3\text{He}$ concentration, x)

${}^3\text{He} - {}^4\text{He}$ Mixtures as Quantum Liquids

If we cool a helium mixture (with $x > 6.6\%$) to temperatures below 0.87 K, the liquid will eventually separate into two phases, one rich in ${}^4\text{He}$ and the other rich in ${}^3\text{He}$ (floating on top of ${}^4\text{He}$ because of its lower density). If the temperature is decreased to close to absolute zero, we see that the ${}^3\text{He}$ -rich liquid becomes pure ${}^3\text{He}$. But the great surprise occurs at the ${}^4\text{He}$ -rich side: here the concentration of the dilute isotope, ${}^3\text{He}$, does not approach zero for $T \rightarrow 0$, but rather reaches a constant concentration of 6.6% ${}^3\text{He}$ in ${}^4\text{He}$ at saturated vapour pressure. This finite solubility, is of utmost importance for ${}^3\text{He}$ - ${}^4\text{He}$ dilution refrigeration technology.

For a classical system, a finite solubility means $S > 0$. It was believed that the liquid helium isotopes – like other two component liquids – have to fully separate into the two pure liquids when the temperature is low enough to fulfil the third law of thermodynamics, so that the entropy of mixing is zero for $T = 0$. Of course, the helium liquids have to fulfil the third law as well, and they can do so even if they do not separate completely for $T \rightarrow 0$ because they are quantum liquids: for $T = 0$ the ${}^3\text{He}$ - ${}^4\text{He}$ mixtures are in their fully degenerate Fermi momentum ground state for the ${}^3\text{He}$ part (“one ${}^3\text{He}$ particle per state”) and with ${}^4\text{He}$ being superfluid, both with $S = 0$.

At $T < 0.5$ K liquid ${}^4\text{He}$ is almost totally condensed into this quantum mechanical ground state; there are essentially no excitations (phonons, rotons) left → its viscosity, entropy and specific heat go to zero. In a helium mixture at these temperatures the component ${}^4\text{He}$ acts as an “inert superfluid background”, which contributes to the volume of the liquid and to the effective mass of the dissolved isotope ${}^3\text{He}$ but has negligible heat capacity. The lighter isotope ${}^3\text{He}$ with its nuclear spin $I = 1/2$ is a Fermi particle, and it has to obey

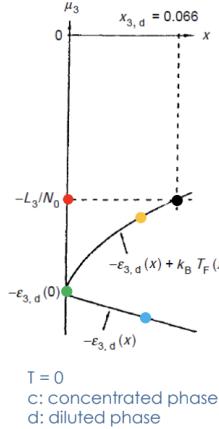


Figure 2.17: Diagram illustrating chemical potential (μ) vs. concentration (x). (Labels: μ_3 , $x_{3,d} = 0.066$, 0 , x , $-L_3/N_0$, $-\epsilon_{3,d}(x) + k_B T_F(x)$, $-\epsilon_{3,d}(0)$, $-\epsilon_{3,d}(x)$, $T = 0$, c: concentrated phase, d: diluted phase, ${}^3\text{He}$ in Pure ${}^3\text{He}$ ($x = 1$))

Fermi statistics and the Pauli principle like the conduction electrons in a metal. However, unlike the Fermi temperature of conduction electrons, which is of the order of 10^4 K, the Fermi temperature of liquid ${}^3\text{He}$ is of order 1K only (because $T_F \propto m^{-1}$). In analogy to the conduction electrons, the specific heat of liquid ${}^3\text{He}$ behaves as

$$c_3 = (\pi^2/2)(T/T_F)R \quad \text{at } T \ll T_F \quad (\text{Fermi-gas-like behavior})$$

or

$$c_3 = (5/2)R \quad \text{at } T > T_F \text{ and } p = \text{const.} \quad (\text{classic-gas-like behavior})$$

2.8.2 Finite Solubility of ${}^3\text{He}$ in ${}^4\text{He}$

The chemical potential (μ - binding energy) of pure liquid ${}^3\text{He}$ is given by the latent heat of evaporation, $\mu_{3,c} = -L_3$. So $\epsilon_{3,c} = L_3/N_A$ is the energy one has to supply to remove one ${}^3\text{He}$ atom from liquid ${}^3\text{He}$ into vacuum.

One ${}^3\text{He}$ Atom in Liquid ${}^4\text{He}$ ($x \approx 0$)

For the dilute phase, the binding energy of a ${}^3\text{He}$ atom in liquid ${}^4\text{He}$ ($x \rightarrow 0$) is given by $\mu_{3,d}(0)/N_A = -\epsilon_{3,d}(0) < -\epsilon_{3,c}$. The inequality holds because due to its smaller mass the ${}^3\text{He}$ atom has a larger zero-point motion than the ${}^4\text{He}$ atom → in the liquid phase ${}^4\text{He}$ atoms occupy a smaller volume. The ${}^3\text{He}$ atom will be closer to the ${}^4\text{He}$ than it would be to ${}^3\text{He}$ atoms or, in other words, its binding – due to the smaller distance – is stronger if it is in ${}^4\text{He}$ than it would be in ${}^3\text{He}$.

Many ^3He Atoms in Liquid ^4He ($x > 0$)

Increasing x , there is an attractive interaction between the ^3He atoms in liquid ^4He , due to:

- a magnetic interaction due to the nuclear magnetic moments of ^3He
- a density effect: ^3He has a larger zero-point motion → needs more space than a ^4He atom → the liquid near to a ^3He atom is more dilute → another ^3He atom feel this low-density region and would like to be combined with the first ^3He atom (to gain enough space for itself)

Due to this attractive interaction the binding energy increase with increasing ^3He concentration $x \rightarrow -\epsilon_{3,d}(x) < -\epsilon_{3,d}(0)$. BUT: ^3He atoms have to obey the Pauli Principle → if we put additional ^3He atoms in the liquid they have to go into successively higher energy states → eventually all the states up to the Fermi energy $E_F = k_B T_F$ are filled → the binding energy of the ^3He atoms has to decrease, due to their Fermi character, if x is increased. Eventually: $\mu_{3,d}(x)/N_A = -\epsilon_{3,d}(x) + k_B T_F(x)$. At $x = 6.6\%$ we have $\mu_{3,d}(6.6\%)/N_A = -L_3/N_A$. For higher x , the excess ^3He prefers to separate in a concentrated phase because it is energetically favorable for the whole system. As for ^4He : for the same reasons as the dilute ^3He atoms, they feel a stronger binding when surrounded by ^4He , and because they do not have to obey the Pauli principle there is no reason to decrease their binding energy if we increase their concentration → full separation.

2.8.3 Cooling Power of the Dilution Process

From measurements of specific heats, we know that the enthalpy of ^3He in the dilute phase is larger than the enthalpy of ^3He in the concentrated phase (its binding energy is higher); we have the heat of mixing:

$$\dot{Q} = \dot{n}_3 [H_d(T) - H_c(T)] \quad (2.1)$$

If we transfer ^3He atoms at the molar flow rate \dot{n}_3 from the concentrated phase into the dilute phase, cooling will result according to the enthalpy difference of the two phases. Because the enthalpy is given by

$$H(T) - H(0) = \int_0^T C(T) dT \quad (2.2)$$

(neglecting pV terms which do not matter here or in the following), we must have $C_{3,d}(T) > C_{3,c}(T)$ for the molar heat capacities of ^3He in the two phases to obtain cooling by the dilution process.

Experimentally, for the liquid ^3He : $C_{3,c} = 2.7 \cdot RT = 22 \cdot T [\text{J}(\text{mol K})^{-1}] \rightarrow H_{3,c}(T) = H_{3,c}(0) + 11 \cdot T^2 [\text{J mol}^{-1}]$.

In the mixtures, where the ^3He is diluted by ^4He , we have a weakly interacting

Fermi liquid, for which it is a good approximation to take the equations for a Fermi gas and replace the bare ${}^3\text{He}$ mass m_3 by the effective mass m^* (reflecting the influence of the neighboring ${}^4\text{He}$ and ${}^3\text{He}$ atoms with which the ${}^3\text{He}$ interacts). We then have at $T < T_F/10$ the following relation for the specific heat of the mixture per mole of ${}^3\text{He}$:

$$C_{3,d} = 0.745 \cdot (m_3^*/m_3) \cdot (V_m/x)^{2/3} \cdot T \quad (2.3)$$

This means that the specific heat per ${}^3\text{He}$ atom in the mixture increases with decreasing concentration! (For one mole of mixture with ${}^3\text{He}$ concentration x we have: $C'_{3,d} = x \cdot C_{3,d} \propto x^{1/3} \cdot T$ at $T < T_F/10$). Thus:

$$C_{3,d}(6.6\%) \cong 106 \cdot T \quad [\text{J}/(\text{mole He}^3\text{K})^{-1}] \quad (2.4)$$

When the two phases are in thermodynamic equilibrium we must have for the chemical potentials

$$\mu_{3,d}(x_d, T) = \mu_{3,c}(x_c, T) \quad (2.5)$$

$$H_{3,d} - TS_{3,d} = H_{3,c} - TS_{3,c} \quad (2.6)$$

$$H_{3,d}(T) = H_{3,c}(T) + T(S_{3,d} - S_{3,c}) \quad (2.7)$$

$$= H_{3,c}(T) + T \int_0^T \left(\frac{C_{3,d}}{T} - \frac{C_{3,c}}{T} \right) dT' \quad (2.8)$$

$$= H_{3,c}(T) + T \int_0^T (106 - 22) dT' \quad (2.9)$$

$$= H_{3,c}(T) + 84T^2[j/\text{mole He}^3] \quad (2.10)$$

So, the heat of mixing is:

$$\dot{Q}(T) = \dot{n}_3[H_{3,d}(T) - H_{3,c}(T)] = 84\dot{n}_3T^2 \quad (2.11)$$

With a typical value of $\dot{n}_3 = 100 \mu\text{mol/s}$ and with $T = 10$ (30) mK we get $\dot{Q} \cong 1(10)\mu\text{W}$.

The cooling power seems small, but at $T < 0.35$ K it is much higher than the cooling power of evaporating ${}^3\text{He}$ cryogenic liquid:

$$\dot{Q} = \dot{n}L \propto p(T) \propto e^{-1/T} \quad (2.12)$$

2.8.4 Osmotic Pressure

As we shall see later, we have isotopic helium mixtures at varying concentrations and temperatures in our refrigerator. In such a situation an osmotic pressure π develops in ${}^3\text{He} - {}^4\text{He}$ mixtures. By considering them as ideal solutions (which is valid for liquid helium mixtures in the classical regime of $T > T_F$, e.g., for $T > 0.15\text{K}$, $x < 0.03$):

$$\pi V_{m,4} \cong xRT$$

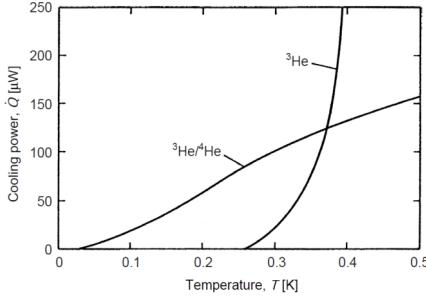


Figure 2.18: Cooling power of a ${}^3\text{He}$ evaporation cryostat and of a ${}^3\text{He} - {}^4\text{He}$ dilution refrigerator, assuming that the same pump with a helium gas circulation rate of 5l s^{-1} is used [7.9] (Diagram shows Cooling power, Q [μW] vs. Temperature, T [K]. Curves labeled ${}^3\text{He}/{}^4\text{He}$ and ${}^3\text{He}$)

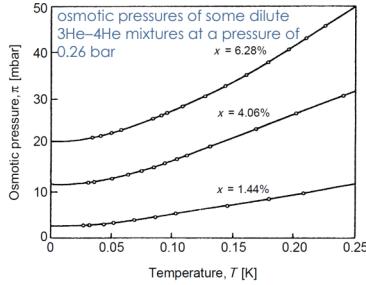


Figure 2.19: Diagram showing osmotic pressures of some dilute ${}^3\text{He} - {}^4\text{He}$ mixtures at a pressure of 40×0.26 bar. (Curves labeled $x = 6.28\%$, $x = 4.06\%$, $x = 1.44\%$)

where $V_{m,4}$ is the molar volume of ${}^4\text{He}$. In a dilution refrigerator a tube connects the mixture in the mixing chamber (where we have the phase separation) with the mixture in the still (where ${}^3\text{He}$ is evaporated), see below. The difference of the osmotic pressures due to the different concentrations and temperatures in the mixing chamber and the still is given by:

$$\pi_{mc} - \pi_{st} \cong (x_{mc}T_{mc} - x_{st}T_{st})R/V_{m,4}$$

If no ${}^3\text{He}$ is pumped from the still, then there is no difference in osmotic pressure between the still and the mixing chamber. Assuming the mixing chamber with 6.6% mixture to be at 10mK and the still to be at 0.7 K, we then have $x_{st} = x_{mc} \cdot (T_{mc}/T_{st}) \cong 0.1\%$. If we pump ${}^3\text{He}$ from the still, the concentration of ${}^3\text{He}$ will decrease there, and an osmotic pressure difference will develop that will drive ${}^3\text{He}$ from the mixing chamber into the still and therefore “sucks” ${}^3\text{He}$ from the concentrated into the dilute phase in the mixing chamber. The max-

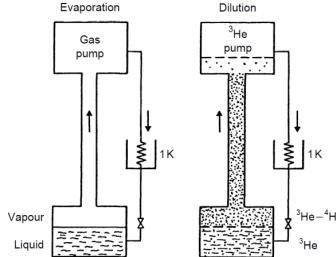


Figure 2.20: Gedanken experiment for comparison of a helium evaporation refrigerator and a ${}^3\text{He}$ – ${}^4\text{He}$ dilution refrigerator (Diagram shows components: Vapour, Liquid, Evaporation (\uparrow), Gas pump (\downarrow), Dilution (\uparrow), ${}^3\text{He}$ pump, 1K, ${}^3\text{He}$ – ${}^4\text{He}$, ${}^3\text{He}$)

imum osmotic pressure will be obtained when the concentration of ${}^3\text{He}$ in the still tends to zero: $x_{st} \cong 0 \Rightarrow \Delta\pi_{max} = x_{mc}RT_{mc}/V_{m,4} \cong 20$ mbar. This pressure corresponds to the hydrostatic pressure of about 1m of liquid helium. In other words, the osmotic pressure will be large enough to drive the ${}^3\text{He}$ from the mixing chamber into the still even if they are separated by a vertical distance of about 1 m.

2.8.5 Realization of a ${}^3\text{He}$ – ${}^4\text{He}$ Dilution Refrigerator

Actually, the circuit cannot operate in this way, because we would have the heavier mixture on top of the lighter liquid ${}^3\text{He}$. In reality we have to design the refrigerator slightly differently. The physics of these cooling processes is quite different:

- In evaporation we rely on the classical heat of evaporation for cooling.
- In dilution refrigeration we rely on the enthalpy of mixing of two quantum liquids. Here the cooling results from quantum mechanical effects: the different zero-point motions of the two helium isotopes and the different statistics which we have to apply to understand their properties at low temperatures.

Cooling occurs when ${}^3\text{He}$ atoms are transferred from the ${}^3\text{He}$ -rich to the ${}^3\text{He}$ -poor side. The ${}^3\text{He}$ is then driven up along the liquid mixture column by a pressure caused by the osmotic pressure difference. It eventually reaches the “ ${}^3\text{He}$ pump”, which in this scheme is our still. The still operates like a distillation chamber, evaporating almost pure ${}^3\text{He}$ (due to the fact that the vapour pressure of ${}^3\text{He}$ is much higher than that of ${}^4\text{He}$ → the mixture, even if dilute, has > 90% ${}^3\text{He}$ vapour above it → by pumping, we extract only pure ${}^3\text{He}$).

The ${}^3\text{He}$ gas coming from the exit of a pump at room temperature will first be precooled by a liquid ${}^4\text{He}$ bath at 4.2 K. It will then be condensed in a second ${}^4\text{He}$ bath at about 1.5 K (continuously operating ${}^4\text{He}$ refrigerator). This

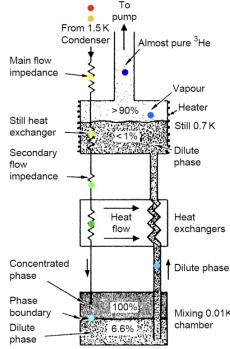


Figure 2.21: Schematic diagram of the dilution refrigerator circuit showing: To pump, From 1.5 K Condenser, Almost pure ^3He , Main flow impedance, Still heat, Vapour, Heater, Still 0.7 K, $> 90\%$, $< 1\%$, Secondary flow impedance, Heat exchangers, Dilute phase, Concentrated phase, Phase boundary, 100%, 6.6%, Mixing chamber 0.01K, Dilute phase

^4He evaporator is also used as a heat sink at which all tubes and leads going to colder parts should be thermally anchored. Below the ^4He refrigerator we need the so-called main flow impedance ($Z \approx 10^{12} \text{ cm}^{-3}$) to establish sufficient pressure ($30\text{--}200 \text{ mbar}$) for the incoming ^3He so that it will indeed condense at 1.5 K. The now liquid ^3He will flow through a heat exchanger which is in thermal contact with (or even inside) the still at a temperature of about 0.7 K. Below the still we have a secondary flow impedance ($Z \approx 10^{11} \text{ cm}^{-3}$) to prevent re-evaporation of ^3He . After leaving this secondary flow impedance the liquid ^3He will flow through one or several heat exchangers to precool it to a low enough temperature before it enters the upper, concentrated phase in the mixing chamber. Then it crosses the phase boundary, giving rise to cooling. A wider tube for the dilute phase in the refrigerator leaves the lower, dilute mixture phase of the mixing chamber, and then goes through the heat exchanger to precool the incoming ^3He . It enters the dilute liquid phase in the still, where we have a liquid ^3He concentration of less than 1%. The vapour above the dilute liquid phase in the still has a concentration at least of 90% ^3He due to the high vapour pressure of ^3He at the temperature of the still. If we then pump on the still and resupply the condensation line continuously with ^3He gas we have a closed ^3He circuit.

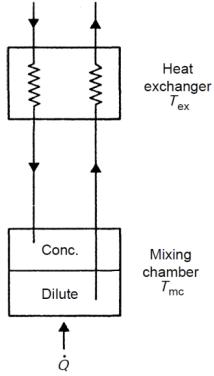


Figure 2.22: Diagram illustrating the Mixing Chamber setup.s

2.8.6 Properties of the Main Components of a ${}^3\text{He} \sim {}^4\text{He}$ Dilution Refrigerator:

Mixing Chamber

From an enthalpy balance applied to the mixing chamber and to the exit of the heat exchanger (see calculat. above):

$$\begin{aligned}\dot{Q} &= \dot{n}_3(H_d(T_{mc}) - H_c(T_{ex})) \\ &= \dot{n}_3(H_c(0) + 95T_{mc}^2 - H_c(0) - 11T_{ex}^2) \\ &= \dot{n}_3(95T_{mc}^2 - 11T_{ex}^2)\end{aligned}$$

The maximum cooling power is achieved for $T_{ex} = T_{mc}$ (since $T_{ex} \geq T_{mc}$):

$$\dot{Q} = 84\dot{n}_3T_{mc}^2 \quad (2.13)$$

giving the minimum reachable temperature of

$$T_{mc,min} = (\dot{Q}/84\dot{n}_3)^{1/2}$$

On the other hand, the minimum cooling power ($\dot{Q} = 0$) is for $T_{ex} \geq 3T_{mc}$. This gives a rather severe requirement for the efficiency of the heat exchangers: the liquid has to leave the last heat exchanger at a temperature which is at the most a factor of three higher than the temperature which we want to obtain in the mixing chamber. This is a very important result, demonstrating the importance of the efficiency of the heat exchangers in a dilution refrigerator. If, for example, we would like to have a cooling power of 1 μw at the mixing chamber, with circulating $\dot{n}_3 = 100 \mu\text{mol He}^3/\text{s}$, we obtain $T_{mc} = 12$ (15) mK for $T_{ex} = 18$ (30) mK.

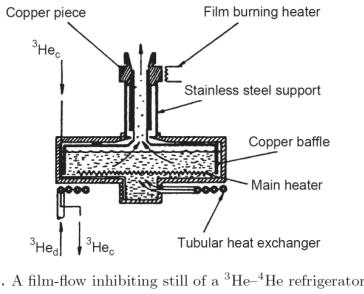


Figure 2.23: A film-flow inhibiting still of a ${}^3\text{He}$ – ${}^4\text{He}$ refrigerator.

Still

The main requirements on the design of the still is that the ratio of ${}^3\text{He}$ to ${}^4\text{He}$ in the vapour phase of the still should be as large as possible. A value of $0.6 - 0.7\text{K}$ is reasonable for the temperature of the still: in this range the ${}^3\text{He}$ vapour pressure is still large enough to keep the circulation rate reasonably large with a typical pump, and the concentration of ${}^4\text{He}$ in the vapour phase is only some percentage. Enthalpy balance for the still, where ${}^3\text{He}$ arrive with the temperature of the condenser and leave with the temperature of the still: cooling is produced by the heat of evaporation of ${}^3\text{He}$ at the temperature of the still.

$$\dot{n}_3 L_3(T_{st}, x_{d,st}) = \dot{n}_3 [H_3(T_{cond}) - H_3(T_{st})] + \dot{Q}_{st}$$

If we again assume a circulation rate of $100 \mu\text{mol s}^{-1}$, a condenser temperature of 1.3K and a still temperature of 0.7 K , we find that the required rate of heat supply \dot{Q}_{st} to the still is several milliwatts; more generally, a good rule is $\dot{Q}_{st}[\text{W}] \cong 40\dot{n}_3[\text{mol/s}]$. This heat will be partly supplied by a conventional heater.

In addition, we can thermally anchor to the still a radiation shield surrounding all the lower, colder parts of the dilution refrigerator to reduce the heat of radiation. To keep the amount of ${}^4\text{He}$ pumped from the still reasonably low (to avoid refrigerator deterioration), one has to suppress, in particular, the superfluid ${}^4\text{He}$ film flow → introduction of a small orifice ($\leq 1\text{mm}$ diameter) with sharp edges in the pumping tube or, for more powerful refrigerators, a so-called ${}^4\text{He}$ film burner.

Heat exchangers

Heat exchangers are the most important and most critical part of a dilution refrigerator; they determine its minimum temperature, for example. The purpose of the heat exchangers is to bring the temperature of the incoming ${}^3\text{He}$ as close to the temperature of the mixing chamber as possible by using the cold mixture leaving the mixing chamber to precool the incoming warmer ${}^3\text{He}$. For

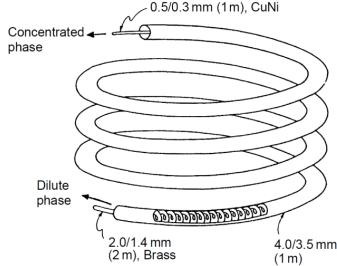


Figure 2.24: Diagram of a continuous counterflow heat exchanger.

this purpose, in general, one needs several heat exchangers between the still and the mixing chamber to achieve the best performance of the refrigerator. An enthalpy balance for the heat exchangers is much more involved than before (temperature gradients in the viscous fluid streams as well as in the heat exchanger body, both in the direction of the flow and perpendicular to it). The requirements (sometimes contrasting) on the heat exchangers can be quite severe for a powerful dilution refrigerator: small volumes, so that the liquids attain temperature equilibrium quickly, small impedances, so that the viscous heating due to flow is small, and there should be a small thermal resistance between the streams to obtain good temperature equilibrium between them → small thermal boundary resistance $\sim T^{-3} A^{-1}$. Usually one requires many square meters of surface area to overcome the Kapitza boundary resistance in the heat exchangers for $T < 0.1\text{K}$. Because the thermal boundary resistance increases with decreasing temperature, it is particularly serious in the mixing chamber, where one needs surface areas between several 10m^2 and several 100m^2 .

Simple Heat Exchangers For a rather simple dilution refrigerator with a minimum temperature of about 30mK , one continuous counterflow tube-in-tube heat exchanger consisting of two concentric capillaries (with diameters of order 0.5 and 1 mm, and with 0.1 or 0.2mm wall thickness) several meters long is sufficient. The concentric capillaries should be of low thermal conductivity (CuNi, stainless steel or brass) and are usually coiled into a spiral so they do not take up too much space. Heat is transferred across the body of the wall, and conduction along the capillary and along the liquid streams should be negligible. The dilute phase moves in the space between the tubes and the concentrated liquid moves in the inner tube.

Step Exchangers For a more powerful dilution refrigerator with a larger cooling power and lower minimum temperature we need more than one heat exchanger. At lower temperatures the earlier-discussed continuous heat exchanger can no longer be used because it does not provide enough surface area to defeat

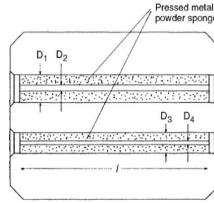


Figure 2.25: Diagram of a sintered metal powder step exchanger.

the increasing thermal boundary resistance. This can only be provided by step exchangers, which in their simplest design can be machined from Cu blocks into which two holes have been drilled. To reduce the thermal boundary resistance, each channel has to be filled with pressed and/or sintered metal powder (into which flow channels for the counterflowing liquids are drilled) to provide the required surface area for adequate heat exchange.

A dilution refrigerator of with $T_{min} = 3 \text{ mK}$ and a maximum value of $\dot{n}_3 = 700 \mu\text{mol s}^{-1}$ used six such step exchangers with the parameters given in the table (the filling factor f is for sinter made from powder of grain size d). Further

Table 2.4: Parameters of Step Exchangers

T_N	[mK]	83	48	32	19	13	10
l	[mm]	62	94	94	94	94	88
D_1	[mm]	8	13	13	18	18	22
D_2	[mm]	2.7	2.7	3.2	4.8	6.5	8.0
D_3	[mm]	6	10	10	14	14	17
D_4	[mm]	2.0	2.0	2.4	3.2	4.8	6.5
Sintered Metal powder	Cu	Cu	Cu	0.2 Cu	0.4 Cu	0.4 Cu	
d	[μm]	100	50	30	44 Cu	44 Cu	44 Cu
					2.5 Ag	0.07 Ag	0.07 Ag
f		0.55	0.65	0.65	0.75	0.70	0.70
A_c	[m^2]	0.05	0.27	0.45	6.5	100	135
A_d	[m^2]	0.09	0.46	0.76	10.6	164	230
V_c	[cm^3]	1.6	7.1	7.0	13.7	12.8	17.1
V_d	[cm^3]	2.8	11.9	11.7	22.2	20.8	29.0

optimization was obtained by rectangular heat exchangers with welded Cu-Ni foil filled with submicrometer silver powder sintered to a silver plated Cu-Ni foil, that are also taken advantage of in powerful commercial dilution refrigerators.

Minimum Temperature Calculation We had (see above) $T_{mc,min} = (\dot{Q}/84\dot{n}_3)^{1/2}$. But, taking the temperature step due to the Kapitza resistance R_K between the helium liquid and the solid with a surface area A into account,

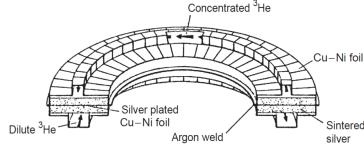


Figure 2.26: Diagram of optimized rectangular heat exchanger.

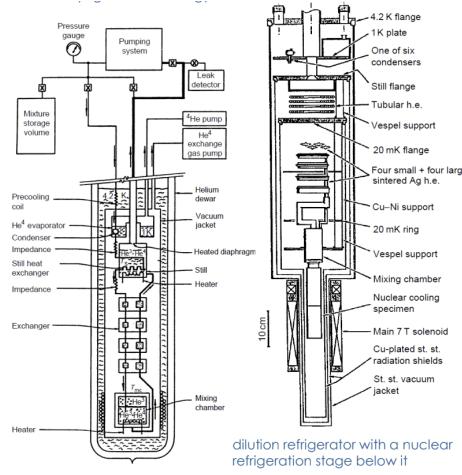


Figure 2.27: Schematic diagrams of successful “home-made” dilution refrigerators

we have to recalculate the minimum temperature in the mc. It turns out that:

$$T_{mc}^2 = 0.011 \frac{\dot{Q}}{\dot{n}_3} + 5.2 R_K \frac{\dot{n}_3}{A} \quad \text{if } R_K \propto T^{-3}$$

$$T_{mc}^2 = 0.011 \frac{\dot{Q}}{\dot{n}_3} + (27 R_K \dot{n}_3 T^2)^2 \quad \text{if } R_K \propto T^{-2}$$

$$T_{mc,min} = 4.5(\dot{Q}/A)^{1/3} \quad [\text{mK}]$$

2.8.7 Examples of ^3He - ^4He Dilution Refrigerators

The most powerful have been built by Frossati et al. in Leiden: it has a maximum circulation rate of $\dot{n}_3 = 10 \text{ mmol s}^{-1}$, a minimum temperature of 1.9 mK at $\dot{n}_3 = 0.85 \text{ mmol s}^{-1}$ and a cooling power of 20 μW at 10 mK. It contains the enormous quantity of 1.6 kg Ag, and $A = 2300 \text{ m}^2$ in total in the heat exchangers!

For use in high (pulsed) magnetic fields, refrigerators with the dilution unit made from plastics often is required to reduce eddy current heating. The lower

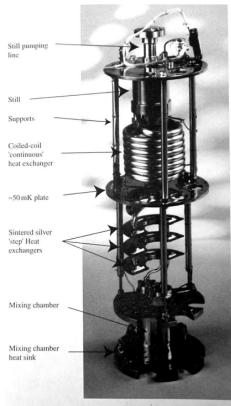


Figure 2.28: A modern dilution refrigerator. Courtesy of Oxford Instruments.

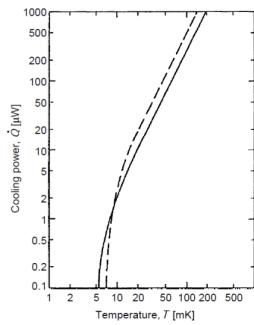


Figure 2.29: Cooling power as a function of temperature of a commercial $^3\text{He} - ^4\text{He}$ dilution refrigerator, for two still heating powers: 8 mW (full line) and 22 mW (dashed line).

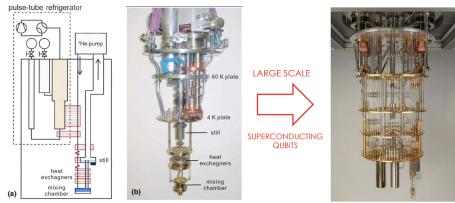


Figure 2.30: Schematic diagram (a) and photograph (b) of a dry dilution refrigerator precooled by a two-stage pulse tube cryocooler. On the right Example from S. Krinner et al. Engineering cryogenic setups for 100-qubit scale superconducting circuit systems EPJ Quantum Technol. 6, 2 (2019; Dilution Refrigerator Bluefors XLD400 with Pulse Tube Coolers (2x Cryomech PT 420))

end of these refrigerators needs to have a small diameter to fit into the bore of the magnet (20 – 40mm). They are made mainly from easily machinable PVC (polyvinyl chloride) parts glued together with Stycast 1266.

2.8.8 Developments

Developments: try to avoid (or to minimize) the expensive and bulky room-temperature gas handling and mechanical pumping system and to avoid the external supply of liquid nitrogen and helium. There are refrigerators in which the circulation of the ^3He gas is achieved by internally regenerating adsorption pumps, using mostly activated charcoal → reduced vibrations → commercially available continuously operating $^3\text{He} \sim ^4\text{He}$ dilution refrigerators. Alternative: use of commercially available two-stage closed-cycle refrigerators to precool the dilution unit and its radiation shields instead of using a cryostat filled with liquid helium. Pulse-tube refrigerator for precooling (no moving cold parts) → smaller level of vibration. The first stage reached 45K to cool the radiation shield and the second stage could be operated at a temperature as low as about 3 K. At flow rates of $0.3 - 1 \text{ mmol s}^{-1}$, the achieved minimum temperature of the dilution refrigerator was 9 mK. Recently they have become commercially available → very promising for quantum computing applications.

2.8.9 $^3\text{He} \sim ^4\text{He}$ Dilution Refrigerator: Addendum: Example

Managing the heat budget of large-scale quantum computers based on superconducting circuits

Example from S. Krinner et al. Engineering cryogenic setups for 100-qubit scale superconducting circuit systems EPJ Quantum Technol. 6, 2 (2019).

When scaling dilute refrigeration from the few qubit level to large scale quantum processors, an increasing number of microwave and DC cables need to be integrated. They connect the classical control electronics at room temperature

(RT) to the quantum processor at the lowest temperature stage of the dilution refrigerator, creating a substantial heat load on the dilution refrigerator due to heat conduction. Besides this passive load, active load due to the dissipation of control signals in cables and attenuators plays a major role. Engineered dissipation is necessary to thermalize the incoming radiation fields and to reduce the number of thermal photons incident on the sample. GOAL: suitable for the operation from 50 to 150 qubits at a temperature of 14 mK.

Typical cabling for experiments with superconducting circuits

When designing the control lines and output lines connecting to a superconducting quantum processor the goal is to provide sufficiently strong coupling rates to the quantum processor, while minimizing decoherence due to coupling of the quantum processor via these lines to its environment. Thermal noise, present due to the connection of the quantum processor to electronics at room temperature, not only leads to qubit dephasing, but can also lead to creation of quasi-particles and thus to dissipation and reduced energy relaxation times. Hence, thorough thermalization of cables, attenuators, and microwave components at the various temperature stages of the dilution refrigerator is not only important for reducing the heat load on the dilution refrigerator, but also for protecting the quantum processor from thermal radiation. In addition to thermal anchoring, filters with stop-bands outside the frequency range of qubits and readout resonators as well as infrared blocking filters further suppress thermal radiation. We distinguish between direct-current (DC) and radio-frequency (RF) cabling:

- DC lines are made from twisted pairs of wires, that are low-pass filtered, and thermalized at each temperature stage.
- RF lines are realized as semi-rigid microwave cables and contain various microwave components such as attenuators, filters and amplifiers.

One typically distinguishes between drive lines, flux lines, and output lines:

- Drive lines are used for controlling the quantum states of qubits with a microwave tone, and for probing the frequency shift of readout resonators. The bandwidth of the drive lines is required to be large enough to cover the typical frequency ranges of qubits (4~6 GHz) and of readout resonators (4~8 GHz).
- Flux lines are used for controlling (through its dynamical flux tunability) the transition frequency of a qubit. Low-pass filters in the flux lines limit the bandwidth to about 1 GHz eliminating thermal noise at qubit frequencies.
- Output lines contain a series of cryogenic and room temperature amplifiers for the detection of readout signals.

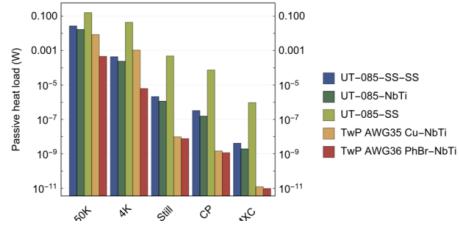


Figure 2.31: Heat flow for indicated cable types. Heat flow to the five temperature stages for indicated cable types and given cable lengths.

Sources of heat loads

We consider three dominant contributions to the heat load on the dilution refrigerator:

1. Passive load is due to heat flow from higher temperature stages to lower temperature stages. Here, we consider only heat conducted through installed cables.
2. Second, active load arises due to the dissipation (Joule heating) of applied microwave signals in attenuators and in the microwave cables themselves.
3. Third, a radiative load arises due to blackbody radiation from stages and shields of higher temperature impinging on stages and shields of lower temperature.

Passive load To minimize passive heat load we use cable materials with low thermal conductivity.

Table 2.5: Dilution refrigerator specifications. Temperatures and available cooling powers on the indicated stages of a Bluefors XLD400 DR. Coaxial cable lengths towards the respective stages are listed as well

Stage name	Temperature (K)	Cooling power (W)	Cable length (mm)
50 K	35	30 (at 45 K)	200
4 K	2.85	1.5 (at 4.2 K)	290
Still	882×10^{-3}	40×10^{-3} (at 1.2 K)	250
CP	82×10^{-3}	200×10^{-6} (at 140 mK)	170
MXC	6×10^{-3}	19×10^{-6} (at 20 mK)	140

Twisted pairs (TwP) consist of either Cu ($RRR = 100$) or PhBr from RT to the 4 K stage, and of NbTi from the 4 K to the MXC stage.

Active load The active load in the dilution refrigerator depends on the level of attenuation of the RF lines and the installed attenuators, and on the signal

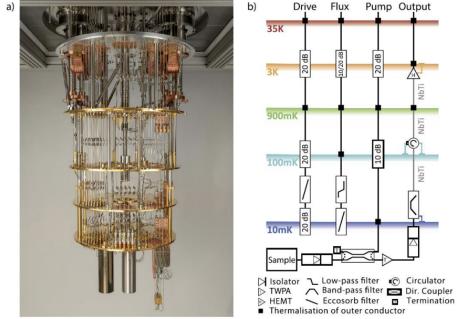


Figure 2.32: Cabled dilution refrigerator (DR). (a) Bluefors XLD DR with 25 drive lines, 25 flux lines, 4 read-out, 6 read-in, and 5 pump lines installed (see end of Sect. 3.1 for details). (b) Schematic of the cabling configuration within the DR

levels required at the chip. The need of attenuation - Although the signals required to drive qubits or to read them out are small (e.g. a peak power of -66 dBm for a 20 ns long π -pulse), the power applied at the input of the dilution refrigerator is orders of magnitude larger. This is because a total attenuation of at least ~ 60 dB is required to reduce blackbody radiation present in cables at room temperature to a level corresponding to a thermal photon occupation number of a few 10^{-3} at the sample. The blackbody radiation in effectively one-dimensional cables is known as Johnson-Nyquist noise in electronics. In a coaxial cable connecting room temperature electronics to base temperature circuits, thermal photons propagate down the line towards the lower temperature stages. To reduce the spectral density of thermal radiation a series of attenuators is installed in the microwave line. An attenuator with an attenuation of $A = 20$ dB = 100 effectively acts as a beamsplitter which transmits 1% of the incident signal and adds 99% of blackbody radiation with the effective temperature T_{att} at which the attenuator is thermalized. We note that 99% of the incident signal is dissipated in the attenuator. To prevent the attenuator from heating up and to keep the effective temperature at which it re-emits thermal radiation low it is efficiently thermalized. This involves installing a cascade of attenuators in the microwave line thermalized at subsequently lower temperature stages.

Required signal levels Powers required at the chip for driving a π -pulse on a qubit and setting a flux bias on a qubit, two important operations on superconducting qubits. To drive a π -pulse on a qubit, we apply an RF pulse at the qubit frequency through a CPW transmission line weakly capacitively coupled to the qubit. -66 dBm is needed. Readout signals used to drive readout resonators to infer the state of qubits are typically an order of magnitude smaller. For the operation of the TWPA (travelling wave Josephson parametric ampli-

fiers) a microwave pump signal with a typical frequency in the range $6\text{--}8$ GHz and a power level at the input of the TWPA of about -55 dBm is required. Concerning dissipation in flux lines, we primarily consider DC biasing currents, which are constantly applied to set the qubit frequency. This corresponds to a current interval of $[-1, 1]$ mA, when using a reasonable mutual inductance of $M = \partial\Phi/\partial l = 0.5\Phi_0/\text{mA}$ between the flux line and the SQUID loop.

Radiative load Each temperature stage (except for the cold plate) is fitted with a dedicated heat shield to protect the next lower temperature stage from radiative load. The heat shields at the 50 K and 4 K stages are made of Aluminium, and of Cu on Still and MXC. They are characterized by their emissivity $\epsilon = 0.06$, as quoted by the manufacturer. Significant contribution only for the 50 K stage, amounting to ~ 50 W. This corresponds to about half of the nominal cooling power of the two pulse tube coolers (2x Cryomech PT 420). This load however has already been taken into account in the available cooling power as specified by the manufacturer ($40\text{--}50$ W). Radiative loads at the lower temperature stages are insignificant.

Cabling and Thermalization

In slide 29 components for thermalization of RF cables and attenuators can be seen, as well as pre-assembly of cable trees. (a) CAD drawing of the Cu plate used to mount and thermalize attenuators. To clamp the attenuators to the copper plate, an adapter as shown in (d) is used. The Cu plates are then fixed on the corresponding plate of the DR using screws. (b) Photograph of a mounting plate with cabling installed in the DR. (c) Photograph of the same plate, here used to clamp RF cables using the radiation tight Cu adapter pieces shown in (d). (d) Two-part Cu adapter which is pressed around an RF cable using the wedge shaped piece in the middle in combination with a screw (see also (a))). (e) Preassembly tool with mounted Cu plates and vacuum flange. (f) Vacuum flange with vacuum tight SMA adapters installed. (g) Mounting the preassembled cable tree into the DR using a custom mounting tool.

Integration of four high-bandwidth output lines. (a) Components mounted on the MXC plate: readout signals from the quantum processor travel through a first 4-channel, magnetically shielded isolator array (right), followed by directional couplers, TWPs and a second, identical isolator array (left). (b) Four HEMT amplifiers mounted on the 4 K plate. (c) Four bandpass filters, mounted and thermalized at the MXC plate, followed by four circulators mounted on the cold plate.

Total Heat Load Budget

Based on the measurements of passive and active loads we estimate the total heat load acting on the different stages of a dilution refrigerator, when operating a 50 qubit processor with individual drive and flux control and with a multiplexed readout architecture allowing for simultaneous readout of sets of 6-7 qubits.

Table 2.6: Passive heat load per line. Passive heat load (HL) of cable types installed in the DR, as inferred from observed temperature increases after the installation of individual cable trees into the DR. The upper and middle sections of the table refer to 0.085" diameter stainless steel coaxial cable (UT-085-SS-SS) with attenuator configurations used in drive lines and flux lines, respectively. Indicated errors include statistical errors between different cooldowns and reflect run-to-run temperature variations on the stages of the DR. The intervals of estimated heat loads correspond to calculations of lower and upper bounds (see text for details)

	50 K	4 K	Still	CP	MXC
Drive line UT-085-SS-SS					
Measured HL	45(34) mW	1.0(5) mW	4(3) μ W	0.4(2) μ W	0.013(6) μ W
Estimated HL	24 – 27 mW	0.4 – 1.9 mW	1.6 – 2.1 μ W	0.33 – 0.60 μ W	0.004 μ W
Flux line UT-085-SS-SS					
Measured HL	56(39) mW	1.2(8) mW	2(1) μ W	0.3(1) μ W	0.029(5) μ W
Estimated HL	24 – 27 mW	0.4 – 1.9 mW	1.6 – 2.1 μ W	0.24 – 0.33 μ W	0.005 – 0.282 μ W
Output line UT-085-NbTi					
Measured HL				0.3(3) μ W	0.020(16) μ W
Estimated HL				0.18 – 0.31 μ W	0.002 – 0.322 μ W

Such a quantum processor requires a total of 124 RF lines (50 drive lines, 50 flux lines, 8 output lines, 8 readout resonator drive lines, 8 TWPA pump lines), corresponding to the operation of the system at its full capacity.

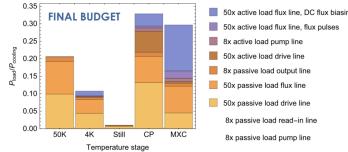


Figure 2.33: Total heat load budget for the operation of a 50 qubit processor. Predicted passive and active heat loads for the operation of a 50 qubit processor with individual drive and flux control. The heat loads are normalized to the measured cooling powers as presented in Table 2.5

The largest relative loads of about 30% occur at the CP and MXC stages, with about an equal share between passive and active loads. The load on MXC corresponds to an operation temperature of 14 mK.

Concluding Note

Google unveiled the world's largest quantum computer processor to date. Dubbed Bristlecone, it's a 72-qubit gate-based superconducting system.

2.9 ^3He solidification - Demagnetization - Others

2.9.1 Outline

- ^3He solidification refrigeration (Pomeranchuk cooling)
- Refrigeration by Adiabatic Demagnetization (electronic, nuclear)
- Other techniques for refrigeration below 1 K (Tunnel junction cooling, Laser cooling, Evaporative cooling)

2.9.2 TODO: Refrigeration by Solidification of Liquid ^3He : Pomeranchuk Cooling

Today this technique is no longer as important because above a few millikelvin this "one-shot" method offers essentially no advantage over continuously operating dilution refrigerators, and at lower temperatures nuclear demagnetization is much more powerful. However, this cooling method offers interesting insight into some important properties of matter, in particular liquid and solid $^3\text{He} \rightarrow$ some hints.

Solidification of matter at the rate \dot{n} usually results in the production of heat according to

$$\dot{Q} = \dot{n}T(S_{sol} - S_{liq}) < 0$$

because the liquid's entropy is usually larger than the solid's entropy, as the liquid state is a state of lower order than the solid state.

In 1950, I. Pomeranchuk predicted that for ^3He on the melting curve below about 0.3K the entropy of the liquid phase is smaller than the entropy of the solid phase (which means the liquid phase is more ordered, due to quantistic reasons connected to the nuclear spin), and that therefore adiabatic solidification of He along the melting line should result in cooling.

image Entropies (divided by the gas constant R) of solid and liquid ^3He along the melting curve. The full disorder nuclear spin entropy of solid He, $S_s/R = \ln(2I + 1) = \ln 2$ is marked. The entropy curves cross at the minimum of the melting curve at 315 mK, 29.31 bar. The anomaly in S_{liq} at 2.44 mK is caused by the superfluid transition of liquid He. The kink in S_{sol} at 0.90 mK is a result of the nuclear antiferromagnetic ordering of solid ^3He .

image Phase diagram for ^3He and melting curve pressure vs. temperature. This unusual negative slope of the melting curve is connected to the fact that here the liquid phase has a smaller entropy than the solid phase. If we move along the melting curve below the minimum, by increasing the pressure and so solidifying more and more ^3He , the temperature has to decrease.

Cooling power (latent heat of solidification):

$$\dot{Q} = \dot{n}_{sol}T(S_{sol} - S_{liq}) = \dot{n}_{sol}T(R \ln 2 - 36T) \approx T$$

when \dot{n}_{sol} moles are solidified per unit time.

image Comparison of specific cooling powers of Pomeranchuk refrigeration and ^3He - ^4He dilution refrigeration.

image Plastic Pomeranchuk cell. Stretched Kapton tube as a pressure transmitter.

Disadvantage: its discontinuous nature. Pomeranchuk cooling is basically a "one-shot" method.

2.9.3 TODO: Refrigeration by Adiabatic Demagnetization of a Paramagnetic Salt

In 1926, a completely new refrigeration technology was proposed: lower temperatures could be reached by using the magnetic disorder entropy of electronic magnetic moments in paramagnetic salts, a method later called adiabatic demagnetization of paramagnetic salts.

Consider paramagnetic ions with an electronic magnetic moment μ_{el} in a solid. Assume the energy \mathcal{E}_m of interaction to be small compared to the thermal energy $k_B T$. This means that we are considering free paramagnetic ions with magnetic moment μ and total angular momentum J with entropy contribution $S = R \ln(2J + 1)$, if they are completely disordered in their $2J + 1$ possible orientations with respect to a magnetic field.

If the temperature is decreased, eventually the interactions between the magnetic moments will become comparable to the thermal energy \rightarrow spontaneous magnetic order, e.g., ferromagnetic or antiferromagnetic orientation of the moments \rightarrow the entropy will decrease and approach zero (as required by the third law of thermodynamics). An externally applied magnetic field will interact with the magnetic moments, at least partially orienting them along its axis to create a magnetized state of higher order. Therefore, in the presence of a field the entropy will start to decrease at a higher temperature.

image Electronic magnetic entropy vs. temperature at different magnetic fields.

Phases (Electronic Demagnetization)

- **Precooling:** Paramagnetic salt in contact with a precooling bath to precool it to a starting temperature T_i .
- **Isothermal Magnetization:** A magnetic field B_i is applied to perform an isothermal magnetization at T_i from $B = 0$ to $B = B_i$. During this process the heat of magnetization is absorbed by the precooling bath.
- **Adiabatic Demagnetization:** Thermal isolation of the paramagnetic salt from the surrounding bath which then allows the adiabatic demagnetization to be carried out by reducing the external field from its starting value B_i to a final field B_f , which may be in the millitesla region \rightarrow temperature decreases.

- **Warming Up:** The cooling agent, the paramagnetic salt, will warm up along the entropy curve at $B_f = \text{const.}$ due to an external heat leak until its cooling power has been used up to refrigerate the sample.

Magnetic refrigeration is a "one-shot" technique, where the demagnetization ends at a low field and then sample and refrigerant warm up.

The heat of isothermal magnetization released when the applied field is increased from zero to B_i and which has to be absorbed by the precooling bath at constant temperature T_i is given by

$$Q(T_i) = nT_i[S(0, T_i) - S(B_i, T_i)]$$

(The heat of magnetization is given by the rectangle $ABDS_\infty$ in the figure).

Adiabatic Demagnetization. For free magnetic moments the entropy is a function of just the ratio of magnetic energy to thermal energy \rightarrow for the adiabatic process $S(B_i/T_i) = S(B_f/T_f)$, which results in

$$\frac{T_f}{B_f} = \frac{T_i}{B_i} \quad (*) \quad (2.14)$$

Of course, we cannot reach $T_f \rightarrow 0$ by letting $B_f \rightarrow 0$ because eventually the condition $k_B T \gg \mathcal{E}_m$ is violated and the internal interactions will align the moments and the entropy vanishes.

Warming-up due to External Heating. The cooling power of the salt after demagnetization to B_f , or the heat it can absorb as it is warming up, is given by

$$Q(B_f) = n \int_{T_f}^{\infty} T (\partial S / \partial T)_{B_f} dT$$

(This hatched area is indicated in the figure).

These results demonstrate that one has to find a compromise between a low final temperature T_f [which would require a low final field B_f ; see (*)], and a large cooling power (which requires a large final field figure).

More accurate calculations \rightarrow

$$T_f = \frac{T_i}{B_i} \sqrt{B_f^2 + b^2}$$

where b is an internal field resulting from the neighboring moments in the paramagnet.

Paramagnetic Salts and Magnetic Refrigerators

- The behavior of a magnetic refrigerator is mainly determined by the experimental starting conditions (B_i, T_i), the heat leaks, and the properties of the paramagnetic salt. Typical starting conditions for paramagnetic refrigeration, $B_i = 0.1 - 1T$, $T_i = 0.1 - 1K$.

- The paramagnetic refrigerant, the salt, should have a low magnetic ordering temperature T_c and a large magnetic specific heat in order to achieve a large cooling power (which means a large angular momentum J).
- The ordering temperature T_c is determined by the interactions between the magnetic moments, which create the internal field b .
- Suitable paramagnetic salts must contain ions with only partly filled electronic shells, i.e., either 3d transition elements or 4f rare earth elements.

image Reduced entropy S/R for different paramagnetic salts.

Examples of Paramagnetic Salts

- "High"-temperature salts:
 - **MAS** (Manganese ammonium sulphate: $Mn^{2+}SO_4 \cdot (NH_4)_2SO_4 \cdot 6H_2O$): $T_c \simeq 0.17K$
 - **FAA** (Ferric ammonium alum: $Fe_2^{3+}(SO_4)_3 \cdot (NH_4)_2SO_4 \cdot 24H_2O$): $T_c \simeq 0.03 K$
- "Low"-temperature salts:
 - **CPA** (Chromium potassium alum: $Cr_2^{3+}(SO_4)_3 \cdot K_2SO_4 \cdot 24H_2O$): $T_c \simeq 0.01 K$
 - **CMN** (Cerium magnesium nitrate: $2Ce^{3+}(NO_3)_3 \cdot 3Mg(NO_3)_2 \cdot 24H_2O$): $T_c \simeq 0.002K$ (most used)

Advantages of Paramagnetic Refrigeration

- The required starting conditions can be fairly easily achieved with simple dilution refrigerators and superconducting magnets.
- The refrigerant can also be used as a thermometer by applying the Curie law $\chi = \lambda/T$ to determine the temperature T by measurements of the susceptibility χ of the salt.

image Schematic drawing of a magnetic refrigerator.

Drawbacks of Paramagnetic Refrigeration

- "One-shot" nature.
- The thermal conductivity of dielectric salts is rather poor (a typical value is $10^{-4}W\ K^{-1}cm^{-1}$ at 0.1 K).

2.9.4 TODO:Nuclear Adiabatic Demagnetization

Nuclear magnetic moments are about 1,000 times smaller than electronic magnetic moments → they can be used for magnetic refrigeration in a similar way as the electronic magnetic moments → avoid the main disadvantages of electronic paramagnetic refrigeration, i.e. low thermal conductivities and "high" magnetic ordering temperatures → refrigeration into the micro-kelvin temperature range and below.

The minimum temperature for magnetic refrigeration is given by spontaneous magnetic ordering. The interaction energy between magnetic moments in the simplest case (dipole-dipole interaction only) is given by $\mathcal{E}_d = \mu b$, where b is the internal dipole field created by neighbors ($\sim \mu/r^3$) → the ordering temperature is $T_c \propto \mu^2/r^3$. The nuclear magnetic moments are of the order of the nuclear magneton $\mu_n = 5.05 \times 10^{-27} \text{ JT}^{-1}$, instead of the Bohr magneton $\mu_B = 9.27 \times 10^{-24} \text{ JT}^{-1}$ for the electronic magnetic moments → the ordering temperature due to nuclear dipole-dipole interaction is much lower, typically of the order of $0.1 \mu\text{K}$ or less.

Further advantages (over electronic demagnetization)

- For nuclear magnetic refrigeration we can use metals and can take advantage of their high thermal conductivity.
- Higher density of moments → large nuclear entropy density.

Advantages are counteracted by experimental problems resulting from the small size of these moments: to achieve a reduction of the nuclear magnetic entropy of, at least, a few percent we need rather demanding starting conditions for the magnetic field and for the temperature.

- E.g., for Cu (most used): with $B_i = 8T$ and $T_i = 10 \text{ mK} \Rightarrow \Delta S$ only $\simeq 5\% R \ln 4$.

image Molar nuclear spin entropy vs. B/T .

The cooling power is given by $Q = n \int T dS$, and because we now also want to work at substantially lower temperature, Q is typically a factor of 1,000 smaller than for electronic magnetic refrigeration.

Transfer of the spin temperature. The nuclear spin system may reach a very low temperature, but how this is transferred to the rest of the system, to the electrons and to lattice vibrations, which determine the "temperature" of our refrigerant? (unless we are only interested in the cold nuclear spin system itself). Metals are suited because of their short spin-lattice relaxation time (see below). Very slow processes (hours, days).

As for electronic demagnetization:

$$T_f = \frac{T_i}{B_i} \sqrt{B_f^2 + b^2}$$

→ e.g., for Cu: $b = 0.36 \text{ mT}$.

Differences in Experimental Procedure (Nuclear vs. Electronic)

There are two ways to precool and polarize the nuclear moments.

- **Superconducting solenoid operating in the persistent mode:**

- No current between the room temperature power supply and the low-temperature part during the long time of precooling, thus avoiding ohmic losses.
- Usually produces a field which is more stable and ripple free than the field of a solenoid connected to a power supply.

The heavy lines indicate the superconducting wire of the magnet and its persistent shunt. A heater is connected which can drive this piece of superconductor into its resistive state, which is necessary for loading and unloading the current of the main solenoid.

image Schematic drawing of a nuclear demagnetization refrigerator setup.

Interaction Between Conduction Electrons and Nuclei

- **Nuclear Cooling vs. Nuclear Refrigeration:** Since $\tau_1 \gg \tau_2$:

- It is possible to refrigerate just the nuclear spin system to low microkelvin temperatures and leave conduction electrons and phonons at higher temperatures, for studies of nuclear magnetic interactions (*nuclear cooling*).
- If the refrigeration is performed in such a way that the nuclei pull conduction electrons and phonons to low temperatures as well, the term used is *nuclear refrigeration*.
- **Nucleus–Electron Coupling:** Electromagnetic interaction between the nuclei and s-electrons, undergoing a mutual spin flip via hyperfine interaction.
 - **Spin–Lattice Relaxation Time (τ_1):** Time needed for thermal equilibrium at $T_e = \text{const}$: $\tau_1 = \kappa/T_e$ (Korringa law), where κ is a material constant.
- **Electron–Phonon Coupling:** The phonon temperature will always follow the electronic temperature within a very short relaxation time. Henceforth we can assume that lattice vibrations and conduction electrons are at the same temperature.
- **Heat Transfer:** The way heat or cooling is transferred is: nuclei → conduction electrons → phonons.
- **Spin–Spin Relaxation Time (τ_2):** Thermal equilibrium among the nuclear magnetic moments at the temperature T_n is established within τ_2 , which is rather short for metals, with typical values of less than 1ms.

image Nuclear spin temperature T_n and electronic temperature T_e as a function of the magnetic field B_f for the demagnetization stage.

The nuclear spin temperature $T_{n,f}$ is about proportional to the demagnetization field. Also, the electronic one T_e is proportional to field as long as it is not too small: for small final fields T_e increases, leading to an increasing difference between nuclear and electronic temperatures due to the heat flow between the two systems.

- **Example:** $T_e = 2T_n$ for 10 moles of Cu at $\dot{Q} = 1 \text{ nW}$ and $B_{f,\text{opt}} = 6 \text{ mT}$.

Influence of an External Heat Load and the Optimum Final Magnetic Field. There is an optimum final demagnetization field $B_{f,\text{opt}}$ if we are interested in achieving the minimum electronic temperature:

$$T_{e,\min} \approx \frac{T_i}{\sqrt{B_i}} \sqrt[4]{\frac{\lambda_n \kappa}{4\dot{Q} B_i}} \quad \text{where } \lambda_n \text{ is the Curie nuclear constant.}$$

Heat Leaks

Because the cooling power of a refrigerator is smaller the lower its temperature regime of operation (nW), more than for all other refrigeration techniques, the success of a nuclear refrigerator depends on the reduction of heat leaks.

- The latest best nuclear refrigerators reach values of total heat leaks of the order of 10^{-10} W .
- Even cosmic rays or natural radioactivity cannot be neglected (e.g. 20 pW from background γ -radiation, 120 pW from cosmic rays).
- **External heat leaks:** Conduction via residual gas atoms, along mechanical supports, via the heat switch, along electrical leads or fill capillaries for helium experiments. Because a nuclear refrigerator contains highly conducting metals in changing magnetic fields, we have to consider eddy current heating effects → slow processes.
- **Internal leaks:** E.g., the exothermic ortho–para conversion of H₂ (typical concentrations of $10^{-1} \dots 10^{-2}$).

image Schematic of the low-temperature part of the nuclear refrigerator at the University of Bayreuth with the two superconducting magnets and their field profiles.

Examples of Nuclear Demagnetization Refrigerators

- **University of Bayreuth:** The first nuclear stage contains 17.5 kg or 275 mol Cu, of which 6.6 kg or 104 mol Cu are effectively in a field of 8 T. The second nuclear stage contains 0.13 kg or 2 mol Cu in a field of 9T. Superconducting Al heat switches and Pt NMR thermometers are used in the low-field experimental spaces.
- **Record:** 2 μK record.

2.9.5 TODO:Other techniques for refrigeration below 1 K

Tunnel Junction Cooling

- The basic principle of direct electronic cooling: an energy filter (gray wall) allows only high energy electrons (red circles) to be removed from the electron system. This ejection leads to sharpening of the electron distribution, i.e. cooling.
- When a proper bias voltage is applied over a NIS (Normal metal-Insulator-Superconductor) junction, the most energetic electrons in the normal metal can tunnel into the superconductor, cooling the normal metal.

Laser Cooling

- **Principle:** Based on the conservation of momentum when a photon is absorbed and emitted.
- Particles are irradiated by a laser beam with a frequency (ω) that is slightly detuned **below** their resonance frequencies (ω_0).
- When a photon traveling opposite to the particles becomes **blueshifted** by $\delta = \omega_0 - \omega$ due to the Doppler effect, the photon can be absorbed by a particle. After that, the particle relaxes to its ground state and the particle experiences a recoil when it emits a photon in a random direction. This slows down the particles, achieving low temperatures.
- **Record:** Ensemble of ^{87}Rb atoms in two dimensions cooled to less than 50 pK.

Evaporative Cooling

- **Method:** This method is based on the preferential removal of those atoms from a confined sample with an energy higher than the average energy, followed by a rethermalization of the remaining gas by elastic collisions.
- **Result:** Those that remain have much lower average energy (temperature) and so they occupy a smaller volume near the bottom of the trap, thereby increasing their density.
- **Forced cooling:** The evaporative cooling can be "forced" by inducing radio-frequency fields that cause only the most energetic atoms to undergo transitions to magnetic states that are not trapped (the depth of the trap itself is gradually decreased).
- **Record:** Bose-Einstein condensate of ^{87}Rb atoms was obtained using evaporative cooling. The condensate fraction had a temperature of 170 nK.

2.9.6 TODO:Summary

Table 2.7: Summary of Cryogenic Techniques

Temperature range	Physical principle	Technique
300 K → 240 K	Evaporation of a proper fluid	Standard home or industrial refrigerator
300 K → 50 K	Stirling or modified Stirling cycle (regenerative cycles)	Single-stage closed-cycle cryocoolers
300 K → 3 K	Stirling or modified Stirling cycle, pulse-tube regenerative cycle	Two-stage closed-cycle cryocoolers
300 K → 4.2 K	Enthalpy or latent heat of a liquefied gas (Ne, O, H, He ⁴)	Cryostat with VTI (Variable Temperature Insert)
4.2 K → 1.3 K	Evaporation of liquid ⁴ He	⁴ He Cryostat with 1K pot
2.5 K → 0.25 K	Evaporation of liquid ³ H	⁴ He cryostat with 1K pot and ³ He evaporation chamber
4.2 K → 2 mK	Mixing of ³ He in ⁴ He	Dilution refrigerator (continuous operation)
3 K → 2 mK	Adiabatic demagnetization of paramagnetic salts	Adiabatic demagn. refrigerator ("one-shot")
10 mK → 0.1 μ K	Adiabatic demagnetization of nuclear spins	Nuclear magnetic refrigerator ("one-shot")
1 K → 0.05 K	Direct electron cooling (Normal metal-Insulator-Superconductor)	Tunnel junction cooling
1 K → 50 pK	Doppler effect of photons (Cs, Rb atoms)	Laser cooling
1 mK → 170 nK	Preferential removal of high energy atoms from a trap	Evaporative cooling

2.10 Superconducting materials and technologies

2.10.1 TODO:Introduction to the phenomenology of superconductivity

Historical notes

- 1908: Kamerlingh-Onnes (Leiden) liquefaction of helium
- 1911: first observation of superconductivity in Hg ("...Mercury had passed into a new state, which on account of its extraordinary electrical properties may be called the superconductive state.")
- 1933: Meissner and Ochsenfeld (Germany) → ideal diamagnetism
- 1941: niobium-nitride was found to superconduct at 16 K
- 1957: The first widely-accepted theoretical understanding of superconductivity (John Bardeen, Leon Cooper, John Schrieffer, USA) → **BCS theory**
- 1962: Brian D. Josephson, a graduate student at Cambridge University → **Josephson effect**
- 1962: first commercial superconducting wire (NbTi)
- 1960s: High-energy, particle-accelerator electromagnets made of copper-clad NbTi (Rutherford-Appleton Laboratory, UK) → first employed in a superconducting accelerator at the Fermilab Tevatron (USA) in 1987
- 1986: Alex Müller and Georg Bednorz (IBM, Switzerland) → first **HTS** (30 K)
- 1987: YBCO ($T_c = 92K$)
- 1989: the first company to capitalize on high-temperature superconductors, Illinois Superconductor (today ISCO International)
- 2001: MgB_2 ($T_c = 39K$), outperform NbTi and Nb_3Sn wires in high magnetic field applications like MRI
- 2003: Nobel prize to Abrikosov, Ginzburg, Legget, for pioneering contributions to the theory of superconductivity
- 2006: pnictides, iron-based superconductors
- 2015: hydrides (high pressure)

Fig. 1 Superconductivity in lutetium-nitrogen-hydrogen at near-ambient pressures. a, Evolution of the superconducting transition temperature (T_c) of recompressed nitrogen-doped lutetium hydride as a function of pressure (P), illustrating a clear dome-shaped peak around 10 kbar with a T_c of 294 K.

Summary of the main properties of superconductors

Perfect conductivity

superconductor normal metal

But superconductors are not just perfect conductors, because:

Meissner effect: perfect diamagnetism

Perfect conductor Superconductor (Diagram comparing cooling ZFC vs FC for a perfect conductor vs a superconductor, showing levitation)

- Conventional SC → BCS theory
 - Phonon-mediated superconductivity
 - Cooper pairs (bosons) formed from electrons (fermions), similar to a Bose-Einstein condensate
 - Characterized by an **Energy gap** (Δ) in the density of states
- Unconventional SC → at least one feature among:
 - multiband superconductivity
 - triplet spin state
 - non-phononic coupling (e.g. spin fluctuations in SC contiguous to magnetic phases or even with a coexisting magnetic order, as for iron-based SCs)
- The **superconducting order parameter** is a complex quantity $\Delta(k)e^{-i\phi(k)}$ where Δ is the magnitude of the superconducting gap and ϕ is the phase.

2.10.2 TODO:Vortex dynamics, vortex pinning

Superconductivity is defined by a critical temperature (T_c), critical magnetic field (H_c), and critical current density (J_c).

Type I vs. Type II Superconductors

- The **Ginzburg-Landau (GL) parameter**, $\kappa(T)$, is microscopically defined as $\kappa = \lambda/\xi$, where λ is the penetration depth of the magnetic field and ξ is the coherence length of the Cooper pairs.
- **Type I superconductor:** $\kappa < 1/\sqrt{2}$. Magnetic flux is expelled up to H_c (normal state).
- **Type II superconductor:** $\kappa > 1/\sqrt{2}$. Magnetic field penetrates in the form of quantized **vortices** between H_{c1} and H_{c2} (mixed state).

CRITICAL CURRENT BY DESIGN

- To have SCs carrying significant currents, the **microstructure must be optimized by introduction of defects to pin the flux lines**, leading to so-called **hard superconductors**.
- Examples of introduced defects: Heavy ion irradiation of HTS to create point-like defects or **columnar defects**. Columnar defects have a diameter ($\sim \xi$) that matches the vortex core, making them highly efficient, but they cause anisotropic pinning.

2.10.3 TODO:Review on superconducting materials (relevant for applications) and preparation methods

None of the good conductors (Cu, Ag, Au, Pt) or the magnetic materials (Fe apart) is a superconductor.

Main classes of superconductors:

1. Simple metals (Nb, $T_c=9$ K)
2. Alloys (Nb_3Ge , $T_c=23$ K)
3. Organic SCs ($T_c=12$ K)
4. Molecular SCs (C_{60} , $T_c=19$ K)
5. MgB_2 ($T_c=39$ K)
6. Graphite intercalation compounds (CaC_6 , $T_c=12$ K)
7. Cuprates (Hg-1223, $T_c=134$ K)

Low-Temperature Superconductors (LTS): Nb, NbTi, Nb_3Sn

- Superconductivity in Type I and Type II superconductor elements and alloys is **phonon-mediated** and the **BCS theory applies**.
- Elemental superconductors are very sensitive to magnetic fields and cannot carry significant currents. Most metals are **Type I** superconductors.
- Only three metals, among them **Nb**, exhibit **Type II** superconductivity → Nb is the only application-relevant superconductor metal.
- **Nb**: Used in sensor technologies, metrology, digital electronics, and RF cavities for particle accelerators. For high RF performance, very pure Nb is needed, produced by multiple e-beam remelting, characterized by a Residual Resistivity Ratio (RRR) of 250–400.
- **NbTi**: Alloying usually changes metals into Type II conductors and/or enhances the upper critical magnetic field. NbTi is an Nb-based alloy.

- Nb_3Sn : Intermetallic compounds of specific composition and crystal structure, for example A15 structure, show much improved properties including higher T_c than alloys. Nb_3Sn is an Nb-based A15-type compound.
- **A15-type intermetallic compounds:** Chemical formula A_3B (where A is a transition metal and B can be any element). Highest T_c values: Nb_3Ge (23 K), Nb_3Ga (20 K), Nb_3Al (19 K), Nb_3Sn (18 K), V_3Si (17 K), and V_3Ga (16 K).
- **NbN:** Within conventional superconductors, NbN has:
 - High critical temperature $T_c \sim 16$ K
 - Energy gap $\Delta \sim 2.5$ meV
 - Upper critical field $B_{c2} \sim 40$ T
 - Large penetration depth $\lambda \sim 200$ nm (vs $\lambda \sim 50$ nm for bulk Nb), resulting in higher kinetic inductance, useful for microwave kinetic inductance detectors
 - Short coherence length $\xi \sim 6.5$ nm (vs $\xi \sim 38$ nm for bulk Nb), allowing fabrication of few nm thick thin films with moderately high T_c
- **Applications of NbN:** Josephson junctions, superconducting hot electron bolometers, superconducting single photon detectors, microwave kinetic inductance detectors.
- **Wire Manufacturing Routes:** Pure Nb, the Nb-based alloy NbTi, and the Nb-based A15-type compound Nb_3Sn are virtually the only SC materials industrially produced in large quantities as **multifilamentary wires** (with SC filaments embedded in a normal conducting matrix).

Magnesium diboride (MgB_2)

- MgB_2 is a simple binary compound whose high- T_c superconductivity was discovered in 2001.
- **Crucial properties for applications:**
 - Very simple structure, made of inexpensive elements
 - Relatively high T_c (39 K)
 - Remarkable absence of **weak-links**, meaning no need for high degree texturing to transport high superconducting current
 - Possibility of enhancing J_c and irreversibility field in thin films
 - Possibility of making superconducting wires by a simple **Powder-In-Tube (PIT) technique**
- **Fundamental physical mechanisms:** MgB_2 is characterized by the presence of **two bands and two energy gaps**.

- **Crystal Structure:** 2D honeycomb layers formed by Boron atoms are sandwiched by the triangular Mg layers (like intercalated graphite).
- **Band Structure:** Two σ bands (2D, hole-type) and two π bands (3D, hole- and electron-type) cross the Fermi level.
- **Two Nodeless Gaps:** These different carriers induce two different s-type nodeless gaps: $\Delta_\sigma \sim 7$ meV and $\Delta_\pi \sim 2$ meV.
- **PIT Technique (Wire Fabrication):**
 - A metallic tube (Ni, Fe, and Ti alloys are preferred) is filled with pre-reacted MgB_2 (ex situ) or Mg and B precursors (in situ).
 - The tube is cold worked (rolling/drawing) into a wire or tape.
 - Intermediate heat treatments are performed to recover mechanical properties and release structural stress.
 - A final heat treatment is performed for the sintering of the MgB_2 powders.

High-Temperature Superconductors (HTS) - cuprates: $YBa_2Cu_3O_{7-\delta}$ (YBCO)

- $YBa_2Cu_3O_{7-\delta}$ (YBCO) has a $T_c = 92$ K.
- **Theory:** Models relate high- T_c superconductivity to the presence of **fluctuating charge stripes**.
 - **BCS-like mechanism:** fluctuating stripes act as the mediating boson ("glue") for the superconducting pairing.
 - **Strongly correlated electron picture:** fluctuations of one-dimensional charge stripes frustrate the dominant Charge Density Wave (CDW) order, allowing the subdominant superconducting order to prevail.
- **Technological Challenges:**
 - HTSs are complex materials with different heavy metals and exotic atoms, limiting deposition techniques that preserve stoichiometry.
 - Deposition of YBCO requires a substrate heated at high temperatures (over 700 K for PLD).
 - Cuprates are ceramic materials: prone to defects, brittle, and sometimes chemically reactive.
 - A crucial aspect of YBCO is **oxygen out-diffusion** (especially for $p < 0.15$) which alters SC properties and is difficult to control at the nanoscale.

2.10.4 Iron-based superconductors (IBS)

The discovery of IBS was unexpected due to the antagonistic relationship between superconductivity and magnetism.

Four main families with distinctive crystallographic structures:

1. "1111" family: $LnFeAsO$ (Ln , lanthanides)
2. "122" family: AFe_2As_2 (A , alkaline earth metal)
3. "111" family: $LiFeAs$
4. "11" family: $FeCh$ (Ch , chalcogen ion)

There are also other families, with more complex structures, such as "42622" ($Sr_4Sc_2O_6Fe_2As_2$) and "32522" ($Ca_3Al_2O_{5-y}Fe_2Pn_2$, $Pn=As, P$).

SC mechanism

- Unconventional nature of the pairing mechanism → a comparison with the high- T_c cuprates is often made.
- **Spin fluctuations** are considered the most likely mediators of the Cooper pairing, leading to an $s\pm$ symmetry of the order parameter: in all cases, magnetism must be suppressed by either doping or pressure to allow superconductivity to emerge, before optimal bulk-phase superconductivity is reached.
- **Pairing Symmetry:** experimental evidence points toward a s-wave symmetry and multiband in most IBSs, with a sign-changing order parameter between different Fermi surface sheets ($s\pm$ symmetry). However, other symmetries have been proposed, depending on the specific material and doping level.

Nowadays, is still a work in progress.

Phase diagram and main properties

Phase Diagram: Reminiscent of high- T_c cuprates, characterized by the **competition between magnetic and superconducting orders** (Spin Density Wave (SDW) order vs. SC order).

- **Coexistence region:** In some IBSs (e.g., 122 family), there is a region of coexistence between SDW and SC orders at low doping.
- **Doping:** The superconducting ground state appears with increasing doping (electron or hole type), below a dome-shaped region, with maximum T_c at optimal doping.

- **Electron doping:** e.g., Co or Ni substitution at the Fe site in 122 compounds.
- **Hole doping:** e.g., K substitution at the Ba site in 122 compounds.
- **Pressure:** In some IBSs (e.g., 11 family), superconductivity can be induced or enhanced by applying external pressure. Both chemical pressure (e.g., isovalent substitution) and physical pressure can modify the electronic structure and lattice parameters, affecting the superconducting properties. The goal is to reduce the anion height relative to the Fe plane, which is correlated with higher T_c values.

Critical magnetic fields

- **Different materials:** Each IBS material has its own challenges in terms of synthesis, stability, and performance. For example, 1111 compounds often require high-pressure synthesis, while 11 compounds are more straightforward to prepare but may have lower T_c values.
- **Critical Currents:** IBSs have high and quite isotropic J_c (weakly varying along crystal directions). J_c shows only a weak decrease on application of a magnetic field at low temperatures (similar to YBCO); there's a low anisotropy in J_c values with respect to the field orientation.
- **Upper Critical Fields:** IBSs exhibit very high upper critical fields (H_{c2}), often exceeding 50 T at low temperatures, with relatively low anisotropy compared to cuprates. This makes them promising for high-field applications.
- **Transparency of grain boundaries:** Unlike cuprates, IBSs show relatively high critical current densities across grain boundaries, even at large misorientation angles. This is advantageous for polycrystalline materials and wire fabrication. Transparency parameter = $J_{c,intern}/J_{c,intra}$, meaning we can consider each grain as a polycrystalline sample, measuring inter- and intra-grain J_c . The closer this parameter is to 1, the better the grain boundary transparency, the lower the weak-link behavior. IBSs have transparency parameters close to 1, even at high misorientation angles (up to 9 times higher than YBCO at 24°).

This is an advantage for applications in high magnetic fields, such as magnets for MRI or particle accelerators, thanks to their high H_{c2} and J_c values.

Material preparation

For the preparation of polycrystalline samples and thin films of IBSs, solid-state reaction methods and high-pressure synthesis methods are commonly used. Large and pure crystals can be grown in fluxes of binary FeAs (e.g., the "self-flux" method for Co-doped 122 compounds). Thin films are not easy to grow due to the volatile nature of As, but high-quality epitaxial films of the 122 or

11 phase can be grown by pulsed laser deposition (PLD). Metal flux methods are less common due to the less predictable incorporation of impurities from the flux into the crystal structure.

Wires and tapes

Powder in tube (PIT) method is commonly used for fabricating wires and tapes of IBSs, particularly for the 122 family. The process involves filling a metallic tube (e.g., Ag, Fe) with IBS powder, followed by mechanical deformation (drawing, rolling) to form wires or tapes. Subsequent heat treatments are performed to promote phase formation and improve superconducting properties. Challenges include controlling the phase purity, texture, and grain connectivity to achieve high critical current densities.

The synthesis of the superconducting compounds occurs either before filling the metal tube (ex situ process) or during the final heat treatment after mechanical deformation (in situ process).

Discovered in 2008, IBS wires and tapes are still in the early stages of development compared to more established superconductors like NbTi or YBCO. However, their promising properties make them a subject of active research for future applications.

Coated conductors based on IBSs are still in the early stages of development compared to more established superconductors like YBCO. However, their promising properties make them a subject of active research for future applications.

They're typically fabricated using techniques similar to those used for YBCO coated conductors, such as pulsed laser deposition (PLD) or chemical vapor deposition (CVD), on flexible substrates with appropriate buffer layers to promote epitaxial growth and enhance performance.

Applications of IBS Wires and Tapes It can be envisaged that IBS wires and tapes have good possibilities for magnet applications at $20\text{-}30\text{K}$ (cryocooler temperatures) where the Nb-based superconductors cannot be used anymore. Potential applications include MRI magnets, high-field research magnets, and magnets for particle accelerators. Their high upper critical fields and relatively low anisotropy make them suitable candidates for these applications, provided that challenges related to fabrication and performance optimization can be addressed.

2.10.5 Hints on Superconducting Film Deposition

Material Requirements vs. HTS:

- The improvement of superconducting properties in HTS (increase of T_c , B_{c2} , Δ) is accompanied by a significant enhancement of: (i) the number of elements, (ii) structural complexity, and (iii) **anisotropy**.
- **Anisotropy:** Whereas LTS materials like Nb or NbN are generally isotropic, the quasi 2D nature of the oxide SCs leads to a large anisotropy in nearly all parameters, with anisotropy factors $\gamma = \xi_{ab}/\xi_c = \lambda_c/\lambda_{ab} \approx 6$.

2.10.6 Cryoelectronic devices

Josephson Junctions-based devices:

- SQUIDs (Superconducting Quantum Interference Devices) for ultra-sensitive magnetometry; used in biomagnetism, geophysics, and nondestructive evaluation, neurology and medical diagnostics. (Nb/AI/AIOx/Nb tunnel junctions)
- Arrays of Josephson junctions for voltage standards (programmable Josephson voltage standard, PJVS); used in metrology laboratories worldwide.
 - SQIFs (Superconducting Quantum Interference Filters) for high precision and wide-band microwave detection and amplification. (Nb/AI/AIOx/Nb tunnel junctions).
 - RSFQ (Rapid Single Flux Quantum) digital electronics for high-speed, low-power digital circuits; used in specialized computing applications. (Nb/AI/AIOx/Nb tunnel junctions)
- Teraheartz (THz) detectors and mixers for high-frequency applications; used in astronomy, security screening, and spectroscopy. (NbN/MgO/NbN tunnel junctions)

Superconducting bolometers and calorimeters: Used to study cosmic microwave background radiation, X-ray astronomy, and particle physics. (Transition Edge Sensors (TES) based on Mo/Cu or Ti/Au bilayers; Kinetic Inductance Detectors (KIDs) based on NbN or TiN thin films). **Microwave devices:** Used for mobile communications, satellite communications, and radar systems. (HTS filters based on YBCO thin films). **Fluxonix devices:** Representing a new class of superconducting electronics based on the manipulation of magnetic flux quanta (fluxons) in long Josephson junctions; potential applications in high-speed digital circuits and signal processing. (Nb/AI/AIOx/Nb tunnel junctions).

Power applications

Coated conductors: Might provide the basis for future HTS power applications, such as power cables, fault current limiters, transformers, and motors/generators. (YBCO thin films on flexible metal substrates with buffer

layers). **Coating of RF cavities:** For particles accelerators, to improve performance and reduce losses. (Nb thin films on copper substrates).

2.10.7 TODO:state of the art

2.10.8 TODO:Material requirements

2.10.9 TODO:Substrate Requirements

2.10.10 Brief overview of applications of superconductivity

Applications include: Electric Power; Transportation; Medical Imaging and Diagnostics; NMR Applications in Pharmaceuticals, Biotechnology, Genomics and Materials Science; Industrial Processing; High-Energy Physics and Other Areas of Research; Wireless Communications; Instrumentation, Sensors, Standards and Radar; Large-Scale Computing; Renewable Energy.

Electric Power

- Power Cables (e.g., AmpaCity Project in Essen, Germany). Cables can carry up to 10 times the current of conventional copper or aluminum cables of the same diameter, thanks to the zero-resistance property of superconductors. It's useful because aerial space for cables is limited in urban areas, and underground cables can be installed.
- Fault Current Limiters (FCLs) (e.g., ABB's 9MVA FCL)
- Transformers (e.g., ABB's 1 MVA HTS Transformer)
- Motors and Generators (e.g., Superconducting Wind Turbine Generator by AMSC)

Nowadays, we are subject to energy fluctuations and shortages, so the use of superconductors in power applications can help improve efficiency and reliability in power transmission and distribution systems.

Transportation

- Marine Ship Propulsion (e.g., 36MW HTS Ship Drive Motor). This propulsion system offers significant advantages over conventional systems, including reduced weight and volume, higher efficiency, and lower maintenance requirements. IT has been successfully tested on a naval vessel, demonstrating its potential for future maritime applications and aerospace propulsion systems.
- Degaussing Coils

- Magnetically Levitated Trains (Maglev) (e.g., JRC L-zero series SCMAGLEV). The SCMAGLEV system uses superconducting magnets to create strong magnetic fields that levitate and propel the train along the guideway. This technology allows for high-speed travel with minimal friction, resulting in a smooth and efficient ride. The SCMAGLEV trains have achieved speeds of over 600 km/h in test runs, showcasing their potential for revolutionizing high-speed rail transportation.

Medical Imaging and Diagnostics

- Magnetic Resonance Imaging (MRI)
- Ultra-Low Field Magnetic Resonance Imaging (ULF-MRI)
- Magnetoencephalography (MEG) and Magnetic Source Imaging (MSI)
- Magnetocardiography (MCG)

It's done measuring the magnetic fields produced by the electrical activity of the brain and heart, respectively, using highly sensitive SQUID sensors. These techniques provide non-invasive ways to study brain function and diagnose cardiac conditions. Current is induced in the body by magnetic fields, and the resulting signals are detected by SQUIDS. From 1990s cryocoolers are used to cool the sensors instead of liquid helium.

NMR Applications in Pharmaceuticals, Biotechnology, Genomics and Materials Science

- High-Resolution Nuclear Magnetic Resonance (NMR) Spectroscopy.
- Magnetic Resonance Spectroscopic Imaging (MRSI)

Industrial Processing

- Magnetic Separation
- Induction Heating

High-Energy Physics and Other Areas of Research

- Particle Accelerators (e.g., Large Hadron Collider at CERN)
- Fusion Energy (e.g., ITER Tokamak)
- Space Applications (e.g., Superconducting Magnetic Energy Storage for satellites)

Wireless Communications

- Superconducting Filters for Base Stations (e.g., NGK Insulators Ltd. HTS Filters)
- Superconducting Mixers and Detectors for Satellite Communications

Instrumentation, Sensors, Standards and Radar

- Superconducting Quantum Interference Devices (SQUIDs) for Ultra-Sensitive Magnetometry
- Josephson Voltage Standards
- Superconducting Mixers and Detectors for Radar Systems

Large-Scale Computing

- Superconducting Digital Electronics (e.g., Rapid Single Flux Quantum (RSFQ) Technology)
- Quantum Computing (e.g., Superconducting Qubits used by IBM, Google, and Rigetti)

TODO:Neuromorphic Computing

- Computer hardware that works according to principles of the human brain.
- The brain's activities of synaptic weighting, spiking, learning, optimization, and networking can all be made with superconducting electronics.
- Spiking and thresholding are similar to neurons, except very fast.
- Coupling through mutual inductance.

Renewable Energy

- Wind Energy (e.g., AMSC SeaTitan Wind Turbine Generator)
- Energy Storage: Superconducting Magnetic Energy Storage (SMES)
- ARC: Affordable Robust Compact fusion reactors.

2.11 The Technology of Josephson Junctions

2.11.1 Josephsons Junctions:main concepts and phenomenology

The Josephson Effect: see the course "Quantum Condensed Matter Physics"

In 1962, Brian Josephson predicted the dissipationless supercurrent flow between two superconductors (SCs) separated by a thin insulating barrier, a phenomenon quickly confirmed experimentally. The Josephson effect is fundamentally rooted in the macroscopic quantum coherence of the superconducting condensate.

Each SC is characterized by a macroscopic wavefunction ψ , defined as:

$$\psi = \sqrt{n}e^{i\phi}$$

where n is the Cooper-pair density and ϕ is the phase.

The phase's gradient governs the supercurrent flow, as the quantum mechanical expression for the probability current density is $\vec{J} \propto \text{Im}[\psi^* \nabla \psi]$.

For a uniform Cooper-pair density, this simplifies to:

$$\vec{J} \propto \nabla \phi$$

When two SCs are weakly coupled through a tunneling barrier, the phase difference $\delta\phi$ drives a dissipationless supercurrent.

Josephson Equations

Due to the single-valuedness and time-reversal symmetry of the condensate wavefunction, the current is a periodic and odd function of the phase difference, $\varphi \equiv \delta\phi$. This leads to the characteristic sinusoidal current-phase relation (d.c. Josephson effect):

$$I = I_c \sin(\varphi) \tag{2.15}$$

where I_c is the critical current.

The Josephson effect also encompasses a dynamical component that links the time derivative of the phase difference to the voltage V across the junction, which follows from the Schrödinger equation and Faraday's law of electromagnetic induction:

$$\dot{\varphi} = \frac{2eV}{\hbar} \tag{2.16}$$

This establishes a direct connection between the quantum phase and observable electrical quantities.

We have the Josephson effect as long as the macroscopic wavefunctions of the two electrodes overlap in the barrier region.

a.c. Josephson Effect and Shapiro Steps

The effect extends beyond Josephson's predictions for SIS junctions, and can exist if SCs are connected by a "weak link" of any physical nature (normal metal, semiconductor, SC with smaller critical temperature, geometrical constriction, etc.)

For a finite voltage $V \neq 0$, the phase φ varies linearly in time: $\varphi = \varphi_0 + (\frac{2e}{\hbar})Vt$. Consequently, an alternating current appears:

$$I = I_c \sin \left(\varphi_o + \frac{2e}{\hbar} V t \right)$$

with an angular frequency $\omega = 2eV/\hbar$. This is known as the a.c. Josephson effect. The ratio between frequency ν and voltage is a fundamental constant:

$$\frac{\nu}{V} = \frac{2e}{\hbar} \approx SI483.6\text{megahertz/microvolt}$$

A direct manifestation of the a.c. Josephson effect is the presence of current steps (Shapiro steps) at constant voltages in the presence of microwave irradiation.

$$V_n = \frac{n\hbar}{2e} \nu_o \quad (n = \pm 1, \pm 2, \dots)$$

where ν_o is the frequency of the applied radiation.

Flux Quantization

The magnetic flux Φ through any closed superconducting ring only appears in quantized values of the flux quantum Φ_0 :

$$\Phi = n\Phi_0, \quad \text{where } \Phi_0 = \frac{\hbar}{2e} \approx SI2.07e - 15\text{weber}$$

TODO:Energy in a Josephson Junction

The energy stored in the junction associated with the phase difference φ is the Josephson energy E_J :

$$E = E_J(1 - \cos \varphi) \tag{2.17}$$

where E_J is defined as:

$$E_J = \int_0^t I(t')V dt' = TODO$$

Josephson Junction as Nonlinear Inductance

A Josephson junction can be considered to have a nonlinear inductance L . Starting from the classical relation $V = L \frac{dI}{dt}$ and substituting the Josephson equations, the inductance is derived as:

$$L = \frac{L_J}{\cos(\varphi)}$$

with the characteristic inductance L_J :

$$L_J = \frac{\Phi_0}{2\pi I_c} = \frac{\hbar}{2eI_c}$$

This also implies the relation $E_J = L_J I_c^2$.

Hamiltonian of a Josephson Junction: see the course "Quantum Devices"

The passage from the classical to the quantum description involves replacing classical variables with corresponding quantum operators. The circuit Hamiltonian is built from three main energy components, all expressed in terms of the phase φ :

- Kinetic energy associated with charging: $K = \frac{1}{2}CV^2$
- Potential energy from Josephson coupling: $U_J = -E_J \cos \varphi$
- Potential energy from external inductance: $U_L = \frac{1}{2}LI^2 = \Phi^2/(2L)$

These are connected by the Josephson relations $\dot{\varphi} = 2eV/\hbar$ and the flux-phase relation $\Phi = \Phi_0\varphi$.

For a current-biased Josephson junction, the Hamiltonian can be expressed as:

$$H = E_{C0}n^2 - E_J \cos \varphi - \frac{\hbar}{2e}I_b\varphi$$

where I_b is the applied bias current, $E_C = E_{C0}n^2 = \frac{(2en)^2}{2C}$ is the charging energy, and n is the Cooper pair charge number operator conjugate to φ .

2.11.2 Current-Phase Relation and Junction Types

Different Types of Josephson Junctions: see the course "Quantum Condensed Matter Physics"

The Josephson effect extends beyond the original Superconductor-Insulator-Superconductor (SIS) prediction to exist wherever SCs are connected by a “weak link.”

- **SIS and S-I-N Junctions:** These are tunnel junctions where the insulating (I) layer acts as a barrier.
- **S-N-S Junctions:** More transmissive barriers are obtained by replacing the I layer with a metallic (N) layer (or a semiconductor or ferromagnet). The Josephson effect occurs as long as the barrier thickness (L) is comparable with the coherence length ξ_n induced in the barrier:

$$\xi_n = \left(\frac{\hbar D_{diff}}{k_B T} \right)^{1/2}$$

where T is the temperature, $D_{diff} = \frac{v_F l}{3}$ is the normal metal diffusion constant, and k_B is the Boltzmann's constant.

- **Consequence of replacing an I with an N as a barrier:**

- Change in the resistance R_n of the junction.
- Change in the effective capacitance of the junction.
- New physical processes dominate over tunneling effects:
 - * **Proximity Effect:** The mutual influence of a superconducting layer in contact with a normal metal (N). The superconducting properties are weaker in the S within a length of the order of its coherence length ξ from the interface. The proximity effect is controlled by the nature of the interface, its effective transparency, and by the boundary conditions, which involve both ξ and ξ_n , as well as the thickness of the N and S layers.
 - * **Andreev Reflection:** A microscopic process where a dissipative electrical current is converted at an N/S interface into dissipation-free supercurrent.
 - For poorly transparent interfaces, an S-N-S junction can effectively behave as an **S-I-N-I-S** type.

- **Alternative Configurations:** The barrier can be realized using materials like graphene, topological insulators, or nanowires. In these cases, the distance between the SC electrodes L is a critical parameter and must be of the order of ξ_n .
- **Edge-Type Bridges:** Configurations where the barrier is deposited on the edge of a superconductor, advantageous for submicron junctions and high-temperature superconductors (HTS).

Microscopic Mechanisms (Proximity Effect and Andreev Reflection)

In SNS junctions, new physical processes occur that dominate over simple tunneling:

- **Proximity Effect:** The mutual influence of a superconducting layer in contact with a normal metal (N). The superconducting properties are weakened in the SC (S) layer within a length of the order of its coherence length ξ from the interface.
- **Andreev Reflection (AR):** A microscopic process where a dissipative electrical current is converted at an N/S interface into a dissipation-free supercurrent. An electron excitation in the normal metal is reflected as a hole excitation, with the missing charge of $2e$ absorbed by the superconducting condensate as a Cooper pair.

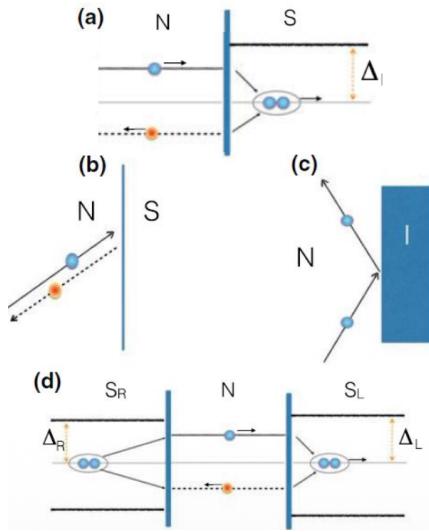


Figure 2.34: (a) Energy representation of the Andreev reflection (AR) process: an electron excitation slightly above the Fermi level in the normal metal is reflected at the interface as a hole excitation slightly below the Fermi level. (b) Spatial representation of AR at the S/N interface. (c) Spatial representation of normal reflection by an insulator. Normal reflection conserves charge but does not conserve momentum, while AR conserves momentum but does not conserve charge. The electron is reflected as a hole with the same momentum and opposite velocity. The missing charge of $2e$ is absorbed as a Cooper pair by the superconducting condensate. (d) Energy representation of Andreev reflections in SNS junctions.

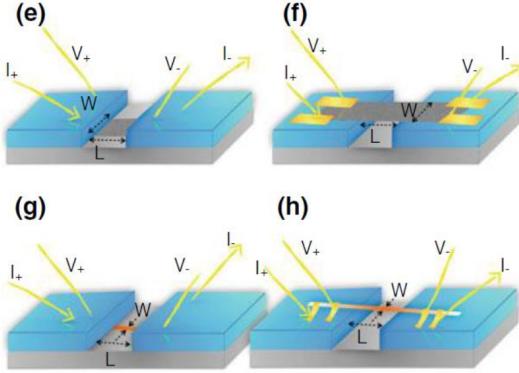


Figure 2.35: (e) A flake of graphene or a topological insulator used as a barrier. (f) A suspended nanowire deposited on the superconductor. (g) A nanowire as a barrier. (h) Suspended configuration with a trench separating the two electrodes. These configurations assemble pre-built blocks and may circumvent material science incompatibilities due to critical deposition conditions. The distance between the electrodes, L , is a critical parameter and must be of the order of the coherence length in N, ξ_n .

Micro-restrictions and Grain Boundaries

- **Point Contacts and Micro-restrictions:** Another practical way to form a Josephson junction is through a point contact or by creating a micro-restriction in a superconducting thin film. Scanning Tunnel Microscopy and atomic point contacts represent advanced and refined versions of the original point contact technique, and they are commonly used for tunneling spectroscopy. For widths of the order of a few times ξ_s , the microbridge will behave as a Josephson weak-link. This type of junction depends very critically on the dimensions of the microbridge and its characteristic lengths, ranging from a Josephson junction to a bridge hosting phase slip events or Abrikosov vortex motion, respectively. A kind of ‘phase diagram’ can be derived as a function of their dimensions (width W , length L). In the limit of long microbridges, Josephson behavior substantially disappears in favor of a regime of strong superconductivity. The critical current is controlled by standard depairing processes without any Josephson coherence.

- **Grain Boundaries (GBs):** Josephson coupling can also take place at grain boundaries (GBs), and this property has been widely used to produce HTS JJs. Any GB can be considered as the result of three fundamental operations:

- One electrode can be tilted with respect to the other electrode around the c-axis (001 tilt),

- Or tilted around the a or b direction (tilt of the c-axis, 100 tilt),
- Or twisted along the junction interface (100 twist).

GBs can realize quite special atomically flat interface configurations. In HTS, a d-wave order-parameter symmetry generates quite unconventional features.

2.11.3 Revisiting the Current-Phase Relation (CPR)

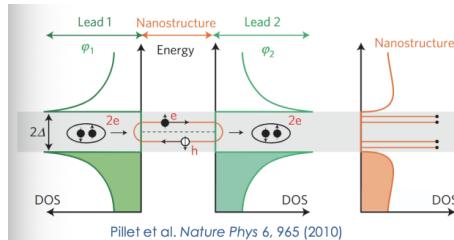
The d.c. Josephson equation $I = I_c \sin(\varphi)$ reflects an idealized tunneling limit in which only the first harmonic is significant. However, in more general weak-links (e.g., more transmissive barriers), the CPR can exhibit substantial higher-order Josephson harmonics, depending on the microscopic properties of the junction.

Andreev Bound States (ABS): A dissipationless Josephson current arises from quasiparticle excitations confined within the weak link. These quasiparticles occupy Andreev bound states (ABS), which form due to multiple Andreev reflections at the interfaces between the superconductors and the weak link. At energies within the superconducting gap, the Andreev reflection process leads to the formation of discrete resonant states (ABS) of entangled electron-hole pairs confined between the superconductors. ABS are electronic analogues to the resonances in an optical Fabry-Pérot cavity. The energies of the ABS depend periodically on the superconducting phase difference $\varphi = \varphi_1 - \varphi_2$, analogous to the optical cavity length.

Generalized Current-Phase Relation: The $I(\varphi)$ expression should include higher-order contributions in the coupling, originating from the ABS. Summing over all conduction channels and applying the second Josephson relation, we arrive at the generalized CPR:

$$I(\varphi) = \sum_i \sum_k I_{ki} \sin(k\varphi), \quad (2.18)$$

where the higher harmonic k -contributions are manifestations of multi-Cooper-pair tunneling processes, depending on the barrier transparency (k -multiple Andreev reflection process).



Anisotropic Order Parameter Symmetry: The CPR may also depend on the angles θ_1 and θ_2 of the crystallographic axes with respect to the junction interface of the left and right electrodes, for anisotropic order parameter symmetry. Examples of the occurrence of a second harmonic contribution are given by high-temperature superconductors (HTS) with d-wave symmetry and by particular junctions with ferromagnetic barriers.

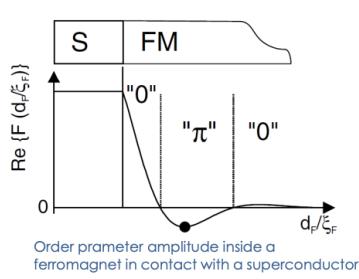
Second Harmonic Contributions: Second-order contributions, which correspond to simultaneous tunneling of two Cooper pairs, are particularly relevant in certain junctions. For instance:

- HTS d-wave Josephson junctions, where the anisotropic order parameter symmetry leads to higher harmonic contributions.
- Superconductor–Ferromagnet–Superconductor (SFS) junctions, where the interplay between superconductivity and ferromagnetism modifies the CPR.

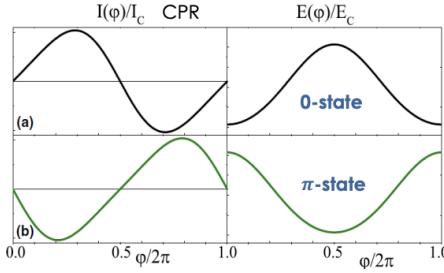
These higher-order terms highlight the need to reconsider the CPR in non-ideal junctions.

2.11.4 SFS Josephson Junctions: The π Junction

Superconductor–Ferromagnet–Superconductor (SFS) junctions exhibit unique properties due to the interplay between superconductivity and ferromagnetism. The superconducting order parameter (OP) penetrates into the ferromagnetic layer via the proximity effect, but with distinct characteristics compared to normal metals.



Spatial Oscillations of the Order Parameter In SFS junctions, the correlated electron-hole pairs, with opposite spin directions, experience the exchange field of the ferromagnet. This leads to spatial oscillations of the superconducting OP within the ferromagnetic layer. The oscillation period is determined by the exchange field. Unlike in NS bilayers, where the OP decays monotonically, the oscillations in the F-layer can result in a sign change of the OP amplitude.



The π Junction State The sign change of the OP amplitude corresponds to periodic $0-\pi$ phase jumps at certain points in the F-layer. By choosing appropriate values for the exchange field and the thickness d_F of the F-layer, the OP amplitude on both sides of the junction can become negative. This results in a π phase shift in the current-phase relation (CPR):

$$I = -I_c \sin \varphi = I_c \sin(\varphi + \pi).$$

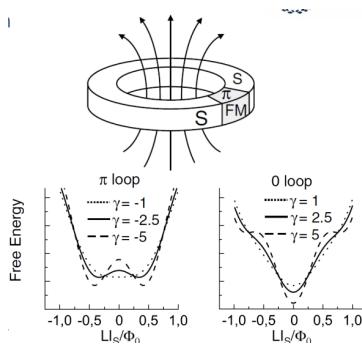
Such junctions are called π junctions because the minimum of the free energy occurs at $\varphi = \pi$ instead of $\varphi = 0$, as in standard 0-state junctions.

Spontaneous Current and Magnetic Flux A superconducting ring with an inserted π junction (or an odd number of π junctions) exhibits a spontaneous supercurrent and a corresponding magnetic flux of half a flux quantum Φ_0 in the ground state. This arises from the fluxoid quantization condition:

$$\varphi_i + \varphi = 2\pi n,$$

where $\varphi = (2e/\hbar)\Phi = 2\pi\Phi/\Phi_0$ is the phase difference associated with the magnetic field, and φ_i is the phase shift across the junction. For a π junction:

$$\pi + \varphi = 2\pi \Rightarrow \varphi = \pi \Rightarrow \pi = \frac{2\pi\Phi}{\Phi_0} \Rightarrow \Phi = \Phi_0/2.$$



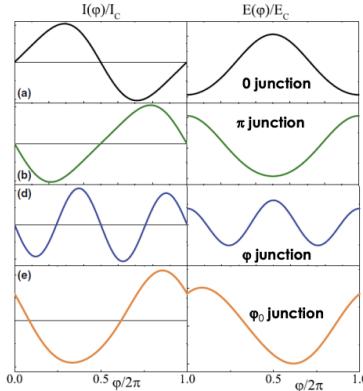
Applications in Superconducting Qubits The ground state of a loop with a single π junction is doubly degenerate with respect to the spontaneous current direction. This property makes π junctions promising candidates for superconducting flux qubits. The free energy of such a loop as a function of the magnetic flux exhibits a double-well structure, enabling the realization of qubits with well-defined states.

2.11.5 More on the SFS Junctions

A ϕ_0 junction is a Josephson junction that exhibits an arbitrary finite phase shift ϕ_0 in the current-phase relation (CPR):

$$I = I_c \sin(\varphi + \phi_0).$$

The energy profile of such a junction shows a minimum at $\varphi = \phi_0 \neq 0, \pi$. This phase shift can arise due to various mechanisms, such as spin-orbit coupling or the presence of magnetic fields in systems with broken time-reversal symmetry. ϕ_0 junctions are of interest for applications in superconducting circuits and quantum computing.



High Harmonics in SFS Junctions

In SFS junctions with transparent boundaries between the superconducting (S) and ferromagnetic (F) layers, the current through the structure can approach the depairing current of the suppressed superconducting regions. This process leads to the appearance of high harmonics in the CPR:

$$I = \sum_{k=1}^{\infty} I_k \sin(k\varphi).$$

The presence of high harmonics significantly modifies the junction's dynamics and can lead to novel effects, such as the coexistence of multiple stable states.

The ϕ Junction

One of the most promising types of SFS junctions is the ϕ junction, which can be realized with a CPR containing two harmonics:

$$I = I_{c1} \sin \varphi + I_{c2} \sin(2\varphi),$$

where $I_{c2} < -|I_{c1}|/2$. The distinctive feature of the ϕ junction is that its energy profile has two degenerate ground states at $\varphi = \pm\varphi'$, where $0 < \varphi' < \pi$. The energy minima correspond to the two stable states of the junction, making it a candidate for use in superconducting qubits.

Critical Currents of the ϕ Junction The ϕ junction has two critical currents, corresponding to the escape of the Josephson phase from the $+\varphi'$ and $-\varphi'$ wells. These critical currents have different amplitudes, reflecting the asymmetry introduced by the second harmonic in the CPR. The ability to tune the CPR and the energy landscape makes ϕ junctions highly versatile for applications in superconducting electronics and quantum information processing.

2.11.6 The π -ring Tricrystal Experiment with d-wave Superconductors

In the case of high-temperature superconductors (HTS) with an anisotropic d-wave order parameter, the intrinsic phase shift of the superconducting wavefunction can lead to π phase shifts in particular configurations. These π shifts result in the presence of spontaneous supercurrents and vortices with fractional flux quanta.

Phase Shift and Supercurrent Orientation The sign of the supercurrent in a junction containing a d-wave superconductor depends on the relative orientation of the crystallographic axes with respect to the junction interface. The pair tunneling current between two superconducting grains i and j is given by:

$$I_{ij} \propto \cos(2\theta_i) \cos(2\theta_j) \sin(\varphi_{ij}),$$

where θ_i and θ_j are the angles of the crystallographic axes of the superconductors relative to the junction interface, and φ_{ij} is the phase difference across the junction. Certain combinations of θ_i and θ_j result in a negative supercurrent, which can be interpreted as a π phase shift at the junction.

Tricrystal Geometry and Half-Integer Flux Quantization In a π -ring tricrystal experiment, three superconducting grains are arranged such that the relative orientations of their crystallographic axes lead to a π phase shift at one or more of the junctions. This configuration creates a closed superconducting loop with a net phase shift of π , resulting in the spontaneous generation of a supercurrent. The fluxoid quantization condition then requires the magnetic flux

through the loop to be quantized in half-integer multiples of the flux quantum Φ_0 :

$$\Phi = \left(n + \frac{1}{2} \right) \Phi_0, \quad n \in \mathbb{Z}.$$

Observation with Scanning SQUID Microscopy High-resolution Scanning SQUID Microscopy has been used to directly observe the half-integer flux quantization in π -ring tricrystal experiments. The presence of vortices with fractional flux quanta provides compelling evidence for the d-wave symmetry of the order parameter in HTS. These experiments highlight the interplay between the anisotropic superconducting order parameter and the geometry of the junctions, leading to novel quantum phenomena.

2.11.7 TODO: Applications of π -JJs: π -SQUIDs and Beyond

Self-Biasing π -SQUIDs The π -SQUID is a superconducting loop incorporating a π -junction, which introduces a built-in fractional phase bias approaching π for large-inductance loops. This self-biasing property eliminates the need for external connections, simplifying design and operation. The absence of external bias reduces size, enhances tolerance to random parameter variations, and minimizes magnetic noise, making π -SQUIDs smaller and more robust than their all-niobium counterparts.

High-temperature superconducting π -junctions are emerging as key components in phase qubits. The permanent phase shift introduced by the π -junction minimizes dephasing interactions with the environment, enhancing qubit coherence. This property is particularly advantageous for scalable quantum computing architectures, where environmental noise is a critical challenge.

Charge and Spin Transport in SFS Junctions There is a continuously growing interest in charge and spin transport in contacts between superconductors and ferromagnets (F). The first experimental observation of supercurrents in SFS junctions and the crossover from 0- to π -state in Nb/Cu_xNi_{1-x}/Nb Josephson junctions was reported by Ryazanov et al. in 2000. A variety of systems exhibiting π -states includes planar SFS proximity effect structures, tunnel junctions with magnetic insulators or magnetically active interfaces, and structures with barriers containing more than one magnetic layer. These systems provide a rich platform for exploring the interplay between superconductivity and magnetism, with potential applications in spintronics and quantum technologies.

Flip-Flopping Fractional Flux Quanta The introduction of additional phase-shifting elements in superconducting loops leads to a twofold degenerate state, which would otherwise require the application of a magnetic flux of exactly $\frac{1}{2}\Phi_0$. This bistable state functions as a flip-flop, a crucial universal circuit

element capable of dividing digital pulse trains by two and acting as a binary storage element. To compensate for this built-in phase shift, a spontaneous circulating current flows in the ring in either the clockwise or counterclockwise direction. The magnetic flux associated with this persistent circulating current is a fraction of a flux quantum, growing asymptotically to a half flux quantum in the large inductance limit. The polarity of this spontaneous flux can be used to store information.

Toggle Flip-Flop (TFF) The TFF toggles between two stable states when a single flux quantum (SFQ) pulse is applied. The SFQ pulses are transferred via a Josephson transmission line (JTL) to the input junction J1 of the flip-flop. Presume that the flip-flop is in the state with a clockwise current in the storage loop. When an SFQ pulse enters through J1, the associated currents give rise to a switching of J3 and J4, causing the flux state in the storage inductor L_{store} to reverse. If a second pulse arrives at the input, it causes a switching of J2 and J5, which toggles the flip-flop back to its original state.

The TFF is typically made of Josephson junctions (JJs), inductors to store magnetic flux, and a bias current source to place the JJ near its switching threshold. When a current or trigger input slightly exceeds the critical current of a JJ, the JJ switches to the voltage state for an extremely brief time (1–2 ps). This creates a voltage pulse whose time integral equals exactly one flux quantum. The inductive loop then returns the JJ to the superconducting state, completing the cycle. This mechanism is fundamental for superconducting digital circuits.

Applications in Superconducting Digital and Quantum Circuits SFS π -shifters are integral to superconducting digital and quantum circuits. Controllable $0-\pi$ Josephson junctions containing ferromagnetic spin valves, which are more complex than simple two-junction π -SQUIDs with additional junctions and drive and readout circuitry, enable advanced circuit designs. In quantum computing, incorporating π -junctions in flux qubit loops addresses the magnetic flux uniformity problem. Large-scale integration of flux qubits is challenging due to the need for external half-flux bias for optimal operation. Realistic variations in qubit loop areas make it practically impossible to bias a large number of qubits on the same chip uniformly. By incorporating π -junctions, the optimal operation can be achieved at zero external flux bias, solving this issue. This approach also reduces magnetic noise, as finite applied fields inherently have fluctuations that act as noise on the qubits, thereby enhancing qubit coherence times and scalability for large-scale integration.

Future Directions The versatility of π -junctions in creating self-biasing circuits, bistable memory elements, and noise-resilient qubits highlights their potential in next-generation superconducting technologies. Continued research into controllable π -junctions and their integration into complex circuits will further expand their applications in quantum computing and digital electronics.

2.11.8 Serial Junctions

Serial junctions with multiple layers connected through thin superconducting buffers have recently gained significant attention. These structures allow for the combination of the mutual properties of the interlayers while preserving the critical current of the individual junctions.

SFsFS and SIsFS Structures

- **SFsFS Devices:** These operate as pseudo spin-valve structures, where the mutual magnetization of the ferromagnetic (F) layers determines the effective exchange field and the critical current of the device. The thin superconducting layer (s) acts as a buffer, enabling the coupling between the F-layers.
- **SIsFS Junctions:** These provide high-performance tunnel SIS junctions with the added capability of switching between 0 and π states. The thin superconducting layer (s) enhances the tunability of the junction's properties.

Complex Current-Phase Relation (CPR) Systems of serially connected junctions exhibit a complex, hysteretic current-phase relation (CPR) with possible multiple branches. This behavior arises from the interplay between the individual junctions and the coupling layers, leading to rich and tunable dynamics.

Applications Serial junctions are promising for advanced superconducting devices, including:

- Tunable qubits for quantum computing.
- High-performance superconducting circuits with controllable $0-\pi$ transitions.
- Spintronic devices leveraging the pseudo spin-valve behavior of SFsFS structures.

2.11.9 Hybrid Josephson Junctions (Hybrid JJs)

Hybrid Josephson Junctions (JJs) represent a class of superconducting devices where the barrier between the superconducting electrodes is made of materials with special functionalities, such as semiconductors, graphene, or topological insulators. These devices combine the properties of superconductors with those of the barrier material, enabling novel functionalities and tunability.

Superconducting Hybrid Devices A hybrid JJ typically consists of a coplanar structure where a barrier material, such as a semiconductor (e.g., $\text{In}_x\text{Ga}_{1-x}\text{As}$, InAs), is deposited and treated before the patterning of the superconducting banks. The induced coherence length ξ_{sm} in the semiconducting barrier depends on the carrier density, which is influenced by the diffusion constant. For high-transmittance S-Sm interfaces, ξ_{sm} can be tuned by applying a gate voltage.

Gate Voltage Control The application of a gate voltage strongly modifies the current-voltage (I - V) characteristics and the amplitude of the hysteresis in hybrid JJs. This tunability has made these devices promising candidates for the realization of a superconducting Josephson field-effect transistor (Jo-FET), a long-standing goal in the field.

Emerging Materials and Layouts Recent advancements in the fabrication and handling of quasi-2D materials, such as graphene, topological insulators, nanotubes, and nanowires, have enabled the development of new families of hybrid JJs. In a standard configuration, a nanowire or a 2D flake is placed on a substrate, and the sample is patterned using e-beam lithography to define the regions where superconducting electrodes will be deposited.

Advantages of Graphene and 2D Materials Graphene and other 2D conducting materials offer the unique advantage of tunability over a wide range via gate voltage. This tunability arises from changes in the chemical potential μ , electron concentration, sheet resistance, and density of states. These properties make graphene-based hybrid JJs highly versatile for various applications.

Table 2.8: Examples of Materials and Layouts for Hybrid JJs

Barrier Material	Superconducting Electrodes	Key Features
Semiconductors (e.g., InAs)	Nb, Al	Tunable ξ_{sm} via gate voltage
Graphene	Nb, Pb	Wide tunability of μ and I_c
Topological Insulators	Nb, Al	Robust edge states, Majorana modes
Nanowires (e.g., InSb)	NbTiN, Al	Strong spin-orbit coupling, gate control

2.11.10 The Resistively Shunted Junction (RSJ) Model

The Resistively Shunted Junction (RSJ) model provides a framework to describe the dynamics of a Josephson junction by considering it as an ideal non-linear element in parallel with a resistance R and a capacitance C . The total current I through the junction is given by:

$$I = I_s + I_n + I_d = I_c \sin \varphi + \frac{V}{R} + C \frac{dV}{dt}, \quad (2.19)$$

where I_s is the supercurrent, I_n is the quasiparticle current, and I_d is the displacement current. Using the Josephson relations, this can be rewritten as:

$$I = I_c \sin \varphi + \frac{\Phi_0}{2\pi R_n} \frac{\partial \varphi}{\partial t} + C \frac{\Phi_0}{2\pi} \frac{\partial^2 \varphi}{\partial t^2}, \quad (2.20)$$

where R_n is the normal state resistance. The dynamics of the phase φ is then governed by:

$$\frac{\Phi_0}{2\pi} C \frac{\partial^2 \varphi}{\partial t^2} + \frac{\Phi_0}{2\pi R_n} \frac{\partial \varphi}{\partial t} + \frac{\partial U}{\partial \varphi} = 0, \quad (2.21)$$

with the potential energy U given by:

$$U = -\frac{\Phi_0}{2\pi} (I_c \cos \varphi + I \varphi). \quad (2.22)$$

Tilted Washboard Potential

The dynamics of the junction can be visualized as a particle moving in a "tilted washboard" potential $U(\varphi)$. The term involving C represents the mass of the particle, the $1/R_n$ term represents damping (friction), and the tilt of the washboard is proportional to the bias current I . The equation of motion is analogous to:

$$m \frac{d^2 x}{dt^2} + \eta \frac{dx}{dt} + \frac{\partial U}{\partial x} = 0, \quad (2.23)$$

where m is the mass, η is the damping coefficient, and x corresponds to the phase φ .

Thermal Activation and Macroscopic Quantum Tunneling

For $I < I_c$, the particle is confined to one of the potential wells. However, due to thermal activation (TA) or macroscopic quantum tunneling (MQT), the junction may switch to the finite voltage state even for $I < I_c$. The relative importance of these escape processes depends on the temperature of the system.

McCumber-Stewart Parameter

The McCumber-Stewart parameter β_c quantifies the damping in the junction:

$$\beta_c = \frac{2\pi I_c R_n^2 C}{\Phi_0}. \quad (2.24)$$

- $\beta_c > 1$: Underdamped regime. High capacitance, tunnel-like junctions, characterized by hysteretic behavior and switching currents.
- $\beta_c < 1$: Overdamped regime. Non-hysteretic behavior.

The presence of a second harmonic in the current-phase relation (e.g., in SFS junctions) modifies the potential $U(\varphi)$, introducing secondary wells and altering the dynamics of the particle.

Temperature Dependence of I_c

In the tunnel limit, the temperature dependence of the critical current I_c is described by the Ambegaokar-Baratoff model:

$$I_c(T) = I_c(0) \frac{\Delta(T)}{\Delta(0)} \tanh\left(\frac{\Delta(T)}{2k_B T}\right), \quad (2.25)$$

where $\Delta(T)$ is the superconducting energy gap, k_B is the Boltzmann constant, and T is the temperature. At $T = 0$, I_c is proportional to the energy gap $\Delta(0)$.

Field Dependence of I_c

The dependence of the critical current I_c on the magnetic induction B (or external magnetic flux Φ) for an ideal rectangular junction geometry follows a Fraunhofer-like pattern:

$$I_c(B) = I_c(0) \left| \frac{\sin(\pi\Phi/\Phi_0)}{\pi\Phi/\Phi_0} \right|, \quad (2.26)$$

where Φ_0 is the magnetic flux quantum. The critical current exhibits a central maximum at $B = 0$ and decreases with oscillatory minima at integer multiples of Φ_0 .

Josephson Penetration Depth λ_J

The Josephson penetration depth λ_J measures the distance over which Josephson currents are confined:

$$\lambda_J = \sqrt{\frac{\Phi_0}{2\pi\mu_0 d J_c}}, \quad (2.27)$$

where μ_0 is the permeability of free space, d is the effective barrier thickness, and J_c is the critical current density. For wide junctions ($W > 10\lambda_J$), the Fraunhofer pattern is smeared out, and the interferometric behavior is lost.

Spatial Distribution of the Current

The spatial distribution of the critical current density $J_c(x)$ across the junction barrier is influenced by:

- The geometry of the junction.
- The nature of the electrodes and the barrier.
- Inhomogeneities, Meissner screening, and vortex formation.

These effects lead to distinctive magnetic fingerprints in the $I_c(B)$ dependence and I - V curves. For example, junctions with d-wave electrodes or ferromagnetic barriers exhibit unique magnetic interference patterns.

Experimental Observations

For d-wave corner junctions (e.g., Pb/YBCO), the $I_c(B)$ dependence shows symmetric maxima at finite magnetic fields. The transition from a Fraunhofer pattern to this behavior depends on the ratio γ between the π and 0 components of the junction.

image

2.11.11 The Evolution of Josephson Junction Technology

The development of Josephson junctions (JJs) has undergone significant advancements since their inception in the 1960s. Below is a summary of the key milestones and materials used in their realization:

Early Developments: Soft Superconductors and Thermal Oxidation

The first SIS tunnel JJs were fabricated using soft superconductors such as Sn, In, and Pb, with thermal oxidation employed to create the insulating barrier. This marked the beginning of Josephson technology.

Lead-Alloy Technology and Sputter-Oxidation

The introduction of lead-alloy technology brought the first self-limiting sputter-oxidation process, which improved the reproducibility and performance of JJs.

Transition to Rigid Superconductors: Nb and Artificial Barriers

The use of rigid superconductors like Nb revolutionized JJ technology. Artificial barriers replaced native Nb oxide barriers, addressing issues such as defects, conducting sub-oxides, and magnetism. Al emerged as the ideal solution, forming a natural, self-limiting, high-quality insulating oxide.

Modern Multilayer SIS Thin-Film Junctions

Modern Josephson electronics, including high-speed digital logic circuits, qubits, SQUIDs, and photon detectors, utilize multilayer SIS thin-film junctions. Examples include:

- Nb/Al-Al₂O₃/Nb
- Nb/Al-Al₂O₃/Al/Nb

The Al overlayer approach has been a key advancement, enabling thermally and mechanically stable tunnel junctions with Nb.

Other Rigid Superconductors

Other rigid superconductors, such as NbN, Nb₃Sn, V₃Si, and Nb₃Ge, have also been explored. These materials require artificial barriers for JJ fabrication.

2.11.12 High-Temperature Superconducting Junctions (HTS JJs)

$\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (YBCO): The Most Studied HTS Compound

YBCO has been extensively studied for JJ fabrication. Most YBCO junctions have their supercurrent flow parallel to the substrate, where the coherence length is more favorable for junction performance. This topology facilitates the production of sub-micron-area junctions.

Challenges and Recent Progress

The lack of a reliable HTS trilayer junction process has hindered the development of applications requiring multi-layer integrated circuitry, such as digital circuits. However, recent progress has been made in this area.

Unique Features of HTS JJs

HTS JJs exhibit unique features, including:

- Supercurrent between phase-coherent electrodes up to ~ 100 K.
- Unexpected energy and length scales, dimensionality, strong correlations, and d-wave order parameter symmetry.

These properties have inspired the development of Josephson-based devices using unconventional superconductors discovered after HTS.

2.11.13 Applications of Josephson Junctions

Voltage Metrology

Josephson junctions are widely used in voltage metrology, providing highly accurate voltage standards.

Nano-JJs and Nano-SQUIDs

Nano-JJs and nano-SQUIDs have been developed for applications requiring ultra-sensitive magnetic field detection and nanoscale precision.

Future Directions

The continued development of Josephson junction technology, including the integration of HTS and unconventional superconductors, holds promise for advancing quantum computing, metrology, and other cutting-edge applications.