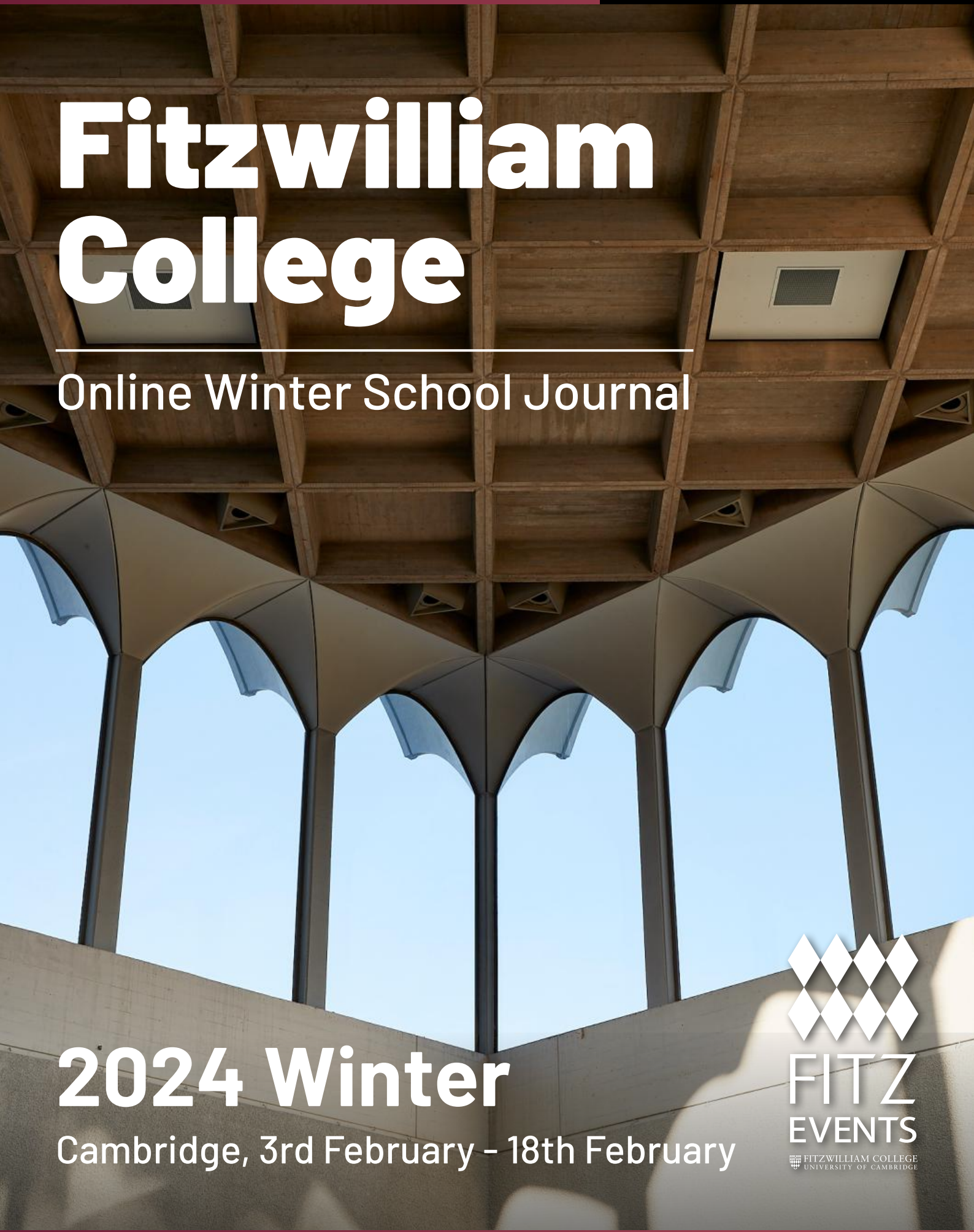
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

阿思丹 | ASDAN China



Fitzwilliam College

Online Winter School Journal

2024 Winter

Cambridge, 3rd February - 18th February

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 of the top-performing students at **the Fitzwilliam College Online Winter School Programme.** The journal covers seven fields ranging from Biology, Chemistry, Computer Science, Engineering, Mathematics, Physics to Neuroscience.

About the Programme

The Fitzwilliam College Online Winter School is **the official Winter School Programme of Fitzwilliam College in the University of Cambridge.** All the courses were taught by academic members of Fitzwilliam College or one of the other colleges of the University of Cambridge. **The same academics who lecture and supervise the undergraduate students at Cambridge taught undergraduate-level content.**The core of Fitzwilliam's academic activities is a desire to retain 'the best of the old', while enthusiastically embracing 'the best of the new'. 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 was **designed to provide students with a flavour of undergraduate study at Cambridge, and an opportunity to explore topics beyond what is covered within the school curriculum.**

In 2023, Fitzwilliam College and ASDAN China have entered a strategic partnership to open the Fitzwilliam College Online Summer School Programme to outstanding high school students in China. Nine subject areas were opened for the summer of 2023 and the winter of 2024, and more than 160 students from across the country participated in these programmes.

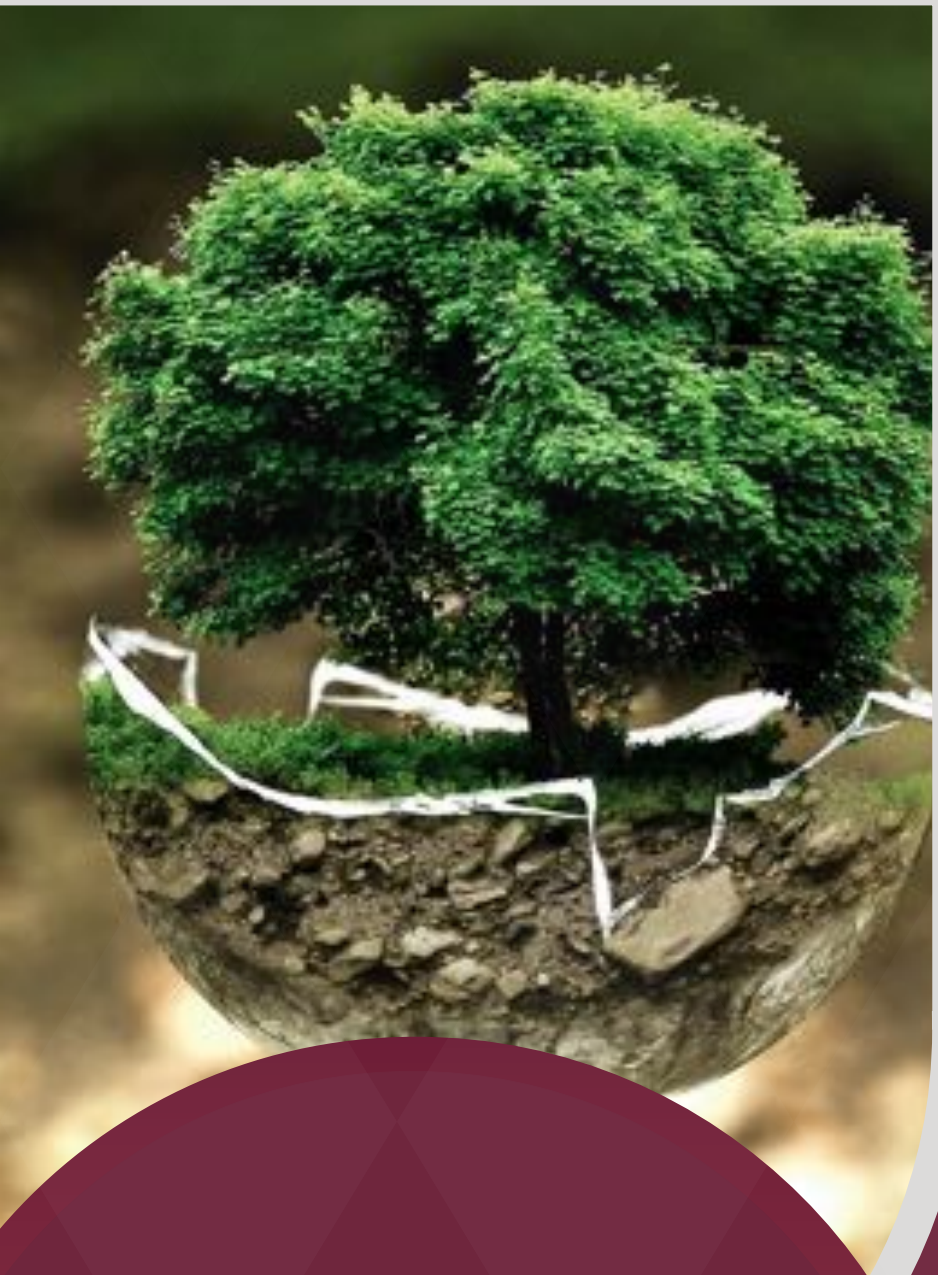
List of Academic Course Instructors

DR ASHRAF ZARKAN
DR ALEXANDRA KRUGLIAK
DR ANDREA GIUSTI
DR GIULIA IADEVAIA
DR JOAO RODRIGUES
DR JOHN FAWCETT
MRS. SERENA POVIA

Table of Contents

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE



01

Biology

Fitzwilliam College Online Winter School Journal · 2024 Winter

Pathogenicity and clinical relevance of Streptococcus pneumoniae

FELICITY JINGYUE ZHANG (Felicity)

1.Introduction

Streptococcus pneumoniae (S. pneumoniae) is the leading cause of community-acquired pneumonia, meningitis, and bacteraemia worldwide. The WHO estimates that over 300,000 children younger than five years die annually due to *S. pneumoniae* infections, making it the largest infectious cause of deaths in children [1, 2]. Although morbidity has significantly dropped since the introduction of PCV7 and other pneumococcal vaccines, the rapid emergence of new antibiotic-resistant serotypes poses a concern. This essay aims to provide a surface-level overview of the pathogenicity and clinical relevance of *S. pneumoniae*.

2. Pathogenicity

2.1. Propulsion system

S. pneumoniae is a non-motile, non-spore-forming, Gram-positive lancet-shaped coccus, usually appearing in pairs, or as diplococci (Figure 1). *S. pneumoniae* can be differentiated from other Streptococci by its alpha haemolytic and optochin-resistant properties. The bacterium has a 6-layer-thick peptidoglycan cell wall containing teichoic acid, lipoteichoic acid, and the enzymes neuraminidase and IgA protease. On the outer surface of the bacterium pili and an antigen-containing capsular polysaccharide (CPS) can be found [3-5]. *S. pneumoniae* serotypes differ by their CPS; 100 CPS types were recognised by 2020 [6].

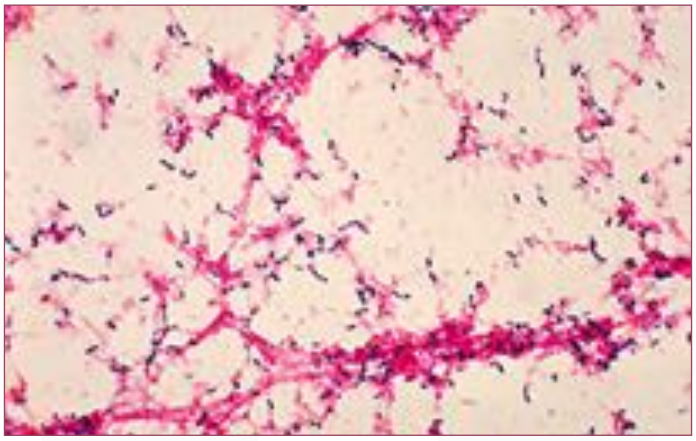


Figure 1 *Streptococcus pneumoniae* gram-stain of blood broth culture CDC [7].
The purple cells show gram-positive *S. pneumoniae* bacterial cells. The pink network in the back is extracellular matrix that has taken up the counterstain safranin. The purple *S. pneumoniae* cells are present in pairs (diplococci) or short chains.

2.2. Virulence factors

Cell-associated

Several cell-associated virulence factors are vital for *S. pneumoniae* to interact with host cells. Pili induce inflammation as well as help the bacterium adhere to surfaces [8]. Teichoic and lipoteichoic acids within the peptidoglycan cell wall contain phosphorylcholine, allowing the bacterium to adhere to choline-binding receptors on host cells [9]. The CPS impedes phagocytosis by inhibiting binding of the C3b complement, which is involved in initiating opsonisation [10].

S. pneumoniae can also actively counteract adaptive immune response by CPS shedding upon interaction with epithelial antimicrobial peptides [11]. LytA, an autolytic enzyme that cleaves peptidoglycan and was thought to regulate autolysis solely,

propels this process [12]. P. Mellroth et al., 2012 indicated that LytA only induces autolysis under certain circumstances and when a 'critical threshold' is met [13]. Referencing this, C. C. Kietzman, G. Gao, B. Mann, L. Myers, and E. I. Tuomanen, 2016 identified a new physiological function of LytA outside the growth zone: shedding and replenishment of CPS [12]. This explains why, despite high evolutionary pressure, *S. pneumoniae* has conserved a 'suicidal' autolytic system [12] – to evade phagocytosis and adherence to antibiotic peptides.

Secretory

IgA is an antibody that coats mucous membranes and neutralises toxins. *S. pneumoniae* combats this by secreting IgA proteases that cleave and destroy IgA [5, 8]. *S. pneumoniae* also expresses neuraminidases, enzymes that remove sialic acids from host cell membranes and facilitate the progression of the bacterium further into infected tracts by cleaving mucous [11]. Removal of sialic acids reduces the binding of regulator Factor H to host cells [14]. Factor H interacts with C3b to ensure that host cells are not mistargeted [15]; neuraminidase disrupts this defence mechanism.

The single exotoxin of *S. pneumoniae* is pneumolysin (Ply). Ply triggers vigorous inflammation and lyses host cells by forming membrane pores. It especially affects endothelial cells and the lung epithelium [3], and is a critical obstruction that leads to invasive diseases [16].

2.3. Transmission

Cell-associated

S. pneumoniae is transmitted through respiratory droplets [7]. Colonisation is present in the oropharynx and nasopharynx of a healthy individual as part of the microbiota [8]. Under compromised conditions, it can become pathogenic and cause pneumonia, meningitis, otitis media and sinusitis [17].

3. Clinical relevance

3.1. Diagnosis

Initial clinical evaluation examines patients' symptoms. The common symptoms of pneumococcal infections are high fever, chills, shortness of breath, cough, chest and neck pains, and aches [17].

Laboratory diagnosis of *S. pneumoniae* is essential before prescribing antibiotics; the most common approaches taken are microscopy, cell culture, serologic testing and PCR. Gram-staining, haemolysis tests and optochin sensitivity tests allow identification of *S. pneumoniae* for its distinct properties. Urinary antigen tests, CIEP and Latex agglutination tests examine CPS antigens, whereas ELISA and fluorescent antibody tests detect antibodies in invasive infections [8, 18]. PCR and qRT-PCR can be used to identify specific serotypes of the bacterium. The standard gene for amplification is LytA due to its prevalence and specificity to *S. pneumoniae* [19, 20].

3.2. Antibiotic Resistance

Antimicrobial resistance is driven by evolution and natural selection; the primary mechanisms by which *S. pneumoniae* develop antibiotic resistance is by altering or inactivating surface binding proteins, enzymes, and efflux pumps through transposons (nicely summarised in Table 2.1 in [8]) [21, 22].

Aside from the rapid growth rate of *S. pneumoniae* [9], the bacterium's natural competence is a chief driver of antibiotic resistance development. With genes involved in the induction of the competent state occupying 10% of the bacterium's genome [23], the endogenous ability to acquire genes via transformation enables quick sharing of genes between serotypes as well as other Streptococci that encode for antibiotic resistance and virulence factors [24].

Penicillin-resistant *S. pneumoniae* are more likely to develop antibiotic resistance towards other antibiotics compared to non-resistant serotypes (see emergence of MDR in Figure 2) [25]. This contributes to increasing multidrug resistance, commonly towards ß-lactams, macrolides, and fluoroquinolones, antibiotics primarily used in treatment, as illustrated in Figure 3 [24, 26]. Recent research also indicates diversely-ranging persistence in *S. pneumoniae* [28], however, exploration

in the field is relatively nascent.

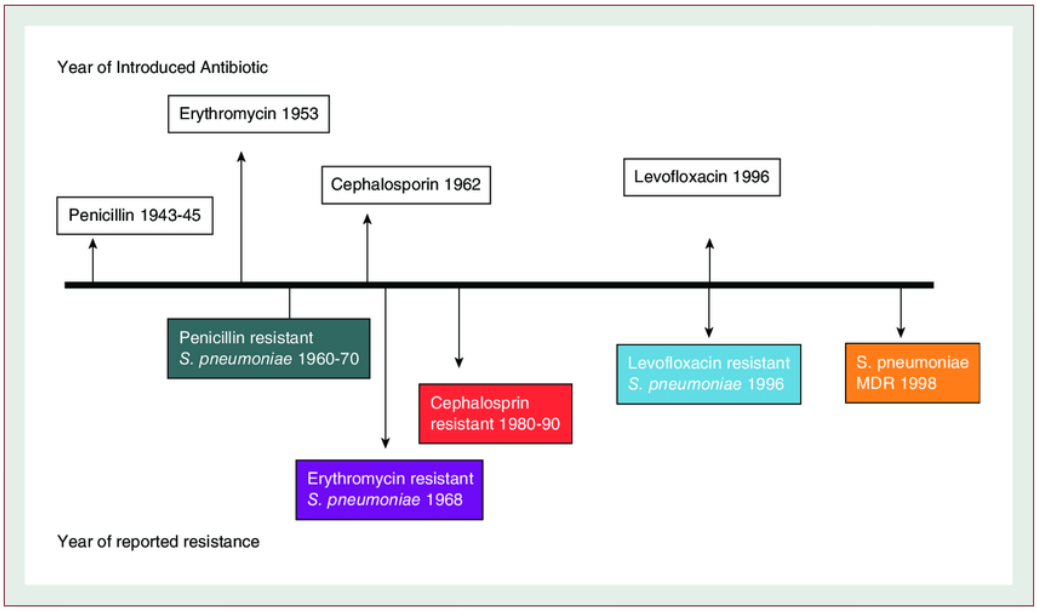


Figure 2 Timeline of antibiotic resistance of *S. pneumonia* [27]
The timeline illustrates how quickly *S. pneumoniae* has developed resistance towards newly introduced antibiotics. MDR (multidrug resistance) emerged in 1998 after introducing ß-lactams penicillin and cephalosporin, the macrolide erythromycin, and the fluoroquinolone levofloxacin. Barcelona Respiratory Network Reviews Rights: 'Excerpts may be used for purposes of research and teaching, providing the use is proportional to the purpose and the author and source are cited.'

Streptococcus Pneumoniae	
Antimicrobial	% Susceptible
Amoxicillin	97
Azithromycin	61
Cefotaxime	99
Cefpodaxime	98
Ceftriaxone	99
Clarithromycin	64
Clindamycin	70
Erythromycin	57
Levofloxacin	99
Meropenem	96
Moxifloxacin	100
Penicillin	77
Tetracycline	75
TMP-SMX	79
Vancomycin	100

Figure 3 *Streptococcus pneumoniae* example antibiogram Contributed by Zachary Sandman, BA, 2024 Jan-.
Resistance is present in those antibiotics towards which *S. pneumoniae* has relatively low susceptibility. In the 57-77% susceptibility range we find the ß-lactam penicillin and the macrolides azithromycin, clarithromycin, and erythromycin, which are all antibiotics that *S. pneumoniae* has resistance to. In the 97-100% susceptibility range we find antibiotics that are comparatively more effective against *S. pneumoniae* such as the ß-lactams cefpodoxime, cefotaxime, ceftriaxone, and amoxicillin.

3.3. Vaccines

As a primary and ubiquitous virulence factor of *S. pneumoniae*, CPS antigens are targeted to provide serotype-specific protection in the two available classes of pneumococcal vaccines: polysaccharide-based pneumococcal vaccines (PPVs) and pneumococcal conjugate vaccines (PCVs) [2, 29] (Table 1). CPS in PPV vaccines are T-cell independent antigens that are generally ineffective in children younger than two years due to failure to induce T-cell memory. CPS has been conjugated to toxoids to elicit a T-cell dependent humoral response, significantly lowering pneumococcus cases as demonstrated in Figure 4.

However, current vaccines lack full serotype coverage and result in vaccine-induced serotype replacement: the rising prevalence of non-vaccine serotypes in infections due to selective pressure (Figure 4) [30, 31]. An interesting non-vaccine serotype to study further is nonencapsulated *S. pneumoniae* (NESp), an unusual serotype that is reported to have caused more infections after the introduction of PCVs [31]. Studies indicate that NESp's pathogenicity significantly differs from encapsulated *S. pneumoniae*; therefore, a potential way forward is to target NESp-specific surface proteins, including Pspk [32, 33].

Table 1. Bullet point summary of the serotypes and indications of pneumococcal vaccines in use. Table compiled from [34, 35]. * = recommended by CDC.

Vaccine	Serotypes	Indications
PCV13	1, 3, 4, 5, 6A and 6B, 7F, 9V, 14, 19A, and 19F, 18C, and 23F	• Children six weeks -five years of age: Prevention of IPD (invasive pneumococcal disease) caused by *Streptococcus pneumoniae* and for the prevention of otitis media. • Children six years through seventeen years of age: Prevention of IPD caused by *Streptococcus pneumoniae*. • Adults ≥ 18 years: Prevention of IPD and pneumonia caused by *Streptococcus pneumoniae*.
PCV15*	1, 3, 4, 5, 6A and 6B, 7F, 9V, 14, 19A, and 19F, 18C, and 23F	• The prevention of IPD caused by *Streptococcus pneumoniae* in individuals six weeks of age and older. • PCV13 and PCV15 can be used interchangeably.
PCV20*	1, 3, 4, 5, 6A, and 6B, 7F, 8, 9V, 11A, 10A, 12F, 14, 15B, 18C, 19A and 19F, 22F, 23F, and 33F	• PCV 20 is indicated for preventing pneumonia and invasive pneumococcal diseases in adults 18 years of age and older.
PCV23*	1, 2, 4, 3, 5, 6B, 7F, 8, 9N and 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A and 19F, 20, 22F, 23F, and 33F	• PPSV23 is approved in individuals ≥ 50 age. • PPSV 23 is also indicated in children ≥ two years of age at high risk of pneumococcal infection.

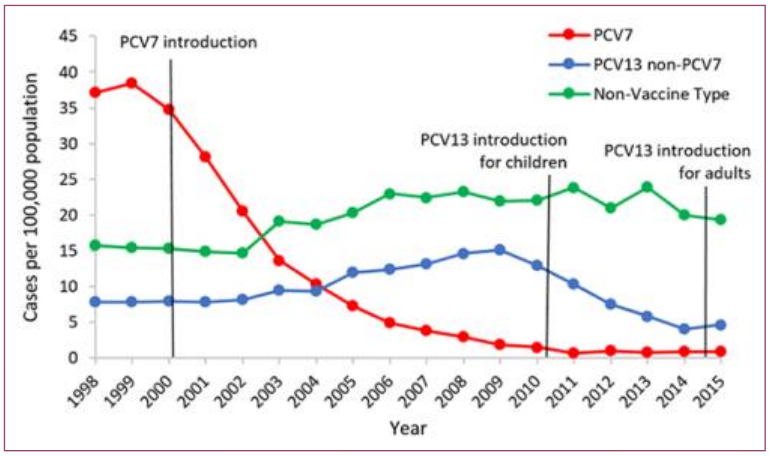


Figure 4 Rates of invasive pneumococcal disease among U.S. adults>65 years of age, 1998-2015. CDC [36]. The number of cases decreased significantly after PCV7 introduction, as well as PCV13 introduction for children (herd immunity) and PCV13 introduction for adults. Despite being small compared to the overall reduction in cases, an increase in disease caused by non-vaccine serotypes of *S. pneumoniae* was observed after the introduction of PCV7. Surveillance of different *S. pneumoniae*-serotype-causing diseases is crucial for detecting trends in drug resistance.

4. Conclusion

Transmitted through droplets, *S. pneumoniae* has several virulence factors that contribute to its pathogenicity, including serotype-variable CPS, LytA-activated CPS shedding, exoenzymes IgA and neuraminidase, and exotoxin pneumolysin. The bacterium's growing resistance to drugs used in treatment is attributed to its natural competence and varying serotypes. Improving existing PCVs by increasing serotype coverage or utilising serotype-independent antigens, ensuring correct prescription of antibiotics using accurate diagnosis, as well as researching further into virulence factors and phenotypic resistance, are essential to combating pneumococcal diseases.

5. References

[1] 'Global Pneumococcal Disease and Vaccination | CDC'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.cdc.gov/pneumococcal/global.html

[2] L. Geddes, 'Vaccine profiles: Pneumococcus | Gavi, the Vaccine Alliance', Gavi The Vaccine Alliance. Accessed: Feb. 15, 2024. [Online]. Available: https://www.gavi.org/vaccineswork/routine-vaccines-extraordinary-impact-pneumococcus

[3] J. N. Weiser, D. M. Ferreira, and J. C. Paton, '*Streptococcus pneumoniae*: transmission, colonization and invasion', Nat. Rev. Microbiol., vol. 16, no. 6, Art. no. 6, Jun. 2018, doi: 10.1038/s41579-018-0001-8.

[4] '15.1.2.2.1: *Streptococcus pneumoniae*', Biology LibreTexts. Accessed: Feb. 15, 2024. [Online]. Available: https://bio.libretexts.org/Courses/Mansfield_University_of_Pennsylvania/BSC_3271%3A_Microbiology_for_Health_Sciences_Sp21_(Kagle)/15%3A_Selected_Pathogens/15.01%3A_Bacterial_Pathogens/15.1.02%3A_Gram_Positive_Cocci/15.1.2.02%3A_The_Streptococci/15.1.2.2.01%3A_Streptococcus_pneumoniae

[5] K. Todar, 'Online Textbook of Bacteriology'. Accessed: Feb. 15, 2024. [Online]. Available: https://textbookofbacteriology.net/index.html

[6] 'Pinkbook: Pneumococcal Disease | CDC'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.cdc.gov/vaccines/pubs/pinkbook/pneumo.html

[7] '*Streptococcus pneumoniae*: Information for Clinicians | CDC'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.cdc.gov/pneumococcal/clinicians/streptococcus-pneumoniae.html

[8] C. F. Dion and J. V. Ashurst, '*Streptococcus pneumoniae* - StatPearls - NCBI Bookshelf'. Accessed: Feb. 15, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK470537/

[9] K. Todar, '*Streptococcus pneumoniae* and pneumococcal pneumonia', Textbookofbacteriology.net. Accessed: Feb. 15, 2024. [Online]. Available: https://textbookofbacteriology.net/S.pneumoniae.html

[10] M. T. Madigan, J. M. Martinko, D. A. Stahl, and D. P. Clark, Brock Biology of Microorganisms, 13th ed. Pearson Education, 2011.

[11] A. J. Loughran, C. J. Orihuela, and E. I. Tuomanen, '*Streptococcus pneumoniae*: Invasion and Inflammation', Microbiol. Spectr., vol. 7, no. 2, p. 10.1128/microbiolspec.GPP3-0004–2018, Mar. 2019, doi: 10.1128/microbiolspec.GPP3-0004-2018.

[12] C. C. Kietzman, G. Gao, B. Mann, L. Myers, and E. I. Tuomanen, 'Dynamic capsule restructuring by the main pneumococcal autolysin LytA in response to the epithelium', Nat. Commun., vol. 7, p. 10859, Feb. 2016, doi: 10.1038/ncomms10859.

[13] P. Mellroth et al., 'LytA, Major Autolysin of *Streptococcus pneumoniae*, Requires Access to Nascent Peptidoglycan', J. Biol. Chem., vol. 287, no. 14, pp. 11018–11029, Mar. 2012, doi: 10.1074/jbc.M111.318584.

[14] S. Syed et al.,'Role of Pneumococcal NanA Neuraminidase Activity in Peripheral Blood,'Front. Cell. Infect. Microbiol., vol. 9, 2019, Accessed: Feb. 15, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcimb.2019.00218

[15] G. Kaiser, '13.2A: Opsonization', Biology LibreTexts. Accessed: Feb. 15, 2024. [Online]. Available: https://bio.libretexts.org/Bookshelves/Microbiology/Microbiology_(Kaiser)/Unit_6%3A_Adaptive_Immunity/13%3A_Humoral_Immunity/13.2%3A_Ways_That_Antibodies_Help_to_Defend_the_Body/13.2A%3A_Opsonization

[16] U. Koppe, N. Suttorp, and B. Opitz, 'Recognition of *Streptococcus pneumoniae* by the innate immune system - Koppe - 2012 - Cellular Microbiology - Wiley Online Library'. Accessed: Feb. 15, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1462-5822.2011.01746.x#b62

[17] A. Jacot, '*Streptococcus Pneumoniae* - Infectious Disease Advisor'. Accessed: Feb. 15, 2024. [Online]. Available: https://www.infectiousdiseaseadvisor.com/ddi/streptococcus-pneumoniae/

[18] S. Aryal,'Laboratory diagnosis, Treatment and Prevention of *Streptococcus pneumoniae*'Accessed: Feb. 15, 2024. [Online]. Available: https://microbenotes.com/laboratory-diagnosis-treatment-and-prevention-of-streptococcus-pneumoniae/

[19] 'Pneumococcus Streptococcus Lab Resources and Protocols | CDC'. Accessed: Feb. 15, 2024. [Online]. Available: https://www.cdc.gov/streplab/pneumococcus/resources.html

[20] 'Identification of *Streptococcus pneumoniae* by a real-time PCR assay targeting SP2020 | Scientific Reports'. Accessed: Feb. 15, 2024. [Online]. Available: https://www.nature.com/articles/s41598-019-39791-1

[21] A. Gaurav, P. Bakht, M. Saini, S. Pandey, and R. Pathania, 'Role of bacterial efflux pumps in antibiotic resistance, virulence, and strategies to discover novel efflux pump inhibitors', Microbiology, vol. 169, no. 5, p. 001333, May 2023, doi: 10.1099/mic.0.001333.

[22] U. Iyer, 'Pneumococcal Infections (*Streptococcus pneumoniae*) Medication', Medscape. [Online]. Available: https://emedicine.medscape.com/article/225811-medication?form=fpf

[23] V. Minhas et al., 'Competence remodels the pneumococcal cell wall exposing key surface virulence factors that mediate increased host adherence', PLOS Biol., vol. 21, no. 1, p. e3001990, Jan. 2023, doi: 10.1371/journal.pbio.3001990.

[24] C. Cillóniz, C. Garcia-Vidal, A. Ceccato, and A. Torres, 'Antimicrobial Resistance Among *Streptococcus pneumoniae*',

Antimicrob. Resist. 21st Century, pp. 13–38, Mar. 2018, doi: 10.1007/978-3-319-78538-7_2.

[25] P. C. Appelbaum, 'Resistance among *Streptococcus pneumoniae*: Implications for Drug Selection', Clin. Infect. Dis., vol. 34, no. 12, pp. 1613–1620, Jun. 2002, doi: 10.1086/340400.

[26] L. M. Bush and M. T. Vazquez-Pertejo, 'Pneumococcal Infections - Infectious Diseases - Merck Manuals Professional Edition'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.merckmanuals.com/professional/infectious-diseases/gram-positive-cocci/pneumococcal-infections

[27] C. Cilloniz and A. Torres, 'Community-Acquired Pneumonia 2000-2015: What is New?', Barc. Respir. Netw., vol. 2, no. 4, p. 365, Oct. 2016, doi: 10.23866/BRNRev:2016-M0031.

[28] N. Geerts, L. De Vooght, I. Passaris, P. Delputte, B. Van den Bergh, and P. Cos, 'Antibiotic Tolerance Indicative of Persistence Is Pervasive among Clinical *Streptococcus pneumoniae* Isolates and Shows Strong Condition Dependence', Microbiol. Spectr., vol. 10, no. 6, pp. e02701-22, Nov. 2022, doi: 10.1128/spectrum.02701-22.

[29] 'Pneumococcal Disease', World Health Organisation. Accessed: Feb. 15, 2024. [Online]. Available: https://www.who.int/teams/health-product-policy-and-standards/standards-and-specifications/vaccine-standardization/pneumococcal-disease

[30] U. Obolski, J. Lourenço, C. Thompson, R. Thompson, A. Gori, and S. Gupta, 'Vaccination can drive an increase in frequencies of antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae*', Proc. Natl. Acad. Sci., vol. 115, no. 12, pp. 3102–3107, Mar. 2018, doi: 10.1073/pnas.1718712115.

[31] K. P. Klugman, 'The Significance of Serotype Replacement for Pneumococcal Disease and Antibiotic Resistance', in Hot Topics in Infection and Immunity in Children V, A. Finn, N. Curtis, and A. J. Pollard, Eds., in Advances in Experimental Medicine and Biology. , New York, NY: Springer US, 2009, pp. 121–128. doi: 10.1007/978-0-387-79838-7_11.

[32] H. Sakatani et al., 'Investigation on the virulence of non-encapsulated *Streptococcus pneumoniae* using liquid agar pneumonia model', J. Infect. Chemother., vol. 28, no. 11, pp. 1452–1458, Nov. 2022, doi: 10.1016/j.jiac.2022.07.003.

[33] L. E. Keller, D. A. Robinson, and L. S. McDaniel, 'Nonencapsulated *Streptococcus pneumoniae*: Emergence and Pathogenesis', mBio, vol. 7, no. 2, p. 10.1128/mbio.01792-15, Mar. 2016, doi: 10.1128/mbio.01792-15.

[34] S. Tereziu and D. A. Minter, 'Pneumococcal Vaccine', in StatPearls, Treasure Island (FL): StatPearls Publishing, 2024. Accessed: Feb. 16, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK507794/

[35] 'About Pneumococcal Vaccine: For Providers | CDC'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.cdc.gov/vaccines/vpd/pneumo/hcp/about-vaccine.html

[36] 'Pneumococcal - Vaccine Preventable Diseases Surveillance Manual | CDC'. Accessed: Feb. 16, 2024. [Online]. Available: https://www.cdc.gov/vaccines/pubs/surv-manual/chpt11-pneumo.html

The Pathogenicity & Clinical Relevance of Escherichia coli

YIYOU WU (Madeleine)

1. Introduction

Escherichia coli (E. coli) is a Gram-negative bacterium which is a normal inhabitant of the healthy gut where it is mostly harmless. However, E. coli can develop into pathogenic variants thereby becoming a human and animal pathogen. E. coli is highly adaptive to gene transfer which results in its ability of causing a range of diseases from mild gut infection to severe meningitis (Matthew Mueller, Christopher R. Tainter, 2023). E. coli has been listed as one of the top important abundant bacteria by World Health Organization (WHO) for its wide range of pathogenic variants (pathovars) with addition to its known multi-drug resistance. In this review, the pathogenicity of E. coli will be discussed, including current understanding of the pathogenic mechanisms of various E. coli pathovars. The clinical relevance of E. coli will also be explored with respective to the related diseases, treatments, and antibiotic resistance.

2. Pathogenicity of E. coli

2.1. Structure & Classification

2.1.1. Structure of E. coli

E. coli possess a bacterial envelope as all Gram-negative bacteria which consists of inner membrane (IM) and outer membrane (OM) that are separated by the periplasmic space (Malvina Papanastasiou et. al, 2013), as shown in Figure 1

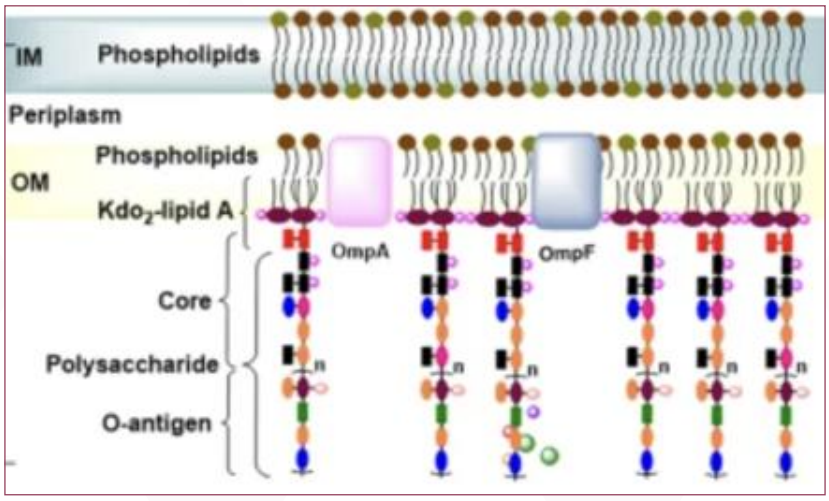


Figure 1. The membrane structure of E. coli, including the inner membrane (IM), periplasm, and outer membrane (OM). The OM include the major porin proteins OmpF and OmpC. The OM consists of phospholipids and lipopolysaccharide (LPS) containing three structural domains: 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo2)-lipid A, the core polysaccharide, and the O antigen (Jianli Wang, Wenjian Ma, Xiaoyuan Wang, 2021).

The OM serves as a layer for protective functions which consists of phospholipid and lipopolysaccharide (LPS) (Lugtenberg, 1981). As shown in Figure 1, 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo2) is the main part of LPS which, along with phospholipids, provide specific characteristics to the OM (Table 1). (Jianli Wang, Wenjian Ma, Xiaoyuan Wang, 2021).

Table 1. The impact of OM components on membrane characteristics of E. coli. Adapted from (Jianli Wang, Wenjian Ma, Xiaoyuan Wang, 2021).

Component of OM	Impact on membrane characteristics
Phospholipids	Membrane fluidity
	Stiffness
Kdo_2-lipid	Bio-renewable tolerance
	Pathogenicity

Interestingly, some OM components are associated with problematic features for industrial and clinical applications (e.g. cytotoxicity, immune response, adhesion, biofilm formation) and subsequently lead to safety and toxicity issues for some industries such as food industry, medical products, and chemical productions (e.g. amino acids, organic acids). In contrast, the IM contains a variety of proteins which function in vital cell processes and enclose the cytoplasm in a dynamic substructure (Malvina Papanastasiou et. al, 2013).

2.1.2. Classification of E. coli

E. coli is classified into 3 major groups, including commensal E. coli, intestinal pathogenic E. coli, and extraintestinal pathogenic E. coli as shown in Figure 2. The classification shows the pathovars of E. coli which vary in their ability to colonise different sites of the human body, leading to a range of infections. The classification of E. coli shown in Figure 2 provides an indication to the wide range of pathogenic mechanisms.

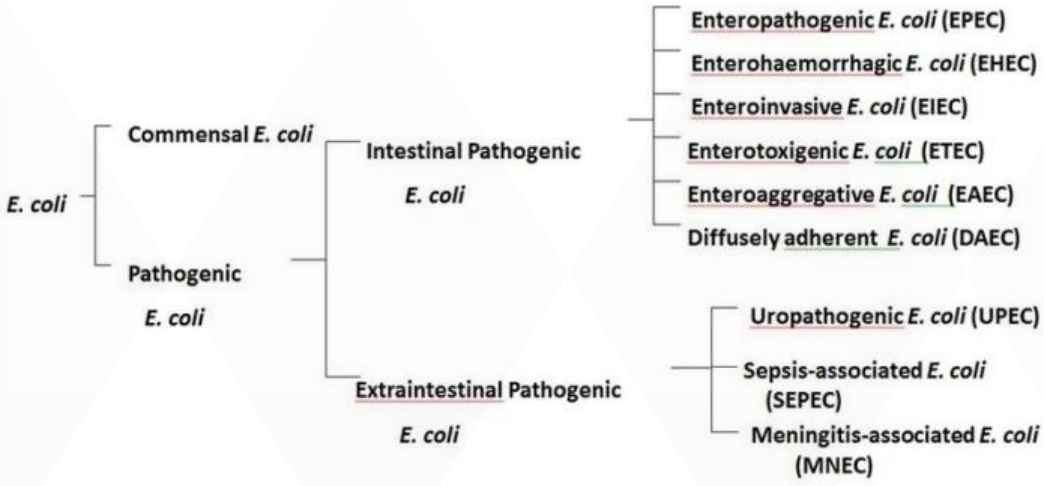


Figure 2. Classification of E. coli into 2 main categories, including commensal and pathogenic E. coli. Pathogenic E. coli can be further classified into intestinal and extraintestinal E. coli with a range of pathovars (James B Kaper et. al, 2004)

2.2. Properties & Transmission

2.2.1. Properties of E. coli

E. coli is a Gram-negative straight short bacillus around 1.0-2.0 micrometers long and has a radius of about 0.5 micrometers. It is the most abundant anaerobe and lactose fermenter that can be differentiated from Salmonella and Shigella via their pink coloured colonies on MacConkey agar. In addition, there are more than 100 antigenic types which can be categorised into O-cell wall antigens, H-flagellar antigens, and K-capsular antigen. (Tankeshwar, 2023). The main virulence factors of E. coli are summarized in Table 2.

Table 2. Virulence factors of E. coli and their respective function. Adapted from (Tankeshwar, 2023).

Virulence Factors	Impact on membrane characteristics
Pilli	Help to adhere organisms to jejunum and ileum when there is a intestinal infection; to urinary tract epithelium when there is urinary tract infection
Capsule	Interferes with phagocytosis and systemic infections
Endotoxin (lipopolysaccharide)	Responsible for features of Gram-negative sepsis (e.g. fever, DIC)
Exotoxin	Verotoxin, Shiga-like toxin

2.2.1. Transmission of E. coli

The transmission of intestinal E. coli is primarily through the consumption of contaminated food, raw or undercooked ground meat products, raw milk, and contaminated raw vegetables and sprouts. The most severe serotype is O157:H7, which is associated with gastroenteritis and designated as Shiga Toxin Producing E. coli (STEC) that differentiates it from other serotypes. (WHO, 2018). Along with the toxin, STEC carries the virulence adherence factor plasmid (pEAF) which encodes the virulence factor bundle-forming pilus (BFP) that mediates microcolony formation and localised adherence on the surface of host enterocytes. (Claudia Silva et. al, 2017). Figure 3 shows the immunopathology of STEC, which is triggered by dysbiosis that is induced by a variety of reasons (e.g. diet, colitis, inflammation), thereby increasing STEC colonisation, the delivery of Shiga toxins and production of cytokines and chemokines in intestinal epithelial cells (IEC). (Kyung-Soo Lee et. al, 2021). STEC can grow in temperatures ranging from 7° C to 50° C and has an optimum temperature of 37° C, but can be eliminated at 70° C or higher temperatures. STEC can also grow in acidic environment (pH of 4.4). STEC is associated with morbidity and large outbreaks. The colonisation with STEC is asymptomatic and fast which tend to be within one month. Reasons for STEC outbreaks have been classified into 3 pathways, including i) transmission through contaminated food and water or swimming water, ii) person to person transmission, and iii) transmission through animal contacts (WHO, 2018).

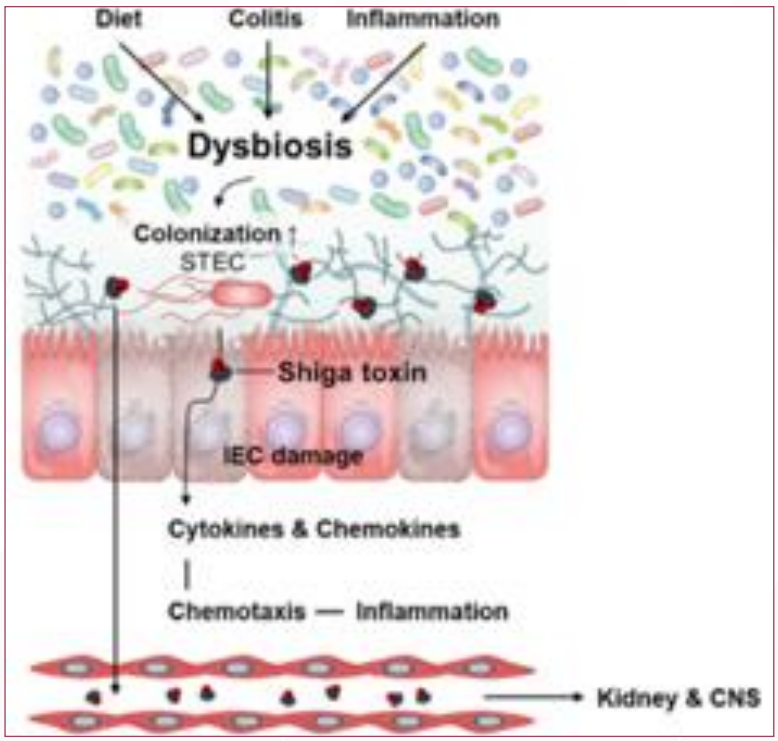


Figure 3. Immunopathology of STEC: induction of dysbiosis; increased colonisation of STEC; production of Shiga toxin leading to IEC damage; increased production of cytokines and chemokines leading to chemotaxis, subsequent inflammation, and further damage (Kyung-Soo Lee et. al, 2021).

Apart from STEC, other pathovars have different sites of colonisation in the human body as shown in Figure 4, leading to a range of diseases. E.g. the neonatal meningitis-causing E. coli (NMEC) crosses the blood-brain barrier into the central nervous system (CNS) and subsequently causes meningitis that mostly occur in newborn babies. (Matthew A. Croxenm B. Brett Finlay, 2009).

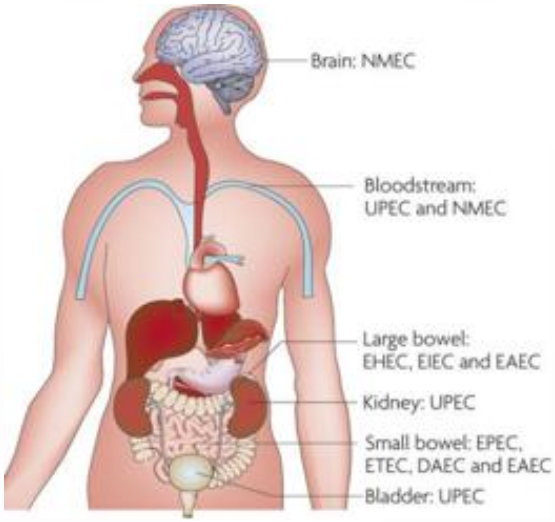


Figure 4. Pathogenic E. coli variants colonise various sites in human body: Enteropathogenic E. coli (EPEC), enterotoxigenic E. coli (ETEC) and diffusely adherent E. coli (DAEC) colonise the small bowel and cause diarrhoea, whereas enterohaemorrhagic E. coli (EHEC) and enteroinvasive E. coli (EIEC) cause disease in the large bowel; enteroaggegrative E. coli (EAEC) can colonise both the small and large bowels. Uropathogenic E. coli (UPEC) enters the urinary tract and travels to the bladder to cause cystitis and, if left untreated, can ascend further into the kidneys to cause pyelonephritis. Septicaemia can occur with both UPEC and neonatal meningitis E. coli (NMEC), and NMEC can cross the blood–brain barrier into CNS, causing meningitis. (Matthew A. Croxenm B. Brett Finlay, 2009).

2.3. Production of toxins

Toxins listed in Table 3 are all produced by uropathogenic E. coli (UPEC) which are the most common causes of urinary tract and blood stream infections (Welch, 2016). However, as mentioned previously, STEC is an example of intestinal pathogenic E. coli capable of producing the Shiga toxin.

Table 3. Toxins secreted by Uropathogenic E. coli (Welch, 2016)

Toxin	Exoprotein Family	Action
Hemolysin	Type I	Membrane pore-formation
Cytotoxic Necotising Factor-1 (CNF-1)	Type I	Deamidation of GTPase, RhoA, Cdc42, and Rac
Secreted Autotransporter Toxin (Sat)	Type V	Serine protease that targets cytoskeletal proteins
Protein involved in intestinal colonisation (PicU)	Type V	Serine protease that tarfets O-linked glycoproteins
Vacuolating autotransporter toxin (Vat)	Type V	Serine protease that causes vacuole formation in cultured bladder cells

3. Clinical Relevance of E. coli

3.1. Diseases & Symptoms

The most common diseases caused by pathovars of E. coli across most populations globally are summarised in Table 4.

Table 4. The most common diseases caused by E. coli and their pathovars, target population, and symptoms.

Disease	Gastroenteritis	Urinary Tract Infection (UTI)	Pneumonia	Meningitis
Pathovar	EPEC	UPEC	UPEC	NMEC
Description	Infection of large intestine caused by certain strains of E. coli	Infection caused by E. coli normally involve lower urinary tract – bladder and urethra	Lung infections involving lower lobes/ abscessation/ empyema	Idiopathic community-acquired E. coli meningitis
Target population	Individuals with bad hygiene	Women has higher risk than man	Chronically ill patients	Newborn babies
Symptoms	• Diarrhea • Vomiting	• Abdominal pain • Pelvic pain • Blood in urine	• High fever • Chest pain • Hemoptysis • Local cellulitis	• Vomiting • Headache • Photo-phobia • Seizure
Reference	(James P. Nataro, Eileen M. Barry, 2013)	(Mazen S Bader et. al, 2020)	(Ilya Berim, Sanjay Sethi, 2012)	(Matthew Mueller, Christopher R. Tainter, 2023)

3.2. Diagnosis

For diagnosis of E. coli intestinal infections (e.g. gastroenteritis), traditional diagnosis requires a stool sample from the patient for laboratory culture, which is usually done on MacConkey agar along with the indole producing test (Marc Roger Couturier et. al, 2011). Diagnosis of UPEC requires a urine sample for culture rather than stool sample. Molecular diagnosis can sometimes be used for detecting the pEAF plasmid or BFP factor in STEC.

3.3. Treatments

Treatment of E. coli infections relies on a range of antibiotics which differ depending on the type of infection. For example, in the case of intestinal infections, fluoroquinolones, azithromycin, and rifaximin are recommended, along with antimotility agents in the case of watery diarrhoea (Table 5) (Matthew Mueller, Christopher R. Tainter, 2023).

Table 4. The most common diseases caused by E. coli and their pathovars, target population, and symptoms.

Toxin	Exoprotein Family	Action
Rehydration		Oral or IV fluids
Antimotility agents	Bismuth salicylate Loperamide	Not recommended
Antibiotics	Fluoroquinolones • Ciprofloxacin • Levofloxacin Azithromycin Rifaximin	

3.4. Antibiotic Resistance

Antibiotics are used extensively for treating E. coli infections. Thus, antibiotic resistance in E. coli is a serious problem, especially in multidrug resistant E. coli. The main reason for the rapid rise of antibiotic resistance in E. coli is the great capacity of E. coli to receive and accumulate resistance genes through horizontal gene transfer (Figure 5) (Lauren Poirel et. al, 2018).

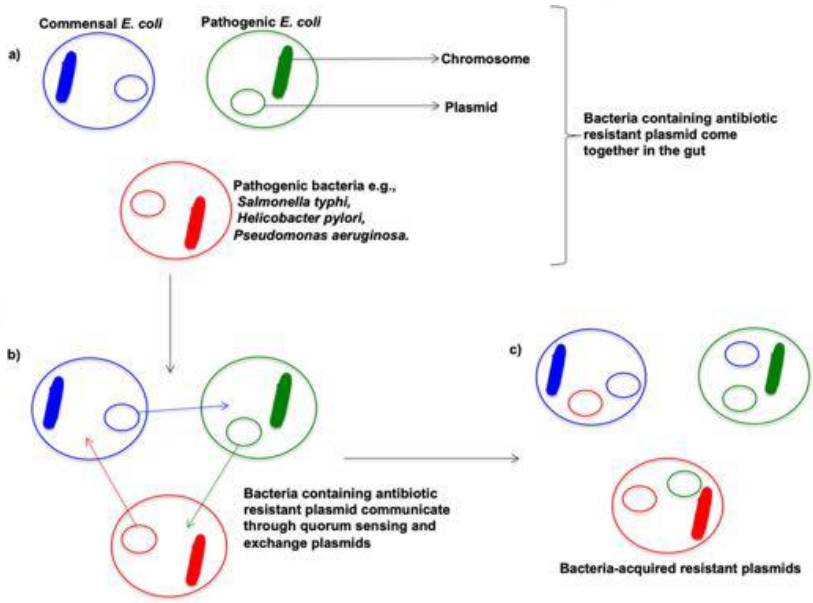


Figure 5. Horizontal gene transfer between gut bacteria, including commensal and pathogenic E. coli as well as a range of other pathogenic bacteria such as Salmonella typhi, Helicobacter pylori and Pseudomonas aeruginosa. (a) Bacteria containing antibiotic resistant plasmids coming together in the gut, (b) Exchange of antibiotic resistant plasmids, and (c) Acquisition of resistant plasmids and spread of antibiotic resistance.(Emmanuel Nji et. al, 2021)

E. coli acts as both a donor and a recipient of resistance genes, thereby acquiring new resistance genes and passing on existing resistance genes to other bacteria simultaneously. E. coli has shown significant resistance to -lactam antibiotics as genes that code resistance for these antibiotics (ESBLs/AmpCs) have emerged widely in E. coli in both human and animal infections. (Lauren Poirel et. al, 2018). In addition to -lactam antibiotics, antibiotic resistance in E. coli against other antibiotics is on the rise, and Figure 6 provides a snapshot of the range of antibiotics and the percentage of resistance across E. coli isolates. Thus, antibiotic resistance in E. coli is increasing at an alarming rate against most widely used antibiotics (Luong Thi Yen Nguyet et. al, 2022). The most common antibiotic resistance mechanisms in E. coli are summarised in Table 6.

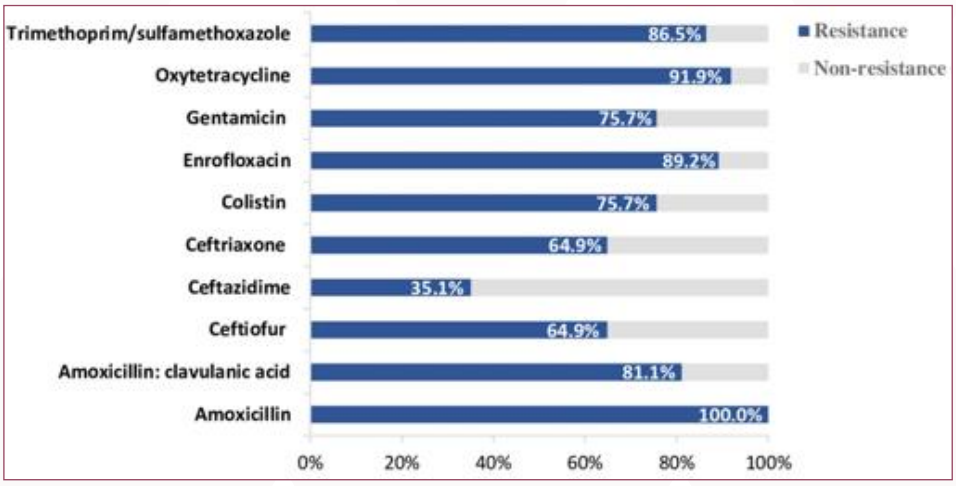


Figure 6. The antimicrobial-resistance percentages of E. coli strains toward a range of antibiotics: trimethoprim/sulfamethoxazole, oxytetracycline, gentamicin, enrofloxacin, colistin, ceftriaxone, ceftazidime, ceftiofur, amoxicillin: clavulanic acid, and amoxicillin. (Luong Thi Yen Nguyet et. al, 2022)

Table 6. Commos antibiotic resistance mechanisms in E. coli and their respective effects on certain antibiotics (Lauren Poirel et. al, 2018)

Mechanisms of resistance	Effect on Antibiotics
Extended-spectrum β-lactamases	Resistance to broad-spectrum cephalosporins
Carbapenemases	Resistance to carbapenems
16S rRNA methylases	Pan-resistance to aminoglycosides
pmgr genes	Resistance to fluoroquinolones
mcr genes	Resistance to polymyxins

4. Conclusion

E. coli as a Gram-negative bacterium that has a wide variety of pathovars, associated with various infections as well as a rise in antibiotic resistance. Sanitation and personal hygiene remain the most effective method of preventing E. coli infections. However, there are few vaccines in development with potential future use such as the one consisting of formalin-inactivated ETEC along with recombinant cholera toxin B subunit (rCTB) which has been found to be safe and immunogenic in both adult and children volunteers (James P. Nataro, Eileen M. Barry, 2013).

5. References

1. Claudia Silva et. al. (2017, October 26). Self-Conjugation of the Enteropathogenic Escherichia coli Adherence Factor Plasmid of Four Typical EPEC Isolates . BioMed Research International.

2. Emmanuel Nji et. al. (2021, February 09). High prevalence of antibiotic resistance in commensal Escherichia coli from healthy human sources in community settings. Scientific Reports, 11, 3372.

3. Ilya Berim, Sanjay Sethi. (2012). Chapter 24 - Community-Acquired Pneumonia. In Clinical Respiratory Medicine (Fourth Edition) (pp. 296-308).

4. James B Kaper et. al. (2004, March). Pathogenic Escherichia coli. Nature Reviews Microbiology, 2(2), 123-40.

5. James P. Nataro, Eileen M. Barry. (2013). 45-Diarrhea caused by bacteria. In Vaccines (Sixth Edition) (pp. 1052-1059).

6. Jason W. Sahl et. al. (2013). Chapter 2 - Comparative genomics of pathogenic Escherichia coli. In Escherichia coli (Second Edition) Pathotypes and Principles of Pathogenesis.

7. Jianli Wang, Wenjian Ma, Xiaoyuan Wang. (2021, March 20). Insights into the structure of Escherichia coli outer membrane as the target for engineering microbial cell factories. Microbial Cell Factories, 20(73).

8. Kyung-Soo Lee et. al. (2021, June 11). Escherichia coli Shiga Toxins and Gut Microbiota Interactions. Toxins, 13(6), 416.

9. Lauren Poirel et. al. (2018, July 12). Antimicrobial Resistance in Escherichia coli. American Society for Microbiology, 6(4).

10. Lugtenberg, B. (1981). Composition and function of the outer membrane of Escherichia Coli. Trends in Biochemical Sciences, 6, 262-266.

11. Luong Thi Yen Nguyet et. al. (2022, May 31). Antibiotic resistant Escherichia coli from diarrheic piglets from pig farms in Thailand that harbor colistin-resistant mcr genes. Scientific reports, 12, 9083.

12. Marc Roger Couturier et. al. (2011, February). Shiga-Toxigenic Escherichia coli Detection in Stool Samples Screened for Viral Gastroenteritis in Alberta, Canada. Journal of Clinical Microbiology, 49(2), 574-578.

13. Malvina Papanastasiou et. al. (2013, March). The Escherichia coli Peripheral Inner Membrane Proteome. Molecular & Cellular Proteomics, 12(3), 599-610.

14. Matthew A. Croxenm B. Brett Finlay. (2009, December 07). Molecular mechanisms of Escherichia Coli pathogenicity. Nature Review Microbiology, 8, 26-38.

15. Matthew Mueller, Christopher R. Tainter. (n.d.). Escherichia coli Infection. In StatPearls .

16. Mazen S Bader et. al. (2020, April). Treatment of urinary tract infections in the era of antimicrobial resistance and new antimicrobial agents. Postgrad Medicine, 132(3), 234-250.

17. Menge, C. (2020, September 21). The Role of Escherichia coli Shiga Toxins in STEC Colonization of cattle. Toxins, 12(9), 607.

18. Min Park et. al. (2015, March). Isolation and characterization of the outer membrane of Escherichia coli with autodisplayed Z-domains. Biochimica et Biophysica Acta (BBA) Biomembranes, 1848(3), 842-847.

19. Natacha Opalka et. al. (2010, September 14). Complete Structural Model of Escherichia coli RNA Polymerase from a Hybrid Approach. Plos Biology.

20. WHO. (2018, February 7). E. coli. Retrieved February 2024, from WHO.int: https://www.who.int/news-room/fact-sheets/detail/e-coli#:~:text=It%20is%20transmitted%20to%20humans,toxins%20produced%20by%20Shigella%20dysenteriae

21. Pranhdeep Kaur, Pradeep K Dudeja. (2023, April 06). Pathophysiology of Enteropathogenic Escherichia coli-induced Diarrhea. Newborn, 2(1), 120-113.

22. Steven L. Percival et. al. (2014). Chapter Six - Escherichia coli. In Microbiology of Waterborne Diseases (Second Edition) Microbiological Aspects and Risks (pp. 89-117).

23. Tankeshwar, A. (2023, December 27). Escherichia coli: Properties and Identification. Retrieved February 2024, from Microbe online: https://microbeonline.com/e-coli-disease-properties-pathogenesis-and-laboratory-diagnosis/#Characteristics_of_E_coli

24. Welch, R. A. (2016, June 24). Uropathogenic Escherichia coli-associated exotoxins. Microbiology Spectrum, 4(3).

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

02

Chemistry

Fitzwilliam College Online Winter School Journal · 2024 Winter

Comparison Between Artificial Molecular Machines and Biological Systems

CHUNSU WANG (Tomn)

Abstract

The field of supramolecular chemistry has been closely related to biological studies. In this minireview, we made connections between artificial molecular machines, a topic in supramolecular chemistry that has gained much interest, and nature's biological machines, from perspectives of structure, function, and mechanism. It is found that artificial molecular machines have successfully achieved several characteristics in structure and function of biological molecular machines, and have even improved upon them, such as the use of mechanical bond. On the other hand, biological molecular machines have more complexed structure and use more advanced mechanisms, much of which still require exploration so that these can be employed in elaborating artificial molecular machines.

1. Introduction

Since the foundation of supramolecular chemistry, it has a close relationship with biology, study of living things. Such connection stems from reciprocal facilitation between the 2 fields —supramolecular chemistry derives inspiration from studies on existing systems found in living things, and biologist uses supramolecular concepts to approach biological problems (Uhlenheuer, Petkau & Brunsveld, 2010). On the one hand, biology incorporates ideas from both disciplines and generates innovative outcomes. Xiong et. al. (2022) use host-guest complexation to reduce the excessive activity of CRISPR, an important biological gene editing tool. Cao et. al. (2021) discussed the combination of site-specific protein modification with supramolecular targeting of protein. On the other hand, many supramolecular systems are bioinspired. Some supramolecular systems are designed to mimic enzymatic function: cyclodextrins (Breslow & Dong, 1998), peptide-based nanozymes (Han et. al., 2021), and systems containing essential coordinated metal ion (Feiters, Rowan & Nolte, 2000).

Other systems are made similar to proteins that exhibit intended motion effects and are called molecular machines. By strict definition, molecular machine would use external stimulus and energy to conduct directional motion and perform net work on surroundings (Steed & Atwood, 2022). The field of molecular machine has achieved great progression in recent years, and biology is playing an increasingly vital role. Zhang et. al. (2018) concluded that biological inspiration would have a more profound effect on molecular machine construction compared to inspirations from classic engineering machinery. This is because the microscopic world is mainly governed by principles different from that in the macroscopic world, such as Brownian motion, thermal fluctuation, and viscous drag (Astumian, 2007). Biological machines have long been adapted to the microscopic world, and therefore they are more related to the molecular machines being designed. Hence, this article compares and contrasts artificial molecular machines with natural biological systems in three perspectives: structure, function, and mechanism, in order to gain understanding on how much of the nature's elegant system has been understood and appreciate characteristics of biological systems that are prospective for application in supramolecular chemistry.

2. Comparison Between Artificial and Biological Molecular Machines

2.1. Structure of Artificial And Biological Molecular Machines

In terms of structural similarity and difference between artificial molecular machine and natural protein machines, a prominent aspect to discuss is the way how components of machine are assembled. Both artificial and natural molecular machines face the need to tightly assemble their components, as the loss of one component could results in disintegration of the whole system. In nature, this is done by various interactions vital for quaternary structure of protein. Except for disulfide bridge, a covalent bond, many interactions are based on non-covalent interactions between R groups of amino

acids, such as hydrogen bond, ionic bond, etc. For instance, FoF1 ATP synthase has a rotor region connected via stalk gamma subunit to stator region, and the assembly of subunit is done using these non-covalent interactions, as shown in Fig.1 (Ruhle & Leister, 2015). In artificial machines, such as the [2] catenane rotary motor, traditional non-covalent interaction is important, as the design carefully manipulated hydrogen bonding between N-H on small macrocycle and C=O on the "track" to make macrocycle be assembled firstly to fumaramide region of track (Fig.2) (Hernandez, Kay & Leigh, 2004). However, artificial systems are assembled with another unique topological technique that is rarely used in natural machines—mechanical bond (Fig.3.). Mechanical bond means that two or more independent molecules, especially rings, are locked together and will not dissociate unless at least 1 covalent bond be broken between them. This makes artificial supramolecular systems more bounded and less susceptible to disintegration. For instance, an aggregation of protein subunits in natural machine will be dissociated upon change in pH or heating, as these machine systems have been evolved to make such elaborated use of weak non-covalent interactions to adapt the organisms homeostatic internal environment, such as about 37 ℃ , pH7 in human body. On contrary, once the mechanical bond is formed in artificial machines, it will require more violent heating and chemical conditions to break covalent bonds that are several orders of magnitude stronger than non-covalent interaction. However, covalent bond is only required to be broken if the components are to dissociate, not during movement of component, so the components are still labile relative to each other.

This use of mechanical bond is an improvement compared to nature's structure, and has implication in making molecular machine applicable in wider range of chemical conditions.

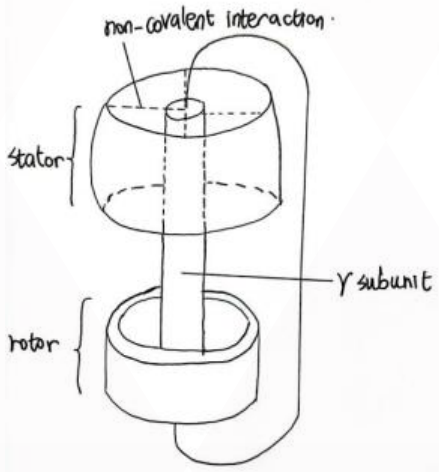


Figure 1. Non-covalent interaction in assembly of FoF1 ATP synthase, adapted from Ruhle & Leister (2015).

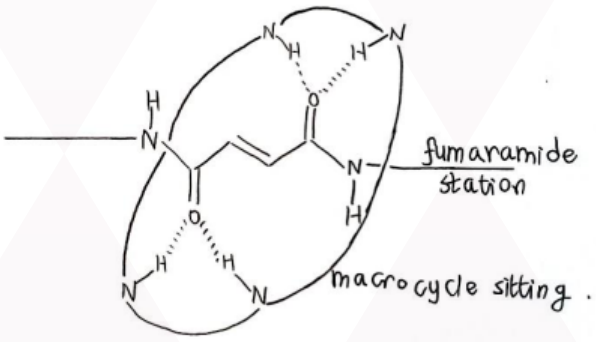


Figure 2. Non-covalent interaction in assembly of artificial supramolecular machine, adapted from Hernandez, Kay & Leigh (2004).

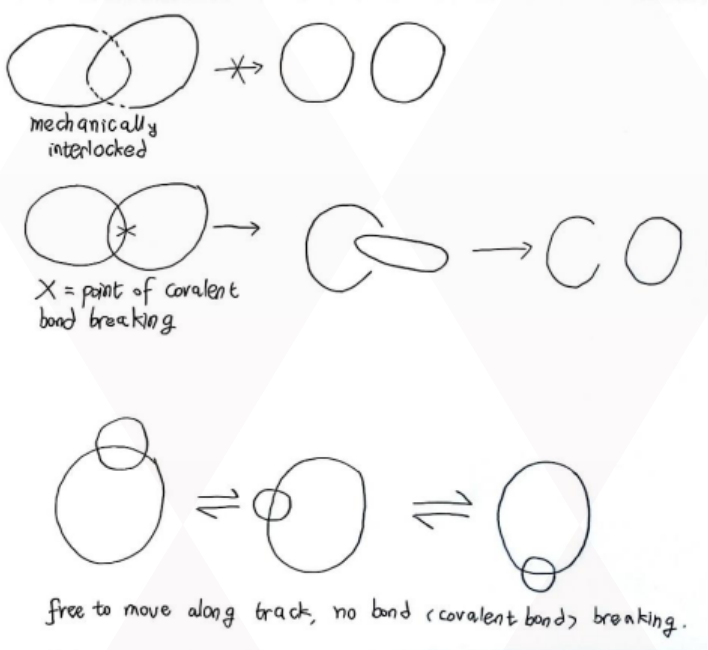


Figure 3. Mechanical bond, process of breaking a covalent bond to dissociate the mechanically interlocked catenane, and flexibility of components mechanically bonded together. Dashed line represent spatially arranged behind from 2-dimensional view.

In addition to mechanical bond, another feature that artificial molecular machine differed from natural machine is the size and complexity of structure. Many artificial molecular machines achieve the basic intended function while retaining only core components crucial for the machine to operate. For instance, Koumura et. al. (1999) designed the molecular rotor that has unidirectional rotation. The function is achieved by using two driving force that alternately operate the system: cis-trans isomerization under light, and tendency to reduce steric hindrance repulsion, and sometimes hydrogen bond formation as well (Kassem et. al., 2016). From this discovery, many more rotors were made based on the simplified 2 "force" pattern, and the core structure maintained is the double bond to be isomerized and, in many cases, the bulky groups generating steric hindrance. On contrary, natural rotors present in bacteria rotating flagella to provide motility of the organism used numerous protein subunits resulting in complexed rotor and stator regions, with diameter up to 45 nm (Sowa et. al., 2005). Another example is the molecular walker made by Delius, Geertsema, and Leigh (2009), which is only a 21 atom two-legged molecular unit. Under comparison, kinesin has two heavy chains, each with relative molecular mass of about 120,000 and two light chains with individual relative mass of about 60,000 to 70,000 (Friedman & Vale, 1999)—a much more complexed and larger machine. These on one side means that artificial molecular machine can be operating at an even more microscopic scale compared to biological scale, while on the other side indicates more studies on comprehending nature's complexed machine system and possibly improving the currently simple and fundamental artificial system toward intricately designed natural system may be needed.

2.2. Function of Artificial And Biological Molecular Machines

Artificial molecular machine are currently mainly used to conduct two type of motion—translational movement and rotation. Translational movement typically involve a labile component having a net change in position relative to a track, and is conducted by rotaxane based molecular pump (Cheng et. al., 2015), rotaxane molecular shuttle (Anelli, Spencer & Stoddart, 1991) (Bissell et. al., 1994), pseudo-rotaxane or rotaxanebased ratchets (Chatterjee, Kay & Leigh, 2006), molecular walker (Delius, Geertsema & Leigh, 2009), and peptide synthesizer (De Bo et. al., 2017). These shares similarity with many proteins having translational movement. For instance, multiple dynein can walk along microtubule to cause cilia beating (Roberts et. al., 2013). Other examples include ribosomes, which move translationally relative to mRNA to assemble polypeptides from aminoacyl tRNA. DNA and RNA polymerase also move along template DNA strands to manufacture new strands of DNA or RNA. Rotational movement are achieved by artificial molecular rotors, mostly the overcrowded alkene rotors (Krajnik et. al., 2017) but also contains catenane rotors (Hernandez, Kay & Leigh, 2004; Wilson et. al., 2016), in which circular work around an axis can be done. In nature, this resembles mainly the flagella rotor which propels prokaryotes.

Regardless of translational or rotational movement, the essential characteristic of functioning is directionality, and better if continuity is achieved. Such feature is indispensable in biological systems as targeted direction and orderness are constantly required. Motor proteins must not transport vesicles in random direction along microtubule, as content of vesicle need to be sent to correct destination for it to function. Ribosome move only from 5' to 3' of mRNA in a single direction manner because codonmust be exposed in correct sequence for amino acid linking so primary structure of protein is specified. Directionality of functioning was not achieved initially by some artificial molecular mechanical devices, but has been improved and now many artificial machines operate unidirectionally. Initially, molecular switches designed do exhibit relative movement of component upon application of external stimuli, but lack of specific single direction in which the system do work. For instance, Livoreil et. al. (1994) reported the rearrangement of [2] catenane using difference in coordination geometry of Cu(I) and Cu(II) upon redox reaction stimulus, but direction of rearrangement of one macrocycle relative to the other, clockwise or anticlockwise, may not be specified (see Fig.4). Many rotaxane based switches with multiple stable stations have some direction in movement upon stimulation, but work is undone by resetting switches (Klein, Kuhn & Beer, 2019; Schroder & Schalley, 2019). Afterward, modification of rotaxane system generate rotaxane ratchets and pumps that operate in single direction with work done on surroundings, though continuity was not reached.

For instance, the macrocycle was, by synthesis, situated on one side of the track, and removing followed by restoring the middle blocking group allow the net distribution of macrocycle to move right therefore achieving simple irreversible

unidirectional movement (Zhang et. al., 2018). Rotor systems also has developed directionality. Koumura et. al. (2000) reported the unidirectional 360 degree counterclockwise rotation with the second generation of alkene rotors, with 2 thermal steps in the rotational cycle generating irreversibility by allowing methyl substituents to adopt favorable axial position (See Fig.5).

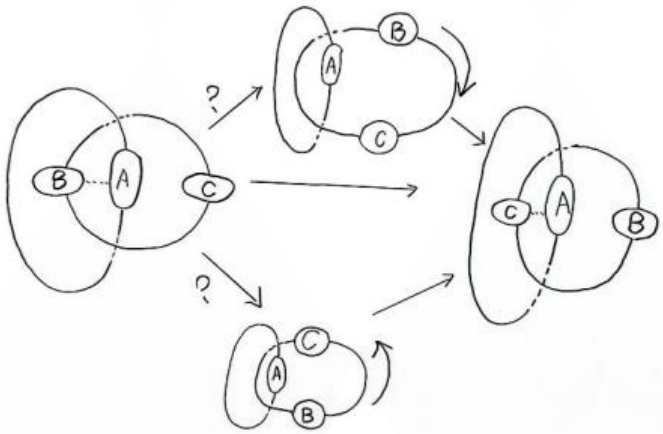


Figure 4. non-unidirectional position switching in a simple molecular switch, adapted from Livoreil et. al. (1994).

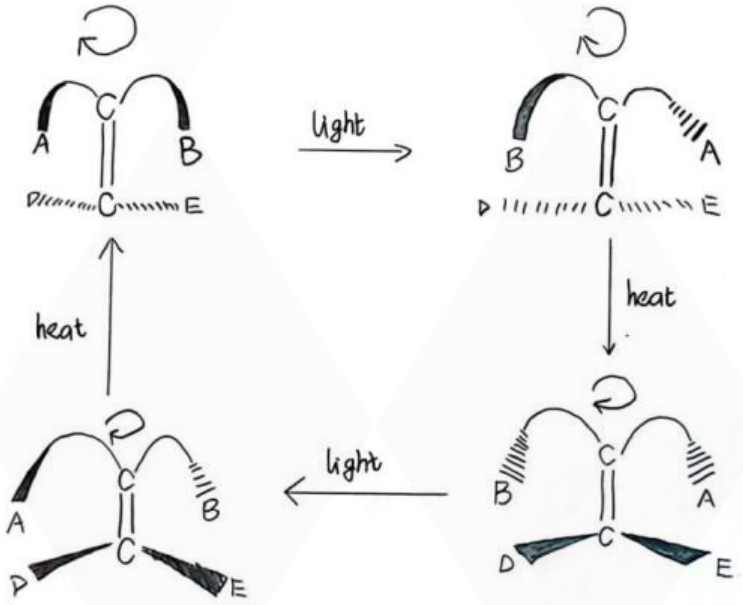


Figure 5. 360 degree rotational switch doing unidirectional rotation, adapted from Koumura et. al. (2000).

Continuity of function means the machine is capable of doing work continuously instead of stopping after few operations. Natural systems have good continuity. For instance, dynein and other motor proteins are capable of transporting vesicles through relatively long distances without detaching from microtubule track (Qiu et. al. 2012); DNA polymerase can continue replication grabbing onto template strand. Continuity in rotaxane based system aimed at translational work has generally less continuity, with some relatively continuous systems such as some molecular pumps, peptide synthesizer (Lewandowski et. al., 2013), and molecular walker (Delius, Geertsema & Leigh, 2009).Pezzato et. al. (2017) reported the threading of another macrocycle after one has been threaded on the rotaxane pump. Binks et. al. (2022) reported that single opening transamidation pump allow the formation of [4] rotaxane, meaning the pumping process occurred up to 3 macrocycle being

threaded. Rotor machines, however, has better continuity since the first generation. The first generation alkene motor by Koumura et. al. (1999) can already perform full 360 degree rotation via the four isomerization steps. Better continuities are displayed by catenane rotor system (See Fig.6) (Wilson et. al., 2016; Lu et. al., 2013).

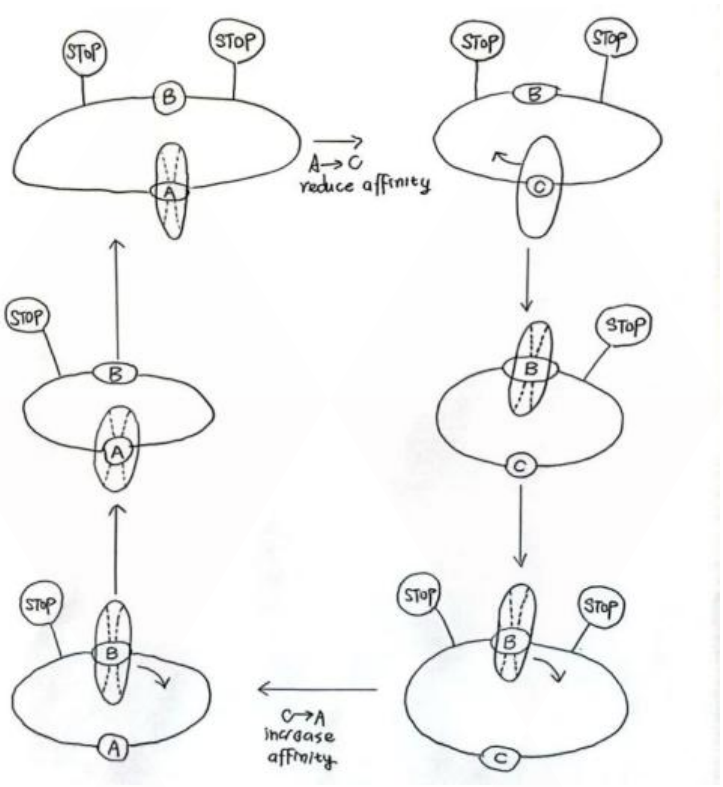


Figure 6. Catenane rotor system with good continuity, adapted from Wilson et. al. (2016).

2.3. Mechanism of Artificial And Biological Molecular Machines

Artificial supramolecular systems, starting with mechanical devices, have been progressing largely toward making use of sophisticated yet intricately designed mechanism used by nature. Initially, mechanical devices had taken much more focus on changing the affinity of specific sites using changes in non-covalent interaction to simply achieve motions. Elizarov et. al. (2002) used the [2] rotaxane that has a dialkylamine and a 1,2-bis-pyridiniumethane site to change the position at which dibenzo[24]crown-8 ring resides by deprotonation and reprotonation. Zubillaga et. al. (2018) use both pH and light condition to achieve the shuttling of cucurbit[7]uril along the track. These are based on the differences in stability of binding to sites on the track at different conditions, and the ring prefer the position favored by equilibrium, so equilibrium is the dominating factor to consider. Later, a change in understanding of the mechanism lead to the development of directionality and systems starts to take advantage of the random Brownian motion, leading to the development of ratchet. Ratchet can harness the seemingly useless random, fluctuated thermal energy, rectifying them thereby doing useful work (Steed & Atwood, 2022). However, such process is not a violation of thermodynamic laws, as there have always been a source of energy input, be it light, electrical, chemical, etc., and it is these energy source which moves toward equilibrium that provides a chance for a weaker interlinked system to go against equilibrium (Astumian, 2007; Aprahamian, 2020; Sangchai et. al., 2023). The more incipient version of ratchet is the energy ratchet, which utilize a driving reaction, the energy source, to vary local minima and local maxima and periodically lower and higher the energy barrier, so as to achieve intended directionality. An example would be molecular pump machine. Cai et. al. (2020) reported that by reduction of bipyridinium, the energy barrier between the opening end and the collecting chain on the track is lowered, allowing the macrocycle to be threaded onto the track, and subsequent oxidation allows kinetically trapping the ring onto the collecting chain by raising the energy barrier. Pezzato et. al. (2018) uses an electrochemical approach, but the principle displayed by energy landscape remain

the same—using electrochemical energy input to first lower the barrier, then recovering it. An analogy could be trapping a fish into a net. The fish itself rarely enters the empty net spontaneously, but if a piece of bread is thrown into the net, the fish would quickly swim into the trap. Subsequently elevating the net above the water trap the fish within the net, just as how macrocycle stayed within the collecting track (See Fig.7). Though energy ratchet seems to be already a satisfactory achievement of realizing endergonic reactions, information ratchet that has been widely applied by biological machines are perhaps more advanced, since energy ratchet rely upon alternating environmental conditions but information ratchet, as in many homeostatic biological environments, occur in steady state conditions (Sangchai et. al., 2023). Biological machines regularly apply information ratchet mechanism (Astumian, 2017; Zhang et. al., 2018). For instance, the energy available for a biological machine depends on the information of distribution of concentration of ATP, instead of energy stored per ATP molecule, and therefore operate autonomously according to concentration of ATP (Sangchai et. al. 2023). Similarly, catenane nanomotor designed by Wilson et. al. (2016) use the preference of adding Fmoc group at the unhindered site without macrocycle, and this information allows the selective blocking of macrocycle's movement to achieve being unidirectional (See Fig.8). Similarly, Serreli et. al. (2007) developed a light responsive rotaxane machine based on the information of the difference in causing a cis-trans isomerization when macrocycle is located at 2 different position—sufficient light energy is only transferred to double bond in the presence of a nearby macrocycle instead of a distant macrocycle. The macrocycle is unable to reverse to nearby the double bond once the cis isomer formed, so the net transport is to move macrocycle toward one end of rotaxane. Recently, ratchet using more various and combined elements of information are being developed which is a closer step toward natural information ratchet. Liu et. al. (2023) use not only the position of the cyclodextrin macrocycle on the axle of [2] rotaxane, but also the chiral nature in the shape of cyclodextrin, making orientation of macrocycle as well as possible selective face functionalization important in creating the kinetic bias. Information ratchet are in a sense selectively controlling the barrier (Steed, Atwood, 2022). Moreover, selectivity depends on information given and therefore forcing the energy input reaction and the reaction being driven in a certain sequence (Sangchai et. al., 2023), which give its unique property of having unidirectional movement when all reagent are added together without subsequent altering (Wilson et. al., 2016). Information ratchet are therefore a promising mechanism at which the designs of artificial supramolecular machines are giving more and more emphasis on, with the help of accruing appreciation on nature's use of information ratchet mechanisms.

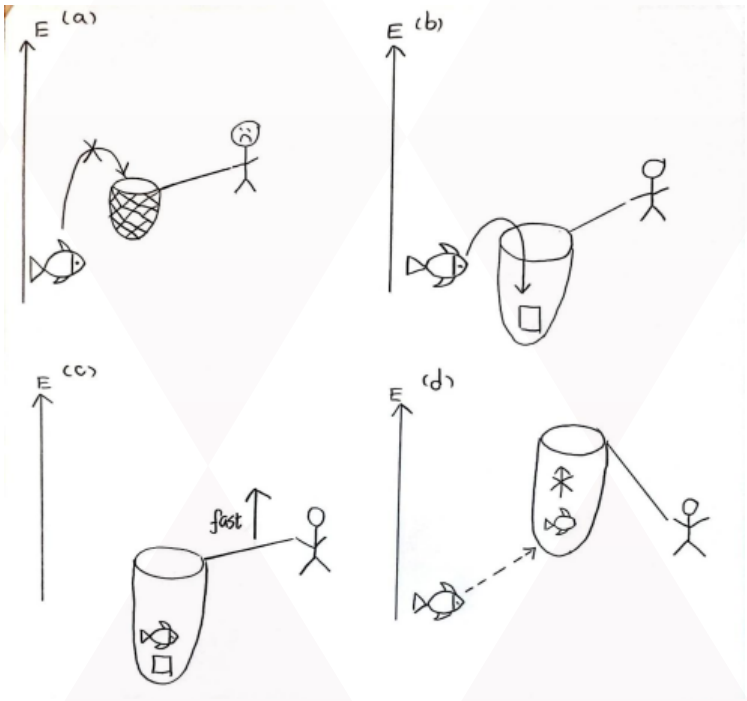


Figure 7. Schematic drawing of the fish-catching analogy to energy ratchet mechanism. Processes are followed in the order from (a) to (d).

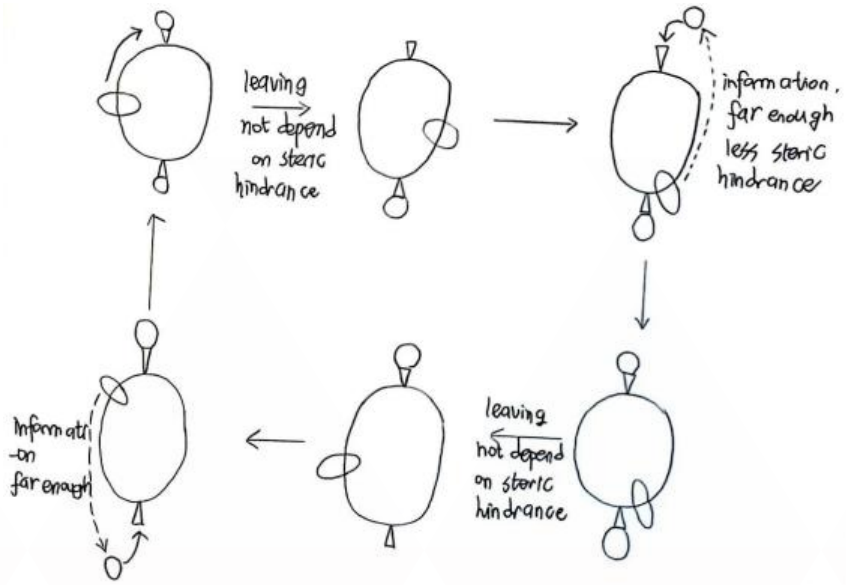


Figure 8. Representation of an informational ratchet, adapted from Wilson et. al. (2016)

3. Conclusion

Artificial supramolecular machines, as a representative field in supramolecular chemistry, are creating stronger and stronger connections with biological studies. Structurally, artificial molecular machines use non-covalent interactions to assemble components, similar to how protein machines form quaternary structures. Artificial molecular machine also use mechanical bond to lock components, which makes it more robust compare to natural way of assembly. Artificial machines are miniaturized, simplified design retaining essential functional region whereas natural machines are larger and much more complexed. In terms of function, artificial machine can achieve 2 basic motion in nature, translational and rotational movement. Progression in artificial machine allowed net work be done via unidirectional and continuous movement. Finally, through deepening in understanding of mechanisms on the operation of molecular machines, directional ratchet machines are made from simple switches, especially energy ratchet. Recently, the more advanced, information ratchet, common in biological machines, has been successfully replicated in artificial machines. Field of supramolecular machine attempted to comprehend biological systems from the microscopic basis, whereas biological studies are focusing more and more microscopic details from the macroscopic picture. The interconnection between these fields would give rise to more profound comprehension of nature, as well as application derived from nature.

4. Reference

1. Anelli, P.L., Spencer, N. and Stoddart, J.F. (1991). A molecular shuttle. Journal of the American Chemical Society, 113(13), pp.5131–5133. doi:https://doi.org/10.1021/ja00013a096.

2. Aprahamian, I. (2020). The Future of Molecular Machines. ACS Central Science, 6(3), pp.347–358. doi:https://doi.org/10.1021/acscentsci.0c00064.

3. Astumian, R.D. (2007). Design principles for Brownian molecular machines: how to swim in molasses and walk in a hurricane. Physical chemistry chemical physics: PCCP, [online] 9(37), pp.5067–5083. doi:https://doi.org/10.1039/b708995c.

4. Aude Livoreil, Dietrich-Buchecker, C. and Sauvage, J. (1994). Electrochemically Triggered Swinging of a [2]-Catenate. Journal of the American Chemical Society, 116(20), pp.9399–9400. doi:https://doi.org/10.1021/ja00099a095.

5. Binks, L., Tian, C., Fielden, S.D.P., Vitorica-Yrezabal, I.J. and Leigh, D.A. (2022). Transamidation-Driven Molecular Pumps. Journal of the American Chemical Society, [online] 144(34), pp.15838–15844. doi:https://doi.org/10.1021/jacs.2c06807.

6. Bissell, R.A., Córdova, E., Kaifer, A.E. and Stoddart, J.F. (1994). A chemically and electrochemically switchable molecular shuttle. Nature, [online] 369(6476), pp.133–137. doi:https://doi.org/10.1038/369133a0.

7. Breslow, R. and Dong, S.D. (1998). Biomimetic Reactions Catalyzed by Cyclodextrins and Their Derivatives. Chemical Reviews, 98(5), pp.1997–2012. doi:https://doi.org/10.1021/cr970011j.

8. Cai, K., Shi, Y., Guo Wei Zhuang, Zhang, L., Qiu, Y., Shen, D., Chen, H., Jiao, Y., Wu, H., Cheng, C. and J. Fraser Stoddart (2020). Molecular-Pump-Enabled Synthesis of a Daisy Chain Polymer. Journal of the American Chemical Society, 142(23), pp.10308–10313. doi:https://doi.org/10.1021/jacs.0c04029.

9. Cao, W., Qin, X. and Liu, T. (2021). When Supramolecular Chemistry Meets Chemical Biology: New Strategies to Target Proteins through Host-Guest Interactions. ChemBioChem, 22(20), pp.2914–2917. doi:https://doi.org/10.1002/cbic.202100357.

10. Chatterjee, M.N., Kay, E.R. and Leigh, D.A. (2006). Beyond Switches: Ratcheting a Particle Energetically Uphill with a Compartmentalized Molecular Machine. Journal of the American Chemical Society, 128(12), pp.4058–4073. doi:https://doi.org/10.1021/ja057664z.

11. Cheng, C., McGonigal, P.R., Schneebeli, S.T., Li, H., Vermeulen, N.A., Ke, C. and Stoddart, J.F. (2015). An artificial molecular pump. Nature Nanotechnology, [online] 10(6), pp.547–553. doi:https://doi.org/10.1038/nnano.2015.96.

12. Cristian Pezzato, Nguyen, M.T., Dong Jun Kim, Ommid Anamimoghadam, Mosca, L. and J. Fraser Stoddart (2018). Controlling Dual Molecular Pumps Electrochemically. Angewandte Chemie, 57(30), pp.9325–9329. doi:https://doi.org/10.1002/anie.201803848.

13. De Bo, G., Gall, M.A.Y., Kitching, M.O., Kuschel, S., Leigh, D.A., Tetlow, D.J. and Ward, J.W. (2017). Sequence-Specific β-Peptide Synthesis by a Rotaxane-Based Molecular Machine. Journal of the American Chemical Society, 139(31), pp.10875–10879. doi:https://doi.org/10.1021/jacs.7b05850.

14. Elizarov, A.M., Chiu, S.-H. and Stoddart, J.F. (2002). An Acid − Base Switchable [2]Rotaxane. The Journal of Organic Chemistry, 67(26), pp.9175–9181. doi:https://doi.org/10.1021/jo020373d. Feiters, M.C., Rowan, A.E. and Nolte, R.J.M. (2000). From simple to supramolecular cytochrome P450 mimics. Chemical Society Reviews, 29(6), pp.375–384. doi:https://doi.org/10.1039/a804252g.

15. Friedman, D.S. and Vale, R.D. (1999). Single-molecule analysis of kinesin motility reveals regulation by the cargo-binding tail domain. Nature Cell Biology, 1(5), pp.293–297. doi:https://doi.org/10.1038/13008.

16. Han, J., Gong, H., Ren, X. and Yan, X. (2021). Supramolecular nanozymes based on peptide self-assembly for biomimetic catalysis. Nano Today, 41, p.101295. doi:https://doi.org/10.1016/j.nantod.2021.101295.

17. HernándezJ.V., Kay, E.R. and Leigh, D.A. (2004). A Reversible Synthetic Rotary Molecular Motor. Science, 306(5701), pp.1532–1537. doi:https://doi.org/10.1126/science.1103949.

18. Kassem, S., Lee, A.T.L., Leigh, D.A., Markevicius, A. and Solà, J. (2016). Pick-up, transport and release of a molecular cargo using a small-molecule robotic arm. Nature Chemistry, [online] 8(2), pp.138–143. doi:https://doi.org/10.1038/nchem.2410.

19. Klein, H.A., Kuhn, H. and Beer, P.D. (2019). Anion and pH dependent molecular motion by a halogen bonding [2]rotaxane. Chemical Communications, 55(67), pp.9975–9978. doi:https://doi.org/10.1039/c9cc04752b.

20. Krajnik, B., Chen, J., Watson, M., Cockroft, S.L., Feringa, B.L. and Johan Hofkens (2017). Defocused Imaging of UV-Driven Surface-Bound Molecular Motors. Journal of the American Chemical Society, 139(21), pp.7156–7159. doi:https://doi.org/10.1021/jacs.7b02758.

21. Lewandowski, B., Bo, G.D., Ward, J.W., Papmeyer, M., Kuschel, S., Aldegunde, M.J., Gramlich, P.M.E., Heckmann, D., Goldup, S.M., D'Souza, D.M., Fernandes, A.E. and Leigh, D.A. (2013). Sequence-Specific Peptide Synthesis by an Artificial SmallMolecule Machine. Science, [online] 339(6116), pp.189–193. doi:https://doi.org/10.1126/science.1229753.doi:https://

doi.org/10.1126/science.1229753.

22. Liu, E., Sawsen Cherraben, Laora Boulo, Troufflard, C., Bernold Hasenknopf, Vives, G. and Matthieu Sollogoub (2023). A molecular information ratchet using a coneshaped macrocycle. Chem, 9(5), pp.1147–1163. doi:https://doi.org/10.1016/j.chempr.2022.12.017.

23. Nagatoshi Koumura, Geertsema, E.M., and Auke Meetsma and Feringa, B.L. (2000). Light-Driven Molecular Rotor: Unidirectional Rotation Controlled by a Single Stereogenic Center. Journal of the American Chemical Society, 122(48), pp.12005–12006. doi:https://doi.org/10.1021/ja002755b.

24. Qiu, W., Derr, N.D., Goodman, B.S., Villa, E., Wu, D., Shih, W. and Reck-Peterson, S.L. (2012). Dynein achieves processive motion using both stochastic and coordinated stepping. Nature Structural & Molecular Biology, [online] 19(2), pp.193–200. doi:https://doi.org/10.1038/nsmb.2205.

25. Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K. and Burgess, S.A. (2013). Functions and mechanics of dynein motor proteins. Nature Reviews Molecular Cell Biology, [online] 14(11), pp.713–726. doi:https://doi.org/10.1038/nrm3667.

26. Rühle, T. and Leister, D. (2015). Assembly of F1F0-ATP synthases. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1847(9), pp.849–860. doi:https://doi.org/10.1016/j.bbabio.2015.02.005.

27. Schröder, H.V. and Schalley, C.A. (2019). Electrochemically switchable rotaxanes: recent strides in new directions. Chemical Science, 10(42), pp.9626–9639. doi:https://doi.org/10.1039/c9sc04118d.

28. Serreli, V., Lee, C.-F., Kay, E.R. and Leigh, D.A. (2007). A molecular information ratchet. Nature, 445(7127), pp.523–527. doi:https://doi.org/10.1038/nature05452.

29. Sowa, Y., Rowe, A.D., Leake, M.C., Yakushi, T., Homma, M., Ishijima, A. and Berry, R.M. (2005). Direct observation of steps in rotation of the bacterial flagellar motor. Nature, 437(7060), pp.916–919. doi:https://doi.org/10.1038/nature04003.

30. Steed, J.W. and Atwood, J.L. (2022). Supramolecular Chemistry. 3rd ed. John Wiley & Sons, pp.923–936.

31. Thitiporn Sangchai, Shaymaa Al Shehimy, Emanuele Penocchio and Giulio Ragazzon (2023). Artificial Molecular Ratchets: Tools Enabling Endergonic Processes. Angewandte Chemie International Edition, 62(47). doi:https://doi.org/10.1002/anie.202309501.

32. Uhlenheuer, D.A., Petkau, K. and Brunsveld, L. (2010). Combining supramolecular chemistry with biology. Chemical Society Reviews, 39(8), p.2817. doi:https://doi.org/10.1039/b820283b.

33. von Delius, M., Geertsema, E.M. and Leigh, D.A. (2009). A synthetic small molecule that can walk down a track. Nature Chemistry, 2(2), pp.96–101. doi:https://doi.org/10.1038/nchem.481.

34. Wilson, M.R., Solà, J., Carlone, A., Goldup, S.M., Lebrasseur, N. and Leigh, D.A. (2016). An autonomous chemically fuelled small-molecule motor. Nature, 534(7606), pp.235–240. doi:https://doi.org/10.1038/nature18013.

35. Xiong, W., Liu, X., Qi, Q., Ji, H., Liu, F., Zhong, C., Liu, S., Tian, T. and Zhou, X. (2022). Supramolecular CRISPR-OFF switches with host–guest chemistry. Nucleic Acids Research, 50(3), pp.1241–1255. doi:https://doi.org/10.1093/nar/gkac008.

36. Zhang, L., Marcos, V. and Leigh, D.A. (2018). Molecular machines with bio-inspired mechanisms. Proceedings of the National Academy of Sciences of the United States of America, [online] 115(38), pp.9397–9404. doi:https://doi.org/10.1073/pnas.1712788115.

37. Zubillaga, A., Ferreira, P., Parola, A.J., Gago, S. and N. Basílio (2018). pH-Gated photoresponsive shuttling in a water-soluble pseudorotaxane. Chemical Communications, 54(22), pp.2743–2746. doi:https://doi.org/10.1039/c8cc00688a.

Supramolecular Chemistry in Everyday Life

GYUBAEK CHUNG (Edward)

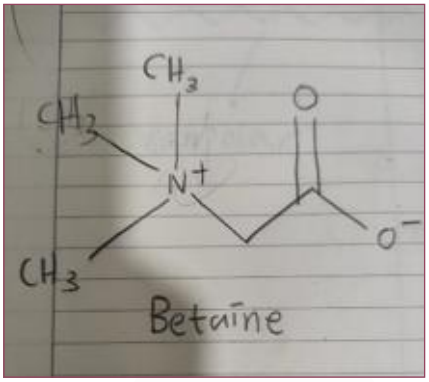
1. Preface: Supramolecular Chemistry

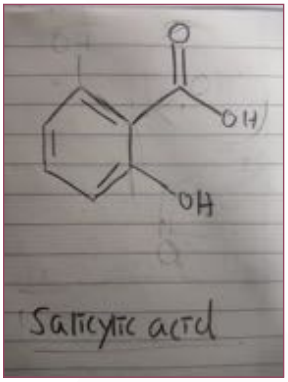
The emergence of supramolecular chemistry can be said to have greatly promoted the development of materials science and modern chemistry. It has also achieved various effects through the interaction between molecules, thereby benefiting people's lives. Supramolecular chemistry is closely related to the research of molecular self-assembly and emerging organic materials. Although its definition may still be a bit vague, it can be sure that this is a very dynamic and with a bright future. Today, the word “supramolecular” is no longer unfamiliar to people, and may be a word that can be seen frequently in science popularization videos or various commodities, I would like to give a few practical examples of supramolecular chemistry in this essay.

2. Cosmetics products and supramolecular chemistry

2.1. Structure & Classification

In fact, when it comes to supramolecular chemistry, what will be mentioned widely now must be various medical supplies. There are a lot of cosmetics advertisements recently that contain this word. Then let's give an example first. Salicylic acid is a chemical component that can be frequently heard. It is an organic acid with the chemical formula C7H6O3. Salicylic acid is an organic acid that can remove excess sebum from the skin surface, inhibit

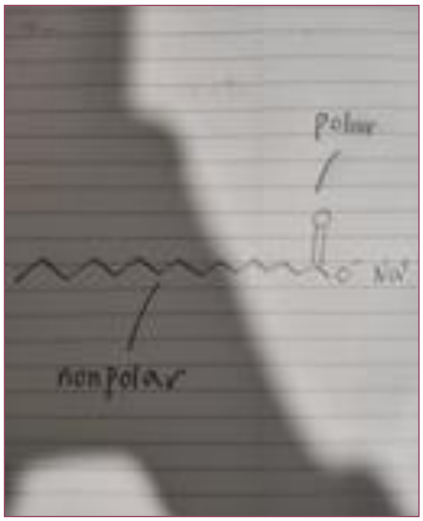




excessive sebum secretion, and play a role in controlling body oil. It is slightly soluble in normal water, easily soluble in hot water, ethanol, ether, and acetone, also, in hot benzene. This molecule itself creates strong stimulation, free molecules can cause certain damage to the skin, and it is not very soluble in water. Ethanol with a polar part and non-polar part can indeed dissolve it, but ethanol itself can also cause harm to the skin. However, those properties can somehow change when creating a supramolecular system with other molecules. This ingredient is composed of salicylic acid and betaine. It can be understood that O on the betaine molecules act as hydrogen bond acceptors, forming two hydrogen bonds with the phenol and carboxyl groups on the salicylic acid molecule as hydrogen bond donors. The remaining atoms in the molecule at the same time, arrange through π - π stacking, forming a stable structure. Salicylic acid, which was originally difficult to dissolve, can now be dissolved in water, while this structure maintains the original effects of salicylic acid and betaine. Also, there are Studies have shown that the cytotoxicity of this molecular structure is significantly reduced compared to their respective precursors, indicating that this approach can also improve the safety as the raw materials.

3. Soap and alcohol

After cosmetics, there is actually a more common supramolecular chemistry, which is soap. At the beginning of the epidemic of COVID-19, people have been emphasizing the importance of using soap, which is actually with clear evidence. Speaking of viruses, they are also related to supramolecular chemistry. Most viruses are composed of three components: RNA, proteins, and lipids. RNA is the genetic material of viruses, similar to DNA. Proteins have several functions, including invading target cells, assisting virus replication, and are essentially key components of virus structure. Lipids form a protective



layer around the virus, which not only plays a protective role but also helps the virus spread and invade cells. RNA, proteins, and lipids self-assemble to form viruses. Their binding is formed through non-covalent interactions between molecules. Although individual non covalent interactions may not be very strong, it is difficult to destroy when they are interlocked one by one to form a supramolecule. A virus is a very stable structure, and soap can make it less stable. Sodium stearate, commonly used in the manufacturing of toothpaste and soap, has a polar head and non-polar tail. It is structurally similar to lipids on the virus and can compete with those lipids on the membrane. Similarly, this also explains how it removes dirt from the skin. Soap molecules can also compete with many other non-covalent bonds that help proteins, RNA, and lipids stick together. With water, can effectively dissolve viruses on the skin.

Another common disinfectant is alcohol, and people usually use a solvent mixed with ethanol and other substances inside. As a solvent, they are more lipophilic than water. Therefore, alcohol does indeed dissolve lipid membranes and disrupt other supramolecular structures in the virus. But it will need a higher concentration of ethanol to quickly dissolve the virus. So, it can be seen that the efficiency of adding water to soap is actually higher.

4. Conclusion

supramolecular chemistry sounds mysterious, and one can also feel that it is not a simple subject to deal with. However, this does not mean that it is a very far-away thing. As mentioned earlier, many of the things we often come into contact with actually have relevant principles. Not only that, molecular self-assembly, molecular recognition, biomedicine, nanotechnology and other recently popular subjects also involve supramolecular chemistry. Believe that with the continuous development and growth of theory and practice in this field, it will promote the emergence and development of many new things. Finally, again, this is really amazing, it is a promising thing in future.

5. References

1. ^The coronavirus is no match for plain, old soap — here's the science behind it -Palli Thordarson.

2. ^ Mi Wang- Zhenyuan Wang,*, jichuan Zhang, Jingbo Zhan, Chengyu Wu, Wen Yu Wenhua FanJinshan Tang, Qiqing Zhang, and Jiaheng Zhang Sustainable Bioactive Salts Fully Composed of Natural Products forEnhanced Pharmaceutical Applicability: Acs Sustainable Chem, Eng 2022, 10, 10369-10382.

3. ^ Riccardo Montis and Michael B. Hursthouse Surprisingly complex supramolecular behaviour in the crystal structures of a family of mono-substituted salicylic acids : CrystEngComm, 2012, 14, 5242–5254.

4. ^ Adler-Abramovich L, Gazit E. The physical properties of supramolecular peptide assemblies: from buildingblock association to technological applications[)]. Chemical Society Reviews, 2014, 43(20): 6881-6893.

5. ^Clin Cosmet Investig Dermatol. 2010; 3: 135–142., Applications of hydroxy acids: classification, mechanisms, and photoactivity

6. ^ Merinville, E., Laloeuf, A., Moran, G., Jalby, O., & Rawlings, A. V. (2009). Exfoliation for sensitive skin with neutralized salicylic acid? International Journal of Cosmetic Science, 31(3), 243–244.doi:10.1111/j.1468-2494.2009.00501_2.x

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

03

Computer Science

Fitzwilliam College Online Winter School Journal · 2024 Winter

EMM386—DOS Memory Management

CHENGXUAN DENG (Sean)

Abstract

In the context of IBM PC-compatible computing, DOS memory management refers to the software and strategies used to enable applications to use more than 640 KiB of "conventional memory". The 640 KiB limit added a significant layer of complexity to the hardware and software designed to overcome it. The physical memory within a system could be organized as a combination of base or conventional memory (including lower memory), upper memory, high memory (not the same as upper memory), extended memory, and expanded memory, each managed in distinct ways. This article delves into one of the memory management techniques—EMM386.

1.DOS Memory Structure

1.1. Conventional Memory

In DOS memory management, conventional memory, also referred to as base memory, is the first 640 kilobytes (KiB) of memory on IBM PC or compatible systems. It is the read-write memory that the processor can directly address for use by the operating system and application programs. This particular section of memory in a DOS system is accessible to all standard DOS programs. In addition to DOS, conventional memory houses programs, routines, and system drivers. The 640 KiB limit was a design decision made due to the memory addressing capabilities of the Intel 8088 CPU used in the original IBM PC. As memory prices rapidly declined, this design decision became a limitation in the use of large memory capacities until the advent of operating systems and processors that rendered it irrelevant.

1.2. Upper Memory Area (UMA)

The Upper Memory Area (UMA) refers to the memory range between 640 kb and 1 Mb. Typically, there is no RAM in this range by default as it is reserved for hardware capable of mapping its own memory to this range. This region usually houses a portion of the graphics card's RAM and the BIOS ROMs of the graphics card and motherboard. Other hardware such as mass storage controllers, network adapters, and more can utilize sections of the UMA for their own BIOS ROMs or buffer RAM.

1.3. High Memory Area (HMA)

The High Memory Area (HMA) is the first 64 KiB of extended memory in an IBM PC or compatible machine. It is part of the system's RAM that exceeds the conventional memory limit of 640 KiB. DOS can use the HMA to load parts of its operating system code, freeing up more conventional memory for DOS applications.

1.4. Extended Memory

Extended memory refers to memory above the first megabyte of address space in an IBM PC or compatible. It is accessed using a different method compared to conventional memory, and DOS applications cannot directly access this memory. However, it can be used by DOS extenders, which allow DOS applications to run in protected mode and access more memory.

1.5. Expanded Memory

Expanded memory is a system that allows a computer to access more than 1 MiB of memory in a DOS real-mode application. It is a bank-switched system where data is moved into and out of the conventional memory area as needed.

This allows applications to handle larger amounts of data even though they are running in real mode.

2.Introduction of EMM386

2.1. However, the conventional memory brings a question - What happens if we need more RAM?

When the IBM PC/AT was introduced, the segmented memory architecture of the Intel family processors had the byproduct of allowing slightly more than 1 MiB of memory to be addressed in the "real" mode. Since the 80286 had more than 20 address lines, certain combinations of segment and offset could point into memory above the 0x0100000 (220) location. The 80286 could address up to 16 MiB of system memory, thus removing the behavior of memory addresses "wrapping around". Since the required address line now existed, the combination F800:8000 would no longer point to the physical address 0x0000000 but the correct address 0x00100000. This is the basis contributing into the building of EMM386.

2.2. Working Mechanism of EMM386

EMM386 worked its magic through a process known as paging. It swapped memory blocks in and out of the EMS page frame, which is located in the first megabyte of RAM and directly accessible by real-mode DOS applications. This was accomplished by reprogramming the page tables, thus controlling which physical memory pages mapped to a given linear address. No memory copying was involved in this process. The mechanism was very similar to how 8086 EMS boards worked, but it relied solely on the CPU's built-in memory management facilities instead of external hardware.

2.3. Key Functions of EMM386

EMM386 had several key functions in DOS memory management:

Creates Expanded Memory: EMM386 used the extended memory of Intel 80386 CPUs to create expanded memory. This allowed MS-DOS computers to have more than 640 KB of memory.

Maps Memory: EMM386 could map memory into unused blocks in the upper memory area (UMA). This allowed device drivers and terminate-and-stay-resident programs to be "loaded high", preserving conventional memory.

Enables EMS and UMBs: The page table entries used by the MMU were configured by EMM386 to map certain regions in upper memory to areas of extended memory. This technique enabled both EMS (Expanded Memory Specification) and UMBs (Upper Memory Blocks) - both of which appeared to DOS applications to be memory in the upper area but were in fact mapped to physical memory locations beyond 1MB.

2.4. Pros and Cons of EMM386

Like any technology, EMM386 had its advantages and disadvantages:

Advantages:

Creates Expanded Memory: EMM386 allowed an 8088 or 8086 CPU to address far more memory than its original design allowed. This was particularly useful for power users who put two or more megabytes of RAM in their PC and XT-class machines.

Preserves Conventional Memory: By mapping memory into unused blocks in the UMA, EMM386 allowed device drivers and terminate-and-stay-resident programs to be "loaded high", preserving conventional memory.

Enables EMS and UMBs: EMM386 enabled both EMS (Expanded Memory Specification) and UMBs (Upper Memory Blocks), which appeared to DOS applications to be memory in the upper area but were in fact mapped to physical memory locations beyond 1MB.

Disadvantages:

Limited Access to Expanded Memory: Expanded memory had to be accessed in comparatively tiny quantities. It was much

faster than swapping out to disk, but it had more overhead than using the standard 640K of RAM.

Lack of Protected Mode: While EMM386 did provide some level of memory management, it didn't offer the same level of protection as modern operating systems do. In modern systems, each process runs in its own protected space to prevent it from interfering with other processes. EMM386 lacked this feature, which could lead to system instability if a process behaves unexpectedly.

3.Conclusion

In conclusion, DOS memory management, including the use of EMM386, was a complex but necessary system for maximizing the use of memory in early IBM PC compatible computers. While these techniques are largely obsolete in modern computing, they represent an important part of the history of computer technology. Understanding these techniques provides valuable insight into the evolution of memory management and the challenges faced by early computer engineers.

True Random Number Generators

SI HUA ROBERT IP (Robert)

Random Number Generators (RNG) are algorithms or systems that generate random numbers and are widely used in many branches of computer science. In particular, the generation of truly random numbers is crucial in areas such as cryptography, as it is essential that no attacker can guess or calculate the secret key and deduce the original message based on the ciphertext. This report will investigate various methods of generating true random numbers and compare their advantages and weaknesses.

A True Random Number Generator (TRNG) is one that generates each subsequent number independently, with no correlation to any previous numbers generated. More specifically, there cannot exist an algorithm that can predict the next generated number any better than guessing by random chance. In addition, it is often required that the generated numbers are uniformly distributed, so that there is no output that is more likely than any other. Unfortunately, computers cannot generate truly random numbers as they are inherently deterministic, that is, any given input will always produce the same output. Instead, complex mathematical algorithms called Pseudorandom Number Generators (PRNG) are used to generate numbers that appear random enough for most applications. A PRNG takes in a random number, known as a seed, and outputs a stream of numbers that satisfies various statistical tests for randomness, but is ultimately determined by the initial seed. PRNGs are commonly used as they are fast, reliable, and easy to implement, but they possess a great security risk as any attacker that acquires the seed will be able to easily calculate the generated numbers. As such, the seed must be truly random, thus requiring external TRNGs. Due to this, TRNGs are essential to any field that depends on randomness to succeed.

1. Pseudorandom Number Generators

As computers cannot produce TRNGs, physical methods must be used to generate entropy, or randomness. While the exact entropy source may differ, the majority of TRNGs operate based on the following structure:

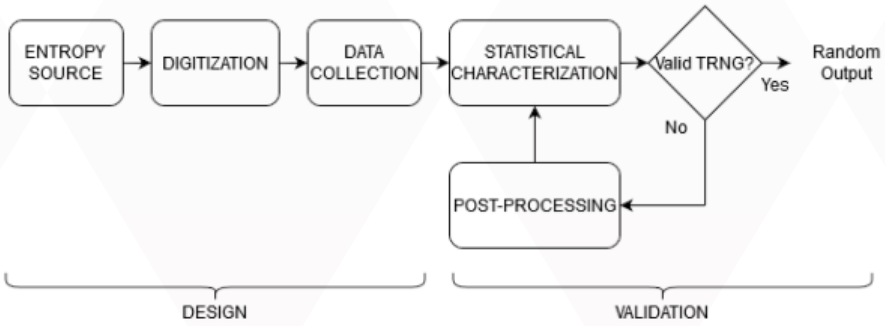


Fig 1. True Random Number Generator System Flowchart

(Rojas-Muñoz, L.F.; Sánchez-Solano, S.; Martínez-Rodríguez, M.C.; Brox, P. True Random Number Generation Capability of a Ring Oscillator PUF for Reconfigurable Devices. Electronics 2022, 11, 4028. https://doi.org/10.3390/electronics11234028)

The entropy source is measured and digitalized into binary numbers, before checked to see if it passes statistical tests of randomness. Any potential flaws, such as outside interference, are checked and removed with post-processing algorithms, before outputting the random sequence.

Common sources of entropy include:

1. Electrical noise, where noise generators create irregular fluctuations in currents that can be measured by comparators.

2. Free running oscillators, where jitters caused by environmental noise cause oscillators to pulse randomly.

3. Chaos, where unpredictable physical systems are measured and digitalized.

4. Quantum mechanics, which utilizes the random nature of quantum effects in phenomena such as nuclear decay.

5. Imprecision, where the least significant digits of a precise measurement such as temperature are taken such that the imprecision exhibit random behavior.

This report will be focusing on examples from the first three types of entropy.

2. Electric Noise

Electric noise generators are typically integrated into analogue circuits, utilizing the imperfections of circuit hardware to generate random fluctuations in voltage. Common sources of entropy include Johnson-Nyquist noise, where tiny thermal agitations in the charge carriers generates small, random currents, and Shot noise, where the random (quantum) motion of electrons cause fluctuations detectable at low currents. For each of them, an amplifier is used to amplify the noise to significant levels, which is then fed into a comparator. Every time a clock ticks, the comparator compares the voltage from the noise to a pre-set threshold, outputting a 1 if the voltage is above the threshold and 0 otherwise.

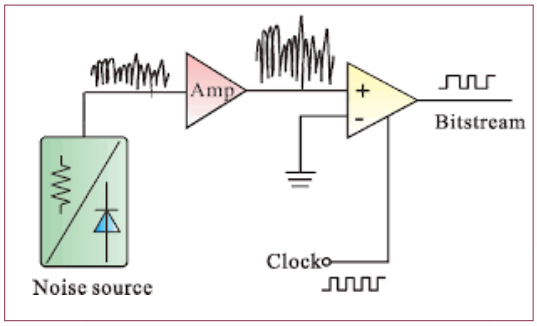


Fig 2. Electric-noise-based TRNG diagram

(Gong, Lishuang & Zhang, Jianguo & Liu, Haifang & Sang, Luxiao & Wang, Yuncai. (2019). True Random Number Generators Using Electrical Noise. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2939027.)

Noise-based generators are used for their simple structure and relatively compact size. However, they also exhibit several drawbacks, including the following:

1. Amplifiers are power-hungry and sensitive to external interference.

2. Precise calibration is required to set the threshold at a suitable level to ensure high entropy.

3. Noise levels are hard to control and change with environmental factors, affecting the entropy of the output.

Post-processing may be needed to correct biases and remove patterns that arise from external influences such as non-random noise.

3. Free Running Oscillators

Free running oscillators (FRO) are circuits that constantly emit an oscillating signal. These are typically constructed using Ring Oscillators, which are odd chains of inverters that constantly invert the signal between on and off.

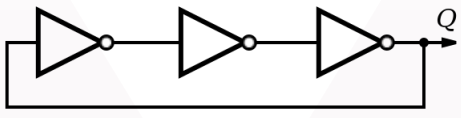


Fig 3. Circuit diagram of a Ring Oscillator

(Agarwal, Tarun. Ring Oscillator : Layout, Circuit Diagram and Its Applications. ElProCus - Electronic Projects for Engineering Students, 27 Nov. 2019, https://www.elprocus.com/ring-oscillator-working-and-its-applications/.)

While Ring Oscillators are by nature periodic, the constant inverting of the signal is affected by noise that causes tiny changes in the period known as phase jitter, which can be amplified by increasing the number of inverters or using components that are more susceptible to noise. Multiple Ring Oscillators can then be connected to form a TRNG.

One approach is to use two ring oscillators of different frequencies. The slower oscillator acts as a sampler, where it activates a d flip-flop every period that records the current state of the faster oscillator. The combined effect of both oscillator's jitter creates the source of entropy, which can be controlled by changing the frequency of the sampling oscillator, where a higher frequency has less entropy but a higher bit rate and vice versa.

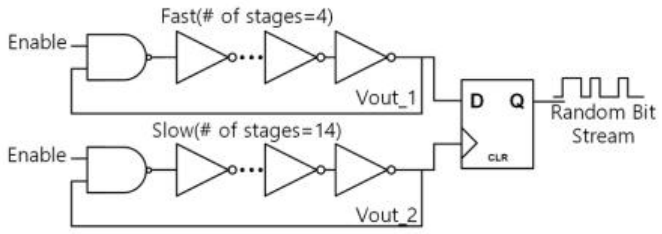


Fig 4. Circuit diagram of a TRNG with 2 oscillators of different frequencies
(Nam, Jae-Won & Kim, Jaewoo & Hong, Jong-Phil. (2022). Stochastic Cell- and Bit-Discard Technique to Improve Randomness of a TRNG. Electronics. 11. 1735. 10.3390/electronics11111735.)

Another design is to use multiple ring-oscillators of the same frequency and connect them to each other with an XOR gate. A clock is then used with a flip flop to sample the output regularly. Since any of the multiple oscillators may experience jitter and fall out of sync, the output increases in entropy over time. The entropy can be increased by simply adding more ring oscillators, increasing the chance of one of them jittering.

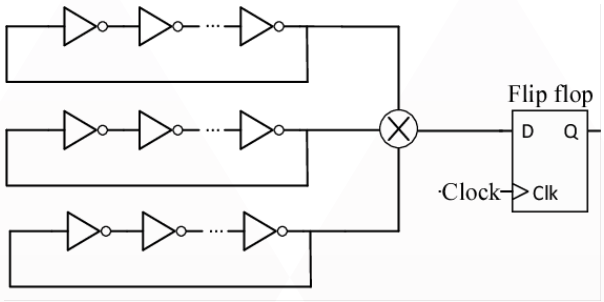


Fig 5. Circuit diagram of a TRNG with multiple ring oscillators of the same frequency
(Li, Chaoyang et al. A metastability-based true random number generator on FPGA. 2017 IEEE 12th International Conference on ASIC (ASICON) (2017): 738-741.)

FROs are useful as they have high bit rates and are very easy to implement due to only using standard electrical components. However, they are not perfect, as they are still susceptible to external interference, including thermal noise, low-frequency noise, and electromagnetic interference from nearby circuits causing them to synchronize with each other. As such, it is possible to attack them by inducing a large electromagnetic field that eliminates the jitter.

4. Chaos

Chaos is a general term used for any chaotic system that generates unpredictable behavior. Any physical system sufficiently complicated can theoretically be used, with current examples including unstable semiconductor lasers, images of the sky, seismic waves, lava flow, and dual pendulums. Chaos systems vary in practicality; in the aforementioned list, the semiconductor lasers are shown to be very fast with good potential, while the rest are much slower and used in a more casual setting.

While different systems differ in structure, the basic principle is the same. A sensor, typically a camera, records the chaotic system and digitalizes it, such as from colors into binary numbers, before transferring the data to a computer. The data is then processed to eliminate patterns and biases, often with use of a different entropic source or a hashing algorithm, depending on the entropy of the original data. The output is then used as seed for a PRNG, such that a single number can be used to generate an endless stream of random number, improving its bit rate.

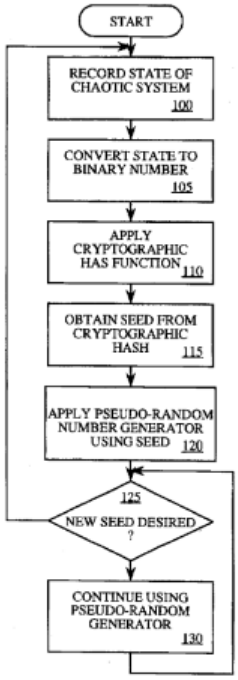


Fig 6. Flowchart of a Chaos-based TRNG
(Noll, Landon Curt, et al. Method for Seeding a Pseudo-Random Number Generator with a Cryptographic Hash of a Digitization of a Chaotic System. 24 Mar. 1998, https://patents.google.com/patent/US5732138/en.)

Chaos-based TRNGs, while perhaps less efficient than other integrated methods, are practical enough for real use cases to be an interesting solution to the randomness problem. Global cloud service provider Cloudflare uses both lava lamps and dual pendulums in their offices as TRNGs, functioning mainly as backup processes that also attracts tourist and media attention for marketing purposes. However, large systems such as lava lamps are unlikely to be used in larger settings due to their lack of scalability and that a proof of randomness is unlikely.

The last issue with the proof of randomness is not limited to Chaos-based systems; rather, it is a question that most TRNGs, and computers in general, struggles with. Apart from quantum-based designs that utilize the inherently random nature of quantum mechanics, it is very difficult to prove any system's randomness, only that they exhibit statistically random properties. Fortunately, in the practical world, it is only sufficient that TRNGs are uncrackable by human means within the time it takes to finish the job at hand.

5. Citations

1. Allini, Elie Noumon, et al. "Evaluation and Monitoring of Free Running Oscillators Serving as Source of Randomness." IACR Transactions on Cryptographic Hardware and Embedded Systems, Aug. 2018, pp. 214–42. tches.iacr.org, https://doi.org/10.13154/tches.v2018.i3.214-242.

2. L'Ecuyer, Pierre. "History of Uniform Random Number Generation." 2017 Winter Simulation Conference (WSC), IEEE, 2017, pp. 202–30. DOI.org (Crossref), https://doi.org/10.1109/WSC.2017.8247790.

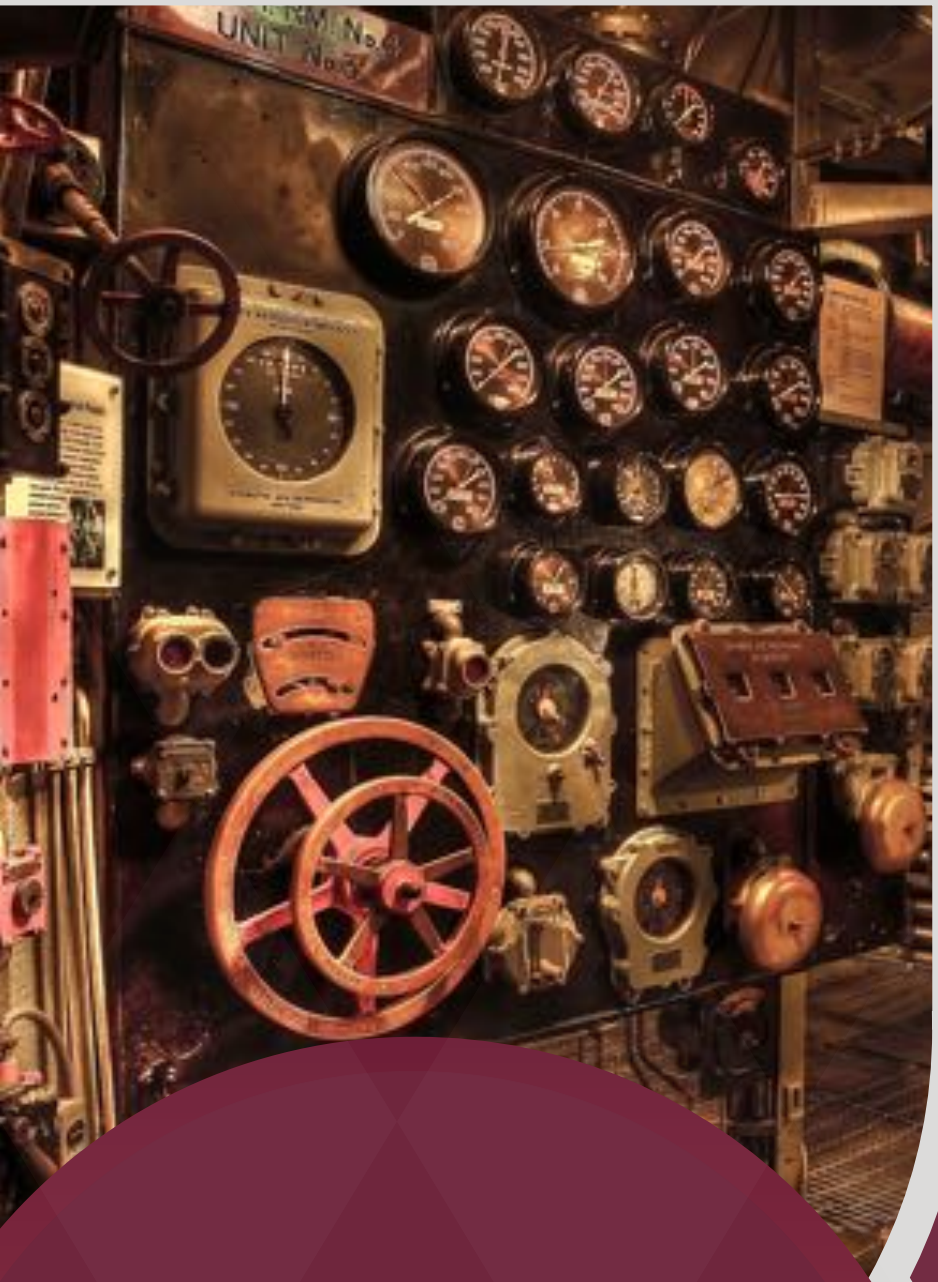
3. Noll, Landon Curt, et al. Method for Seeding a Pseudo-Random Number Generator with a Cryptographic Hash of a Digitization of a Chaotic System. US5732138A, 24 Mar. 1998, https://patents.google.com/patent/US5732138/en.

4. Stipčević, Mario, and Çetin Kaya Koç. "True Random Number Generators." Open Problems in Mathematics and Computational Science, edited by Çetin Kaya Koç, Springer International Publishing, 2014, pp. 275–315. DOI.org (Crossref), https://doi.org/10.1007/978-3-319-10683-0_12.

5. Sunar, Berk, et al. "A Provably Secure True Random Number Generator with Built-In Tolerance to Active Attacks." IEEE Transactions on Computers, vol. 56, no. 1, Jan. 2007, pp. 109–19. DOI.org (Crossref), https://doi.org/10.1109/TC.2007.250627.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE



04

Engineering

Fitzwilliam College Online Winter School Journal · 2024 Winter

Cutting Through the Air: Enhancing Truck Efficiency with Aerodynamics

DONGLAI LI (Leo)

1. Introduction

In our rapidly globalizing world, the transportation of goods across diverse regions has escalated to meet the burgeoning demands of consumers. Amidst this backdrop, trucks play a pivotal role in the logistical chain, bearing the brunt of goods transportation. Therefore, enhancing the aerodynamic efficiency of trucks has emerged as a critical priority, not only to elevate fuel efficiency and safety but also to minimize the environmental impact. This essay explores strategies to improve truck aerodynamics, thereby contributing to a more sustainable and efficient future in goods transportation.

There are three main strategies I will use: streamlining the body, removal of rearview mirrors and underbody modifications, which help to reduce air resistance air turbulence.

2. Basic Principles of Aerodynamics

Aerodynamics delves into the dynamics of air, especially as it interacts with solid objects, including vehicles like cars and the wings of airplanes. By applying aerodynamic principles, I scrutinize the air flow surrounding trucks to engineer an optimal and appropriate shape.

The dynamics of a truck can be decomposed into longitudinal dynamics, lateral dynamics, and vertical dynamics, and in the longitudinal dynamics, the truck is influenced by the air drag (FD). To be more specific, the drag force consists of pressure drag and friction drag. The friction drag arises from the friction between the air and the surface of the moving object, while the pressure drag is caused by the shape of the object and how it displaces the air. To calculate this force, you have to apply the drag coefficient (C_d), which is determined by the shape and design, size and frontal area and Reynolds number.

3. Current Challenges

Today's trucks feature a large cross-sectional area and bluff body design, leading to a high drag coefficient value. 'The average drag coefficient was determined as 0.608 for truck.' [1] By comparison, 'The average modern automobile achieves a drag coefficient of between 0.25 and 0.3.' [2]

Figure 1: Modern truck with sharp edges

Figure 2: Modern truck in wind tunnel test

The figure above demonstrates a modern truck with numerous sharp edges, which can induce turbulence, separation of the flow and low pressure. Consequently, the separation of the flow and turbulence contribute to a large pressure drag.

4. Aerodynamic Improvement Strategies

I will employ three primary strategies: streamlining the body, installation of rearview cameras and modifying the underbody, which are instrumental in decreasing air resistance and minimizing flow separation.

4.1. Streamlining the Body

By gradually transitioning the cross-sectional areas, it enables smoother airflow adaptation and minimizes the wake's size. Consequently, this leads to a reduction in wind resistance experienced by the truck.

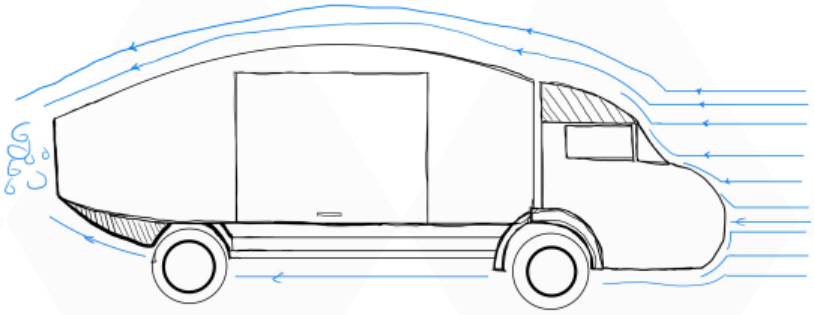


Figure 3: Designed truck

I implemented a total of four modifications to the truck design. The first noticeable change is at the front of the truck. The second is the shaded area above the driving cab. It is an add-on device serving as a roof fairing. Furthermore, I modified the shape of the cargo box to feature a curved roof. As the area at the rear of the truck will be reduced, I moved the cargo door to the side of the cargo box. By changing the cross-sectional area gradually, it allows the flow of the air to adapt and reduce the size of the wake.

4.2. Underbody Modifications

Another important alteration is an underbody modification; I installed an underbody fairing at the rear of the truck, which is designed to minimize the turbulence that forms beneath the vehicle. Furthermore, as can be noticed on Figure 1, the uneven surfaces under the truck, such as the transmission and fuel tanks, which can induce turbulence and may have a high friction coefficient value. Therefore, I installed underbody panels, which is like a shield to create a smoother underbody surface.

Figure 4: Underbody of the improved truck

4.3. Remove the Rearview Mirror

The two rearview mirrors are important to cover the blind points, but they usually take up an unnecessary large area that could increase air drag. The following equation demonstrates the relationships between the cross-sectional area and drag force.

$$F_D = C_D \frac{1}{2} \rho V_\infty^2 S$$

where C_D is the drag coefficient, is the density of air, is the velocity of the object relative to the air, and is the cross-sectional area. By substituting the rearview mirrors with cameras, we can achieve a reduced cross-sectional area, resulting in a smaller drag force, while the rear view can be showed on the screen.

It is suggested that the Actros (Mercedes-Benz truck) achieve a lower fuel consumption by using the MirrorCam which cuts the fuel consumption by up to 1.3% compared to conventional mirrors as standard. [3]

5. Conclusions

In conclusion, enhancing the aerodynamics characteristics of trucks presents a significant opportunity to revolutionize the trucking industry by improving fuel efficiency and reducing emissions. Through the modifications such as streamlining the truck body, integrating rearview camera, and modifying the underbody, the industry can achieve substantial gain in performance and sustainability. Overall, the adoption of these aerodynamics improvements not only aligns with environmental goals but also offers economic benefits to operators.

6. References

[1] Bayındırlı, Cihan & Akansu, Yahya & Salman, M.. (2016). The Determination of Aerodynamic Drag Coefficient Of Truck and Trailer Model By Wind Tunnel Tests. International Journal of Automotive Engineering and Technologies.

[2] Wikipedia contributors (2024). Automobile Drag Coefficient. The Free Encyclopedia.

[3] Mercedes-Benz RoadEfficiency (2024) Low total Costs https://www.mercedes-benz-trucks.com/en_GB/buy/mercedes-benz-roadefficiency/low-total-costs.html

Bird-inspired Winglet Design to Improve Aerodynamic Efficiency of Aircraft

YUETONG WANG (Tim)

1. Introduction

Aerodynamic efficiency is a critical factor in the performance of aircraft. It influences the overall fuel consumption, environmental impact, and operating costs of the aircraft. In the pursuit of enhancing this efficiency, engineers have turned to nature for inspiration, particularly the elegant and efficient flight of birds. One area of particular interest is the design of winglets, which are aerodynamic devices attached to the tips of aircraft wings to reduce drag and improve fuel efficiency. The winglet shapes observed in different bird species are adaptations to their specific flying behaviors and environments. For instance, eagles, known for soaring at high altitudes and attacking at high velocities, often have wingtips that curve upward to generate more lift. This design helps minimize drag and improve stability during soaring flights. The upward-curving tips reduce the effects of turbulence and enhance the bird's efficiency in maintaining altitude (1). On the other hand, birds with more planar wingtips, like swifts or swallows, are typically adapted for agile and maneuverable flight, often at lower altitudes. These straighter wings allow for quick changes in direction and better control during complicated flight patterns (2). This theory of winglets can be applied to modern vehicles with different operating requirements. By studying and equipping the winglet shapes of birds with modern airplanes, aerodynamic efficiency can improve, thus saving fuel.

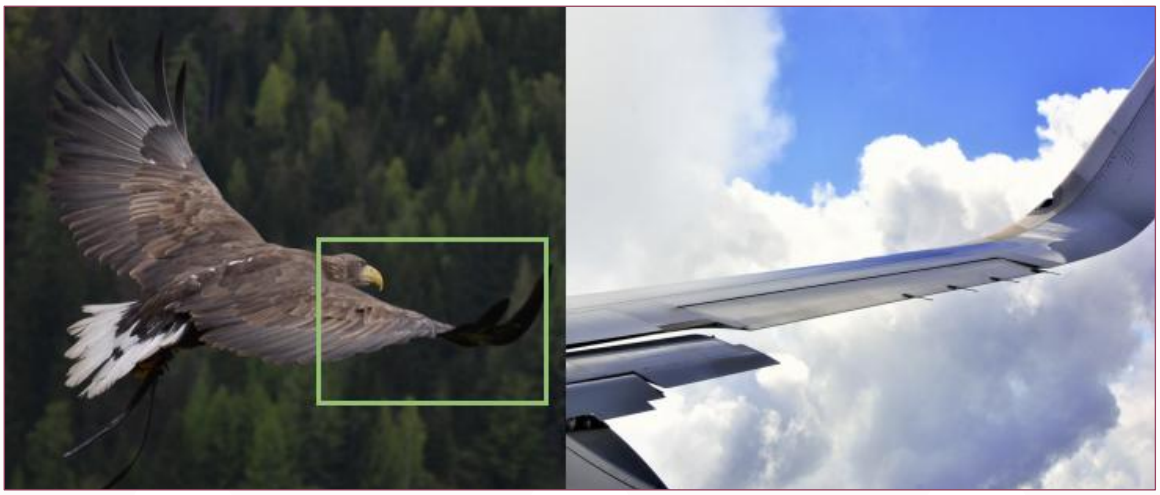
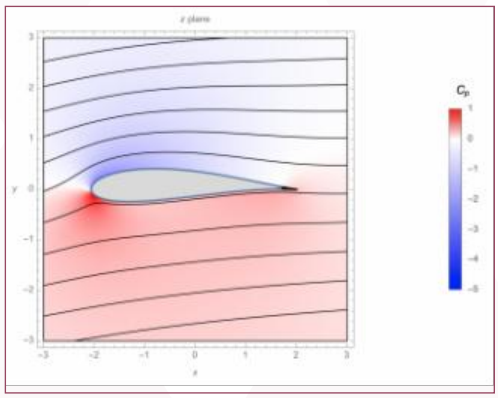


Fig.1 Wingtip of an Eagle and the Winglet on an Airplane (pictures from pixabay.com)

2. Theory of Flight and The Use of Winglets

During an aircraft flight, four main forces are acting on the body, which is the weight of the aircraft, the driving force of the engine, the lift, and the drag. In applied aerodynamics, the magnitude of the lift (L) and drag (D) depends on their separate coefficients, C_L as lift coefficient and C_D as drag coefficient, which are shown by their corresponding equations (3):

$C_L \equiv \frac{L}{\frac{1}{2}\rho_\infty V_\infty^2 S}$ and $C_D \equiv \frac{D}{\frac{1}{2}\rho_\infty V_\infty^2 S}$, where $\frac{1}{2}\rho_\infty V_\infty^2$ is the dynamic pressure and S is the surface area of the wing. The aerodynamic efficiency of an object is defined by the ratio of its total lift force to the drag force (4).



Graph.1 Pressure Distribution of an Airfoil (6)

The ultimate purpose of using the winglets is to reduce the induced drag, also known as vortex drag (5). This kind of drag is produced due to the pressure difference between the top and the bottom of the airfoil during a flight, as illustrated in Graph.1 (simulated using the Mathematica Joukowski Airfoil model (6)). This pressure difference between the upper and lower surfaces of the airfoil pushes the airplane upward, generating lift. However, looking at the airplane as a whole, air flows from the region of low pressure to the region of high pressure. Therefore, the air will flow around the wingtips from below to above, which creates the wingtip vortices (see Fig.2). This effect causes additional rotation of the air and produces a tornado-like flow pattern following behind the aircraft. The vortices themselves require energy to generate, which is bad for the aircraft's efficiency.

Fig.2 Wingtip Vortices Generated From Pressure Difference

Inspired by birds, the solution to this vortex drag is by adding a winglet similar to the wingtip of an eagle that is curved upwards. Winglets reduce the pressure gradient at the tips of the wings, thus restricting the circulation of air and reducing the size of the vortices. Furthermore, it improves the aircraft's efficiency by reducing drag, which enables aircraft to carry more loads and travel a longer distance as fuel consumption is reduced. A study by NASA (7) reveals that fuel efficiency increased by 6% to 7% after the addition of winglets. Moreover, after the winglets were first invented in the 1990s, there were about 10 billion gallons of fuel and over 130 million tons of carbon dioxide emissions that had been reduced, as observed on about 10,000 aircraft.

3. Winglet Types and their effectiveness

Winglets have evolved in a variety of ways since their invention. In 1985, Airbus first introduced its 'wingtip fences' on the A310 (8), with winglet surfaces extending both upwards and downwards. Similar designs were applied to the later A320 and A380. After that, the A320 Neo used sharklets to further reduce induced drag due to the smoother connection between the wingtips and the wing, saving an additional 3.5% of the fuel (9). Other winglet designs, such as the blended structure on the Boeing 757, split-scimitar winglets on the Boeing 737, additional intriguing devices on the Boeing 777x, and even multi-tipped winglets that resemble bird feathers (10), are all carefully designed and tested for particular aircraft.

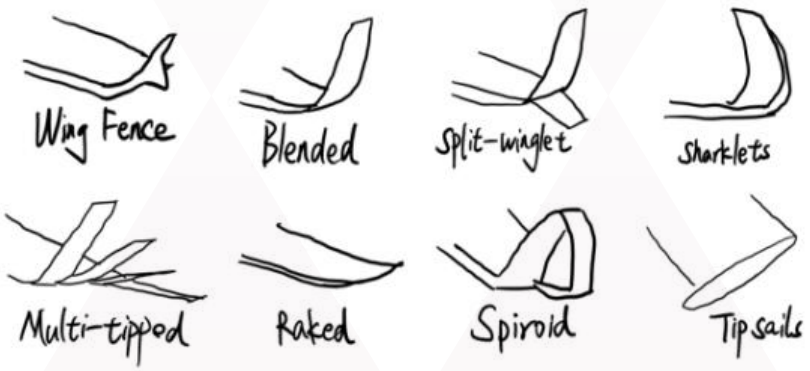


Fig.3 Different Winglet Shapes

4. Conclusions

The invention of winglets inspired by birds has reduced the drag coefficient and fuel consumption of aircraft. This adaptation is especially beneficial for commercial airplanes flying at cruising altitudes. The application of bird-inspired wingtip designs in aircraft showcases how nature's solutions can be translated into practical engineering solutions for various purposes in aviation. Other features of birds such as wing structures and flying postures can be further studied to improve the aerodynamic performance of aircraft. To build on the promising results of this research, numerical simulation and analysis can be conducted to compare the effectiveness of the addition of different winglets on airplanes.

5. References

1. These Masters of the Sky Can Fly for Hours (or Days) While Barely Flapping. Audubon. [Online] [Cited: 2 15, 2024.] https://www.audubon.org/news/these-masters-sky-can-fly-hours-or-days-while-barely-flapping.

2. The Common Swift Is the New Record Holder for Longest Uninterrupted Flight. Audubon. [Online] [Cited: 2 15, 2024.] https://www.audubon.org/news/the-common-swift-new-record-holder-longest-uninterrupted-flight.

3. Joel Guerrero, Marco Sanguineti and Kevin Wittkowski. CFD Study of the Impact of Variable Cant Angle Winglets on Total Drag Reduction. Aerospace. 2018, Vol. 5, 4.

4. Michael Lynch, Boris Mandadzhiev and Aimy Wissa. Bioinspired wingtip devices: a pathway to improve aerodynamic performance during low Reynolds number flight. Bioinspiration & Biomimetics. 2018, Vol. 13, 3.

5. Urrutia, Luis. Design framework for a bio-inspired adaptive wingtip using bend-twist coupled composites. University of Illinois at Urbana-Champaign.

6. Joukowski Airfoil: Flow Field. WOLFRAM Demonstrations Project. [Online] 12 12, 2016. [Cited: 2 14, 2024.] https://demonstrations.wolfram.com/JoukowskiAirfoilFlowField/.

7. Dinius, Dede. On the Winglets of Innovation with NASA Armstrong. NASA. [Online] [Cited: 2 14, 2024.] https://www.nasa.gov/centers-and-facilities/armstrong/on-the-winglets-of-innovation-with-nasa-armstrong-2/#:~:text=Over%20the%20course%20of%2048%20test%20flights%2C%20winglets,reduced%20CO2%20emissions%20by%20over%20130%20million%20tons..

8. MEMON, OMAR. What Was The First Airliner To Have Winglets? Simple Flying. [En ligne] [Citation : 14 2 2024.] https://simpleflying.com/first-airliner-with-winglets/.

9. Rory Hensey and Ana Magdalina. A320 NEO vs. CEO comparison study. FPG Amentum. 2018.

10. Gautham Narayan and Bibin]ohn. Effect of winglets induced tip vortex structure on the performance of subsonic wings. Aerospace Science and Technology. 2016, Vol. 58.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

05

Mathematics for the Natural Sciences

Fitzwilliam College Online Winter School Journal · 2024 Winter

Bessel Function

WILLIAM LIU (William)

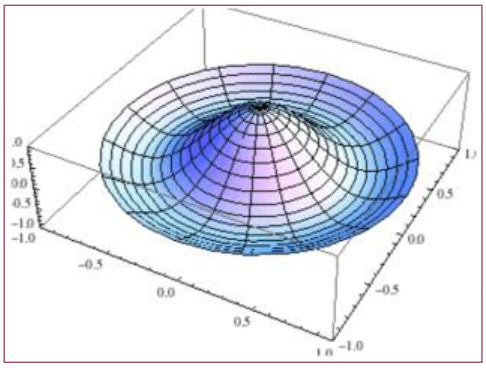
1. The Bessel Equation

- The Bessel Equation is a second order differential equation with strong significance in spherical and cylindrical coordinates.
- The solutions of the Bessel Equation are known as the Bessel Functions.

The Bessel Equation has the form:

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0, \quad \nu \geq 0$$

Where ν is a real constant and denotes the order of the Bessel Equation



Using Bessel Function to demonstrate vibration mode of a circular membrane

https://www.quora.com/Why-do-we-use-the-Bessel-function-in-physics

2. Singular-Points & Frobenus Method

- To solve the Bessel Equation, one needs the technique to utilize power series and back substitution, which is known as the Frobenus Method.
- But Frobenus Method cannot be used for every second order differential equation, so the definition of singular points needs to be involved.

For a second order differential equation of the form:

$$y'' + p(x)y' + q(x)y = 0 \quad (1)$$

The point $x = \alpha$ is called a singular point if:

$$\lim_{x\to\alpha} |p(x)| = \infty \quad or \quad \lim_{x\to\alpha} |q(x)| = \infty$$

Moreover, $x = \alpha$ is a regular singular point if and only if:

$$y'' + p(x)y' + q(x)y = 0$$

Can be written as:

$$(x-\alpha)^2 a(x)y'' + (x-\alpha)b(x)y' + c(x)y = 0$$

3. Frobenus Method

- The Frobenus Method claims that if x = α is a regular singular point of the second differential equation (1), then:

$$y(x) = (x-\alpha)^r \sum_{k=0}^{\infty} a_k (x-\alpha)^k, \quad a_0 \neq 0$$

- Is a special solution to differential equation (1), and the constant r and the coefficients can be determined by re-substituting the solution back into the differential equation.

$$y'(x) = (k+r)\sum_{k=0}^{\infty} a_k x^{k+r-1}$$

$$y''(x) = (k+r-1)(k+r)\sum_{k=0}^{\infty} a_k x^{k+r-2}$$

4. Applying Frobenus Method to solve The Besse Equation

- Notice that the Bessel Equation has a regular singular point at x = 0, and by applying Frobenus Method, we obtain:

The Bessel Equation has the form:

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0, \quad \nu \geq 0$$

Where ν is a real constant and denotes the order of the Bessel Equation.

$$Let \quad y(x) = \sum_{k=0}^{\infty} a_k x^{k+r} \Rightarrow x^2 \sum_{k=0}^{\infty} (k+r)(k+r-1)a_k x^{k+r-2}$$

$$+x\sum_{k=0}^{\infty}(k+r)a_k x^{r+k-1} + (x^2-\nu^2)\sum_{k=0}^{\infty} a_k x^{k+r} = 0$$

$$\Rightarrow \sum_{k=0}^{\infty} a_k x^{k+r}[(k+r-1)(k+r) + (k+r) - \nu^2] + \sum_{k=2}^{\infty} a_{k-2}x^{k+r} = 0$$

$$\Rightarrow x^r(r^2-\nu^2)a_0 + x^{r+1}[(r+1)^2-\nu^2]a_1 + \sum_{k=2}^{\infty} x^{k+r}[a_k((k+r)^2-\nu^2) + a_{k-2}] = 0$$

5. Computing the Coefficients

- Using the fact that the result is the zero function, all the coefficients of x must be 0, and thus we derive a recursive formula for the coefficients of the special solution of the Bessel Equation.
- N here represent the set of all natural numbers.

$$Let \quad y(x) = \sum_{k=0}^{\infty} a_k x^{k+r} \Rightarrow x^2 \sum_{k=0}^{\infty} (k+r)(k+r-1)a_k x^{k+r-2}$$

$$+x\sum_{k=0}^{\infty}(k+r)a_k x^{r+k-1} + (x^2-\nu^2)\sum_{k=0}^{\infty} a_k x^{k+r} = 0$$

$$\Rightarrow \sum_{k=0}^{\infty} a_k x^{k+r}[(k+r-1)(k+r) + (k+r) - \nu^2] + \sum_{k=2}^{\infty} a_{k-2}x^{k+r} = 0$$

$$\cdot x^r(r^2-\nu^2)a_0 + x^{r+1}[(r+1)^2-\nu^2]a_1 + \sum_{k=2}^{\infty} x^{k+r}[a_k((k+r)^2-\nu^2) + a_{k-2}] = 0$$

Since x is the variable, and $a_0 \neq 0$, the coefficients of x must be 0 in order to satisfy the equation.

$$\Rightarrow (r^2-\nu^2) = 0, r = \pm\nu$$

We first consider the case when $r = \nu$, plugging in, we obtain:

$$(2\nu+1)a_1 = 0, a_k(2k\nu + k^2) + a_{k-2} = 0$$

And since $2\nu + 1 > 0$, we have:

$$a_1 = 0, a_k = -\frac{a_{k-2}}{2k\nu + k^2} \Rightarrow a_{2n+1} = 0, n \in N$$

6. The Bessel Function of the First Kind

Therefore, we now only consider the case when $k = 2n$:

$$a_{2n} = -\frac{a_{2n-2}}{2^2 n(\nu+n)} = (-1)^2 \frac{a_{2n-4}}{2^4 n(n-1)(\nu+n)(\nu+n-1)} = \ldots = (-1)^n \frac{a_0 \Gamma(1+\nu)}{2^{2n} n! \Gamma(\nu+n+1)}$$

Where Γ is the gamma function, with properties:

$$\Gamma(x+1) = x\Gamma(x), \quad \Gamma(n) = (n-1)!, n \in N, \quad \Gamma(1) = 1, \quad \Gamma(0) = \Gamma(-n) = \infty, n \in N.$$

Notice that we get to choose the first constant a_0, by convention, we choose $a_0 = \frac{1}{2^\nu \Gamma(1+\nu)}$, and thus we get the first special solution $y(x)$ for the Bessel Equation:

$$J_\nu(x) = y(x) = \sum_{n=0}^{\infty} a_{2n} x^{2n+\nu} = \sum_{n=0}^{\infty} (-1)^n \frac{a_0 \Gamma(1+\nu)}{2^{2n} n! \Gamma(\nu+n+1)} x^{2n+\nu} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(\nu+n+1)} (\frac{x}{2})^{2n+\nu}$$

This solution is denoted as $J_\nu(x)$, and is called the Bessel Function of the first kind of order ν.

7. The Second Solution

- We have derived the first special solution of the Bessel Equation by using r=v , now hopefully the linearly-independent second special solution can be obtained by using r=-v.
- But there is an issue: When -v is a negative integer, n − v+1 can be a negative integer, and Γ(n − v+1) becomes infinite, which affects the linearly-dependency.

For $r = -\nu$, by similar process, we can get:

$$J_{-\nu}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n-\nu+1)} (\frac{x}{2})^{2n-\nu}$$

And in the case when $\nu \notin N$, $J_{-\nu}(x)$ is indeed linearly-independent with $J_\nu(x)$, namely, $\frac{J_\nu(x)}{J_{-\nu}(x)} \neq C, \nu \notin N$ since:

$$J_0(0) = 1, \quad J_\nu(0) = 0, \nu > 0, \quad J_{-\nu}(0) = \infty, \nu \notin N$$

8. The Whole Number Situation

- As a result, we need to investigate the Bessel Function of the first kind when v is a positive integer.
- Unfortunately, the result turns out to be that there is still another unknown special solution for the Bessel Equation when v is a positive integer.

When $\nu \in N$, using $\Gamma(n+1) = n!$, $n+1 \in N$, we have:

$$J_\nu(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+\nu)!} (\frac{x}{2})^{2n+\nu}$$

Notice in the case of $J_{-\nu}(x)$, when $n - \nu + 1 < 0$:

$$\frac{1}{\Gamma(n-\nu+1)} = 0 \Rightarrow J_{-\nu}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n-\nu+1)} (\frac{x}{2})^{2n-\nu} = \sum_{n=\nu}^{\infty} \frac{(-1)^n}{n!(n-\nu)!}$$

Using the substitution $s = n - \nu$, we get:

$$J_{-\nu}(x) = \sum_{s=0}^{\infty} \frac{(-1)^{s+\nu}}{(s+\nu)!s!} (\frac{x}{2})^{2s+\nu} = (-1)^\nu \sum_{s=0}^{\infty} \frac{(-1)^s}{(s+\nu)!s!} (\frac{x}{2})^{2s+\nu} = (-1)^\nu J_\nu(x)$$

Therefore, when $\nu \in N$, the two solutions $J_\nu(x), J_{-\nu}(x)$ are linearly dependent.

9. The Bessel Function of the Second Kind

- The second linearly-independent special solution of the Bessel Equation when the order is a positive integer is known as The Bessel Function of the Second Kind.
- To construct such solution, the basic idea is to start with a positive non-integer order, then using limits to approach to the positive integer.

For instance, we could construct:

$$Y_n(x) = \lim_{\varepsilon \to 0} \frac{J_{n+\varepsilon}(x) - (-1)^n J_{-n-\varepsilon}(x)}{\varepsilon}$$

For positive integer n, and we could use the L'Hopital's law to compute the limit explicitly:

$$Y_n(x) = [\frac{d}{d\varepsilon} J_{n+\varepsilon}(x) - (-1)^n \frac{d}{d\varepsilon} J_{-n-\varepsilon}(x)]_{\varepsilon=0}$$

By using the power series form of $J_{n+\varepsilon}(x)$ and $J_{-n-\varepsilon}(x)$, we could calculate the derivatives. And then we can verify the linear dependency and that $Y_n(x)$ is indeed a solution when $n \in N$.

10. The Neumann Function

A well-known second kind Bessel Function is the Neumann Function.

The Neumann function is defined as:

$$Y_\nu(x) = \lim_{\alpha \to \nu} \frac{\cos \alpha\pi J_\alpha(x) - J_{-\alpha}(x)}{\sin \alpha\pi}$$

If $\nu \notin N$, then:

$$Y_\nu(x) = \frac{\cos \nu\pi J_\nu(x) - J_{-\nu}(x)}{\sin \nu\pi}$$

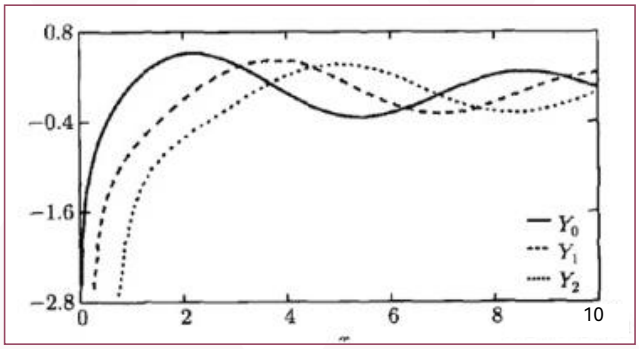
Since in this case $J_{-\nu}(x)$ is linearly-independent with $J_\nu(x)$, $Y_\nu(x)$ is also linearly-independent with $J_\nu(x)$ and also solves the equation.

But if $\nu \in N$, the function becomes a $\frac{0}{0}$ limit, and we use L'Hopital law to compute the explicit form of the limit:

$$Y_\nu(x) = \lim_{\alpha \to \nu} \frac{(\cos \alpha\pi J_\alpha(x) - J_{-\alpha}(x))'}{(\sin \alpha\pi)'} = \lim_{\alpha \to \nu} \frac{-\pi \sin \alpha\pi J_\alpha(x) + \cos \alpha\pi \frac{d}{d\alpha} J_\alpha(x) - \frac{d}{d\alpha} J_{-\alpha}(x)}{\pi \cos \alpha\pi}$$

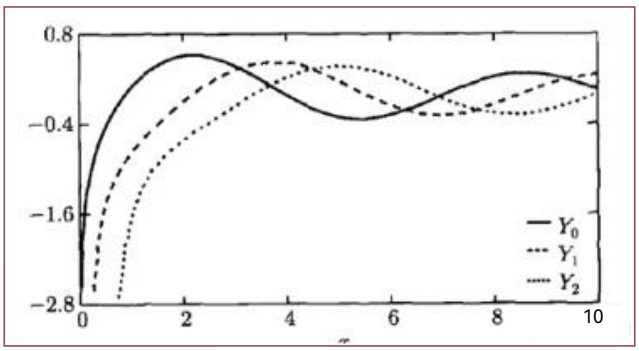
$$= \frac{1}{\pi} [\frac{d}{d\alpha} J_\alpha(x) + (-1)^{\nu+1} \frac{d}{d\alpha} J_{-\alpha}(x)]_{\alpha=\nu}$$

Similarly, we could derive the explicit form of $Y_\nu(x)$, and the graph looks like:



https://zhuanlan.zhihu.com/p/662781878

11. Construction of the Neumann Function



https://zhuanlan.zhihu.com/p/662781878

Noticing the fact that:

$$\cos n\pi = (-1)^n,\ n \in N, \quad \sin n\pi = 0,\ n \in N$$

So the construction of Neumann function also follows the basic idea we mentioned before.

We now verify that the Neumann function solves the equation when $n \in N$:

We start by taking the derivative of α both sides when plugging in $J_\alpha(x)$ into the equation:

$$\frac{d}{d\alpha}[x^2\frac{d^2}{d^2x}J_\alpha(x) + x\frac{d}{dx}J_\alpha(x) + (x^2-\alpha^2)J_\alpha(x)] = 0$$

$$\Rightarrow x^2\frac{d^2}{d^2x}\frac{d}{d\alpha}J_\alpha(x) + x\frac{d}{dx}\frac{d}{d\alpha}J_\alpha(x) + (x^2-\alpha^2)\frac{d}{d\alpha}J_\alpha(x) - 2\alpha J_\alpha(x) = 0$$

$$\Rightarrow x^2\frac{d^2}{d^2x}\frac{d}{d\alpha}J_\alpha(x) + x\frac{d}{dx}\frac{d}{d\alpha}J_\alpha(x) + (x^2-\alpha^2)\frac{d}{d\alpha}J_\alpha(x) = 2\alpha J_\alpha(x)$$

Similarly, we have:

$$x^2\frac{d^2}{d^2x}\frac{d}{d\alpha}J_{-\alpha}(x) + x\frac{d}{dx}\frac{d}{d\alpha}J_{-\alpha}(x) + (x^2-\alpha^2)\frac{d}{d\alpha}J_{-\alpha}(x) = 2\alpha J_{-\alpha}(x)$$

Then, plug in $Y_\nu(x)$ into the equation, we obtain:

$$x^2\frac{d^2}{d^2x}Y_\nu(x)+x\frac{d}{dx}Y_\nu(x)+(x^2-\nu^2)Y_\nu(x) = \frac{1}{\pi}[2\alpha J_\alpha(x)-2\alpha(-1)^\nu J_{-\alpha}(x)]_{\alpha=\nu} = 0$$

Moreover, the graph shows that $\lim_{x\to 0^+} Y_\nu(x) = -\infty$ and we have

$$J_0(0) = 1, \quad J_\nu(0) = 0,\ \nu > 0,$$

thus $Y_\nu(x)$ is linearly-independent with $J_\nu(x)$.

12. Citations

1. 热带鱼学生 . “贝塞尔函数推导（Bessel Function）.” 知乎专栏 , 26 Oct. 2023,https://zhuanlan.zhihu.com/p/662781878.

2. Dobrushkin, Vladimir. “MATHEMATICA Tutorial, Part 2.7: Second Kind Bessel Functions.” Cfm.Brown, https://www.cfm.brown.edu/people/dobrush/am34/Mathematica/ch7/besselY.html. Accessed 14 Feb. 2024.

3. Faculty of Khan. “Bessel Function of the 2nd Kind.” YouTube, 10 July 2017, https://www.youtube.com/watch?v=TW_20TBF0XY.

4. Neumann function. Encyclopedia of Mathematics. URL:http://encyclopediaofmath.org/index.php?title=Neumann_function&oldid=43795

5. Dubey, Prosenjit. “Why Do We Use the Bessel Function in Physics?” Quora, 2018, https://www.quora.com/Why-do-we-use-the-Bessel-function-in-physics.

6. Niedziela, Jennifer. “Bessel Functions and Their Applications.” Sces.Phys, 29 Oct. 2008, sces.phys.utk.edu/~moreo/mm08/niedzilla.pdf.

7. “Bessel Equations and Bessel Functions.” Faculty.Fiu, faculty.fiu.edu/~meziani/Note11.pdf. Accessed 14 Feb. 2024.

8. Garcia, Markel Epelde. “Bessel Functions and Equations of Mathematical Physics.” Addi, 25 June 2015, addi.ehu.es/bitstream/handle/10810/17969/Bessel Funtions_Epelde García.pdf?sequence=2&isAllowed=y.

9. Weisstein, Eric W. "Bessel Function of the Second Kind." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/BesselFunctionoftheSecondKin d.html

10. Mainardi, Francesco. “The Bessel Functions.” Fractional Calculus and Waves in Linear Viscoelasticy, 2010, appliedmath.brown.edu/sites/default/files/fractional/35%20TheBesselFun ctions.pdf.

11. Borji, Amir. “Bessel Functions.” Sharif.Edu. Accessed 13 Feb. 2024.

12. Rimmer. Microsoft PowerPoint - Ch12sc6frbess. Accessed 13 Feb. 2024.

13. Weisstein, Eric W. “Indicial Equation -- from Wolfram MathWorld.” Wolfram Mathworld, https://mathworld.wolfram.com/IndicialEquation.html. Accessed 13 Feb. 2024.

14. Brar, Jatinder Singh, and David Lemon. “Application of Bessel Function.” Mathematics Stack Exchange, 19 Aug. 2014, https://math.stackexchange.com/questions/902655/application-of-bessel-function.

15. Contributors to Wikimedia projects. “Bessel Function.” Wikipedia, 8 Feb. 2024, https://en.wikipedia.org/wiki/Bessel_function.

16.Libretexts. “7.3: Singular Points and the Method of Frobenius.” Libretexts, 7 Nov. 2013, https://math.libretexts.org/Bookshelves/Differential_Equations/Differential_Equations_for_Engineers_(Lebl)/7%3A_Power_ser ies_methods/7.3%3A_Singular_Points_and_the_Method_of_Frobenius.

17. Gantumur, Tsogtgerel. “Frobenius Series Solutions.” Math.Mcgill. Accessed 14 Feb. 2024.

18. “The Big Theorem on the Frobenius Method, With Applications.” Howellkb.Uah. Accessed 12 Feb. 2024.

19. University of Guelph. “Chapter 5: Bessel Functions.” Physics, https://www.physics.uoguelph.ca/chapter-5-bessel-functions. Accessed 14 Feb. 2024.

20. Libretexts. “5.5: Fourier-Bessel Series.” Libretexts, 14 Nov. 2021, https://math.libretexts.org/Bookshelves/Differential_Equations/Introduction_to_Partial_Differential_Equations_(Herman)/05%3A_Non-sinusoidal_Harmonics_and_Special_Functions/5.05%3A_Fourier-Bessel_Series.

21.Libretexts. “4.6: Bessel Functions.” Libretexts, 15 Nov. 2021, https://math.libretexts.org/Bookshelves/Differential_Equations/A_First_Course_in_Differential_Equations_for_Scientists_and_ Engineers_(Herman)/04%3A_Series_Solutions/4.06%3A_Bessel_Functions.

22.Libretexts. “4.6: Bessel Functions.” Libretexts, 15 Nov. 2021, https://math.libretexts.org/Bookshelves/Differential_Equations/A_First_Course_in_Differential_Equations_for_Scientists_and_ Engineers_(Herman)/04%3A_Series_Solutions/4.06%3A_Bessel_Functions.

23. University of Guelph. “Chapter 8: Expansion in Orthogonal Functions.” Physics, https://www.physics.uoguelph.ca/chapter-8-expansion-orthogonal-functions. Accessed 14 Feb. 2024.

The Beta Function

SHAOJIONG YAO (Alexandra)

1. What is the Beta Function

▶ Definition :

The integral, $\int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ (1), is called the Eulerian integral of the first kind by Legendre and Whittaker and Watson (1990).

The solution is the beta function $B(a,b)$. $(a>0, b>0)$

▶ $B(a,b) = B(b,a)$

let $y = 1-x$ in (1) $\int_0^1 x^{a-1}(1-x)^{b-1}\,dx$

$$B(a,b) = \int_1^0 (1-y)^{a-1} y^{b-1}(-dy) = \int_0^1 (1-y)^{a-1} y^{b-1}\,dy = B(b,a)$$

▶ $B(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$

From the definition of gamma function, we know

$$\Gamma(a)\,\Gamma(b) = \int_0^\infty \int_0^\infty x^{a-1} y^{b-1} e^{-(x+y)}\,dx\,dy. \quad (2)$$

?

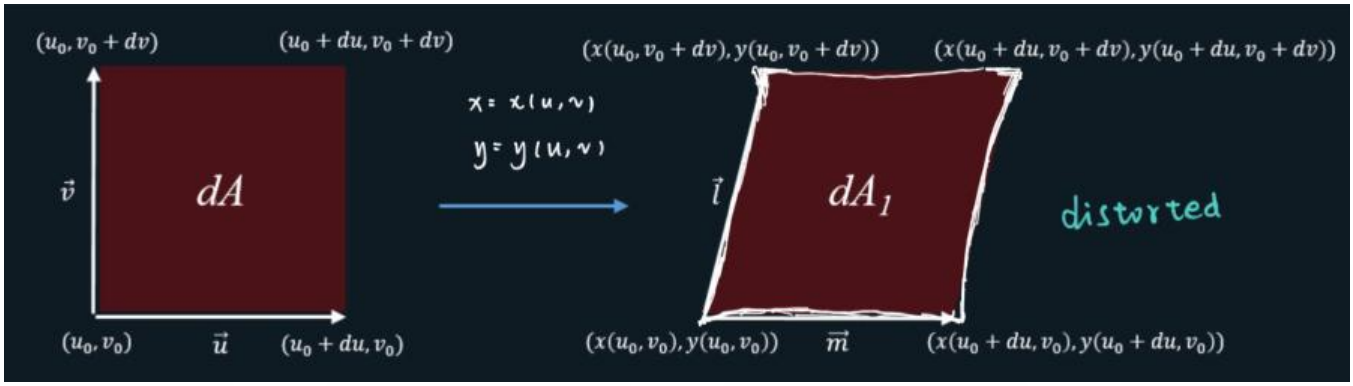
Change variables $x = uv$, $y = u(1-v)$,

▶ Jacobian Tips: 1 Upper and lower limits of integral.

2 Integrand function

3 Integral variables.

Jacobian



shape changes.

$dx\,dy = dA$ $\qquad$ $du\,dv = dA_1$

$dA = |\vec{u} \times \vec{v}| = du\,dv$, $dA_1 = |\vec{l} \times \vec{m}|$

$l = (x(u_0, v_0 + dv) - x(u_0, v_0),\ y(u_0, v_0 + dv) - y(u_0, v_0))$

$m = (x(u_0 + du, v_0) - x(u_0, v_0),\ y(u_0 + du, v_0) - y(u_0, v_0))$

- Think of Multivariable calculus.

$x(u+du, v+dv) - x(u,v) = x'_u\,du + x'_v\,dv = \frac{\partial x}{\partial u}\Big|_v + \frac{\partial x}{\partial v}\Big|_u$

$y(u+du, v+dv) - y(u,v) = y'_u\,du + y'_v\,dv = \frac{\partial y}{\partial u}\Big|_v + \frac{\partial y}{\partial v}\Big|_u$

So $l = (x'_v\,dv,\ y'_v\,dv)$, $m = (x'_u\,du,\ y'_u\,du)$

Finally, $dA_1 = |(x'_v\,dv, y'_v\,dv) \times (x'_u\,du, y'_u\,du)|$

$= |x'_v y'_u - y'_v x'_u|\,du\,dv = dA$

Absolute value

$$dA_1 = |J|\, dA$$

Method :

$$J = \begin{vmatrix} x_u' & x_v' \\ y_u' & y_v' \end{vmatrix} = x_u' y_v' - x_v' y_u'$$

$$\Gamma(a)\,\Gamma(b) = \int_0^\infty \int_0^\infty x^{a-1} y^{b-1} e^{-(x+y)}\, dx\, dy. \quad (2)$$

Change variables $x = uv$, $y = u(1-v)$,

$$x_u' = v \qquad x_v' = u$$

$$y_u' = 1-v \qquad y_v' = -u$$

$$J = v(-u) - u\cdot(1-v)$$

$$= -u$$

Remember to change the limits

$$(2) = \int_{v=0}^{1} \int_{u=0}^{\infty} (uv)^{a-1} [u(1-v)]^{b-1} \cdot e^{-[uv+u(1-v)]} \cdot u \; du\,dv$$

$$= \int_{v=0}^{1} \int_{u=0}^{\infty} u^{a-1} \cdot v^{a-1} \cdot u^{b-1} \cdot (1-v)^{b-1} \cdot e^{-u} \cdot u \; du\,dv$$

$$= \int_{u=0}^{\infty} u^{a+b-1} \cdot e^{-u}\, du \int_{v=0}^{1} v^{a-1} (1-v)^{b-1}\, dv$$

Double Integrals + Simple Limits Separate!

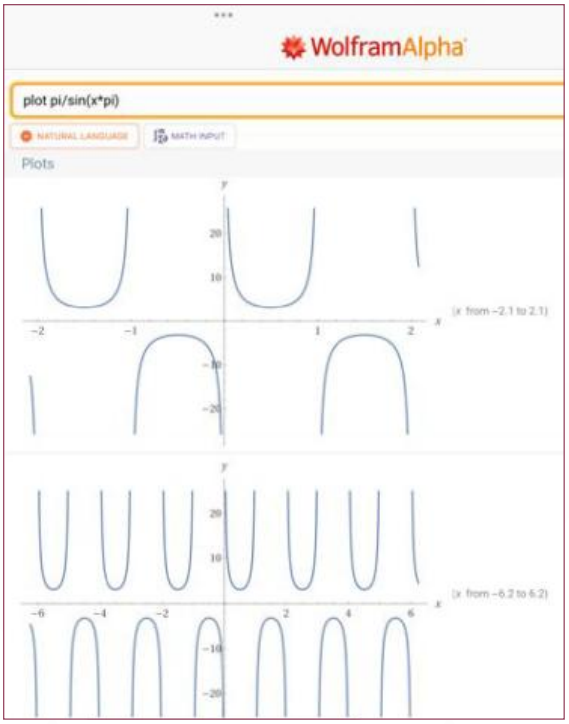
$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t}\, dt \quad (z>0)$$

$$B(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1}\, dt \quad (a>0, b>0)$$

$$(2) \quad \Gamma(a)\,\Gamma(b) = \Gamma(a+b) \cdot B(a,b)$$

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin(p\pi)} \quad (0<p<1)$$



2. References

1. https://zhuanlan.zhihu.com/p/496073353

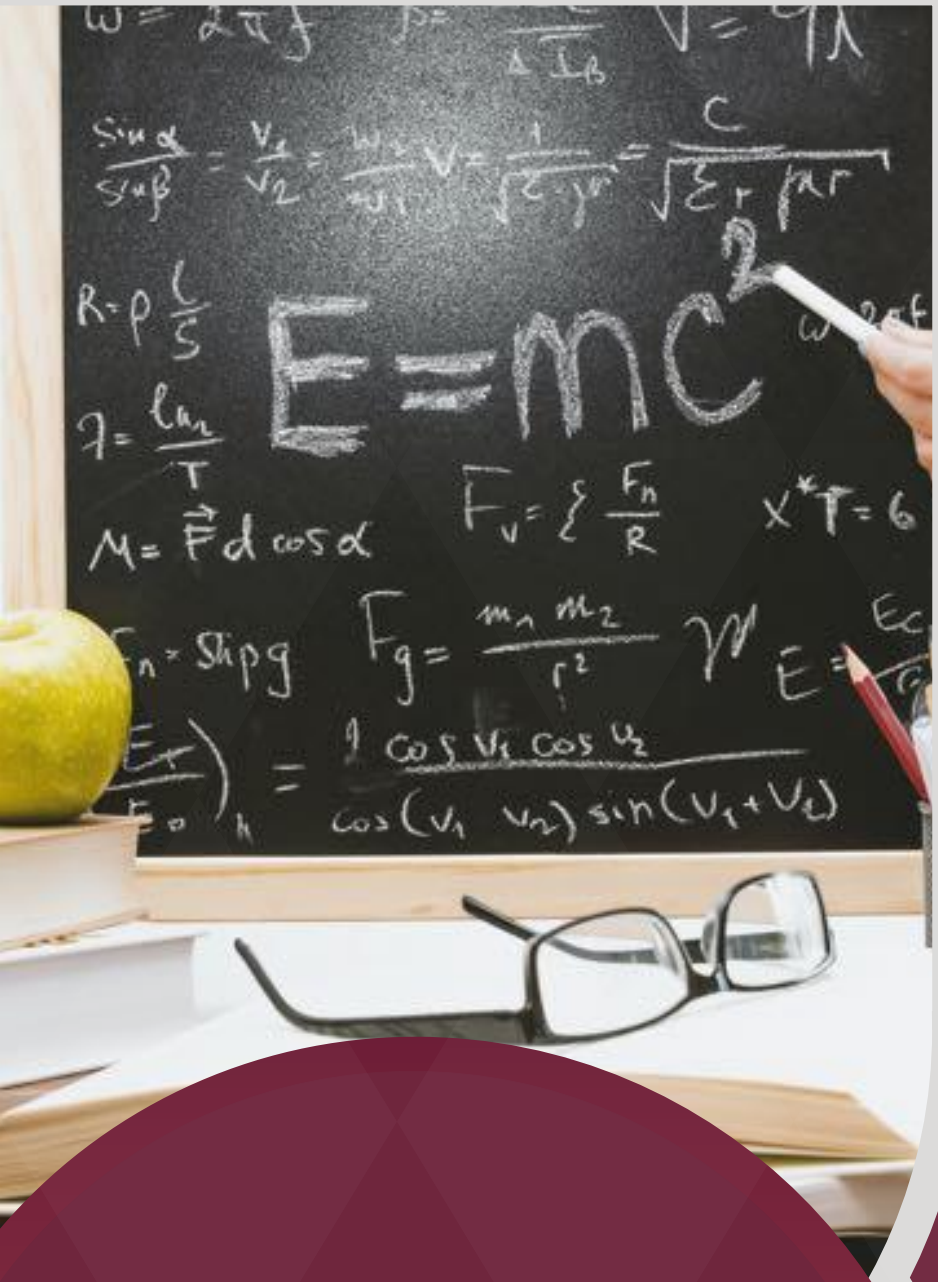
2. https://zhuanlan.zhihu.com/p/382301310

3. https://baike.baidu.com/item/ 贝塔函数 /4643050

4. https://blog.csdn.net/lucien_zong/article/details/50041341

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE



90

Physics

Fitzwilliam College Online Winter School Journal · 2024 Winter

Relativistic Electron Behavior and Magnetic Confinement in Synchrotron Radiation Facility

ALBERT MA (Albert)

Abstract

This essay provides an in-depth study of the role of relativistic effects in the operation of synchrotron radiation facilities. Synchrotrons accelerate charged particles (such as electrons) to velocities close to the speed of light, requiring the application of the Lorentz transformation of special relativity for accurate momentum calculations. When electrons reach relativistic speeds, their behavior results in the emission of synchrotron radiation. This essay describes the basic principles that govern these particles, including the influence of the Lorentz force and the modification of particle mass and energy by relativistic effects. The Shanghai Synchrotron Radiation Facility (SSRF) serves as a case study illustrating the practical implementation of these principles in a state-of-the-art research institution. The relationship between relativistic physics and synchrotron engineering design would highlight the profound impact of Einstein's theory on modern scientific exploration and technology.

1. Introduction

A synchrotron is a type of cyclic particle accelerator in which charged particles like electrons or protons are propelled to high speeds along a fixed closed-loop path (Lim and Lewis 2013). Synchrotron light sources, pivotal in disciplines such as materials science and molecular biology, rely on the principles of special relativity to accelerate electrons to speeds where they emit a broad spectrum of radiation (Lim and Lewis 2013). The role of relativity in the function of synchrotron radiation facilities is a testament to the profound application of Einstein's theory of relativity. The exploitation of relativistic effects is increasingly essential, leading to significant enhancements in the capabilities of these instruments (Kolomenski and Lebedev 1956).

This essay will center on the intricacies of relativistic effects and their influence on the energy, velocity, and power of an electron within a synchrotron. It will also consider how the synchrotron's magnetic field and curvature radius impact the power of emitted radiation, demonstrating the significant interplay between physics at relativistic speeds and the engineering design of these advanced research facilities.

1.1. Synchrotron radiation

The principle of a synchrotron is about electromagnetic fields used to accelerate charged particles to high energies along a circular path. The central concept is based on Lorentz force, which dictates the motion of charged particles in electromagnetic fields. Precisely, a charged particle moving in a magnetic field experiences a force perpendicular to both its velocity and the magnetic field, causing it to follow a circular or spiral path.

If an electron's speed is substantially less than the speed of light, it produces gyro radiation, characterized by $v \ll c$. When electrons possess kinetic energies like their rest mass energy $m_e c^2$, they emit cyclotron radiation. Conversely, electrons with kinetic energies greatly exceeding their rest mass energy generate synchrotron radiation and these electrons are called ultrarelativistic electrons.

1.2. Lorentz transformation

In a synchrotron, the charged electrons travel at velocities approaching the speed of light; therefore, accurate momentum calculations for these electrons necessitate the application of the Lorentz transformation from special relativity.

Lorentz transformations are linear transformations that relate two coordinate frames in spacetime, one stationary and the

other moving at a constant velocity, with the transformations parameterized by the negative of this relative velocity, and were formulated by the Dutch physicist Hendrik Lorentz to describe how events are observed in both frames (Condon and Ransom, 2016).

The Lorentz transform connects the coordinates of an event (x, y, z, t) in a stationary frame to those (x', y', z', t') in a frame moving at velocity v along the x-axis.

$$x = \gamma(x' + vt'),\ y = y',\ z = z',\ t = \gamma(t' + \beta x'/c) \quad (1)$$

Or

$$x' = \gamma(x - vt),\ y' = y,\ z' = z,\ t' = \gamma(t - \beta x/c) \quad (2)$$

Where:

$$\beta \equiv v/c \quad (3)$$

And

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \quad (4)$$

γ is also called Lorentz factor.

1.3. Relativistic masses

According to Einstein's famous mass-energy equation,

$$E = mc^2 \quad (5)$$

The ultrarelativistic electrons have an adjust mass-energy equation.

$$E = \gamma mc^2 \quad (6)$$

1.4. The orbit of electrons in synchrotron radiation facility

The popular design of a synchrotron radiation facility is like a ring, such as the Shanghai Synchrotron Radiation Facility (SSRF) in Pudong area of Shanghai, China (Wu 2022). Electrons are stored in the storage ring to keep the kinetic energy of the electrons to be constant.

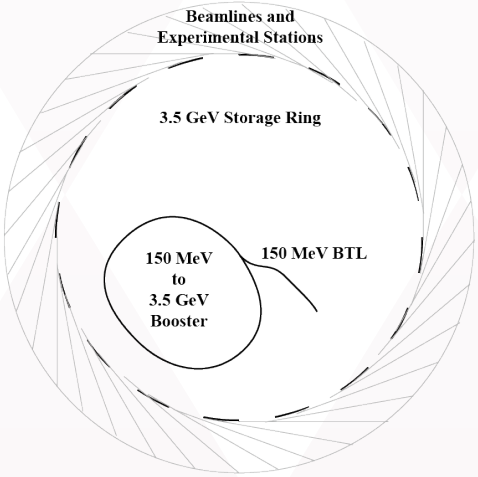


Figure 1. Layout of the SSRF storage ring

Ultrarelativistic electrons in a synchrotron follow spiral paths along magnetic field lines. The angular frequency ω is vital for defining their rotational speed and energy radiation. The lower angular frequencies (ω_B) are a result of increased inertial mass, amplified by the Lorentz factor (γ), and are essential for characterizing electron behavior and emitted synchrotron radiation.

$$\omega_B = \frac{qB}{mc} = \frac{eB}{\gamma m_e c} = \frac{\omega_G}{\gamma} \tag{7}$$

And

$$v_B = \frac{\omega_B}{2\pi} \tag{8}$$

1.5. Energy of a single electron in synchrotron

When a particle with electric charge q is at rest or moves with a constant velocity, its electric field lines are purely radial: E = Er. If a charged particle initially at rest is accelerated to a small velocity Δv < < c in a short time Δt, it disturbs the lines of force, and the disturbance travels outward at the speed of light c. At time t after the acceleration, the disturbance will have propagated to r=ct, and the perpendicular component of the electric field will have a magnitude (Figure 2) (Condon and Ransom, 2016).

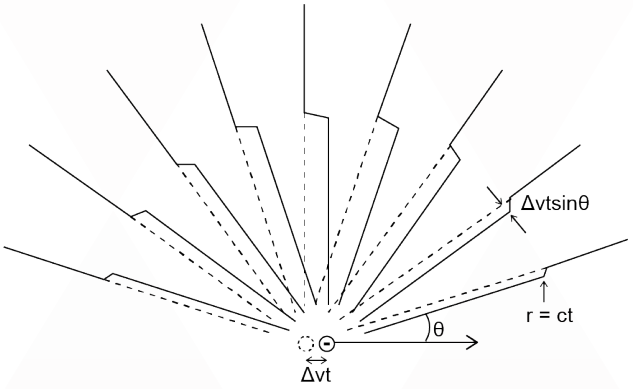


Figure 2. The electric field lines from an accelerated electron. The dotted circle shows the initial position of the electron, and the dotted lines are the radial lines of force emanating from that position. At time t after a small acceleration Δv/Δt, the electron position has moved by Δvt and its lines of force have shifted transversely by Δvtsinθ.

$$\frac{E_\perp}{E_r} = \frac{\Delta v t sin\theta}{c\Delta t} \tag{9}$$

According to Coulomb's law for the radial component Er of the electric field (electric force per unit charge) a distance r from a stationary charge q is

$$E_r = \frac{q}{r^2} \tag{10}$$

Therefore,

$$E_\perp = \frac{q}{r^2} \cdot \frac{\Delta v_\perp t sin\theta}{c\Delta t} = \frac{qvsin\theta}{rc^2} \tag{11}$$

where

$$v \equiv \lim_{\Delta t \to 0} \left(\frac{\Delta v}{\Delta t}\right)$$

1.6. The total power emitted by the charged particle

What the power pattern of a charged particle is? It's a dipolar power pattern shaped like a doughnut and the axis is parallel to the acceleration a (Figure 3) (Condon and Ransom, 2016).

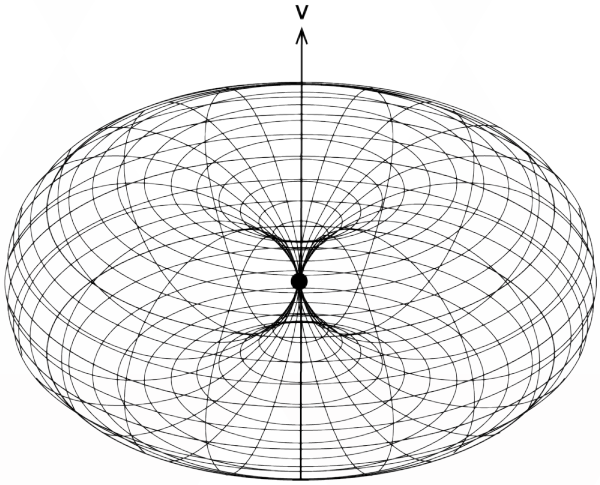


Figure 3. The power pattern of a charged particle. The power received from any direction is proportionate to the component of v that is perpendicular to the observer's line of sight.

The Poynting flux (power per area) is

$$S = \frac{c}{4\pi}E^2 \tag{12}$$

Insert equation (11) to (12) gives

$$S = \frac{c}{4\pi}\left(\frac{qa\sin\theta}{rc^2}\right)^2 = \left(\frac{q^2a^2}{4\pi c^3}\right)\frac{\sin^2\theta}{r^2}$$

Therefore, the total power is

$$P = \int_{sphere} SdA = \frac{q^2a^2}{4\pi c^3}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi}\frac{\sin^2\theta}{r^2}r\sin\theta\, d\theta r d\varphi = \frac{q^2a^2}{2c^3}\int_{\theta=0}^{\pi}\sin^3\theta\, d\theta$$

The integral , so $\int_{\theta=0}^{\pi}\sin^3\theta\, d\theta = \frac{4}{3}$,

$$P = \frac{2}{3}\frac{q^2a^2}{c^3} \tag{13}$$

1.7. The synchrotron power emitted by a single electron

Because

$$a_\perp = \frac{dv_\perp}{dt} = \frac{dv_\perp}{dt'}\frac{dt'}{dt} = \frac{1}{\gamma}\frac{dv_\perp'}{dt'}\frac{dt'}{dt} = \frac{a_\perp'}{\gamma^2} \tag{14}$$

Inserting equation (14) into (13) gives

$$P' = \frac{2e^2(a')^2}{3c^3} = \frac{2e^2{a_\perp}^2\gamma^4}{3c^3} \tag{15}$$

Additionally, inserting equation (5) into (14) gives

$$a_{\perp} = \frac{dv_{\perp}}{dt} = \omega_B v_{\perp} = \frac{eBv_{\perp}}{\gamma m_e c} \tag{16}$$

To calculate the electron power P, we have

$$P = \frac{dE}{dt} = \frac{dE}{dE'}\frac{dE'}{dt'}\frac{dt'}{dt} = \gamma \frac{dE'}{dt'}\frac{1}{\gamma} = \frac{dE'}{dt'} = P' \tag{17}$$

So

$$P = P' = \frac{2e^2 a_{\perp}{}^2 \gamma^4}{3c^3} \tag{18}$$

Then insert equation (16) to equation (18), we can get

$$P = \frac{2e^2 a_{\perp}{}^2 \gamma^4}{3c^3} = \frac{2e^2\gamma^2}{3c^3}\left(\frac{eBv_{\perp}}{m_e c}\right)^2 = \frac{2e^2\gamma^2}{3c^3}\left(\frac{eBv\sin\theta}{m_e c}\right)^2 \tag{19}$$

Let

$$\sigma_T = \frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2 = \frac{8\pi}{3}\left(\frac{(4.8\times10^{-10} statcoul)^2}{9.1\times10^{-28} g\cdot(3\times10^{10} cms^{-1})^2}\right)^2 = 6.65\times10^{-25} cm^2$$

$$U_B = \frac{B^2}{8\pi}$$

Then

$$P = \left[\frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2\right]\cdot 2\cdot\left(\frac{B^2}{8\pi}\right)c\gamma^2\frac{v^2}{c^2}\sin^2\theta = 2\sigma_T\beta^2\gamma^2 c U_B \sin^2\theta \tag{20}$$

Also,

$$\sin^2\theta = \frac{\int \sin^2\theta\, d\Omega}{\int d\Omega} = \frac{1}{4\pi}\int \sin^2\theta\, d\Omega = \frac{1}{4\pi}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi} \sin^2\theta\sin\theta\, d\theta r d\varphi$$
$$= \frac{1}{4\pi}2\pi\frac{4}{3} = \frac{2}{3}$$

Therefore,

$$P = \frac{4}{3}\sigma_T\beta^2\gamma^2 c U_B \tag{21}$$

2. Result

2.1. Energy of an electron at rest status

According to equation (5), the energy corresponding to the electron's rest mass m_e is

$$E_0 = m_e c^2 = 9.1\times10^{-31} kg\cdot(3\times10^8 ms^{-1})^2 = 8.2\times10^{-14} J$$
$$= \frac{8.2\times10^{-14} J}{1.6\times10^{-19} JeV^{-1}} = 5.1\times10^5 eV = 0.51 MeV$$

There, the energy of an electron at rest status is 0.51MeV.

2.2. The Lorentz factor of an electron in SSRF

The usual operational status of the Shanghai Synchrotron Radiation Facility is 3.5GeV energy in the storage ring (Wu 2022). Now we need to find the Lorentz factor γ for the relativistic electron in SSRF.

According to equation (6), we get

$$\gamma = \frac{E}{mc^2} = \frac{3.5 \times 10^9 eV \times 1.6 \times 10^{-19} JeV^{-1}}{9.1 \times 10^{-31} kg \cdot (3 \times 10^8 ms^{-1})^2} = 6.8 \times 10^3$$

According to equation (4), we can obtain the value of β.

$$\beta = \sqrt{1 - (\frac{1}{\gamma})^2} = \sqrt{1 - (\frac{1}{6.8 \times 10^3})^2} = 0.99999999$$

2.3. The velocity of a relativistic electron

So, the electron velocity is

$$v = \beta c = 0.99999999 \times 3 \times 10^8 ms^{-1} = 2.99999997 \times 10^8 ms^{-1}$$

According to equation (19),

2.4. Relativistic effect on the power of an electron in synchrotron facility

According to equation (21), $P \propto \beta^2, \gamma^2, and\ U_B$. When $\gamma \gg 1$, the $\beta^2 = 1 - \gamma^{-2} \approx 1$.
So β^2 can be ignored, and $P \propto \gamma^2, and\ U_B$.

Moreover, $U_B = B^2/8\pi$ (B is the magnetic field strength). We know that

$$r = \frac{mv}{qB} \qquad (22)$$

where r is the radius of the magnetic field. So the value of B is related to the radius of the synchrotron.

$$P \propto \gamma^2 U_B$$

$$P \propto \gamma^2 B^2$$

$$P \propto \frac{B^2}{c^2 - v^2}$$

$$P \propto \frac{B^2}{c^2 m_e{}^2 - q^2 B^2 r^2}$$

Accordingly, the power radiated by electrons moving within a synchrotron is contingent upon the magnetic field's radius and its strength. As such, an increase in either the magnetic field strength or the radius leads to a corresponding increase in the power emitted.

3. DISCUSSION

In this essay, we understand the relativistic effects on electron dynamics based on the data of synchrotron radiation facility of Shanghai Synchrotron Radiation Facility (SSRF). The relativistic effects is crucial for enhancing the functionality and efficiency of the electrons.

The rest mass energy of an electron is 0.51MeV, while the published energy for SSRF is 3.5MeV, indicating the substantially increased relativistic mass of the electrons compared to their rest mass, which is about 7000-fold.

According to the operational energy levels of SSRF, the Lorentz factor far exceeds unity ($\gamma=6.8\times10^3$), which impacts the electrons' velocity to approaching the speed of light, with β (v/c) nearing unity.

Understanding the velocity of relativistic electrons is fundamental for calculating the synchrotron radiation power. The power radiated by an electron moving in a synchrotron magnetic field is proportional to the square of the Lorentz factor and the strength of the magnetic field, as indicated by $P \propto \gamma^2 B^2$. It becomes apparent that P is influenced by both the magnitude of the magnetic field and the geometry of the radius. The greater the magnetic field strength and the radius, the greater the power emitted by the electrons in a synchrotron.

Synchrotron design is not just a matter of maximizing the magnetic field strength and the size of the ring to increase the power of the emitted radiation. Rational design must also account for cost, construction complexity, and operational efficiency. Hence, embracing innovation is key, such as considering the application of wigglers and undulators in fourth-generation synchrotron sources (Mills and Dejus 2023). These devices can enhance beam properties without necessarily scaling up the size of the synchrotron.

4. REFERENCES

1. Condon, James J., and Scott M. Ransom. Essential Radio Astronomy. SCH-School edition. Princeton University Press, 2016. https://doi.org/10.2307/j.ctv5vdcww.

2. Kolomenski, A. A., and A. N. Lebedev. "The effect of radiation on the motion of relativistic electrons in a synchrotron." (1956). https://cds.cern.ch/record/1241617/files/p447.pdf

3. Lim, K.F., and S.W. Lewis. “Spectroscopic Techniques.” Encyclopedia of Forensic Sciences, 2013, 627–34. https://doi.org/10.1016/b978-0-12-382165-2.00255-5.

4. Mills, D. M., and R. J. Dejus. “Synchrotron Sources.” International Tables for Crystallography, May 11, 2023. https://doi.org/10.1107/s1574870720007624.

5. What is synchrotron light? Accessed February 14, 2024. https://www.ansto.gov.au/education/nuclear-facts/what-is-synchrotron-light.

6. Wu, Xu, Shunqiang Tian, Xinzhong Liu, Wenzhi Zhang, and Zhentang Zhao. “Design and Commissioning of the New SSRF Storage Ring Lattice with Asymmetric Optics.” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 1025 (February 2022): 166098. https://doi.org/10.1016/j.nima.2021.166098.

Twin Paradox: A Review of Controversial Solutions

ZIAO ZHOU (Ziao)

1. Introduction

In the renowned special relativity paper Albert Einstein published in 1905, he raised an example of lagging clocks as a consequence of time dilation as follows: if two clocks started at the same position and synchronized, and then one clock is moved in a constant speed, away in one direction for some distance and back in the reversed direction, the travelling clock would be delayed compared to the unmoved clock when they meets up again in the initial position. This idea was later expanded in 1911 by Einstein and Paul Langevin and others, in which the space traveller was no longer a clock but a living organism.

Later in the decade, controversies rose in discussing the paradox through different approaches. Langevin, as one of the first to explain the paradox, used Doppler shift to illustrate the different aging rates and attribute the paradox to the asymmetry between the two twins due to change in velocity and acceleration being absolute. Max von Laue compared the world lines of the travelling twin and the Earth twin with Minkowski's space-time diagrams, displaying the effect of changing inertial reference frames have on asymmetrical ageing. However, soon the acceleration solution is rebutted in various ways and Lord Halsbury formulated a three-brother technique to eliminate the effect of acceleration, whereas many others proposed new solutions to the twin paradox.

This article will focus on two mainstream but controversial solutions that resolves the twin paradox to an asymmetry between the twins but in different approaches: the acceleration explanation, clock synchronisation, and the doppler effect. While many invoked general relativity in a great calculation complexity, this paper will demonstrate how special relativity alone is capable of resolving the paradox, but also providing an example of solution via general relativity. Additionally, some original responses to the solutions are provided.

2. The scenario: naïve application of time dilation

Zoe and Andrew are two twins on Earth. One day Zoe leaves the Earth and, in a spaceship, drives toward Alpha Centauri, which is L = 4 light years away in the Earth-star frame. The spaceship travels at speed . Once Zoe reached Alpha Centauri, she immediately turns around and heads back to Earth. The twins compared their clocks and found Zoe younger than Andrew.

In Andrew's perspective, Zoe's journey takes

$$\Delta t_A = \frac{2L}{v} = \frac{8\ ly}{0.8\ c} = 10\ years$$

Andrew will predict the time taken for Zoe with the special relativity time dilation formula, calculating Zoe aging 6 years.

$$\Delta t_Z = \frac{1}{\gamma}\Delta t_A = \Delta t_A\sqrt{1-(\frac{v^2}{c^2})} = 10 \times \sqrt{1-0.8^2} = 6\ years$$

However, the question lies that according to concepts of relativity, that all reference frames are equal, Zoe could as well consider herself at rest while thinking Andrew and the star is the one moving. In this case, because Zoe is in motion, the distance between Earth and Alpha Centauri will be contracted for her, measuring only 2.4 light years.

$$L' = \frac{1}{\gamma}L = L\sqrt{1-(\frac{v^2}{c^2})} = 4 \times \sqrt{1-0.8^2} = 2.4\ light\ years$$

Zoe would therefore experience a time of

$$\Delta t_Z{}' = \frac{2L'}{v} = \frac{4.8\ ly}{0.8\ c} = 6\ years$$

In Zoe's view, she sees the Andrew leaving her and coming back again at speed 0.8c and predicts the time of Andrew as dilated, as she reasons Andrew as the moving one in her frame, calculating his time as

$$\Delta t_A{}' = \frac{1}{\gamma}\Delta t_z{}' = \Delta t_z{}' \sqrt{1 - \left(\frac{v^2}{c^2}\right)} = 6 \times \sqrt{1 - 0.8^2} = 3.6\ years$$

Zoe would think Andrew is the younger one.

This generates two questions to be resolved, a) what accounts for the contradictory perceptions, and b) who, Zoe or Andrew, is truly younger. In order to explain the paradox fully, the solution should answer both two questions.

3. Acceleration

3.1 Asymmetry in the twin's views

The most apparent explanation is that Zoe has to experience an acceleration when her velocity is reversed and changed a direction. Since acceleration is absolute*, observers in any other frames including Andrew could not have this acceleration. An asymmetry lies between the situation of the two twins, causing Zoe to have a jump in time as she accelerates, while in Zoe's rest frame Andrew couldn't experience this jump. Furthermore, this acceleration pushes Zoe into a non-inertial reference frame in which special relativity can't apply as claimed.

3.2 Gravitational red shift

In theories of general relativity, there is a formula discovered by Einstein that shows gravitational fields can be accounted for changes in observed frequencies, known as the gravitational red shift. Reinterpretation of this phenomena would give a formula for the time dilation effects due to gravity, that states time runs more slowly in regions with lower gravitational potential, as shown below.

$$\Delta t_{high} = \Delta t_{low}\left(1 + \frac{gy}{c^2}\right)$$

According to the Equivalence Principle, Zoe's acceleration is indistinguishable from effects of gravity and can be made equivalent to a gravitational field centred on her, so Andrew obtains a higher gravitational potential than Zoe. Applying the gravitational red shift and the basic kinematic formula, equation (6) could be easily obtained.

$$\Delta t'_{acc} = \Delta t_{acc}\left(1 + \frac{aL}{c^2}\right) = \Delta t_{acc} + \frac{a\Delta t_{acc}L}{c^2} = \Delta t_{acc} + \frac{2vL}{c^2}$$

where is the time interval of Zoe's acceleration in Zoe's frame, and is the corresponding time interval in Andrew's frame. If acceleration is infinitely large, would be negligible compared to , so the total time passed for Andrew calculated by Zoe would be

$$\Delta t_A = \frac{1}{\gamma}\frac{2L'}{v} + \frac{2vL}{c^2} = 3.6 + 2(0.8)(4) = 10\ years$$

which matches with Andrew's perception.

However, note that this approximation can only be made when it is assumed that Zoe takes an infinitesimal time to reverse her velocity, leading to . As stated by R.C. Tolman in his solution, "the treatment of the problem without approximation would involve the full apparatus of the general theory of relativity."

3.3. Controversies

There are various opposing claims throughout literature against the solution of acceleration. Some argues that acceleration shouldn't change the effect of time dilation. (Feynman,) Special relativity could as well be applied, even if acceleration exists. As shown below the integral form of the time dilation formula, the effect is only dependent of the relative velocity of the frame, which can be expressed in terms of t if the object is accelerated and the formula is still valid. Additionally, Richard Feynman in his Lectures of Physics mentioned an example of a decaying particle. The lifespan of a particle forced into centripetal motion by magnets will be extended due to relativity to just the same extent as if it is in uniform straight motion, indicating that acceleration has no influence on change in time dilation calculations.

$$\Delta t' = \int \sqrt{1 - \frac{v(t)^2}{c^2}} dt$$

Other responses to this solution include claiming acceleration to be relative, that Zoe will also observe a turn and thus acceleration in Andrew. However, it is evident that when Zoe accelerates, she is experiencing a force. She feels the acceleration as her body inclines when she turns and items in her spaceship bang to the wall, while it is impossible for Andrew to feel such forces in another frame.

4. Doppler effect

The idea is comparatively less renowned, raised by Wolfgang Rindler in his book of Essential Relatvity, and formulates as follows. Suppose Andrew and Zoe both possess a transmitter constantly emitting light pulses at , one signal per second, in its rest frame. According to the relativistic Doppler Shift formula, during Zoe's outward journey, Andrew receives signals at frequency

$$f_{outward} = f_0 \sqrt{\frac{1 - v^2/c^2}{1 + v^2/c^2}} = \frac{1}{3}$$

And when Zoe is returning, Andrew receives signals at frequency

$$f_{return} = f_0 \sqrt{\frac{1 + v^2/c^2}{1 - v^2/c^2}} = 3$$

Time for the last signal emitted when Zoe is moving away takes 4 years to arrive on Earth, so during these years Andrew is still receiving signals at frequency f1, which makes a total of 9 years of f1 frequency signals. Let Y be the number of seconds in a year. The total number of signals received by Andrew is

$$9 * \frac{1}{3} * Y + (10 - 9) * 3 * Y = 6Y$$

Since he knows that Zoe possesses the same transmitter, Andrew presumes that she would only age 6 years.

On the other hand, in Zoe's perspective, she has been receiving light signals that move away from her for 3 years, before she turns the direction and starts receiving signals moving towards her, for another 3 years. Similarly, she will receive 10Y signals in total, and for the same reason concludes that Andrew aged 10 years.

$$3 * \frac{1}{3} * Y + 3 * 3 * Y = 10Y$$

Thus, Andrew and Zoe reached an agreement on Zoe being younger.

5. Discussion

Apart from the criticisms to the acceleration solution mentioned in 3.3, modifications to the paradox were made to also demonstrate the invalidity of the acceleration solution, such as by introducing a third brother. The author also came up with a modification. Three twins, one stays on Earth, while the two blasts off in opposite directions from the Earth, drives for the same distance (in Earth's frame) and comes back. Both travelling twins could experience the same acceleration if their spaceships are completely identical, so ideally no difference should there be between the two brothers, but the paradox can still exist as the two brothers each think themselves as stationery. This can show that acceleration is not the only cause for the asymmetry.

For the Doppler Effect explanation, the author of this modest article believes that it does not fully solve the paradox on its own, but rather only displays the story in another way. This formulation of the paradox pre-assumes that Andrew is stationary. In Zoe's reasoning, as Andrew is moving in her view, she calculates with her knowledge on relativistic doppler effect and predicts the amounts of signals she would receive from Andrew as

$$(2.4 + 3) * \frac{1}{3} * Y + 0.6 * 3 * Y = 3.6Y$$

(Since the last moving-away signal will need L'/c = 2.4 ly / c = 2.4 years to reach Zoe)

And therefore expect a 3.6-year duration for him, while in reality she receives 10Y signals. This could be taken further by completely reversing the situation, assuming both situations are symmetrical, Zoe is at rest and Andrew and Alpha Centauri is moving away and back at speed v. In this case, Zoe would think that Andrew, now the moving one, receives

$$1.8 * \frac{1}{3} * Y + 1.8 * 3 * Y = 6Y$$

Hence, in both cases Andrew and Zoe could have reached a consensus on who is younger; however, the solution does not explicitly show why Zoe's reasoning will not match with what she calculated and why one situation is preferred over the other. Thus, the Doppler Shift approach should only be considered as a supplementary material for the already existing claims on why the situations are not symmetrical.

6. Conclusion

Solutions to twin paradox is controversial. The most widely discussed explanation being asymmetry caused by acceleration experienced solely by the travelling twin, relying on the premise that acceleration is absolute in reference frames. Some also used general theories of relativity and Einstein's Equivalence Principle to further elaborate the effect of acceleration. However, these face criticisms that proved the irrelevance of acceleration and time dilation and by modifying the twin paradox to eliminate the asymmetry. Other approaches include the less popularised Doppler Shift resolution, but is considered to be dependent on analysis from other solutions.

7. References

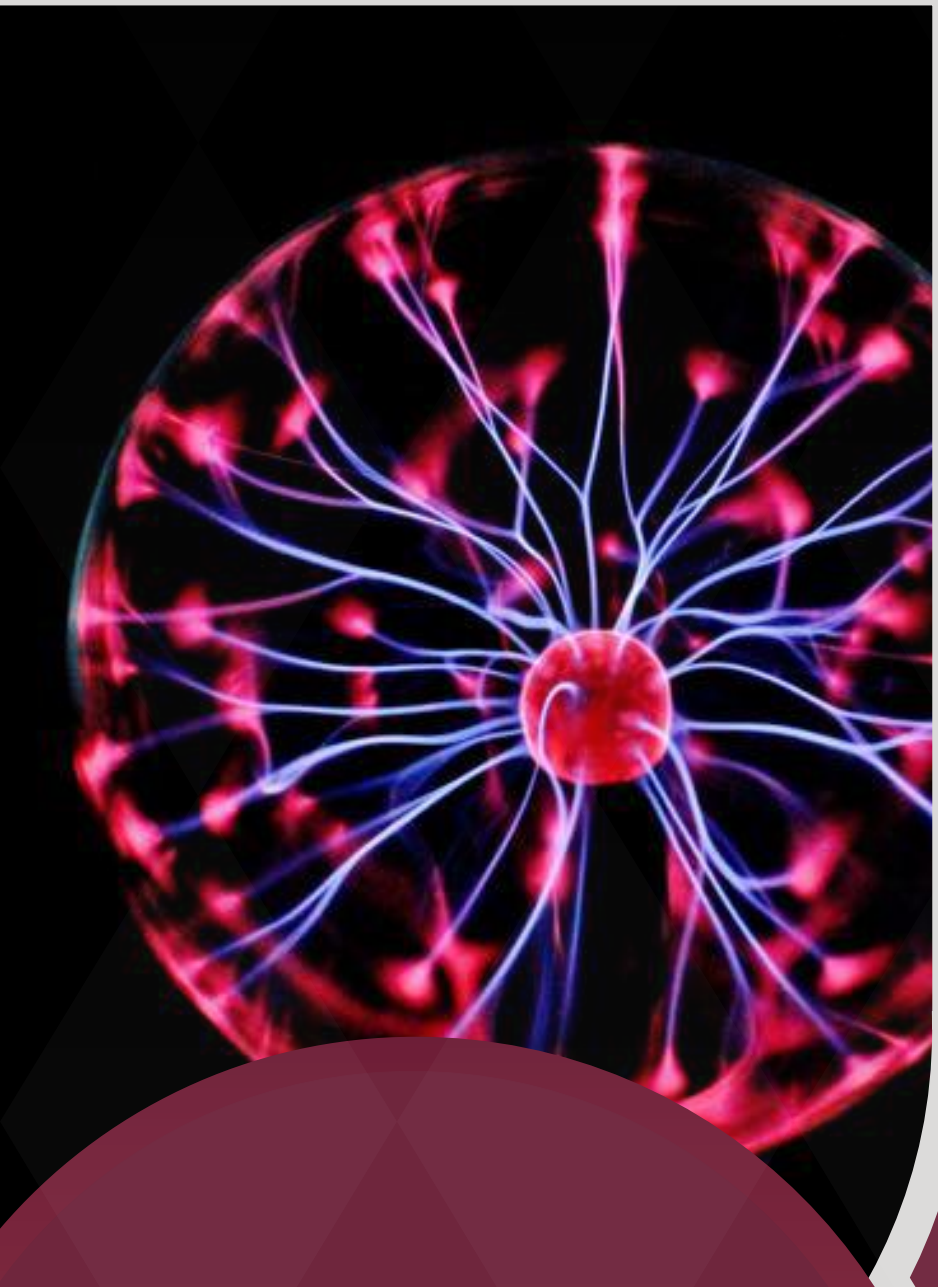
1. A. Einstein, "The Foundations of the General Theory of Relativity", translated from "Die Grundlage der allgemeinen Relativitätstheorie", Annalen der Physik, 49 (1916), in The Principle of Relativity, 111-164, Dover Publications (1952).

2. R.C.Tolman, Relativity, Thermodynamics and Cosmology, (Dover Publications, New York, 1987; reprint of 1934), pp. 194–197.

FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

FITZ
EVENTS
FITZWILLIAM COLLEGE
UNIVERSITY OF CAMBRIDGE

07



Psychology and Neuroscience

Fitzwilliam College Online Winter School Journal · 2024 Winter

Perceptions equal to reality? – Exploring sensory illusions

BEIXI CHEN (Bessy)

1. Introduction

As a Chinese proverb goes, “seeing is believing," which encapsulates that vision provides us with a reliable window to the external world. However, is everything we perceive always an accurate representation of reality? The phenomenon of sensory illusions challenges this notion, which is defined as perceptual distortions that result from the brain's inaccurate processing of sensory input (Gregory, 1997). Any of the human senses are capable of generating Ilusions, but it is visual illusions that a large amount of research focuses on because vision is the most dominant sense in humans (Stokes, 2014). Optical illusions occur when we are deceived by the visual system and are characterized by visual perceptions that differ from reality. Richard Gregory categorized optical illusions into physical, physiological, and cognitive illusions, which are distortionary perceptions caused by the physical environment, visual pathway and eye structure, unconscious awareness respectively (Gregory, 1997). Among these three classifications, physiological and cognitive illusions are invaluable tools for investigating neural mechanisms under visual perceptions. Using neuroimaging, specific patterns of neural activity are elicited, and regions and circuits involved in visual illusions are revealed (Eagleman, 2001). This means that visual illusions are a meaningful aspect of neuroscience. In this essay, I will discuss one classical type of physiological and cognitive illusion, respectively, further explain an example of a cross-modal illusion, and elaborate on what these illusions tell us about how our brain represents the world.

2. Main Body of Essay

2.1. Hermann Grid

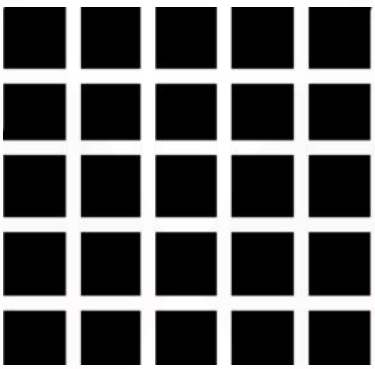


Figure 1. Classic Hermann grid

The Hermann grid is a typical example of a physiological illusion. When viewers stare at one of the intersections of this grid, gray shadows shaped like spots appear at other intersections. But if you look through one illusory spot directly, it will just disappear. Why would this happen? This could be explained by using neuroscience knowledge. One theory proposes that the illusion is a result of lateral inhibition. According to this theory, neurons in the visual system not only respond to stimuli but also inhibit neighboring neurons, enhancing contrast between adjacent regions of the visual field. In the Hermann grid, the intersections of high-contrast black background and white lines activate neurons, which then inhibit nearby neurons responsible for brightness perception. This inhibition reduces the perceived brightness of adjacent regions, enhances the contrast between black and white stimuli, and creates the illusion of dark blots at the intersections (Spillmann, 1994). However, there is evidence against this explanation by showing that Hermann grid illusions persist when lateral inhibition is experimentally disrupted (Rossi & Paradiso, 1999). The mechanism by which illusions are formed in the Hermann grid are still a matter of debate.

2.2. Rubin's vase

Figure 2. Rubin's vase

In terms of cognitive illusions, Rubin's vase is a well-known example. In this image, viewers perceive either a vase or two men facing each other. This face-vase illusion is also known as ambiguous illusions, which is defined as that the appearance of an illusion changes when viewers' visual attention shifts (Leopold, 2002). There are some cognitive mechanisms underlying this illusion. Firstly, when we are confronted with the ambiguous shapes in the image, the brain engages in a process of figure-ground separation, where it attempts to distinguish between the foreground and background elements (Koffka, 1935). Next, if the figure and ground are equally likely to be engaged in cognitive pattern matching, the brain "shapes" what it sees. In the case of Rubin's vase illusion, the central region of the image can be interpreted as either the foreground (the vase) or the background (the faces), leading to two distinct perceptual interpretations. What we see actually depends on which part of the image we are focusing on. The ambiguity between faces and vase arises because their contours in the image can be equally compatible with both interpretations (Gregory, 1972).

2.3. McGurk illusion

What's impressive is that illusions are not only generated by a single sense but also by combined senses. This is called cross-modal illusions, which refers to the mismatch between information from different modalities (Eagleman, 2001). One famous visual-auditory illusion named the McGurk effect demonstrates a hearing illusion when the auditory component of one sound is paired with the visual component corresponding to another sound, leading to the perception of a third sound (Nath & Beauchamp, 2012). When lip motions of the "ga" syllable are coordinated with the sound of "ba", listeners would usually identify the syllable as "da". This means that what we see significantly affects what we hear. The illusion results from a mismatch between the auditory information about speech sounds and the visual information about lip movements. Multisensory integration processes in the brain, which combine inputs from several sensory modalities to create a single, coherent perception, are thought to be the mechanism behind the McGurk illusion (Nath & Beauchamp, 2012). The brain integrates the contradictory information of lip movements and syllable sounds when they are presented simultaneously and tries to reconcile this difference, leading to sensory illusion.

3. Conclusion

In conclusion, the Hermann grid highlights the role of lateral inhibition, while Rubin's vase illustrates how different perceptions appear with different figure-ground combination. Additionally, the McGurk illusion showcases the brain's remarkable ability to integrate conflicting sensory information to form a unified percept. Together, these illusions remind us that perception is not a direct reflection of reality but rather a dynamic interpretation, shaped by various cognitive mechanisms. Studying these illusions provides valuable insights into the field of neuroscience and cognitive psychology.

7. References

1. Eagleman, D. M. (2001). Visual illusions and neurobiology. Nature Reviews Neuroscience, 2(12), 920-926.

2. Gregory, R. L. (1972). Cognitive contours. Nature, 238(5358), 51-52.

3. Gregory, R. L. (1997). Knowledge in perception and illusion. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352(1358), 1121-1127.

4. Koffka, K. (2013). Principles of Gestalt psychology (Vol. 44). Routledge.

5. Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. Nature neuroscience, 5(6), 605-609.

6. Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage, 59(1), 781-787.

7. Rossi, A. F., & Paradiso, M. A. (1999). Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. Journal of Neuroscience, 19(14), 6145-6156.

8. Spillmann, L. (1994). The Hermann grid illusion: a tool for studying human perceptive field organization. Perception, 23(6), 691-708.

9. Stokes, D. (Ed.). (2015). Perception and its modalities. Oxford University Press, USA.

The Multifunction of Memory Reveals the Unreliable, Constructive, and Adaptive Nature of Memory

QING QIN (Yoyo)

1. Introduction

False memory is defined as remembering details of events that have never occurred in the first place. Memory distortion is a common phenomenon that can be implanted under any circumstance. For example, a research team has demonstrated that 76% of the population has once experienced a Déjà vu – a false memory that provokes a sense of familiarity with what people are currently experiencing (Adachi, 2003). The widespread occurrence of memory illusion fascinates both researchers and the public. Thus, a lot of research on false memory is carried out to determine the mystery of memory formation. Currently, the study of the multifunction of memory is considered a crucial factor when investing in the nature of memory. The purpose of this essay is to illustrate how the research on false memory contributes to a more thorough understanding of the nature of memory.

2. The deceitful nature of memory

The deceitful nature of memory is revealed as our understanding of false memory progresses. The first experimental investigation of memory distortions carried out by Bartlett (1932) revealed the probability of memory illusions. His experiment demonstrated that repeatedly recalling events resulted in the formation of false memories. As time elapses, Deese's (1959) replicable paradigm adds more credibility to the entrenched unreliable nature of memory. He presented list of words and then asked for participants' free recall in a single trial. In the replication of this experiment, false identifications existed in 40% of the cases (Roediger, 1995). The remarkably high rate indicates the universality of memory inaccuracy. Researchers also discovered that a majority of critical non-presented words recalled by subjects have a semantic correlation with the known words. This is probably due to the associative response generated at the primary stage of memory – the encoding process (Underwood, 1965). The unconscious relation of studied items and critical cues of interrelated notions leads to the memory illusion. In addition, Loftus' (1996) psychological investigation on false childhood memory creation showed that imagination acts as another influencing factor as well. She demonstrated that the likelihood of the appearance of memory illusion increases by imaging counterfactual events. It reveals that false memories are implanted in the subconscious while experiencing robust interference from the external environment, which makes the memory less reliable. Therefore, the validity of memory is influenced by suggestive forces including association and imagination, suggesting the unreliability of memory.

3. Constructive and adaptive nature of memory

Memory is constructive in nature. Memory construction refers to the process by which memories are modified based on new experiences. On one hand, some researchers indicate that it discloses the inaccuracy of memory, whereas, on the other hand, memory construction is regarded as an implication of the adaptivity of memory. As Loftus' (1996) work revealed, illusions of memory can result from external manipulation and imagination. The research showed that memory was not only the retrieval of factual events stored in the mind but also a coherent reconstructing process based on new experience and association. Furthermore, the adaptive nature of memory was discovered as the study of memory reconstruction developed. It was exemplified in John Briere and Jon Conte's study (1993) that 59% of subjects, who suffered from sexual abuse in their childhood and received proper treatment before 18, have forgotten their experiences. The distortion shows that some traumatic memories are repressed by our bodies unconsciously. It could be conducted for protective purposes, since forgetting relieved people from recalling the overwhelming past. Therefore, we can conclude that memory is highly adaptive in that we can actively reconstruct information during retrieval.

4. Conclusion

In conclusion, the fact that people will falsely recall non-presented words in the experiments caused by semantic association and exterior misleading information (e.g. Imagination of counterfactual events) shows the unreliability of memory. Furthermore, the distortion of remembering attributed to the consistent reconstruction of memory based on accumulating experience indicates the constructive nature of memory. The adaptive nature of memory can be shown in the research on memory construction as well. Our memory can be reconstructed when experience exterior manipulation and irreversible trauma for protection purposes.

5. References

1. Adachi, N., Adachi, T., Kimura, M., Akanuma, N., Takekawa, Y., & Kato, M. (2003). Demographic and psychological features of déjà vu experiences in a nonclinical Japanese population. The Journal of nervous and mental disease, 191(4), 242-247.

2. Neuschatz, J. S., Lampinen, J. M., Toglia, M. P., Payne, D. G., & Cisneros, E. P. (2014). False memory research: History, theory, and applied implications. In Handbook Of Eyewitness Psychology 2 Volume Set (pp. 239-260). Routledge.

3. Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of experimental psychology: Learning, Memory, and Cognition, 21(4), 803.

4. Loftus, E. F. (1996). Memory distortion and false memory creation. PubMed, 24(3), 281–295. https://pubmed.ncbi.nlm.nih.gov/8889130

5. Ramirez, S., Liu, X., Lin, P., Suh, J., Pignatelli, M., Redondo, R. L., Ryan, T. J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. Science, 341(6144), 387–391. https://doi.org/10.1126/science.1239073

6. Pardilla-Delgado, E., & Payne, J. D. (2017). The Deese-Roediger-McDermott (DRM) Task: A Simple Cognitive Paradigm to Investigate False Memories in the Laboratory. Journal of visualized experiments : JoVE, (119), 54793. https://doi.org/10.3791/54793

7. Conway, M. A., & Howe, M. L. (2022). Memory construction: a brief and selective history. Memory, 30(1), 2–4. https://doi.org/10.1080/09658211.2021.1964795

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 Winter School
https://www.seedasdan.com/en/fitz/

阿思丹 | ASDAN
China