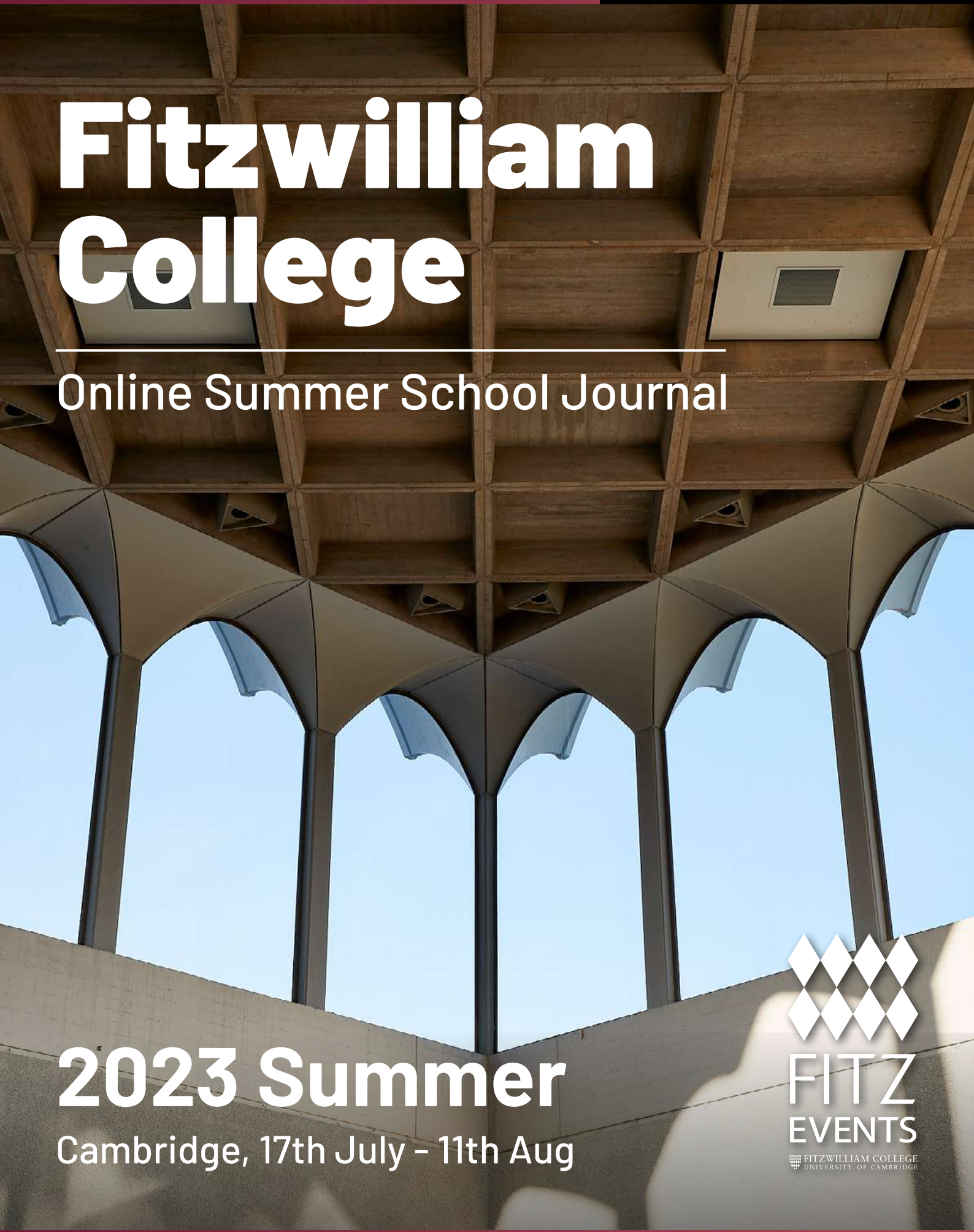
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

阿思丹 | ASDAN China



Fitzwilliam College

Online Summer School Journal

2023 Summer

Cambridge, 17th July - 11th Aug

FITZ EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ EVENTS
FITZWILLIAM COLLEGE UNIVERSITY OF CAMBRIDGE

About the Journal

This journal is a semi-annual journal publishing the best essays written by some top-performing students at **the Fitzwilliam College Online Summer School Programme.** The journal will cover nine fields ranging from Engineering, Physics, Computer Science, Biology, Chemistry, Medicine, Neuroscience and Mathematical Finance to Mathematics.

About the Programme

The Fitzwilliam College Online Summer School Programme is officially run by Fitzwilliam College, one of the constituent colleges of the University of Cambridge. All the courses were taught by academics teaching undergraduate students at the University of Cambridge. Fitzwilliam has always been characterised by discussion, debate and creativity of ideas, and full participation should form a positive, rewarding and sustainable part of an academic course. This programme is designed to provide students with a flavour of undergraduate study and an opportunity to explore topics beyond what is covered within the school curriculum.

Starting in 2023, Fitzwilliam College and ASDAN China have entered into a strategic partnership to open the Fitzwilliam College Online Summer School Programme to outstanding high school students in China for the first time.

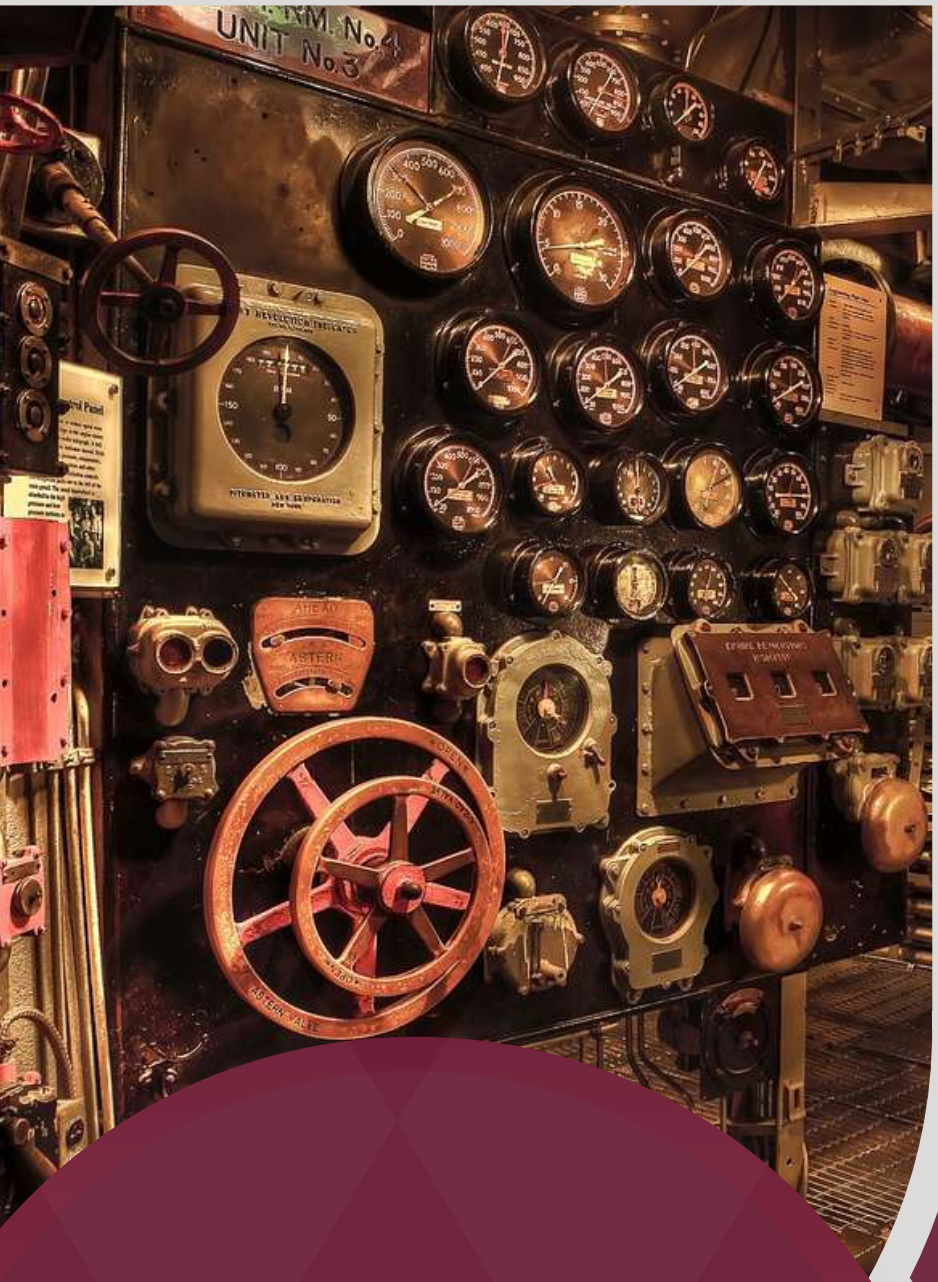
List of Academic Course Instructors

DR ASHRAF ZARKAN
DR JOAO RODRIGUES
MRS.SERENA POVIA
DR VIHANGA MUNASINGHE
DR ALEXANDRA KRUGLIAK
DR ANDREA GIUSTI
DR JOHN FAWCETT
DR VASILEIOS KOTSIDIS
PROFESSOR MATTHEW J. MASON
DR AARON D'SA
DR EMILY LEES
PROFESSOR KOUROSH SAEB-PARSY
DR AUDE RAUSCENT
DR SAEED KAYHANIAN
DR RICHARD DYBALL

Table of Contents

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE



01

Engineering

Fitzwilliam College Online Summer School Journal · 2023 Summer

Full electric airplanes

KEMING GONG (Kimi) & SIQI LUO (Peter) & ANNING HU (Annie)

1. Introduction

Peter

The aviation industry has been witnessing a great shift with the rapid advancement of electric propulsion technology. As concerns about environmental sustainability and carbon emissions grow, there is an increasing interest in developing all-electric and hybrid-electric aircraft. However, the success of these technologies largely depends on overcoming critical challenges, such as improving battery energy density, reducing weight, and extending flight range. This essay explores the current development of electric and hybrid aircraft while suggesting a possible way to improve the range using existing technology through changes in the propulsion system and the airplane's shape.

Annie

Electrification has become one of the most promising methods for a more ecologically sustainable solution for its outstanding characteristic in low carbon emissions, noise pollution and dependent on fossil fuels. There have been various attempts and prototypes of full electric aircrafts. However, due to the low specific energy of lithium-ion batteries, existing vehicles have common characteristic of short range, low maximum take-off mass, high battery mass ratio and seating capacity of usually 1 to 2 passengers.

Kimi

Nowadays, electric aircraft is a topic that is being widely discussed. Large aviation industries such as Airbus already proposed their electric airplane designs. In fact, at 2016, a lightweight solar powered but bizarre looking aircraft, with a wingspan similar to Boeing 747 successfully circumnavigated the earth [3]. Even though this aircraft totally differs from commercial aircrafts, it does provide some useful insights about the possible electric airliners' design. For example, the gigantic wing with solar panels directly on it could be the most feasible idea right now. Meanwhile, the electric motor won't be as powerful as jet engines at the current stage, so it is important to keep the aircraft light weighted. My design will be a short-range commercial aircraft powered by electric propellers, and it can be used in general aviation or as a private aircraft.

2. Design proposed by Peter

Peter

2.1. Propulsion system

Due to the current battery technology, batteries are heavy and have low energy density, limiting the airplane's range. The existing fully electric airplane with the longest range is the Eviation Alice, which can fly up to 650 miles at a cruise speed of 276 mph and accommodate only 9 passengers [1]. However, this performance falls short of the demands for commercial flights. To extend the range without relying on extraordinary advancements in battery technology, a hybrid configuration may be introduced.

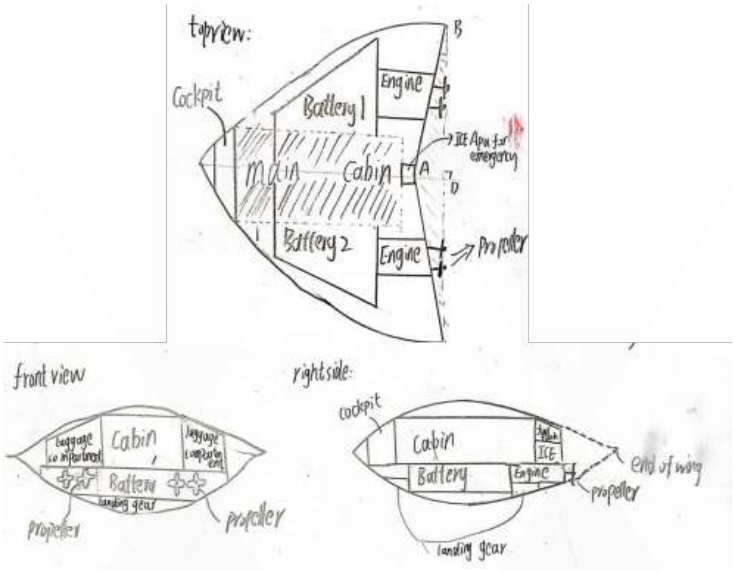
Two electric-driven wingtip propellers can be added to each wing, along with a conventional combustion engine. The conventional airplane uses a fuel cell to provide power, offering a very high energy density and a heating value of 11.9 kWh/kg [2]. However, the chemical-to-mechanical energy conversion in this process has a relatively low efficiency of approximately 35%.

In comparison, the electric-driven propeller may achieve higher efficiency. The direct drive capability of the electric

motor eliminates the complex burning process and minimises chemical energy losses. Moreover, the electric-driven propeller reduces the need for fuel systems, including fuel, fuel tanks, and fuel delivery systems. It also eliminates the requirement for cooling and exhaust systems, significantly reducing weight and creating space for batteries and cargo.

2.2. Shape and design of the airplane

The energy density of existing battery materials is relatively low, with the highest being about 0.4kWh/L, while jet fuel has a high density of 9.6kWh/L. This 24 times difference results in the need for larger space to accommodate the battery and an increase in weight. Improving the energy density of the battery is the most effective way to enhance the battery range, but achieving this in current technology development may be challenging. Therefore, to extend the range of the airplane and make room for the battery, a redesign of the aircraft may be necessary.



To create space for the battery and maintain sufficient aerodynamic performance, I have designed the airplane as shown in the figure above. The airplane has a drop-shaped body in the right side view, which may reduce flow separation and minimise drag. The front part of the airplane serves as the cockpit, while the main body functions as the main cabin. Below the cabin is the space allocated for the battery. The battery is connected to four electric-driven propellers on the rear wings, providing forward thrust for the airplane. At the back of the aircraft, there is a small auxiliary power unit housing

an internal combustion engine that can supply electricity and power in case of emergencies. The shaded region BAC represents the area cut from the semicircle to reduce the weight of the wings while creating a large wingspan to enhance lift.

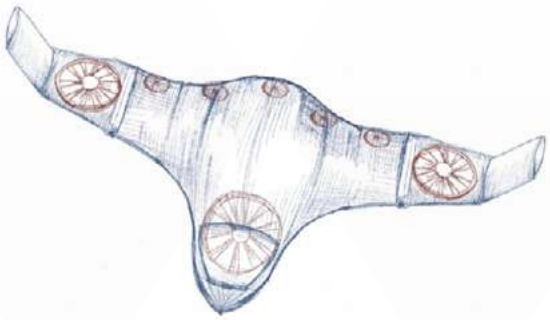
In the front view, the aircraft consists of three layers: cabin, battery, and gear. Since the battery occupies the original cargo and luggage space, the area between the main cabin and the drop- shaped body can be utilised for storage. Furthermore, since electric motors do not require air inlets for combustion gases, the engine can be positioned inside the airplane's body, reducing frontal area and flow separation, thus improving aerodynamic performance.

Additionally, to extend the airplane's range, weight reduction can be achieved through other components. As stated Yu LiMing [4], 'there are four types of secondary power systems on an aircraft: hydraulic, pneumatic, electrical, and mechanical systems for primary propulsion.' [7] By replacing hydraulic actuators with electric actuators, gear-driven hydraulic pumps and fuel pumps with electric pumps, and pneumatic air conditioning compressors with electric compressors, the complex mechanical structures can be simplified into wires and electric motors, leading to weight reduction and an increased aircraft range.

3. Design proposed by Annie

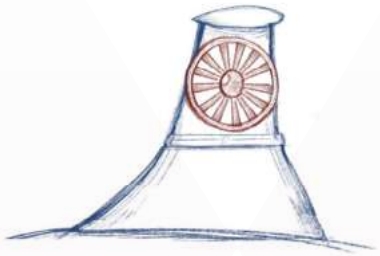
Annie

For electric propulsion systems, the number of propellers is no longer limited to 2. Because, unlike combustion engines, electric motors' performance varies linearly with size [10], we may break them into smaller ones to satisfied different needs respectively. In addition, better efficiency can be achieved as propellers can operate at lower tip speeds hence lower induced drag. In this design, 5 small electric propellers are distributed across tail of the vehicles to provide thrust. They lie between the upper and lower surface of the aircraft. This configuration blend into the vehicle perfectly while it doesn't add to frontal area. Two larger electric motors are embedded in the side wings. Joints of the parts of the wings are designed to be rotatable to provide propulsion in different direction under different circumstances.



Configuration 1: horizontal side wings with embedded rotor

This configuration is applied during the hover state. The two rotors provide vertical thrust. The additional rotor under the vehicle balances the torque and keeps the vehicle from tilting during the process.



(view from above)

Configuration 2: horizontal side wings with covered rotors

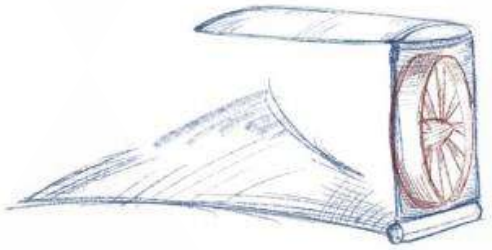
This configuration is also used in the state of taking off, but not vertically. The propeller group along the tail provides thrust and the vehicle takes off by the upward lift.

(view from above)

The two different choices of take-off method also arise from the special characteristic of electric vehicle. For a personal vehicle, there is not always a runway available when needed. Meanwhile, vertical flying is more energy consuming, resulting in a significant increase in battery load. Decisions are based on the surrounding, state of charge of battery, distance of the tour etc

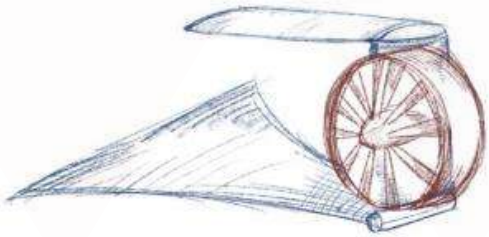
Configuration 3: Vertical side wings with embedded rotor

The Vertical side wings form a c-wing. Side-to-side motion can be achieved by turning on motor on only one side, giving a quicker response to obstacle than turning.



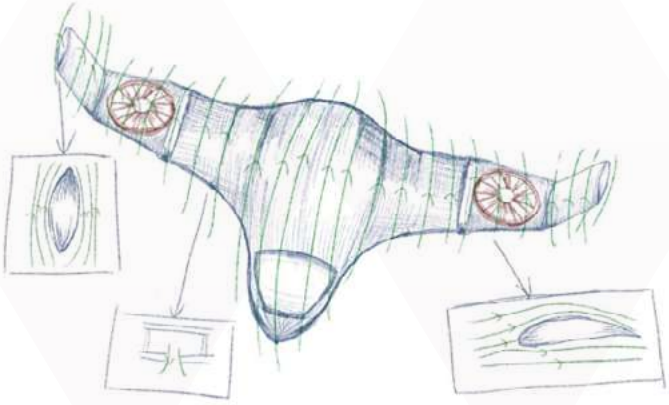
Configuration 4: vertical side wings with rotated rotor

The rotor rotates 90 degrees and act as another propeller providing horizontal thrust. This configuration can be applied when some rotors malfunctions or to accelerate.



Aerodynamic design

A good aerodynamic design significantly improves the vehicle’s energy efficiency, extending flight range and allow for greater load. The overall shape for this design blends the wings into the shape of the body, avoiding sharp edges causing flow separation.



Smaller engines blend into the shape easily and avoid adding to frontal area and increased drag. Teardrop shape, lowest drag coefficient, is applied to the top wing. In any state of the flight, the top wing is always vertical, aiming to reduce the circulation around the wingtip, hence decrease the wingtip vortices. During the cruise state, the side wing is held vertical, forming a C shape for wing configuration. It generates extra thrust on the top wings by using the downwash created by the main wings, it causes a forward induced drag, which is thrust. Additional thrust act like a propeller, but generated by downwash of fluid. This technology lightens the load for batteries.

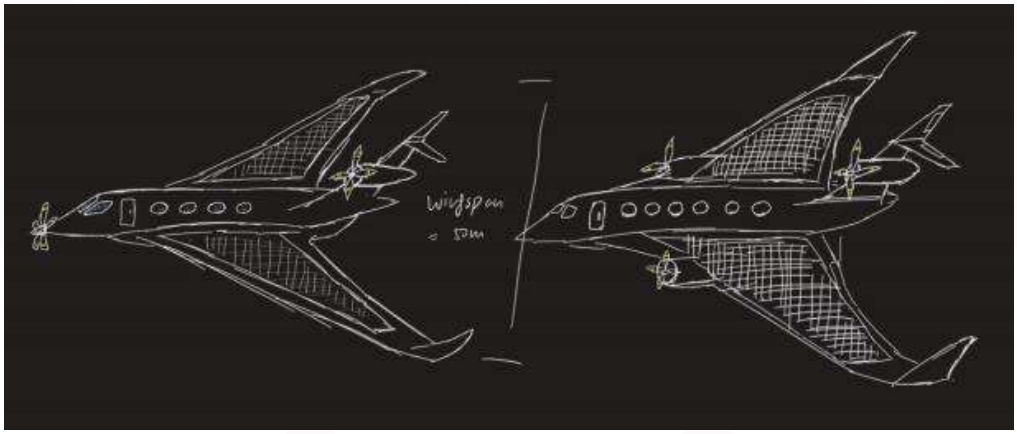
A boundary layer control system is added at the wings . It sucks the air with low speed near the surface continuously. With a thinner boundary layer, the air flow are less likely to detach from the surface, delaying flow separation. However, the traditional suction is heavy and complex to manufacture, not suitable for an electric vehicle. This design adopt a simplified version proposed by Horstmann and Schröder using simply a double skin structure.

4. Desing proposed by Kimi

Kimi

The powertrain of my design is mainly divided into the following parts: electric propellers (fans), solar panels and batteries. During flight, the solar panels absorb energy from sunlight, then the energy will be delivered to the fans and thus create a thrust.

The aircraft has a 15 meters long fuselage, but its wingspan must be long enough to fit sufficient amount of solar panels that could generate enough power. Nowadays, Canadian Solar's newest product TOPBiHiKu7: a 3.106 square meters solar panel with 22% efficiency (converting solar energy into electricity) can generate 695 W power [5]. In other words, this solar panel can produce 223.76 W per every squared meter. However, a newest technology can further increase the power of solar panels by a significant amount. Scientists from University of California stated that by spraying a layer of multi-scale surface (which is composed by particles from 10 nanometers to 10 micrometers large), those particles with variant sizes can increase the surface area that are used to absorb sunlight, therefore the efficiency of converting energy can be incremented to 90% [6]. Therefore, this variant of solar panel can generate more than 900 W of energy per meter squared. The energy supply that comes from the solar panel will be directly used to power the propellers. However, another propeller is located at the end of the aircraft. It is powered by 3000 kg of 700Wh/kg rechargeable lithium-ion battery (which is the battery with the highest energy density) [7], which can provide a consistent 2100 kW energy output for 1 hour. The aim of this engine is to provide additional thrust during the takeoff and ascending stage, since the horizontal drag force will increase massively as a result of lift "leaning backward", and by decomposing lift into horizontal and vertical forces, a horizontal drag force that equals to the magnitude of lift times sin θ (where θ means banking angle) will be produced. Based on my estimation, more than 50 kN of additional drag will be produced when the aircraft ascends. The aim of this engine located at the rear end is to produce enough thrust to deal with the drag force, and it will be shut down when the airplane stopped ascending. The battery can be recharged by the solar panel during flight, this is because as the height increases, density of air decreases, and the horizontal air resistance reduces (since air resistance is directly proportional to the density of air), therefore, the engines won't need to work with full power to provide enough thrust, instead, energy that comes from the solar panels will be used to charge the battery system.



The wing of the aircraft is extraordinarily large (approximately 200 square meters), since massive solar panels are used to generate energy. The most important advantage of a large wing is that it allows the aircraft to cruise at slow speed without stalling (this is because a larger wing can generate more lift) [7]. Therefore, the propeller doesn't need to be as powerful as traditional petrol engines to make the plane fly. Additionally, even though a large wing isn't as maneuverable as short wings, but it can provide the best performance at low speed (which is suitable for electric aircrafts) [8]. However, a larger wing might bring additional weight to the aircraft, so it is important to use lightweight materials as much as possible.

5. Conclusions

Peter

In conclusion, electric and hybrid aircraft hold promising potential for greener aviation. Despite current challenges, advancements in technology offer hope for improved range and efficiency. By embracing innovation and collaboration, we can pave the way for a sustainable and eco-friendly future in air travel.

Annie

In conclusion, a fully electric design provide more freedom when no longer confined to two main fuel engines. This design of electric vehicle offers better control, improving the flexibility and safety for the tour and leave more room for decision. Breaking down propellers allows more possibilities of shape and configuration of the vehicle. A blended-wings and C-wing configuration is adopted to improve aerodynamic. Safety is ensured due to redundancy offered by multiple propellers. If one malfunctions, the vehicle continues to fly.

However, this design has obvious limitation. The additional propeller under the vehicle is exposed to foreign objects which may damage the propeller. Although better efficiency and aerodynamic performance is achieved, in addition to extra thrust generated by the c-wing configuration, range anxiety still exists. The most promising solution is to improve the performance of batteries. Although the current battery used have 50 times lower specific energy compared with fuels a better efficiency of electric motors make up for this. It is predicted by Elon Musk that a specific energy of 400 kWh/kg will be able to catch up with kerosene around 2024.

As technology continues to advance, the future of fully electric aircrafts gets closer within reach. They hold the potential to revolutionize the way we travel and shape the future of air transportation. By keep innovating the fully electric aircrafts, we leave a lasting legacy for generations to come.

Kimi

Overall, my fully electrical totally differs from the traditional layout of a passenger plane. The large wing can provide sufficient power (by installing solar panels on it). The power comes from the solar panels will be then allocated to the propellers adjacent to the wings, and an additional propeller (that is used to give additional power during ascend) is located at the rear of the aircraft, and it is powered by efficient lithium-ion battery. The wing can also produce lift effectively so that the aircraft can cruise at low speed. Its major concerns are reducing weight to 30000 kg maximum; minimizing drag coefficient and maximizing the lift coefficient of the wing; and enhancing the capacity of the battery so that it can fly without sunlight.

6. References

1. Future flight (2022). Available from:https://www.futureflight.aero/aircraft-program/alice

2. Julian Hoelzon (2018), Conceptual Design of Operation Strategies for Hybrid Electric Aircraft, available from:https://www.mdpi.com/1996-1073/11/1/217

3. Prisco, J. (2022). This solar-powered plane could stay in the air for months. Available from: https://edition.cnn.com/travel/article/skydweller-solar-powered-plane-solar-impulse-climate-scn-spc-intl/index.html[Accessed 25th July 2023].

4. Yu Li Ming (1999). Technological Improvements and Development Status of All-Electric Aircraft, 1999, available from:http://www.cqvip.com/qk/97907x/199903/4816774.html

5. Santos, B. (2023). Canadian Solar launches 700 W bifacial TOPCon solar modules. Available from:https://www.pv-

magazine.com/2023/06/14/canadian-solar-unveils-700-w-bifacial-topcon- solar-modules/[Accessed 25th July 2023].

6. Stone, K. (2014). Solar Power Material 90 Percent Efficient. Science Connected Magazine. Available from:https://magazine.scienceconnected.org/2014/10/solar-power-material-90- percent-efficient/[Accessed 26 July 2023].

7. Dumé, I. (2023). Lithium-ion batteries break energy density record. Available from: https://physicsworld.com/a/lithium-ion-batteries-break-energy-density-record/#:~:text=Researchers%20have%20succeeded%20in%20making [Accessed 26 July 2023].

8. Loftin, L. (1985). Quest for performance: the evolution of modern aircraft. Washington DC, U.

S. Government Printing Office.

9. Caron, J. (n.d.). The Wing's the thing: How Wing Design Affects Your Ride, and How to Preflight for Safety. Available from: https://iflyamerica.org/safety_wing_design.asp#:~:text=It's%20more%20aerodynamically%20ef ficient%2C%20generates,forgiving%20of%20improper%20pilot%20techniques. [Accessed 26 July 2023].

10. M.D. Moore, B. Fredericks (2014). Misconceptions of electric aircraft and their emerging aviation markets

The Future of Internal Combustion Engines: Towards Sustainable Solutions

SHUNHANG ZHOU (Brandon)

1. Introduction

Due to the worldwide requirement to minimize the use of fossil fuels and prevent climate change, the future of internal combustion engines (ICE) is in a precarious position. The United Nation's Sustainable Development Goal number 7, affordable and clean energy and number 13, climate action is the most prominent when it comes to the future of automobiles, as sustainable development has been the main goal across the board. It is therefore of particular significance to determine the possibilities of using ICE technology in a sustainable way after the transition to alternative, renewable fuels. This research project will examine the prospective trajectory of ICE technology. The paper covers the current state of ICE technology, the difficulties it faces, and proposed solutions to make it more ecologically friendly.

For more than a century, internal combustion engines have provided power for a wide range of vehicles, including cars, trucks, ships, and airplanes. Fossil fuels are widely used, but this has prompted questions about how they affect the environment, particularly on how they affect emissions of greenhouse gases and global warming. It is critical to assess the viability of continuous use of ICE technology with the implementation of sustainable fuels as we progress towards a more sustainable future.

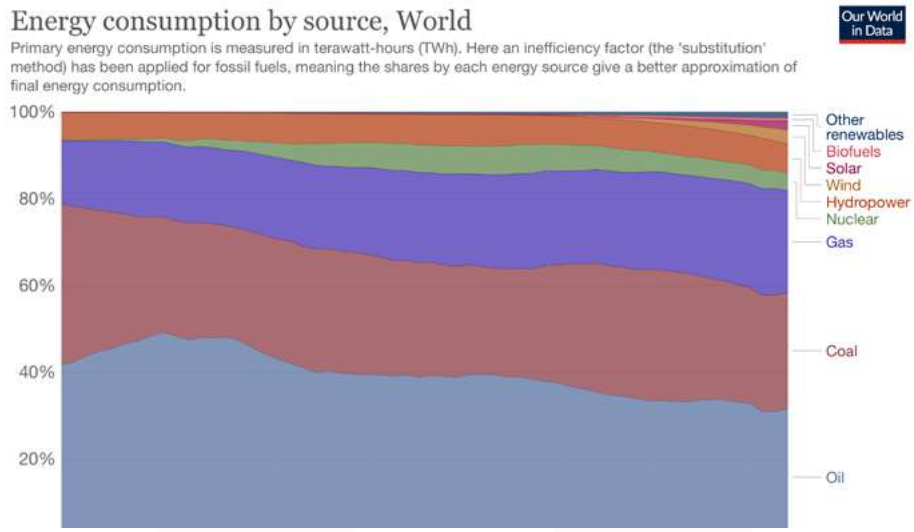


Figure 1. World energy consumption by source (millions of petroleum equivalent) in the last 25 years [1].

2. Current State of Internal Combustion Engines

Internal combustion engines produce mechanical power by burning fossil fuels (e.g. gasoline or diesel) [2]. Despite improvements in efficiency and technology, the fundamental idea has not changed. However, the effects of their operations on the environment have drawn increasing attention. The primary concern with conventional ICEs is their dependency on fossil fuels, which causes the release of pollutants including nitrogen oxides (NOx) and particulate matter (PM) as well as greenhouse gases such as carbon dioxide (CO2). Approximately 3,000 out of 13,000 million tons of oil equivalent are produced annually. ICEs that run on fossil fuel oil accounts for about 25% of the world's primary energy consumption and 10% of its greenhouse gas emissions (Figures 1 and 2). These emissions have a severe negative impact on health, air pollution, and global warming. These issues must be resolved in order to guarantee a sustainable future for the further development of human kind.

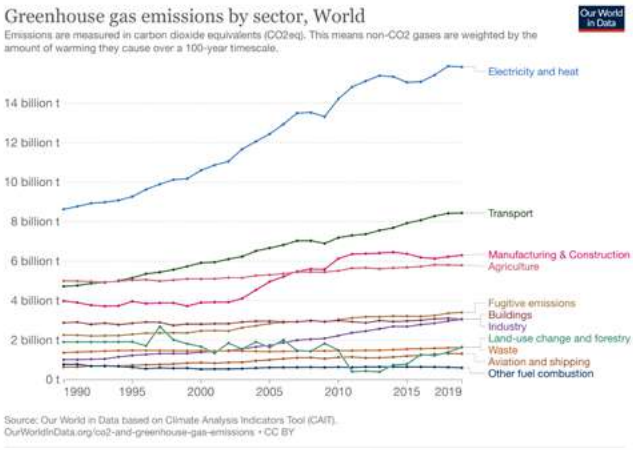


Figure 2. Global warming potential (GWP) in CO2 equivalent tons by sector [3].

3. Transition to Sustainable Fuels

Researchers have worked to increase the fuel efficiency of internal combustion engines for decades. Concerns about climate change gained traction, while ensuring the best possible use of scarce fuel supplies for current and future generations. Shifting to sustainable fuels, commonly referred to as biofuels, is one way to extend the usage of ICE technology. These fuels can be produced from organic materials including biomass, food waste, algae, or synthetic sources. There are various benefits when comparing sustainable fuels to traditional fossil fuels counterparts [4]. Since the carbon released during combustion is a natural component of the carbon cycle, they can dramatically lower the net greenhouse gas emissions. Moreover, they only require minor adjustments to the current ICE infrastructure.

4. Future Solutions

Conventional biofuel sources were cereals such as corn and sugar cane, but they were severely limited by food rivalry. The second-generation biofuels that have garnered interest are the "next- generation" materials that do not compete with food. Next-generation biofuels made from waste materials like food andother organic materials are now being researched, developed, and commercialized [5].

Microalgae, are described as "promising strains" over the medium to long term. Its abundance of components, including fatty acids and hydrocarbons, which are valuable as fuel, is one of the factors that make it a promising stock. During the cultivation phase, wax esters (oils and fats) develop in the body of the euglena. Fatty acid methyl esters (FAME), often known as crude oil, are created when the material is dried to extract the wax esters and hydrogen is added to take the place of any additional oxygen. This FAME has been broken down and refined by Euglena to be used as fuel (figure 3). It is more productive and grows more rapidly than cellulosic biomass [6].

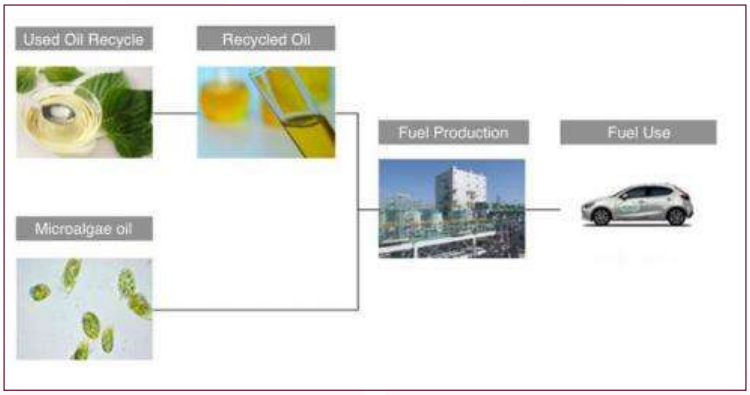


Figure 3. Next-generation biodiesel fuel raw material production (microalgae oil and recycled oil made from used edibleoil), supply, and utilization [6].

5. Challenges and Limitations

While sustainable fuels have the potential to revolutionize ICE vehicles, there are a number of obstacles in the way of their widespread implementation. First, consumer acceptance may be hampered by present production and distribution costs that are greater than those of conventional fuels. Secondly, the overall efficiency of ICEs may be affected since sustainable fuels may have a lower energy density than conventional fossil fuels (figure 4).

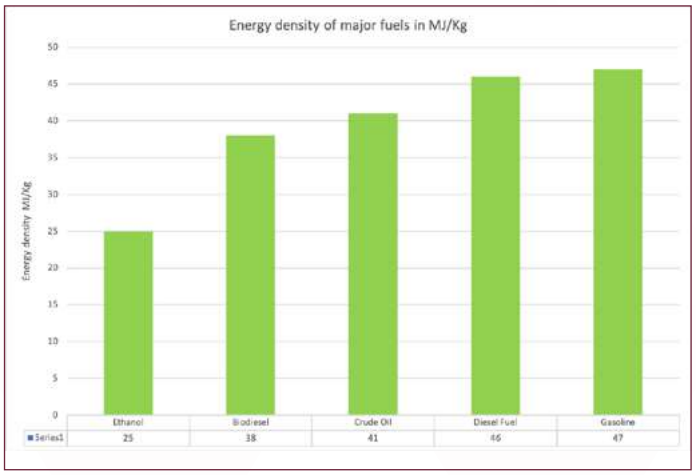


Figure 4. Energy density of major fuels in MJ/Kg [9]

However, the main challenge is still cost. Although it is claimed that biomass from microalgae like Euglena will be far more productive than biomass from terrestrial plants, a significant amount of energy must still be expended during the cultivation stage to replace water, supply carbon as a nutrition source, and add nutrients. Some of the algae production companies announced in 2020, aimed to build a commercial-scale plant in 2025 and increase the production of "biofuel" by 2000 times. Production volume will reach 250,000 kL by 2025, further reducing manufacturing costs. The target price at that time is also concrete, with 0.72 USD per litre, which is 1/100 of the current price. Allowing a carbon neutral future at a low cost without the concern of carbon emissions (figure 5) [6].

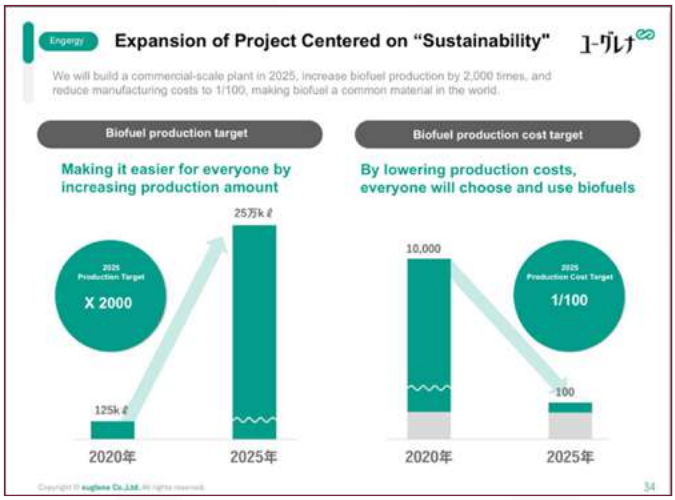


Figure 5. euglena Co., Ltd.'s Goal for construction of a commercial-scale plant for 2025 (future) [6].

6. Conclusions

The ability of internal combustion engines to adopt environmentally friendly solutions will determine their future. The use of alternative fuels can increase the practical application of conventional fossil fuel-based ICEs while contributing to a more sustainable transportation sector. In order to encourage the development, application, and adoption of sustainable fuels, governments, businesses, and researchers must collaborate closely for their

common objective.

ICE technology can contribute to a low-carbon future by switching to renewable fuels and improving engine designs. However, it is crucial to recognize that ICEs alone cannot address all environmental issues. To create a genuinely sustainable transportation system, an integrated approach that incorporates electrification and environmentally friendly urban planning should be further explored.

7. References

1. "Energy Consumption by Source." Our World in Data, ourworldindata.org/grapher/energy-consumption-by-source-and-country.

2. Wikipedia Contributors. Internal combustion engine [Internet]. Wikipedia. Wikimedia Foundation; 2019. Available from:https://en.wikipedia.org/wiki/Internal_combustion_engine

3. "Greenhouse Gas Emissions by Sector." Our World in Data, ourworldindata.org/grapher/ghg-emissions-by-sector.4. SCHAFER A, HEYWOOD J, WEISS M. Future fuel cell and internal combustion engine automobile technologies: A 25-year life cycle and fleet impact assessment. Energy. 2006 Sep;31(12):2064–87.

5. Mekonnen D, Bryan E, Alemu T, Ringler C. Food versus fuel: examining tradeoffs in the allocation of biomass energy sources to domestic and productive uses in Ethiopia. Agricultural Economics. 2017 Jan 24;48(4):425–35.

6. 次世代**バイオディーゼル**燃料**が**内燃機関**の**未来**を救う。マツダが** 2022 年**スーパー**耐久**シリーズに** 100%**サステオで**参戦 - Web **モーターマガジン** [Internet]. web.motormagazine.co.jp.

[cited 2023 Jul 27].Available from:https://web.motormagazine.co.jp/_ct/17519156/p3#content-paging-anchor-17519156

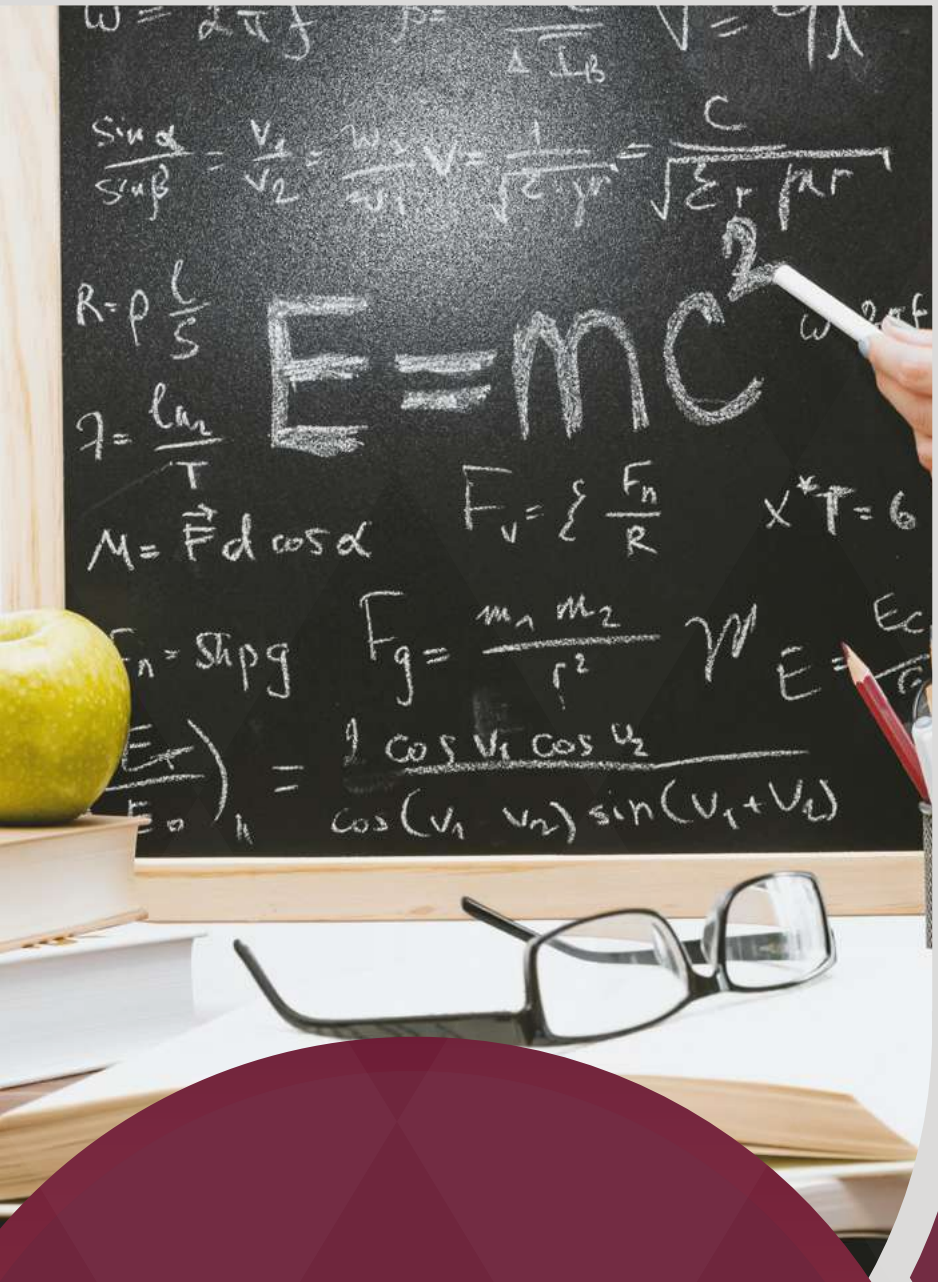
7. Stephen JD, Mabee WE, Saddler JN. Will second-generation ethanol be able to compete with first-generation ethanol? Opportunities for cost reduction. Biofuels, Bioproducts and Biorefining. 2011 Nov 4;6(2):159–76.

8. Reitz RD, Ogawa H, Payri R, et al. IJER editorial: The future of the internal combustion engine. International Journal of Engine Research. 2020;21(1):3- 10. doi:10. 1177/1468087419877990

9. Ibrahim, Asst.Prof.Dr. Abdullahi. (2020). Journal of Fundamentals of Renewable Energy and Applications. Journal of Fundamentals of Renewable Energy and Applications.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE



20

Physics

Fitzwilliam College Online Summer School Journal · 2023 Summer

The difference between classical rocket motion and relativistic rocket motion

DAKAN LI (Ken)

Abstract

Rockets are having a very different state in high speed, and we called this situation relativistic case. This article will use formula derivation to evaluate the difference of rocket motion between relativistic case and classical case. The challenges of high-speed rocket will also be mentioned, and some future aspect will be proposed. The data used are from other essays or textbooks.

1. Introduction

Nowadays rockets are widely used to explore the universe or execute other special missions, and their motion can be explained by Tsiolkovsky rocket equations. However, this equation will be incomplete to describe the motion if the rocket speed is approaching light speed. The energy, momentum, mass, and velocity of rocket will be different in high speed case.

Normally the speed to significantly show relativistic effect is above 0.3c. "At 30% c, the difference between relativistic mass and rest mass is only about 5%, while at 50% it is 15%, (at 0.75c the difference is over 50%)" [1] Nevertheless, the current technology can't let rocket achieve this speed. "Relativistic rockets require huge advances in spacecraft propulsion, energy storage, and engine efficiency which may or may not ever be possible. Nuclear pulse propulsion could theoretically reach 0. 1c using current known technology" [1] This means the relativistic rocket is only possible theoretical right now.

More than to show the different performance of relativistic rocket, this article will also mention the challenges for high speed motion. Moreover, to propose some methods that may help.

2. Equations

2.1. Tsiolkovsky rocket equation

In an isolated rocket system, the momentum is always conserved.

$$P_i = P_f \quad (1)$$

We can rewrite it as $Mv = -dM\,U + (M + dM)(v + dv)$ (2)

Then, let us say that the relative speed between rocket and exhausted products is vrel , and the speed between exhaust products and the rest frame is U. We can get this formula:

$$(v + dv) = v_{rel} + U \quad (3)$$

$$U = v + dv - v_{rel} \quad (4)$$

Then, substitute U into (2): $-dMv_{rel} = M\,dv$ (5)

And divide dt for both side: $-\frac{dM}{dt}v_{rel} = M\,\frac{dv}{dt}$ (6)

This equation showed the thrust force is proportional to the mass consumption rate and rocket speed relative to exhaust products.

According to the equation above, the velocity can be derived.

$$dv = -v_{rel}\frac{dM}{M} \quad (7)$$

Then we can integrate it. $\int_{v_i}^{v_f} dv = -v_{rel}\int_{M_i}^{M_f}\frac{dM}{M}$ (8)

Finally, we get the velocity change: $v_f - v_i = v_{rel}\, ln\frac{M_i}{M_f}$ (9)

From this equation, we can derive mass: $M_i = M_f e^{\Delta v/v_{rel}}$ (10)

2.2. Special relativity formulas

relative mass will increase as speed increase: $m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$ (11)

The velocity can be expressed as:

$$u'_x = \frac{u_x - v}{1-\frac{u_x v}{c^2}} \quad (12) \qquad u_x = \frac{u'_x + v}{1+\frac{u'_x v}{c^2}} \quad (13)$$

Momentum and energy can be expressed as:

$$p = m\gamma(u)u \quad (14) \qquad E = m\gamma c^2 \quad (15) \qquad T = m\gamma c^2 - mc^2 (16)$$

2.3. Relativistic rocket motion

The momentum in relativistic case can be written in: $P(t+dt) = (m+dm)\,\gamma\,(v+dv)\,(v+dv) + dm_e\gamma\,(v_e)v_e$ (17)

The y derived relative to v is: $\frac{d\gamma}{dv} = \frac{v}{c^2}\gamma^3$ (18)

We can write y(v+dv) as: $\gamma(v+dv) = \gamma(v) + \frac{d\gamma}{dv}dv$ (19)

Then, we can get: $m\gamma v + m\gamma dv + m\frac{v^2}{c^2}\gamma^3 dv + dm\gamma v + dm_e\gamma(v_e)v_e$ (20)

According to momentum conservation, which means P(t)=P (t + dt), we can cancel myv in both side:

$$0 = m\gamma dv + m\frac{v^2}{c^2}\gamma^3 dv + dm\gamma v + dm_e\gamma(v_e)v_e \quad (21)$$

And we know that y ²F² = y²-1, so the equation will become: $0 = m\gamma^3 dv + dm\gamma v + dm_e\gamma(v_e)v_e$ (22)

The energy can be expressed as: $E(t+dt) = (m+dm)\gamma(v+dv)c^2 + dm_e\gamma(v_e)c^2$ (23)

Then, according to energy conservation, we can deduce that: $0 = m\frac{v}{c^2}\gamma^3 dv + dm\,\gamma + dm_e\gamma(v_e)$ (24)

We can substitute (24) into (22), then we get: $0 = m\gamma^2\left(1-\frac{vv_e}{c^2}\right)dv + dm(v-v_e)$ (25)

And we know: $v'_e = \frac{v_e - v}{1-\frac{vv_e}{c^2}}$ (26)

Then we can get a differential equation: $m\frac{dv}{dm} + v'_e\left(1-\frac{v^2}{c^2}\right) = 0$ (27)

We assume v' e is constant, and we move constants to the right: $\frac{dv}{1-\beta^2} = -v'_e\frac{dm}{m}$ (28)

Then we divided by c for both side: $\frac{d\beta}{1-\beta^2} = -\frac{v'_e}{c}\frac{dm}{m}$ (29)

Now we can integrate it and get: $\frac{1}{2}log\frac{1+\beta}{1-\beta} = -\frac{v'_e}{c}\log m + C$ (30)

We can let m = m0, and F = 0 which is the initial condition to get C .

$$C = \frac{v'_e}{c}\log m_0 \quad (31)$$

Finally, we can express velocity as: $\beta = \frac{1-(\frac{m}{m_0})^{\frac{2v'_e}{c}}}{1+(\frac{m}{m_0})^{\frac{2v'_e}{c}}}$ (32)

Where F= $\frac{v}{c}$

Compare to (9), we can clearly see the velocity difference between classical rocket and relativistic rocket.

Moreover, according to Lorentz transformation and time dilation, we can get the relationship below:

$$-\frac{dm'}{dt'}\,t' = -\frac{dm}{dt}\,t \quad (33)$$

$$t' = \gamma\left(t - \frac{v}{c^2}r\right) \quad (34)$$

$$-\frac{dm'}{dt'}\left(\frac{t}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{v}{c^2\sqrt{1-\frac{v^2}{c^2}}}r\right) = -\frac{dm}{dt}\,t \quad (35)$$

3. Discussion

The special relativity declares the maximum information transmit speed is light speed. "As an object approaches the speed of light, more and more energy are needed to maintain its acceleration, with the result that to reach the speed of light, an infinite amount of energy would be required." [2]This is the main difference between classical equations and relativistic equations, in classical formula, the speed theoretically has no limit, but from relativistic rocket equations we found that F is definitely less than 1, which follows the special relativity rule above. Einstein theory make the rocket motion completer and more precise, although our rocket does not need these equations now, these equations pave our path to the future.

4. Future aspect

The current technology still has a large gap to reach a speed that have significant relativistic effects. There are challenges that block us, but there are possible methods that may solve the problem.

For instance, the energy efficiency for current propellant is relatively low. Then, if we want to rise the speed, the rocket will be extremely heavy. We cannot reach 0.01c right now, as I mentioned before, the mass at least needs to be above 0.3 c to have an obvious relativistic effect. So, how can we increase the propellant efficiency? There are two possible solutions. The first one is nuclear fusion. The specific impulse of nuclear fusion engine can reach 1000---30000, which is far more efficient than current engine. On the other hand, because its large amount of energy, it is dangerous to put them into application. We have to make nuclear fusion controllable, and this may achieve in decades. Another way is anti-matter annihilation rockets. It is about 50 times stronger than nuclear fusion, which can 100% convert mass to energy. Other antimatter rockets in addition to the photon rocket that can provide a 0.6c specific impulse (studied

for basic hydrogen-antihydrogen annihilation, no ionization, no recycling of the radiation. However, now we cannot make a large quantity of antimatter, it is expensive and require specific condition. As a consequence, it probably takes longer time to be widely used than nuclear fusion.

The material is also a challenge if a rocket wants to reach a high speed. The high speed will cause heat effect because of high kinetic energy. Our material cannot take this heat. This means we have to synthesis some new substance artificially or to discover some new material in the future which is unpredictable. Now equipment like LHC are devoting to do this.

5. Conclusion

The rocket motion in relativistic case is different from classical one for velocity, momentum, energy. The special relativity makes the equations for rocket motion more complete, but if we want relativistic rocket exist in real life, there are problems that we need to overcome.

6. Reference

1. Relativistic rocket - Wikipediaa

2. https://math.ucr.edu/home/baez/physics/Relativity/SR/Rocket/rocket.html

The visual appearance of the black hole

RUINAN SU (Nebula)

Abstract

This research aims to describe the visual appearance of black holes and explain the physical principle behind it. The evidence used to support the research is collected from secondary methods due to the theoretical nature of the topic being investigated. The topic was explained by first analyzing the structure of a black hole with accretion disk, then combining with the bending of light rays and relativistic beaming effect to provide the description of the visual appearance of it.

1. Introduction

A black hole is one of the most mysterious object in the universe. Time and space are distorted to the extreme inside the event horizon, and the gravitational pull there is so great that nothing, not even light, can escape. It is hard to imagine that this black, deadly celestial body is the ultimate fate of a star. The stars like sun release energy by the hydrogen-helium nuclear fusion during about 90% of their lifetime. Stars in this stage are called main sequence stars. As long as the hydrogen in the star is exhausted, the main sequence stage will end and the star will start contracting because of the gravity. Then the gravitational energy will convert into heat energy and temperature will rise. Along with the increasing of temperature, a new round of nuclear fusion will start and this time the fuel is helium. When most of the helium is converted into carbon and oxygen, the stars with mass over 25$M_\odot$ ($M_\odot$ means solar mass) will repeat this circle, and keep releasing energy through nuclear fusion with new elements with increasing nuclear mass until iron is generated. Iron is the element with the most binding energy which means no more energy can be released during nuclear fusion, thus, the nuclear fusion will stop and then there is no radiation pressure against gravity. As a result, the massive star will compact and experience supernova explosion, ultimately becoming a black hole which nothing can escape from, not even light.

Since black hole itself is invisible, people cannot investigate black holes by detecting the light emitted from it like the other stars. Therefore, people can only observe black holes through the space around them. Nowadays, even though there are multiple ways to find black holes, most of the black holes detected are still rely on the X-ray detection method. This method is based on the detection of the X-ray from the accretion disk around the black hole. **Stephen E. Schneider and Thomas T. Arny** suggested that when a black hole and its companion star is close enough that the gas from the companion star may be drawn toward the black hole by its gravity, the falling matter swirls around the black hole and forms an accretion disk. The disk orbits at nearly the speed of light when it closes to the black hole, therefore, turbulence and friction heat the swirling gas to a furious 10 million K, causing it to emit X-ray and gamma ray. This accretion disk not only makes black holes easier to be detected, but also provides an interesting visual appearance of the black hole.

2. Literature review

2.1. The types of black holes and radius of their event horizon

According to the no-hair theorem, for a stationary black hole, it can be completely characterized by only three independent externally observable classical parameters: mass, electric charge and angular momentum **(Misner, C. et al., 1973).** Thus, there are four different types of black hole: Non-rotating and uncharged black holes (Schwarzschild black holes), rotating and uncharged black holes (Kerr black holes), non-rotating and charged black holes (Reissner-Nordstrom black holes), and rotating and charged black holes (Kerr-Newman black holes). This research mainly focuses

on Schwarzschild black holes.

When considering black holes, people might first think of a sphere with black surface, this surface is called event horizon, which means any events inside this surface cannot affect the observer. The concept of black holes has already been cited by Pierre-Simon **Laplace** in **1796** and one hundred years later, in the **Schwarzschild** solution to **Einstien**'s field equations states that: for a non-rotating and uncharged black hole, the radius of the event horizon equals to the Schwarzschild radius (R_s), which is calculated by the equation:

$$R_s = 2GM/c^2$$

where G is the gravitational constant, M is the mass of the black hole and c is the speed of light.

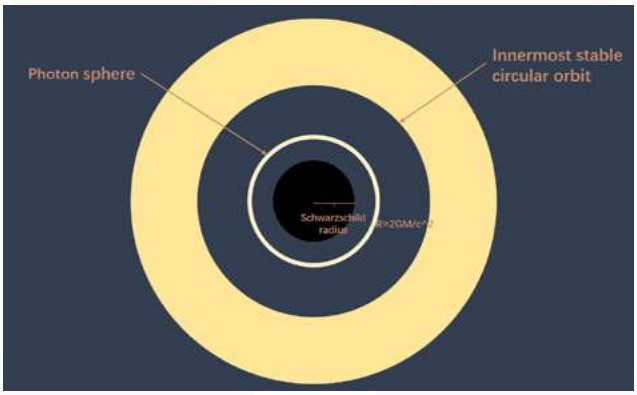
2.2. Photon sphere and the innermost stable circular orbit

Now consider further away from the event horizon, a sphere formed by photons will appear in our sight. This is called photon sphere, where photons here experience an unstable circular motion which will eventually end up falling into the black hole or escape it **(Teo, E. 2003)**. If considering standing at the photon sphere, ones can even see the back of their head because of light's circular motion. The radius of this circle is 1.5Rs for non-spinning black holes.

Continue moving away from the black hole, people will arrive the inner edge of the accretion disk where no matter with mass spinning around the black hole inside this edge. This edge is called the innermost stable circular orbit (ISCO), which means there is no object can maintain a stable circular motion if it cross the ISCO **(Misner, C. et al., 1973)**. This is because of two main reasons. First of all, when an object wants to maintain a stable circular motion, the centripetal acceleration of the object need to be equal to the gravitational force of the central body. As it getting closer to the central body, due to the inverse square relationship between the radius and the gravitational force, the centripetal force required will increase rapidly and so as the linear speed required to provide the centripetal force. Therefore, there will eventually be a radius which the linear speed required is equal to light speed which is the upper limit of the speed in general relativity. Moreover, in general relativity, compact mass will warp spacetime and change the path of the object travels. Combine these factors together, ones can evaluate the value the radius of ISCO.

For Non-spinning black holes, the radius of ISCO (Risco) gives as: Risco=3Rs **(Jefremov et al., 2015).** For Kerr black hole, ISCO will be closer to the event horizon, the faster the black hole spins, the closer the ISCO gets. Most of the celestial body have their ISCO inside their surface, but for extremely compact body like black holes and some neutron stars will have the ISCO outside their surface and are able to form photon sphere **(Nemiroff et al., 1993).**

3. Result and discussion



Graph 1: structure of the black hole with photon sphere and accretion disk

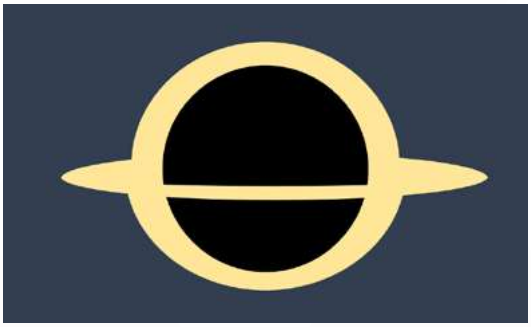
Graph 1 shows the rough model of a black hole according to the structure that has been discussed in the literature review. This section will focus on finding the visual appearance of the black hole based on this model.

3.1. The shadow of the black hole

When considering parallel light rays shooting at the black holes, the shadow appeared will actually larger than the event horizon. This is because in general relativity, light follows the curvature of space-time and black hole warps the space-time around it, thus, bends the light rays. For the light rays slightly higher than the event horizon will still end up falling into the black hole. Thus, in order to make the parallel light not getting absorbed by the black hole, it actually need to be at least 2.6Rs away from the singularity of the black hole and it will go around the photon sphere and move to infinity. Thus, the shadow that black hole appeared is actually 2.6 times bigger than the event horizon.

3.2. The shape of the accretion disk

Assuming the observer is looking at the black hole with its accretion disk edge on, so part of the accretion disk in hidden behind the black hole. If the light follows a straight path, the observer will not be able to see the light comes from the part of the accretion disk behind the black hole. But due to the changing of space time by the black hole, the photon sent from the hidden part of the accretion disk will follow a curved path and end up receiving by the observer. Thus, both the top and bottom surface of the hidden part of the accretion disk can be seen by the observer and forming a ring shape around the black hole like the image shown in Graph 2.



Graph 2: Visual appearance of the black hole without considering the relativistic beaming effect

3.3. Relativistic beaming effect

The figure obtained now is already close to the visual appearance of the black hole, but there is still one important phenomenon needs to be considered. As the accretion disk is spinning around the black hole with its speed close to the speed of light, which means the light emit from the accretion disk will experience a strong Doppler effect. According to special relativity, the equation of observed frequency and wave length is that:

$$f=f_0*[(1-\beta)/(1+\beta)]^{1/2}$$
$$\lambda=\lambda_0*[(1+\beta)/(1-\beta)]^{1/2}$$

where f° is the frequency of the source, λ° is the wave length of the source and $\beta = v/c$. The accretion disk will have one side spin towards the observer and one side spin away from the observer as long as the observer is not looking perpendicular to the accretion disk. Thus, the frequency will be larger on the side spin towards the observer and smaller on the other side. The light intensity is proportional to the value of frequency, as a result, one side of the accretion disk will be brighter than the other side. People can also tell the direction of spinning by analyzing the difference in brightness.

Moreover, the light aberration also contributes in forming this phenomenon. Assume in the reference frame of the observer, the light source is moving with a speed v and the direction of motion of the source has an angle θ to the line join the source and observer together as the light is sent. The aberration of the light source, θ_o, is calculated by this equation:

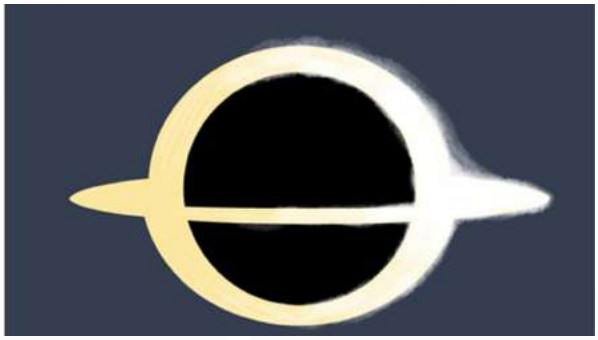
$$\cos\theta_o=(\cos\theta_s-\beta)/(1-\beta\cos\theta_s)$$

where β is calculated from the speed v of the source. This means that the rays of light emitted by a moving object is concentrated conically towards its direction of motion. Thus, there are more photons received on the side of the accretion disk moving towards the observer and there are less photons received on the side moving away from the observer.

This is the physical principle behind the relativistic beaming effect which causing the difference in brightness of the accretion disk.

4. Conclusion

In conclusion, the visual appearance of the black hole should look like Graph 4, where comes from the structure of a black hole with accretion disk, the bending of light rays, and the relativistic beaming effect. This visual appearance also has been verified by the pictures of the M87 and Sgr A* black holes taken by the Event Horizon Telescope. Because of the great distortion of space-time, the black hole with accretion disk provides a beautiful and counterintuitive figure. In the future, with the development of technology, the telescope with a higher level of precision will be used and provide more photos of black holes with better quality.



Graph 3: Visual appearance of the black hole.

Graph 4: Photo of the M87 black hole taken by Event Horizon Telescope

5. Bibliography

1. Bethe, H., Brown, G. and Lee, C., 2003. Formation and evolution of black holes in the galaxy. Singapore [etc.]: World scientific, pp.71-74.

2. Bowyer, S., Byram, E., Chubb, T. and Friedman, H., 1965. Cosmic X-ray Sources. Science, 147(3656), pp.394-398.

3. Cavalleri, G. (1968) ‘Solution of ehrenfest’s paradox for a relativistic rotating disk’, Il Nuovo Cimento B Series 10, 53(2), pp. 415–432. doi:10. 1007/bf02710244.

4. Chruściel, P.T., Costa, J.L. and Heusler, M. (2012) ‘Stationary black holes: Uniqueness and beyond’, Living Reviews in Relativity, 15(1). doi:10. 12942/lrr-2012-7.

5. Einstein, A (June 1916). "Näherungsweise Integration der Feldgleichungen der Gravitation". S itzungsberichte der Königlich Preussischen Akademie der Wissenschaften Berlin. part 1: 688 – 696.

6. Gro/n, O. (1975) ‘Relativistic description of a rotating disk’, American Journal of Physics, 43(10), pp. 869–876. doi:10. 1119/1.9969.

7. Jefremov, P.I., Tsupko, O.Yu. and Bisnovatyi-Kogan, G.S. (2015) ‘Innermost stable circular orbits of spinning test particles in Schwarzschild and Kerr space-times’, Physical Review D, 91(12). doi:10. 1103/physrevd.91. 124030.

8. Nemiroff, R.J., Becker, P.A. and Wood, K.S. (1993) ‘Properties of ultracompact Neutron Stars’, The Astrophysical Journal, 406, p. 590. doi:10. 1086/172471.

9. Schneider, S. and Arny, T., 2019. Pathways to astronomy, 5e. Harbin: Harbin Institute of Technology press, p.546.

10. Sparks, W.B. et al. (1992) ‘A counterjet in the elliptical Galaxy M87’, Nature, 355(6363), pp. 804–806. doi:10. 1038/355804a0.

11. Su, Y. and Su, C., 2019. Tian wen xue xin gai lun. 5th ed. Beijing: Ke xue chu ban she, p.355.

12. Teo, E. (2003) ‘Spherical photon orbits around a Kerr Black Hole’, General Relativity and Gravitation, 35(11), pp. 1909–1926. doi:10. 1023/a:1026286607562.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

03

Computer Science

Fitzwilliam College Online Summer School Journal · 2023 Summer

A Brief Introduction of Self-balancing Trees

TIANYI YANG (Stephen)

1. Introduction

1.1. Tree

definition 1.1 In graph theory, a tree is an undirected graph in which any two vertices are connected by exactly one path or, equivalently, a connected acyclic undirected graph.

Trees can help humans to organize the relationship between the data, but for most trees, it's pretty hard for a computer to understand and deal with. However, we have some specific trees with specially designed structures so that computers can interact with data quickly with them.

This essay will discuss these tree structures' concepts and how they are im- plemented. Some comparison of their performance, as well as their applications in real life, would also be included.

1.2. Binary Search Tree (BST)

definition 1.2 A binary tree is a tree data structure in which each node has at most two children, the left and right child.

A binary tree is a structure with great potential. It's easy for a computer to represent with a handful of memory. More importantly, storing data in a specific order to a binary tree makes searching for the data later in it easy.

definition 1.3 In computer science, a binary search tree (BST), also called an ordered or sorted binary tree, is a rooted binary tree data structure with the key of each internal node being greater than all the keys in the respective node's left subtree and less than the ones in its right subtree .

With a BST, binary search algorithm can be used to search data. In the best situation, we only need O(log N) time to get the result. With the same data, BST's height is expected to be as short as possible, which means it should be as bushy as possible.

But here's a problem.Consider adding 1 to 7 into a BST. By adding them in ascending order, we get BST shown in figure 1. By adding numbers in the order of 4, 2, 6, 1, 3, 5, 7, we get BST shown in figure 2. It's easy to find that The

Figure 1: BST1

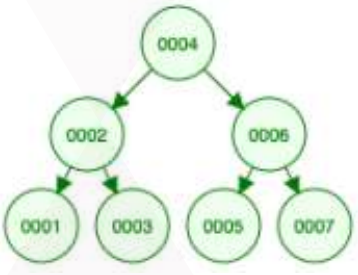


Figure 2: BST2

worse query time complexity for BST 1 is O(N), for BST 2 is O(log N).

If the data are inserted randomly, the query time complexity is O(log n). But if a data set is inserted in ascending/ descending order to a BST, the tree would degenerate into a linked list.

In daily life, many data are well-organized, so if we directly add them into a BST, the tree would be pretty spindly, which leads to a bad performance. Therefore, some trees which can avoid this situation are required.

2. Self-Balancing Tree

This section will introduce several kinds of self-balancing trees; they are de- signed to avoid the appearance of very bad structures. Some have special con- struction, and some use a special operation called tree rotation(2.1.2) to trans- form a spindly tree into a bushy one.

2.1. AVL tree

Named after inventors Adelson-Velsky and Landis, AVL tree is the first self- balancing tree invented. It has a relatively fast performance for lookup-intensive applications since it's strictly balanced, but it needs more time to organize the order of data in response.

2.1.1. Tree traversal

Trees may be traversed in multiple ways. Unlike linked lists, one-dimensional arrays and other linear data structures are canonically traversed in linear order. They may be traversed in depth-first or breadth-first order. There are three common ways to traverse them in depth-first order: in-order, pre-order, and post-order.

For in-order traversal specifically, it would recursively traverse the current node's left subtree, then visit the current node and finally recursively traverse its right subtree.

It's worth noticing that for a BST, in-order traversal retrieves the keys in ascending sorted order, and if the result of in-order traversal of a tree is in ascending order, the tree is a BST.

2.1.2. Tree Rotation

definition 2.1 In discrete mathematics, tree rotation is an operation on a bi-nary tree that changes the structure without interfering with the order of the elements. A tree rotation moves one node up in the tree and one node down.

There are two kinds of rotations in detail.

definition 2.2 rotateLeft(G): Let x be the right child of G. Make G the new left child of x.

definition 2.3 rotateRight(G): Let x be the left child of G. Make G the new right child of x.

The rotateLeft(G) process is shown in figure 3. G's right child, P, merges with G, bringing its children along. P then passes its left child to G, and G goes down to the left to become P's left child—the structure of the tree changes as well as the number of levels. Rotation can also be applied on a non-root node by temporarily disconnecting the node from the parent, rotating the subtree at the node, then reconnecting the new root.

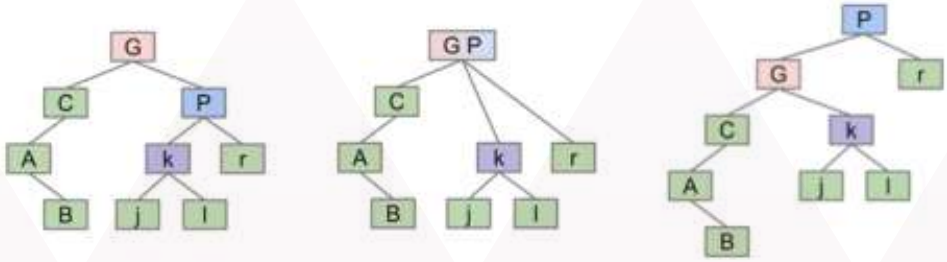


Figure 3: process of rotateLeft(G)

Below is the code in Java to implement these functions.

```
//rotateLeft
private void  leftRototate(){
//create  a new node with the value  of root
Node newNode = new Node(this .value);
//new node’s  left  child  -> root’s  left  child
newNode .left = this .left;
//new node’s right  child  -> root’s right  child’s  left  child newNode .right = this .right .left;
//set root’s value to  its right  child’s value
this .value = right .value;
//root’s right  child  ->  it’s right  child
right = right .right;
//root’s  left  child  -> the new node
left = newNode;
//rotateRight
private void rightRotate(){
//create  a new node with the value  of root
Node newNode = new Node(this .value);
//new node’s right  child  -> root’s right  child
newNode .right = this .right;
//new node’s  left  child  -> root’s  left  child’s right  child
newNode .left = this .left .right;
//set root’s value to  its  left  child’s value
this .value = this .left .value;
//root’s  left  child  ->  it’s  left  child
this .left = this .left .left;
//root’s right  child  -> the new node
right = newNode;
```

Tree rotation is used to change the tree’s shape and, in particular, to decrease its height by moving smaller subtrees down and larger subtrees up, resulting in improved performance of many tree operations.

Since applying rotation won’t affect the in-order of the tree, a BST is still a BST after rotation.

2.1.3. Unbalanced Situation

definition 2.4 In a binary tree, the balance factor of node X (BF(X)) is de- fined as the height difference of its two child sub-trees.

definition 2.5 An unbalanced tree contains a node X with $|BF(X)| > 1$.

Though there are all kinds of unbalanced trees, there are only 4 kinds of unbalanced situations. (shown as figure 4 below)

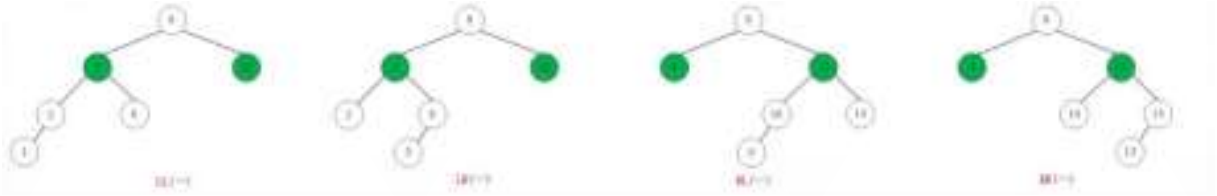


Figure 4: unbalanced situations(LL LR RL RR)

With tree rotation, we can fix all these kinds of unbalanced trees into balanced trees:

It turns into a balanced tree by applying rotateRight() to the root node in LL.

It turns into a balanced tree by applying rotateLeft() to left child(now it becomes LL) and then rotateRight() to the root node in LR.

It turns into a balanced tree by applying rotateRight() to right child(now it becomes RR) and then rotateLeft() to the root node in RL.

It turns into a balanced tree by applying rotateLeft() to the root node in RR.

What AVL tree does is detect whether an unbalance situation takes place when adding or deleting data. If it does happen, then fix them throw rotation

2.1.4. Performance

AVL tree is an absolutely balanced binary search tree, which can ensure ef- ficient query time complexity, that is, O(log N). However, the performance is very low to do some structural modification operations on the AVL tree, such as: maintaining its absolute balance when inserting, and the number of rota- tions is relatively large. When deleting, letting the rotation continue to the root position is possible.

2.2. B-tree

Besides using tree rotation, another idea to implement a self-balancing tree is using some ingenious construction. B-tree follows this idea and has a good performance.

2.2.1. Naive Approach

Consider when the balance situation would be broken. When a new leaf node is added to the tree, the height of one subtree will increase, which means the balance may be ruined.

To avoid this, the tree just put the data in our existing node instead of adding a new leaf node to store data.

For example, in figure 5 below, a complete BST is on the left side. To insert 17 into the tree, we put it into the node

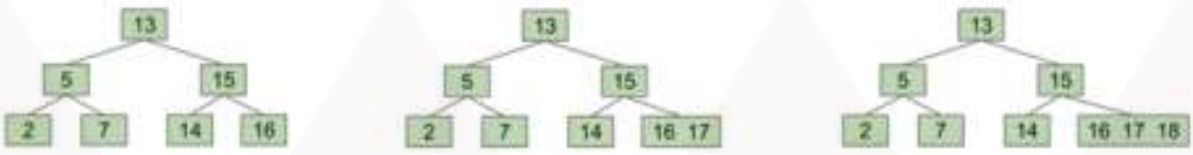


Figure 5: a naive idea of B-tree

containing 16 directly. To insert 18, we put it into the node containing 16 and 17.

The tree uses a list as a node to solve the problem. It’s useful when the list is short. Binary search can be applied to

the tree, and by iterating the list, it's easy to find the data stored in the list node.

But when the list gets longer, this approach is not efficient enough. The time taken to iterate the list is way much than the time for using binary search to find the list node, so the tree approximately degenerates into a linked list.(figure 6)

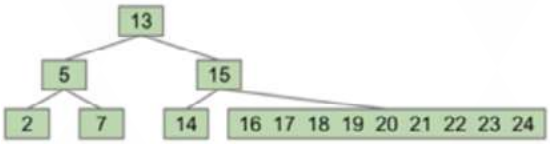


Figure 6: a naive B-tree with a long list

2.2.2. Solution

A limitation to the maximum list length is set to avoid the degeneration problem.

For example, if the maximum length of the list of the tree in figure 7 is 3, then the list has to be shortened.

In B-tree, the data in the middle of the list(the left middle is the length is even) would be moved to the parent node of the list node.

For instance, the tree in figure 7 would be transformed into the tree in figure 8.

But the result isn't satisfying. In the second tree, 17's child node is a list containing 16, 18, 19, which has numbers both larger and smaller than it. This way, no efficient algorithm can be used to search data here.

While moving the data in the middle, B-tree also split the list from that position.

Thus, the tree in figure 8 would turn into the tree in figure 9. Now that data are put in order, data in this tree can be searched using a combination of binary and linear search.

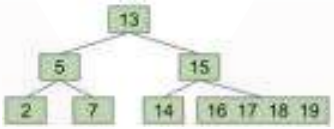


Figure 7: bad form

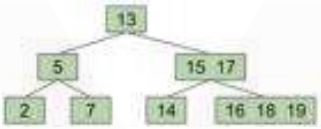


Figure 8: try to change

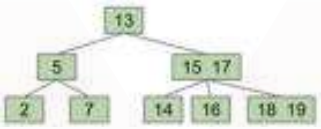


Figure 9: good form

Figure 10: B-tree transforming

A B-tree of order[1] m has 4 properties according to its operation form.

property 2.1 Every node except the root must contain at least m/2-1 keys .The root may contain a minimum of 1 key.

property 2.2 All nodes may contain at most m - 1 keys.

property 2.3 All leaves must be the same distance from the source.

property 2.4 A non- leave node with k items must have exactly k +1 children.

There are 2 situations to consider while deleting data from a B-tree: non-leaf node and leaf node. It's relatively complicated. The web here explains the logic of B-tree deletion well.

2.2.3. B+ tree

Though B-tree is good enough, B+ tree is invented to adapt the usage in database better.

[1] The order of the tree represents the maximum number of children a tree's node could have. Through property 2.4, if the maximum length of list of a B-tree is m, the order of the tree is also m

A B+ tree has all properties that a B-tree has, but there are several differences between the structure of a B-tree and a B+ tree:

In B+ tree, only leaf nodes store actual data. Therefore, a B+ tree can store more keys per node than a B-tree if they have the same order. This is important since the capability of pages is fixed for hardware. B+ tree stores more keys in a page[2] than b-tree so that it can store more data with the same layers.

What's more, since all data are stored in leaf nodes in a B+ tree, and all leaves are the same distance from the source(property 2.3), so the time it takes to search different data is similar, which is more fair compares to B-tree.

The leaves are not connected with each other on a B-tree, whereas they are connected on a B+ tree. Thus, a B+ tree can deal with range queries more efficiently. In B+ tree, after finding the beginning data, the rest data can be obtained through the extra link. Contrastingly, continuous data may not be stored in the continuous memory area, which means it's hard to respond to a range query.

2.2.4. Performance

If the order of the B-tree is large, when there is little data, the tree is similar to a linked list, which would be very slow. With a reasonable order, B-tree is fast.For a B-tree of order m with N nodes, inserting or searching data takes between $\log_{m-1} N$ to $\log_{m/2} N$ of comparison.

2.3. Red Black tree

Red Black tree(RBtree) is a kind of self-balancing tree that also uses tree rotation(2.1.2). Still, it has some other properties, so less time is needed to maintain balance when inserting/deleting data. However, it's not as balanced as the AVL tree, so a longer time is required to respond to queries.

There are two kinds of nodes in a RBtree: red node and black node. All RBtree types correspond to B-tree; a red node and its parent node(which must be a black node) can be seen as 'glued', which corresponds to list nodes in B-tree.

2.3.1. Left Leaning Red Black treeh

There are many variations of RBtree, and left-leaning RBtree(LLRB) is one of them. It's easier to implement and has similar performance compared to RBT.

property 2.5 1- 1 correspondence with 2- 3 trees[3].

property 2.6 No node has 2 red links.

property 2.7 Every path from the root to leaf has same number of black links (because 2- 3 trees have same number of links to every leaf) .

property 2.8 Height is no more than 2x height of corresponding 2- 3 tree.

Since property 2.5, by inserting into a 2-3 tree and converting it to let the tree has properties above, the result of inserting into LLRB can be obtained.

However, this is stupid since there is no point in spending extra time to trans- form a usable 2-3 tree into a LLRB. So the proper method is to insert into the LLRB the same as a normal BST and then use rotations to fix its 1-1 map to 2-3

[2] For B-tree and B+ tree which are used in database, a node corresponds to a page(see 2.4)
[3] A 2-3 tree means a B-tree of order 3

tree.

Here are the tasks that need to be done when inserting into a LLRB:

In 2-3 trees, data always be added into a leaf node, so the colour of the link added to LLRB should always be red.

According to the name, a red link should never appear on the right side. A rotation should be applied to maintain the LLRB property.

However, if the node has two red links, it needs to be allowed to exist for a while.

If there are 2 left red links, it corresponds to a list with a length of 4. After rotating to create a tree with two red links, flipping the colours of all edges touching the node.

Though it seems complex to fix the format, only 3 if-sentences are needed to update an AVL tree to an LLRB.

```
private Node put(Node h, Key key, Value val)
{if  (h == null)  {return new Node(key, val, RED);}
int  cmp = key .compareTo(h .key);
if  (cmp  <  0)  {h .left   = put(h .left,   key, val);}
else  if  (cmp  >  0)  {h .right = put(h .right, key, val);}
else  {h .val = val;}
if  (isRed(h .right) &&  !isRed(h .left))
{h = rotateLeft(h);}
if  (isRed(h .left) &&  isRed(h .left .left))
{h = rotateRight(h);}
if  (isRed(h .left) &&  isRed(h .right))
{flipColors(h);}
return h;}
```

Because a left-leaning red-black tree has a 1-1 correspondence with a 2-3 tree and will always remain within 2x the height of its 2-3 tree, the runtimes of the operations will take log N time.

2.3.2. Normal RBtree

An LLRB can only have one red link per node, corresponding to a 2-3 tree. In a normal RBtree, each node can have 2 red links, corresponding to a 2-3-4 tree [4] .

Here are the properties that an RBtree should have:

property 2.9 The root node of RBtree is black.

property 2.10 The leaf nodes [5] are all black.

property 2.11 Red nodes' children and parent are black nodes.

Classification by the node case after the conversion of RBtree to B-tree can be made use of the equivalence property of the 2-3-4 tree and red-black tree.(see figure 15)

[4] A 2-3-4 tree means a B-tree of order 4
[5] The leaf node in RBtree is slightly different, it refers to the lowest empty nodes (external nodes)

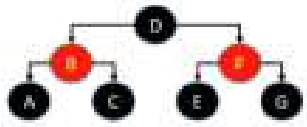


Figure 11: red black red

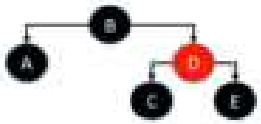


Figure 12: black red

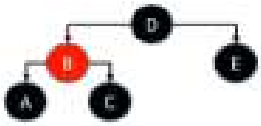


Figure 13: red black

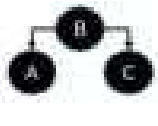


Figure 14: black

Figure 15: 4 situations of RBtree node

Same as LLRB, the colour of a new node being inserted is red.(see 2.3.1)

Considering how to insert a new node into the tree in figure 16 makes it easier to understand how an RBtree maintains its properties.

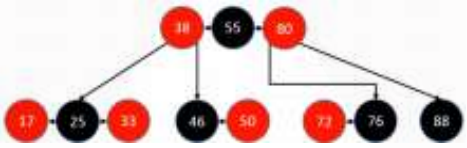


Figure 16: a RBtree

The RBtree in figure 16 contains all the previously classified situations. 12 places in this tree can insert new nodes. Among them, 4 situations don’t require any further operation to maintain the properties(cases shown in figure 17)

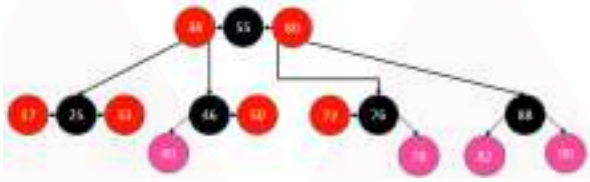


Figure 17: 4 places that don’t need further operation

8 situations do not satisfy the property of red-black tree 2.11, and 4 situations on the left belong to the overflow situation of B-tree nodes (a 4-order B-tree node can store a maximum of 3 numbers, and these 4 cases already have 3 numbers, and another one is inserted, which beyond the capacity range of the 4-order B-tree node). These 8 cases require additional processing. (figure 18)

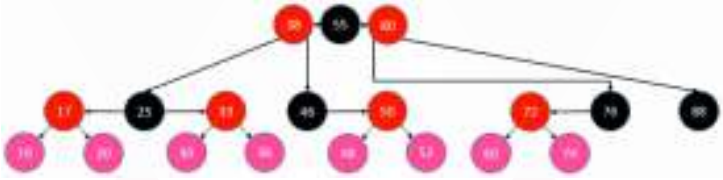


Figure 18: other places that need further operation

The situations can be named by their unbalance situations(see 2.1.3). For example, for the tree in figure 18, insert node 10 or 60 forms LL; insert node 36 or 52 forms RR; insert node 20 or 74 forms LR; insert node 30 or 48 forms RL

For cases without overflow:
To fix LL/RR, dye parent black and grand red.
To fix LR/RL, dye inserted node black and grand red.

Then apply rotation mentioned in 2.1.2.

For cases with overflow:
Dye parent and uncle black and grand red.

2.3.3. Performance

RBtree ensures that the longest path is not more than twice the shortest path, so it is approximately balanced (the shortest path is an all-black node, the longest path is a red node and a black node, and when the path from the root node to the leaf node has the same black node, the longest path is exactly twice the shortest path).

RBtree's search, insert, and delete operations are all O(log N) in time com- plexity.

2.4. Applications

2.4.1. AVL tree and Red Black tree

An AVL tree is suitable when a data structure that is query efficient and ordered and the number of data is static is needed, but a system that changes frequently is not a good fit.

In real life, most data in memory changes frequently. RBtree is used to store data there mostly [6]. For database system, tree structure will fewer layers are better since hard disk IO is too expensive.(see B-tree and B+ tree 2.4.2)

As a result, though AVL tree can solve the unbalance problem, it's too slow to be used. It's more common for AVL tree to appear in data structure course book to let students understand the idea and commemorate the progress from nothing to something.

Though RBtree is not as balanced as AVL tree, it has better overall perfor- mance, which means in most situations, an RBtree would be used rather than AVL tree

2.4.2. B-tree and B+ tree

When data is little, Red Black tree is faster than B-tree. But storing a large amount of data on hard disks makes putting a page into memory time- consuming, so using fewer pages is better. B-tree is suitable for doing this. The order of B-tree should be as large as possible to get fewer layers, so each node corresponds to a page. This way, only a few pages need to be mentioned to deal with a query. For instance, within 4 pages, an element would be found among 6.2 1010 elements.

However, most databases are now using B+ tree, since it has some improve- ments compared to B-tree.

The leaves are not connected on a B-tree, whereas they are connected on a B+ tree. Thus, a B+ tree can deal with range queries more efficiently. In B+ tree, after finding the beginning data, the rest data can be obtained through the extra link. Contrastingly, continuous data may not be stored at continuous memory area, which means it's hard to respond to a range query.

2.5. Comparison

Though the time complexity of these trees can be calculated, it's more visu- alized to run the code and see the result. And in real life, sometimes the result may differ from the theoretical result.

[6] Many ADTs like hashmap, treemap, and treeset use RBtree to store data since RBtree can significantly shorten the searching operation time for any data size.

By searching the implementations online and writing a program to compare the time these trees needed to insert and search data, the result in appendix A came out.

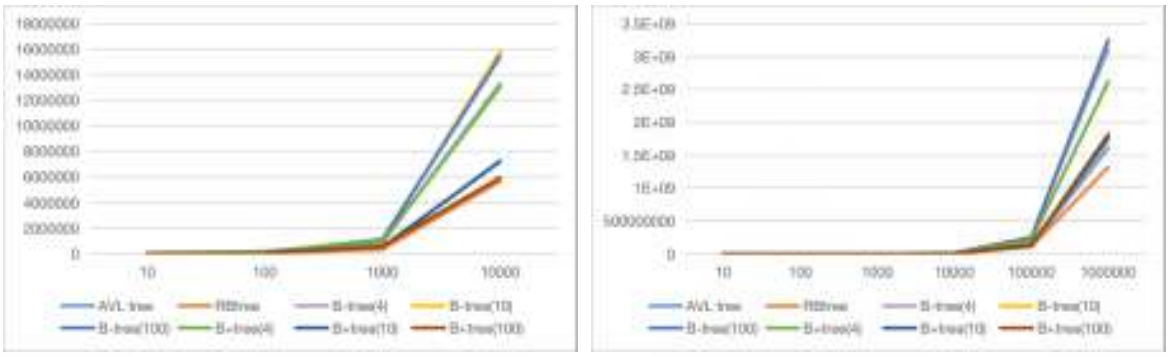


Figure 19: data size – insert time(ns)

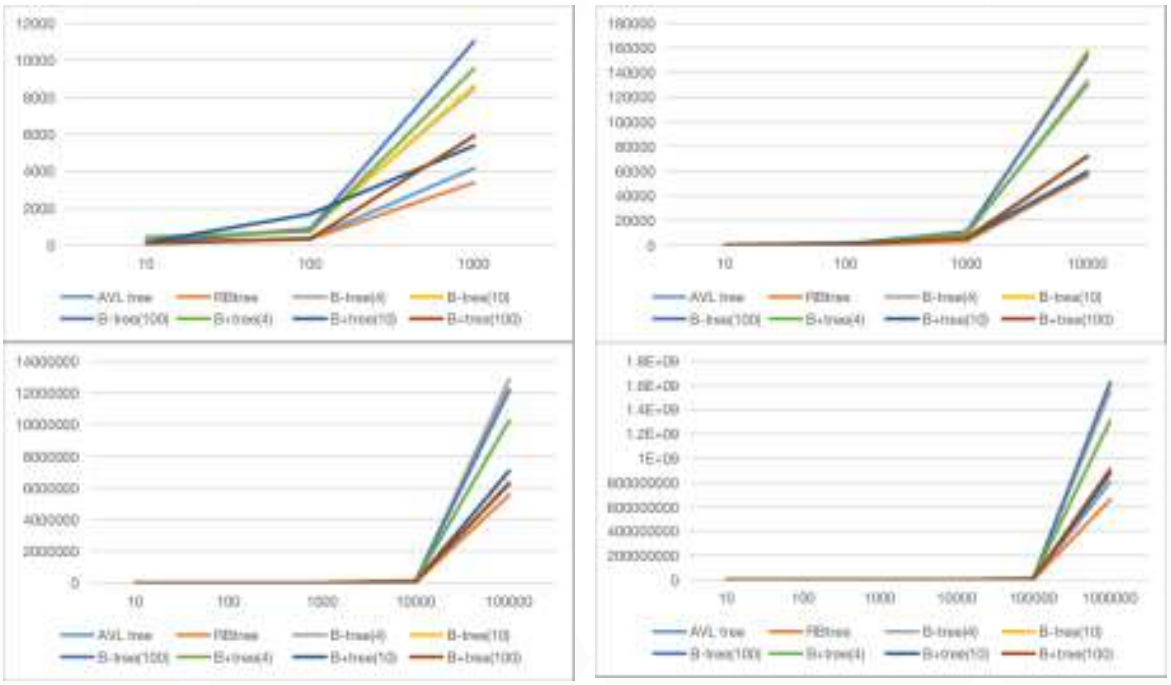


Figure 20: data size – query time(ns)

The line charts in figure 19 and figure 20 are drawn by organizing the data.

From figure 19, when the data set is relatively small(n*10^3), B-tree of order 10 and B-tree of order 100 perform the worst. The difference is more significant when the data set grows larger(n*10^5). B-trees have the worst performance, B+ trees are better, and RBtree, faster than AVL tree, performs best. This is roughly consistent with theoretical speculation.

In figure 20, AVL tree and RBtree are generally still the best, B+ tree follows and B-tree is the worst. But this is different from the theory since logically AVL tree should be faster than RBtree when dealing with queries. This is probably because the optimize of the code for RBtree is better. Also, for B+ tree, the code uses array as leaf node, so binary search can be applied. But in B-tree the nodes are linked list, which means it will use linear search, which is much slower.

The code for this test can be found here.

3. Reference

1. Caxton C Foster. A generalization of avl trees. Communications of the ACM, 16(8):513 – 517, 1973.

2. Goetz Graefe et al. Modern b-tree techniques. Foundations and Trends @ in Databases, 3(4):203 – 402, 2011.

3. Sabine Hanke. The performance of concurrent red-black tree algorithms. In Algorithm Engineering: 3rd International Workshop, WAE' 99 London, UK, July 19 – 21, 1999 Proceedings 3, pages 286 – 300. Springer, 1999.

4. Josh Hug. Hug61B Spring 2019 Edition. gitbook, 2021.

5. Erkki M ¨ akinen. A survey on binary tree codings. The Computer Journal,34(5):438 – 443, 1991.

6. Robert Sedgewick. Left-leaning red-black trees. In Dagstuhl Workshop on Data Structures, volume 17, 2008.

7. Robert Sedgewick and Kevin Wayne. Algorithms (4th edn). Google Scholar Google Scholar Digital Library Digital Library, 2011

4. A Original result of code

```
insert 10 nodes to AVL tree:                    10667 ns
insert 10 nodes to RBtree:                      19292 ns
insert 10 nodes to B-tree of order 4:           24000 ns
insert 10 nodes to B-tree of order 10:          17334 ns
insert 10 nodes to B-tree of order 100:         13833 ns
insert 10 nodes to B+ tree of order 4:          22583 ns
insert 10 nodes to B+ tree of order 10:         31459 ns
insert 10 nodes to B+ tree of order 100:        12541 ns
fast <------------insert------------> slow
AVL B+(100) B(100) B(10) RBt B+(4) B(4) B+(10)
search in AVL tree:                             3500 ns
search in RBtree:                               3792 ns
search in B-tree of order 4:                    5625 ns
search in B-tree of order 10:                   2250 ns
search in B-tree of order 100:                  2000 ns
search in B+tree of order 4:                    3917 ns
search in B+tree of order 10:                   1750 ns
search in B+tree of order 100:                  1208 ns
fast <------------search------------> slow
B+(100) B+(10) B(100) B(10) AVL RBt B+(4) B(4)
averagetime_insert:
{B(10)=12329 .16, B+(100)=8404 .08, B+(4)=43430 .38,
B(100)=12608 .37, RBt=12471 .66, B(4)=19601 .28,
AVL=9443 .35, B+(10)=15146 .59}
averagetime_search:
{B(10)=123 .2916, B+(100)=84 .0408,
B+(4)=434 .30379999999997, B(100)=126 .08370000000001,
RBt=124 .7166, B(4)=196 .0128, AVL=94 .43350000000001,
B+(10)=151 .4659}
-----------------------------------------------------
```

insert 100 nodes to AVL tree: 34958 ns
insert 100 nodes to RBtree: 30959 ns
insert 100 nodes to B-tree of order 4: 141916 ns
insert 100 nodes to B-tree of order 10: 80333 ns
insert 100 nodes to B-tree of order 100: 94500 ns
insert 100 nodes to B+ tree of order 4: 126542 ns
insert 100 nodes to B+ tree of order 10: 65709 ns
insert 100 nodes to B+ tree of order 100: 42875 ns
fast <----------insert----------> slow
RBt AVL B+(100) B+(10) B(10) B(100) B+(4) B(4)
search in AVL tree: 1625 ns
search in RBtree: 1500 ns
search in B-tree of order 4: 3208 ns
search in B-tree of order 10: 2042 ns
search in B-tree of order 100: 1584 ns
search in B+tree of order 4: 1958 ns
search in B+tree of order 10: 791 ns
search in B+tree of order 100: 708 ns
fast <----------search----------> slow
B+(100) B+(10) RBt B(100) AVL B+(4) B(10) B(4)
averagetime_insert:
{B(10)=82433 .71, B+(100)=34530 .4, B+(4)=168120 .84,
B(100)=86569 .96, RBt=32285 .04, B(4)=89461 .76,
AVL=38018 .79, B+(10)=74490 .4}
averagetime_search:
{B(10)=824 .33710000000001, B+(100)=345
.30400000000003,
B+(4)=1681 .2084, B(100)=865 .6996,
RBt=322 .85040000000004, B(4)=894 .6175999999999,
AVL=380 .1879, B+(10)=744 .904}
---
insert 1000 nodes to AVL tree: 457375 ns
insert 1000 nodes to RBtree: 401417 ns
insert 1000 nodes to B-tree of order 4: 885834 ns
insert 1000 nodes to B-tree of order 10: 900209 ns
insert 1000 nodes to B-tree of order 100: 845959 ns
insert 1000 nodes to B+ tree of order 4: 929208 ns
insert 1000 nodes to B+ tree of order 10: 589458 ns
insert 1000 nodes to B+ tree of order 100: 424625 ns
fast <----------insert----------> slow
RBt B+(100) AVL B+(10) B(100) B(4) B(10) B+(4)
search in AVL tree: 1916 ns
search in RBtree: 1959 ns
search in B-tree of order 4: 3291 ns
search in B-tree of order 10: 1458 ns
search in B-tree of order 100: 1750 ns
search in B+tree of order 4: 2250 ns
search in B+tree of order 10: 875 ns
search in B+tree of order 100: 416 ns
fast <----------search----------> slow
B+(100) B+(10) B(10) B(100) AVL RBt B+(4) B(4)
averagetime_insert:
{B(10)=859769 .16, B+(100)=590647 .05, B+(4)=955079 .98,
B(100)=1102582 .06, RBt=338150 .86, B(4)=847384 .17,
AVL=415199 .65, B+(10)=537789 .11}
averagetime_search:
{B(10)=8597 .6916, B+(100)=5906 .4705, B+(4)=9550 .7998,
B(100)=11025 .820600000001, RBt=3381 .5085999999997,
B(4)=8473 .8417, AVL=4151 .9965, B+(10)=5377 .8911}
---
insert 10000 nodes to AVL tree: 9670000 ns
insert 10000 nodes to RBtree: 8590000 ns
insert 10000 nodes to B-tree of order 4: 18246166 ns
insert 10000 nodes to B-tree of order 10: 20345250 ns
insert 10000 nodes to B-tree of order 100: 19274125 ns
insert 10000 nodes to B+ tree of order 4: 15648833 ns
insert 10000 nodes to B+ tree of order 10: 10381458 ns
insert 10000 nodes to B+ tree of order 100: 7639125 ns
fast <----------insert----------> slow
B+(100) RBt AVL B+(10) B+(4) B(4) B(100) B(10)
search in AVL tree: 1375 ns
search in RBtree: 2292 ns
search in B-tree of order 4: 5292 ns
search in B-tree of order 10: 3583 ns
search in B-tree of order 100: 3583 ns
search in B+tree of order 4: 4833 ns
search in B+tree of order 10: 3167 ns
search in B+tree of order 100: 792 ns
fast <----------search----------> slow
B+(100) AVL RBt B+(10) B(100) B(100) B+(4) B(4)
averagetime_insert:
{B(10)=1 .58285672E7, B+(100)=5938541 .24,
B+(4)=1 .303157998E7, B(100)=1 .541014462E7,
RBt=5652890 .43, B(4)=1 .334035707E7, AVL=7253795 .84,
B+(10)=7203287 .06}
averagetime_search:
{B(10)=158285 .672, B+(100)=59385 .4124,
B+(4)=130315 .79980000001, B(100)=154101 .4462,
RBt=56528 .904299999995, B(4)=133403 .5707,
AVL=72537 .9584, B+(10)=72032 .8706}
---
insert 100000 nodes to AVL tree: 146884792 ns
insert 100000 nodes to RBtree: 120585791 ns
insert 100000 nodes to B-tree of order 4: 277470166 ns
insert 100000 nodes to B-tree of order 10: 255380625 ns

insert 100000 nodes to B-tree of order 100: 245922541 ns
insert 100000 nodes to B+ tree of order 4: 369338292 ns
insert 100000 nodes to B+ tree of order 10: 133651708 ns
insert 100000 nodes to B+ tree of order 100: 107554834 ns
fast <----------insert----------> slow
B+(100) RBt B+(10) AVL B(100) B(10) B(4) B+(4)
search in AVL tree: 3125 ns
search in RBtree: 2875 ns
search in B-tree of order 4: 7791 ns
search in B-tree of order 10: 4834 ns
search in B-tree of order 100: 4041 ns
search in B+tree of order 4: 4958 ns
search in B+tree of order 10: 4166 ns
search in B+tree of order 100: 1208 ns
fast <----------search----------> slow
B+(100) RBt AVL B(100) B+(10) B(10) B+(4) B(4)
averagetime_insert:
{B(10)=2 .448318646E8, B+(100)=1 .2629613555E8,
B+(4)=2 .0513267075E8, B(100)=2 .4331373115E8,
RBt=1 .108997041E8, B(4)=2 .5701895625E8,
AVL=1 .2480924385E8, B+(10)=1 .416284334E8}
averagetime_search:
{B(10)=1 .224159323E7, B+(100)=6314806 .7775,
B+(4)=1 .02566335375E7, B(100)=1 .2165686557500001E7,
RBt=5544985 .205, B(4)=1 .28509478125E7,
AVL=6240462 .1925, B+(10)=7081421 .67}
---
insert 1000000 nodes to AVL tree: 1784261125 ns
insert 1000000 nodes to RBtree: 1420085958 ns
insert 1000000 nodes to B-tree of order 4: 3438281834 ns
insert 1000000 nodes to B-tree of order 10: 3788487625 ns
insert 1000000 nodes to B-tree of order 100: 3752132584 ns
insert 1000000 nodes to B+ tree of order 4: 2927331208 ns
insert 1000000 nodes to B+ tree of order 10: 2162745500 ns
insert 1000000 nodes to B+ tree of order 100: 1330336083 ns
fast <----------insert----------> slow
B+(100) RBt AVL B+(10) B+(4) B(4) B(100) B(10)
search in AVL tree: 9750 ns
search in RBtree: 23958 ns
search in B-tree of order 4: 73917 ns
search in B-tree of order 10: 72917 ns
search in B-tree of order 100: 44417 ns
search in B+tree of order 4: 101875 ns
search in B+tree of order 10: 5541 ns
search in B+tree of order 100: 2791 ns
fast <----------search----------> slow
B+(100) B+(10) AVL RBt B(100) B(10) B(4) B+(4)
averagetime_insert:
{B(10)=3 .2371293125E9, B+(100)=1 .815506875E9,
B+(4)=2 .615794021E9, B(100)=3 .246175146E9,
RBt=1 .315888708E9, B(4)=3 .1055527505E9,
AVL=1 .613829729E9, B+(10)=1 .7530696875E9}
averagetime_search:
{B(10)=1 .61856465625E9, B+(100)=9 .077534375E8,
B+(4)=1 .3078970105E9, B(100)=1 .623087573E9,
RBt=6 .57944354E8, B(4)=1 .55277637525E9,
AVL=8 .069148645E8, B+(10)=8 .7653484375E8}

IEEE754

SIDA LUO (Michael)

1. Introduction

IEEE754 is the most common representation today for real numbers on computers. Apart from simple numbers like 1 and 2, it can also represent floating point numbers, negative numbers and even other sorts of mathematical objects. However, there are a lot of calculations that can never be defined or be calculated mathematically, such as finding the square root of a negative number, or finding the angle with a sine value that is more than 1. As this international standard is such an interesting thing that helps with representing numbers and allows calculators to work, this research is going to focus on how IEEE754 works in the correct way with proper math, and the mathematical objects that it can represent. In addition to that, this research project would also find out what would happen if the user tries to do something that is mathematically wrong.

2. Methodology

The aim of this mini-research-project is to find out how the IEEE754 system works when it represents mathematical concepts. As a result, the most important data would be the theoretical features of IEEE754, which basically supports the research on all the aspects. As IEEE754 is a new concept for me, most of the data would be secondary data. Apart from that, this research also requires primary data, that are collected after using the IEEE754 and act as examples that show how IEEE754 works and show the situations when this kind of representation would not work. The quantitative data such as the results after using IEEE754 could be collected simply by finding websites that can generate the results, which is okay if the operand is an actual number or floating point number, but might not work if the operand is basically a mathematical expression. The only way for this research project to collect secondary data is look around the website, which might or might not be an efficient process, since it is easy to find articles about IEEE754, but the information needs to be analyzed or understood properly with no argument among them.

3. MAIN BODY

The IEEE Standard for Floating-Point Arithmetic (IEEE754) is a standard established by the Institute of Electrical and Electronics Engineers(IEEE) in 1985, and it is a standard for floating-point computation. Among all kinds of representations of floating-point numbers, IEEE754 is the most commonly used and the most efficient one. It is available on platforms such as Intel-based PC’s, Macs, and etc. IEEE754 is the most efficient way of representing floating-point numbers since it includes the following three basic components: 1.The Sign of Mantissa, which is a bit in front of the whole number in binary, if the number is positive, the bit would be 0, and it would be 1 if the number is negative. 2. The Biased exponent, which adds a bias to the exponent in order to get the stored exponent. 3. The Normalized Mantissa, which means that the left of the decimal point contains only one “1”. With all those components, the IEEE754 numbers are divided into, and they are called “Single Precision” and “Double Precision”. The single precision numbers have 32 bits in total, and they are distributed like this:

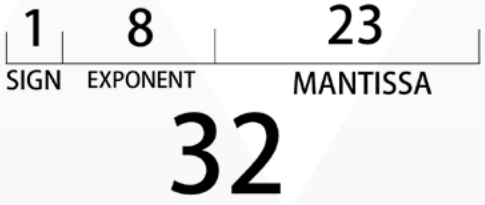


Figure1, the distribution of bits in a floating point number with single precision.

Similarly, there are 64 bits in total in a floating point number with 64 bits, and they are distributed like this:

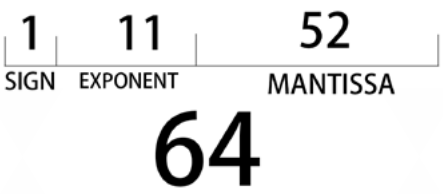


Figure2, the distribution of bits in a floating point number with double precision.

Those would be the basic structures of floating-point numbers that are represented in terms of IEEE754 standard. These structures have different biases for the exponent, and the bias is 127 for single precision, 1023 for double precision. Let's see how the standard work with an example. Take 10.1 as an example, in order to implement the IEEE754 standard, we need to convert the number into binary:

10 = 1010, 0.1 = 0.00011;
So, 10.1 = 1010.00011 = 1.01000011 * 2^3;
Because 10.1 is positive, the sign is positive;
So sign = 0.
In terms of single precision, the biased exponent is 127+3=130;
130 = 1000 0010;
Normalized mantissa = 01000011001100110011010
Thus, the IEEE754 single precision is 0 10000010 01000011001100110011010 for the number 10.1 in terms of binary.

How IEEE754 works is now known, but the special values that might be shown by IEEE754 hasn't been shown currently, so this research would be about the other sorts of mathematical content that is can represent. First of all, there are reserved values by IEEE754 that can ambiguity: 1. Zero, which can be both positive or negative, and this would be wrong mathematically. 2.Denormalized, which shows up if the exponent is 0 but the mantissa isn't. 3. Infinity, which has an exponent of 255 or 2049 in single or double precision when mantissa is 0. 4. Not A Number(NAN), which represents errors. Unfortunately, the converter that has been chosen doesn't support the inputs that are not numbers or decimal points. However, if any sort of mathematical is turned into a number, than it is possible for the converter to convert it into a IEEE754 number. As a result, if the programs that are available can identify the exact form of things like square roots and logarithmic number and so on, IEEE754 standard is fully capable with them. For example, sine 30。equals 0.5, and 0.5 can be turned into binary and into IEEE754 number, and it turns out to be 0 01111110 00000000000000000000000 with single precision.

According to both of the structures, there is always a position for the positive or negative sign of the number. As a result, it is possible for the appearance of +0 and -0, without having any error with both of them. However, if the user tries to divide by 0, calculate 0 * ±infinity, or calculate ±infinity / ±infinity, the result would always be NAN, which means "Not A Number".

4. BIBLIOGRAPHY

Source of information:
https://www.geeksforgeeks.org/ieee-standard-754-floating-point-numbers/
https://www.geeksforgeeks.org/introduction-of-floating-point-representation/
https://flylib.com/books/en/1.330.1.34/1/#:~:text=Almost%20any%20unnormalized%20value%20can%20be%20normalized%20by,exponent%2C%20you%20multiply%20the%20floating-point%20value%20by%20two.

IEEE converter:
https://www.toolhelper.cn/Digit/FractionConvert

Carrier Sense Multiple Access With Collision Detection

SHUOQI LI (Leo)

The full name of CSMA/CD is Carrier Sense Multiple Access with Collision Detection, which is carrier sense multiple access technology based on collision detection. CSMA/CD is also the core of the original 802.3, used in 10M/100M half-duplex wired network, now CSMA/CD application scenarios are much less, most of the directly based on full-duplex work. CSMA/CD is conceptually derived from 1-persistence CSMA, or 1-persistentes CSMA. On the basis of it, a Collision Detection mechanism is added. Conflict detection, that is, the mechanism of CD is mainly used to find conflicts and resolve conflicts. We describe how CSMA/CD works as follows: "Nodes need to continuously monitor the channel before sending data, and as soon as they discover that the channel is idle, they send data. While sending data, the node continuously listens to the channel and "detects" whether another node is also sending data at that time. If no other node's transmission is detected during the transmission process, the transmission is successful. After a successful transmission, the node needs to wait for the inter-frame interval IFG time before it can make the next transmission. If in the transmission process, detected other nodes are also transmitting, then the collision is detected. After the JAM occurs, the node immediately stops the current transmission, and sends a specific interference sequence to strengthen the collision to ensure that all other nodes detect the collision. After the jam sequence is sent, the node randomly selects a time countdown to backoff. When the backoff is complete, the node can try to retransmit again." In the above description, we can find that CSMA/CD is very similar to 1-persistent CSMA. Compared with the traditional aloha, CSMA/CD not only adds the LBT mechanism, but also introduces the collision detection mechanism to detect the collision immediately in the transmission, instead of relying on the feedback of ACK to determine whether there is a collision. Thus, the efficiency of the network is improved. In CSMA, we need to describe a few more details:

• carrier sense (CS) : In wired networks, CS actually means receiving information on the channel and parsing it. In this way, it judges whether there is a node transmitting information on the shared channel, so as to achieve the use of listening. The name carrier detection actually comes from AM/FM reception, that is, the carrier is an analog signal carrying modulation information, so that carrier sensing is to listen to whether there is an AM/FM signal.

• collision detection (CD) : In some theories, collision detection is introduced as sending data while receiving data on the same channel, and comparing sending data Tx with receiving data Rx. If Tx=Rx, no conflict occurs, and if Tx ≠ Rx, a conflict is identified. In some engineering introductions, the method introduced to detect conflicts is the "medium dependence method". The connection segment medium has separate paths for transmitting and receiving data, and collision detection is performed in the same receiver segment transceiver with the help of activities that occur on both the transmission and receiving data paths. On the coaxial cable medium, the transceiver detects the spike by detecting the DC signal level of the coaxial cable. When two or more base stations transmit simultaneously, the average DC voltage on the coaxial can reach the collision detection voltage level in the coaxial transceiver. The coaxial transceiver continuously detects the average voltage level on the coaxial cable and sends a JAM signal to the Ethernet interface if the average voltage level indicates that multiple base stations are simultaneously transmitting content. This process of sending the JAM signal takes longer than the collision detection time, and the extra time includes the time calculated from the total signal latency over a 10Mbps Ethernet network.

Time slot and capture effect: In wired networks, the capture effect is defined as no collision occurs within 1 time slot length, while the time slot is defined as the sending time of 512bit in 10 MBPS or 100 MBPS network. 512bit is chosen as the reference value for a slot in wired networks because of the maximum roundtrip time needed to send a signal to the peer. This maximum round trip time includes the round trip time of the electromagnetic wave through the physical layer and the time taken to transmit the JAM signal. If the node has captured the channel, that is, it has sent 512 bits, then the other party will not necessarily interrupt your current transmission. Make sure they can detect you, 2. And that

the feedback JAM signals get to you. At the same time, if from the point of view of electromagnetic wave transmission, 512bit transmission time in 10M converted into electromagnetic wave propagation distance is about 2800 meters, 512bit transmission time in 100M converted into electromagnetic wave propagation distance is about 200 meters, compared with the twisted pair length of wired network, these parameters are still acceptable. On and above 1000Mbps networks, where the default physical layer is in full-duplex mode, the slot length is defined as 512byte if CSMA/CD mode is used.

Retransmission mechanism (Backoff and BEB mechanism) : If the node detects that the collision occurred in the first 512 bytes, that is, within a time slot, then the node first backoff and retransmit. Here, the backoff algorithm is BEB (Binary Exponential Backoff Algorithm), that is, in a random window, a random number is selected and multiplied by the time slot to backoff. In the process of 0~10 backoff, the random window was doubled in each backoff. In the process of 11 ~ 16 times, backoff was still carried out and the data packet was resent, but the window size was not enlarged. If the 17th time failed, the packet was lost. According to the general case, the collision does not occur in the part after 512 bytes, that is, the capture effect has occurred, that is, the node has captured the channel. However, there are some cases, such as different steps in time, that cause the collision to occur after 512 bytes, so direct packet loss is chosen after the timeout.

The reason why Carrier Sense Multiple Access with Collision Detection is rarely seen nowadays is that most networks such as wireless network transmission have used full-duplex transmission protocol. What is full-duplex? Full-duplex When the data sending and receiving are separated and transmitted by two different transmission lines, both sides of the communication can transmit and receive at the same time, such a transmission mode is full duplex. In full-duplex mode, each end of the communication system is equipped with a transmitter and receiver, so that data can be controlled to transmit in both directions at the same time. Full-duplex mode does not need to switch directions, so there is no time delay caused by switching operation, which is very beneficial for interactive applications that cannot have time delay, such as remote monitoring and control systems. This method requires a transmitter and receiver on both sides of the communication, at the same time, two data lines are needed to transmit data signals.

For example, the host computer uses a serial interface to connect to the display terminal, and the display terminal has a keyboard. In this way, on the one hand, the characters entered on the keyboard are sent to the host memory; On the other hand, information from the host memory can be sent to the screen for display. Usually, after typing a character on the keyboard, it is not displayed, and the computer host receives the character, immediately sends it back to the terminal, and then the terminal displays the character. In this way, the previous character is sent back and the next character is input at the same time, that is, in full-duplex mode.

Therefore, this protocol does not produce data collisions in SCMA/CD.

CSMA/CD is a contention-based medium access control protocol. It originates from the content-type protocol adopted by the ALOHA network developed by the University of Hawaii, and has been modified to have higher medium utilization than ALOHA protocol. It is mainly used in field bus Ethernet. Another improvement is that for each station, as soon as it detects a collision, it drops its current transmission task. In other words, if both stations detect that the channel is idle and start transmitting at the same time, they will detect that a collision has occurred almost immediately. They should not continue to send their frames, as this will only generate garbage; Instead, they should stop transmitting as soon as a collision is detected. Quickly terminating corrupted frames saves time and bandwidth.

SCMA/CD seems to be an obsolete protocol at the beginning OF the presentation BUT even now SCMA/CD has some advantages when it comes to not using full duplex

1. It can improve the utilization of network bandwidth, because nodes can listen to the data transmission of other nodes while transmitting data, avoiding the occurrence of data collision, thereby improving the success rate of data transmission. 2. In the case of small network topology, the use of CSMA/CD can reduce the operation cost of the network and improve the reliability of the network.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

04

Biology

Fitzwilliam College Online Summer School Journal · 2023 Summer

Clostridium Difficile

ZHIHUAN ZHANG

Keywords: Clostridium difficile, Clostridium difficile infection, tcdA and tcdB, metronidazole, fidaxomicin, vancomycin, and fecal microbiota transplantation.

1. Introduction

Clostridium difficile (C. difficile) is a Gram-positive, spore-forming, anaerobic, toxin-producing bacillus, causing Clostridium difficile infection (CDI) and pseudomembranous colitis characterized by yellow-white plaques on the colonic mucosa (1, 2). Although C. difficile spores stay intrinsically intact under many antibiotics and alcoholic disinfectants, C. difficile is usually antagonized by colonization resistance of the gut microbiome, specifically the inhibition of secondary bile salts produced by Clostridium scindens, microbiota-derived antimicrobials, and so on (3, 4). However, the microbiota can be disrupted. Previous research has shown that C. difficile tends to prosper and induce antibiotic-associated diarrhea when certain antibiotics are administered (5). Notably, CDI tends to reoccur eight weeks after the first episode (rCDI) due to the persisting dysbiosis of the microbiota and limited levels of serum immunoglobins (6). C. difficile remains the leading cause of nosocomial diarrhea and has spread to communities consisting of young people (<65 years) without short-term hospitalization contact or high comorbidity scores, accounting for 41% of 385 definite CDI cases in a population-based cohort study (7, 8). The hypervirulent C. difficile BI/NAP1/027 strain causes 30,000 deaths and $4.8 billion in inpatient costs yearly in the United States (9). CDI clinical manifestations range from asymptomatic carrier stages, abdominal pains, mild diarrhea, fever, nausea and vomiting, dehydration, abdominal distension, hypoalbuminemia, pseudomembranous colitis, toxic megacolon, colon perforation, intestinal paralysis, kidney failure, systemic inflammatory response syndrome, septicemia, and death (10). Major risk factors for CDI include senior age (>65 years) and the use of antibiotics, such as ampicillin, amoxicillin, cephalosporins, clindamycin, and fluoroquinolones (11). In this review, we will interpret the virulence of the pathogen, symptoms, diagnosis, and treatments of CDI to summarize the current understanding and containment of CDI.

2. Pathogenicity

The pathogenicity of C. difficile partly depends on its ability to secrete toxins into epithelial intestinal cells via receptor-binding endocytosis (12). Most C. difficile expresses the pathogenicity locus (PaLoc), which is a 19.6 kb long DNA section, encodes two types of toxins, tcdA and tcdB, as well as two regulatory proteins, tcdR, tcdC, and holin-encoding gene tcdE (13) (Figure 1). Toxins A and B both inactivate Rho GTPases, independently leading to the pathogenicity of CDI (14). Or they work in a synergistic manner when tcdB concentration ≥10 ug/mL for rabbit small intestine samples (15). TcdB is eventually shown to be a more potent cytotoxin than enterotoxin tcdA. Riegler et al. found that tcdB at a low concentration of 3nM reduced the potential difference, short-circuit current, and resistance of human colonic mucosa to a similar extent caused by tcdA at a concentration of 32nM (76, 58, and 46% for tcdB, respectively; 76, 55, and 47% for tcdA, respectively) (16). Lyras and his colleagues showed that hamsters infected by tcdA+/tcdB- C. difficile were more likely to survive than those with tcdA-/tcdB+ strains (79% vs. 6%, P=.001) (17). Notably, the lack of anti-sigma factor tcdC negatively regulating the gene expression of PaLoc also contributes to the toxigenicity of C. difficile (18). Meanwhile, some strains of C. difficile possess a 6.2 kb long chromosomal site called CdtLoc coding for binary toxin C. difficile transferase (CDT), which depolarizes actin filaments and interferes with actin cytoskeletons in the colon (19). Moreover, the spore-forming lifecycle of C. difficile contributes to its virulence (20). Transmitted oral-fecally, C. difficile spores primarily survive the acidic gastric juice in the stomach, germinate or outgrow after being stimulated by bile acids in the small intestine, and colonize and produce toxins in the colon, thereby leading to the death of intestinal cells (21) (Figure 2)

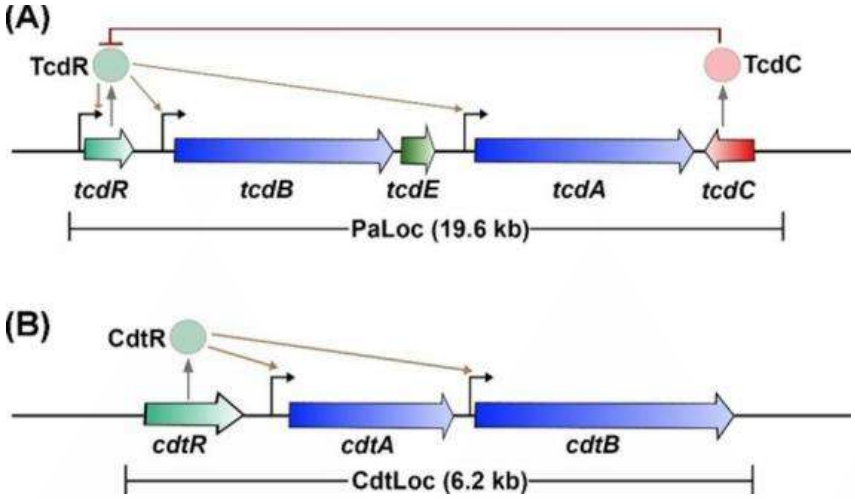


Figure 1. Organization of toxin genes. a. TcdA and tcdB are indicated by blue arrows;regulatory genes are in light green (tcdR) or red (tcdC); and holin-encoding gene tcdE is in dark green. The direction of the arrows reflects the direction of transcription. TcdR positively regulates expressions of itself, tcdA, and tcdB (brown arrows). TcdC is an anti-sigma factor that negatively regulates toxin expression. b. CdtA and cdtB are in blue. The positive regulatory gene cdtR is in light green. (Credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5812492/#bib119)

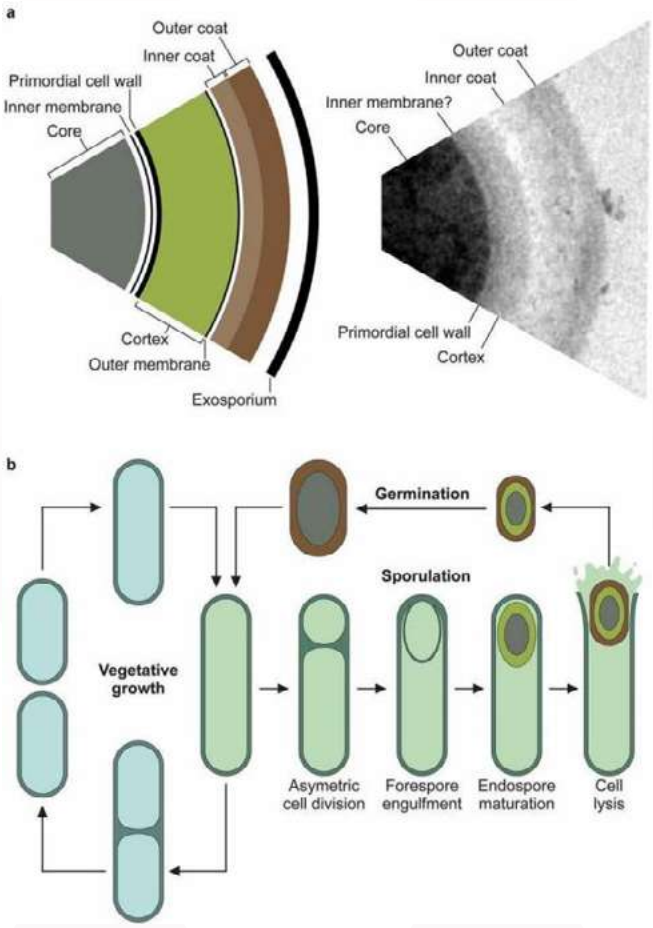


Figure 2. Structure of the spore and life cycle. a. The robust properties of the spore are due to its multi-layered structure, including the dehydrated dense core, the extremely impermeable inner membrane and the germ cell wall, the cortex, the second membrane, the protein coat, and the exosporium, from the inside to the outside. b. The step-by-step sporulation pathway includes asymmetric septation, genome transfer into the nascent spore, the forespore being engulfed by the mother cell, DNA compact, core dehydration, cortex and protein coat synthesis, and the lysis of

the mother cell. (Credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9815241/)

3. Diagnoses

The diagnosis of CDI is based on the presence of particular toxins, antigens, or pathogens in stool samples (22). The majority of detecting techniques include glutamate dehydrogenase immuno-enzymatic assays (EIA for GDH) with high NPV but interpretation errors, nucleic acid amplification testing (NAAT) with high sensitivity but the potential of overdiagnosis, cell cytotoxicity that is the gold standard, stool cytotoxicity assay (CTA) which is rarely used because of the lack of standards and long turnaround time (23). According to the European Society of Clinical Microbiology and Infectious Diseases (ESCMID), the reasonable algorithm of different methods obtains the most accurate result (23) (Figure 3, Table 1).

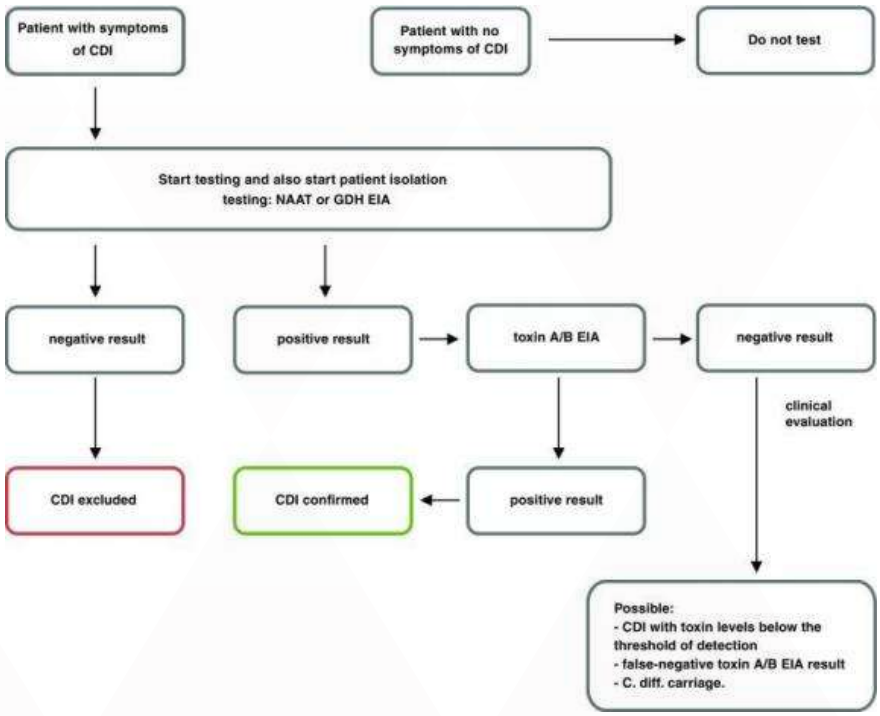


Figure 3. Flow chart of CDI diagnosis. (Credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6570665/)

Table 1. The percent sensitivity, percent specificity, and comment for immuno-enzymatic assays for glutamate dehydrogenase, nucleic acid amplification testing, cell cytotoxicity, toxigenic culture, immuno-enzymatic assays for toxins A and B, and colonoscopy.

Test	*Percent Sensitivity*	*Percent Specificity*	*Comment*
EIA for GDH	71-100%	67-99%	Compared w/stool culture
NAAT	88-100%	88-97%	Most rapid sensitive test, most expensive
Cell Cytotoxicity	77-86%	97-99%	Considered a gold standard
Toxigenic Culture	95-100%	96-100%	The more sensitive of the 2 " gold standards"
EIA for Toxins	67-92%	93-99%	vs. cell cytotoxicity
A and B	60-89%	93-99%	vs. toxigenic culture
Colonoscopy	50%	100%	For detection of pseudomembranous colitis and diagnostic sample

EIA for GDH: immuno-enzymatic assays for glutamate dehydrogenase; NAAT: nucleic acid amplification testing; EIA for toxins A and B: immuno-enzymatic assays for toxins A and B.

4. Containment

Once CDI is confirmed, the infection is controlled via quarantine, gloves and gowns, hand hygiene with soap and water, and terminal room cleaning (24). Treatments for CDI comprise traditional antibiotics, including metronidazole, fidaxomicin, vancomycin, and fecal microbiota transplantation (FMT), recommended by the Infectious Diseases Society of America (IDSA) and the Society for Healthcare Epidemiology of America (SHEA) (24). The first-line antibiotic for nonsevere CDI or severe rCDI cases in ESCMID guidelines is metronidazole diffusing into anaerobic bacteria and deactivating cysteine-bearing enzymes (31). A meta-analysis showed that metronidazole had clinical success rates of 72%, which was inferior to vancomycin (79%) (25). As a narrow-spectrum, bactericidal antibiotic for Gram-positive bacteria, fidaxomicin is more effective than vancomycin (71% vs. 61%) and reduces CDI recurrence percentage to 15 to 20% (25, 26). In severe cases, vancomycin is prescribed with a 25% chance of developing rCDI and takes bactericidal effects by attacking peptidoglycans in bacterial cell walls (26, 27). Fecal microbiota transplantation (FMT) can correct gut dysbiosis and induce probiotics by placing stool from healthy donors into patients' guts with low rCDI rates (28). A randomized controlled study illustrated that 81% of patients with ≥3 episodes of CDI had resolution after one FMT infusion, significantly higher curing rates than oral vancomycin alone (31%, $P<0.001$) and vancomycin and bowel lavage (23%) (29). However, FMT potentially transmits other infectious agents; therefore, step-by-step cautious preparation of stool and recipient is necessary for FMT effectiveness (30) (Table 2).

Table 2. The first choice, second choice, third choice, and nonantibiotic treatments for the first episode of nonsevere CDI, the second episode of CDI, the severe episode when oral treatment is not possible, the first recurrence of CDI, and multiple recurrences of CDI.

Episode	*First Choice*	*Second Choice*	*Third Choice*	*Nonantibiotic Treatment*
First episode of nonsevere CDI	Metronidazole orally 500 mg three times a day for 10 days	Vancomycin orally 125 mg four times a day for 10 days	Fidaxomicin orally 200 mg two times a day for 10 days	For mild cases; stop inducing antibiotics and observe clinical response at 48 hours
Severe episode of CDI	Vancomycin orally 125 mg four times a day for 10 days	Fidaxomicin orally 200 mg two times a day for 10 days		In case of colon perforation or severe systemic inflammation, surgery is indicated
Severe episode when oral treatment is not possible	Metronidazole 500 mg three times a day and oral vancomycin 500 mg four times a day for 10 days			In case of colon perforation or severe systemic inflammation, abdominal surgery is indicated
First recurrence of CDI	Vancomycin orally 125 mg four times a day for 10 days	Fidaxomicin orally 200 mg two times a day for 10 days		
Multiple recurrences of CDI	Fidaxomicin orally 200 mg two times a day for 10 days	Vancomycin orally 125 mg four times a day for 10 days, followed by vancomycin pulse strategy or taper strategy		FMT added to antibiotic treatment

CDI: C. difficile infection.

5. Conclusions

In summary, we reviewed the pathogenicity of Clostridium difficile, which is attributable to PaLoc and CdtLoc gene expression, the resistance of spores, and oral-fecal transmission, various symptoms, ranging from mild diarrhea to death, risk factors, such as age and antibiotic therapy, four types of diagnosis, and two groups of treatments of CDI. However, current therapies reveal different issues and require improvement. Specifically, fewer antibiotics should be applied. FMT is a promising solution. Moreover, scientists should establish a global monitoring network to summarize existing antibiotics and inspire new ones.

6. References

1. Czepiel J, DróżdżM, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, Goldman S,Wultańska D, Garlicki A, Biesiada G. Clostridium difficile infection: review. Eur J Clin Microbiol Infect Dis. 2019 Jul;38(7):1211- 1221. doi: 10. 1007/s10096-019-03539-6. Epub 2019 Apr 3. PMID: 30945014.

2. Farooq PD, Urrunaga NH, Tang DM, von Rosenvinge EC. Pseudomembranous colitis.
Dis Mon. 2015 May;61(5):181-206. doi: 10. 1016/j.disamonth. 2015.01.006. Epub 2015 Mar 11. PMID: 25769243.

3. Paredes-Sabja D, Shen A, Sorg JA. Clostridium difficile spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol. 2014 Jul;22(7):406- 16. doi: 10. 1016/j.tim.2014.04.003. Epub 2014 May 7. PMID: 24814671.

4. Sorbara MT, Pamer EG. Interbacterial mechanisms of colonization resistance and the
strategies pathogens use to overcome them. Mucosal Immunol. 2019 Jan;12(1):1-9. doi: 10. 1038/s41385-018-0053-0. Epub 2018 Jul 9. Erratum in: Mucosal Immunol. 2019 May;12(3):840. PMID: 29988120.

5. Guo Q, Goldenberg JZ, Humphrey C, El Dib R, Johnston BC. Probiotics for the prevention of pediatric antibiotic-associated diarrhea. Cochrane Database Syst Rev. 2019 Apr 30;4(4):CD004827. doi: 10. 1002/14651858.CD004827. pub5. PMID: 31039287.

6. Deshpande A, Pasupuleti V, Thota P, Pant C, Rolston DD, Hernandez AV, Donskey CJ, Fraser TG. Risk factors for recurrent Clostridium difficile infection: a systematic review and meta-analysis. Infect Control Hosp Epidemiol. 2015 Apr;36(4):452-60. doi: 10. 1017/ice.2014.88. Epub 2015 Jan 28. PMID: 25626326.

7. Cunha BA. Nosocomial diarrhea. Crit Care Clin. 1998 Apr;14(2):329-38. doi: 10. 1016/s0749-0704(05)70398-5. PMID: 9561820.

8. Khanna S, Pardi DS, Aronson SL, Kammer PP, Orenstein R, St Sauver JL, Harmsen WS,Zinsmeister AR. The epidemiology of community-acquired Clostridium difficile infection: a population-based study. Am J Gastroenterol. 2012 Jan;107(1):89-95. doi: 10. 1038/ajg.2011.398. Epub 2011 Nov 22. Erratum in: Am J Gastroenterol. 2012 Jan;107(1):150. PMID: 22108454.

9. McCormick WL, Jackson G, Andrea SB, Whitehead V, Chargualaf TL, Touzard-Romo F. Impact of mandatory nucleic acid amplification test (NAAT) testing approval on hospital-onset Clostridioides difficile infection (HO-CDI) rates: A diagnostic stewardship intervention. Infect Control Hosp Epidemiol. 2023 Jul 10:1-4. doi: 10. 1017/ice.2023.92. Epub ahead of print. PMID: 37424227.

10. Abad CLR, Safdar N. A Review of Clostridioides difficile Infection and Antibiotic - Associated Diarrhea. Gastroenterol Clin North Am. 2021 Jun;50(2):323-340. doi: 10. 1016/j.gtc.2021.02.010. PMID: 34024444.

11. Pépin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, Leblanc M, Rivard G, Bettez M, Primeau

V, Nguyen M, Jacob CE, Lanthier L. Emergence of fluoroquinolones as the predominant risk factor for Clostridium difficile-associated diarrhea: a cohort study during an epidemic in Quebec. Clin Infect Dis. 2005 Nov 1;41(9):1254-60. doi: 10. 1086/496986. Epub 2005 Sep 20. PMID: 16206099.

12. Chandrasekaran R, Lacy DB. The role of toxins in Clostridium difficile infection. FEMS Microbiol Rev. 2017 Nov 1;41(6):723-750. doi: 10. 1093/femsre/fux048. PMID: 29048477.

13. Buddle JE, Fagan RP. Pathogenicity and virulence of Clostridioides difficile. Virulence. 2023 Dec;14(1):2150452. doi: 10. 1080/21505594.2022.2150452. PMID: 36419222.

14. Leffler DA, Lamont JT. Clostridium difficile infection. N Engl J Med. 2015 Apr 16;372(16):1539-48. doi: 10. 1056/NEJMra1403772. PMID: 25875259.

15. Lima AA, Lyerly DM, Wilkins TD, Innes DJ, Guerrant RL. Effects of Clostridium difficile toxins A and B in rabbit small and large intestine in vivo and on cultured cells in vitro. Infect Immun. 1988 Mar;56(3):582-8. doi: 10. 1128/iai.56.3.582-588.1988. PMID: 3343050.

16. Riegler M, Sedivy R, Pothoulakis C, Hamilton G, Zacherl J, Bischof G, Cosentini E, Feil W, Schiessel R, LaMont JT, et al. Clostridium difficile toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. J Clin Invest. 1995 May;95(5):2004- 11. doi: 10. 1172/JCI117885. PMID: 7738167.

17. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R, Adams V, Vedantam G, Johnson S, Gerding DN, Rood JI. Toxin B is essential for virulence of Clostridium difficile. Nature. 2009 Apr 30;458(7242):1176-9. doi: 10. 1038/nature07822. Epub 2009 Mar 1. PMID: 19252482.

18. Carter GP, Douce GR, Govind R, Howarth PM, Mackin KE, Spencer J, Buckley AM, Antunes A, Kotsanas D, Jenkin GA, Dupuy B, Rood JI, Lyras D. The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of Clostridium difficile. PLoS Pathog. 2011 Oct;7(10):e1002317. doi:10. 1371/journal.ppat. 1002317. Epub 2011 Oct 13. PMID: 22022270.

19. Ahn SW, Lee SH, Kim UJ, Jang HC, Choi HJ, Choy HE, Kang SJ, Roh SW. Genomic characterization of nine Clostridioides difficile strains isolated from Korean patients with Clostridioides difficile infection. Gut Pathog. 2021 Sep 16;13(1):55. doi: 10. 1186/s13099-021-00451-3. PMID: 34530913.

20. Martin JS, Monaghan TM, Wilcox MH. Clostridium difficile infection: epidemiology, diagnosis and understanding transmission. Nat Rev Gastroenterol Hepatol. 2016 Apr;13(4):206- 16. doi: 10. 1038/nrgastro.2016.25. Epub 2016 Mar 9. PMID: 26956066.

21. Buddle JE, Fagan RP. Pathogenicity and virulence of Clostridioides difficile. Virulence. 2023 Dec;14(1):2150452. doi: 10. 1080/21505594.2022.2150452. PMID: 36419222.

22. Carroll KC, Mizusawa M. Laboratory Tests for the Diagnosis of Clostridium difficile. Clin Colon Rectal Surg. 2020 Mar;33(2):73-81. doi: 10. 1055/s-0039-3400476. Epub 2020 Feb 25. PMID: 32104159.

23. Gateau C, Couturier J, Coia J, Barbut F. How to: diagnose infection caused by Clostridium difficile. Clin Microbiol Infect. 2018 May;24(5):463-468. doi: 10. 1016/j.cmi.2017. 12.005. Epub 2017 Dec 18. PMID: 29269092.

24. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, Dubberke ER, Garey KW, Gould CV, Kelly C, Loo V, Shaklee Sammons J, Sandora TJ, Wilcox MH. Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare

Epidemiology of America (SHEA). Clin Infect Dis. 2018 Mar 19;66(7):e1-e48. doi: 10. 1093/cid/cix1085. PMID: 29462280.

25. Nelson RL, Suda KJ, Evans CT. Antibiotic treatment for Clostridium difficile-associated diarrhoea in adults. Cochrane Database Syst Rev. 2017 Mar 3;3(3):CD004610. doi: 10.1002/14651858.CD004610.pub5. PMID: 28257555.

26. Oksi J, Anttila VJ, Mattila E. Treatment of Clostridioides (Clostridium) difficile infection. Ann Med. 2020 Feb-Mar;52(1-2):12-20. doi: 10. 1080/07853890.2019. 1701703. Epub 2019 Dec 13. PMID: 31801387.

27. Patel S, Preuss CV, Bernice F. Vancomycin. 2023 Mar 24. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan – . PMID: 29083794.

28. Wang JW, Kuo CH, Kuo FC, Wang YK, Hsu WH, Yu FJ, Hu HM, Hsu PI, Wang JY, Wu DC. Fecal microbiota transplantation: Review and update. J Formos Med Assoc. 2019 Mar;118 Suppl 1:S23-S31. doi: 10. 1016/j.jfma.2018.08.011. Epub 2018 Sep 1. PMID: 30181015.

29. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JF, Tijssen JG, Speelman P, Dijkgraaf MG, Keller JJ. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med. 2013 Jan 31;368(5):407- 15. doi: 10. 1056/NEJMoa1205037. Epub 2013 Jan 16. PMID: 23323867.

30. Dinleyici M, Vandenplas Y. Clostridium difficile Colitis Prevention and Treatment. Adv Exp Med Biol. 2019;1125:139-146. doi: 10. 1007/5584_2018_322. PMID: 30689174.

Discuss the Pathogenicity and Clinical Relevance of the Following Bacteria: Mycobacterium Tuberculosis

YUNZHEN GONG (Jenica)

Discuss the pathogenicity and clinical relevance of the following bacteria: Mycobacterium tuberculosis.

Mycobacterium tuberculosis (Mtb) is an intracellular bacterial pathogen first discovered in 1882 by Robert Koch (19). As the etiological agent of tuberculosis (TB), Mtb is a leading cause of human disease and death (5). According to the report released by WHO, an estimated 10 · 6 million people and 10 · 1 people became ill with TB in 2021 and 2020 respectively (24) (Fig. 1).

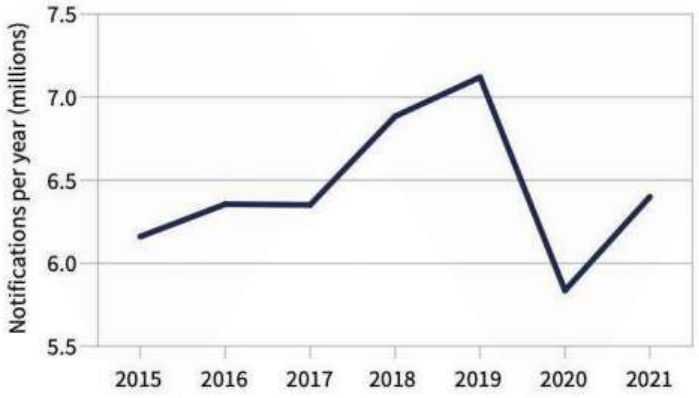


Fig. 1 "Global trend in case notifications of people newly diagnosed with TB, 2015-2021" (World Health Organization)

Mtb most commonly affects the lungs, but it can also affect other parts of the body, including the kidney, spine, and brain. There are two TB-related conditions: latent tuberculosis infection (LTBI) and TB disease, which can be fatal if not treated properly (15). This essay aims to review previous and recent research into Mtb in terms of pathogenicity and clinical relevance, with a focus on its structure, properties, toxin, and transmission, as well as providing an overview of the diagnosis and symptoms of TB disease, along with its treatment and antimicrobial resistance.

1. Pathogenicity

1.1. Classification and Structure

Mtb is a non-spore-forming, aerobic, rod-shaped bacilli that is neither gram-positive nor gram-negative due to its cell wall that is rich in lipids and waxes, primarily mycolic acid, which makes the cells impervious to Gram staining. Acid-fast techniques are used instead where Mtb appears bright red to intensive purple with green or blue background (1). The Mtb cell wall is comprised of mainly mycolic acid, cord factor, and Wax-D, and consists of two segments: the outer part and core of the cell wall (7). The core of the cell wall is made up of peptidoglycan (PG), covalently attached with arabinogalactan (AG) and mycolic acids subsequently, forming the mycolyl arabinogalactan-peptidoglycan (mAGP) complex. The upper part is composed of free lipids which are linked with fatty acids. Mostly this part is made up of different cell wall proteins, the Phosphatidylmyo-inositol mannosides (PIMs), Lipomannan (LM), and Lipoarabinomannan (LAM). These proteins along with lipids and glycoconjugate lipids act as effector molecules of the signaling process, and the insoluble core is essential for the viability of the cell (8) (6).

1.2. Toxins and Virulence factors

Mtb does not produce a known exotoxin or endotoxin. The tuberculosis necrotizing toxin (TNT), which is the C-terminal domain of the outer membrane protein CpnT, a NAD+glycohydrolase, is its key cytotoxicity component in

macrophages. Aside from the C-terminal, CpnT contains another terminal called the N-terminal, which is essential for TNT secretion (9). TNT reaches the cytoplasm of infected cells after being released and triggers necroptotic cell death by depleting NAD+ (2).

The key virulence factors of Mtb, rather than toxin, are associated with its cell wall composition and ability to resist the host immune response, including the mycolic acid, LAM, and ESX 1 secretion system (26). Mycolic acid generates a waxy layer around the bacteria that protects it against several host factors as well as many antibiotics, including beta-lactamases. LAM's biological activities include, inhibition of interferon-gamma-mediated macrophage activation, induction of macrophage cytokine production and release, and suppression of T-cell proliferation (27). The ESX 1 secretion system aids in the secretion of proteins and enzymes which help in phagosome survival and mediates the delivery of bacterial products such as TNT into the macrophage cytoplasm (2).

1.3. Transmission

Mtb is mainly transmitted through the air. When an infected individual coughs, sneezes, talks, or spits, saliva droplets containing Mtb are propelled into the air and can be inhaled by a neighboring person. These tiny particles then enter the lower airways, followed by phagocytosis of the Mtb, intracellular multiplication, latent contained phase of infection, and finally the active lung infection (10).

2. Clinical relevance

2.1. Prevalence

As mentioned above, the disease caused by Mtb—which typically affects the lungs— is tuberculosis. Young adults, people living in developing countries, and people with weakened immune systems, such as those who smoke or have HIV, are at higher risk of contracting TB (16). The world health organization estimates that among 10.6 million individuals worldwide contracted TB in 2021, 6.7% of them were HIV-positive individuals.(25) (Fig. 2).

		Number of cases (in thousands)					
		Total			HIV-positive		
Region or country group	Population	Best estimate	Low	High	Best estimate	Low	High
African Region	1 160 000	2 460	2 180	2 760	485	420	555
Region of the Americas	1 030 000	309	287	332	32	30	35
South-East Asia Region	2 060 000	4 820	4 340	5 320	102	84	121
European Region	931 000	230	211	251	29	25	33
Eastern Mediterranean Region	767 000	860	687	1 050	17	12	22
Western Pacific Region	1 930 000	1 890	1 490	2 320	38	27	50
High TB burden countries	4 870 000	9 180	8 470	9 920	568	500	641
Global	**7 880 000**	**10 600**	**9 850**	**11 300**	**703**	**633**	**776**

Fig. 2 “Global and regional estimates of TB incidence, numbers (in thousands) and rates (per 100 000 population) in 2021” (World Health Organization)

2.2. Diagnosis

Tuberculosis skin tests (TST) and Interferon-Gamma Release Assays (IGRA) are the standard methods for identifying infected individuals (13). A positive TST or IGRA only indicates that an individual has been infected with Mtb so additional tests such as a chest x-ray and sputum analysis are needed to determine whether the infection is latent (3).

Moreover, patients with positive TB tests should also be tested for HIV within 2 months of diagnosis since it is a major risk factor (12).

2.3. Treatment and Antimicrobial resistance

The type of treatment applied depends on whether LTBI or active TB is detected. Isoniazid, rifampin, rifabutin, rifapentine, pyrazinamide, and ethambutol are common drugs used to treat tuberculosis (11). The standard drug therapy regimen for latent TB infection typically involves the use of a single antibiotic, usually isoniazid or rifabutin, for a specified duration, which lasts 9 months for most adults and infants and 6 months for certain individuals, such as those without HIV infection (17) (18). The standard treatment for TB disease involves a combination of antibiotics, typically referred to as Directly Observed Therapy. In the initial phase of treatment which lasts for two months, isoniazid, rifampin, pyrazinamide, and ethambutol are administered simultaneously. After the initial phase, the continuation phase which typically involves isoniazid and rifampin begins, lasting for an additional four months (28).

Three strains of TB are resistant to the drugs mentioned above: multi-drug resistant TB which is resistant to both isoniazid and rifampicin, which are first-line drugs; Extensively drug-resistant TB which is resistant to isoniazid, rifampicin, a fluoroquinolone, and at least one of three injectable second-line drugs (23); Total drug-resistant TB where the TB strain is resistant to all first-line and second-line drugs (21). Although Total drug-resistant TB is not yet recognized by the WHO, extreme cases like this still pose a significant challenge to TB control and treatment.

2.4. Vaccine

Bacille Calmette-Guérin (BCG) is currently the only licensed vaccine for TB disease. The BCG vaccine is an attenuated strain of Mycobacterium bovis that activates the innate immune system and induces changes in the pattern of histone modifications of specific genes in innate immune cells (14). However, the BCG vaccine only partially protects infants and children from severe types of TB, but efficacy in adults and certain strains varies (20). Therefore, to implement the WHO’s "End TB strategy", a vaccine that is more consistently efficacious than BCG in both adolescents and adults is required. However the design of a new replacement or booster vaccine is difficult, because of a lack of understanding to the variable efficacy of BCG and the complex host-pathogen interaction, as well as a lack of animal models that accurately represent the human heterogeneous response against Mtb (4).

3. References

1. Kalscheuer R, Palacios A, Anso I, Cifuente J, Anguita J, Jacobs WR, et al. The Mycobacterium tuberculosis capsule: a cell structure with key implications in pathogenesis. Biochemical Journal [Internet]. 2019 Jul 18;476(14):1995 – 2016. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6698057/

2. Pajuelo D,Tak U, Zhang L, Danilchanka O, Tischler AD, Niederweis M. Toxin secretion and trafficking by Mycobacterium tuberculosis. Nature Communications [Internet]. 2021 Nov 15 [cited 2022 May 9];12(1):6592. Available from: https://www.technologynetworks.com/immunology/news/revealing-how-tuberculosis-toxin-is-secreted-and-trafficked-356184

3. Brodie D, Schluger NW. The Diagnosis of Tuberculosis. Clinics in Chest Medicine. 2005 Jun;26(2):247 – 71.

4. Kaufmann SHE. Is the development of a new tuberculosis vaccine possible? Nature Medicine [Internet]. 2000 Sep 1 [cited 2022 Apr 24];6(9):955 – 60. Available from: https://www.nature.com/articles/nm0900_955

5. Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, et al.Physiology of Mycobacteria. Advances in Microbial Physiology [Internet]. 2009;81 – 319. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/

PMC3728839/

6. Maitra A, Munshi T, Healy J, Martin LT, Vollmer W, Keep NH, et al. Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen. FEMS Microbiology Reviews [Internet]. 2019 Jun 10;43(5):548 – 75. Available from: https://academic.oup.com/femsre/article/43/5/548/5513445

7. Alderwick LJ, Birch HL, Mishra AK, Eggeling L, Besra GS. Structure, function and biosynthesis of the Mycobacterium tuberculosis cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets. Biochemical Society Transactions. 2007 Oct 25;35(5):1325 – 8.

8. Brennan PJ. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis. 2003 Feb;83(1-3):91 – 7.

9. Izquierdo Lafuente B, Ummels R, Kuijl C, Bitter W, Speer A. Mycobacterium tuberculosis Toxin CpnT Is an ESX-5 Substrate and Requires Three Type VII Secretion Systems for ntracellular Secretion. mBio. 2021 Apr 27;12(2).

10. Churchyard G, Kim P, Shah NS, Rustomjee R, Gandhi N, Mathema B, et al. What We Know About Tuberculosis Transmission: An Overview. The Journal of Infectious Diseases. 2017 Oct 1;216(suppl_6):S629 – 35.

11. Cruz-Knight W, Blake-Gumbs L. Tuberculosis. Primary Care: Clinics in Office Practice. 2013 Sep;40(3):743 – 56.

12. Agyeman AA, Ofori-Asenso R. Tuberculosis—an overview. Journal of Public Health and Emergency. 2017 Jan 6;1:7 – 7.

13. Horsburgh CR, Barry CE, Lange C. Treatment of Tuberculosis. Longo DL, editor. New England Journal of Medicine. 2015 Nov 26;373(22):2149 – 60.

14. Covián C, Fernández-Fierro A, Retamal-Díaz A, Díaz FE, Vasquez AE, Lay MK, et al. BCG-Induced Cross-Protection and Development of Trained Immunity: Implication for Vaccine Design. Frontiers in Immunology [Internet]. 2019 Nov 29;10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6896902/

15. Chai Q, Zhang Y, Liu CH. Mycobacterium tuberculosis: An Adaptable Pathogen Associated With Multiple Human Diseases. Frontiers in Cellular and Infection Microbiology. 2018 May 15;8(158).

16. Narasimhan P, Wood J, MacIntyre CR, Mathai D. Risk Factors for Tuberculosis. Pulmonary Medicine [Internet]. 2013;2013(828939):1 – 11. Available from: https://dx.doi.org/10.1155%2F2013%2F828939

17. Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent Mycobacterium tuberculosis Infection. Campion EW, editor. New England Journal of Medicine. 2015 May 28;372(22):2127 – 35.

18. Huaman MA, Sterling TR. Treatment of latent tuberculosis infection – An Update. Clinics in chest medicine [Internet]. 2019 Dec 1;40(4):839 – 48. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7043866/

19. Cambau E, Drancourt M. Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases [Internet].2014;20(3):196 – 201. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24450600

20. Davenne T, McShane H. Why don't we have an effective tuberculosis vaccine yet? Expert Review of Vaccines [Internet]. 2016 May 3;15(8):1009 – 13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4950406/

21. Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J, ZiaZarifi AH, et al. Emergence of New Forms of Totally

Drug-Resistant Tuberculosis Bacilli. Chest. 2009 Aug;136(2):420 – 5.

22. WHO editors of. Tuberculosis: Totally drug-resistant TB [Internet]. www.who.int. 2015. Available from: https://www.who.int/news-room/questions-and-answers/item/tuberculosis- totally-drug-resistant-tb

23. Palomino J, Martin A. Drug Resistance Mechanisms in Mycobacterium tuberculosis. Antibiotics [Internet]. 2014 Jul 2;3(3):317 – 40. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4790366/

24. COVID- 19 and TB [Internet]. www.who.int. Available from:

https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis- report-2022/covid-19-and-tb

25. 3. 1 Case notifications [Internet]. www.who.int. [cited 2023 Aug 3]. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022/tb-diagnosis-treatment/3- 1-case-notifications

26. Bhuwan M, Arora N, Sharma A, Khubaib M, Pandey S, Chaudhuri TK, et al. Interaction of Mycobacterium tuberculosis Virulence Factor RipA with Chaperone MoxR1 Is Required for Transport through the TAT Secretion System. mBio. 2016 Mar 1;7(2).

27. Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, Santangelo M de la P, et al. Virulence factors of theMycobacterium tuberculosiscomplex. Virulence [Internet]. 2013 Jan;4(1):3 – 66. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3544749/

28. CDC. Treatment for TB Disease [Internet]. CDC. 2019. Available from: https://www.cdc.gov/tb/topic/treatment/tbdisease.htm

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

05

Chemistry

Fitzwilliam College Online Summer School Journal · 2023 Summer

Non-Covalent Interactions in Drug Delivery Systems: Delivery of mRNA Vaccines

ANG LI (Max)

1. Discussion

There are more and more applications of non-covalent interactions in the area of drug delivery, one example is in the delivery of mRNA vaccine. mRNA vaccine is one of the various vaccines developed during the covid-19. Now the developed mRNA-based coronavirus vaccine from Pfizer and BioNTech has been given approval from the US Food and Drug Administration.

To protect the weak mRNA molecules from the repulsion between mRNA molecules and cell membrane and the easy degradation of mRNA molecules, a delivery vector is developed, called lipid nanoparticles (LNPs). This vector consists of four components, an ionizable lipid, cholesterol, a helper phospholipid and a PEGylated lipid.

Among them, the most important component is the ionizable lipid. When preparing LNPs in vitro, it is under acidic environment. There will be protons in the solution, which will combine with the amino group in the ionizable lipid, make it charged positively (proton combines with the lone pair on nitrogen atom by dative bond). This positively charged ionizable lipid can form electrostatic interaction with the negatively charged phosphate group on the mRNA. Now they are combined in vitro.

LNPs have an outer membrane structure similar to the cell membrane consisting of the above four components. There are some strong covalent bonds, but also a hydrophobic interaction between the hydrophobic carbon chains of the three lipid components and the cholesterol which all contain hydrophobic part. Through this hydrophobic interaction and some other covalent bonds, the whole LNPs are able to be stable.

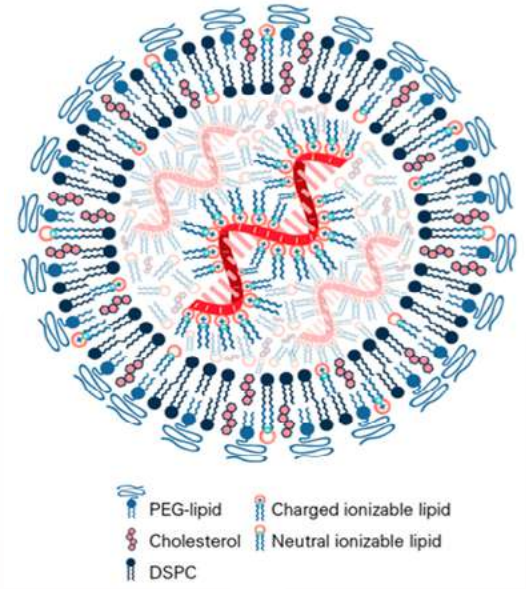
The above operation is all under acidic environment. To make it suitable for human body’s environment, the solution will carry out diffusion with neutral solution against a semipermeable membrane, to make the solution containing LNPs become neutral.

When the LNPs carrying mRNA molecules enter human body, the whole particles will enter target cell by endocytosis. After endocytosis, the particles inside cell will be surrounded by a layer of cell membrane, the whole structure is called endosome. Inside the endosome, the solution environment is becoming more and more acidic. This provides protons for the ionizable lipids in the outer layer of LNPs. Then the ionizable lipids are positively charged again, which enable them to form electrostatic interaction with the negatively charged phosphate group of the membrane of endosome. After the fusion, the endosome is opened and mRNA molecules are able to come out to the cytoplasm.

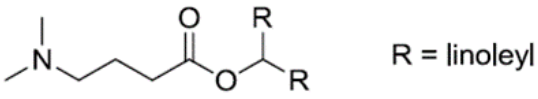
As mRNA molecules enter the target cell, they will be translated to produce proteins, which are the antigen of the certain virus. Then the immune response can be triggered to produce memory cells for long period precaution.

For all the non-covalent interactions in this topic, no matter in the outer membrane of LNPs or between the ionizable lipids and mRNA molecules, they all show the characteristics of this kind of interactions: easy to form and easy to break.

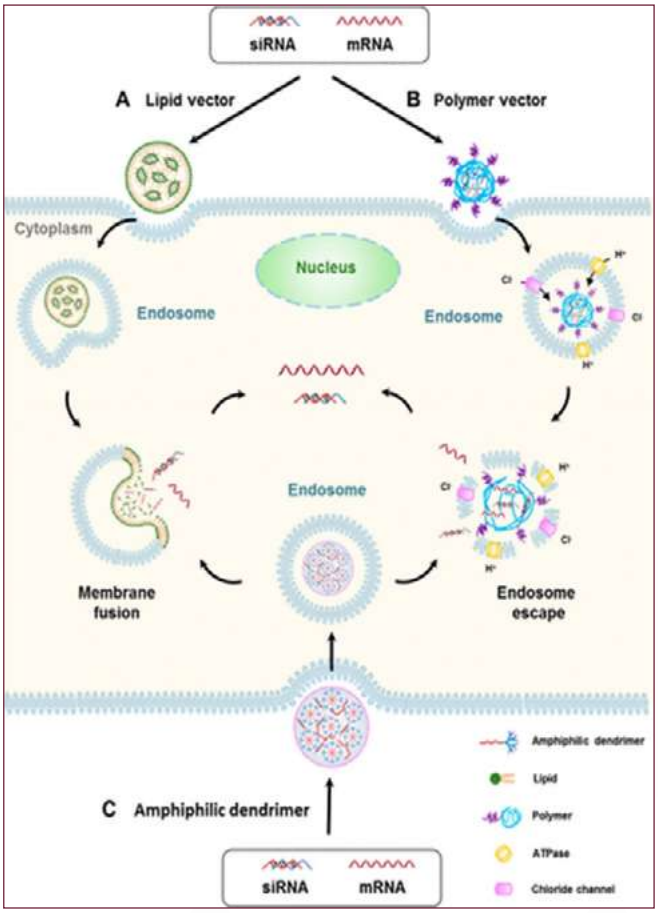
The delivery of mRNA molecules is just a temporary process. The combination between LNPs and mRNA needs to be easy to form, the mRNA molecules inside the target cells need to be easy to release. This shows the significance of supramolecular chemistry.



Structure of LNPs encapsulating mRNA molecules



Structure of one popular ionizable lipid (R groups are hydrophobic carbon chains)



The process of endocytosis (the left side of the picture is the endocytosis of LNPs)

2. Reference

1. Aldosari, B.N, Alfagih, I.M, Almurshedi, A.S. (2021) Lipid Nanoparticles as Delivery Systems for RNA-Based Vaccines. Pharmaceutics. Available from: https://www.mdpi.com/1999-4923/13/2/206

2. Buschmann, M.D et al. (2021) Nanomaterial Delivery Systems for mRNA Vaccines. Vaccines. Available from: https://www.mdpi.com/2076-393X/9/1/65

3. Katy, M (2020) Cell Membrane. Biology dictionary. Available from: https://biologydictionary.net/cell-membrane/

4. Jayaraman, M. et, al. (2012) Maximizing the Potency of siRNA Lipid Nanoparticles for Hepatic Gene Silencing In Vivo. Angewandte Chemie. Available from: DOI: 10.1002/anie.201203263

5. Chen, J. et al. (2022) Amphiphilic Dendrimer Vectors for RNA Delivery: State-of-the-Art and Future Perspective. Accounts of material research. Available from: Acc. Mater. Res. 2022, 3, 5, 484 – 497 https://doi.org/10.1021/accountsmr.1c00272

Supermolecular Inks: Novel applications in printing and beyond.

Naminrina

Supramolecular Inks

NAMINRINA

BACKGROUND: The problem

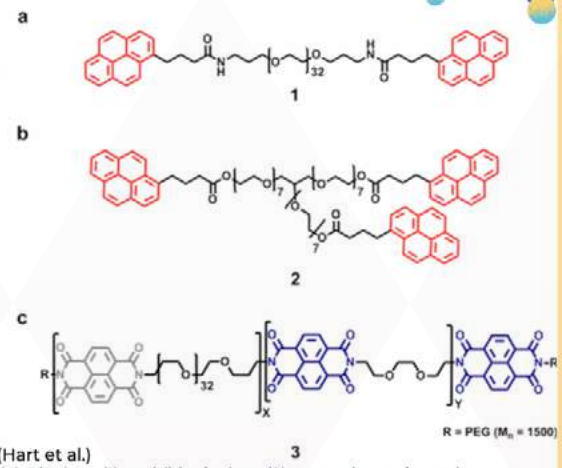
Usually, continuous inkjet or drop-demand technology uses ink contains molecules with high molecular-weighted polymers. However, these high molecular-weighted polymers would cause the printer nozzle to block resulting in lower efficiency for the consumer.

INTRODUCTION: Solution

Supramolecular ink contains molecules with low molecular weight are used as substitute by material scientists to overcome this issue. Low molecular weight can more easily self-assemble on substrate surfaces due to the promotion of supramolecular chemistry. Resulting polymers that have the same chemical and physical property as the high molecular-weighted polymers. They would also print out a lot of vibrantly colored materials and take smaller volumes than high molecular-weighted polymers in the pigments and dyes would.

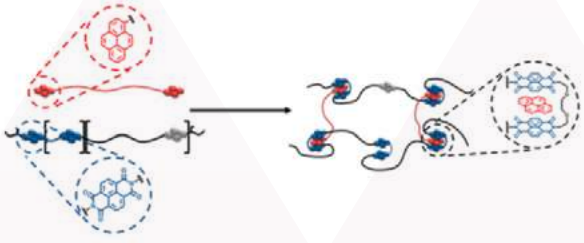
EXPERIMENT: Further explanation of the solution

Supramolecular ink contains molecules with low molecular weight are used as substitute by material scientists to overcome this issue. Low molecular weight can more easily self-assemble on substrate surfaces due to non-covalent interaction. Resulting polymers have the same chemical and physical property as the high molecular-weighted polymers. They would also print out a lot of vibrantly colored materials and take smaller volumes than high molecular-weighted polymers in the pigments and dyes would.



(Hart et al.)
(a) Divalent (1) and (b) trivalent (2) pyrenyl-terminated π-electron rich polymers (c) naphthalene–diimide residues (blue) (Hart et al.)

From bellow, the red molecules are pyrenyl polymer and the blue molecule is naphthalene polymer. They form π-π stacking shown on the left. Because non-covalent interactions between the constituent parts are reversible, supramolecular polymers can be temporarily disassembled "on command" by applying the right external stimulus thus this stacking can form healable supramolecular networks, and this approach can be used in the development of new materials for inkjet printing. (Montero et al.)



Reference
Hart, Lewis R., et al. "Supramolecular Approach to New Inkjet Printing Inks." Supramolecular Approach to New Inkjet Printing Inks, vol. 7, no. 16, Apr. 2015, pp. 8906–14, https://doi.org/10.1021/acsami.5b01569. Accessed 3 Aug. 2023.Montero, Lucas, et al. "Healable Supramolecular Polymer Solids." Healable Supramolecular Polymer Solids, vol. 49–50, Oct. 2015, pp. 60–78, https://doi.org/10.1016/j.progpolymsci.2015.04.003. Accessed 12 June 2023.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

90

Medicine

Fitzwilliam College Online Summer School Journal · 2023 Summer

Signs of a Head Injury and Their Relationship to Cranial Nerve Functions

Jasmeh Kaur Thethi

1. Introduction

There are many signs and symptoms a patient might develop after head injury, which relates to the functions of the cranial nerves. The cranial nerves are vulnerable following head trauma because many of them run across the surface of the skull and are only shielded by the muscles and tissues of the face. Nerves can also be damaged by broken facial and skull bones, and depending on the type of the injury, the effects of cranial nerve injury might be short-term or permanent (Hvingelby, 2022). These injuries may result in numerous symptoms, and understanding this association is important for giving patients the best treatment and recovery possible. To begin with, before discussing the relationship between head injuries and cranial nerves, it is necessary to understand their basic functions and injuries associated with it. A Traumatic Brain Injury (TBI) is a common head injury associated with disrupting the functions of the cranial nerves which results from a blow to the head. The cranial nerves are a group of twelve-paired nerves in the back of the head that transmit electrical signals between the brain, face, neck, and torso, and are frequently represented by roman numerals. Each nerve pair divides among the two sides of the brain and body, assisting with motor, sensory, and autonomic tasks (Cleveland Clinic, 2021).

- Olfactory (I): Sense of smell.
- Optic (II): Transmits visual data from the eye to the brain.
- Oculomotor (III): Controls the size of the pupils and its reaction to light.
- Trochlear (IV): Downward and inward eye movement.
- Trigeminal (V): Touch sensation to the face, controls chewing muscles.
- Abducens (VI): Eye's lateral mobility.
- Facial (VII): Moves the muscles that make facial emotions, gives the front two-thirds of the tongue a feeling of taste.
- Auditory-Vestibular (VIII): Balance and the sense of hearing.
- Glossopharyngeal (IX): Regulates the muscles of the neck, the salivary glands, and the back third of the tongue. It also perceives variations in blood pressure and transmits it to the brain.
- Vagus (X): Controls autonomic processes (heart rate, digestion, breathing).
- Spinal Accessory (XI): Neck and throat muscles and shoulder movement.
- Hypoglossal (XII): Enables speech through tongue movement (Hvingelby, 2022).

Symptoms of brain injury have a strong connection to the functions of the cranial nerves, which are sensitive during head trauma. Many of them run over the surface of the skull and are only protected by facial muscles and tissues (Hvingelby, 2022). Brain damage disrupts neural circuitry by killing neurons and disrupting their connections. This includes the cellular extensions (dendrites and axons) through which neurons receive and transmit messages via neurotransmitters (Pekna, 2012). Damage to a specific nerve can cause the following:

- Olfactory Nerve (I): Anosmia (loss of smell), hyposmia (decreased capacity to smell), and hypersomnia (increased sensitivity to the sense of smell) (Physiopedia, n.d.).
- Optic Nerve (II): Partial vision loss.

- Oculomotor (III), Trochlear (IV), and Abducens (VI) Nerves: Diplopia (double vision) (Kaufman, 2017).
- Trigeminal (V), Facial (VII), Glossopharyngeal (IX), and Vagus (X) Nerves: Facial numbness, issues chewing food, difficulty swallowing.
- Auditory-Vestibular Nerve (VIII): Problems maintaining balance.
- Spinal Accessory Nerve (XI): Weakness in shoulder and neck muscles (Hvingelby, 2022).

While there are direct ways a head injury can affect the cranial nerves, this can also happen indirectly through bleeds and raised intracranial pressure. Intracranial pressure, also known as intracranial hypertension, "is a spectrum of neurological disorders where cerebrospinal fluid (CSF) pressure within the skull is elevated," (Sharma, Hashmi, Kumar, 2023). The optic nerve may swell as a result of the elevated intracranial pressure, which could impair vision (Mayo Clinic, 2019).

There are numerous symptoms a patient might develop after a head injury, all depending on its severity. Although these symptoms can be categorized into different categories, the main symptoms are physical and can be seen in a conscious patient:

- Physical: difficulty maintaining balance, brief losses of consciousness, disorientation, memory loss, attention issues, blurred vision and sensitivity to light, facial weakness or paralysis (injuries to cranial nerve VII), difficulty swallowing/speaking, and hearing impairment.

Some of these symptoms could be challenging to distinguish from those of other medical conditions. Headaches that worsen, numbness/weakness, vomiting repeatedly, odd behavior, speaking in a different manner, and loss of consciousness are some of the warning symptoms to look out for (CDC, 2021).

However, there are signs a patient is experiencing cranial nerve damage while they are unconscious. If the patient's pupils are dilated and unresponsive, it is possible that the oculomotor nerve is being compressed. If the patient's neck is stiff, it may be because of raised intracranial pressure. Furthermore, if the patient is in a coma, it's best to check their motor responses, pupillary light reflexes, and corneal reflexes. An absence of a corneal reflex could indicate damage to the trigeminal nerve and facial nerve (NCBI, n.d.).

Head traumas can result in a variety of symptoms, some of which are directly related to cranial nerve functions. Understanding these connections is critical for healthcare workers to appropriately assess head trauma patients and provide suitable treatments. Thorough examination and personalized therapy can help enhance the recovery process and provide the best outcome possible for patients experiencing head trauma.

Additional Information – Acronym to remember the 12 Cranial Nerves: **On Old Olympus' Towering Top A Fully Armored Greek Ventured A Hop.**

2. Bibliography / Works Cited

1. Cranial Nerve Damage from Head Trauma. (2022). Verywell Health. https://www.verywellhealth.com/cranial-nerve-damage-from-head-trauma-1720018

2. Cleveland Clinic. (2021). Cranial Nerves: Function, Anatomy and Location. Cleveland Clinic. https://my.clevelandclinic.org/health/body/21998-cranial-nerves#:~:text=The%20cranial%20nerves%20are%20a

3. CDC. (2021, May 12). Symptoms of Mild TBI and Concussion | Concussion | Traumatic Brain Injury | CDC Injury Center. Www.cdc.gov. https://www.cdc.gov/traumaticbraininjury/concussion/symptoms.html

4. Pekna, M., & Pekny, M. (2012). The Neurobiology of Brain Injury. Cerebrum: The Dana Forum on Brain Science, 2012. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574788/#:~:text=Brain%20injury%20affects%20neuronal%20circuitry

5. Olfactory Nerve. (n.d.). Physiopedia. https://www.physio-pedia.com/Olfactory_Nerve#:~:text=Lesions%20to%20the%20Olfactory%20Nerve

6. Trochlear Nerve Injury - an overview | ScienceDirect Topics. (2017). Www.sciencedirect.com. https://www.sciencedirect.com/topics/medicine-and-dentistry/trochlear-nerve-injury#:~:text=Damage%20to%20any%20of%20these

7. Sharma, S., Hashmi, M. F., & Kumar, A. (2023). Intracranial Hypertension. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK507811/#:~:text=Introduction

8. Pseudotumor cerebri - Symptoms and causes. (2019). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/pseudotumor-cerebri/symptoms-causes/syc-20354031

9. Walker, M. C., & O' brien, M. D. (1999). Neurological Examination of the Unconscious Patient. Journal of the Royal Society of Medicine, 92(7), 353 – 355. https://doi.org/10.1177/014107689909200706

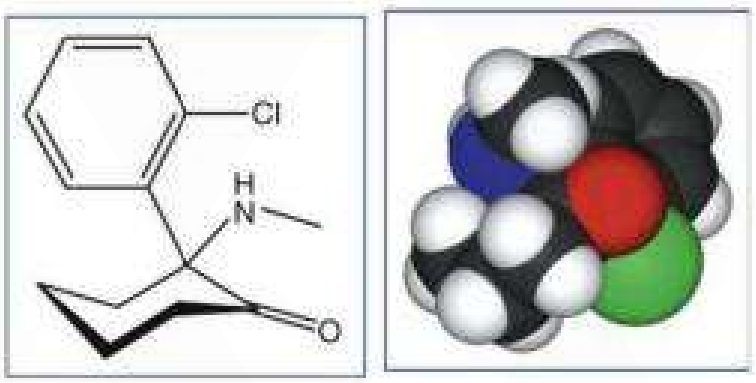
Two Drugs Used in Anesthesia

Naminrina

This essay explores two drugs used in anaesthesia, ketamine and nitrous oxide, focusing on how physical properties reflect upon anaesthetic usefulness.

1. General Properties of Ketamine

Ketamine is commonly seen as a white crystalline powder with a pKa of 7.5 (Orser, Pennefather & MacDonald, 1997) and melting points at approximately 260 degrees centigrade (Sinner & Graf, 2008). The molecule is hydrophilic and lipophilic at physiological pH (Kurdi, Theerth & Deva, 2014). The carbonyl group on the cyclohexanone ring forms hydrogen bonds with water, increasing the molecule's hydrophilicity. The benzene ring, cyclic carbon chain and methyl group add to lipophilicity.



Chemical structure of ketamine' (Sinner & Graf, 2008)

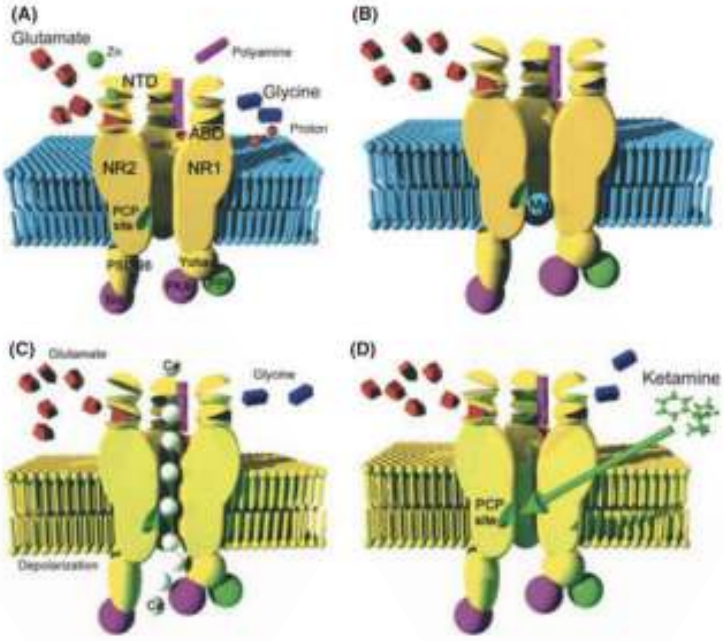
The drug is injected into the body as a solution and takes effect soon after entering circulation. Hydrophilicity allows ketamine to be absorbed thoroughly into the bloodstream, while lipophilicity renders quick crossing of lipid membranes. Metabolism of ketamine occurs mainly in the liver (Panos Zanos et al., 2018), meaning that the drugs retain potency upon reaching the brain.

2. Antagonistic Effects of Ketamine at NMDAR

Ketamine introduces a dissociative state, where the patients appear unconscious, amnesic and deeply analgesic. The effect is commonly attributed to the drug's interactions with NMDA receptors found on most cells in the central nervous system (Mion & Villevieille, 2013). The activation of NMDARs allows an influx of Calcium ions, depolarizing the neuron and initiating complex intracellular processes. Inhibition by ketamine depresses the activity of the cell and the central nervous system.

Agonists of the NMDA receptor open the ligand-gated ion channel. The presence of agonists, however, does not warrant depolarisation. At resting membrane potential, a Magnesium ion is locked in the channel pore by electrostatic force, blocking the passage of ions, and is only ejected when the cell is sufficiently depolarised. Only in the presence of agonists and the absence of the Magnesium blocker will a current be present.

Ketamine antagonises NMDAR by limiting the channel's opening time or opening frequency (Orser, Pennefather & MacDonald, 1997), reducing the extent of depolarization.



Glutamate-Dependent Mechanisms (Mion & Villevieille, 2013)

One antagonising mechanism is uncompetitive inhibition by binding to an ntrachannel site and preventing ion flux. The binding and dissociation of ketamine from the intrachannel site requires an open channel. After agonists bind to NMDAR and depolarisation occurs, the channel activates briefly and is then blocked by ketamine, reducing channel opening time.

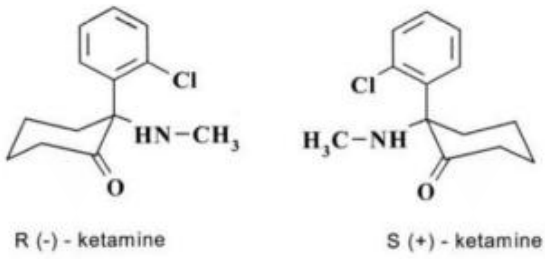
$$\begin{array}{ccc} & \beta[A] & \\ \text{Closed} & \rightleftharpoons & \text{Open} \\ & \alpha & \end{array}$$

$$k_{-B} \uparrow \downarrow k_{+B}[C]$$

$$\begin{array}{ccc} & \beta[A] & \\ \text{Closed} & & \text{Open} \\ \text{Blocked} & \rightleftharpoons & \text{Blocked} \\ & \alpha & \end{array}$$

'A Model of "Drug Trapping"' (Orser, Pennefather & MacDonald, 1997)
'The rates of channel opening and closing are α and β, the rates of association and dissociation of the blocker are k+B and k-B, and [A] and [C] are the concentrations of agonist and blocker, respectively.'

Additionally, ketamine reduces the number of activatable channels and limits opening frequency. On channel opening, the ketamine blocker binds to the intrachannel site. If the agonists disassociate from the receptors while the blocker is still in place, ketamine cannot dissociate and becomes trapped in the channel. The channel is now in a closed and blocked state. Only upon rebinding of agonists and further dissociation of ketamine will an ion channel be reactivated (Orser, Pennefather & MacDonald, 1997). The increase in affinity of the drug to the target protein when the protein is in an active state is known as a use-dependent block, commonly seen in local anaesthetics that block sodium channels. This mechanism potentially maximises drug efficacy by inhibiting signals that significantly contribute to sensory and motor functions, while leaving out the rest.



'Structural Formulas of Two Enantiomers of Ketamine' (Kohrs & Durieux, 1998)

Ketamine solution contains equal amounts of two enantiomers known as S-ketamine and R-ketamine. Studies have shown that S-ketamine is more potent than the latter, possibly due to its greater affinity to the intrachannel binding site (Kohrs & Durieux, 1998).

Another mechanism for NMDAR inhibition is non-competitive inhibition. The lipophilicity of ketamine allows the drug to bind to a hydrophobic site on the channel, reducing the number of excitable channels.

3. General Properties of Nitrous Oxide (N_2O)

Nitrous Oxide is a polar molecule existing as a colourless, sweet-smelling gas used in anaesthesia. It is known for its analgesia and anxiolysis effects (Emmanouil & Quock, 2007), along with a history of safety of use (Becker & Rosenberg, 2008), demonstrated by its fast onset of action, rapid termination of effect and low potency.

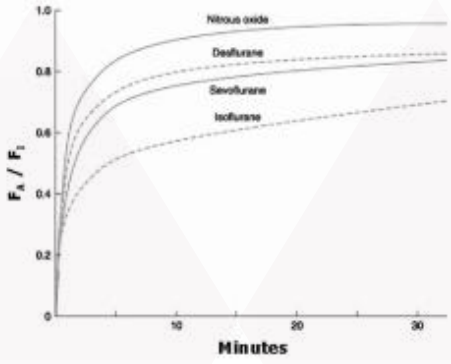
4. Analgesic and Anxiolytic effects of Nitrous Oxide

Nitrous Oxide produces analgesic and anxiolytic effects similar to morphine and benzodiazepines.

N2O relieves pain by stimulating the release of endogenous opioid peptides (Emmanouil & Quock, 2007) instead of directly acting on opioid receptors, supported by the observation that morphine-tolerant animals show cross-tolerance to N2O unilaterally. Opioid peptides no longer produce the same level of effect when binding to desensitised receptors in animals with morphine tolerance (Goodsell, 2005). N2O tolerance, on the other hand, is attributed to the depletion of opioid peptides rather than changes induced on the receptor level.

5. Solubility of Nitrous Oxide, Onset of Action and Reversibility

The onset of action or effect is the time it takes for a drug to produce a noticeable effect. In the case of N2O, it is the time it takes for sufficient amounts of the gas to act upon opioid receptors of the nervous system. This is measured to be within 10 minutes, making nitrous oxide one of the fastest-acting inhalation anaesthetics (Becker & Rosenberg, 2008).



'Relative Onset of Effect' (Becker & Rosenberg, 2008)
FI : inspired gas tension, the partial pressure of the gas throughout body tissues; FA : alveoli gas tension, the partial pressure of the gas in the alveoli. 'Nitrous oxide achieves approximately 90% equilibrium within 10 minutes'

This fast onset of action can be explained by its low blood : gas and fat : blood partition coefficients of 0.47 and 2.3, respectively. Partition coefficients measure the ratio between a substance's concentration in two contacting mediums at equilibrium (Pawson & Forsyth, 2008). The blood : gas and fat : blood partition coefficients show the relative solubility of N2O in respecting mediums and tissues, that being inhaled gases, the blood and neural tissue.

Under a unidirectional partial pressure gradient, the lower the solubility of a gas in the medium of lower partial pressure, the faster the equilibrium is reached and the medium saturated. Upon administration into the alveoli, the partial pressure gradient pushes N2O from the lungs into the bloodstream. N2O concentration in the proxy bloodstream increases until partial pressures equalise. Blood saturated with nitrous oxide then pushes into circulation until it reaches neural tissue, where the process repeats. Upon reaching nerve tissues concentrated with lipids, N2O acts on receptors and the anaesthetic effect takes place.

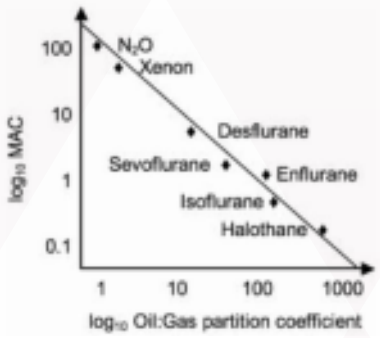
Elimination of the gas occurs in reverse. When administration halts, the direction of the partial pressure gradient reverses and the drug is rapidly removed from the system. This makes nitrous oxide suitable for operations that require intermittent alterations (Becker & Rosenberg, 2008) with its fast onset and termination of action.

6. Potency of Nitrous Oxide

N2O is considered the least potent inhalation anaesthetic as it has little influence on respiratory and cardiovascular functions compared with more potent anaesthetics (Becker & Rosenberg, 2008). This is reflected by the gas's minimum alveolar concentration, an indicator of relative potencies of anaesthetic gases, and low lipid solubility (Becker & Rosenberg, 2008).

The minimal alveolar concentration, or MAC, measures the concentration of the drug required to suppress motor response to a noxious stimulus. With a MAC of 104, nitrous oxide is often administered with other central nervous system depressants to produce reliable anaesthetic effects (Becker & Rosenberg, 2008).

The linear structure of nitrous oxide produces momentary dipoles, making the molecule sparingly soluble in lipids. Since the mechanism of action of nitrous oxide involves passing through lipid membranes to stimulate the release of opioid peptides, this low lipophilicity hinders diffusion across the membrane, limiting the efficacy of nitrous oxide.



'Meyer-Overton graph of potency versus lipid solubility' (Cross & Plunkett, 2014)
log10MAC is inversely related to drug potency;
log10 Oil:Gas partition coefficient measures lipid solubility.

The unreachable MAC and low lipophilicity are in line with observations that gave rise to the Meyer-Overton hypothesis: For most molecules that diffuse through membranes, lipid solubility relates to membrane permeability, predicting the potency of the anaesthetic. The low potency and reversible nature make nitrous oxide one of the safest inhalant anaesthetics used in medical settings.

7. References

1. Becker DE, Rosenberg M. Nitrous oxide and the inhalation anesthetics. Anesth Prog. 2008 Winter;55(4):124-30; quiz 131-2. doi: 10.2344/0003-3006-55.4. 124. PMID: 19108597; PMCID: PMC2614651.

2. Cross, M., & Plunkett, E. (2014). The Meyer – Overton hypothesis. In Physics,Pharmacology and Physiology for Anaesthetists: Key Concepts for the FRCA (pp. 146- 147). Cambridge: Cambridge University Press. doi:10. 1017/CBO9781107326200.059

3. Emmanouil DE, Quock RM. Advances in understanding the actions of nitrous oxide. Anesth Prog. 2007 Spring;54(1):9- 18. doi: 10.2344/0003-3006(2007)54[9:AIUTAO]2.0.CO;2. PMID: 17352529; PMCID: PMC1821130.

4. Goodsell, D.S. (2005), The Molecular Perspective: Morphine. STEM CELLS, 23: 144- 145. https://doi.org/ 10. 1634/ stemcells.FCM1

5. Kohrs R, Durieux ME. Ketamine: teaching an old drug new tricks. Anesth Analg.

6. 1998 Nov;87(5):1186-93. doi: 10. 1097/00000539- 199811000-00039. PMID: 9806706.

7. Kurdi MS, Theerth KA, Deva RS. Ketamine: Current applications in anesthesia, pain, and critical care. Anesth Essays Res. 2014 Sep-Dec;8(3):283-90. doi: 10.4103/0259- 1162. 143110. PMID: 25886322; PMCID: PMC4258981.

8. Sinner B, Graf BM. Ketamine. Handb Exp Pharmacol. 2008;(182):313-33. doi:

9. 10. 1007/978-3-540-74806-9 15. PMID: 18175098.

10. Mion G, Villevieille T. Ketamine pharmacology: an update (pharmacodynamics and molecular aspects, recent findings). CNS Neurosci Ther. 2013 Jun;19(6):370-80. doi: 10. 1111/cns. 12099. Epub 2013 Apr 10. PMID: 23575437; PMCID: PMC6493357.

11. Orser BA, Pennefather PS, MacDonald JF. Multiple mechanisms of ketamine blockade of N-methyl-D-aspartate receptors. Anesthesiology. 1997 Apr;86(4):903- 17. doi: 10. 1097/00000542- 199704000-00021. PMID: 9105235.

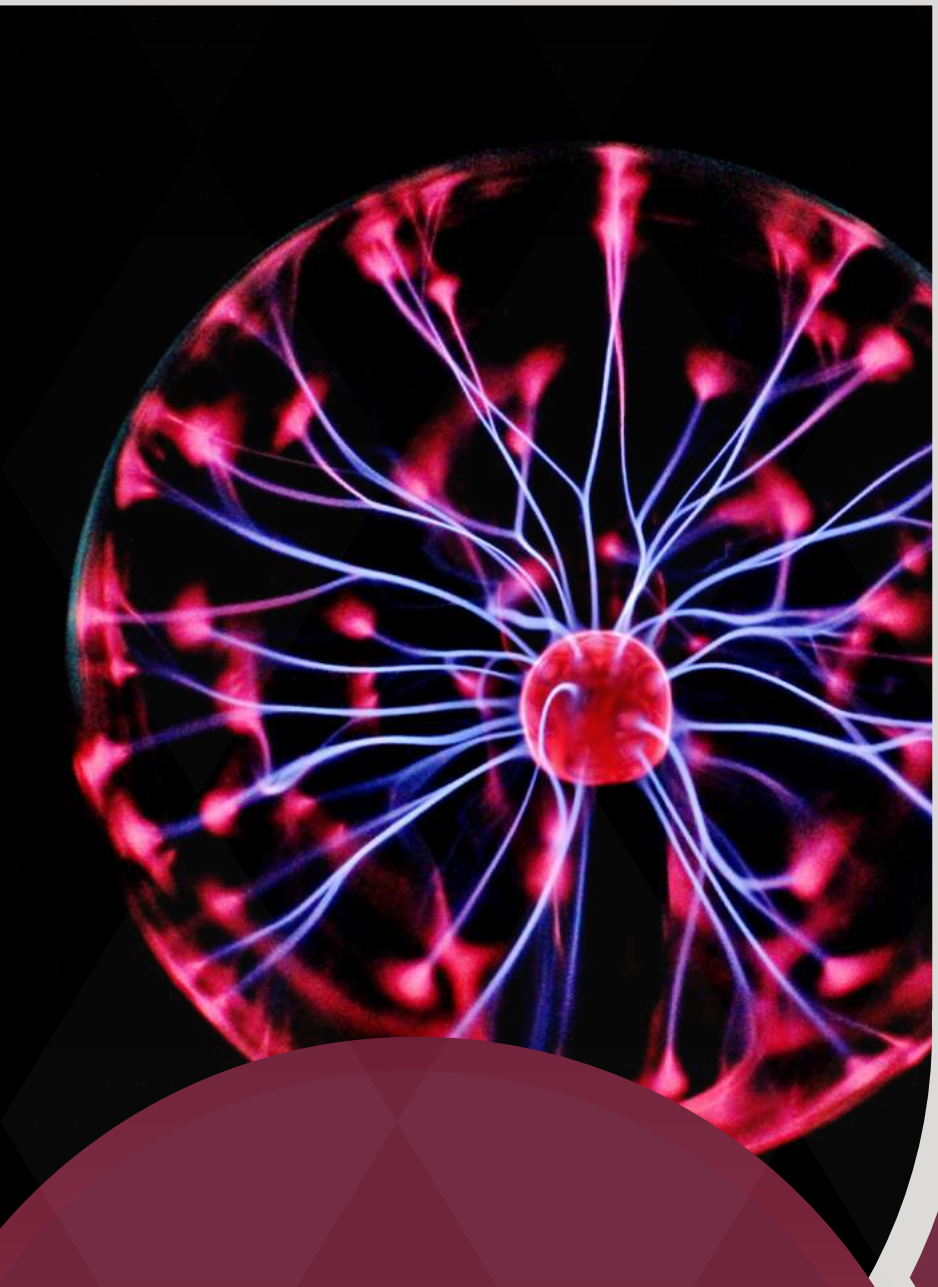
12. Patricia Pawson, Sandra Forsyth, Chapter 5 - Anesthetic agents, Editor(s): JILL E MADDISON, STEPHEN W PAGE, DAVID B CHURCH, Small Animal Clinical Pharmacology (Second Edition), W.B. Saunders, 2008, Pages 83- 112, ISBN 9780702028588, https://doi.org/ 10. 1016/B978-070202858-8.50007-5. (https://www.sciencedirect.com/science/article/pii/B9780702028588500075)

13. Zanos P, Moaddel R, Morris PJ, Riggs LM, Highland JN, Georgiou P, Pereira EFR, Albuquerque EX, Thomas CJ, Zarate CA Jr, Gould TD. Ketamine and Ketamine Metabolite Pharmacology: Insights into Therapeutic Mechanisms. Pharmacol Rev. 2018 Jul;70(3):621-660. doi: 10. 1124/pr. 117.015198. Erratum in: Pharmacol Rev. 2018 Oct;70(4):879. PMID: 29945898; PMCID: PMC6020109.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

07



Psychology and Neuroscience

Fitzwilliam College Online Summer School Journal · 2023 Summer

Delving Into Representativeness Heuristics: Influences on Decision-Making

Sicun Lin (Amy)

1. Introduction

Suppose you are confronted with a decision that needs to be made in seconds, it would be impossible to pore over each piece of data and forecast potential outcomes. This is precisely when heuristics - mental shortcuts or "rules of thumb" - become invaluable for processing information and guiding our decision-making process. In the context of cognitive psychology, three primary types of heuristics significantly influence the way decisions are made. These include representativeness heuristics, availability heuristics, and anchoring bias (Tversky & Kahneman, 1974). Although they all function as cognitive shortcuts, the underpinning mechanisms of each are diverse and distinct from one another. To delve deeper and have a better understanding, this essay will specifically illuminate how representativeness heuristics operate and how they manipulate the mind of decision-making.

2. Ways representativeness heuristics influence decision-making

The first way that representativeness heuristics impact how decisions are made is by predisposing people to favor elaborate, scenario-specific descriptions over more general ones, despite the fundamental probability indicating that the latter is more likely. A clear example of this bias can be found in the Reader Survey conducted by Tversky and Kahneman (1982). In this survey, a description of a woman named Linda is provided:

> "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations."

Following this, survey participants were asked to select the most plausible option from two possibilities:

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

In defiance of basic probability principles, a significant proportion of people felt that Linda was more likely to be a feminist bank teller rather than just a bank teller. This intuition starkly contradicts the conjunction principle in Statistics, which states that the probability of two events occurring simultaneously (in this case, Linda is a bank teller and a feminist) is always lesser than the probability of either event happening independently, which is also illustrated in the Venn Diagram below. Despite this, our representativeness heuristic leads us to believe that Linda is a feminist bank teller because the added detail makes this description more representative of the image we've formed of her. In essence, people tend to opt for the more detailed option as it aligns more closely with their mental representation of the situation, overlooking the greater likelihood of a more generic scenario.

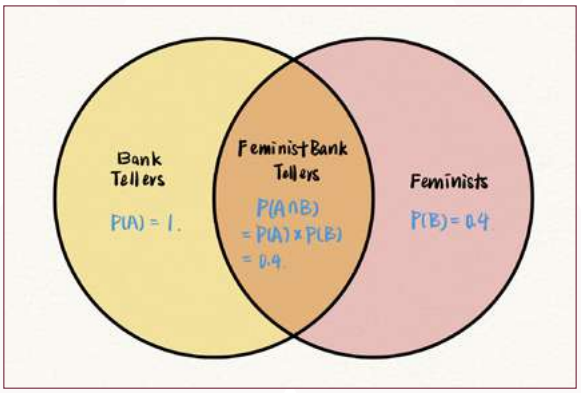


Figure 1. Venn Diagram showing the possibility of each scenario with P(B) randomly assigned to a value between 0 and 1

Another notable way that the representativeness heuristic influences decision-making is by leading us to heavily rely on how well something conforms to our existing prototypes while overlooking base rate information. For instance, if we encounter a slim, short man who wears glasses and enjoys poetry, we might immediately presume that he's likely a professor due to the stereotype that these characteristics fit. However, statistically speaking, he's more likely to be a truck driver, simply because there are more truck drivers than professors in this world. This example underlines how we tend to omit considering base rate information when making decisions, often resulting in inaccurate judgments. Such tendencies are not exclusive to laypeople but can also affect professionals, including doctors. For example, as demonstrated by Brannon and Carson (2003), physicians are not immune to the pitfalls of representativeness heuristics. When tasked with diagnosing patients who present symptoms indicative of a heart attack but had recently experienced stress, they were more likely to provide less severe diagnoses, attributing these symptoms to the recent stressful event rather than a potential heart condition. They neglected the high probability of cardiovascular damage, illustrating how easily representativeness heuristics can lure even trained professionals into diagnostic errors.

While representativeness heuristics can lead to biases and inaccuracies, they also play an important role in decision-making, particularly when time and cognitive resources are limited. Humans need to make innumerable decisions every day, and not all decisions require or allow for a detailed analysis of every single option. In these circumstances, heuristics act as mental shortcuts that allow us to make reasonably accurate decisions in a limited time. For instance, if we have to decide whether or not to cross a busy street, it would be impractical and even dangerous to calculate the exact speeds and trajectories of all the vehicles. Instead, we quickly categorize the situation based on past experiences and prototypes (e.g., "cars are moving too fast, it's not safe to cross") and make a decision accordingly. At the same time, education and awareness are key in mitigating the biases caused by representativeness heuristics. Understanding the role that stereotypes and base rates play in our perceptions, can help us make more informed judgments. It's not about eliminating the use of heuristics, but rather, being mindful of their limitations and improving our decision-making correspondingly.

3. Conclusion

Therefore, representativeness heuristics, like any other cognitive tool, is a double-edged sword: it can lead to biases and inaccuracies but also enables an efficient decision-making process; specifically, we know that biases are highly based on our tendencies to ignore fundamental probability laws. However, the goal should always be to strike a balance, by leveraging its benefits while minimizing its downsides. Eventually, representativeness heuristics and all other kinds of cognitive traps will no longer be a trap, instead, they will become our best toolkit to be utilized in our everyday lives.

4. References

1. Brannon, L. A., & Carson, K. L. (2003). The representativeness heuristic: influence on nurses' decision making. Applied Nursing Research, 16(3), 201 – 204. https://doi.org/10.1016/s0897-1897(03)00043-0

2. Cornell (PhD), D., & Drew (PhD), C. (2023, January 18). 11 Representativeness Heuristic Examples (2023). Helpfulprofessor.com. https://helpfulprofessor.com/representativeness-heuristic-examples/#:~:text=We%20have%20seen%20that%20the

3. Gerrig, R. J. (2014). Psychology and life. Pearson.

4. Nikolopoulou, K. (2022, December 28). Representativeness Heuristic | Example & Definition. Scribbr. https://www.scribbr.com/research-bias/representativeness-heuristic/

5. Plous, S. (1993). The psychology of judgment and decision making (p.233). New York McGraw-Hill Higher Education.

6. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. Science, 185(4157), 1124 – 1131. https://doi.org/10.1126/science.185.4157.1124

Can We Create Humanoid Robots With Human Emotions?

Menghan Ren (Vicky)

1. Introduction

The emergence and rapid development of artificial intelligence (AI) have sparked intriguing debates about the possibility of creating humanoid robots that can possess emotions similar to humans. This prompts two fundamental questions: Can we truly create machines capable of experiencing emotions akin to humans? And, are human emotions truly equivalent to machine emotions? Addressing these questions requires a deep exploration of the complexities of human emotions, the advancements in AI technology, and an understanding of the intersection between human and machine cognition. This paper aims to delve into these issues, analyzing the potential of creating human-like emotionally intelligent robots while examining the similarities and disparities between human and machine emotions. By shedding light on the current understanding of emotions in both humans and machines, this study aims to contribute to the ongoing discourse on the boundaries and possibilities of emotional intelligence in artificial beings.

2. Can we create humanoid robots with human emotions?

This pursuit involves integrating various disciplines, such as robotics, neuroscience, psychology, and computer science. Researchers have made significant progress in developing emotionally intelligent machines that can recognize and mimic facial expressions, make empathic responses, and even exhibit simulated emotional states. Emotional computing, machine learning, and cognitive architecture have paved the way for advances in emotion recognition and synthesis. However, despite these achievements, the question remains: Can machines really experience emotions in the same way that humans do, or are they simply programmed algorithms and reactions to simulate emotions? To answer this question, it is important to understand the nature of human emotions and the underlying mechanisms that produce them. Human emotion encompasses a complex set of cognitive and physiological processes that are influenced by cultural, social and personal factors. Emotions intertwine with subjective and introspective experiences to form an intricate tapestry of human consciousness. These experiences produce not only emotional states such as joy, sadness, anger, fear, love, and surprise, but also situation-dependent emotional responses. In contrast, the emotions expressed by machines are based on algorithms, patterns, and programmed responses derived from the analysis of vast amounts of data. While machines can recognize and mimic human emotions to a certain extent, they lack the subjective richness and deeply personal aspects that accompany human emotional experiences. This discrepancy raises questions about the authenticity and depth of the emotions displayed by the machines. As we delve further into the quest to create emotionally intelligent humanoid robots, it becomes critical to consider the limitations and challenges inherent in this pursuit. These challenges include the complexity of replicating conscious subjective experience in machines, the nuances of non-verbal communication behind human emotional expression, and the ethical implications of creating machines that can manipulate human emotions. By addressing these questions, we can gain a deeper understanding of the boundaries and possibilities of creating human-like emotionally intelligent machines.

3. Are Human emotions equivalent to Machine emotions?

In order to assess the equivalence between human emotion and machine emotion, it is important to examine the similarities and differences between the two. Human emotion is a complex, multifaceted phenomenon that is influenced by a variety of cognitive, physiological and sociocultural factors. Human emotions have important adaptive functions, facilitating social interaction, decision-making processes, and subjective experience. Machine emotion, on the other hand, is the product of computational processes and algorithms that mimic human emotional responses.

While machines can recognize and respond to specific emotional cues, they lack the subjective experience and awareness associated with human emotions. Machine emotion is based on objective analysis and pattern recognition, driven by programmed algorithms rather than subjectively perceived internal states. One similarity between human and machine emotions is the ability to recognize and classify emotional states. Advances in techniques such as emotional computing and natural language processing have enabled machines to detect and interpret facial expressions, tone of voice, and verbal content associated with different emotions. Machines can learn to categorize emotions using machine learning techniques and translate those classifications into appropriate responses. A fundamental difference, however, lies in the subjective experience attached to human emotions. Human emotions are intertwined with personal beliefs, values, memories, and background experiences, resulting in subjective and introspective elements that machines cannot replicate. While a machine can simulate the outward appearance of emotions, it lacks the inner subjective experience that gives depth and meaning to human emotional states. Another difference comes from the social and cultural aspects of human emotion. Human emotions are influenced by socio-cultural norms, personal experience and social context. Cultural differences in the expression and definition of emotions highlight the connection between emotions and the environment experienced. In contrast, machines lack the ability to embed emotions in cultural and social contexts, limiting their understanding and expression of emotions. While machines have advanced in their ability to recognize and respond to emotions, there are fundamental differences between human and machine emotions. Human emotions include subjective experiences, cultural influences, and social contexts that do not exist in machine emotions. These differences raise questions about the true equivalence of the two, and ethical considerations about the use of emotionally intelligent machines and the potential impact they could have on human emotional experiences.

In conclusion, while we may not be able to fully replicate human emotions in machines, ongoing research and advancements in artificial intelligence and robotics hold the potential to create machines that exhibit programmed emotional responses. However, it is important to acknowledge and respect the uniqueness of human emotions and avoid conflating machine responses with genuine human feelings. As the field progresses, ethical considerations and responsible development should be at the forefront to ensure the integration of emotionally advanced robots aligns with societal needs and values.

4. References

1. Hawkins, J. (2004). On Intimacy, Authenticity, and Autoplastic Adaptation: Emotional Robots as Therapeutic Agents. CyberPsychology & Behavior, 7(2), 135-143. - The study discusses the use of emotional robots in therapy. And their potential for fulfilling an individual's emotional needs.

2. Picard, R. W. (2000). Affective Computing: Challenges. International Journal of Human-Computer Studies, 59(1-2), 55-64. - Challenges. International Journal of Human-computer Studies, 59(1-2), 55-64. - Challenges. International Journal of Human-computer Studies, 59(1-2), 55-64. And how to give machines emotional intelligence.

3. Breazeal, C. (2002). Designing Sociable Robots. MIT Press. - This book explores how to design robots that can establish an emotional connection with humans.

4. Bartneck, C., & Forlizzi, J. (2004). Affective Robots Need Genuine Emotions. Proc. of CHI 2004, 669-670. - This paper discusses that the design of emotional robots must have true emotions and corresponding behaviors.

5. Parrenas, R. S. (2018). Emotional Cyborgs: Designing Robots for Emotional Labor. Social Robots from a Human Perspective, 143-156. - Designing robots for emotional labor. Social robots from a human perspective, 143-156.

Also Google Wikipedia and Baidu search.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

80

Mathematics for the Natural Sciences

Fitzwilliam College Online Summer School Journal · 2023 Summer

Fourier Transformation and its application

KINTAT NG (Arvin)

1. Introduction

FT is an extension of Fourier series that transfer from time domain to frequency domain, which is useful for example when we want to filter waves. It is difficult to see what the signal contains in time domain but is direct in frequency domain.

So it is useful to first look at what Fourier series is.

2. Fourier Series

We've known a way of using simple functions to mimic other functions which is Taylor series

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(a)\frac{(x-a)^k}{k!}$$

But it requires f (x) has nth derivative at least and this only describe the neighbourhood around x = a. Is there any other expansion?

In the set of [1,sin(nx),cos(nx)], the integral of product of two different element is 0 in one period where as the product of two same element will be π, (or 2π if that is 1)

For example,

$$\int_{-\pi}^{\pi} \cos(3x)\cos(2x) \quad dx = 0$$

$$\int_{-\pi}^{\pi} \cos(3x)\cos(3x) \quad dx = \pi$$

In 1822, Fourier claimed that any function can be expanded into a series of cosines and sines.

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx))$$

where f(x) is of period 2π

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx) \quad dx$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx) \quad dx$$

by the Orthogonality of the [1,sin(nx),cos(nx)]

Then we change the variable $x = \frac{T}{2\pi}t, t = \frac{2\pi}{T}x$, and now $f(\frac{\pi}{2T}t) = g(t)$ is of period 2π

Expand the g(t) and change to f (x),(skip derivation) we get:

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n\cos(\frac{2n\pi}{T}x) + b_n\sin(\frac{2n\pi}{T}x))$$

$$a_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)\cos(\frac{2n\pi}{T}x) \quad dx$$

$$b_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)\sin(\frac{2n\pi}{T}x) \quad dx$$

Recall Euler's Formulae $e^{i\theta} = cos(\theta) + isin(\theta)$ and change all the cos and sin into exponential (skip derivation) , and we have :

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad , \quad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

$$f(x) \sim \sum_{n=-\infty}^{+\infty} \frac{1}{T} C_n e^{in\omega_0 x}$$

,where $\omega_0 = \frac{2\pi}{T}$, $C_n = \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-in\omega_0 x} \quad dx$ (Pretty!)

3. Fourier Transformation

In other words the T will tend to infinity. $T \to \infty$, in order to find f (t) we can do the following steps:

1. Take the $\left[-\frac{T}{2}, \frac{T}{2}\right]$ of f(t) and expand it to periodic function $f_T(t)$ on [$-\infty$, $+\infty$] (such that it is of period T)
2. expand $f_T(t)$ by Fourier series
3. let $T \to \infty$

consider $\Delta\omega = (n+1)\omega_0 - n(\omega_0) = \omega_0 = \frac{2\pi}{T}$, then $\frac{1}{T} = \frac{\Delta\omega}{2\pi}$

$$f(t) = \lim_{T\to+\infty} f_T(t) \sim \lim_{\Delta\omega\to 0} \sum_{n=-\infty}^{+\infty} \frac{1}{T} C_n e^{in\omega_0 t}$$

$$= \lim_{\Delta\omega\to 0} \sum_{n=-\infty}^{+\infty} \frac{\Delta\omega}{2\pi} C_n e^{in\omega_0 t} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} C_n e^{i\omega t} d\omega$$

where C_n will become

$$\int_{-\infty}^{+\infty} f(t)e^{-i\omega x} dt$$

Bring them together we have :

$$f(t) \sim \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(t)e^{-i\omega t} dt \quad e^{i\omega t} d\omega$$

We call

$$\int_{-\infty}^{+\infty} f(t)e^{-i\omega t} dt$$

Fourier Transformation F(t)

and the outer part inverse Fourier Transformation

4. Application

There are some properties that are useful and here I'll list one of them to help solve ODEs and PDEs

F(f')=iωF(f)

SHM: $\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x(t) = \frac{f(t)}{m}.$

To solve for x(t) , we first take the Fourier transform of both sides

$$-\omega^2 X(\omega) - 2i\gamma\omega X(\omega) + \omega_0^2 X(\omega) = \frac{F(\omega)}{m},$$

The Fourier transform has turned our ordinary differential equation into an algebraic equation which can be easily solved

$$X(\omega) = \frac{F(\omega)/m}{-\omega^2 - 2i\gamma\omega + \omega_0^2}$$

then use inverse Fourier Transformation

$$x(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega t} F(\omega)/m}{-\omega^2 - 2i\gamma\omega + \omega_0^2},$$

where $F(\omega) = \int_{-\infty}^{\infty} dt\ e^{i\omega t} f(t).$

Laplace Transform

SIRUI CHEN (Colin)

1. What is Laplace transform?

The Laplace transform is a mathematical tool used to transform a function of real variable (usually time) into a function of a complex variable known as the Laplace domain.

It is widely used in engineering, physics, and applied mathematics to solve differential equations, analyze linear systems, and study the behavior of signals and systems.

In particular, it effectively converts an ordinary differential equation into an algebraic equation which can be manipulated easily.

2. How laplace Transform works

Laplace transform is the integral transform of the given derivative function with real variable t to convert into a complex function with complex variable s.

For t ≥ 0, let f(t) be given and assume the function satisfies certain conditions to be stated later on. The Laplace transform of f(t), that is denoted by L{f(t)} or F(s) is defined by the Laplace transform formula:

$$\mathcal{L}\{f\}(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

3. Laplace transform table

Table of Laplace Transforms

	$f(t) = \mathcal{L}^{-1}\{F(s)\}$	$F(s) = \mathcal{L}\{f(t)\}$		$f(t) = \mathcal{L}^{-1}\{F(s)\}$	$F(s) = \mathcal{L}\{f(t)\}$
1.	1	$\frac{1}{s}$	2.	e^{at}	$\frac{1}{s-a}$
3.	$t^n,\ n = 1, 2, 3, \ldots$	$\frac{n!}{s^{n+1}}$	4.	$t^p,\ p > -1$	$\frac{\Gamma(p+1)}{s^{p+1}}$
5.	$\sqrt{t}$	$\frac{\sqrt{\pi}}{2s^{\frac{3}{2}}}$	6.	$t^{n-\frac{1}{2}},\ n = 1, 2, 3, \ldots$	$\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)\sqrt{\pi}}{2^n s^{n+\frac{1}{2}}}$
7.	$\sin(at)$	$\frac{a}{s^2+a^2}$	8.	$\cos(at)$	$\frac{s}{s^2+a^2}$
9.	$t\sin(at)$	$\frac{2as}{(s^2+a^2)^2}$	10.	$t\cos(at)$	$\frac{s^2-a^2}{(s^2+a^2)^2}$
11.	$\sin(at) - at\cos(at)$	$\frac{2a^3}{(s^2+a^2)^2}$	12.	$\sin(at) + at\cos(at)$	$\frac{2as^2}{(s^2+a^2)^2}$
13.	$\cos(at) - at\sin(at)$	$\frac{s(s^2-a^2)}{(s^2+a^2)^2}$	14.	$\cos(at) + at\sin(at)$	$\frac{s(s^2+3a^2)}{(s^2+a^2)^2}$
15.	$\sin(at+b)$	$\frac{s\sin(b) + a\cos(b)}{s^2+a^2}$	16.	$\cos(at+b)$	$\frac{s\cos(b) - a\sin(b)}{s^2+a^2}$
17.	$\sinh(at)$	$\frac{a}{s^2-a^2}$	18.	$\cosh(at)$	$\frac{s}{s^2-a^2}$

(Boyd, n.d., p. 1)

4. Proof of some useful formulas

For f(t) = 1, F(s) = 1/s

$$f(t) = 1$$
$$F(s) = \int_0^\infty 1 \cdot e^{-st}\,dt$$
$$= \left[-\frac{1}{s}e^{-st}\right]_0^\infty$$
$$= 0 - \left(-\frac{1}{s}\right)$$
$$= \frac{1}{s}$$

For f(t) = e^(at) , F(s) = 1/(s-a)

$$f(t) = e^{at}$$
$$F(s) = \int_0^\infty e^{at} \cdot e^{-st}\,dt$$
$$= \int_0^\infty e^{-(s-a)t}\,dt$$
Let $s-a = S$
$$F(s) = \int_0^\infty e^{-St}\,dt$$
$$= \frac{1}{S}$$
$$= \frac{1}{s-a}$$

For f(t) = sin(bt), F(s) = b/(s^2+b^2)

For f(t)= cos(bt), F(s) = s/(s^2+b^2)

$f(t) = \sin(bt)$, $f(t) = \cos(bt)$

By Euler's formula, $e^{i\theta} = \cos\theta + i\sin\theta$

$$e^{ibt}e^{-st} = \cos(bt)e^{-st} + i\sin(bt)e^{-st}$$
$$\int_0^\infty e^{ibt}e^{-st}\,dt = \int_0^\infty \cos(bt)e^{-st}\,dt + i\int_0^\infty \sin(bt)e^{-st}\,dt$$
$$\int_0^\infty e^{-(s-ib)t} = \mathcal{L}(\cos(bt)) + i\mathcal{L}(\sin(bt))$$
$$\frac{1}{s-ib} = \mathcal{L}(\cos(bt)) + i\mathcal{L}(\sin(bt))$$
$$\frac{s+ib}{(s-ib)(s+ib)} = \frac{s+ib}{s^2+b^2}$$
$$\therefore \mathcal{L}(\cos(bt)) = \frac{s}{s^2+b^2}, \quad \mathcal{L}(\sin(bt)) = \frac{b}{s^2+b^2}$$

5. Region of convergence

Since the implementation of Laplace transform requires the integration over an infinite domain, it is necessary to confirm whether the function converges or not.

This means that the following limits must exist when applying Laplace transform.

$$\lim_{R\to\infty} \int_0^R f(t)e^{-st}\,dt$$

6. Region of convergence example

The region of convergence(ROC) for f(t)=e^(at) to apply laplace transform can be interpreted as shown beside.

$$f(t) = e^{at}$$
$$F(s) = \int_0^\infty e^{at} \cdot e^{-st}\,dt \quad \text{where } s = \phi + i\omega$$
$$= \left[\frac{1}{a-s}e^{(a-s)t}\right]_0^\infty$$
$$= \frac{1}{a-s}\left[e^{(a-\phi)t}e^{-i\omega t}\right]_0^\infty$$

1° $a - \phi \geq 0$
$\lim_{t\to\infty} e^{(a-\phi)t}e^{-i\omega t}$ does not exist,
and the integral cannot be evaluated

2° $a - \phi < 0$
$\lim_{t\to\infty} e^{(a-\phi)t}e^{-i\omega t} = 0$
$$\therefore F(s) = 0 - \frac{1}{a-s}\cdot 1 = \frac{1}{s-a}$$
Hence, $Re[s] > a$ is the region of convergence for e^{at} in Laplace transform

7. Basic properties of Laplace transform

	t-domain function	s-domain function
1. Linear	$af_1(t) + bf_2(r)$	$aF_1(s) + bF_1(s)$
2. Scaling	$f(at)$	$\frac{1}{a}F\left(\frac{s}{a}\right)$
3. Time Shift	$f(t-t_0)\,u(t-t_0)$	$e^{-st_0}F(s)$
4. Frequency Shift	$e^{at}f(t)$	$F(s-a)$
5. Reverse Time	$f(-t)$	$F(-s)$
6. Time Differentiation	$\frac{df(t)}{dt}$ $\frac{d^n f(t)}{dt^n}$	$sF(s) - f(0)$ $s^nF(s) - s^{n-1}f(0) - \ldots - f^{(n-1)}(0)$
7. Time Integral	$\int_0^t f(\tau)d\tau$	$\frac{F(s)}{s}$
8. Frequency Differentiation t-multiplication	$tf(t)$ $t^n f(t)$	$-\frac{dF(s)}{ds}$ $(-1)^n\frac{d^nF(s)}{ds^n}$
9. t-division Frequency Integration	$\frac{f(t)}{t}$	$\int_s^\infty F(u)du$
10. Periodic	$f(t)$	$\frac{F_0(s)}{1-e^{-Ts}}$
11. Initial Value	$f(0)$	$\lim_{s\to\infty} sF(s)$
12. Final Value	$f(\infty)$	$\lim_{s\to 0} sF(s)$

(M.S, 2012, p. 1)

8. Worked example of solving second order ODE

Example:

$\mathcal{L}\{f'(t)\} = sY(s) - y(0)$

$\mathcal{L}\{f''(t)\} = s^2Y(s) - sy'(0) - y(0)$

$y'' - 10y' + 9y = 5t$, $y(0) = -1$, $y'(0) = 2$

Apply Laplace transform to both sides,

$s^2Y(s) - sy'(0) - y(0) - 10sY(s) + 10y(0) + 9Y(s) = \frac{5}{s^2}$

$Y(s)(s^2 - 10s + 9) - 2s + 2 - 10 = \frac{5}{s^2}$

$Y(s) = \frac{\frac{5}{s^2} + 2s + 8}{(s-1)(s-9)}$

$= \frac{5 + 2s^3 + 8s^2}{s^2(s-1)(s-9)}$

Use partial fraction method,

$Y(s) = \frac{A}{s} + \frac{B}{s^2} + \frac{C}{s-1} + \frac{D}{s-9}$

$A(s-1)(s-9)s^2 + B\cdot s(s-1)(s-9) + C\cdot s^3(s-9) + D\cdot s^3(s-1) = 2s^3 + 8s^2 + 5$

$A = \frac{50}{81}$, $B = \frac{5}{9}$, $C = \frac{31}{81}$, $D = -2$

Using inverse Laplace transform,

$Y(s) = \frac{50}{81} + \frac{5}{9}t + \frac{31}{81}e^t - 2e^{9t}$

Example 2:

$2y'' + 3y' - 2y = te^{-2t}$, $y(0) = 0$, $y'(0) = -2$

$2s^2Y(s) - 2sy(0) - 2y'(0) + 3sY(s) - 3y(0) - 2Y(s) = \frac{1}{(s+2)^2}$

$Y(s)(2s^2 + 3s - 2) = \frac{1}{(s+2)^2} - 4$

$Y(s) = \frac{1 - 4(s+2)^2}{(2s-1)(s+2)^2}$

$Y(s) = \frac{A}{2s-1} + \frac{B}{s+2} + \frac{C}{(s+2)^2} + \frac{D}{(s+2)^3}$

By comparing coefficient,

$A = -\frac{192}{125}$, $B = \frac{96}{125}$, $C = -\frac{2}{25}$, $D = -\frac{1}{5}$

By Inverse Laplace transform,

$Y(s) = \frac{1}{125}\left(-96e^{\frac{t}{2}} + 96e^{-2t} - 10te^{-2t} - \frac{25}{2}t^2e^{-2t}\right)$

9. Bilateral Laplace Transform

The Laplace transform can also be defined as bilateral Laplace transform. This is also known as two-sided Laplace transform, which can be performed by extending the limits of integration to be the entire real axis.

The bilateral Laplace transform F(s) is defined as follows:

$$F(s) = \int_{-\infty}^{\infty} e^{-st} f(t)\, dt \quad \text{(Eq.2)}$$

10. Bilateral Laplace transform vs Laplace transform

Laplace transform

- Domain: For $t \geq 0$, meaning that only applicable for positive time value

- ROC: One single ROC for each function ($0 < t < \infty$)

- Application: used to analyze causal system that only has positive time values

Bilateral Laplace transform

- Domain: For $-\infty < t < \infty$, meaning that also works for negative time values

- ROC: Two separate ROC for domain $-\infty < t < 0$ and $0 < t < \infty$

- Application: used to analyze non-causal system with both past and future dependencies

11. Inverse Laplace Transform

Inverse Laplace transform allows us to obtain the initial function from its Laplace transform.

$$f(t) = \mathcal{L}^{-1}\{F\}(t) = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma-iT}^{\gamma+iT} e^{st} F(s)\, ds \quad \text{(Eq.3)}$$

12. Application of Laplace Transform

Besides solving differential equations, its significance also lies in the following fields:

- Control Systems Analysis: Laplace transform plays a crucial role in the analysis and design of control systems. It allows engineers to analyze the behavior of the system in the frequency domain, determine stability, and design controllers. The Laplace domain representation simplifies the analysis of feedback control systems and helps in stability analysis using techniques like the transfer function.

- Signal Processing: In the field of signal processing, the Laplace transform is used to analyze and process continuous-time signals. It allows for the representation of signals in the frequency domain, enabling tasks such as filtering, convolution, and system analysis. The Laplace transform is particularly useful in studying the behavior of linear time-invariant systems.

- Network Analysis: The Laplace transform is applied in analyzing complex networks, such as communication networks or transportation networks. It helps in modeling and analyzing the behavior of interconnected systems, determining system response, and optimizing network performance.

13. Reference

1. https://byjusexamprep.com/laplace-transform-i

2. https://www.quora.com/What-exactly-is-the-ROC-region-of-convergence-in-a-Laplace-transform-and-Z-transform

3. http://www.math.utah.edu/~gustafso/f2010/laplaceTableProofs.pdf

4. https://en.wikipedia.org/wiki/Laplace_transform

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ EVENTS
FITZWILLIAM COLLEGE UNIVERSITY OF CAMBRIDGE

60

Mathematical Finance

Fitzwilliam College Online Summer School Journal · 2023 Summer

Financial Goals and Strategies for Achieving Long-term Financial Success

LIYUE YANG (Lisa)

1. Introduction

In this article, I aim to explore the path to attaining long-term profits in business investments. The context of this paper is set within the current era of rapid globalization and informatization. It begins by analyzing the business strategies and risks involved in traditional industries, followed by an examination of the appropriate business strategies in emerging industries to achieve financial success and mitigate risks. Finally, a synthesis of the findings is presented, along with an acknowledgment of the study's limitations and suggestions for future research directions.

2. Global Context: Globalisation and informatization

Since the middle of the 20th century, the scientific and technological revolution, particularly driven by information technology, has transcended the barriers of time and space, transforming the world into a "global village" .After the outbreak of the novel coronavirus pneumonia, offline cross-border activities have declined sharply, but online international activities have become more frequent, indicating that globalization has taken on a new era with the accelerated development of information technology.(Zhu Xianqiang & Xu Bijun,2021)Under the background of enterprise business globalization, there are some problems such as lack of good development environment, low penetration rate and insufficient integration.(Wang Dong, 2017)Consequently, an increasing number of enterprises are now focusing on identifying business strategies to ensure long-term financial success.

When considering financial goals and strategies for achieving long-term success, different perspectives emerge. Some believe that executives should employ mathematical calculations to determine the business costs and approximate profits, aiming to reach a break-even point to avoid losses. On the other hand, some argue that companies should prioritize market expansion during commercial development, even if it entails early-stage losses, as the successful development of the industry can yield substantial returns later on. In my view, achieving long-term financial success necessitates a balanced approach: investing the majority of funds in industries likely to thrive in the future and allocating a portion of the funds to industries with short-term profit potential to mitigate losses. However, how can we ensure long-term success rather than incurring losses?

3. Profit returns in traditional industry

The volatility of traditional industries in the stock market tends to be relatively stable, allowing investors to assess their performance by analyzing the return rate of past stocks and calculating turnover to ensure profitability. As a result, businesses can potentially achieve a financial return rate of 5% or even 10%.

Investing in traditional industries carries a lower level of risk, but there are still some inherent risks involved, especially in the securities market. The market's efficiency depends

on the ability of stock prices to respond to relevant information freely, as well as the full disclosure and even distribution of securities-related information. Stock prices serve as one of the most effective indicators to reflect all known information in the capital market. This information, in turn, determines the market price of the stock, enabling investors to promptly adjust their asset investments in response to new information. A pertinent example is the recent media exposure of "Pinduoduo" selling counterfeit goods, which led to a swift decline in its stock price (Yin Jingyi, 2020). Therefore, achieving long-term financial success through investments in traditional industries, such as stocks or securities, necessitates careful selection of relevant industries and constant monitoring of news to ensure

profitable outcomes.

4. Profit returns in emerging industry

In today's rapidly evolving world, advancements in science and technology, such as VR, AI, and other emerging technologies, have led to the emergence of promising industries. With the ongoing trend of global informatization, these industries hold significant development prospects for the future. As astute investors recognize the potential, they invest in these nascent industries to secure market share and aim for substantial returns of 15%, 50%, or even 100% as these sectors mature.

However, these lucrative opportunities come with considerable risks. The overall securities market can be influenced by irrational investor sentiment, leading to a speculative atmosphere. Moreover, if the company does not have the consistent growth and profitability after the issuance at a high price, the situation of continuous slump, long-term downturn or weakness after the listing of new shares is unavoidable. It is not conducive to the long-term, stable and healthy development of the securities market.(Wu Lili, 2022). Therefore, the key to success lies in making rational investment choices in these emerging industries and knowing when to sell to maximize profits. A prudent approach will help investors navigate the risks and capitalize on the growth potential of these technologies and industries.

5. Conclusion

Based on the discussions above, the most robust strategy for achieving long-term financial success is a combination of investments in both traditional and new industries. Investing in traditional industries provides a source of relatively stable funds, which can act as a protective measure for investments in emerging industries, thereby achieving a balance and forming a hedge to secure profits in various market situations.

In conclusion, this paper highlights the advantages and disadvantages of both traditional and emerging markets, offering valuable insights for investors. However, it's essential to acknowledge that this study is limited to the Chinese market. Future research will aim to delve into the international market, providing a broader perspective on investment

opportunities and strategies beyond national boundaries. Expanding the analysis to the global context will offer more comprehensive guidance for investors seeking long-term financial success.

The impact of stock changes in the hands of large share holders on individual investors

JINKUN JI (Lofan)

Abstract

Large shareholders make up a large proportion of the company, and changes in the number of shares held often have a significant impact on individual investors. However, at present, there is no in-depth research on the specific impact of the change of major shareholders' stocks on individual investors. This paper will discuss the impact of stock changes in the hands of major shareholders on individual investors, and explore the influence mechanism and economic effects.

1. Introduction

The influence of stock changes in the hands of major shareholders on individual investors is a relatively important research field. In the stock market, large shareholders usually hold a high proportion of stocks, and the cost is low. When major shareholders reduce their holdings through trading, the number of shares sold is large and the transaction price is relatively low. This behavior often has an impact on the psychology of individual investors who hold shares that have been sold by large shareholders, interfering with their valuation of the company. In addition, arbitrage traders in the market will also use block trading to take over the shares sold by major shareholders at a lower trading price, and gradually sell them through the continuous bidding market to obtain profits. Since the cost of holding stocks is low for carry traders, they may exert some pressure on the stock price when they sell stocks later, especially for some stocks with poor liquidity and small circulating market value.

In the modern stock market, individual investors increase their wealth and diversify their risks by buying stocks. Therefore, it is of great significance for individual investors to make reasonable decisions in the stock market to understand the influence of major shareholders' stock changes on individual investors.

The research work of this paper mainly includes the following parts:

(1) The first part introduces the relationship between major shareholders' stock changes and individual investors.

(2) The second part discusses the reaction of individual investors to the changes of major shareholders' stocks.

(3) The third part is empirical research and analysis.

(4) The fourth part is case study.

(5) The fifth part, summary and prospect.

2. Literature review

Methodology

2.1. Analysis of the influence of major shareholders' stock movements on individual investor

The stock movements of major shareholders have an important impact on individual investors. First, the stock movements of major shareholders may trigger uncertainty in the market, leading to mood swings among individual investors. When large shareholders reduce their holdings, the market may panic, and individual investors may sell their shares, causing prices to fall. On the contrary, if large shareholders increase their holdings of stocks, the market may

be optimistic, and individual investors may buy more stocks, driving the price up. Therefore, individual investors need to pay close attention to the stock movements of major shareholders, and the impact on market sentiment.

Second, changes in the shares of major shareholders may change the corporate governance structure and affect the rights and interests of individual investors. An increase in stock holdings by major shareholders could mean they have more confidence in the company and greater control over it. This may lead to changes in the corporate governance structure, giving individual investors less say. In addition, the reduction of shares by major shareholders may make the company out of the control of major shareholders, resulting in the instability of the corporate governance structure, and then affecting the interests of individual investors. Therefore, individual investors need to pay attention to the stock changes of major shareholders in order to understand the changes in corporate governance structure and make corresponding investment decisions.

In addition, the changes of major shareholders' stocks may also include the prediction of the company's performance and development prospects, which will have an impact on the investment returns of individual investors. If large shareholders sell down their shares because the company is in trouble or poorly run, individual investors may face potential risks. An increase in shares by large shareholders could mean that they are optimistic about the company's prospects, which could provide investment opportunities for individual investors. Therefore, individual investors need to make a comprehensive analysis of the company's performance and development prospects according to the stock changes of major shareholders to decide whether to buy or sell stocks.

2.2. Study on the reaction of individual investors to the changes of major shareholders' stocks

As one of the participants of the market, the individual investor's reaction to the stock changes of major shareholders is an important part of the investment decision. In this section, the reaction of individual investors to the changes of major shareholders' stocks and the influencing factors will be studied.

First, the reaction of individual investors to changes in the shares of major shareholders may be influenced by different access to information. According to past studies, individual investors mainly obtain relevant stock information through media, Internet and research reports. However, there are often differences in individual investors' ability to obtain and interpret information, which leads to differences in individual investors' reactions to major shareholders' stock changes. Some individual investors may make quick decisions to adjust their portfolios, while others may be too conservative or too aggressive, missing out on investment opportunities or taking on too much risk.

Secondly, the reaction of individual investors to the changes of major shareholders' stocks is also affected by psychological factors. Investment decision is often affected by personal preference, risk attitude, cognitive bias and other factors. For example, some individual investors may blindly follow the decisions of major shareholders and change themselves out of trust in them or pursuit of short-term profits. This behavior can easily lead to excessive trading and bad decisions, which can negatively impact the financial situation of individual investors.

In addition, the reaction of individual investors to changes in the stocks of major shareholders is also affected by the market environment. Market risk appetite, liquidity, volatility and other factors will affect the attitude and behavior of individual investors in the face of major shareholders' stock changes. When market optimism is high, individual investors may be more inclined to follow the buying behavior of major shareholders, believing that it can bring good investment returns. When market panic spreads, individual investors may be more susceptible to the selling behavior of major shareholders, and then quickly follow the trend to sell stocks, resulting in further market declines.

Therefore, the reaction of individual investors to the changes of major shareholders' stocks will be influenced by the

information access channels, psychological factors and market environment.

3. Conclusion

The following conclusions can be drawn from the analysis of the impact of changes in the shares held by major shareholders on individual investors.

First, the stock movements of major shareholders have a significant impact on the psychology of individual investors. According to the results of the research individual investors generally show anxiety and anxiety when major shareholders reduce their stock holdings. This is because the reduction means that the major shareholders' confidence in the company is declining, which may indicate that the company is facing some bad situations. Individual investors, affected by such news, are likely to choose to reduce their holdings in order to avoid risk.

Secondly, the operation behavior of major shareholders has a certain impact on the trading behavior of individual investors. The study found that when large shareholders increase their holdings of stocks, individual investors are generally positively influenced to increase their own investments. This is because the increase of major shareholders is often seen as an optimistic expectation of the company's prospects, and individual investors will follow the pace of major shareholders and increase their confidence in the company's investment. On the contrary, the reduction behavior of major shareholders will cause panic among individual investors, leading them to choose to reduce or suspend investment.

In addition, the stock changes of major shareholders also have an impact on the investment returns of individual investors. According to research results, when major shareholders increase their holdings of stocks, the investment returns of individual investors tend to increase. This is because the increase of major shareholders is often interpreted by the market as confidence in the future development of the company, which leads to expectations of rising stock prices. Individual investors will therefore choose to increase their investment in the company and get better returns. However, when major shareholders reduce their holdings of stocks, the investment income of individual investors is often suppressed to a certain extent, because the behavior of reducing holdings often brings negative expectations of the market, resulting in a decline in stock prices.

Finally, we also find that the investment behavior of individual investors also has a certain feedback effect on the operation of major shareholders. Specifically, when individual investors actively follow the actions of major shareholders, some major shareholders may take advantage of the situation to operate, thereby affecting the market. This feedback effect shows the mutual utilization between major shareholders and individual investors, and further deepens the connection between the two.

To sum up, the stock movements of major shareholders have a significant impact on individual investors. Individual investors will be affected by emotions and behaviors when facing the operation of major shareholders, and consider the operation behavior of major shareholders and market conditions in future investment decisions. Therefore, in the decision-making process of individual investors, it is necessary to conduct appropriate analysis and judgment on the operational behavior of major shareholders in order to reduce risks and obtain better investment returns. At the same time, major shareholders should also be fully aware of the impact of their own operational behavior on individual investors, actively guide market expectations, and promote the stability and healthy development of the market.

4. Reference

1. Luo Jinhui, Xiang Yuan-Gao, Jin Sijing. Major shareholders' equity pledge and stock suspension manipulation: Based on the "thousand shares suspension" event [J].Finance and Economics,2020

2. Wang Xiaotong. An Empirical Study on the impact of investor Sentiment on stock returns [J]. New Business Weekly,2018

3. Yue Shizhong, Zhu Xiaoxi. Major shareholder reduction, investor sentiment and stock price crash risk [J]. Social Science & Technology,2020

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

Fitzwilliam College
Cambridge
CB3 0DG
+44 (0)1223 332000

Fitzwilliam College
https://www.fitz-events.com/academic-programmes
Fitzwilliam College Online Summer School
https://www.seedasdan.com/en/fitz/

阿思丹 | ASDAN
China