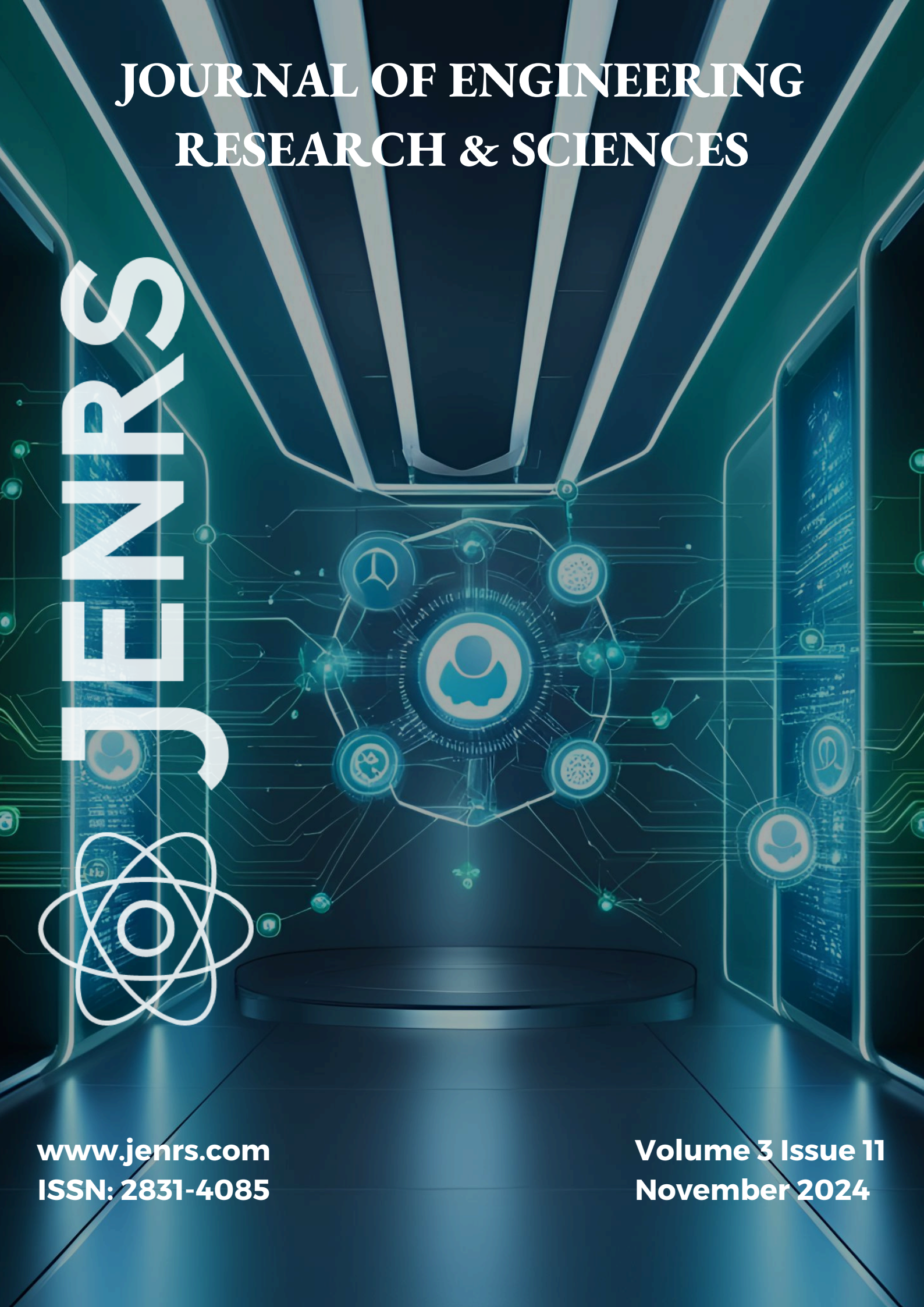
JOURNAL OF ENGINEERING RESEARCH & SCIENCES

JENRS



www.jenrs.com
ISSN: 2831-4085

Volume 3 Issue 11
November 2024

EDITORIAL BOARD

Editor-in-Chief

Editorial Board Members

Editorial

The continuing evolution of technology, from the historical development of computing systems to the implementation of modern intelligent solutions, reveals the multidisciplinary character of contemporary research. These papers collectively illustrate the critical intersections of security, transport, networking, and optimization in today's dynamic environments. Whether addressing road safety, data privacy, energy consumption, or network reliability, each study represents a unique contribution to the progression of engineering knowledge and its practical implications for society.

A detailed narrative on the evolution of computing technology is presented through the development of a dedicated exhibit at the University of Florence. Tracing the lineage from rudimentary calculating tools such as slide rules to the advent of personal computers, microprocessors, and electromechanical calculators, the study serves as a curated journey through the milestones of computer history. Unlike traditional museums, this exhibit is tailored for engineering students, offering them insight into the rapid pace and depth of technological advancement. The expanded material enriches the academic experience by contextualizing past innovations that form the foundation of present-day computing [1].

Secure and anonymous acknowledgment mechanisms in communication protocols are explored in a study that focuses on ad-hoc and Delay-Tolerant Networks (DTNs). Unlike conventional TCP acknowledgments, the proposed model uses cryptographically generated acks embedded within encrypted messages to ensure both message integrity and anonymity. These acks are distributed through peer-to-peer forwarding without revealing sender or receiver identities. An innovative feature includes the use of hashed message IDs, enabling nodes to manage cache memory efficiently by identifying delivered messages without compromising privacy. This system demonstrates how privacy-preserving communication can coexist with robust data delivery across intermittently connected networks [2].

Traffic classification under real-world conditions benefits from a fuzzy logic-based framework that interprets flow, occupancy, and speed data to determine road congestion levels. Applied to the freeway between Padua and Venice, the fuzzy approach is shown to outperform deterministic models in delivering more intuitive feedback to drivers. By embracing human-like reasoning and uncertainty, the system aligns with the way drivers perceive traffic, thus enhancing its practicality for real-time traffic management. The MATLAB implementation confirms the system's ability to improve classification accuracy and offers a new perspective for managing dynamic transport environments [3].

Ensuring sufficient pavement friction is vital for road safety, particularly in braking and turning scenarios. This study compares the performance of Calcined Bauxite an expensive but high-friction material with several cost-effective local alternatives such as Meramec River Aggregate, Flint Chat, and Steel Slag. Through laboratory tests, including dynamic friction, British Pendulum measurements, and aggregate imaging, the research identifies viable substitutes with comparable frictional properties. Key variables such as aggregate size and surface texture are linked to friction performance, offering a cost-efficient pathway for infrastructure development without compromising road safety [4].

The performance of interior gateway protocols (IGPs) in large-scale enterprise networks is critically analyzed through simulation using a tri-connected topology. The study evaluates protocols such as RIP, EIGRP, OSPF, and IS-IS across metrics like convergence time, delay, and jitter. Notably, EIGRP exhibits superior performance in delay and convergence, while IS-IS outpaces OSPF in convergence speed. This investigation enhances current understanding by offering a statistical computation framework for jitter analysis and optimizing protocol selection in complex network scenarios. The experimental setup, simulated in GNS3, provides a realistic foundation for future enterprise-scale deployments [5].

To address the persistent challenge of rear-end collisions due to unsafe following distances especially under adverse weather conditions—a novel Arduino-based intelligent driver-assistance system is developed. By integrating sensors for distance and rain detection, along with a computational algorithm for speed adjustment, the system dynamically adapts recommended speeds in real time. Trials conducted in various environments across Bahrain confirm the model's adaptability and effectiveness. This intelligent system not only enhances vehicular safety but also introduces an accessible, affordable solution for reducing accident rates on busy roads [6].

The pressing need for energy efficiency in wireless sensor networks (WSNs) is met by this bibliometric analysis focused on energy optimization algorithms (EOAs). Drawing data from the Web of Science and analyzing trends between 2019 and 2023, the study identifies key research themes and evaluates popular protocols such as PSO and LEACH. Using tools like VOSviewer, the analysis maps out collaborative networks, keyword occurrences, and co-citation patterns, shedding light on the depth and direction of research in this domain. The findings highlight promising avenues for improving WSN performance and call attention to underexplored optimization techniques for future exploration [7].

This collection reflects the dynamic and multifaceted nature of engineering research today. From preserving the legacy of early computing machines to engineering smart, safe, and efficient systems in the modern world, each study affirms the importance of bridging theoretical understanding with real-world applications. Together, they exemplify the enduring relevance of innovation, interdisciplinary thinking, and systemic resilience in shaping a technologically responsible future.

References:

[1] G. Bucci, I. Zaza, "An Educational Exhibit Aimed at Demonstrating the Rate of Growth of Computer Technology to Graduate Students," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 1–23, 2024, doi:10.55708/js0311001.

[2] E. Biagioni, "Secure Anonymous Acknowledgments in a Delay-Tolerant Network," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 24–30, 2024, doi:10.55708/js0311002.

[3] G. Erdinc, C. Colombaroni, G. Fusco, "Fuzzy-Based Approach for Classifying Road Traffic Conditions: A Case Study on the Padua-Venice Motorway," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 31–40, 2024, doi:10.55708/js0311003.

[4] E. Deef-Allah, K. Broaddus, M. Abdelrahman, "Comparing the Performance of Recycled Calcined Bauxite vs. Locally Available Aggregate as Components in High Friction Surface Treatment Applications," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 41–59, 2024, doi:10.55708/js0311004.

[5] R. Taha, E. Awadh Ben Srity, A.T. Abu Raas, "A Comparative Analysis of Interior Gateway Protocols in Large-Scale Enterprise Topologies," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 60–73, 2024, doi:10.55708/js0311005.

[6] Y. Rabie Farag, B. Jasson, "Smart Vehicle Safety System Using Arduino: An Experimental Study in Bahrain's Driving Conditions," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 74–80, 2024, doi:10.55708/js0311006.

[7] M. Idah Masopoga, M. Sumbwanyambe, Z. Wang, N. Reginald Netshikweta, "A Bibliometric Analysis Review on Energy Optimisation while Designing Wireless Sensor Networks," *Journal of Engineering Research and Sciences*, vol. 3, no. 11, pp. 24–30, 2024, doi:10.55708/js0311007.

Editor-in-chief

Prof. Paul Andrew

JOURNAL OF ENGINEERING RESEARCH AND SCIENCES

Volume 3 Issue 11 **November 2024**

CONTENTS

An Educational Exhibit Aimed at Demonstrating the Rate of Growth of Computer Technology to Graduate Students 01
Giacomo Bucci and Imad Zaza

Secure Anonymous Acknowledgments in a Delay-Tolerant Network 24
Edoardo Biagioni

Fuzzy-Based Approach for Classifying Road Traffic Conditions: A Case Study on the Padua-Venice Motorway 31
Gizem Erdinc, Chiara Colombaroni and Gaetano Fusco

Comparing the Performance of Recycled Calcined Bauxite vs. Locally Available Aggregate as Components in High Friction Surface Treatment Applications 41
Eslam Deef-Allah, Korrenn Broaddus and Magdy Abdelrahman

A Comparative Analysis of Interior Gateway Protocols in Large-Scale Enterprise Topologies 60
Saleh Hussein Al-Awami, Emad Awadh Ben Srity and Ali Tahir Abu Raas

Smart Vehicle Safety System Using Arduino: An Experimental Study in Bahrain's Driving Conditions 74
Youmna Rabie Farag and Bintu Jasson

A Bibliometric Analysis Review on Energy Optimisation while Designing Wireless Sensor Networks 81
Makwena Idah Masopoga, Mbuyu Sumbwanyambe, Zenghui Wang and Nthambeleni Reginald Netshikweta

JENRS

Received: 19 June 2024, Revised: 02 September 2024, Accepted: 06 September 2024, Online: 04 November 2024

DOI: https://dx.doi.org/10.55708/js0311001

An Educational Exhibit Aimed at Demonstrating the Rate of Growth of Computer Technology to Graduate Students

Giacomo Bucci*, Imad Zaza

Dipartimento di Ingegneria dell'Informazione (DINFO), University of Florence, Italy

*Corresponding author: Giacomo Bucci, University of Florence, Italy & Email: giacomo.bucci@unifi.it

ABSTRACT: This paper is an extended version of that presented at the conference Histelcon 2021 (IEEE). It provides a deeper illustration of the elements of the exhibit under development at the Faculty of Engineering at the University of Florence (Italy). The paper presented at Histelcon 2021 focussed on the birth of microprocessors, on the 8086, on an example of core memory and punched tape. In the following, though these arguments are still dealt with, much material has been added, including electromechanical calculators, medium-sized computer systems, personal computers, 8-bit microprocessors, disks, and other elements on the exhibit. The organization of the paper is quite different. It starts with a description of old computing machines (including a slide rule) before passing to the illustration of some more recent computer systems. The last part of the paper is dedicated to the illustration of the mentioned elements. Though there are many computer museums around the world, usually open to the general public, this exhibit is directed to the students of the Faculty, to enlighten the pace of development in computers.

KEYWORDS Digital museums, digital heritage, computer systems, microprocessors, computer devices, technological progress.

1. Introduction

At the School of Engineering of the University of Florence (Italy), a sort of museum of Computers is being set up. Transliterating the Italian name, the exhibit would be named "An exposition path through technology", just to point out the pace of technology growth. The aim is to make it clear to the students, essentially young people, how the rate of improvement has been impressive. While there are many computer museums [1] for the general public, this exhibit is directed to the students of the Faculty.

The exhibit is made up of some display cabinets where the various pieces are exposed. These include microprocessors, memories, disks, personal computers, both desktop and laptop, as well as some (parts of) systems of medium and low size. An example of a display case is in Figure 1. Most of these objects come from the same School of Engineering in Florence, where, at the end of their usage period, they were declared down and made available. Some materials come from private individuals.

Microprocessors are by far the largest class of devices. Unfortunately, the first microprocessor in history, the Intel 4004 [2, 3], is not available at the moment; on the other hand, there is the 8008, a 8-bit microprocessor, Intel's second micro. Though the largest class of microprocessors belongs to the X86 family, there are also devices from other manufacturers (AMD, Motorola, Zilog, and others). Several types of memory are on display, including a magnetic core memory that is presumably part of an Olivetti Elea computer. The auxiliary storage devices section is also well stocked, including an interchangeable plate of a 14" hard disk drive, plus numerous disk drives of different generations. Among the devices on display, there is no shortage of complete systems or parts of systems. The collection includes a variety of personal computers: from the first 8-bit devices (ZX Spectrum, Olivetti M10, and others) to modern laptops.

As a complement to the exhibit, a document (in Italian, about 150 pages) [4], downloadable from the Internet, is provided; it contains a more detailed exposition of the exposed elements. It has a chapter entitled "How a laptop can be disassembled and re-assembled". The document is regularly updated when new pieces are added. It aims to be an instructional aid for the students of Information Technology, Electronics, and Telecommunications. We hope that this document can be useful to those who, not being specialists in the subject matter, will find information in it that is not always easily available. Of course, the document is downloadable from the Internet, at the address `https://sites.google.com/unifi.it/calcolatori-elettronici`.

The objects described in the rest of this article are a subset of those on display. We illustrate the specimens of the greatest historical value and/or the elements that have had an important technological impact. The description accompanying them is essentially the same as that contained in the explanatory leaflet [4], though certain general considerations have been omitted. Part of the following has been already presented in [5].

2. Mechanical, electromechanical and electronic calculators

The exhibit contains some old calculating machines. In the following, we begin with a slide rule, then a couple of electromechanical machines are described, plus a couple of electronic calculators.

Figure 1: Example of a display case.

2.1. Slide rule

Until the advent of pocket electronic calculators, the slide rule was the calculation tool of choice for the technical-scientific community. In its pocket form, it often peeked out from the breast pocket of engineers' jackets. The ruler allows, with a few maneuvers, calculation of multiplications, divisions, squares, cubes, exponentials, and some trigonometric operations. It is based on the properties of logarithms. Since the logarithm of the product is equal to the sum of the logarithms of the two factors [log(xy)=log(x)+log(y)], it follows that if the numbers are represented in a logarithmic scale their product is obtained by positioning in a sum fashion the two logarithmic scales on which the numbers are represented. To this end, the slide rule is made up of three elements: (i) a body on which there are fixed scales; (ii) a sliding ruler, inside the body, also provided with numerical scales; (iii) a cursor with one or more reference lines.

The image of Figure 2 shows the rule with the sliding rod positioned so that the 1 of its upper scale (scale B) corresponds to the 1.5 of the adjacent fixed scale (scale A) . The result of the products, $1.5 \times 2 = 3$ and $1.5 \times 3 = 4.5$, can be read on the A scale in correspondence with the 2 and 3 on the B scale.

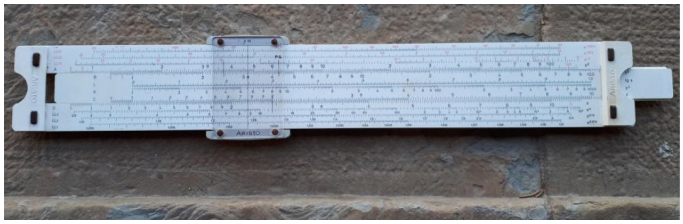


Figure 2: Example of a slide rule.

In the close-up image of Figure 3, the cursor is positioned on π of scale B, in order to facilitate reading the result of the product $1.5\pi = 4.7$ on scale A. Obviously, all results are to be considered approximate unless it is obvious that we are dealing with products that give an exact result (as in the preceding examples).

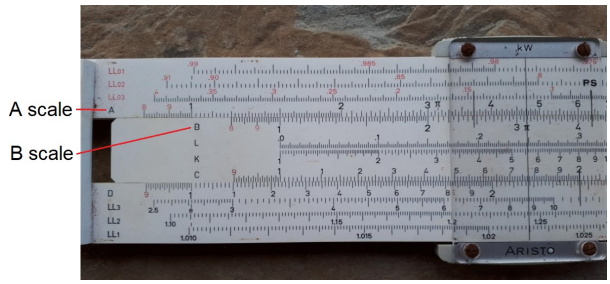


Figure 3: Close view of the ruler, set up for the multiplication $1,5 \times \pi$.

The first slide rule is attributed to the Englishman William Ougthred (1575-1660), who built a model composed of two rulers in 1622, based on the studies of John Napier (1550-1617).

Starting from 1970, with the advent of affordable pocket electronic calculators, the slide rule was abandoned by the technical-scientific community. Not only because it provides approximate results, but also because it does not allow addition and subtraction operations, which any trivial calculator does. The German city of Darmstadt had a centuries-old, thriving slide-rule manufacturing industry. With the arrival of pocket computers, this industry disappeared in a few years. Production of the slide rule ended around 1978.

2.2. Monroe LA7-200 calculating machine

In 1820 the Frenchman Charles Xavier Thomas de Colmar built the first completely mechanical calculating machine, called Arithmomètre. Unlike the prototypes of Pascal and Leibniz a century before, it was a commercial product. Towards the end of the 19th century, other manufacturers entered the market. The Monroe Calculating Machine Company was founded in New York in 1912 by Jay R. Monroe.

Figure 4: The Monroe LA7-200 calculating machine.

The Model LA7-200 is an electromechanical calculator, capable of performing the four fundamental arithmetic operations (addition, subtraction, multiplication, and division). It was produced by the Dutch branch of Monroe in 1950. You can set numbers of (up to) 10 digits on the keypad. Internally there are 10 counters and 20 accumulator registers. The results appear in the cart (two distinct lines depending on the operation). The trolley crank is used to reset it. The machine weighs 7.6 kg. A video on its functioning can be found at the address https://www.youtube.com/watch?v=HFGcv7a8l5M.

2.3. The Olivetti Tetractys calculating machine

This electromechanical computer was introduced in 1956 together with the Divisumma 24, from which it was derived, with some extensions, including a double totalizer. It performed the basic arithmetic operations. The keyboard had 10 digits in addition to the keys for the calculation functions; the printing was on a roll of scrolling tape. Tetractys represented the state of the art of mechanical calculation of those years. The two totalizers allowed moving from one operation to the next while preserving the products and quotients of the previous calculation.

The dimensions are 24; 42; and 24.5 (W; D; H) centimeters. It weighs 15kg.

Figure 5: The Tetractys calculating machine (Olivetti).

2.4. Casio Digital Diary SF-4300A

The Casio SF-4300A is a rather rudimentary device (Figure 6), with a few more features of a calculating machine. The SF-4300A was marketed started in the early 1990s. It had one memory of 32 kB and could perform the following functions: (corresponding to the same number of buttons located to the right of the display): phone book; storing texts and other data (memos); to-do list (schedule); calendar; clock (local and world); alarm (reminder); calculation.

2.5. Olivetti 600 Solar

The Olivetti 600 Solar (Figure 7) was introduced in 1989 (approximately). It is a small pocket electronic computer (125x82x15mm), whose main feature is that it is equipped with solar cells which allow its use in practically any light condition. The display is liquid crystal.

Electronic calculators equipped with solar cells began to appear in the late 1970s. But they have not achieved great success, probably because the always lowering consumption of electric devices does not make the presence of solar cells important. Similar calculators are currently on sale at a cost of around ten dollars. However, there is a more reason that makes use of electronic calculators unattractive: any mobile phone has at least one App that carries out the operations of such calculators.

Figure 6: The Casio Digital Diary

Figure 7: The Olivetti 600 Solar

3. Portable computers

There are almost thirty portable computers on display. Here we present a selection: the Sinclair ZX Spectrum, the Commodore 64, the Olivetti M10, the Toshiba T1600, the Compaq Contura Aero 4/33c, the Apple PowerBook G4,

3.1. ZX Spectrum

The ZX Spectrum was a family of small portable computers ($23.3 \times 14.4 \times 3$ cm), introduced by Sinclair in 1982 and produced until 1986 by the same manufacturer, and, from 1986 to 1992, by Amstrad.

The ZX was based on the Z80 microprocessor. In the initial version at 3.5 MHz, with a RAM of 16 kB, was managed through a sort of operating system called Sinclair Basic, essentially a version of Basic to which some commands such as RUN for starting programs had been added. The version on display is the ZX 48K model, i.e. with 48 kB of RAM. The ZX has been very succesful: 5 million of the ZX Spectrum have been sold worldwide, not counting imitations.

Although the ZX Spectrum is no longer produced, small cards are on sale which allow the implementation of a ZX Spectrum, and other various small computers in the same category as the ZX. Figure 8 shows one such small card the ZX-Uno (V4.1). The card size is 86 by 56 millimeters. It is preferentially oriented towards the creation of the ZX. But it also allows the construction of the Apple II, the VIC-20

(Commodore), the Acorn Atom, the Atari 2600, and many others. Anyway, this requires an appropriate setting of an FPGA (*Field Programmable Gate Array*), which can be defined via VHDL or Verilog languages.

The top image of Figure 9 shows the board without components; the image in the middle the placement of the connectors, while the lower the mounted components. Basically, anyone who wants to build one ZX Spectrum just needs to assemble the required components, connect a keyboard, a television (depending on the video), and connect the power supply (via USB).

Figure 8: The ZX Spectrum

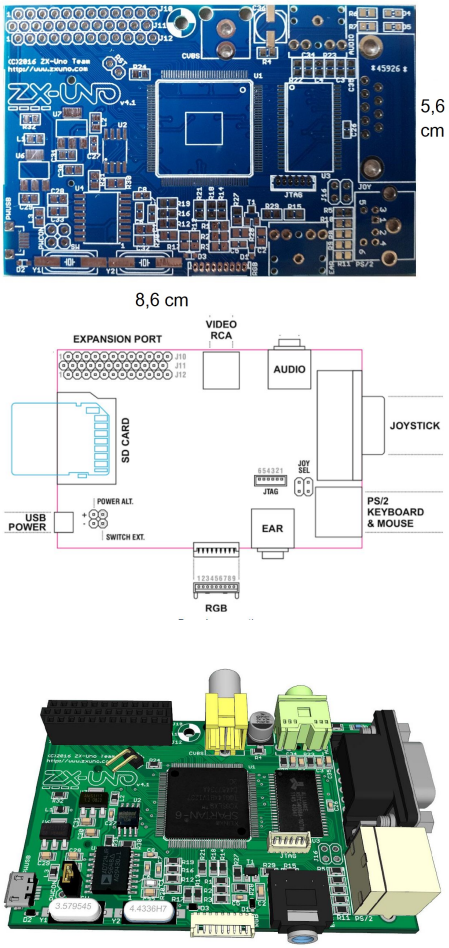


Figure 9: Above the ZX-Uno V 4.1 card, note its dimensions; center schematization of the interfaces; below an example of implementation.

3.2. *Commodore 64*

Produced by Commodore Business Machines, from 1982 to 1994, the Commodore 64 (C64, 64 indicates 64 kB of memory, Figure 10) was a very successful product. 17 million units were sold, surpassing the sales numbers of the Apple II and other personal computers very popular at the time (including, in Europe, the ZX Spectrum).

Figure 10: The Commodore 64.

The C64 was the first truly mass-market PC, thanks to its affordable price and high performance. It was marketed with an aggressive pricing policy: it was sold in department stores and even toy stores. The C64 looks like a headboard, quite massive by today's standards, in Italy was called "Biscottone" (big cookie). On the back, there were connectors for audio cassettes, serial or parallel printers, and TV sets (as video output).

The processor was an MOS 6510 from MOS Technology, operating at a frequency of approximately 1000 Hz to adapt to both the American and European television standards (1023 Hz for NTSC, 985 Hz for the PAL).

The 6510 was a 40-pin DIP (*Dual In-line Package*) IC. It was a derivative of the 6501, an 8-bit microprocessor that appeared in 1975 and was designed by a couple of engineers who had participated in the development of the Motorola 6800. The 65xx family processors were produced by several electronics industries and had considerable success. The 6501 had a bus compatible with that of the 6800, but a simplified internal architecture and instruction set. At the time the 6501 was introduced, while the Motorola 6800 was selling for $300, the 6501 was selling for $25 – consequently, Motorola had to lower the price to $35. The successor to the 6501, the 6502, was widely used, for instance, in the Apple II.

From 1985 to 1989 the Commodore C128 was produced (128 kB of memory and 8502 microprocessor, a higher performance version of the 6510), but it was not so successful: the era of 16-bit microprocessors had begun.

3.3. *Olivetti M10*

The Olivetti M10 (Figure 11) was introduced in 1983 and withdrawn in 1985. It employed an 80C85 CPU at 2.4 Mhz, CMOS version of Intel's then widespread 8085.

The LCD had 8 lines of 40 characters (240 x 64 pixels, black and white); the keyboard with 94 keys it had the QZERTY layout (Italian layout used by Olivetti on its typewriters). It used a proprietary operating system from Mi-

crosoft, equipped with Basic, a word processor, an agenda, an address book, and Multiplan (Microsoft's first spreadsheet before Excel), all on ROM. It is said that Bill Gates participated in the development of the programs and that this was the last time he directly programmed.

Figure 11: The Olivetti M10.

The RAM was static (non-volatile) 24 kB, and served as mass storage, but to be able to run Basic it was necessary to expand it to 32 kB. The M10 featured an RS232C serial port 25-pin, a 26-pin Centronics parallel port, and a connector for an external cassette recorder. Consumption was low enough to be powered by 4 normal 1.5V AA batteries. It cost from a million Lire to around 2.4 million Lire (today 1 million Lire corresponds to about €500) for the most equipped version.

It had the dimensions of 30x21x5 cm and weighed 1.7 kg. At the time, "lightweight" laptops weighed just under 10kg, were extremely bulky, and had to be attached to the power socket. These factors, together with the captivating aesthetics, determined its immediate success. In numerical terms, sales in 1984 amounted to approximately 24,000 pieces; on the Italian market, the M10 reached 70% of laptop sales and 22% on the European market. It went out of production the following year when sales dropped dramatically. It is interesting to note that the M10 was the Olivetti version of the Kyotronic KC-85, a portable introduced in 1983 by the Japanese Kyocera. Probably the first real laptop. Three industries: Olivetti, Tandy, and NEC, had obtained the license to produce their own version of the Japanese laptop; Tandy operated on the American market, NEC on the Asian, and Olivetti in Italy and Europe.

3.4. Toshiba T1600

The Toshiba T1600 (Figure 12) was introduced in 1987. it was based on the 80C286 processor at 12 MHz. Memory was expandable to 5 MB. It could mount a hard disk of either 20 or 40 MB, plus a 3.5" flexible disk with 1440/720 kB capacity. It had an EGA standard video (16 colors from a palette of 64 and a resolution up to 640×350 pixels. MS-DOS was the operating system, but it was possible to install Windows 3.0.

Its weight was 5.2 kg (with only one battery installed). Suggested retail price was $5,000.

Figure 13 shows the components standing under the keyboard. The keyboard was easily removable and was connected via a flat cable to the indicated connector; a connector was also provided for mounting a possible memory expansion.

Figure 12: The Toshiba T1600.

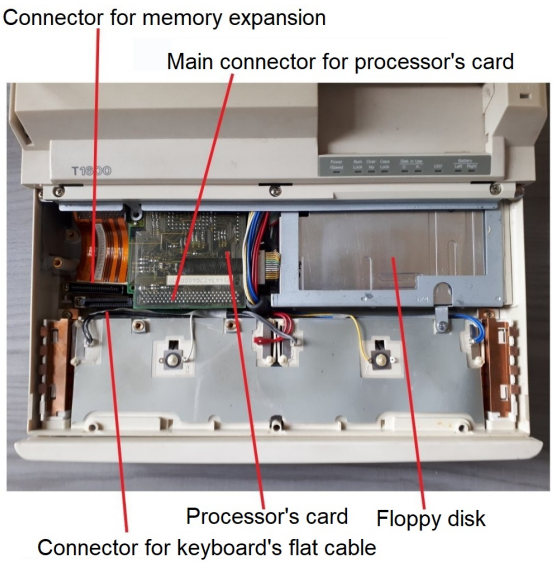


Figure 13: The T1600 after the keyboard has been removed.

The microprocessor was mounted on the board which appears upside down in Figure 14. In turn, this card was mounted via two connectors on the system board (the second connector is seen from the soldering side in Figure 14); on this board, there was an empty socket for the insertion of the 80287 coprocessor.

A comment from back then: *The T1600 is a truly portable machine; the enlarged backlit Supertwist EGA display is a marvel of readability.*

3.5. Portatile Compaq Contura Aero 4/33c

The Compaq Aero line was produced in various versions, with different processors, disks, and memories from the second half of the 1980s until around the mid-1990s.

The model on display (Figure 15) has these characteristics: Intel 486SX processor at 33-MHz, with 8 KB of integrated cache; 8" monochrome display; 4 MB RAM, expand-

able to 20 MB; 171 MB hard drive (Seagate ST9190AG). Size was 19.5 x 26.7 x 4.3 cm, while it weighed less than 2 kg.

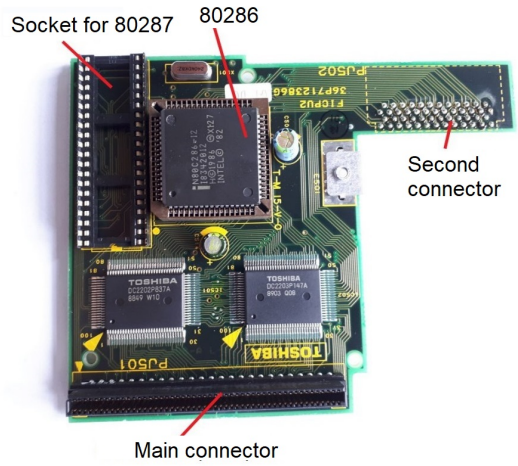


Figure 14: The microprocessor's card of the T1600.

Figure 15: The Contura Aero 4/33c of Compaq.

Furthermore, this laptop was equipped with a PCMCIA card that allowed the connection of an external 3.5-inch, 1.44 MB floppy disk. In subsequent models, the PCMCIA interface was replaced by a USB interface, the newborn industrial standard. The characteristic of the Compaq Aero was the trackball that can be seen on the right under the keyboard, together with the two buttons, barely visible in the photo on the right side (opposite the shift key). The trackball and the two buttons acted as the mouse.

3.5.1. *The motherboard of the Compaq Contura Aero 4/33c*

Figure 16 shows the motherboard from the side that, when mounted, looked downwards; Figure 17 the other side.

On the side facing downwards, there is the processor, the system memory, and a series of ASICs (*Application Specific Integrated Circuits*) devices, used by the manufacturer to implement accessory functions. The memory expansion was piggyback-mounted on the motherboard, it used eight 1M by 4 bits Hitachi HM514400B/BL ICs, for a total of 4MB. There was also a connector for dock expansion.

On the side facing upwards is the slot for the PCMCIA card.

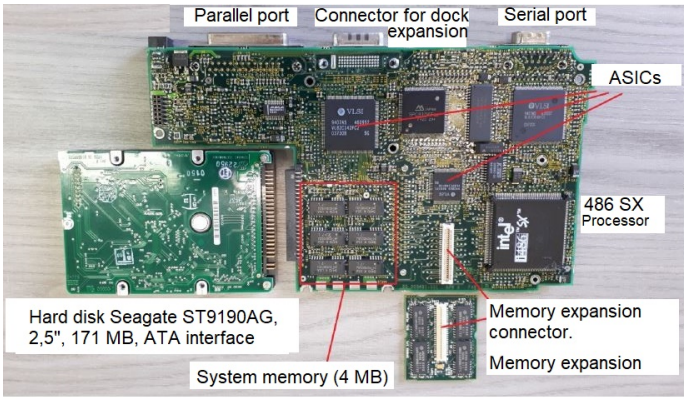


Figure 16: The Contura Aero 4/33c motherboard seen from its back side

Figure 17: The other side of the card of Figure 16. Notice the PCMCIA card slot

3.6. *Apple PowerBook G4*

The PowerBook G4 family was produced from 2001 to 2006. The family included a large number of performance models starting from the one produced in 2001. The acronym G4 indicates that the family uses PowerPC G4 processors. The one on display is the A1010 model introduced in January 2005.

The microprocessor was a PowerPC G4 (mod 7447) at 1.33 GHz, with 256 MB of RAM, and 256 kB L2 Cache. The hard disk had a capacity of 60 GB. The video was a 12" with 1024x768 pixels. Dimensions 3x27,7x21,8 cm; weight was 2.1 kg.

The PowerBook G4s were the last Apple laptops to use PowerPC processors (see the discussion at the last paragraph of section 4.3).

Figure 18: The PowerBook G4

4. Desktop computers

In the following some desktop computers, out of those in the exhibition are described.

4.1. Olivetti M24 New P100

Olivetti's M24 line was introduced in 1983. Olivetti had previously produced the M20, based on a Z8000 processor and with its operating system. The M20 did not hit the mainstream market, which then, as now, was oriented towards PCs compatible with the original IBM PC. For this reason, Olivetti switched to the production of the compatible M24. However, unlike the latter, which used the micro 8088 at 4.7 MHz, the M24 was based on an 8086 at 8 or 10 MHz.

Figure 19: The Olivetti M24 desktop computer.

The Computer on display (Figure 19) is a much more recent version of the initial one and is based on a 100 MHz Pentium. This PC is contained in a 36x41x11 cm box. Figure 20 shows (part of) the motherboard.

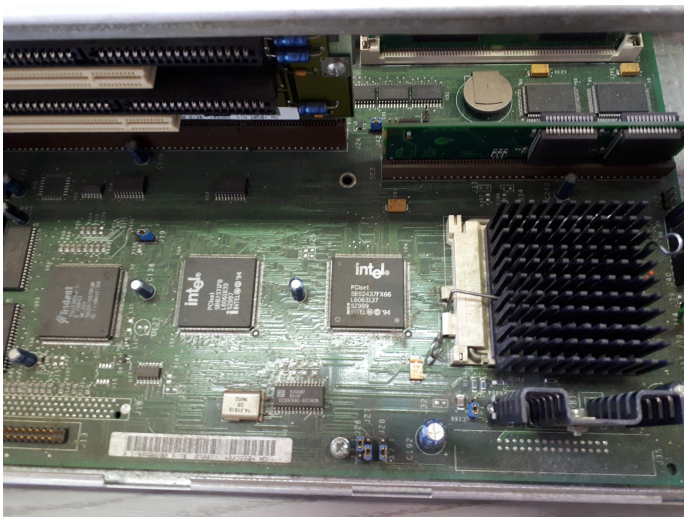


Figure 20: Details of the M24 new P100 motherboard. The processor is located under the cooling winglets mounted on it. The two Intel ICs act as the *north* and *south bridge*.

4.2. Apple MacIntosh 512

This computer was introduced in September 1984 and produced until April 1986. It represented the first update to the original MacIntosh 128 introduced in early 1984.

At the time of launch, its price was $2,795. It looked as in Figure 21. It had a driver for 3.5 diskettes that had to be inserted into the opening under the screen; no hard drive was provided. The video was 9" B&W, with a resolution of 512×342 pixels. The measurements were 35x24x28 (HLD); it weighed 7.5 kg. At introduction, version 1.1 of the MacOS operating system was pre-installed; Before its retirement, the Mac 512 had reached version 4.1 of MacOS.

Figure 21: The MacIntosh 512.

4.2.1. Motherboard of the MacIntosh 512

The motherboard, shown in Figure 22, was mounted horizontally under the floppy disk driver. The 4 connectors (female) that can be seen in Figure 22 protruded on the back. Orderly, from left to right, they represent these 4 doors: mouse port (9 pin), proprietary parallel port (19 pins) for a possible external floppy disk, serial port 422 (9 pins) for a printer, serial port 422 (9 pins) for a modem.

These were the main characteristics of the motherboard: the processor was a Motorola 68000 (in this specific case it is a Hitachi, produced under Motorola license), at 8 MHz (to be precise 7.8336 MHz); it had 512 kB memory, consisting of 16 integrated circuits of 256 Kx1bit, plus two integrated ones for parity control; ROM memory was 64 kB (two chips). In addition, there was a Z8530 serial port controller (produced by AMD under a Zilog license), and a VL6522 Parallel Port/Timer Controller (VLSI). Dimensions were 22x27 cm.

Between the serial port controller and the ROM, there is a custom VLSI integrated circuit. All the integrated circuits are soldered directly onto the board, except the two ROM integrated circuits, which are instead mounted on sockets. They contain (part of) the operating system. Evidently, there was a prediction that the content of the ROM could be subject to evolution.

4.3. Apple Power Macintosh G3 M3979

This computer was introduced in November 1997 and produced until April 1998 (replaced by updated versions). At the time of introduction, its price was $2,400. Figure 23 shows the computer box with its motherboard on top of it.

The following were the main features of the Power Macintosh G3 M3979. A PowerPC 750 microprocessor (page 20), operating at 233 MHz, mounted on a ZIF (*Zero Insertion Force*) socket; standard RAM 32 MB, expandable up to 192 MB; L1 Cache 64KB; L2 Cache 512KB (on 117 MHz bus); bus speed 66 MHz; cache bus speed 117 MHz; ROM/Firmware size 4 MB; Standard VRAM 2 MB, expandable up to 6 MB; 1024x768 pixel color display; 3.5" hard disk, 4.0 GB capacity, IDE/ATA-2 interface optical disk; floppy disk. It had also

a SCSI interface for external and internal additional disks. There were a 10Base-T Ethernet port and 3 PCI (*Peripheral Component Interface*) slots at 33 MHz. The operating system was MacOS 8.0. Motherboard dimensions: 28x20.5 cm.; dimensions of the case: 6.3 x 14.4 x 16.9 cm; weight approx. 10 kg.

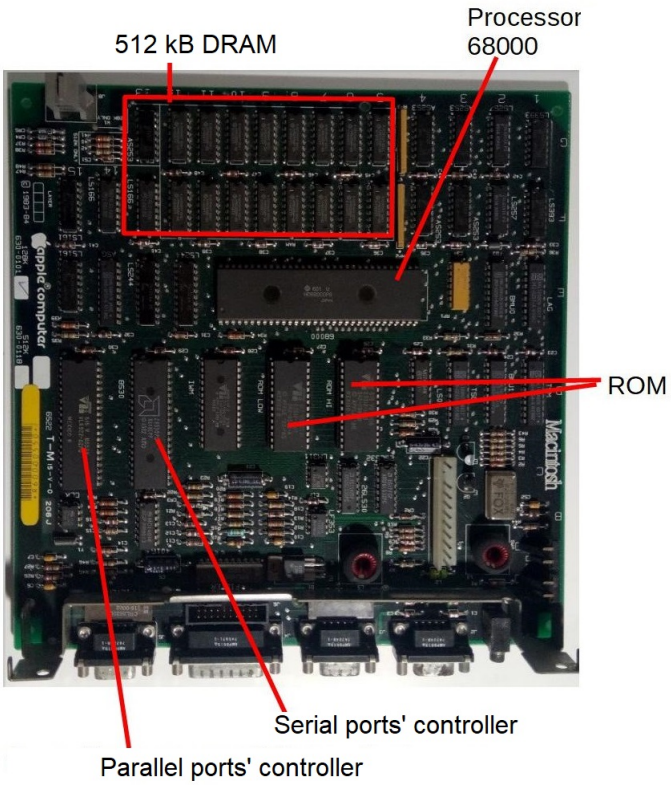


Figure 22: The MacIntosh 512 motherboard

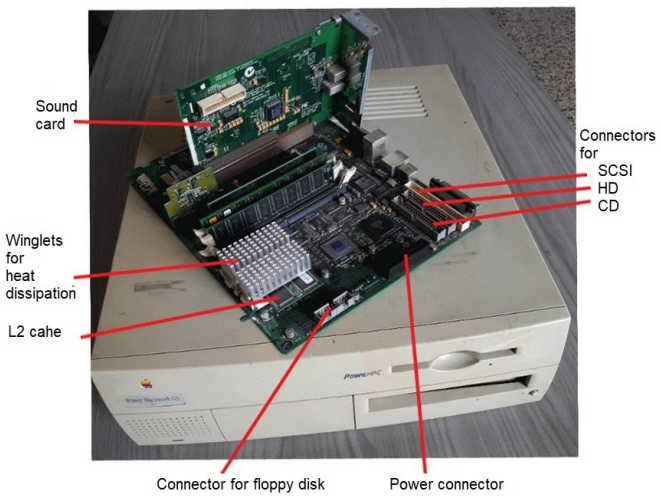


Figure 23: The Macintosh G3 M3979. The motherboard with its extensions rests on the computer casing. The PowerPC 750 processor is under the winglets of the heat sink, the two integrated second-level caches partially protrude from them. A special sound card can be seen on the back.

Apple stopped making machines with G3 processors in October 2003.

Excluding the small machines of the debut, in which the 8-bit 6502 microprocessor was used, Apple initially used the 68000 processors (Motorola); between 1994 and 1996 it switched to PowerPC; between 2005 and 2006 it switched to Intel processors. Recently, Apple has also abandoned Intel processors switching to in-house production [6], [7].

4.3.1. Power Macintosh G3 M3979 motherboard

Figure 24 shows further details of the motherboard on display.

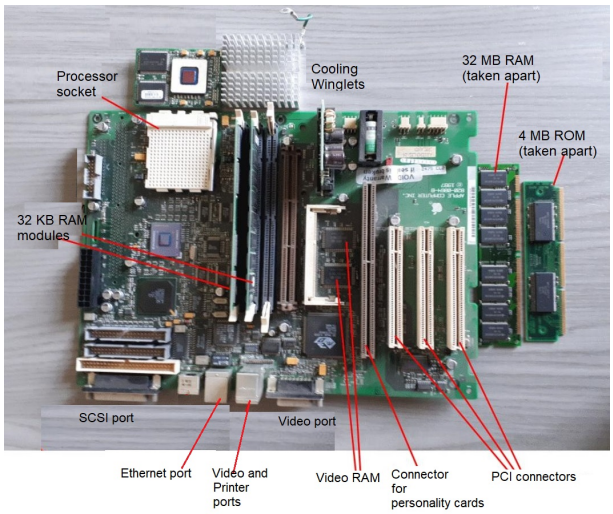


Figure 24: The motherboard of the Macintosh G3 M3979, after having removed the plate with the processor.

The processor and cooling winglets have been disassembled to reveal the socket on which the processor is mounted.

The DRAM memory consists of 3 modules of 32 MB for a total of 96 MB. A module has been disassembled and photographed on the right of the motherboard. The DRAM modules are all DIMMs (*Dual In-line Memory Modules*). The disassembled one and that appearing at the center of the photo have 16 integrated 2 MB. The third module (the left-most memory card seen vertically) instead is made of 8 ICs by 4 MB. modules. The system ROM is photographed to the right of the removed DRAM module; it is a SIMM (*Single In-line Memory Module*) that carries 2 ROM ICs by MB. The card features two integrated 128K 32-bit Video RAMs.

The SCSI port is for external SCSI disks (additionally, a SCSI connector is also provided for mounting disks inside the box). The system is equipped with an Audio card (Figure 23), which is mounted on the connector for personality cards.

4.4. Desktop PC motherboard with Pentium II

This card was produced in 1999. It is interesting because it belongs to the era in which some new standards began to establish themselves while some old standards held. Apart from the Pentium II, the card features 3 connectors for the traditional PC ISA (*Industry Standard Architecture*); 4 PCI connectors; 2 IDE (PATA) hard drive connectors; 1 connector for Floppy disk; Dimensions: 30x19 cm.

On the board, the integrated circuit that acts as a PCI controller (Intel FW 82371AB PCI set) can be seen. The

processor hides the RAM controller (Intel Fw82443lx Pentium Dram Controller). Behind the processor, there are 4 connectors for SIMM memory modules.

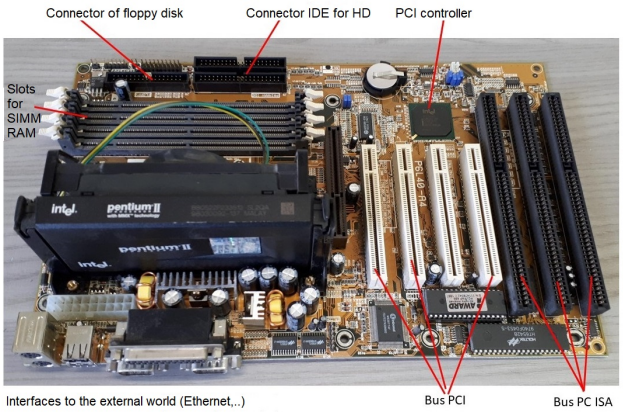


Figure 25: The motherboard of a Pentium II-based system

4.5. *Motherboard ASUS P5D2-VM with Pentium Dual Core E2140*

This board was produced for a few years starting in 2006. It has an LGA 775 socket on which Core 2 Duo, Pentium D, Pentium 4, and Celeron D processors can be mounted. In the specific case of Figure 26 the board featured the Pentium Dual Core E2140 (described on page 20). The board carries two 1 GB DDR2 DIMMs.

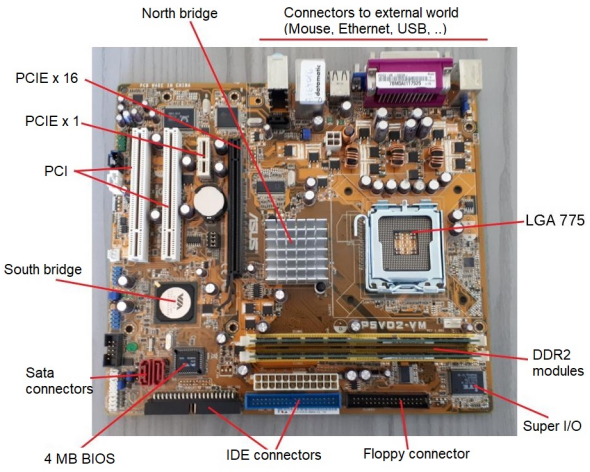


Figure 26: The ASUS P5D2-VM motherboard

To illustrate the organization of this system let us refer to the scheme of Figure 27.

This organization was essentially introduced with the Pentium and was also adopted for the class of processors like the Core 2 Duo. It is no longer used for more modern microprocessors, for which serial buses, such as QPI, are in use. From the CPU the so-called *Front-Side Bus* (FSB) comes out connected to the so-called Northbridge. This creates the AGP bus (graphics) or the PCI express bus, plus the bus to the memory. The Northbridge is connected to the Southbridge which in turn creates a series of (relatively) low-speed buses, i.e. the external buses (IDE, SATA, Ethernet, USB, ..), the PCI bus, plus the traditional PC bus. The BIOS ROM is connected to the Southbridge, as well as traditional low-speed buses (parallel port, serial port, keyboard, mouse, ..). Note that in this way, the BIOS can sit on ROM memory which is seen exactly as in the original PC. (Northbridge and Southbridge form the so-called Chipset.)

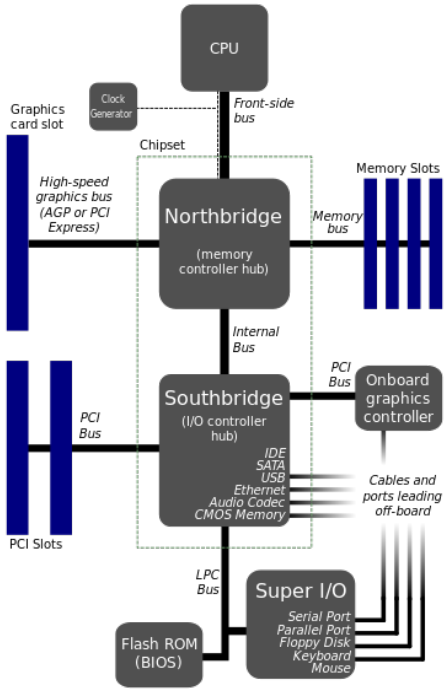


Figure 27: The organization based on the Front-Side Bus, introduced (more or less) with the Pentium.

4.6. *Motherboard of the "SuperServer" 6013A-T*

The SuperServer 6013A-T was produced around 2005 by the company Super Micro Computers. The system had the classic blade shape for rack mounting. Dimensions 4.3 x 42.6 x 57.4 cm (HLD); weight 15.9 kg. Figure 28 shows a blade.

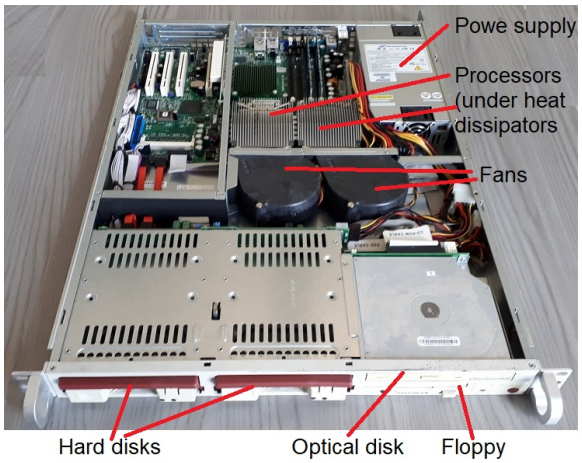


Figure 28: A blade of the server 6013A-T by Supermicro

A blade was a dual-processor server. Figure 29 presents the main elements of the system. Observe the two fans, whose purpose is to force the circulation of air between the cooling winglets of the processors. Up to two hard drives can be mounted (SATA) hot swap. There is an optical disc drive and a floppy disk. The motherboard (X5DPA-TGM+) occupies approximately half of the horizontal surface of the blade.

Figure 29 shows both the image of the card and the diagram of the layout of the components. The meaning of some components is illustrated below. There are two sockets on the board for mounting Xeon processors. In the picture one of the sockets is empty, while the processor is mounted on the other, hidden by the winglets of the heat sink. In the specific case of the board shown, two Xeon 2800dp/512/533 were mounted.

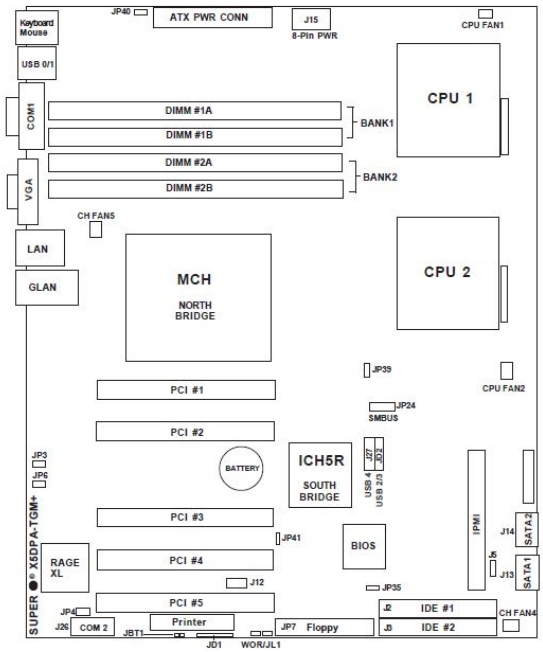


Figure 29: The 6013A-T server motherboard. Dimensions are 28x 25 cm

In summary, the main features of the motherboard were as follows: up to 2 Xeon processors (usable with socket 604), with frequencies up to 3.2 GHz; up to 8 GB of memory (using 4 x 2 GB DIMMs); two SATA ports for connecting two RAID disks (the ones on the front of Figure 28), hot-swappable; blade dimension was 30x25 cm.

4.7. SPARCstation 1

The SPARCstation 1, aka the Sun 4/60, was the first SPARCstation, sold by Sun Microsystems. It was marketed in April 1989, production ended in 1995. The SPARCstation 1 represented Sun's fourth computer model using SPARC processors. The box containing the electronics (Fig. 30), due to its very flattened shape, was defined as *pizza box*.

Figure 30: The "pizza box" containing the Sparcstation 1

SPARC stands for *Scalable Processor ARChitecture*, a RISC architecture defined by Sun Microsystem in the early eighties. The SPARCstation 1 sold for around $9,000 in the version without the hard drive, up to $20,000 in the most extended version. The motherboard (Figure 31) has connectors for the Sbus, the bus adopted by Sun in its systems, one Ethernet port and one for SCSI bus (at 5 MB/s). The Sbus remained in vogue for years, until it was supplanted by PCI and PCIe even on Sun machines. From the photo the processor is not visible, nor the arithmetic coprocessor Weitek 3167 coupled to it, as they lay under the card of video control logic.

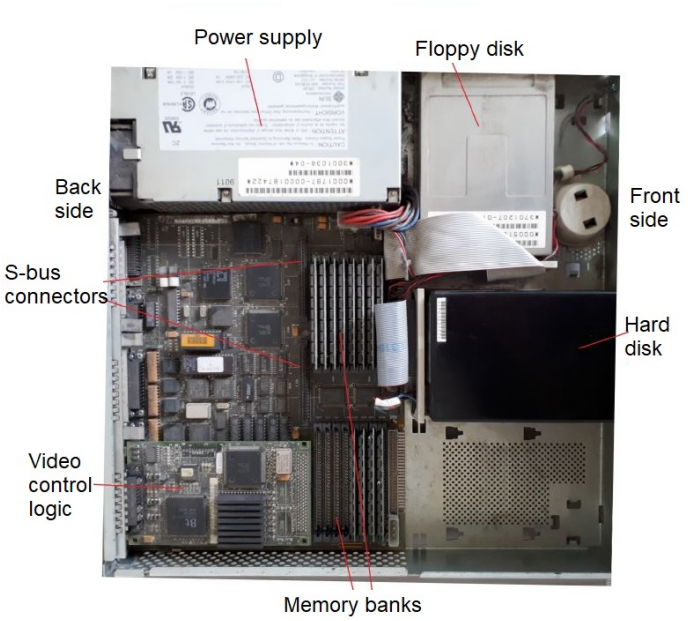


Figure 31: Components of the SPARC station 1

5. Electronic memories

The exhibit contains a magnetic core memory, some integrated ROMs and EPROMs, a couple of memory expansion cards, and several modern memory modules for PCs or Servers.

Below we describe the magnetic core memory, the HP and Multibus expansion cards, two more recent memory cards, plus a cache expansion.

5.1. A core module

Core memory was developed at MIT in 1951 [8] and has been used as random access memory (RAM) from the mid-1950s to the mid-'70s. It became obsolete when semiconductor RAM became available at a convenient price.

Figure 32 shows the core memory in the exhibit. The nameplate shows the diction "Memoria C.B.N. 16x8x7", indicating that it is organized as a three-dimensional matrix of 18x8x7 bits. It was part of an Olivetti computer, presumably an Elea [9]. Apparently, the module was built by Olivetti with Mullard material. The latter was an English company that produced vacuum tubes (valves) and iron cores.

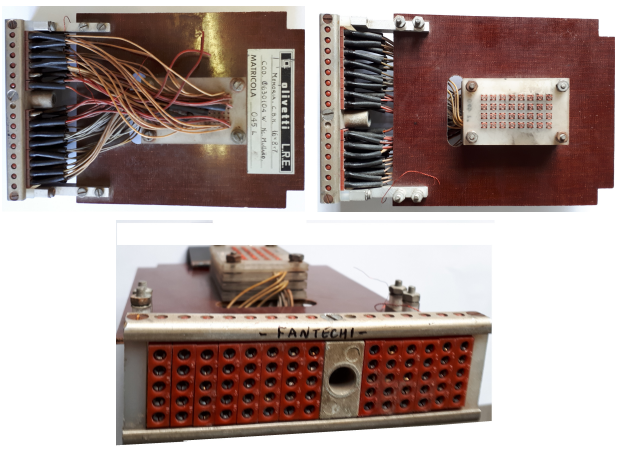


Figure 32: A "core" memory module, presumably of an Elea Olivetti computer. The front panel (below) measures 10x3 cm

The organization of the memory module of Figure 32 is somewhat unusual. Normally, core memories were formed by a matrix of magnetic cores in a toroidal shape, as shown in Figure 33. Each nucleus had the function of a single bit and it was crossed by three wires, as per the diagram of Figure 33. The nuclei were arranged with their axis parallel to the plane and inclined by 45°, in such a way as to make the passage of the three wires. Referring to the diagram, the single ring is addressed via a pair of lines (X, Y), while the third wire (SENSE) serves as a reading and inhibition line. The operation is based on the hysteresis of ferromagnetic materials: a core subjected to an adequate magnetic field tends to maintain its state as long as another magnetic field does not reverse it.

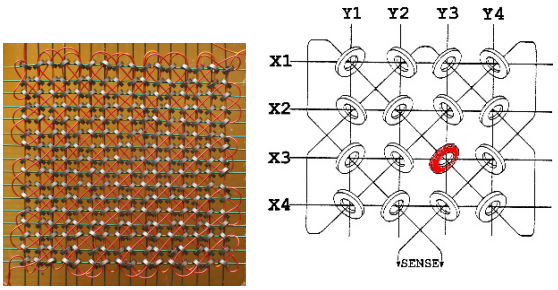


Figure 33: Usual core layout and addressing scheme.

To bring a nucleus to 1 it was necessary to make a flow of current through the relative X and Y wires. The value of the current was such that it affected only the nucleus at the intersection, while all the other nuclei on X and Y did not change the state.

The reading mechanism was quite complex. It began writing a 0: if the addressed core contained 1, a pulse appeared on the SENSE wire, due to the change in polarization of the nucleus; if instead, the bit contained 0 the pulse did not appear. Note that this technique is destructive of the nucleus content since its status is always brought to 0, therefore a subsequent write was necessary to bring back the state to 1, if that was the previous state. But if the state of the core had to remain at 0 it was necessary to inhibit the writing. In the first versions of these memories there was a specific Inhibition wire. Later, the SENSE wire was used (since the recognition function was not used while writing), to issue a current that generated a field opposite to that of the pair (X, Y), leaving the nucleus at 0. The controller of the memory had to switch the function of the line depending on whether it was reading or writing.

Referring to parallelism, if, for example, a parallelism of 8 bits was to be implemented, 8 matrices (planes) of nuclei were required.

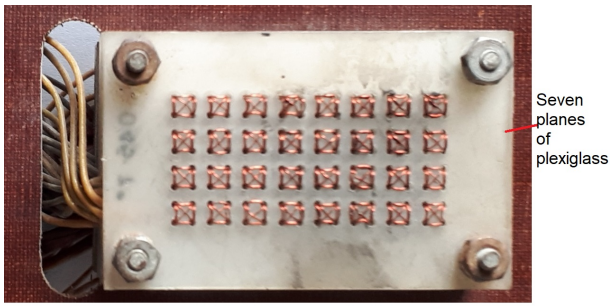


Figure 34: A close look at wiring

The 16x8x7 memory module is organized on 7 floors. Looking at the device closely (Figure 34, or, better, the lower image of Figure 32), 7 planes of plexiglass are seen. On each plane, there are 16x8 small holes (Figure 34), each of which is crossed by 4 wires. It can be deduced that the magnetic cores are placed and embedded in the plexiglass, with the hole of the ring corresponding to the small holes in the plexiglass, as to be crossed by the 4 wires; it can also be deduced that both the SENSE and the Inhibition lines were present.

On the front of the module (lower image in Figure 32 there is an array of female plugs. This suggests that the address lines and the two SENSE and Inhibition lines were brought to the module through pins or through a matrix of pins on which the front panel was fitted.

Note: Elea is the name of a series of computers developed by Olivetti in the second half of the fifties. The term "Elea" was chosen regarding to the colony of Elea in Magna Greece, home of the Eleatic school of philosophy (Parmedine, Zeno of Elea). The Elea 9003 was announced in 1959; the first example was delivered in September 1960. According to Olivetti publications, the Elea 9003 was the first commercial computer in the world entirely made with solid-state components (transistors). However, this statement can be hardly shared; exist, in fact, photos advertising IBM's 7070, introduced in 1958, the first computer using only Transistor.

5.2. 16 kB 16-bit word memory expansion card for HP 2100 computer

The expansion card in Figure 35 was mounted on the internal (proprietary) bus of the HP2100 minicomputer. The contacts towards the bus are the ones at the bottom. The dimensions of the card are approximately 19x20 cm. It uses the 4027 memory chip, containing 4K x 1 bit. On the card, 4 rows of 17 chips can be seen; they give rise to 16K 16-bit words. The seventeenth serves for parity control on each row. The HP 2100 minicomputer was introduced in 1971 and was the first of the HP2xxx series to be equipped with electronic memory (dynamic RAM). The previous models, HP 2116 and HP 2115, introduced in 1996 and 1997 respectively, used magnetic core memories.

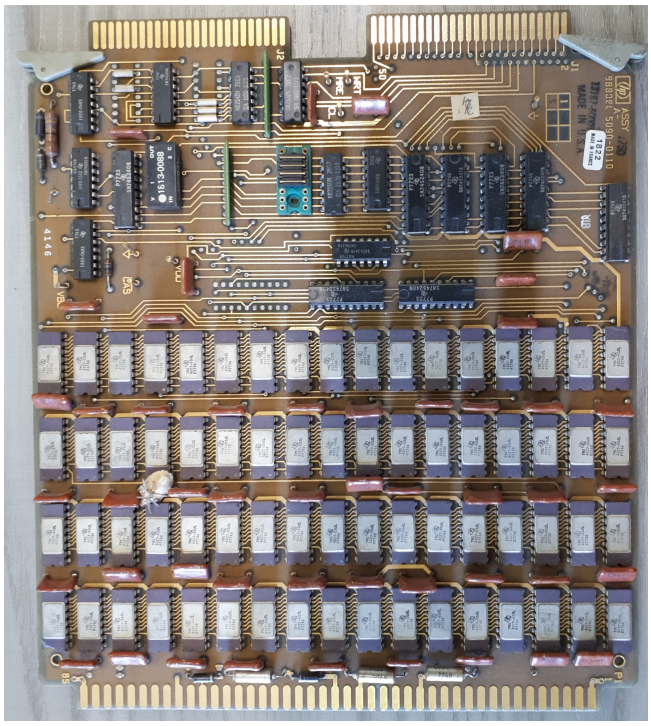


Figure 35: The 16 kB 16-bit word memory expansion card for HP 2100 computer.

5.3. 0.5 MByte Memory Expansion Card for Multibus

This board was produced by the English company Plessey in the early 1980s (Figure 36). It was designed to be mounted on the Multibus. This bus, defined by Intel, was very popular from the late 1970s through the 1980s. It was standardized by IEEE as the IEEE 796 bus. In 1982 there were over 100 industries producing cards for the Multibus. The board uses M3764 memory ICs, 64K x 1-bit dynamic RAM, in a 16-pin DIP.

The matrix of 17 by 4 memory chips forms a 512 KB memory bank, plus parity (16x4x64K/8). The integrated 5 by-4 memory matrix serves for error detection and correction. The cycle time was 500 nanoseconds, with an access time of no more than 300 nanoseconds. There were 4 devices (Dip switches) to configure the operating mode of the board. In addition to addressing, they allowed the selection of 8 or 16-bit operating modes. This was typically derived from the fact that on the Multibus could be used cards with 8 or 16-bit CPUs, typically the 8088 or 8086. Also, note the AMD 2964 dynamic memory controller.

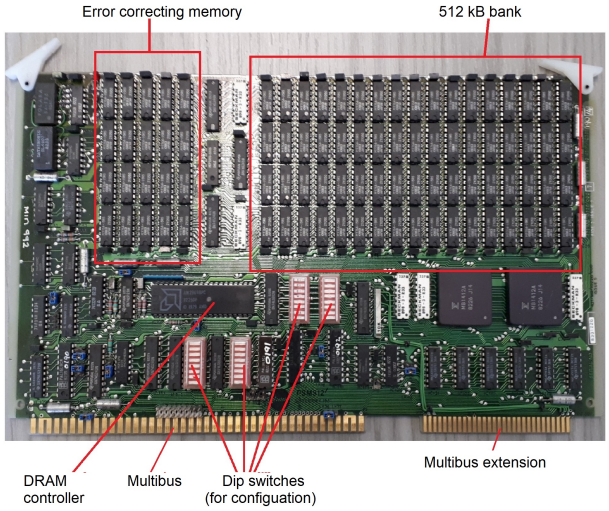


Figure 36: The 512 KB expansion card for Multibus. The dimensions are approximately 30x17 cm.

5.4. 256 kB Memory module

The part in figure 37 has been one of the first memory modules in a modern format. It was manufactured by MBS (Japanese) in the early 1990s and is marked as NMBS TCC-T-H7V. Capacity was limited to 256 kB. In the upper image, the two chips form a 256 kB block (their exact characteristics could not be traced); the third chip is presumably the memory controller, specially produced for those memory devices.

Note that the size of the plate is about 9 cm, and there is no notch. Subsequently, modern memory modules have been standardized, first with SIMMs and DIMMs. These have a size of less than 11 cm and a notch for their correct placement.

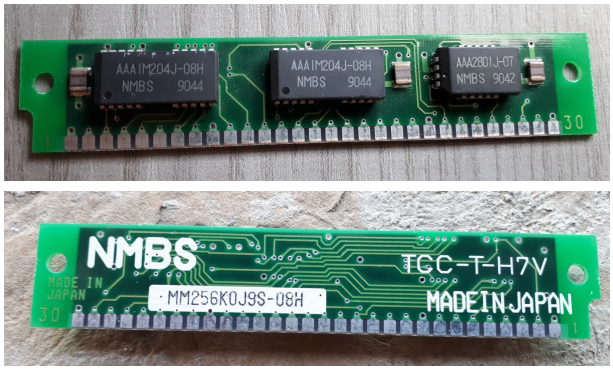


Figure 37: A 256 kB memory module (front and rear), built using chips of 256 k per 4. This does not corresponds to SIMM or DIMM standards.

5.5. 8GB DIMM Memory module, DDR3 (Samsung M378B1G73EB0-YK0)

The DIMM M378B1G73EB0-YK0 (Figure 38) features 16 Samsung K4B4G0846E 4 Gbit DDR3 memory chips, 800 MHz operating frequency, 1.35 V supply voltage. The nameplate says 8GB 2Rx8 PC3L-12800U. The transfer rate is 12.8 GB/s. Currently, this is the highest integration density memory module among those on display. The module was used on an HP laptop. It does not have the extra bits for error checking.
Note the notch.

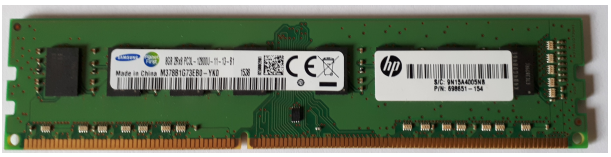


Figure 38: The Samsung M378B1G73EB0-YK0 DIMM memory module

5.6. 256 KB cache memory module

The part in Figure 39 is a DIMM module (160 pins) functioning as a 256KB second-level cache memory. It is not known which system it came from (probably from a system based on the micro 80486). The production is Chinese. The ICs were produced by the no longer active Utron Technology company (production in Taiwan). The two UT6132C32AQ ICs (on the right in Figure 39) are synchronous SRAM (CMOS), organized as 32K 32-bit words. The UT6164JC integrated circuit (on the left) is instead a 32K byte SRAM (CMOS).

Presumably, it is a direct mapping cache organized as follows: the two integrated 32K x 32bit ones give rise to a data RAM of 32K by 8 bytes (i.e. 256KB of data RAM), while the third IC (32KB, equal size of the data RAM) is used to contain the Tags.

Figure 39: An example of an old cache module (DIMM).

6. Storage Media

The exhibit contains storage media from the past such as punched cards, punched tapes, floppy disks, magnetic tape cassettes, and several hard disks. The following describes punched cards and tapes, flexible disk drives, and several hard disk drives.

6.1. Punched cards

Until the 1970s, punched cards were perhaps the main support for programs and data. All computer centers were equipped with card readers and punchers. The format card standard was the one defined by IBM in 1928: 80 columns by 12 rows, on a cardboard measuring 1187.325x82.55 mm. One character could be encoded on each column through drilling. The cut corner, at the top left, was introduced to avoid possible errors regarding the position of the cards (on puncher or reader). Punched cards have had a long history. In 1801 Joseph Jacquard used cardboard punched cards for controlling a weaving loom that became known as a Jacquard loom. In 1837 Charles Babbage, the creator of the first programmable computer, adopted Jacquard's punched cards for controlling the sequence of computation for his analytical machine; it was still a mechanical machine. In 1885 Herman Hollerith used his tabulating machine to read and count data punched on index cards to take a census of the population of the United States. The company founded by Hollerith in 1896 Hollerith, the Tabulating Machine Company, through some steps, became IBM (International Business Machines) in 1924).

Examples of cards are pictured together with floppy disks in Figure 41.

6.2. Punche tape

Until the 1970s, while traditional computer centers used punchers and card readers, minicomputers were equipped with punchers and tape readers. The tape was made of paper or plastic material (Mylar).

Figure 40: A sample of punched tape and of a Teletype.

The reader was normally of the optical type. Even teleprinters, often used as consoles, could be equipped with a tape reader/puncher. Figure 40 shows one of these machines (a *Teletype*). Note that the tape reader/puncher is part of the teleprinter. A piece of punched tape is shown above the teleprinter. On the tape, each character was punched across 8 bits, usually in ASCII coding for alphanumeric characters, or directly in binary. Note that in addition to the 8 (possible) holes for the character, a much smaller hole was also produced in correspondence with each of them. This was to give synchronization. In the case of optical readers, the tape was pulled by a conveyor; the little hole gave the synchronization to read the 8 bits at the right time, providing a time window centered on the character's transit window. The striping of the little holes divided the characters into 3 and 5 bits.

6.3. Floppy disks

Flexible (floppy) disks were the inevitable elements in any kind of system until the first five years of the 21st century. Subsequently, they began to disappear, replaced by memory sticks.

In 1967 IBM began studying the development of floppy disks. In 1971, IBM also released the first floppy disk, in an 8-inch format, to transferring the microcode (i.e. the microprogram that gave rise to the hardware behavior of the machine) into the control memory of the System/360. Shortly thereafter, the IBM employee responsible for the project (Alan Shugart) founded his own company to produce 8" floppies. This format was abandoned in 1976 in favor of

the 5.25” format. In both of these, formats the container was a plastic bag. Subsequently, the industries began to produce disks with rigid cartridges measuring around 3"; but it was only in 1981 that Sony introduced the 3.5" disk in the shape and structure that later became standard. Figure 41 shows an example for each of the three formats. The photo also shows some punched cards and a miniature CD.

Figure 41: Flexible disks, punched cards and a miniature CD.

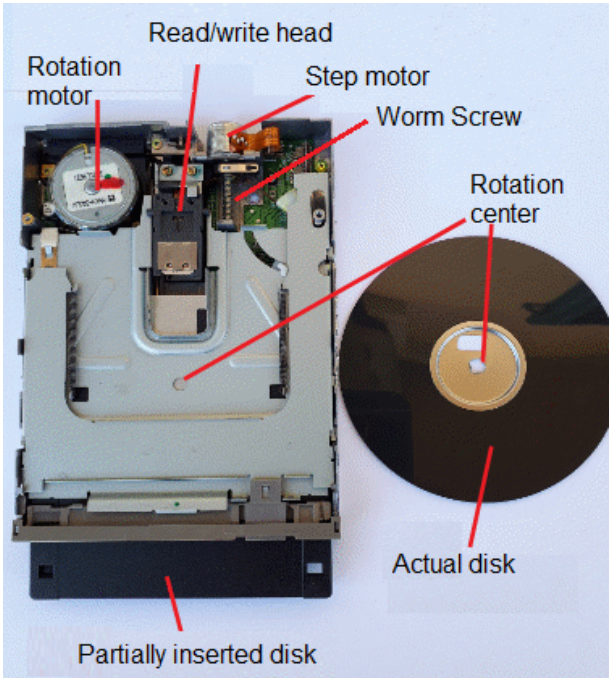


Figure 42: 3.5-inch flexible disk drive. The actual disk is contained in the plastic cartridge

3.5” flexible disks

Figure 42 presents the details of the 3.5” floppy driver on display.

Next to the driver, there is the actual “disk”, deprived of its container. It is a disk of plastic material, Mylar, on which the magnetic layer that acts as an information support is deposited. Note that a disk is partially inserted in the driver, and the metal part still covers the window on which the reading/writing head moves. The metal cover is automatically moved when the disk is completely inserted, to allow the head to face the actual disk.

When inserted, the disk is kept in continuous rotation, driven by a transmission belt guided by the rotation motor; the belt follows a non-direct path between the motor and the rotating element. The read/write head is driven by a step motor. Looking at the image of Figure 42 a worm screw in the axis with the step motor is seen. This screw engages with the head and causes it to move forward/backward (in the vertical direction in the image).

6.4. Hard Disks

Until recent years hard disks have been the most important auxiliary storage devices. Currently, SSDs (Solid-State Drives) tend to replace them. The relatively low cost per byte and the non-volatility of the information stored in hard disks make them suitable for long-term storage. The following description refers to modern hard drives, i.e. disks contained in a sealed cartridge. The first version of these disks was introduced by IBM in 1973. The drive was, in reality, made up of a fixed part and a removable part, both 30 MB. The device was called "Winchester 30-30" from the name of a rifle, the Winchester 30-30, owned by one of its designers, and thus the related technology was called *Winchester technology*, but this term disappeared since all hard drives are now in sealed cartridges.

The diagram in Figure 43 shows that a modern hard disk is made up of a certain number of plates (aluminum) whose surfaces are covered with a layer of magnetic material which acts as a storage medium. The diameter of the plates ranges from 1 to 5.25 inches. 3.5" disks are normally used in desktop computers and server systems; 2.5" are used in laptops. The plates are in constant rotation, with a constant speed normally included between 3600 and 15000 g/m. The read/write heads are integral to each other. The dishes and the read/write heads are inside a sealed, vacuum container. This requires extremely refined mechanics. Each magnetized face is subdivided in tracks, i.e. in concentric circles on which the data are stored. The number of tracks per face ranges from 1000 to 5000. Each track is divided into sectors. The number of sectors per track ranges from 64 to 200. For a given disk the number of sectors per track is constant, regardless of the diameter of the trace. As a result, there is a decreasing storage density from the center towards the periphery. The sector is the smallest read/write unit. Traditionally, the typical sector size is 512 bytes, but also larger sizes (1024, 2048 bytes) are in use. The average time required by the head to position itself on the addressed sector is called *Average positioning time*.

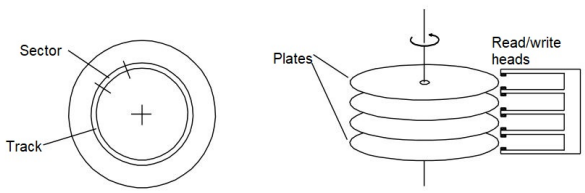


Figure 43: Schematization of a hard disk

The most common hard drives present one of the following interfaces:

- ATA (*Advanced Technology Attachment*), introduced in 1986, for which they are also used other names (IDE, EIDE). Since it is a parallel interface today it is normally indicated as PATA. In 2003, the serial version of it, named SATA, has practically completely replaced PATA.
- SCSI (*Small Computer System Interface*), introduced in 1978 under the name SASI (*Shugart Associates System Interface*) from the name of the disk manufacturer who proposed it, became soon a very popular industry standard. The original SCSI was a parallel interface, designed to connect a high number of devices, not necessarily disks. Successively, the interface has undergone a significant number of updates. The serial version (SAS, *Serial Attached SCSI*) was introduced in 2003; it is also replacing the traditional parallel.

6.5. The Diablo 40 Series Hard Drive

As stated previously, before the emergence of modern sealed drives, hard disk drives included a fixed and a mobile disk. The dimensions of the plates were much larger than the current ones.

The image on the left of Figure 44 shows the platter of one of these disks. The diameter is 14", or 35.56 cm. The platter was made of aluminum and covered with magnetic material. The moving plate was itself enclosed within a cartridge which could be removed and replaced with another cartridge. Given its size and shape, the removable cartridge was called (at least in Italy) *frying pan*.

The image on the right of Figure 44 shows one of the most successful minicomputers of the seventies: the Data General Nova 3, introduced in 1976. A cartridge rests on the top shelf of the rack. The lower part of the image shows the compartment where the cartridge had to be inserted.

This hard disk belonged to the Series 40 Diablo. Diablo was a sub-brand of Xerox, and these disks were widely used with minicomputers up to the seventies. They had a capacity of 5+5 MB (5 fixed and 5 mobile). Interchangeability potentially allowed an unlimited disk space.

Figure 44: On the left, the plate of a Diablo disc; on the right, Data General's Nova 3 minicomputer equipped with a Diablo disk.

Data General was founded in 1968 by leavers from Digital Equipment Corporation (DEC), including Edson de Castro, the designer of the first commercial minicomputer in history the Digital's PDP-8. Both Digital and Data General are long gone. Digital was purchased in 1998 by Compaq, which was later merged into Hewlett-Packard in 2002; Data General was purchased in 1999 by EMC, which later became Dell.

6.6. Vertex V130

The platters of the Vertex V130 had a diameter of 5.25", a format that is now abandoned. The hard drive in question (Figure 45) was part of an HP9000 station; these were his characteristics: 5.25 inch format; capacity 26 MB; number of sectors 50,337; Bytes per sector 512; average positioning time (seek time) 30 ms; transfer rate (max) 0.625 MB/S; rotation speed 3600 rpm; number of disks 2; number of read/write heads 3; production period early 1980s.

Figure 45: The Vertex V130 (5.25").

6.7. IBM DSAA-3540

The disk in Figure 46 was part of a series called Dekstar produced by IBM. The series was introduced in 1994 by IBM, passed to Hitachi at the end of 2003, and then to Western Digital which abandoned its production in 2018. The DSAA 3540 model is from the second half of the first decade of the 2000s.

Figure 46: The disk opened, and, below, the control logic.

These were the main features: size 3.5"; capacity 527 GB; IDE-AT interface; 12.0 ms average positioning time (seek time); 1,240 cylinders; 63 sectors per track; 512 bytes per sector; 8.3 MB/s transfer speed; 4,500 rpm rotation speed; 2 plates; 3 read/write heads (the central head operates on two faces); weight 530g.

6.8. Quantum Fireball Plus LM

Figure 47 shows the Quantum Fireball Plus LM. It was introduced in 2000, and manufactured by Quantum Corporation. These were the main characteristics: size 3.5"; capacity 20.5 GB; interface type ATA-66; average head positioning time (seek time) 8.5 ms; transfer speed 66 MB/s; rotation speed 7,200 rpm; number of plates 3; number of read/write heads 6.

Please note the connection cable for the PATA interface (flat cable on the left side). Compare this cable with the SATA cable in Figure 48.

Figure 47: The Quantum Fireball Plus M.

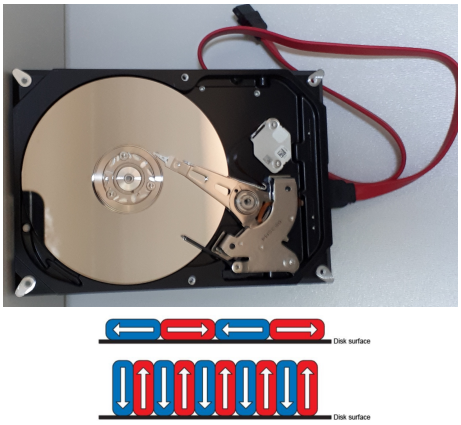


Figure 48: Above the Seagate Barracuda 7200.10. Below: the "vertical" bit orientation compared with the traditional "horizontal" bit orientation (above).

6.9. Seagate Barracuda 7200.10 Modello ST 3250310 AS

The Barracuda 7200.10 series (from Seagate Technology) includes drives ranging from 80GB to 720 GB. Both 1.5 Gb/s or 3 Gb/s SATA interfaces and a PATA 100 interface, are available for them. This series was introduced around 2006. The main characteristics of the specimen in the show are as follows: size 3.5"; capacity 250 GB; SATA type interface; average rotation latency time 4.16 ms; average transfer speed 125 MB/s; rotation speed 7,200 rpm; number of plates 2; number of read/write heads 3.

The SATA connection cable was left on the specimen of Figure 48. Compare this cable with the PATA cable of the Quantum Fireball Plus of Figure 47).

It is worth noting that, to increase storage density, this disk adopts vertical storage, differing from previous products where the bit orientation was horizontal (refer to Figure 48).

7. Microprocessors

In November 1971 in a then very widespread electronics magazine (Electronic News), this ad appeared: *Announcing a new era of integrated electronics. A micro-programmable computer on a chip!* Since then the term "microprocessor" has been part of the common lexicon.

The exhibit contains a substantial number of microprocessors, largely belonging to the x86 family. There is also a precursor of microprocessors technology, that is SLT logic and we start describing it.

7.1. Solid Logic Technology

The *Solid Logic Technology* (SLT) was the precursor to integrated circuits in the form we know today. It was introduced in 1964 by IBM in System/360. It represented the first example for miniaturized semiconductor circuits of large industrial production (Figure 49).

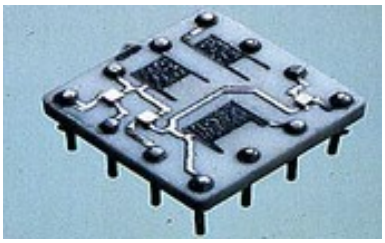


Figure 49: An example of an SLT component

SLT devices were denser and faster than devices using separate transistors and consumed less energy. These were hybrid circuits, including transistors and diodes, mounted on a ceramic base having a surface area of half an inch square, as illustrated in Figure 49. Figure 50 shows one such component.

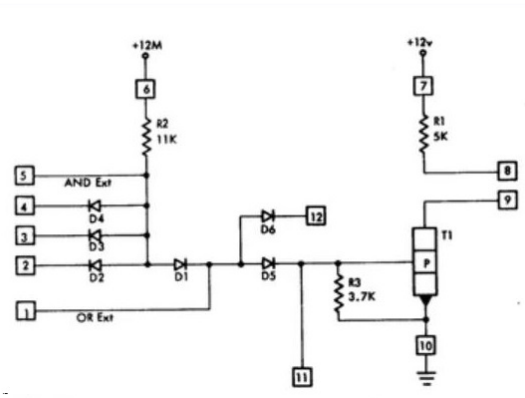


Figure 50: The electrical diagram of IBM component 361493 used in the S/360 system (present in the module of Figure 51). This SLT component was cataloged as an AND-OR-INVERTER (AOI), basically a NOR gate.

Figure 51 shows an SLT module used in the S/360 system. Various components appear mounted on the module; four of them correspond to the mentioned IBM component 361493.

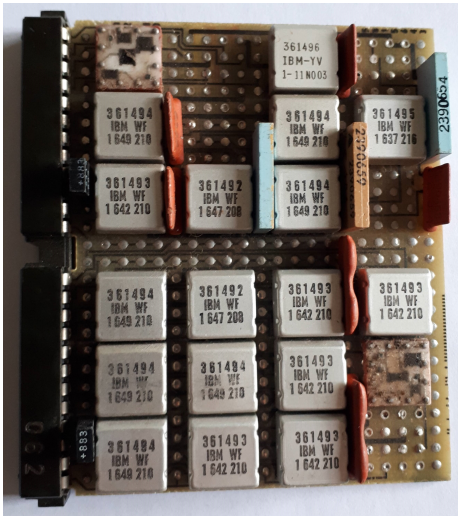


Figure 51: An SLT logic module belonging to the IBM S/360 system. Note the component in the top left corner and the one in the bottom right corner, without the cover; as you can see, the two have different elements.

7.2. The 8008 microprocessor

The 8008 (Figure 52) was produced by Intel and released in April 1972, just over a year after the first microprocessor, the Intel 4004 [2], produced with the fundamental contribution of Federico Faggin [3].

These were the main characteristics 8008: 0.2 MHz clock frequency, 3,500 transistors in an 18-pin DIP package. The address space was 16 kB (14-bit address lines). There was a single 8-bit bus (for data and addresses); as a result, to access memory, several clock cycles were required. In fact, it was necessary to present the 14 address bits in two successive phases, so that the address could be captured on an external latch (first 8 bits and then the remaining 6); the data (8 bits) passed on subsequent machine cycles.

A detailed description of the 8008, as well as of the 8080, 8085, and 8086 can be found in [10].

Note that the 8008 was commissioned by the Datapoint company, aiming at its use in the Datapoint 2200 programmable terminal. However, the device was delivered too late and did not meet the client's expectations in terms of performance. It was therefore not used in the Datapoint 2200.

Figure 52: The 8008 integrated circuit.

7.3. The 8085 microprocessor

This was an 8-bit microprocessor introduced by Intel in 1976; production continued until 1990. These were its main characteristics: clock frequency 3.5 MHz; 6,500 transistors in a 40-pin DIP package; address space 64 KB; 8-bit data bus, 16-bit address bus. The data bus is multiplexed with the lower part of the addresses.

The 8085 was the improved version of the 8080, a highly successful 8-bit processor, introduced in 1974. Specifically, the 8080 was the first widely used microprocessor, used not only to replace wired logic but also for the creation of personal computers (then called microcomputers), being able to address a memory space 64 kB of memory space, which was considered even excessive at the time. The instruction set of the 8080 was significantly influenced by that of the 8008.

The 8085 had a single supply voltage of 5V, hence the "5" as the last digit in its name. Whereas, the 8080 required two additional supply voltages, -5V and 12V. Furthermore, the 8085 had a more refined interrupt system than that of the 8080. In its use in personal computers, the 8085 and 8080 were outclassed by the Z80.

7.4. The Z80 microprocessor

The Z80 was introduced in 1976. It was designed and developed by Zilog, a company founded by F. Faggin, after he left Intel. These were its features: clock frequency 2.5 MHz; 8,500 transistors in a 40-pin DIP package; 4-micron MOS technology; 64 KB address space; 8-bit data bus, 16-bit address bus, distinct; It had a compatible but more extensive repertoire of instructions than that of the 8008/85 and had a more advanced I/O management system.

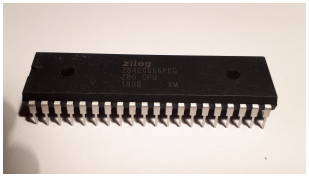


Figure 53: The Z80 integrated circuit.

The Z80 became very popular in the construction of personal computers. At the time, the dominant operating system (also working on 8080/85) was CP/M, which was swept away by the arrival of PC-DOS for 8088.

The Z80 was produced by a large number of electronics companies, including Hitachi, SGS-Thomson, NEC, Sharp, Toshiba, National Semiconductor, and Mostek. It is still produced today. Furthermore, there are various SoCs (*System on a Chip*) in circulation that incorporate the Z80, as well as various IP (*Intellectual Property*) *core* containing the Z80. IP cores are designs or specifications of logic that are licensed to build devices; they are typically used as building blocks for ASIC (*Application-Specific Integrated Circuit*) circuits.

7.5. CDP 1802 microprocessor

This microprocessor (Figure 54) was introduced by RCA (*Radio Corporation of America*) in 1976. RCA has been one of the most important American companies in the field of electronics. Currently the RCA brand belongs to a French multinational.

Originally the CDP 1802 was designed for use in personal computers (*home computers*), but, irony of history, due to its peculiar characteristics (low consumption and static logic) it found great use in *embedded applications*.

The address bus was 8 bit-wide, therefore the 16 bits of the bus were transmitted on two successive clock cycles.

The two characteristic aspects that differentiated the CDP 1802 from the microprocessors of the time, normally

in nMOS or MOS technology, were the CMOS technology and the static logic. The CMOS technology led to reduced consumption, while the static logic allowed the clock frequency to be brought to zero; in this case, the processor maintained its state as if it were in a halt state, but without the consumption corresponding to the dynamics imposed by the clock. For this reason, it had some success in space applications, as static logic was considered more resistant to radiation. A *radiation hardened* version was also built, specifically for space missions.

Figure 54: The CDP1802ACD3.

The CDP 1802 (also known as COSMAC, from the name RCA gave its CMOS process: *complementary silicon/metal-oxide semiconductor*) was used in the Galileo mission. The spacecraft was launched in October 1989 and reached Jupiter in December 1995; orbited the largest planet in the solar system for almost 8 years. Out of curiosity, we report the dimensions of the software: 650,000 lines of code for orbital control, 1,615,000 lines for telemetry, and 550,000 lines for navigation. Obviously, the spacecraft had a multitude of microprocessors.

The example on display (Figure 54) is the model CDP 1802ACD3, an improved version of the initial 1802.

7.6. The 8086 microprocessor

The 8086 [10] represented a milestone in the development of computing. It was Intel's first 16-bit microprocessor and defined an architecture that is still referred to as "x86". Almost any personal computer contains a processor whose roots lie in the distant progenitor 8086. The use of x86 architecture processors is not limited to personal computers: they are the basis of many server systems and many supercomputers.

Over the years, the original 16-bit version was extended to 32 and later to 64-bit. The 64-bit extension, with which segmentation was practically put aside, was defined by AMD – not by Intel. Given the great popularity of the x86 architecture and the great acceptance of its extension to 64 bits, the giant Intel was somehow forced to adopt it, abandoning other projects for 64-bit machines. Intel continues to be the largest manufacturer of x86 processors, followed, at some distance in terms of numbers, by AMD.

The 8086 was introduced in June 1978. These were the features of the introduction model: Clock frequency of 5 MHz; 29,000 transistors in a 40-pin DIP package; 1 MB address space; 16-bit data bus, 20-bit address bus data and address bus multiplexed.

The state of technology of the time, the constraints imposed on the design (presumed compatibility with the 8085 8-bit microprocessor, and use of a 40-pin integrated circuit), and the intent to define a sophisticated protection system had a notable influence in determining the architecture and making it a little complicated; specifically, the 8086 featured a segmented memory model, a model that was substantially put aside by the 64-bit architecture. The limited number of pins also imposed multiplexing of the data and address bus.

The average number of *Clock cycles Per Instruction* (CPI) could be estimated at 15. Thus, at a clock rate of 5 MHz, the average time to execute an instruction is $15 * 200ns = 3\mu s$, corresponding to a performance level of approximately 0.33 MIPS (*Million Instructions Per Second*).

The 8088 microprocessor

The 8088 (Figure 55) is the 8-bit version of the 8086. It was introduced about a year after the 8086. The 8088 is a CPU completely compatible with the 8086: it has essentially the same internal structure, like the 8086 it has a memory space of 1 MB, but has an 8-bit external data bus.

Figure 55: The 8088 integrated circuit and the original IBM PC.

In 1981 IBM introduced a personal computer, the IBM PC, based on the micro 8088, with a clock frequency of 4.77MHz. The choice of the 8088 essentially arose from cost reasons: the 8-bit data bus made it possible to reduce the costs of the electronics compared to those of an equivalent 8086-based system. Moreover, at the time the peripherals that could be used on Personal Computers (PCs) were practically only 8-bit, while the speed of the 8088 (although lower than that of the 8086) was even excessive for the typical applications of the time.

The IBM PC was not the first personal computer to make its appearance on the market. For years, a variety of personal computers based on 8-bit microprocessors such as the 8085 and the Z80 (all these kinds of machines had an address space limited to 64 kB) had been on the market. To get an idea of why the IBM PC enjoyed a huge favor, we must keep in mind that, until the beginning of the eighties, in the USA, which practically represented the computer market almost exclusively, the acronym "IBM" and the word "computers" were practically synonymous. When IBM introduced the PC, to the general American public it was as if the PC was born at that moment. Large users, such as banks, industries, state apparatus, etc., who had until then been refractory to the use of PCs, found their use completely natural, also by the supposed possibility of integration with existing IBM mainframes.

Time magazine, which traditionally dedicates the cover of the last issue of each year to the most distinguished person of the year, in 1981, to the surprise of its readers, dedicated the cover to the PC instead of to a person.

The adoption of the 8088 in the IBM PC and the success of the latter determined the success of the x86 architecture.

7.7. The 80386 microprocessor

The 80386 was the first 32-bit x86 architecture microprocessor [1]. It is usually indicated as 386. Produced from October 1985 to September 2007.

These were the characteristics at the time of introduction: clock frequency 12.5 MHz; 275,000 transistors in a 132-pin *Pin Grid Array* (PGA) package; IA-32 architecture (x86 extension); separate data bus and address bus, both 32-bit.

In addition to doubled parallelism, the 386 introduced virtual memory paging after segmentation. The maximum virtual space was up to 64 TB. The 32-bit address bus allows a maximum physical addressing of 4 GB (compared to 64 TB virtual).

The 386 represented a milestone in the development of the x86 family, implementing the transition to 32-bit architecture. The 386 redefined the format of the instructions and extended the repertoire of the 8086 (for example by making the registers substantially interchangeable with each other, unlike the 8086 where they were instead specialized and therefore subject to a more limited use). The 386 did not yet have an integrated cache or floating point arithmetic. For the cache, the manufacturer made some ICs to build it externally; for floating point arithmetic, the 80387 coprocessor had to be used.

The photos in Figure 56 show two 386s on display: one Intel and one AMD. This second one is surface-mounted and is mounted on a socket to make it compatible with the Intel pinout.

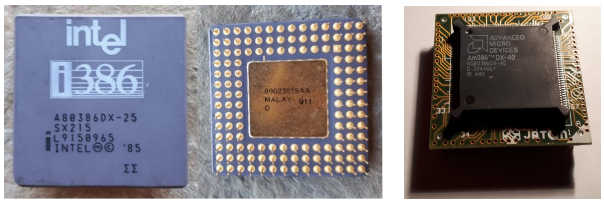


Figure 56: Two examples of the 80386 integrated circuit

7.7.1. The 80387 coprocessor

A specific coprocessor was introduced for the 386, the 80387. Figure 57, shows three specimens on display.

Figure 57: Examples of the 80387 integrated circuit. From left to right the frequencies are 25, 40, and 33 MHz respectively

7.8. The 80486 microprocessor

The 80486 (486 for short) was introduced by Intel in April 1989, production lasted until 2007. These were the characteristics at introduction time: clock frequency 25 MHz; 1 micron CMOS technology; approximately 1.2 million transistors into a 168-pin PGA.

With 486, a 5-stage pipeline was introduced: Prefetch (PF), Decoding (D1), Address Generation (D2), Execution (EX) and Write Back (WB). The creation of the pipeline required overcoming several problems related to the *Complex Instruction Set Computer* (CISC) structure of the instruction set.

The 486 was the first device of the x86 family to integrate the floating point unit (FPU) and a cache memory (8 kB) on the same chip. The actual acronym of the processor with FPU was 486DX. At the same frequency, the performance of the 486 was 2 or 3 times higher (depending on the type of processing) than the performance of the 386. With the 66 MHz DX2 version the performance of the 486 reached 54 MIPS.

Figure 58 shows 486's die on display.

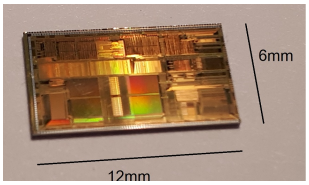


Figure 58: The die of the 486 (presumably a DX2).

In mid-1991 Intel introduced the lower-cost version of the 486, called the 486SX. It differed from the 486DX in that it lacked floating point (FPU) arithmetic. It is interesting to note that the first 486SX were substantially identical to the 486DX, i.e. they contained the FPU which, however, was disabled. The price was lower only for commercial reasons. Subsequently, towards the end of 1992, the FPU was removed to reduce production costs.

The 486 introduced a differentiation between the frequency of the processor and that of the external bus which, depending on the models, operated at a frequency equal to half or a third of that of the processor. On display, there are 2 examples of 486 (Figure 59), one produced by Texas Instruments and one produced by AMD (the pinout of these is also shown here, obviously identical). The dimensions are 4.45x4.45 cm. The second IC states that a heat sink and fan are required.

486 Texas
Frequency 100 MHz
Bus frequency 33 MHz
8 KB cache
Voltage 3.45 V

486 AMD
Frequency 120 MHz
Bus frequency 40 MHz
8 KB cache
Voltage 3 V

Figure 59: The two examples of 486 on display

7.9. The Pentium microprocessor

The Pentium (Figure 60), introduced in March 1993, was the first microprocessor of the x86 family to cross the 100 MIPS threshold. These were the characteristics at the time of introduction: clock frequency 60 MHz; 3.1 million transistors, across 2.16 square inches, in a 273 PGA; 32-bit architecture.

[1] A short description of the 80386 and other processor of that epoch can be found in [11].

Figure 60: The Pentium integrated circuit.

While maintaining software compatibility with previous models, the Pentium provided some additional instructions, including those that allow the implementation of the MESI (*Modified, Exclusive, Shared, Invalid*) protocol for cache coherence in multiprocessor environments.

The Pentium introduced significant innovations: 64-bit external data bus; superscalar architecture; pipelined floating point calculation; separate internal cache for instructions (8 kB) and data (8 kB); prediction of conditional jumps; speculative execution (limited to the presence of conditional jumps).

It is worth noting that with the Pentium, Intel abandoned the numerical denomination of its processors because numbers are not patentable and therefore any manufacturer could introduce a device and call it, for example, 386. The term Pentium was used to indicate that it was the successor to the 486. The Pentium acronym was used for many subsequent x86 models, until the adoption of the acronym "Core".

7.10. Athlon 64 X2

This is the dual-core version of the Athlon 64 (this was the first 64 bit-processor of the x86 family). It has been produced from 2005 to 2009 by AMD (Figure 61).

Figure 61: The Athlon 64 X2.

Main features of the processor are [12]: 64-bit architecture; operating frequency 2.7 GHz; 16 bit bus HyperTransport 1000 MHz (2 GT/s); FSB frequency 800 MHz; number of cores 2; 221 million transistors; 65 nm technology; two second-level caches of 512; compatible with Intel extensions MMX, SSE, SSE2, SSE3.

7.11. The Intel Pentium Dual Core E2140 microprocessor

The Pentium dual-core E2140 was introduced in 2007. The Pentium name was abandoned for this processor and is now referred to as Core 2 Duo. Main characteristics: 64-bit architecture; operating frequency 1.6 GHz; bus frequency (FSB) 800 MHz; number of cores 2; 105 million transistors; 65 nm technology; 1 MB second-level cache; *Land Grid Array* (LGA) 775 type socket (i.e. with 775 possible contacts); extensions: MMX, SSE, SSE2, SSE3, SSSE3.

Despite having the name Pentium, this microprocessor, designed for desktop uses, adopts the Core microarchitecture, introduced by Intel in 2006. This microarchitecture is typically multicore and constitutes the continuation of the P6 microarchitecture, the one defined with the PentiumPro processor in 1995.

The example on display (Figure 62) was mounted on the ASUS P5D2-VM motherboard described on page 9 on which the LGA 775 socket can be seen. A heat sink and the related air movement fan were mounted on the microprocessor (see Figure 68).

Figure 62: The Pentium Dual Core 2140

7.12. The Core 2 Duo T5500 microprocessor

The Intel Core 2 Duo T5500 was introduced in 2008 for mobile computers. The example on display (Figure 63), was taken from a Vaio VGN-1GZ laptop computer (see Figure 69).

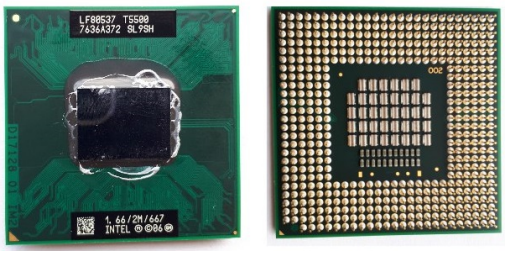


Figure 63: The Core 2 Duo T5500

Main characteristics: 64-bit architecture; operating frequency 1.66 GHz; bus frequency (FSB) 667 MHz; number of cores 2; no hyperthreading; 291 million transistors; 65 nm technology; 2 MB second-level cache; socket PPGA478 (PPGA: Plastic Pin Grid Array; power supply 1,0375V-1,300V; extensions: MMX, SSE, SSE2, SSE3, SSSE3.

Main characteristics: 64-bit architecture; operating frequency 2.66 GHz; bus frequency (FSB) 133 MHz; number of cores 4; no hyperthreading; 473 million transistors; 45 nm technology; 6 MB second-level cache; socket 775; (PPGA: Plastic Pin Grid Array; power supply 0.8500V-1.3625V; extensions: MMX, SSE, SSE2, SSE3, SSSE3.

7.13. The PowerPC 750 microprocessor

The PowerPC 750 was introduced in 1997 by IBM [13] and Motorola. It was one of the members of the 7xx family, also called the third generation (G3) of PowerPC processors, designed and produced by the AIM (Apple-IBM-Motorola) consortium, born in 1991 to define a new architectural standard, capable of counteracting Intel's dominance. It was a RISC (*Reduced Instruction Set Computer*) architecture, based

on IBM's previous architecture called Power. When introduced, the PowerPC had the best ideas and implementation that could be put together.

7.14. The Core 2 Quad

The Intel Core 2 Quad was introduced in 2008. The example on display (Figure 64), was taken from an Asus PSQL PRO mother board.

Figure 64: The Core 2 Quad mounted on the Asus PSQL PRO mother board.

At first, these CPUs were very favorably received by the market. Despite this, the PowerPC failed to counter the dominance of the x86 architecture, although it was used for a long time in Apple's Macintosh (Figure 65). Since 2006, Apple has abandoned PowerPC processors in favor of Intel processors; more recently Apple has switched the entire Mac lineup to its own designed chips. IBM continued to produce PowerPCs and use them in various product lines (including some supercomputers).

Figure 65: The PowerPC 750

The relevant features of the 750 model are as follows. Clock frequency at introduction 233 MHz; 6.35 million transistors, in a 360-pin CBGA (*Ceramic Ball Grid Array*) package; 64-bit data bus at 66 MHz; separate data and instruction caches of 32 kB each; dedicated 117 MHz external bus for a (unified) second-level cache.

Figure 66 shows the microprocessor and a module that mounts it together with a second-level cache memory.

Apple used the 750 model in the Power Macintosh G3, specifically in the M3979 model, described on page 7.

Figure 66: A module containing the microprocessor PowerPC 750 and two memory chips implementing a second-level cache memory.

8. Other elements on the exhibit

The exhibit contains other items related to computer technology, such as heat sinks, mice, various expansion, and interface cards. Below we describe a heat sink and an optical mouse.

8.1. Heat sinks

Heat sinks are often mounted over modern microprocessors. With miniaturization and the growth of frequencies, heat dispersion has become a significant problem. To understand this, refer to Figure 67 which shows the power density as a function of the technology.

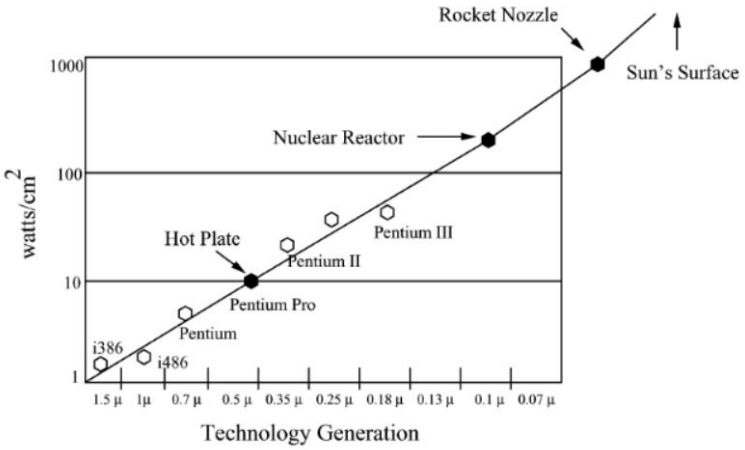


Figure 67: This diagram is from V. Venkatlachan and M. Franz "Power reduction techniques for microprocessor systems" ACM Computing Surveys, 37(3): pp 195–237, September 2005.

The ASUS board of section 4.5, described on page 9, had the heat sink of Figure 68. It was mounted in contact with the microprocessor, with a thermal paste interposed for heat conduction. The heat sink is made up of a radial pattern of winglets, between which the air is moved by the fan. A heat-conducting paste is always placed between the processor and the face of the heat sink that comes into contact with it. This is to determine a uniform contact surface.

Figure 68: The heat sink mounted on top of the processor (Core 2 Duo T5500) of the Asus board described on page 9, as well as on the processor (Core 2 Quad) of the Asus board PSQL PRO (Figure 64).

Figure 69 shows the overall heat sink system mounted on the microprocessor Core 2 Duo T5500 (inside the Vaio laptop VGN-1CZ). The heat sink consists of (i) a part that acts as a cover for the processor, (ii) a copper duct; and (iii) a sort of radiator over which the fan passes a flow of air. Schematically it works like this: the radiator is cooled by the airflow generated by the fan; consequently, heat is removed from the assembly formed by the duct and the processor cover.

Note that there is an additional heat duct that abuts the fan shell. This conductor comes from a heat sink located near the memory (not visible in the picture).

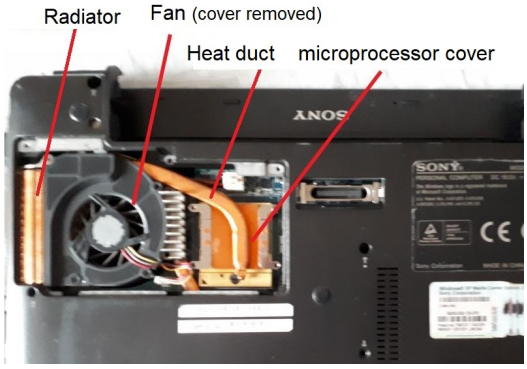


Figure 69: Part of the back of the Vaio VGN-1CZ laptop after removing the cover that hides the heat sink system.

8.2. An optical mouse

The first mice were mechanical. A portion of a ball (with 2 degrees of freedom) emerged from the underside of the mouse; by moving the mouse the ball rotated and transmitted its movement to two wheels with orthogonal axes to which two encoders were associated, measuring the rotation speed (and therefore the movement along the axes of the two wheels). The mouse logic transmitted trains of pulses corresponding to the movements to the computer and therefore the mouse driver could reconstruct the position of the cursor. Mechanical mice had a significant flaw: they filled with dust, resulting in an impoverishment of their functioning.

Early optical mice used an LED and a photodiode to detect movement on the surface. To work, they required the supporting surface to be metallic with a dense network of thin blue lines.

Current optical mice still use an LED to generate light (the most refined ones use a laser to illuminate the support surface) and a very small camera for image acquisition. The camera is a CCD (*Charge Coupled Device*) sensor with a resolution of approximately 15x15 pixels on a surface of a few square millimeters. The output of the CCD is taken to a DSP (*Digital Signal Processor*) which carries out image processing. In practice, a comparison is made between the images detected at subsequent clocks, and the movement is calculated from the comparison, which is passed to the computer. All the logic in question, including the CCD and the interface to the computer is contained in a specific integrated circuit.

To function, unlike the first optical mice, the support surface must have irregularities, so that subsequent acquisitions differ from each other and the logic can determine if there has been movement. If the mouse is placed on a perfectly smooth surface, such as a sheet of glass, it will not work. Likewise, if you remove the mouse from the support surface and hold it up, it stops working.

In Figure 70 the cover is removed. The figure highlights (a) the LED lamp, with a small prism in front that reflects the light towards the surface below; (b) the integrated circuit (Agilent ADNS-1610) containing the motion recognition logic; and (c) the mouse wheel with which the page is scrolled (it can also be clicked). The figure also indicates the contact points relative to the right or left click, as well as the contact point that recognizes the click on the wheel.

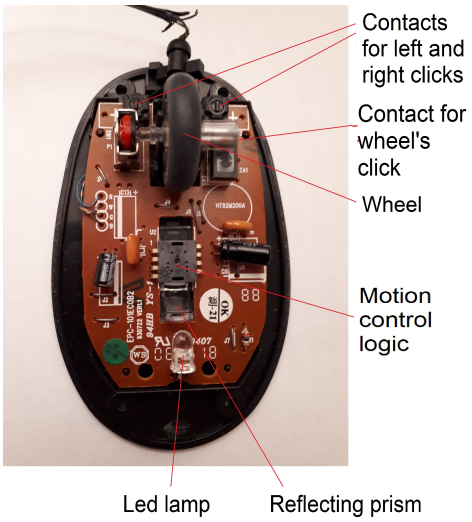


Figure 70: An optical mouse with the cover removed.

9. Conclusions

The described exhibit has been set up to evidenziate, the pace of advancement of the technologies related to computers. A large part of the described elements belonged, in the past, to the teachers in the Faculty; they were incorporated into the exhibit after they were abandoned due to obsolescence. The perspective is to continue to enlarge the collection, adding newer and, possibly, outstanding devices of the past.

Conflict of Interest The authors declare no conflict of interest.

References

[1] Wikipedia, "List of computer museums", https://en.wikipedia.org/wiki/List_of_computer_museums, accessed on 2024-07-31.

[2] F. Faggin, M. E. Hoff, S. Mazor, M. Shima, "The history of the 4004", *IEEE Micro*, vol. 16, no. 6, pp. 10–20, 1996.

[3] F. Faggin, "The making of first microprocessor", *IEEE Solid-state circuits*, vol. 1, no. 1, pp. 8–21, 2009.

[4] G. Bucci, "Un percorso espositivo attraverso la tecnologia", https://sites.google.com/unifi.it/calcolatori-elettronici, accessed on 2024-01-18.

[5] G. Bucci, "From a personal collection to an educational exhibit of the advancement of computer technology", "Proceedings of Histelcon 2021", pp. 54–57, 2021.

[6] Apple, "Apple unleashes M1", https://www.apple.com/newsroom/2020/11/apple-unleashes-m1/, accessed on 2024-01-18.

[7] Apple, "Apple introduces M4 chip", https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/, accessed on 2024-07-31.

[8] MIT, "Development of magnetic core memory, 1950", https://archivesspace.mit.edu/repositories/2/archival_objects/142711, accessed on 2024-01-18.

[9] Wikipedia, "Olivetti elea", https://en.wikipedia.org/wiki/OlivettiElea, accessed on 2024-01-18.

[10] Morse, Raveiel, Mazor, Pohiman, "Intel Microprocessors–8008 to 8086", *Computer*, vol. 13, no. 10, pp. 42–60, 1980.

[11] Intel, "Intel® 64 and ia-32 Architectures Software Developer's Manual, Combined volumes", https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html, accessed on 2024-07-31.

[12] AMD, "Amd Athlon™ 64 x2 Dual-Core Processor Product Data Sheet", January 1980.

[13] IBM, "Powerpc 750 microprocessors", https://www.ic72.com/pdf_file/i/23704.pdf, accessed on 2024-07-31.

GIACOMO BUCCI is an emeritus professor at the University of Florence (Italy). He graduated from the University of Bologna, Italy. From 1970 to 1982, he was with the University of Bologna. In 1975, he was a visiting researcher at IBM T.J. Watson Research Center, Yorktown Heights, New York. Since 1986, he has been a full professor at the University of Florence, Faculty of Engineering, where he taught a course in computer architecture and a course in software engineering. Prof. Giacomo Bucci has held several academic positions. His research interests included computer architecture, real-time systems, distributed systems, software development methodologies, and computer performance evaluation.

IMAD ZAZA is a system engineer and network administrator at the Department of Information Engineering (DINFO), University of Florence, since 2020. He is responsible for designing and developing systems that meet the needs of the research laboratory, collaborating with software developers, network engineers, and cybersecurity experts to ensure that the various systems are integrated and working together effectively. Imad Zaza got his Ph.D. degree in 2017, with a Ph.D. thesis entitled "Ontological Knowledge-base for Railway Control System and Analytical Data Platform for Twitter". Before joining DINFO he cooperated with the University of Florence and has been the author of more than 10 papers written in collaboration with professors therein.

Received: 13 September 2024, Revised: 17 October 2024, Accepted: 18 October 2024, Online: 04 November 2024
DOI: https://dx.doi.org/10.55708/js0311002

Secure Anonymous Acknowledgments in a Delay-Tolerant Network

Edoardo Biagioni*
Edoardo Biagioni, University of Hawai'i at Mānoa Department of Information and Computer Sciences, Honolulu, HI 96822, USA
*Corresponding author: Edoardo Biagioni, 1680 East-West Road, Honolulu HI 96822, USA, +1-808-956-3891 esb@hawaii.edu

ABSTRACT: TCP and many other protocols use acknowledgments to provide reliable transmission of data over unreliable media. **Secure** acknowledgments offer a cryptographic guarantee that valid acknowledgments for a given message can only be issued by the intended receiver. In the context of an ad-hoc network, **anonymous** acknowledgments make it hard for an attacker to determine which device issued a particular acknowledgment. And unlike TCP, the acknowledgments described here work well even for connectionless communications. This acknowledgment mechanism assumes that message data is protected by secure encryption. The sender of a data message includes in the encrypted part of the message a randomly-generated acknowledgment. Only the intended receiver can decrypt the message and issue the acknowledgment. The acknowledgment is issued by sending it out to its peers, who will forward it until it reaches the sender of the data being acknowledged. Such randomly-generated acknowledgments in no way identify senders and receivers, providing a degree of anonymity. This paper describes the use of such acknowledgments in both ad-hoc networks and Delay-Tolerant Networks. In such networks every peer participates in forwarding data, including both the routing and the end-host functionalities of more conventional networks. In a Delay-Tolerant Network, peers may cache messages and deliver them to other peers at a later time, supporting end-to-end delivery even when peers are only connected intermittently. Caches have limited size, so peers must selectively remove cached messages when the cache is full. As an additional aid to selecting messages to be removed from a cache, peers can remove messages for which they have received a matching ack. This can be done while preserving both security and anonymity, by including in every message, unencrypted, a message ID computed as the hash of the message ack sent encrypted with the message. A peer seeing a new ack can then hash it and discard any cached message whose message ID matches the hash of the ack.

KEYWORDS: Ad-Hoc Networks, Delay-Tolerant Networks, Security, Anonymity, Confidentiality.

1. Introduction

The acknowledgement, commonly abbreviated **ack**, is common in networking protocols that seek to provide reliable transmissions over an unreliable medium. TCP relies on acks to confirm receipt of data transmitted on a connection, and also to grant permission to send further data on that connection. Because of this, acks are an essential part of TCP data transmission. TCP acks each have a 32-bit ack number indicating the sequence number of the next byte of data expected on the connection. In a connection transmitting data in just one direction, acks flow in the direction *opposite* the flow of data.

The expected value of a TCP ack can be computed by anyone observing the flow of TCP traffic, so attackers may create and transmit spoofed acks [1].

This paper describes an ack mechanism that provides:

- a guarantee that a message has been delivered, since only the intended receiver can transmit a valid ack for the message (Sections 3 and 4). The intended receiver is securely identified by its ability to decrypt the message.
- a measure of anonymity, in that the ack message in no way identifies either the sending or the receiving device (Section 3). This offers more protection against traffic analysis compared to protocols where the ack carries source and destination addresses.
- the ability to reliably and securely delete acked messages from message caches, even on peers that have no access to any secret key (Section 3).
- communication to the sender that partial receipt of a message is sufficient for the receiver (Section 5).

2. Background

2.1. Secure Hashing

A hash function is a function that computes a fixed-length bitstring, called the hash, from a variable-length input bitstring. In most applications it is beneficial if the hash bears no resemblance to the input. Such hashes are used, for example, in hash tables to evenly distribute the indices to which keys are assigned.

A secure hash or cryptographic hash [2] is a hash that is hard to invert, which means that, given a hash, it is hard to create an input that hashes to that given hash value. Such hashes have many security applications, and several hash algorithms have been standardized by government

agencies such as the US National Institute of Standards and Technology (NIST) [3].

In brief, the result of hashing a bitstring with a cryptographic hash is the hash value or *digest*. As long as the security of the hash is strong, it would be very hard for an attacker who has no access to the original message, to create a bitstring that hashes to the same hash value. In contrast, anyone with access to the original bitstring can easily hash it and verify that it indeed matches the hash value that was received.

The algorithm described in this paper hashes an acknowledgment and sends in the clear the corresponding hash value, which in this paper is called a *message ID*. In the same message as the message ID, the acknowledgment (ack) is sent encrypted, such that only the intended recipient can decrypt it.

After the receiver has decrypted an ack, it can broadcast it. Any peer receiving this decrypted ack can hash it and verify that the previously seen message ID matches the acknowledgment. As long as the hash is hard to invert, only the intended receiver (and the original sender, who also has access to the unencrypted version of the ack) can broadcast a valid ack.

This is an algorithm that can be used with any encryption algorithm and any secure hash function, and in that sense is very general.

2.2. Delay Tolerant Networks

In an ad-hoc network, every peer communicates wirelessly with every other peer in its range. The network is ad-hoc because its connectivity may change with changing conditions. Each peer contributes what it can to data forwarding as well as to creating and receiving data, behaving as a combination of the roles of host and router in other networks.

Delay-Tolerant Network (DTN) technologies [4, 5, 6] support communication among peers that may only occasionally be in communication range of each other. Such conditions are common among mobile peers that communicate through an ad-hoc network. The purpose of DTNs is to deliver data even in the absence of any simultaneous end-to-end path between sender and receiver.

In the 1970s and 1980s email was often delivered even to peers that were never directly connected to the Internet [7]. For example, a peer H (Host) would from time to time dial up another peer G (Gateway) with better connectivity and which had agreed to cache email to and from H. The connection used a protocol called UUCP (unix-to-unix copy) [8] to download emails addressed to the users of H, and to upload emails originated by users of H. H might then in turn forward emails to other peers that depend on it for connectivity.

To support this intermittent email delivery, G had to save messages addressed to users of H and other peers that connected to H, delivering them on request. G would likewise save outgoing messages originating from H and others until G itself could connect to its upstream peer (or directly to the wider Internet) to deliver these messages.

These techniques allowed email to be delivered even if neither sender nor receiver were ever directly connected to the Internet, as long as they were able to connect to someone else with either an Internet connection, or closer to another host that was connected to the Internet.

DTNs have similar goals as the old uucp email system, but delivering general-purpose messages rather than just email. Intermediate peers in a DTN cache messages and deliver them on request. Typically the intermittent connection is established when one of the peers moves into wireless range of another peer.

To accomplish this delivery, in a DTN all devices are peers, so each device must include all the functionality that in a more conventional network is divided among data sources, data sinks, and routers.

When a sending and a receiving DTN node both have access to the Internet, messages can be delivered directly from the sender to the receiver. When the sender, the receiver, or both are not connected to the Internet, they may still be able to communicate directly with each other over ad-hoc links. If at any given time there is no path between sender and receivers, all peers reachable by the sender cache the message, in case they are able to deliver it later.

As connectivity changes and allows communication with new nodes, peers can forward their cached messages to any new peers, with the goal of eventually delivering each message to its final destination.

In their most extreme form DTNs are only useful for data that is not delay-sensitive. Delay-sensitive communications can still occur over ad-hoc networks as long as the devices have an end-to-end simultaneous path over any combination of the Internet and ad-hoc networks, whereas delay-tolerant communications can be supported even without simultaneous end-to-end connectivity.

2.3. A Useful Delay Tolerant Network: AllNet

One useful application of DTNs is delivery of text messages (chat) among mobile devices. Users carry their mobile devices even to locations with no or intermittent connectivity[1]. It would be desirable for users to be able to communicate with at least their neighbors even in the absence of Internet connectivity. Where such connectivity is intermittent, cached messages can be forwarded once connectivity is available. This way of sending text messages resembles at a high level the uucp email delivery described above, but the details are very different, especially the unpredictable availability of wireless channels compared to the scheduled uucp connections over telephone modems.

The benefit of a chat application is that data requirements are moderate, while the usefulness can be very substantial even in situations where delivery of individual chat messages is delayed.

Creating and using ad-hoc networks and DTNs for such applications is the main goal of the AllNet project [9]. AllNet is designed to work whenever devices can communicate directly among each other even in the absence of cellular or Internet service. Available technologies include point-to-point (infrastructure-less) WiFi and the many variants of Bluetooth, particularly Bluetooth Low Energy (BLE). All

[1]The lack of connectivity may be due wilderness adventure, foreign travel, emergency situations, rural areas, or any other reason, such as being outside of the provider's coverage.

of these are available on popular mobile devices, though in many cases the operating system imposes idiosyncratic restrictions on their usage.

Since peers that might forward a message also have the ability to inspect the message, AllNet encrypts the contents of interpersonal messages to prevent eavesdropping. Communication in AllNet needs to identify neither sender nor receiver. Messages may carry optional addresses, only used to improve efficiency of message delivery.

AllNet authenticates users to each other when they are within direct communication range of each other, or when they both know a secret string that allows them to authenticate to each other over the Internet. Such an exchange creates keys that the parties can use at any time thereafter, with assurance that they really are communicating with each other.

AllNet provides some anonymity of communication both with the addresses being optional, and to a lesser extent by the use of ad-hoc communication. AllNet provides such anonymity without being vulnerable to DDoS amplification attacks [10].

The optional nature of addresses in AllNet supports reasonable tradeoffs between anonymity and performance. A message with 0 significant bits of address is anonymous, and so may be delivered to all peers within reach. On the other hand, intermediate peers that care about performance, including especially bandwidth and battery life, may be more willing to forward messages that carry more significant bits of destination address than messages with fewer bits of destination address, since such messages may be delivered more precisely, ultimately consuming fewer network resources. The sender of a message can then choose fewer bits of address to give greater anonymity, or more bits of address to give greater likelihood of message delivery. This choice may be made dynamically, based for example on network traffic, assuming that less overall traffic implies more chances of delivery for anonymous messages.

Ethical Statement: Like every other security feature, encryption, authentication, and anonymity are intended to protect some people from other people. Like every other security feature, anonymity can be used to shield ethical behavior or unethical behavior. This paper does not attempt to distinguish such uses. In general, purely technical work cannot favor ethical over unethical uses of technology.

As is true for many other protocols (including `https`), encryption, anonymity, and authentication are likely to encourage people to use the technology, and their absence would likely discourage people from using technologies such as AllNet. For an ad-hoc network these security properties are essential since there is no assurance that ad-hoc peers, who might forward all of one's messages, will be friendly.

3. Acknowledgments and Caching in an Ad-Hoc Network

Since DTNs work best with all-to-all delivery of anonymous messages, conceptually each peer in a DTN has to cache all messages. Not only do caches have storage limitations, exchanging cached data with every peer that one encounters may require substantial spectrum and too much energy from a limited battery. For all these reasons a peer that caches messages should be informed when one of its cached messages has been delivered to its final destination. AllNet does this by sending a small ack message that confirms receipt for each data message that has reached its final destination.

There are many differences between TCP acks and AllNet acks:

- an AllNet ack gives evidence that an application, rather than the transport layer, has received the message. In TCP, even data that has been acked by the receiving system may never be delivered if the application crashes or stops reading the socket.
- attackers cannot spoof AllNet acks, since only the intended receiver can decrypt the message and issue the corresponding ack. Technically, the sender of the message could also issue the ack, but a legitimate sender by definition is not an attacker.
- in TCP, acks are normal TCP segments that may or may not carry user data. In AllNet acks are 16-byte (128-bit) random strings. As long as the acks are randomly generated, by the birthday paradox the chances of a collision are small as long as there are substantially fewer than $2^{64} = 18,446,744,073,709,551,616$ acks.
- AllNet acks are anonymous. Nothing in the 16-byte random string identifies either the sender or the receiver.
- since an AllNet ack message may carry multiple ack values, a single ack message can acknowledge data messages from different conversations at once. A TCP cumulative ack may acknowledge multiple segments at once, but all such data segments were sent on the same connection.

Each AllNet peer receiving an ack message caches the acks it contains. The peer also hashes the ack, giving the message ID of the message it is acking. If the message ID matches any message that this peer has originated, that message is marked as acknowledged. Also, any matching data message in the peer's cache no longer needs to be cached.

AllNet peers forward ack messages to their ad-hoc peers and across the Internet. Unlike data messages, acks from different messages and from the ack cache may be combined and forwarded together in a single ack message.

When an AllNet peer retransmits a data message for which it has not received an acknowledgment, it may get a matching ack from a peer other than the final receiver, if that peer has cached an ack for that message.

Ack transmission in AllNet is no more efficient as the delivery of the data message, but acks are much smaller than most data messages, so any overhead is less, and likewise the need to evict acks from caches is less – a same-sized cache can hold many more acks than data messages.

Since acks in most systems (including both TCP and AllNet) are idempotent, meaning that receiving the same ack once has the same effect as receiving it multiple times, duplicate transmission of acks has no consequences beyond the cost of transmission.

Acks are also less important than data messages. In contrast to a dropped data message, the worst possible result of

a dropped ack is the sender not knowing that the message has reached its destination and transmitting a duplicate copy of the message, whereas a dropped data message may have the potentially much more serious result of failure to communicate.

4. Technical Details: Secure Acknowledgments

An attacker can easily generate spoofed TCP acks by observing any part of the connection traffic [1]. As a result TCP is not secure, and higher layers such as TLS must be used to provide some assurance of delivery to the intended party.

AllNet provides by design many of the features provided by the combination of TCP and TLS, but for connectionless decentralized ad-hoc networks and DTNs. The security of these transmissions can only be guaranteed if the intended receiver of a message is the only system able issue the corresponding ack.

In AllNet, each receiver holds one or more cryptographic private keys that it uses to decrypt messages addressed to itself. As mentioned above, AllNet certifies keys based on interpersonal interactions among users, whereas TLS relies on hierarchically issued certificates that are vulnerable to hackers penetrating the systems that issue certificates [11].

A sender generating a data message for a specific receiver includes the ack in the plaintext part of the message before encrypting it, at which point the ack is included in the encrypted part of the message.

The sender then adds to the unencrypted part of the data message the cryptographic hash of the ack.[2] This hash is known as the message identifier or **message ID**. The message ID, like the ack, is very likely to be unique for each message.

The entire process is shown in Figure 1.

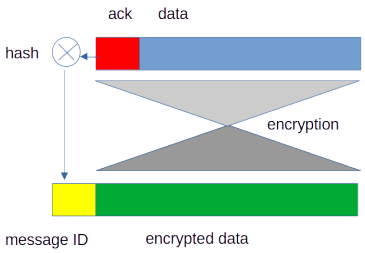


Figure 1: Relationship between ack and message ID

Every peer can see the message ID, but cannot generate a valid ack without access to the receiver's key, so only the intended receiver [3] can generate a valid ack.

A peer that receives an ack can verify whether it matches any of its cached messages or any incoming message by hashing the ack to give the corresponding message ID, then comparing this message ID computed from the ack to the message ID which is in the unencrypted part of every cached or received message.

Since the original sender also has access to the cleartext ack, the sender could also cancel messages that are cached, but this feature is unused both currently and in the foreseeable future.

4.1. Anonymity of Messages and Acknowledgments

Each source and destination address in AllNet messages is defined by a number of significant bits specified as part of the message header. A message whose addresses have no significant bits carries no information about the sender or intended receiver, and as such every peer on the network will try to decrypt it, so that any peer able to decrypt a message is an intended recipient.

The concern with such broadcast messages is the resource usage in having all peers forward and attempt to decrypt every message. The significant bits mechanism of AllNet addresses allows any number of bits to be specified, allowing a sender to specify a small number of bits to reduces the resource usage of the network while still preserving some anonymity. Senders are incentivized to provide as many bits of address as will still retain anonymity, and thereby minimize network resource usage, since packets with more bits of address are more likely to be delivered.

Just as messages can be anonymous, so acks in AllNet are also anonymous in that the ack itself carries no information about the sender and receiver of the ack. In addition, since any peer in the network might have cached the corresponding message, to the extent possible acks are distributed to every peer in the network. This universal distribution makes acks more resistant to traffic analysis than if they were only delivered back to the originator of the data message.

5. Acknowledgments for Messages Larger than the Maximum Transmission Unit (MTU)

The secure acks described so far allow for interesting possibilities when the size of a message is greater than a network's Maximum Transmission Unit (MTU), requiring that the message be sent as a collection of packets instead of a single message. Such large messages can be used to send multimedia data such as audio, images, and video.

The Internet Protocol (IP) uses a mechanism called **fragmentation** [13], and this paper uses the same term. The receiver can reconstruct the larger message once it has received all of the fragments.

In AllNet, each fragment of a larger message carries two encrypted acks, one for the message as a whole and the other for the specific fragment. Correspondingly, the unencrypted part of each fragment contains both a message ID obtained from hashing the message ack, and a fragment ID (which AllNet calls a packet ID) obtained from hashing the fragment ack.

A receiver receiving fragments of a larger message may ack them individually. Once the receiver has received all the fragments of a message, it then issues the ack for the entire message. Each peer receiving a message ack can clear from its cache every fragment it has of the larger message, even if it is only caching some of the fragments.

[2] Normally such hashes produce more than 16 bytes of data, so the sender only includes the first 16 bytes of the hash.

[3] Or anyone who can invert the hash function, which is expected to be challenging for strong cryptographic hash functions such as the SHA-512 hash [12] used by AllNet.

5.1. Acknowledging Partial Transmission

The combination of message acks and fragment acks provides some functionality beyond what traditional TCP segment acks provide. Specifically, if the content can be delivered without delivering all the fragments, then the receiver can issue the message ack immediately even with some fragments still missing.

As one example, email is often sent as both a plain text version and an html version. If either part of the message is received in its entirety, that part can be displayed to the user without having to receive every fragment of the other part, and the receiver can issue the message ack immediately.

When sending image or video data, the resolution of the image or video usually does not need to exceed the resolution of the device. Sending a low-resolution image first, followed by the high-resolution image, allows a low-resolution device to immediately send the message ack, without having to wait for or process the high-resolution image, and likewise allows the sender to only send the low-resolution version. The sender can immediately start sending the high-resolution version if it gets fragment acks for all the fragments of the low-resolution version, but no message ack.

For transmission of more general data (beyond videos and images) there are many schemes for forward error correction (FEC) that involve sending redundant data. If this data can be fragmented appropriately so that the message can be reconstructed even while some fragments are still missing, then a receiver that sends the message ack as soon as it is able to reconstruct the message can avoid having the sender engage in retransmission of any missing fragments.

There are many FEC algorithms. A particularly simple (and inefficient!) FEC scheme simply transmits each fragment 3 times. A receiver that obtains at least one copy of each fragment can immediately issue the message ack, providing reliable transmission without retransmission even in the case of substantial packet loss or transmission delay. Avoiding retransmission is particularly useful in Delay-Tolerant Networks.

If synchronous communication is available, the sender may receive the message ack for such a triply-redundant transmission before sending all three copies of all of the fragments, and can immediately stop transmitting the duplicate/triplicate information. On the other hand in a DTN, a receiver may over time receive a random subset of the fragments, and can then deliver the message to its application and issue the ack as soon as it receives all the fragments needed to reconstruct a complete message. This message ack lets other peers remove from their caches even fragments that the destination has never received.

6. Comparison to TCP Acknowledgments

Section 4 described how AllNet secure acks reliably assure the sender that messages have indeed been received by the intended recipient [4]. This is substantially different from TCP acks, which can be issued by any attacker that knows the sequence numbers in use on the connection and can spoof source IP addresses. A TCP sender has no way to know that such acks are not from the legitimate intended receiver.

Assurance that the ack was issued by the intended receiver is especially valuable in ad-hoc networks and DTNs: since the network is not organized by an authority, there is no reason to believe that intermediate peers are benign. AllNet, as other secure ad-hoc networks, had to be designed assuming that attackers may control at least some of the peers that are forwarding messages and acks.

Further differences between AllNet secure acks and TCP acks follow:

TCP acks carry the sequence number following the last byte that was received. This makes TCP acks cumulative, meaning that a single ack can acknowledge many segment's worth of data, and that loss of a single ack in a continuing stream of data transmission is not likely to lead to retransmission. On the other hand, since TCP acks count bytes rather than messages/packets/segments, and since TCP ack numbers are 32 bits, an ack in the original TCP can acknowledge up to 2^{32} different bytes, or an ack using the Protection Against Wrapped Sequences [14] (PAWS) mechanism can theoretically acknowledge up to 2^{64} different bytes of data.

Unlike TCP acks which count bytes, AllNet secure acks identify packets, so a single 1,000-byte message requires only one distinct secure ack in AllNet but consumes 1,000 sequence and ack numbers in TCP. This is of interest when we consider how many outstanding unacknowledged messages can be handled by each protocol.

The AllNet secure acks are not counters, so cannot be used as cumulative acks (where one ack potentially acknowledges many, many data packets) as in TCP, but one single AllNet message ack can acknowledge many different fragments.

Since there are 2^{128} possible different randomly selected AllNet message acks, the birthday paradox tells us that the chances of collision is extremely low until the number of acks begins to approach $\sqrt{2^{128}} = 2^{64}$.

Unlike TCP, these secure acks might collide with acks from any sender, whereas the TCP connection mechanism ensures that sequence and ack numbers can only overlap within a connection. These differences are summarized in Table 1.

Table 1: Comparison of what TCP and AllNet acks can distinguish.

Protocol	Can reliably distinguish
Original TCP	2^{31} bytes in a window
TCP with PAWS	2^{63} bytes in a window
AllNet	2^{64} simultaneous messages

6.1. Performance Analysis

Given a 64-bit sequence number space for TCP and with reasonable assumptions against delivery of old packets, TCP is guaranteed not to have sequence or ack collisions as long as 2^{63} or fewer bytes are transmitted on a connection within a two-minute Maximum Segment Lifetime (MSL) period, leading to a maximum bandwidth of over 2^{63} / $120s$ = 7×10^{16} bytes/second for each TCP connection.

For the secure acks described in this paper, to stay well away from the birthday paradox we assume that it would be desirable to have no more that about 2^{60} unacknowledged

[4] As long as the recipient's key and the encryption algorithm have not been compromised.

messages in the network at any given time. We further assume message sizes of 1,000 bytes and a maximum message lifetime (as specified by the message expiration option in AllNet) of about a week or 604,800 seconds. Satisfying these assumptions limits the entire network to about $2^{60} \times 1000$ / 604800 = 10^{15} bytes per second.

For communication across the Internet a message lifetime of a week is excessive. Using the same 2 minutes as for TCP, the network can support almost $2^{60} \times 1000$ / 120 = 10^{19} bytes per second.

While this throughput for the entire network cannot be directly compared to the TCP per-connection throughput, it is quite adequate for both the current, largely experimental AllNet, and for any foreseeable developments. Ad-hoc networks are typically small and relatively inefficient [15], and even in the imaginable future are unlikely to scale to large sizes and large amounts of traffic. Therefore for the ad-hoc side of the AllNet communications, even the lower network throughput derived by assuming a 1-week message lifetime is very abundant.

Should the capacity of AllNet ever become an issue due to the limited number of bits in an ack, AllNet could evolve, as TCP already has, to use more acknowledgments bits.

6.2. Performance Results

As in TCP, AllNet acks are sent unreliably, and can therefore be lost. Again as in TCP, the mechanism for requesting an ack retransmission is to retransmit the original message.

Unlike in TCP, for messages that are retransmitted, any intermediate host that has cached the ack can resend it immediately, without the original receiver having to respond. As compared to networks that support TCP and therefore require continuous end-to-end connectivity, for DTNs this increases reliability of ack delivery.

Decrypting a message to generate the ack is more time consuming than the process to generate a TCP ack. TCP has the additional performance advantage of being implemented in the kernel.

With these caveats, we compared the time to return an ack in AllNet between two hosts connected at distant locations across the Internet, with a ping time of 83ms. The first ack (really, SYN+ACK) from the TCP connection establishment 3-way handshake took 125ms (measured using `tcpdump`). Sending a message to an existing contact with AllNet between the same two hosts returned the ack in 154ms (measured using AllNet's `trace` program). These are comparable results.

7. Future Work and Conclusions

It is clear from the above analysis that if AllNet ever becomes as popular as TCP, it will have to be redesigned or extended for higher performance, just as TCP has been and likely will be again in the future. Moving from 16-byte acks to 32-byte acks would address any foreseeable performance limitation due to ack size.

The secure acks described in this paper provide many advantages over conventional acks such as used in TCP. This paper has explored a few, including the guarantee that the ack can only be issued by a receiver that has the correct cryptographic key, the ability to use the combination of message ack and fragment ack to let the sender know how much of the data the receiver actually needs, and the use of the acks to securely enable removing acknowledged messages from peer caches.

Secure acks as described above only work when sent encrypted. Encryption of data messages is pervasive in AllNet, so sending an encrypted ack with the encrypted data of a message requires no additional effort.

Should there be a reason to send the data unencrypted, the ack itself could still be encrypted, as long as keys are distributed in such a way that only the intended recipent can issue the ack.

When encryption is not an option, one could instead imagine having identical ack generators (for example, some kind of secure random number generator) based on identical secret seeds on each pair of sender and receiver, such that the receiver can generate the same sequence of acks as the sender. In such a case, the acks do not need to be transmitted at all. Instead, the sender would include a counter or an identifier for the ack associated with a message, and the receiver can independently generate the ack and hash it to compare the locally generated acks to any received message IDs. The receiver can then issue a valid ack that can be verified by any peer that is caching messages. While these acks do acknowledge receipt, they cannot be used by peers to discard cached messages.

Alternatively, in a scheme somewhat resembling the Bitcoin blockchain [16], and if anonymity is not a concern, a receiver could digitally sign a received message ID with a widely known public key. This scheme requires an infrastructure (perhaps a blockchain?) to record and distribute the public keys, but does not require encryption. Since a shared blockchain requires a persistent connection to the Internet, this scheme is more suitable for systems that can rely on Internet connections than for systems that use ad-hoc and delay-tolerant communications.

We have explored some of the design space for secure and anonymous acknowledgments. While this section has indulged in speculation, the mechanisms in Sections 3 and 4, and in the initial part of Section 5, are fully implemented and live on the AllNet network.

References

[1] Hastings, McLean, "TCP/IP spoofing fundamentals", 1996 International Phoenix Conference on Computers and Communications, doi: 10.1109/PCCC.1996.493637

[2] NIST, "Secure Hash Standard (SHS)", FIPS 180-4, August 2015.

[3] NIST, "SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions", FIPS 202, August 2015. https://csrc.nist.gov/pubs/fips/202/final

[4] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets", SigCOMM, Aug 2003.

[5] Benhamida, Bouabdellah, Challal, "Using delay tolerant network for the Internet of Things: Opportunities and challenges", 2017 8th International Conference on Information and Communication Systems (ICICS), 2017, doi: 10.1109/IACS.2017.7921980

[6] Mallorqui, Zaballos, Serra, "A Delay Tolerant Network for Antarctica", IEEE Communications Magazine, August 2022, doi: 10.1109/MCOM.007.2200147

[7] Partridge, "The Technical Development of Internet Email", IEEE Annals of the History of Computing, vol. 30, no. 2, April-June 2008, doi: 10.1109/MAHC.2008.32

[8] Nowitz, "Uucp Implementation Description", Unix Manual Version 7. https://web.archive.org/web/20180221100921/http://a.papnet.eu/UNIX/v7/files//doc/36_uucpimp.pdf

[9] Biagioni, "Ubiquitous Interpersonal Communication over Ad-Hoc Networks and the Internet", 47th Hawaii International Conference on Systems Sciences), in January 2014, and other papers at https://alnt.org/

[10] Biagioni, "Preventing UDP Flooding Amplification Attacks with Weak Authentication", International Conference on Computing, Networking and Communications (ICNC 2019), February 2019, Honolulu, Hawaii.

[11] Google Security Blog, "An update on attempted man-in-the-middle attacks", August 2011. https://security.googleblog.com/2011/08/update-on-attempted-man-in-middle.html

[12] Penny Pritzker, "Specifications for the Secure Hash Standard", U.S. Federal Information Processing Standards Publication 180-3, October 2008.

[13] Information Sciences Institute, "Internet Protocol", RFC 791 (section 3.2.1.4), September 1981.

[14] Borman, Braden, Jacobson, Scheffenegger, "TCP Extensions for High Performance", RFC 7323, September 2014.

[15] Gupta, Kumar, "The Capacity of Wireless Networks", IEEE Transactions on Information Theory, vol. 46, March 2000.

[16] Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", made public May 24 2009. http://bitcoin.org/bitcoin.pdf

Edoardo Biagioni did his bachelor's degrees at MIT in 1982, diplom at ETH Zürich in 1985, and PhD degree in computer science at the University of North Carolina, Chapel Hill, in 1992.

He is currently an associate professor in Information and Computer Sciences at the University of Hawai'i at Mānoa, where his research includes networking and systems. He has been working on the AllNet project (www.alnt.org) since 2011.

Received: 11 September, 2024, Revised: 14 October, 2024, Accepted: 15 October, 2024, Online: 07 November, 2024

DOI: https://doi.org/10.55708/js03110003

Fuzzy-Based Approach for Classifying Road Traffic Conditions: A Case Study on the Padua-Venice Motorway

Gizem Erdinc*, Chiara Colombaroni, Gaetano Fusco

Department of Civil, Constructional and Environmental Engineering, Sapienza University of Rome, Rome, 00184, Italy

*Corresponding author: Gizem Erdinc, Dept. of Civil, Constructional and Environmental Engineering Sapienza University of Rome, Italy, +905434493938 & gizemerdincc@gmail.com

ABSTRACT: This study offers a fuzzy-based method for determining the variety of traffic conditions on roads. The fuzzy approach appears more appropriate than the deterministic technique for giving drivers qualitative information about the present traffic condition, as drivers have a shaky understanding of the traffic status. It was used in an analysis that included flow, occupancy, and speed measurements from the Italian freeway that runs between Padua and Venice. MATLAB is used in the application's development. The empirical findings demonstrate how effectively the suggested study performs in classification. The experiment can offer a straightforward and distinctive viewpoint for induction and traffic control on motorways.

KEYWORDS: K Fuzzy traffic congestion estimation, Qualitative traffic state classification, Fuzzy approach

1. Introduction

Transportation plays a crucial role in enabling us to move from one place to another, facilitating the exchange of goods and services, and connecting communities worldwide. There are many vital reasons for the importance of transportation, such as mobility, economic growth, globalization, supply chains, emergency services, tourism, recreation, urban development, and innovation and technology. As a result of the growing economy, people become more inclined to have a car, which leads to jam conditions in traffic on expressways, which is a fundamental element of the transportation system. According to the global economy, traffic congestion may cost billions of dollars by 2023, and this amount tends to grow [1]. Hence, estimating traffic conditions has become a significant problem in transportation engineering.

Severe congestion conditions negatively impact the level of serviceability (LoS) of road networks, leading to poor traffic flow efficiency, environmental degradation, economic loss, and road safety concerns [2]. In this regard, understanding the road traffic conditions as a whole and classifying them may guide traffic engineers to develop effective solutions to corresponding congestion mitigation measures or provide other feasible routes for travelers. Simultaneously, accurate expressway congestion identification can help improve the service level of the road network and encourage the intelligent and refined management of expressways. Therefore, further in-depth research on this topic is required. With in-depth research on traffic conditions, scholars have generally believed that the traffic state has fuzzy characteristics [3, 4]. Defining the traffic flow parameters that mirror the traffic state according to different data characteristics is crucial for traffic state identification [5- 7].

However, it remains a vague notion for which there is currently no agreed-upon definition or measurement. Defined it as a traffic flow state characterized by high density and low speed in [8]. Alternatively, it can be defined as a time-based measure, such as the additional time a vehicle spends in the network when it is unable to navigate at the free-flow speed level in [10] or as a definition of delay in [9]. Conversely, congestion has been interpreted as inadequate road space as an instance of a circumstance where demand surpasses supply on a roadway [11], or when cars collide due to an imbalanced speed-flow relationship [12]. Not to mention, the terms demand-capacity, delay-time, and cost have been used to characterize congestion in over ten distinct ways [13].

Current studies on traffic state identification can be classified into two groups: data- and model-driven. Data-

driven methods can be used for state identification problem without assumptions yet performing only with a large amount of reliable data. Some of common methods include the Bayesian Information Criterion (BIC) [14], Neural Network (NN) [15–16], Support Vector Machine (SVM) [17–19], and K-nearest neighbor (KNN) [20, 21]. While they can predict congestion levels by using artificial intelligence, they are less effective than fuzzy logic for classifying traffic congestion because of its limitations in dealing with real-time uncertainty, adaptability, and data requirements. For example, SVM operates on binary classification (e.g., congested vs. non-congested) or multi-class classification. It assumes clear boundaries between classes, which are not always present in traffic data. This makes SVM less suited for dealing with ambiguous or overlapping congestion levels. On the other hand, adapting a neural network to new traffic conditions requires retraining the model, which can be time-consuming and computationally expensive. Also, their internal workings, especially in deep neural networks, are difficult to interpret, which makes it harder for engineers to understand why the model classifies congestion in a particular way. This lack of transparency can be a major drawback in safety-critical applications like traffic management.

On the other hand, model-driven methods are used to form traffic flow depending on the traffic flow theory, such as Macroscopic Fundamental Diagram (MFD) [22], Lighthill–Whitham–Richards (LWR) [23-27], and Kalman Filter (KF) [28–31] models. Parameters such as speed or density are used to determine the condition of the traffic. Model-driven algorithms exhibit high explanatory power [32]. However, these models generally require idealistic presumptions that might not apply to certain real-world traffic situations and inefficient under fuzzy conditions.

Moreover, traffic is in a state of traffic flow characterized by fundamental variables together, that should be considered in the assessment. In the congested road conditions, the speed may change rapidly, and the congestion estimation based on only the detected speed value may be unstable. Since all three flow variables are known, this application represents the ability of fuzzy logic to handle the relationship between flow, density, and speed. Through all feasible solutions of those variables, the driver's behaviour can be modelled with linguistic definition. For that reason, this paper suggests that the fuzzy methodology might be more appropriate to provide qualitative information to drivers than a simple deterministic threshold-based method because the influence factors of traffic conditions are complicated, and each traffic condition has a certain similarity.

The present paper is structured as follows. After this Introduction, we gave traffic flow characteristics in the Second Section. The Third Section introduces the methodology for quantification. The case study and the findings are discussed afterward. The Conclusion and Discussions section resumes the main assumptions of the methodology implemented and the outcomes produced by the method while comparing them with LOSs by HCM.

2. Traffic Flow Characteristics

Mathematically, traffic is modelled as a flow through a fixed point on the route, analogous to fluid dynamics. In traffic theory, three parameters are typically used to describe traffic flow characteristics: flow (f = vehicle/h), density (k = vehicle/km), and speed (v = km/h).

Flow (f) refers to the number of cars that transit a specific cross-section in a given time unit in a particular interval.

As a measure of the efficiency of highway operations and a direct indicator of how easily cars can move through the traffic stream, density (k) is the characteristic most closely associated with the phenomena of congestion. This symbolizes the quantity of automobiles on a unitary long-road segment at a given time instant. In this study, the density is derived from the loop detectors' occupancy, and the flow and speed values are obtained from the sensors.

Another popular congestion indicator that illustrates the volume of vehicle movement on the road is Speed (v). It should be noted that congestion is a function of speed. Hence, considering speed as an index of congestion is straightforward and intuitive [33].

Equation 1 usually describes the general relationship between density and speed as a linear relationship, as in the Greenshields model [34]:

$$\bar{v}_s = v_l - \frac{v_l}{k_i} k \quad (1)$$

where v_s indicates the space mean speed, v_l indicates the free-flow speed, k indicates the density, and k_i indicates the jam density. The appropriate relationships for the flow and speed can be established as follows using this assumed linear relationship:

$$\bar{v}_s^2 = v_l \bar{v}_s - \frac{v_l}{k_i} f \quad (2)$$

where f is the flow rate. Similarly, the corresponding relationships for flow and density are as follows:

$$f = v_l k - \frac{v_l}{k_i} k^2 \quad (3)$$

Eqs. (2) and (3) show that parabolic relations are found between the flow and density as well as between flow and speed, assuming that the relationship is linear in the shape of Eq. (1) for speed and density. In short, Equation provides a formal expression for the relationship between all of the parameters at the point of equilibrium:

$$f = k * v \quad (4)$$

Traveler information systems find it difficult to give people precise advice since road congestion is not well defined, but a traffic status is characterized with a particular pair of values of the fundamental variables. There isn't a single, unified regulation as of yet [33]. However, there is a vast amount of data on traffic observations, and the phenomenon of traffic is widely recognized. The sole purpose of the example that follows is to serve as an instructive explanation of the ideas behind the fuzzy method that is described in the paper. As is widely known, direct traffic data show two distinct trends in the flow-density plane (Figure. 1a): The flow displays a very noisy and dense structure throughout the range of higher values of density, with an average declining tendency as the density increases. In the range of low levels of density, the flow is practically linear with density and presents few deviations from the average speed. It is commonly recognized that the high-density regime is typified by erratic flow conditions that are dictated by the microscopic mechanism that underlies the traffic stream. These fluctuations include even minor driving movements that result in a stop-and-go traffic pattern.

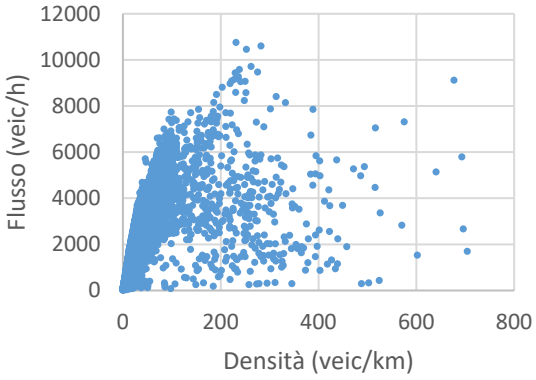


Figure 1a: Density-Flow relationship [35]

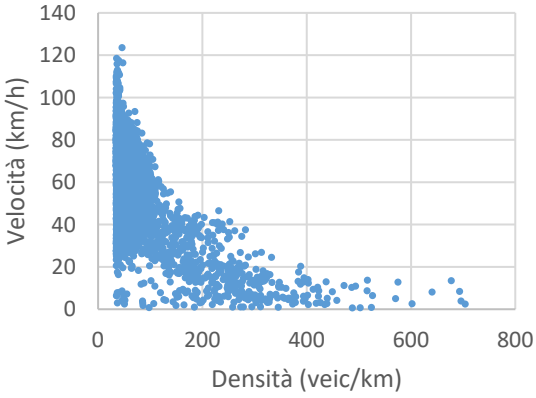


Figure 1:. Density-Speed relationship[35]

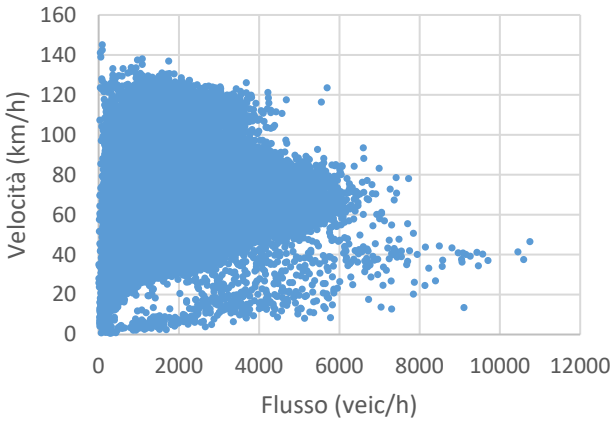


Figure 1c: Flow-Speed relationship [35]

Figures 1a, 1b and 1c illustrate a 15-min observation of the traffic of one of the sections for several months and it is evident from the figures that collecting traffic data for a long period continuously introduces a huge noisy component. The scope of the proposed model is to simulate the general state of the road and derive a relationship between the fundamental variables by applying the fuzzy Mamdani inference approach regardless of their value [33]. Each part has been modelled and executed independently because they could each have a distinct general state. The network consists of two main branches: the North (Highway) and the South (Tangential), the results of which are provided in Section 4.

3. Methodology

A traffic stream can be characterized by its current state and the change in its condition. To identify in which state the traffic is in, congestion should be quantified and detected how it is spreading. From this point of view, this paper is an evaluation of the average situation of traffic process in the related time interval as an index between [0-1]. The main purpose is to answer the questions *what is happening*, and *how critical is the current situation in terms of congestion condition level* which reflects the severity of traffic. It has been built with a fuzzy model in MATLAB.

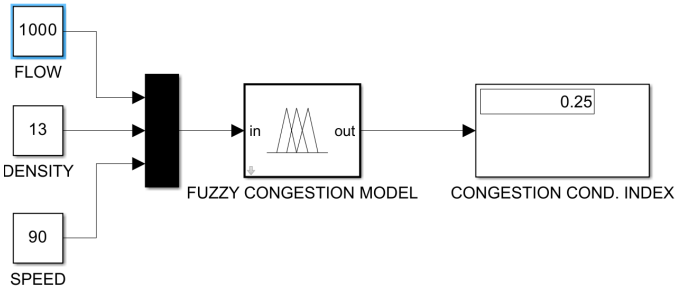


Figure 2: Evaluation of the current traffic condition.

The fuzzy model named 'Fuzzy Congestion Model', represented in the middle of the Figure 2, computes the average congestion level by using average flow, average density, and average speed information. All input and output variables have been defined by setting linguistic

terms which help us represent them with a wide range. We set $n_{x1} = n_{x2} = n_{x3} = 5$ linguistic terms as:

$$F = \{FF, RFF, AF, CF, VCF\},$$
$$K = \{VLD, LD, MD, HD, VHD\},$$
$$V = \{VS, S, A, F, VF\}.$$

The output variable is also classified as:

$$CCI = \{VL, L, M, H, VH\}$$

Explanation of variables are: FF: Free Flow, RFF: Reasonably Free Flow, AF: Average Flow, CF: Congested Flow and VCF: Very Congested Flow, VLD: Very Low Density, LD: Low Density, MD: Medium Density, HD: High Density and VHD: Very High Density, VS: Very Slow, S: Slow, A: Average, F: Fast and VF: Very Fast, VL: Very Low, L: Low, M: Medium, H: High and VH: Very High.

While linguistic definitions give the variables a thinking and understanding way to close humans, they need to represent numerically. In order to complete the process, all variables are converted from crisp numbers into membership degrees of the fuzzy set, or fuzzified. Membership functions are utilized in this process, providing the numerical value's degree of belongingness to the associated fuzzy set inside a closed interval [0,1]. Here, 1 expresses full membership and 0 states non-membership. In this paper, triangular membership (Equation 5) functions are used, since they are one of the most used examples and they are effective in reflecting the characteristics of the fuzzy sets used here.

$$\mu(x) = \begin{cases} 0, x \leq \alpha min \; or \; x \geq \alpha max \\ \frac{x-\alpha min}{\beta-\alpha min}, \; x \in (\alpha min, \beta) \\ \frac{\alpha max-x}{\alpha max-\beta}, \; x \in (\beta, \alpha max) \end{cases} \quad (5)$$

As an example of the mathematical form, the membership function of the *Medium level of Density (MD)* variable is provided in Eq. 6 and its demonstration is given with Figure 3.

$$\mu MD(k) = \begin{cases} 0, k \leq 11 \; or \; k \geq 30 \\ \frac{k-11}{20-11}, \; x \in (11,20) \\ \frac{30-k}{30-20}, \; x \in (20, 30) \end{cases} \quad (6)$$

Figure 3 shows the membership function plots of the density input variable. All the Very Low, Low, Medium, High, and Very High-density levels are the set members of K. Hence, the *MD* is a set member of the K fuzzy set. It has a membership degree of K fuzzy set that Eq. 6 gives us. This means it has a certain quantity of numerical belongingness to the K fuzzy set. That numerical value in [0-1] is calculated by Eq. 6.

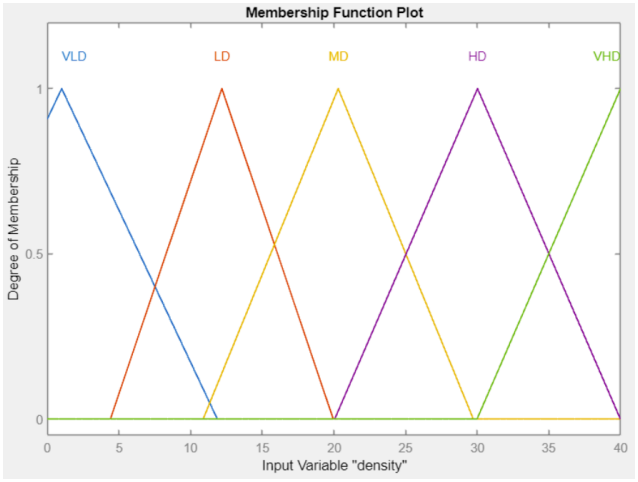


Figure 3: Membership functions for density variable

Similarly, Figure 4 shows the membership function plots of congestion condition level index output variable. In both plots (Figure 3 and 4), the horizontal plane shows the value range of the related variables, while the vertical plane gives the membership function value of it in the range [0-1]. The range intervals of the related variables are set after a close investigation according to the data. It is to note that the terms are imprecisely defined following the fuzzy logic concept, and there are no clear cuts between the fuzzy congestion levels.

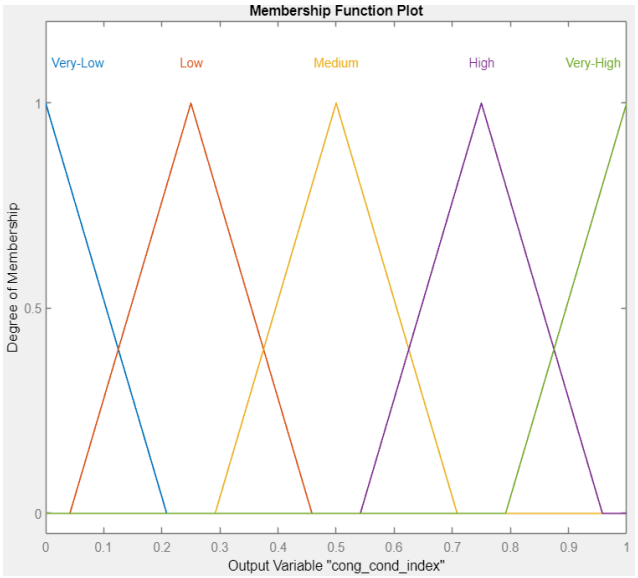


Figure 4: Membership functions for congestion condition variable

The fundamental aspect of the approach is to represent the input-output connection, which shows how input variables reflect onto the output universe, in order to construct the nonlinear surface model and make an inference. We define 45 rules for this application and they can be stated mathematically as:

$$(F \bullet F(f)) \; \theta min \; (K \bullet K(k)) \; \theta min \; (V \bullet V(v)) \; \rightarrow \; (A \bullet A(f, k, v))$$

where the symbol • states the linguistic term is, symbol θmin states the logic operator AND. Some of fuzzy if-then rules are given in Table 1 below. 45 rules have been modelled for this application.

Table 1: If-then rules for congestion condition index

IF	AND	AND	THEN
Flow is Free Flow	Density is Low	Speed is Fast	CCI is Very Low
Flow is Free Flow	Density is Medium	Speed is Average	CCI is Low
Flow is Congested	Density is High	Speed is Slow	CCI is High
Flow is Very Congested Flow	Density is Very High	Speed is Very Slow	CCI is Very High

The framework of the method is given in Figure 5. The figure shows the structure of the model, with input parameters, assigned linguistic variables to our crisp inputs, applied fuzzy rules and operators, chosen implication and aggregation method as well as calculated value of congestion as a defuzzied output.

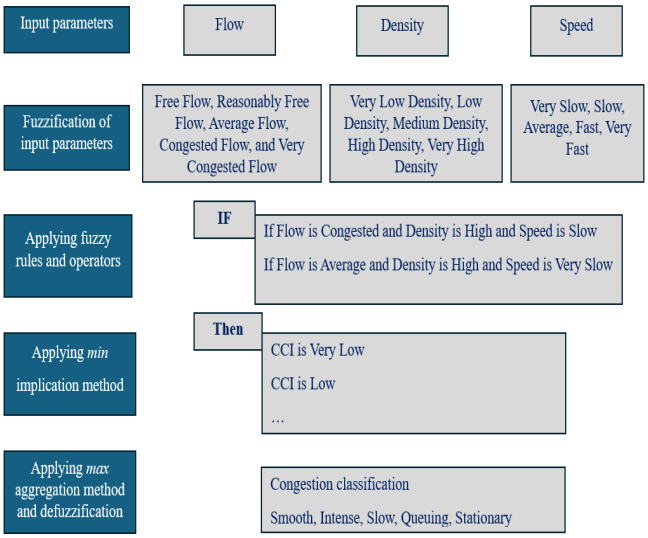


Figure 5: General framework of the study

Detailed information about methodological steps such as variable definitions, fuzzification, rules formulations and defuzzification can be seen in [36].

4. Application Results

The main scope of this study is to identify the traffic condition ranges of Padua-Venice motorway network, which is composed by 4 branches of 3-lane roads with separate carriageways, the motorway has a total length of about 74 km and 16 inter-sections. Real traffic data collected from 31st December 2018 to the 30th of August 2019 contains the following information: local unit code, code section counting, day type, road (section) id, date, flow, density, and harmonic speed aggregated every 15 minutes. The whole data was examined statistically section-by-section; however, it is important to mention that, due to temporary failures of some detectors, some parts of the data are missing. We have neglected them in the application step. In this research, we investigated the congestion from Monday through Saturday for eight months, during the hours of 6:30 am to 9:00 am. The application is run on MATLAB fuzzy logic toolbox R2020b and SIMULINK.

4.1. Traffic situation in the motorway (Northern part)

The first step of the application consisted in computing the speed predictions for 15-minute time intervals between 6:30-9:00 for all sections along the northern part of the network. Table 2 shows the congestion levels with average observed intervals, and Figure 6 depicts the traffic condition for all sections of the motorway.

A similar pattern, with intense or slow state levels, may be seen in the observations of traffic states with average intervals along the entire freeway. At section 43, the change to a slow traffic status is smooth. Sections 42 and 44, on the other hand, move to a more intensive condition. The point at which an intense condition transitions to a slow state is reached when flows exceed 1400 cars per hour and density passes 17 vehicles per kilometres.

Table 2: Traffic state results with average observed intervals

Section number	Road section	Average of the Observed Intervals			Congestion Condition Index	Traffic State
id	id	Flow (vehicle/hour /lane)	Density (vehicle/km /lane)	Observed Speed (km/h)	[0-1]	
24-1	742308	602	7	101	0.21	Intense
3-1	742456	1534	21	88.2	0.52	Slow
15-1	742250	1534	21	88.2	0.52	Slow
4-1	742463	930	11	86.6	0.25	Intense
5-1	742240	848	10	88.9	0.25	Intense
6-1	742483	1775	24	87.5	0.53	Slow
16-1	742476	1775	24	87.5	0.53	Slow
17-1	742230	1775	24	87.5	0.53	Slow
18-1	742215	2368	27	92.5	0.61	Slow
19-1	742496	2368	27	92.5	0.61	Slow
9-1	742202	1578	20	87.4	0.52	Slow
10-1	742508	1578	20	87.4	0.52	Slow
20-1	742512	1768	22	83.3	0.55	Slow
21-1	742194	1768	22	83.3	0.55	Slow
41-1	741467	1531	17	97	0.38	Slow
42-2	740855	1480	15	107	0.27	Intense
42-1	741527	1475	15	108	0.27	Intense
43-1	741555	1453	17	93	0.37	Slow
44-2	740700	1145	14	106	0.25	Intense
44-1	741682	1042	13	105	0.25	Intense
45-1	741708	1432	15	99	0.25	Intense
46-1	741741	973	11	98	0.24	Intense

In this case, the average speed decreases by around 15% from its previous level. This means that drivers can drive at around 67% of the free-flow speed level of the road. In such a traffic state, flow and density rates decrease to around 1150 vehicles/h and 14 vehicles/km, respectively. This allows the speed level to grow up to around 106.0 km/h, which equals more than 75% of its free-flow level, and it gives drivers more freedom to manoeuvre. There is no severe congestion and no Stationary state along the highway with this subset of data, because having an average level of flow and low level of density allows Fast level of speed with a high membership and Average and Very Fast levels of speed

with low memberships, so this situation occurs in an Intense state generally. It was coloured in Figure 6.

Figure 6: The condition of the motorway using a subset of data for the hours of 6:30 am to 9:00 am

4.2. Traffic situation in the highway (Tangential part)

The same computations have been done for all sections in the Tangential part of the network as well. This application is run with the same subset of data with 5-minute time intervals to criticize the effectiveness of the timeframe in characterizing the trend of rate changes. The observations of traffic states with average intervals all along the highway show a similar pattern, with intense or slow state levels. However, after having an intense situation, there is a clear transition to a slow state at sections 3 and 15 with an almost double amount of congestion indices. In this case, the average levels of flow and density go up to 2.5 and 3 times higher than the previous section's levels and that intensity causes a medium congestion and it is classified in the next congested state by the proposed model. Like the motorway part, there is no severe congestion and no Stationary state along the highway, as vehicles can move in a Fast level of Speed with a range between 85.5 km/h to 101km/h with a low membership, so this situation occurs in a slow state generally. Identified traffic states have been colored in Figure 7.

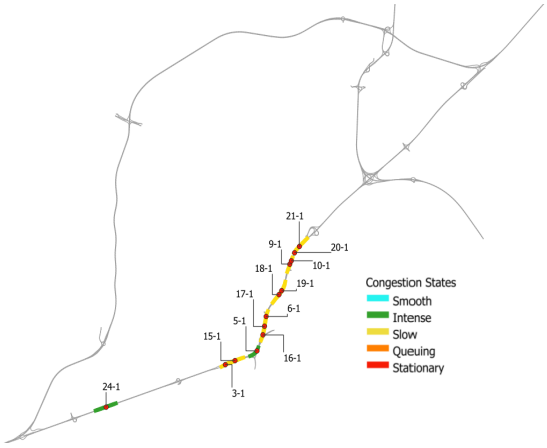


Figure 7: The condition of the highway using a subset of data for the hours of 6:30 am to 9:00 am

5. Discussion and Conclusion

5.1. Evaluation of congestion index by the fuzzy congestion calculator with flow, density, and speed variables

Table 3 summaries the results in evaluating congestion condition index (CCI) for some selected intervals of the field data including variables for density, speed, and flow. The intervals are chosen at random so that the traffic conditions encompass the entire spectrum of the traffic level index. The following lists the LOSs established by the Highway Capacity Manual [37] are shown for comparison. The detailed comparison with the LOSs was given in the next section.

Here, Table 3's observation (with a focus on average values) enables the following important conclusions to be drawn:

- Table 3 demonstrates that there is a comparatively large variation of density for the same speed. For instance, the densities for the sampling numbers 8, 9, 17, 24, and 99 at the speed level with v = 88-89 (km/h) are 8, 20, 18, 42, and 22 (veh. /lane/km), respectively.
- Similarly, there is a comparatively large range in speed for a given value of density. For instance, sample numbers 46, 52, and 89 have a speed range of [118–77] (km/h) for density k = 31 (veh./lane/km), but their flow range (veh./lane/hr) is rather wide (3588, 495, 375). These show that the traffic state cannot be adequately represented by a single variable.
- Additionally, Table 3 demonstrates the speed is not highly responsive to density alone at the "Very Low" and "Low" levels. For densities k = 2, 3, and 12, the corresponding average speeds (veh./lane/km) are 111.3, 106.4, and 99.0 (km/h). Speed drops by only 11.05% as densities rise six times from 2 to 12 (veh./lane/km). While the flow increases along with the density, the speed also reduces dramatically. For instance, when the flow hits 1800–2000 (veh. /lane/hr), as in indices of 69, 73, and 89, speed drops 26.6% to the level of 73–77 (km/h). This corroborates the idea that the traffic status is a specific tuple of the basic variables' values. As such, it requires contemporaneous attention.

The congestion indices generated by the FIS system are displayed in Figure 8 across with the LOS as described by [37]. Despite the fact that LOSs are connected to a comparatively broad range of congestion condition indices (CCIs), Figure 8 illustrates a general correlation between the CCIs and service levels. The CCIs of the various LOSs significantly overlap. It is possible to state that Very Low congestion (μVL = {-0.208 0 0.208}) is mainly linked with LOS A and partially with LOS B in

Table 3: Evaluation results

No	Section id	Unix time	f (veh/ hour /lane)	v (km/h)	k (veh/ km /lane)	CCI	traf_ state	LOS
1	742308	450	1872	112.2	17	0.390	slow	D
2	742202	0	1044	87.9	12	0.250	intense	C
3	740855	1140	2480	114.2	22	0.581	slow	F
4	742250	0	648	113.0	6	0.168	intense	A
5	742215	0	1155	96.1	12	0.250	intense	C
6	742496	0	520	101.5	5	0.157	intense	A
7	742230	400	1368	87.7	16	0.335	slow	D
8	742508	0	708	88.6	8	0.237	intense	B
9	742230	395	1716	88.4	20	0.500	slow	E
10	742250	80	120	100.7	1	0.899	smooth	A
11	742463	5	168	111.3	2	0.080	smooth	A
12	742496	1285	1520	97.0	16	0.375	slow	D
13	742250	0	312	90.7	4	0.100	smooth	A
14	742476	495	3696	75.0	51	0.500	queuing	F
15	742496	520	1440	97.8	15	0.356	slow	D
16	742194	1430	1296	90.6	14	0.306	slow	C
17	742240	945	1572	88.8	18	0.411	slow	D
18	742463	640	732	84.3	9	0.242	intense	B
19	742483	5	468	90.3	5	0.223	intense	A
20	742240	1435	1000	90.2	11	0.250	intense	B
21	742456	0	240	99.0	3	0.092	smooth	A
22	742496	245	600	78.7	8	0.224	intense	B
23	742202	415	2064	89.6	23	0.504	slow	E
24	742483	1050	3708	88.5	42	0.752	stationary	F
25	742508	1260	780	82.2	9	0.246	intense	B
26	742508	0	3084	18.9	25	0.678	queuing	F
27	741741	1065	852	112.0	8	0.215	intense	B
28	742512	0	270	90.4	3	0.100	smooth	A
29	742476	805	1812	76.8	24	0.589	slow	F
30	742194	1430	1692	82.1	21	0.538	slow	F
31	742240	1435	336	93.1	4	0.098	smooth	A
32	742308	0	360	107.3	3	0.083	smooth	A
33	742463	95	324	52.4	9	0.218	intense	B
34	742308	450	2424	110.4	22	0.575	slow	F
35	742230	395	1176	99.0	12	0.254	intense	C
36	742215	0	1440	90.7	16	0.356	slow	D
37	742463	610	1716	82.8	21	0.537	slow	F
38	742308	450	1290	108.7	12	0.297	intense	C
39	742496	520	3600	92.5	39	0.750	queuing	F
40	742215	0	2220	90.9	25	0.530	slow	E
41	742202	0	108	100.7	2	0.089	smooth	A
42	742476	1065	2892	76.4	38	0.774	queuing	F
43	742230	400	768	99.2	8	0.231	intense	B
44	740765	30	316	106.4	3	0.084	smooth	A
45	742215	5	600	95.6	6	0.216	intense	B
46	741527	510	3588	118.1	31	0.750	queuing	F
47	740700	150	64	125.0	4	0.071	smooth	A
48	742512	350	936	69.1	14	0.250	intense	C
49	742508	680	3048	71.8	43	0.785	stationary	F
50	742240	1430	444	93.0	5	0.223	intense	B

relation to the fuzzy partitioning of the CCI in Figures. 8 and 4. Low congestion (μL = {0.041 0.25 0.458}) is mostly linked to LOS C, sporadically to LOS B, and barely to LOS A. A little amount of medium congestion (μM = {0.30 0.5 0.708}) is linked to LOS C, and most of it to LOS D and E. Finally, Very High congestion (μVH = {0.791 1 1.208}) is strictly associated with LOS F.

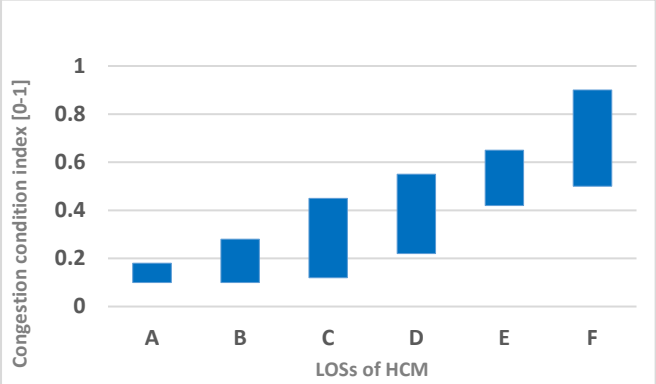


Figure 8: Correspondence between congestion condition levels and levels of service defined by HCM.

5.2. Comparison with the US Highway Capacity Manual's Service Level definitions

To provide additional insight into the significance of the suggested approach, the outcomes have been contrasted with the similar LOSs estimations derived using the HCM method [37], which establishes service traffic levels using the figure 9 diagram as a basis. Since the HCM speed-flow curves derive from observations conducted in the USA, they should only be used as a means of comparing different approaches and applied very carefully to an Italian motorway. The speed limit on the freeway under investigation is 130 km/h, which is higher than the ranges taken into account by the HCM. On the other hand, a LOS can be applied to the actual traffic states if a projection of the patterns of HCM curves is accepted. This experiment, however, shows that the combined observations of all three state variables are inconsistent with the theoretical model, which focuses on a univariate relationship between flow and speed and the state equation (3). It's true that if density is taken into account as the LOS identification indicator, then the recorded state of traffic on the road falling between 11 veh/km/lane (18 veh/mi/lane) and 17 veh/km/lane (27 veh/mi/lane) would be classified as either LOS C or LOS D. Nevertheless, the associated LOS will change from LOS B to LOS C if the flow values—which span an array of values from 973 veh/h/lane to 1531 veh/h/lane—are taken into account. More importantly, the density ranges from 7 veh/km/lane to 27 veh/km/lane on the highway would be associated with all levels from A to F with density guidance, while they would be considered in levels B to F for a huge range of flows between 602 veh/h/lane and 2368 veh/h/lane. These traffic conditions were categorized using the fuzzy technique as either slow or intense. The fuzzy method overcomes classification ambiguity caused by inherent mistakes created by the steady-state principles underpinning the HCM methodology and encompasses the inevitable ambiguity of traffic state classification because it takes into account all three state variables at once.

Furthermore, because the qualitative description makes use of common descriptors and is better suited for informing drivers, it is easier than a scale classification. A sensitivity analysis was conducted in terms of the condition of traffic range index between [0-1] to observe the impact of the chosen membership function of the suggested fuzzy model. For that analysis, five cases with triangular, trapezoidal and Gauss, Gaussbell, and Gauss2 function shapes have been carried out and the triangular membership function is stepped out as the best function type [36].

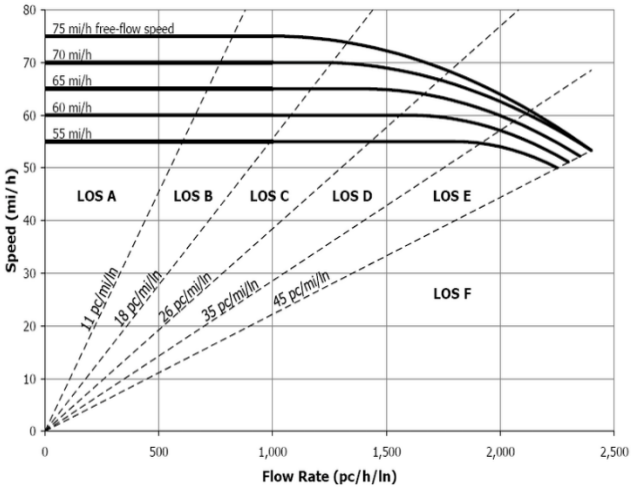


Figure 9: Speed-flow-density relationship [37]

5.3. Conclusion

This article handles the traffic state identification problem and proposes a Mamdani-based fuzzy logic application with a much bigger real dataset from a motorway in Italy. To address the value of multi-dimensional evaluation and we already possess all three-traffic flow variables, we examined the capability of fuzzy reasoning to simulate the relationships between variables and driving behaviour modelling. Moreover, predicted states have been discussed by comparing the LOSs by HCM. The fuzzy technique, which is based on a non-univocal range characterization of traffic circumstances, looks to be a more acceptable strategy for traffic congestion classification because LOS is a stepwise assessment and does not convey a concept of congestion level clearly. Since the fuzzy method considers all three

state variables, it takes into account the inevitable uncertainty in traffic state recognition and removes any ambiguity in categorization caused by inaccuracies resulting from the steady-state assumptions that underlie the HCM methodology. Furthermore, because the qualitative description makes use of common descriptors and is better suited for informing drivers, it is more straightforward than a scale classification. When properly designed using rules, the fuzzy-based congestion analyser can be used as a tool to help control actions on expressways under boisterous conditions. The current applications are in places like Singapore and Los Angeles [38-39]. These systems continuously monitor the state of the roads and modify the length of the signals. As a result, traffic flows more smoothly, stop-and-go situations are less frequent, and fuel consumption is decreased. Moreover, some Indian towns are experimenting with fuzzy logic-based systems that enable traffic signal modifications in real-time upon detection of emergency vehicles [40]. This facilitates the clearance of space for emergency vehicles and, by cutting reaction times in urgent situations, may save lives. By incorporating intelligence into traffic management, these solutions enhance effectiveness, security, and environmental sustainability.

However, due to the case study's static environment, this analysis limited to show severe and/or temporal congestion. For future works, the model will be extended with a consideration of this issues. Additionally, unanticipated events like accidents, weather changes, etc., will be handled to make developing control actions on motorways more convenient.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work has not been supported by any foundation.

References

[1] Y. Liu, C. Zhen, D. Huili, "Highway traffic congestion detection and evaluation based on deep learning techniques." *Soft Computing* vol. 27, no.17, 12249-12265, 2023, doi.org/10.1007/s00500-023-08821-6.

[2] C. Wang, C. Yuting, W. Jie, Q. Jinjin, "An improved crowddet algorithm for traffic congestion detection in expressway scenarios." *Applied Sciences*, vol. 13, no. 12, 2023, 7174, doi.org/10.3390/app13127174.

[3] T. G. Huang, L. H. Xu, and X. Y. Kuang. "Urban road traffic state identification based on fuzzy C-mean clustering." *Journal of Chongqing Jiaotong University (Natural Science)*, vol. 34, no. 2, 101-106, 2015.

[4] H.W. Liu, Y.T. Wang, X.K. Wang, Y. Liu, Y. Liu, X.Y. Zhang, F. Xiao, "Cloud Model-Based Fuzzy Inference System for Short-Term Traffic Flow Prediction", *Mathematics*, vol. 11, no. 11, 2509, 2023, doi.org/10.3390/math11112509.

[5] Z. Liu, C. Lyu, Z. Wang, S. Wang, P. Liu, Q. Meng, 2023 "A Gaussian-process-based data-driven traffic flow model and its application in road capacity analysis." *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 2, 1544-1563, 2023, doi.org/10.1109/TITS.2022.3223982.

[6] K. Zhao, G. Dudu, S. Miao, Z. Chenao, S. Hongbo, "Short-term traffic flow prediction based on VMD and IDBO-LSTM." *IEEE Access* 2023, doi.org/10.1109/ACCESS.2023.3312711.

[7] J, Ying, "Research on methods of urban road traffic condition identification", *Qingdao Univ. (Eng. Technol. Ed.)*, vol. 27, no. 3, 84–87, 2012, doi.org/10.1007/978-981-19-1057-9_26.

[8] P.H. Bovy, I. Salomon, "Congestion in Europe: Measurements, patterns and policies." *Travel behaviour: spatial patterns, congestion and modelling*, 143-179, 2002, doi.org/10.4337/9781781951187.

[9] A. Almeida, S. Brás, S. Sargento, I. Oliveira, "Exploring bus tracking data to characterize urban traffic congestion". *Journal of Urban Mobility*, vol. 4, 100065, 2023, doi.org/10.1016/j.urbmob.2023.100065.

[10] M.A.P. Taylor, "Exploring the nature of urban traffic congestion: concepts, parameters, theories and models." In *Proceedings, 16th Arrb Conference, 9-13 November, Perth, Western Australia*, Vol. 16, No. 5. 1992.

[11] I. Pedram, S. Salek, R. Omrani, *Defining and Measuring Urban Congestion—Guidelines for Defining and Measuring Urban Congestion*, Transportation Association of Canada: Ottawa, ON, Canada. 2017, Access (December 2020).

[12] ECM, *The spread of congestion in Europe*, 11th Round Table on Transport Economics, OECD Publication Service, 1999.

[13] M. Aftabuzzaman, "Measuring traffic congestion - a critical review", *Proceedings of the 30th Australasian Transport Research Forum*, 16, 2007.

[14] J. Xia, W. Huang, J. Guo, "A clustering approach to online freeway traffic state identification using ITS data", *KSCE Journal of Civil Engineering*, vol. 16, 426-432, 2012, doi 10.1007/s12205-012-1233-1.

[15] B. Priambodo, A. Azlin, A.K. Rabiah, "Prediction of average speed based on relationships between neighbouring roads using K-NN and neural network.", 18-33, 2020, doi.org/10.3991/ijoe.v16i01.11671.

[16] N. Tilahun, P. Thakuriah, "Incorporating weather information into real-time speed estimates: Comparison of alternative models", *J. Transp. Eng.*, vol. 139, no.4, 379–389, 2012, doi.org/10.1061/(ASCE)TE.1943-5436.0000506.

[17] L. Zhang, J. Ma, X. Liu, M. Zhang, X. Duan, Z. Wang, "A novel support vector machine model of traffic state identification of urban expressway integrating parallel genetic and C-means clustering algorithm", *Tehnički vjesnik*, vol. 29, no. 3, 731-741, 2022, doi.org/10.17559/TV-20211201014622.

[18] Y. Zhu, Y. Zheng, "Retracted article: traffic identification and traffic analysis based on support vector machine." *Neural Computing and Applications*, vol. 32, no. 7, 1903-1911, 2020, doi.org/10.1007/s00521-019-04493-2.

[19] S. Dong, "Multi class SVM algorithm with active learning for network traffic classification." *Expert Systems with Applications* vol. 176, 114885, 2021, doi.org/10.1016/j.eswa.2021.114885.

[20] D. Mladenović, S. Janković, S. Zdravković, S. Mladenović, A. Uzelac, "Night traffic flow prediction using K-nearest neighbors algorithm", *Operational research in engineering sciences: theory and applications*, vol. 5, no. 1, 152-168, 2022, doi.org/10.31181/oresta240322136m.

[21] F. Harrou, Z. Abdelhafid, S. Ying, "Traffic congestion monitoring using an improved kNN strategy." *Measurement*, vol. 156, 107534, 2020, doi.org/10.1016/j.measurement.2020.107534.

[22] T. Seo, T. Kusakabe, Y. Asakura, "Calibration of fundamental diagram using trajectories of probe vehicles: Basic formulation and heuristic algorithm", *Transportation Research Procedia*, vol. 21, 6-17, 2017, doi.org/10.1016/j.trpro.2017.03.073.

[23] C.G. Claudel, A.M. Bayen, "Guaranteed bounds for traffic flow parameters estimation using mixed Lagrangian-Eulerian sensing," *46th Annual Allerton Conference on Communication, Control, and Computing*, 636–645, 2008, doi.org/10.1109/ALLERTON.2008.4797618.

[24] J. Jiang, N. Dellaert, T. Van Woensel, L. Wu, "Modelling traffic flows and estimating road travel times in transportation network under dynamic disturbances", *Transportation*, vol. 47, 2951-2980, 2020, doi.org/10.1007/s11116-019-09997-3.

[25] G.B. Whitham, *Linear and Nonlinear Waves*, Wiley-Interscience, 1974.

[26] R. Pradhan, S. Shrestha, D. Gurung, "Mathematical modeling of mixed-traffic in urban areas." *Mathematical modeling and computing*, vol. 9, No. 2, 226-240, 2022, doi: 10.23939/mmc2022.02.226.

[27] A. Iftikhar, Z.H. Khan, T.A. Gulliver, K.S. Khattak, M.A. Khan, M. Ali, N, Minallahe, "Macroscopic traffic flow characterization at bottlenecks", *Civil Engineering Journal*, vol. 6, no. 7, 1227-1242, 2020, doi.org/10.28991/cej-2020-03091543.

[28] J. Junfeng, H. Wang, "Traffic behavior recognition from traffic videos under occlusion condition: a Kalman filter approach." *Transportation research record,* vol. 2676, no. 7, 55-65, 2022, doi.org/10.1177/03611981221076426.

[29] F. Zheng, S.E. Jabari, H.X. Liu, D. Lin, "Traffic state estimation using stochastic Lagrangian dynamics", *Transportation Research Part B: Methodological*, vol. 115, 143-165, 2018, doi.org/10.1016/j.trb.2018.07.004.

[30] Y. Liu, Z. Cai, H. Dou, "Highway traffic congestion detection and evaluation based on deep learning techniques", *Soft Comput*, vol. 27, no. 20, 12249–12265, 2023, https://doi.org/10.1007/s00500-023-08821-6.

[31] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data", *Transportation Research Part C: Emerging Technologies*, vol. 66, 99-118, 2016, doi.org/10.1016/j.trc.2015.07.005.

[32] C. Wang, W. Zhang, C. Wu, H. Hu, H. Ding, W. Zhu, "A traffic state recognition model based on feature map and deep learning" *Physica A: Statistical Mechanics and its Applications*, vol. 607, 128-198, 2022, doi.org/10.1016/j.physa.2022.128198.

[33] T.D. Toan, Y.D. Wong, "Fuzzy logic-based methodology for quantification of traffic congestion", *Physica A: Statistical Mechanics and its Applications*, vol. 570, p.125784, 2021, doi.org/10.1016/j.physa.2021.125784.

[34] A. Skabardonis, P. Varaiya, K.F. Petty, "Measuring recurrent and nonrecurrent traffic congestion", *Transportation Research Record*, vol. 1856, no. 1, 118-124, 2003, doi.org/10.3141/1856-12.

[35] Erdinç, G., Colombaroni, C. and Fusco, G., "Two-stage fuzzy traffic congestion detector", *Future transportation*, vol.3, no. 3, 840-857, 2023, doi.org/10.3390/futuretransp3030047.

[36] G. Erdinç, C. Colombaroni, G. Fusco, "Application of a Mamdani-Based Fuzzy Traffic State Identifier to a Real Case Study", In *2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* 1-6, IEEE, 2023, doi.org/10.1109/MT-ITS56129.2023.10241428.

[37] National Academies of Sciences, Engineering, and Medicine, *Highway Capacity Manual 7th Edition: A Guide for Multimodal Mobility Analysis*, Washington, DC: The National Academies Press, 2022, doi.org/10.17226/26432.

[38] G. Hewei, A.S. Sadiq, M.A. Taheir, "Adaptive fuzzy-logic traffic control approach based on volunteer IoT agent mechanism", *SN Computer Science*, vol. 3, no. 1, p.68, 2022, doi.org/10.1007/s42979-021-00956-3.

[39] B. Yulianto, "Adaptive Traffic Signal Control Using Fuzzy Logic Under Mixed Traffic Conditions", *Proceedings of the 5th International Conference on Rehabilitation and Maintenance in Civil Engineering. ICRMCE 2021. Lecture Notes in Civil Engineering*, vol 225. Springer, Singapore, doi.org/10.1007/978-981-16-9348-9_59.

[40] A. Agrahari, M.M. Dhabu, P.S. Deshpande, A. Tiwari, M.A. Baig, A.D. Sawarkar, "Artificial Intelligence-Based Adaptive Traffic Signal Control System: A Comprehensive Review", *Electronics*, vol.13, no. 19, p.3875, 2024, doi.org/10.3390/electronics13193875.

Gizem Erdinc has done her bachelor's and master's degree from Erciyes University in 2014 and 2016 respectively. She completed her first PhD at Erciyes University in 2020 and the second one at Sapienza University of Rome in 2024.

She has experience in fuzzy logic applications in civil engineering problems, transportation planning and construction management.

Chiara Colombaroni has done his bachelor's, master's and PhD degree from Sapienza University of Rome.

Her research activities are addressed to transportation planning and transportation system modelling, driver behaviour modelling, traveller information systems, road traffic control and network design.

Gaetano Fusco has done his degree from Sapienza University of Rome. He is a Full Professor of Transport at the Faculty of Engineering of the University of Roma "La Sapienza.

JENRS

Received: 17 September, 2024, Revised: 23 October, 2024, Accepted: 24 October, 2024, Online: 07 November, 2024

DOI: https://doi.org/10.55708/js03110004

Comparing the Performance of Recycled Calcined Bauxite vs. Locally Available Aggregate as Components in High Friction Surface Treatment Applications

Eslam Deef-Allah [*1], **Korrenn Broaddus** [2], **Magdy Abdelrahman** [2]

[1] Department of Construction Engineering and Utilities, Zagazig University, Egypt

[2] Department of Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, U.S.A.

*Corresponding author: Eslam Deef-Allah, Email: emddkc@gmail.com

ABSTRACT: For safe driving, the proper level of pavement friction must be maintained. The capacity of a pavement surface to offer friction to vehicles during cornering maneuvers or crucial braking is improved by high friction surface treatment. Calcined Bauxite is the most often utilized high friction aggregate. Calcined Bauxite, a byproduct, was the priciest aggregate relative to other locally accessible aggregates due to its restricted availability. Bauxite is one source of aggregate that when recycled (processed) has high friction properties. The researchers therefore assessed the frictional performance of Calcined Bauxite compared to five alternatives: Meramec River Aggregate, Rhyolite, Earthworks, and two byproducts (Flint Chat and Steel Slag). These alternatives were collected because they were locally available and less expensive when compared to Calcined Bauxite. The dynamic friction, the British pendulum, and the aggregate image measurement system were among the performance tests. The impact of aggregate sizes (#4 – #6 and #6 – #8) on British pendulum number (BPN) values was examined for Calcined Bauxite and three alternatives (Meramec River Aggregate, Rhyolite, and Steel Slag). Aggregates of larger size exhibited higher BPNs than smaller-size aggregates with differences in values ranging from 0.8 to 7. The correlation between the coefficients of friction measured by the dynamic friction tester and the number of polishing cycles was examined, and it revealed an exponential relationship. Alternatives to Calcined Bauxite can be screened using percentages of texture and angularity indices changes following Micro-Deval abrasion. Meramec River Aggregate, Flint Chat, and Steel Slag were effectively recommended alternatives for Calcined Bauxite.

KEYWORDS: Calcined Bauxite, Friction, HFST, Recycled, Byproduct, British pendulum, Micro-Deval, Dynamic friction, Abrasion, Polishing

1. Introduction

Friction, also known as skid resistance, refers to the force generated when a vehicle's tire slides down a paved surface without spinning [1]. Friction between pavement and vehicle's tires can shorten braking distance and reduce vehicle-related collisions, particularly during wet weather [2,3]. The complicated phenomena of friction on a roadway are influenced by a variety of elements, including the environment, traffic, and materials. Different macro- and micro-textures are produced by material characteristics (such as aggregate size and shape), which contribute to variations in friction. It has been demonstrated that at low speeds, good pavement micro-texture is critical for friction, whereas, at high speeds, good pavement macro-texture is more significant [2,4]. When pavement surfaces lose their friction, surface treatments are introduced as a cure to increase friction and prolong pavement life [5]. One of these surface treatments is known as high friction surface treatment (HFST), which is widely utilized in the United States [5,6].

High friction surface treatment was introduced as a remedy for locations with friction deficits, inadequacies of horizontal curves with tiny radii, and modest superelevation rates [5,6], or places where relocating immovable objects to expand visibility is difficult [5,7]. It

was found that HFST is a safety application comprised of a resin layer that bonds the pavement with a 3 to 4 mm high-angularity, high-abrasion, high-texture, and polish-resistant aggregates (such as Calcined Bauxite, Slags, Flint/Chert, or Granite) [5,6,8–12]. Calcined Bauxite is one of the high-friction aggregates that is frequently utilized in the HFST.

Calcined Bauxite is currently the predominant aggregate utilized in Missouri for HFST applications. Bauxite is mined largely for its aluminum content. The ore is refined using the Bayer process, in which it is treated with sodium hydroxide at high temperatures and pressure to produce alumina (aluminum oxide). After alumina extraction, the alumina is calcined to form aluminum metal. This is accomplished by heating alumina in rotary kilns or fluidized bed calciners to temperatures of 1000–1500 degrees Celsius. During this process, some of the alumina may be transformed into Calcined Bauxite. In this situation, whereas alumina is the primary product of Bauxite processing, Calcined Bauxite is considered a byproduct of alumina calcination. With over 3.7 billion tons of Bauxite reserves, China is a major producer [13]. The top producers of Calcined Bauxite are China, Australia, Guinea, Jamaica, Brazil, and India [6,14]. As of 2023, China is the world's third-largest producer of Bauxite, accounting for over 22% of global production. Calcined Bauxite typically has a bulk density of around 3 gm/cm3, an alkali content of less than 0.4%, and an alumina content of more than 82% [5,14,15]. This type of aggregate is stable, dense, and very pure, making it suitable for a variety of applications such as abrasives, anti-skid protections, and surface treatments [5,6].

The Missouri Department of Transportation (MoDOT) has been using HFST since 2013 to enhance wet crash locations and to increase friction on the pavement where traffic has worn down surface particles [16]. When analyzing the performance of HFST application, the main objective is frictional performance, which is determined by the aggregates' micro-textures and the surfaces' macro-textures [8,17]. Aggregate micro-texture is influenced by its shape (texture and angularity), whereas aggregate macro-texture is determined by its gradation, compaction degree, and mixture design [17–21]. To determine the aggregates' shape (angularity and texture) both after and before abrasion in the Micro-Deval device, Texas A&M University created the aggregate image measurement system (AIMS) [17,22–24]. The two-dimensional form of the aggregates is used to establish their shape, the wavelet analysis method is used to analyze grayscale images to determine the surface texture, and the irregularity of a particle's surface is used to estimate the angularity of the aggregate [17]. The British pendulum assesses the aggregates' resistance to polishing after and before a 10-hour polishing session using the British accelerated polishing equipment named the British wheel. The British pendulum F-scale is used for the 1.25-inch-wide slider, whereas the main (primary) scale is utilized for the 3-inch-wide slider, which is utilized for flat surfaces [25]. Furthermore, a dynamic friction tester (DFT) is used to investigate the friction of aggregates after and before polishing with a three-wheel polishing device (TWPD) [26].

Using the Micro-Deval test, the abrasion resistances of Steel Slag and Calcined Bauxite with a 9.5-mm nominal maximum aggregate size were assessed [25]. Compared to the percentage of mass losses for Calcined Bauxite, the mass loss percentage for Steel Slag was 17% greater. Additionally, they investigated the polish value of Steel Slag and Calcined Bauxite, with aggregate sizes of 1–3 mm and 6.3–9.5 mm, both after and before polishing. The initial polish value and 10-hour polish value values rose with decreasing aggregate size. The finding was explained by the researchers due to the preparation process: As opposed to the 6.3–9.5 mm aggregate size, the smaller aggregate size (1–3 mm) was too small to be put individually in a single layer. As a result, there was more variation between the 1–3 mm samples. Additionally, the polish values for Steel Slag were lower than those for Calcined Bauxite. The researchers [25] did advise using Steel Slag as an aggregate substitute for Calcined Bauxite, though. Other researchers [26] compared seven friction aggregates with Calcined Bauxite and examined how well they performed in terms of friction. Based on the DFT results, none of the seven frictional aggregates had friction similar to that of Calcined Bauxite. The researchers determined that friction aggregates exhibited comparable friction losses in cycles ranging from 70k to 140k. Furthermore, a circular texture meter (CTM) was used to assess the impact of aggregate size—Calcined Bauxite, Slag, Flint, and Taconite—on the frictional performance. According to this study [26], the surface texture was reduced when the aggregate's particle size was reduced. For example, when the particle size was changed from #6 to #16, the CTM macro-texture diminished from 2.3 mm to 1 mm.

In the U.S.A., the principal aggregate for HFST is Calcined Bauxite. While Calcined Bauxite has significance in many applications due to its unique qualities, it confronts several concerns that may jeopardize its market viability and development prospects in the future. Concerns about the continuous use of Calcined Bauxite include availability and cost, which are greatly affected by the worldwide distribution of raw Bauxite supplies, as well as environmental regulations. Calcined Bauxite's cost issues originate from its energy-intensive production process, expensive capital expenditures for production facilities, supply chain vulnerabilities due to raw material sourcing, importing expenses, and rising environmental compliance costs. Stringent environmental rules in China, a major producer of Bauxite, have forced manufacturing

facilities and mines to close temporarily. The shutdown of facilities owing to environmental compliance diminishes current supply while also creating uncertainty about future availability. These regulatory acts can cause major variations in market supply and costs of Calcined Bauxite.

Calcine Bauxite has been the sole source of aggregate used in HFST, but it is expensive and is imported. Thus, assessing the frictional performance of alternatives to Calcined Bauxite was one of the primary goals of this study. Extensive comparative frictional investigations between alternatives with different sizes and Calcined Bauxite were lacking. Consequently, this study investigated how aggregate size affects the frictional performance of high-friction aggregates. Another major objective was to strengthen the principle of sustainability by including recycled aggregates in the selected alternatives.

2. Materials and Methods

2.1. Materials

Five alternatives and Calcined Bauxite were chosen for evaluation. These aggregates were chosen as potential replacements for Calcined Bauxite, and they were available locally. Meramec River Aggregate, Rhyolite, Steel Slag, Earthworks, and Flint Chat were the alternatives (note Figure 1). The NJSP-15-13B HFST requirements set out by the MoDOT and anticipated modifications to the HFST standard were assessed for each of these aggregates through testing.

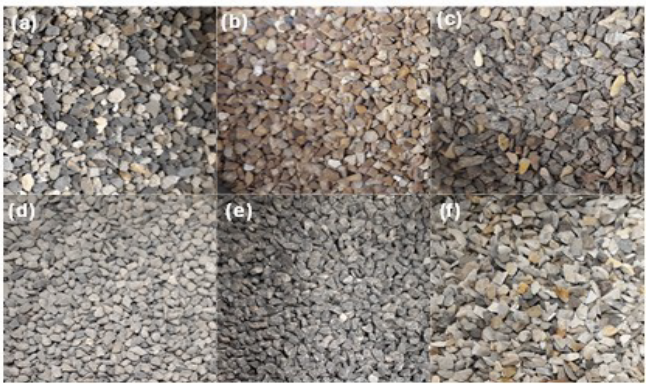


Figure 1: Calcined Bauxite and Alternatives: (a) Calcined Bauxite; (b) Meramec River Aggregate; (c) Rhyolite; (d) Steel Slag; (e) Earthworks; (f) Flint Chat

Bauxite is mined because it is usually always found on the surface of the ground. Bauxite is one type of aggregate that, when recycled or processed, has high friction properties. Calcined Bauxite is created by sintering high-alumina Bauxite at high temperatures (Great Lakes Minerals, LLC, Wurtland, Kentucky, U.S.A.). Calcined Bauxite is a recycled material derived from naturally occurring Bauxite ore, and it makes use of previously mined materials that were overlooked during earlier processing stages. The average aluminum oxide percentage for Calcined Bauxite was 87.5%, somewhat higher than the 87% minimum specified in most HFST requirements. It included 88.65% aluminum oxide. China is the world's top producer of Calcined Bauxite, leveraging its vast reserves of Bauxite ore, notably those containing the Diaspore mineral.

Natural Calcined Bauxite, known as Earthworks, was blasted, mined, and crushed at Earth Work Solutions Quarry (Gillette, Wyoming, U.S.A.). The rock formation is layered. Meramec River Aggregate is coarse manufactured sand obtained from Winter Brothers Material Company (Saint Louis, Missouri, U.S.A.). The material is a thousand-years-old sand and gravel deposit in the Meramec basin. The layers of material include overburden (topsoil), gravel, often a thin coating of sticky mud, and sand at the bottom. Fred Weber Inc. acquired Iron Mountain Trap Rock, also known as Rhyolite, an igneous, volcanic rock (Lead Belt region of Missouri, Ironton, U.S.A.). It was produced during significant volcanic activity in Missouri. The Lead Belt region is noted for iron ore, lead smelting and manufacturing, granite, and trap rock.

Flint Chat and Steel Slag were chosen as byproducts. Flint is a sedimentary rock made up of microscopic quartz crystals (silica or SiO_2). It is frequently discovered as a byproduct in the dumps of abandoned lead and zinc mines (Williams Diversified Materials, Baxter Springs, Kansas, U.S.A.). It can be considered recycled material. As mining develops, undesired byproducts and remaining minerals are usually placed in dumps or tailings. Tailings are waste materials that remain after valuable minerals have been extracted from ore. In many cases, these tailings contain considerable amounts of Flint that were not originally intended for extraction but are still there owing to a geological relationship with other mined minerals. These Flint accumulations in tailings or landfills have the potential to be recycled material over time. Furthermore, there is now more interest in recycling minerals from mining activities due to environmental concerns. Reusing accumulated Flint and repurposing tailings are two ways that businesses may reduce its environmental impact and support sustainable mining methods. Steel Slag is a byproduct of the steelmaking process that occurs when impurities in molten iron are removed during the manufacturing of steel. The business operations of Harsco Inc. (Muscatine, Iowa, U.S.A.) provide an example of how businesses may use byproducts like Steel Slag to satisfy market demands while reducing waste and making a positive impact on environmental sustainability.

FasTrac CE330 epoxy binder, an A and B two-component epoxy polymer binder with low modulus, was investigated to be used in this study. The two components are an epoxy (hardener) and a resin (extender). The epoxy binder exhibited a 40% tensile elongation and a bond strength of 2 ksi after two days and 2.8 ksi after fourteen

days. Loose dense-graded asphalt mixtures, with a 12.5-mm nominal maximum aggregate size, were supplied by an asphalt plant (Pullman, WA, U.S.A.) to create asphalt slabs (20 × 20 × 2 inches). The epoxy binder was utilized for HFSTs on hot mix asphalt (HMA) slabs, accelerated friction testing, and coupon preparation for the British pendulum test.

2.2. Methods

2.2.1. AIMS

Following the Micro-Deval test (105-minute and 180-minute abrasion periods) on aggregate size 3/8" – #4, the aggregate samples were examined in the AIMS alongside samples before abrasion (ASTM D6928 – 17 [27]). AIMS was used to investigate two sizes, following the HFST specifications, of aggregate: 3/8" – 1/4" and 1/4" – #4. The AIMS analysis was performed to investigate the changes in aggregate shape (angularity and texture indices) following Micro-Deval abrasion. Black and white images were used to identify the angularity indices based on particle surface irregularities, while grayscale images were analyzed using the wavelet analysis approach to find the surface texture indices [17].

2.2.2. Accelerated Friction Testing

To compare the surface frictional properties of alternate aggregates to Calcined Bauxite, accelerated friction testing was carried out. The sand patch test for mean texture depth (MTD) measurement and a DFT for coefficient of friction (COF) measurement were accelerated friction testing techniques.

I. Applying HFST on Hot Mix Asphalt Slabs

The plant mixes were warmed, and a small plate compactor was utilized in the laboratory to create and compact the HMA slabs (20 × 20 × 2 inches). The HMA slabs were coated with an A and B two-part epoxy binder before the application of size #6 – #8 aggregates. Selected aggregate size is following the HFST specifications. According to the supplier's specifications, the epoxy's part A to part B weight ratio was 1:18 to 1.00.

II. Sand Patch Test: A Volumetric Approach to Assessing Pavement Macro-texture

The sand patch test was used to determine the MTD values of the manufactured test slabs following ASTM E965 – 15(2019) [28]. The average MTD for each surface was computed using Equation 1. The average of two replicates—two test slabs—was the basis for the results.

$$MTD = \frac{4V}{\pi D^2} \quad (1)$$

where,
D is the sand patch circle's average diameter (mm), V is the volume of sand (mm^3), and MTD is the mean texture depth (mm).

III. Dynamic Friction Test

The test slabs were polished with a TWPD to resemble the polishing of aggregates used in HFST under field traffic conditions. The TWPD included three turntable-mounted pneumatic rubber wheels as well as a water spray system that mimicked wet conditions to lessen rubber wheel wear and remove surface fines to enable extra polishing. The overall weight of the wheel, involving the metal plates (six in total) and the wheel cluster, was 149 lbs. The COF was tested by the researchers at three distinct polishing cycle numbers: 0 cycles (initial), 70k cycles, and 140k cycles (terminal). Following ASTM E1911-19, the COF was evaluated by a DFT at various speeds (20, 40, and 60 km/hr) [29]. Three rubber pads were fastened to a circular disk that made up the DFT. The circular disk revolved at speeds of up to 100 km/hr. Once the disk had attained the desired speed, it was dropped to the paved surface, and the COF was determined as the speed of the circular disk as it steadily reduced. The friction was tested under wet circumstances. The average of two duplicates—two test slabs—was the basis for the results.

2.2.3. Assessing the Surface Friction of Aggregate Coupons with the British Pendulum

I. Creating Aggregate Coupons

The metal mold's bottom (Figures 2-a and 2-b) was covered with a 12g ready-mix plaster, on which the aggregates were inserted. After experimenting with various ready-mix plaster weights, it was discovered that a weight of 12g prevented the epoxy binder from leaking into the gaps between the aggregate particles for both the #4 – #6 and #6 – #8 sizes. Selected aggregate sizes are following the HFST specifications. To prevent the epoxy from sticking to the metal molds, more plaster was applied entirely to the sides of the molds using a tiny brush (see Figure 2-c). A single layer of aggregates was carefully applied to the plaster.

An epoxy binder consisting of two parts, A and B, was produced. The weight ratio of epoxy part A to part B was 1:18 to 1.00. To fill the remaining space in the metal mold, the aggregates were covered with the prepared epoxy binder. For four to six hours, the aggregate coupons were kept in the metal molds at room temperature. Ultimately, the plaster coating was removed from the aggregate coupons by taking them out of the metal molds and giving them a water wash. Samples of prepared aggregate coupons are shown in Figure 3. The created coupons were first examined for their British pendulum number (BPN), and then after ten hours of polishing in the British wheel, they were again tested for their BPN.

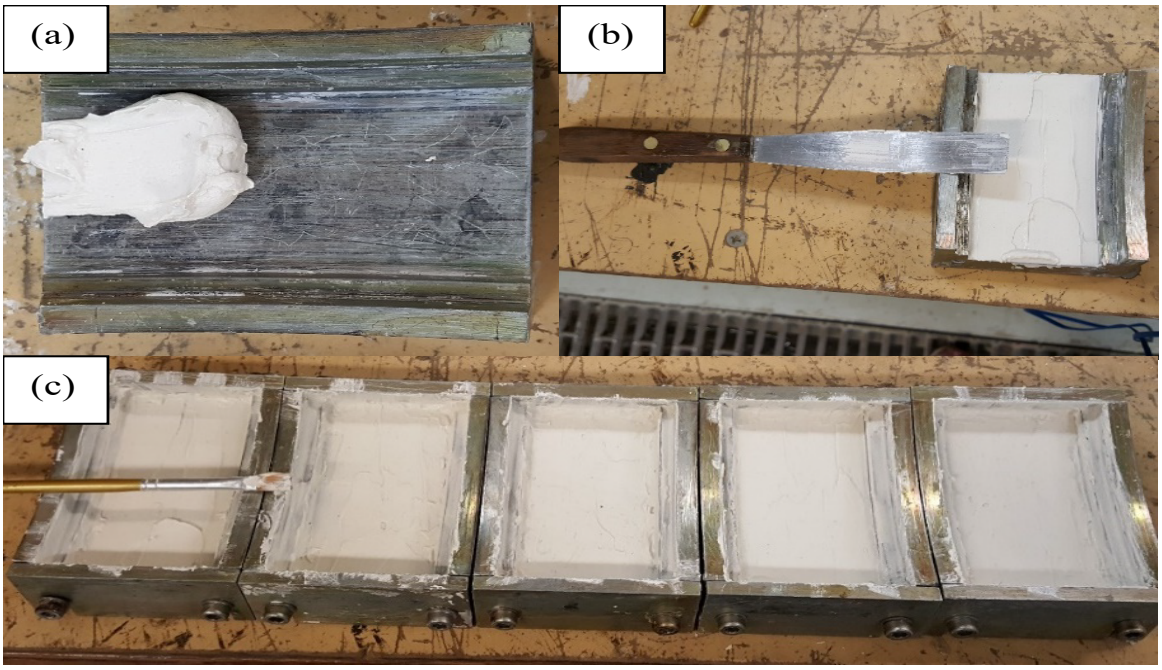


Figure 2: Plaster Used in Metal Molds: (a) Plaster Before Spreading; (b) Plaster After Spreading; (c) Additional Plaster After Painting onto the Sides of the Metal Molds

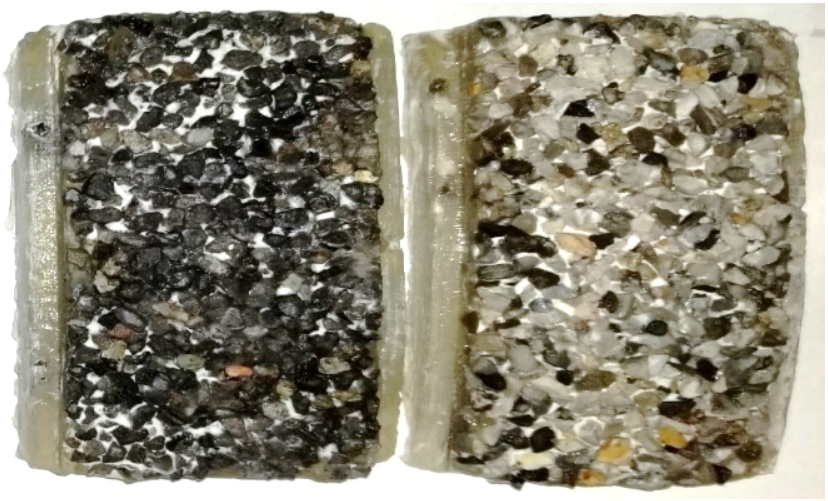


Figure 3: Samples of the Aggregate Coupons

II. British Pendulum Test

The test's objective was to use the British pendulum to quantify the surfaces' frictional characteristics. The AASHTO T 278-90 (2017) [30] was followed in the preparation of the tester shown in Figure 4 with slide length adjustments (Section 7.3) and zero adjustments (Section 7.2). A slider measuring 1/4 inch by 1 inch by 1 1/4 inches was employed. The test employed two aggregate sizes. The first size was #6 – #8, while the second was #4 – #6. Two aggregate coupons were created for each aggregate size. Each aggregate coupon's BPN was measured five times on the F-scale (note Figure 4), and the average BPN before polishing was computed. Following that, the British wheel (Figure 5-a) was used to polish the aggregates on the coupons for ten hours, and the BPN values were recorded five times before being averaged to determine the BPN after polishing.

III. Polishing Aggregates Using the British Wheel

The coupons' aggregates were polished after being tested in the British pendulum per AASHTO T 279-18 [31] using the British wheel, note Figure 5-a. The test imitated the polishing operation that occurs on aggregates in the field. For every run, 14 aggregate coupons were secured around the road wheel's periphery (Figures 5-b and 5-c). The road wheel speed was adjusted to 320 ± 5 rpm. With a total load of 391.44 ± 4.45 N, the pneumatic-tired wheel was lowered to bear on the aggregate coupons' surface. For ten hours, the aggregates were polished in the presence of water and #150 silicon carbide grit (as a polishing agent).

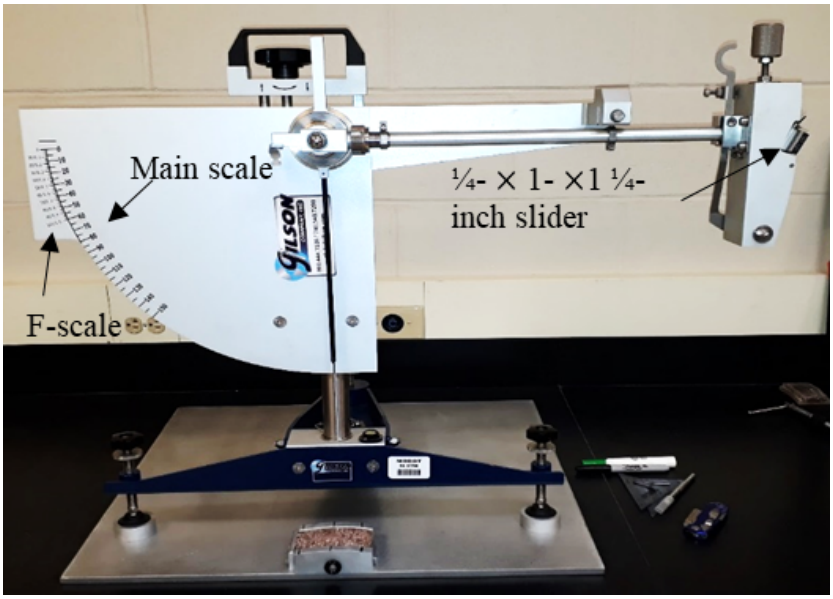


Figure 4: British Pendulum Tester

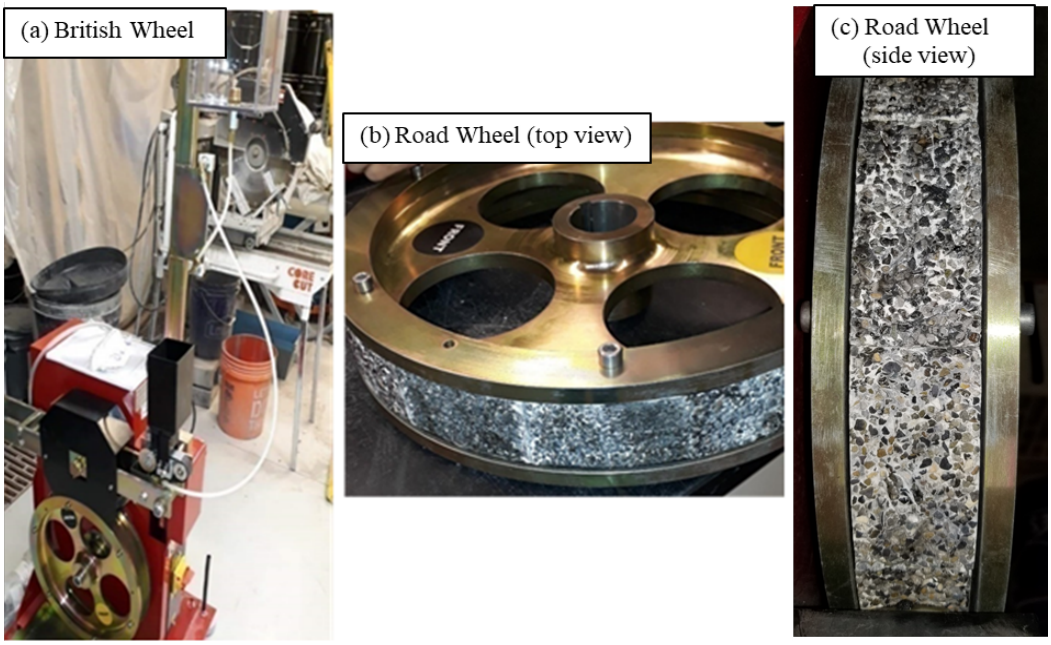


Figure 5: Polishing HFST Coupons: (a) The British Wheel; (b and c) Road Wheel

3. Results and Discussion

3.1. AIMS

This section compares the AIMS angularity and texture indices for aggregates before Micro-Deval abrasion (BMD), after Micro-Deval abrasion time of 105 minutes (AMD 105), and after Micro-Deval abrasion time of 180 minutes (AMD 180). There were two aggregate sizes examined: 3/8" – 1/4" and 1/4" – #4.

3.1.1. Aggregate Size's Impact on Angularity and Texture Indices

Figure 6 displays the AIMS angularity and texture indices for two sizes of aggregates: 3/8" – 1/4" and 1/4" – #4—for BMD and AMD (105 or 180). The texture and angularity indices for BMD and AMD (105 or 180) were lowered by reducing the aggregates' sizes from 3/8" – 1/4" to 1/4" – #4. Nonetheless, when the aggregate size was reduced from 3/8" – 1/4" to 1/4" – #4, Steel Slag provided greater texture indices for BMD. This was attributed to the greater surface area and irregularities associated with smaller-sized aggregates. Furthermore, when aggregate sizes decreased from 3/8" – 1/4" to 1/4" – #4, Earthworks and Flint Chat demonstrated increases in angularity indices concerning BMD. When the size of the aggregate was reduced, Calcined Bauxite showed an increase in the angularity indices for AMD (105 or 180). To clarify the influence of aggregate size on texture indices, the indices were averaged for each size using the results of BMD, AMD 105, and AMD 180. A similar procedure was followed for the angularity indices. The percentages of decrease in these indices were computed when the

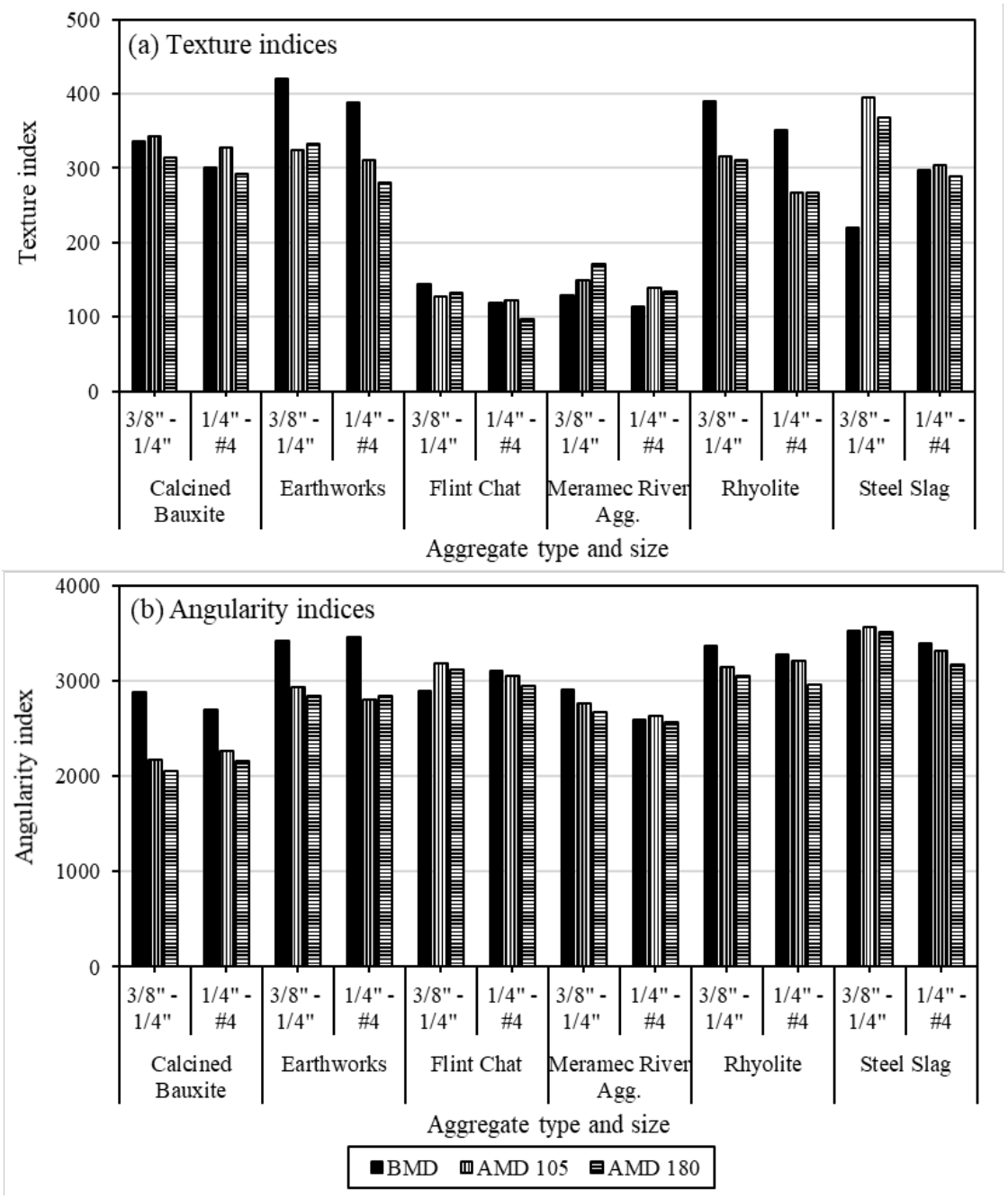


Figure 6: AIMS Analysis: (a) Texture Indices; (b) Angularity Indices

aggregate size decreased from 3/8" – 1/4" to 1/4" – #4. Reducing aggregate size lowered texture indices by −7.3% to −16.5% and angularity indices by −1% to −6.9%. Increasing the Micro-Deval abrasion time from 105 to 180 minutes reduced the angularity indices in 100% of the samples and decreased the texture indices in 75%. The polishing mechanism caused by Micro-Deval decreased sharp edges and smoothed surfaces, lowering the texture and angularity indices [23,24,32].

The percentages of change in the angularity and texture indices of AIMS for aggregates for AMD (105 or 180) are depicted in Figure 7. Figure 7-a demonstrates that for three different types of aggregates—Earthworks with two sizes, Flint Chat with 3/8" – 1/4" size, and Rhyolite with two sizes—the texture indices dropped while employing AMD. The percentages of decrease in texture indices AMD for these aggregates reached values lower than −18%. This was ascribed to mechanical abrasion leading to smoother surfaces and particle breaking down resulting in less textured surfaces [12]. Using AMD 105 for Calcined Bauxite, 1/4" – #4 size Flint Chat, and Steel Slag led to higher texture indices. The percentage of increase in texture indices AMD 105 for these aggregates attained values ranging from 2% to 79%. AMD experienced higher texture indices for Meramec River Aggregate and 3/8" – 1/4" Steel Slag. The percentage of increase in texture indices AMD for these aggregates reached levels ranging from 15% to 79%. The angularity indices decreased with AMD for six different types of aggregates, as shown in Figure 7-b: Two sizes of (Calcined Bauxite, Earthworks, and Rhyolite),1/4" – #4 size (Flint Chat and Steel Slag), and 3/8" – 1/4" size of Meramec River Aggregate. The percentages of decreases in angularity

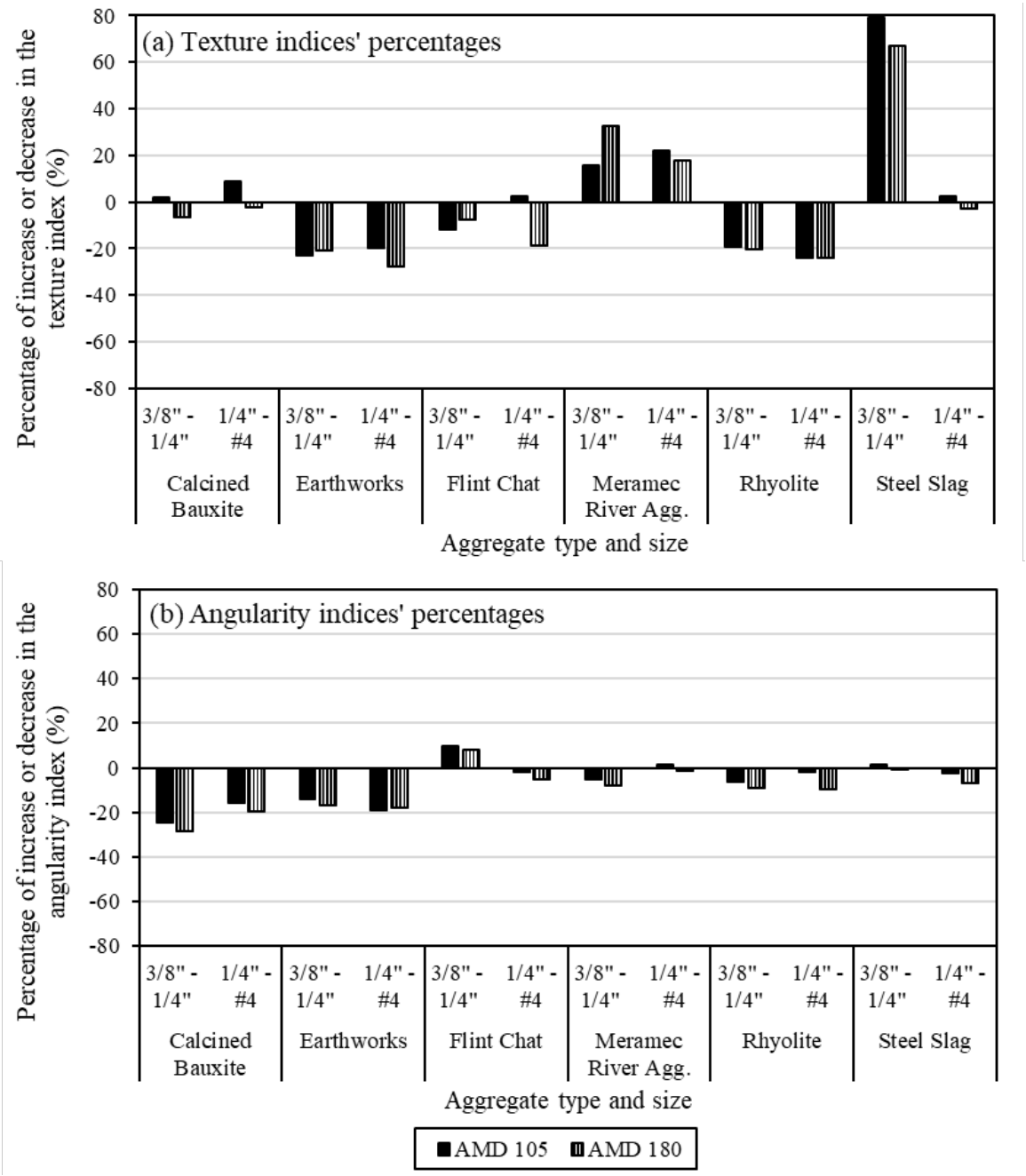


Figure 7: Percentages of Increase/Decrease in the (a) Texture Indices; (b) Angularity Indices

indices AMD for these aggregates ranged from −1.8% to −28.5%. On the other hand, the angularity indices rose with AMD 105 for the 3/8" – 1/4" Steel Slag and the 1/4" – #4 Meramec River Aggregate. The percentage of increase in angularity indices AMD 105 for these aggregates exceeded 1.2%. Furthermore, the angularity indices for the 3/8" – 1/4" Flint Chat increased for AMD (105 or 180). T he percentage of increase in angularity indices AMD for this size of Flint Chat surpassed 8%. Following Micro-Deval abrasion time, the aggregates' texture and angularity indices rose, most likely for one or more of the reasons listed below [12, 22, 24, 32]:

1. Particles were broken rather than abraded (e.g., Steel Slag and Calcined Bauxite), revealing interior surface textures;
2. Micro-Deval abrasion unveiled additional rough areas that were previously hidden by smoother surfaces (e.g., Steel Slag and Meramec River Aggregate);
3. Mineralogies in particular aggregates (e.g., Meramec River Aggregate) revealed new textured surfaces during Micro-Deval abrasion.

3.1.2. *Analyzing Average Angularity and Average Texture Indices*

AIMS indices were calculated as an average for two aggregate sizes (3/8" – 1/4" and 1/4" – #4). Figure 8 illustrates the average angularity and average texture indices for BMD and AMD (105 or 180). During BMD abrasion, Earthworks had the greatest average angularity index (3437.3) and average texture index (404.6), whereas Meramec River Aggregate had the lowest average indices (2749.6 for angularity and 121.4 for texture). The greatest average angularity and average texture indices among

AMD (105 or 180) were found in Steel Slag (3440 for angularity AMD 105, 3344.3 for angularity AMD 180, 349.4 for texture AMD 105, and 327.9 for texture AMD 180). For AMD (105 or 180), Calcined Bauxite displayed the lowest average angularity indices (2221.4 for AMD 105 and 2109.6 for AMD 180), whereas Flint Chat exhibited the lowest average texture indices (124.3 for AMD 105 and 114.7 for AMD 180). When AMD 105 was utilized for Meramec River Aggregate, average texture indices rose (from 121.4 for BMD to 144.1 for AMD 105); when AMD 180 was employed, this rise persisted reaching a value of 152.5. This occurred because the aggregates possessed mineralogies using Micro-Deval abrasion that revealed new textured surfaces and/or the Micro-Deval abrasion revealed a more textured area that had been concealed by a smoother surface. The average texture indices of both Steel Slag and Calcined Bauxite rose with AMD 105 and dropped with AMD 180. During AMD 105, particles were broken rather than abraded, exposing their interior surface textures. On the other hand, AMD 180 underwent abrasion on both the old and new exposed internal surface textures. The average texture indices for Rhyolite, Earthworks, and Flint Chat declined with AMD (105 and 180). Aggregates' average angularity indices declined and reached their lowest value with AMD 180.

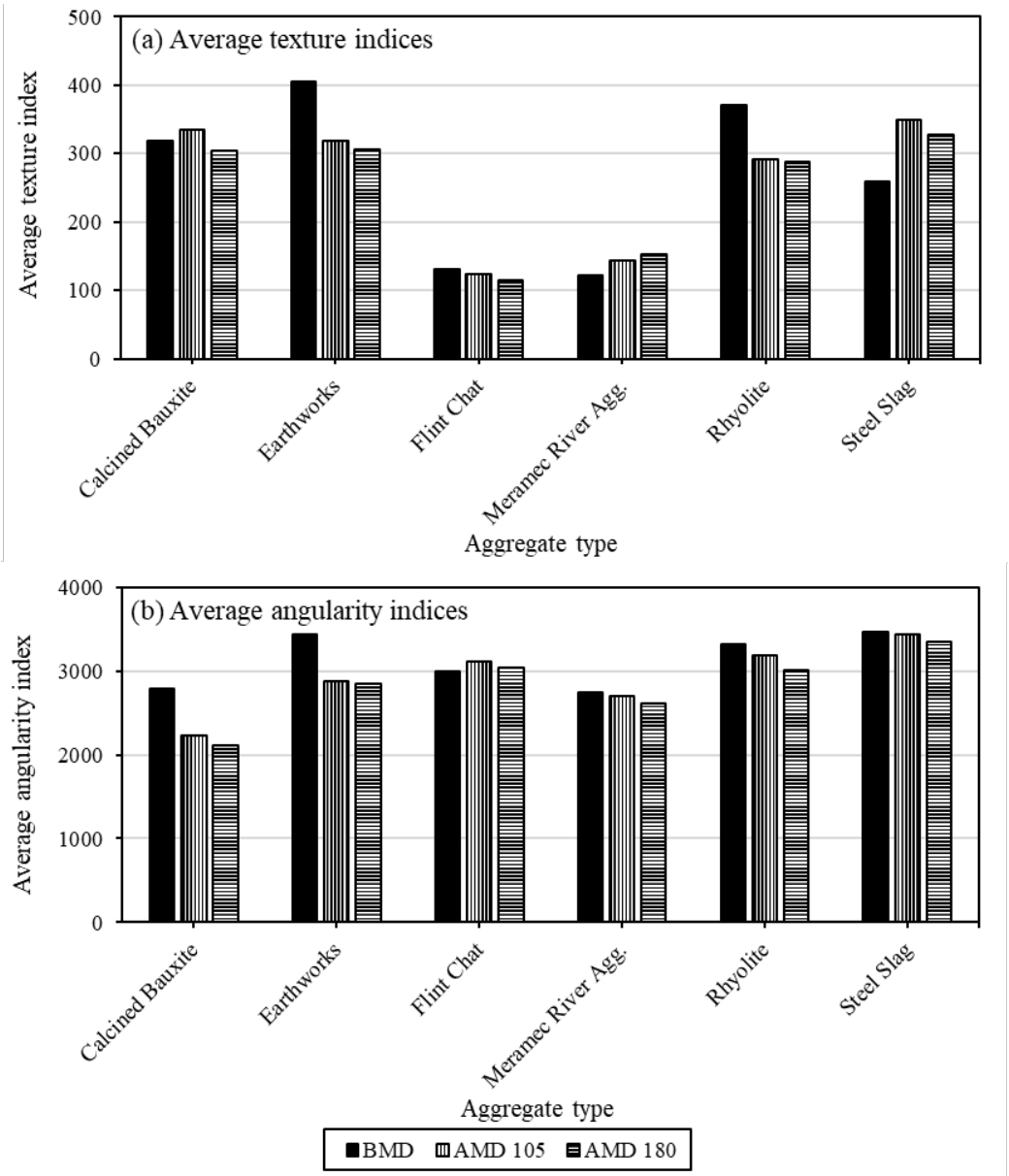


Figure 8: AIMS Analysis: (a) Average Texture Indices; (b) Average Angularity Indices

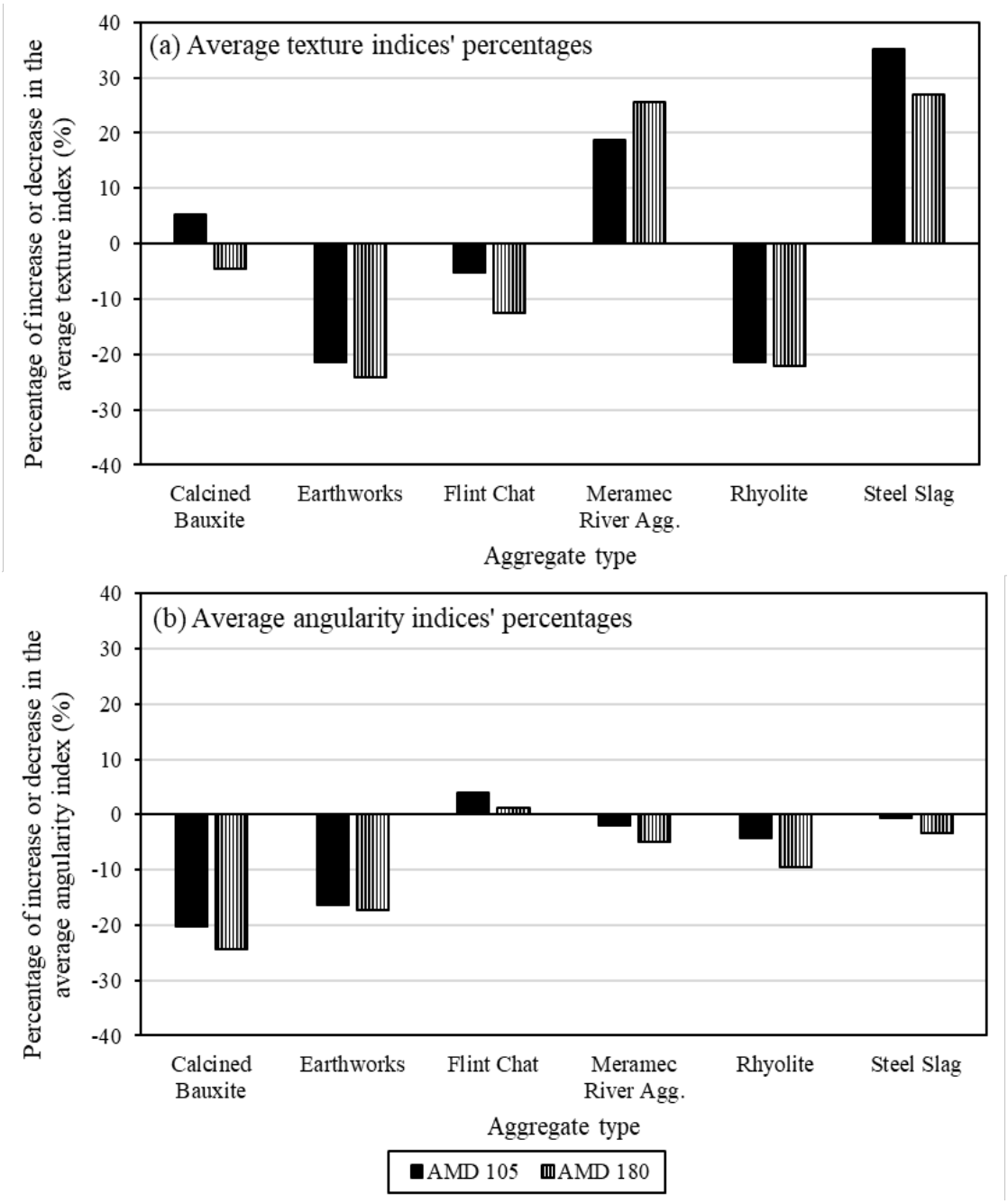


Figure 9: Percentages of Decrease/Increase in (a) Average Texture Indices; (b) Average Angularity Indices

Figure 9 shows the percentages of decrease or increase in the average texture and average angularity indices. Figure 9-a indicates that Calcined Bauxite's average texture indices percentage for AMD 180 dropped the least (−4.7%). The highest percentages of average texture indices increase for AMD (105 and 180) were seen in Steel Slag reaching values of 35.1% and 26.8%, respectively. In contrast, the average texture indices for AMD (105 or 180) decreased the most for Earthworks aggregate (−21.4% for AMD 105 and −24.2% for AMD 180). Except for Flint Chat, all aggregates showed a reduction in the average angularity indexes among AMD (105 or 180) (see Figure 9-b). Calcined Bauxite had the highest decrease percentages in average angularity indices (−20.4% for AMD 105 and −24.4% for AMD 180), as shown in Figure 9-b. Steel Slag, however, showed the lowest decrease in average angularity indices (−0.7% for AMD 105 and −3.4% for AMD 180).

3.2. *Accelerated Friction Testing*

3.2.1. *Sand Patch Test*

The MTD results for the HFSTs employing aggregates before polishing are shown in Figure 10. The MTD was the two replicates' average, with four measurements collected on each test surface. All test surfaces had measured MTDs between 2.19 and 2.69 mm. The results revealed that the MTDs of Calcined Bauxite and Rhyolite surfaces were 2.19 mm and 2.33 mm, respectively, lower than those of Flint Chat (2.69 mm) and Steel Slag (2.64 mm). MTD values were higher for the other surfaces (e.g., Earthworks and Meramec River Aggregate) than for Calcined Bauxite and Rhyolite. The MTD values of Steel Slag and Flint Chat

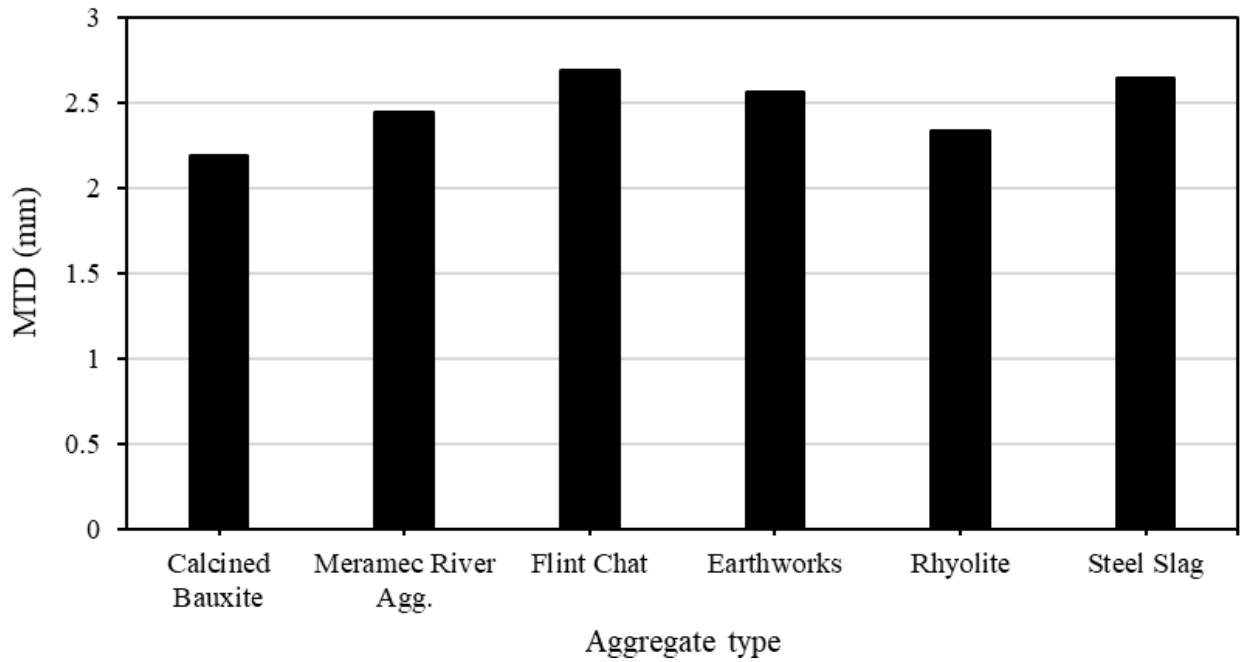


Figure 10: The Values of MTD

were greater than those of Earthworks and Meramec River Aggregate. Aggregate surface characteristics, such as aggregate angularity, led to different MTD values being noted.

3.2.2. Dynamic Friction Test

The average of two replicates was used to represent the COF values in this section. For each condition, polishing speed and cycle, a single friction measurement obtained using DFT was recorded. The average COF values for the two repetitions at 20, 40, and 60 km/hr, as determined by the DFT, are displayed in Figure 11. As predicted, the data revealed that the COF dropped with polishing. At corresponding DFT speeds and polishing cycles, Calcined Bauxite exhibited greater COF values (initial and terminal) in comparison to the remaining alternative aggregates. The Meramec River Aggregate exhibited the lowest initial friction among all aggregates, and its terminal friction was equivalent to that of Earthworks.

Figure 12 shows the initial and terminal COF values determined by DFT at various speeds for Calcined Bauxite and alternatives. According to the initial COF values, the greatest value was found in Calcined Bauxite, which was followed by Flint Chat, Earthworks, Rhyolite, and Steel Slag. The aggregate from Meramec River had the lowest initial COF. Calcined Bauxite, Flint Chat, Steel Slag, Rhyolite, and Meramec River Aggregate were deemed the best options based on the terminal COF values. Earthworks had the lowest terminal COF value recorded. This was ascribed to the action of polishing and abrasion, which exposed new smoother surfaces underneath the abraded ones. This was agreed with the AIMS results: Earthworks had the highest percentage decrease in the texture indices AMD; however, it showed the highest texture and angularity indices BMD. For the Meramec River Aggregate, the initial COF values were not significantly affected by the DFT speed. As DFT speeds increased, the COFs for the Earthworks (before polishing) and Calcined Bauxite (both after and before polishing) declined. Nonetheless, for the other feasible aggregates, increasing the DFT speed raised the initial and terminal COF values. Rhyolite (after polishing) showed the greatest rise in COF values, with a 12.5% increase when the DFT speed was increased from 20 km/hr to 60 km/hr.

In comparison to the values of initial friction at 0 cycles, Figure 13 displays the percentage of COF value losses calculated using DFT at 20 km/hr after 70k and 140k polishing cycles. After 70k and 140k cycles, Earthworks had the highest percentage of COF loss. After 70k and 140k cycles, Steel Slag showed the lowest percentage of COF loss, followed by Meramec River Aggregate, Calcined Bauxite, Flint Chat, and Rhyolite.

There was a noticeable correlation between the polishing cycles' number and the COF values determined by DFT. Figure 14 shows the correlation between the polishing cycles' number and COFs recorded by the DFT at 20 km/hr (DFT_{20}). COF readings declined exponentially as the polishing cycles' numbers increased. Therefore, an exponential regression was developed for predicting DFT_{20} based on the polishing cycles' number, as illustrated in Equation 2. Fitting parameters, for the (DFT_{20}-N) model, were determined using Excel by minimizing the sum of squared error (SSE), as shown in Table 1. Figure 15 depicts the predicted and measured values of DFT_{20} for the (DFT_{20}-N) model.

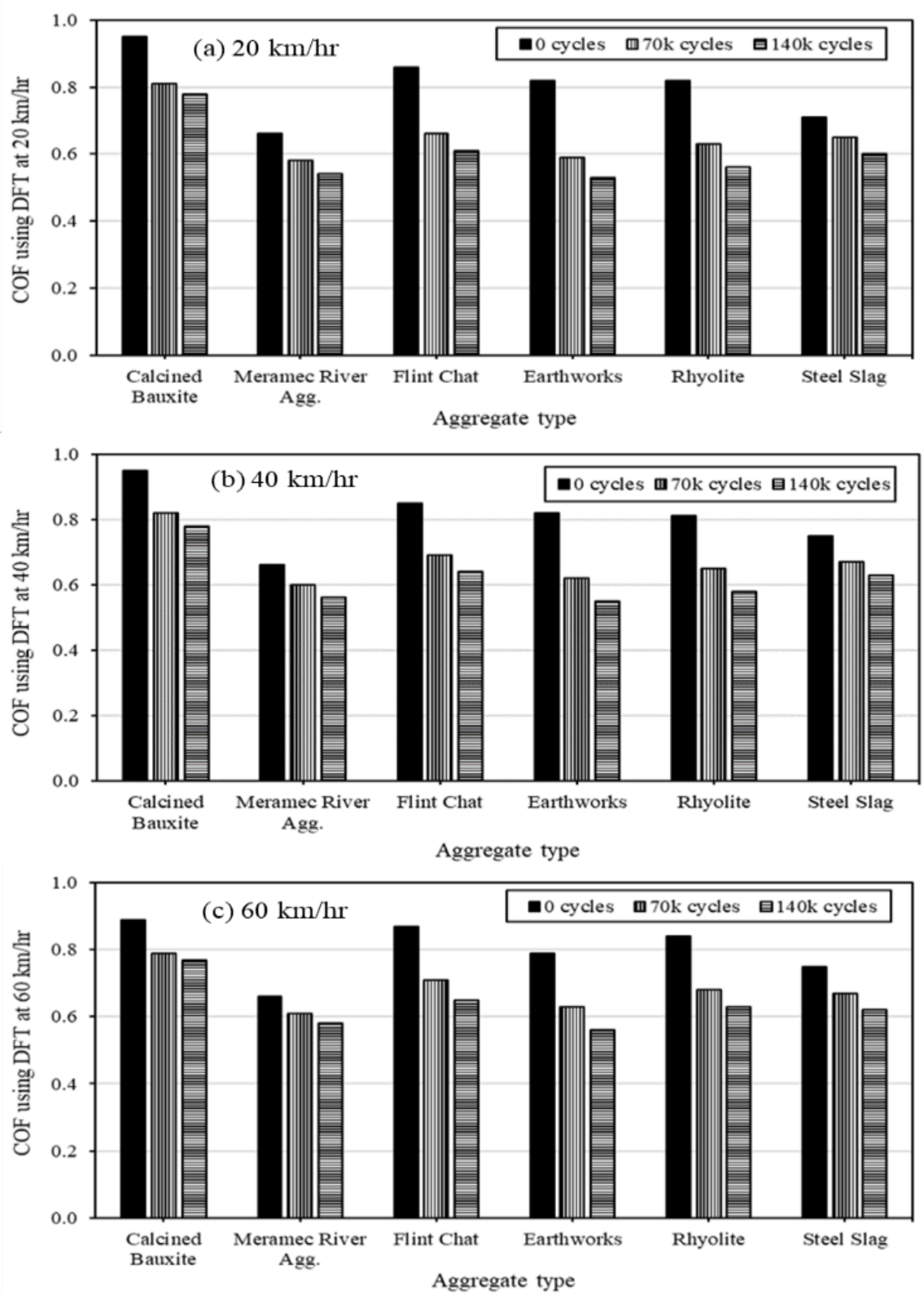


Figure 11: COF Values at (a) 20 km/hr; (b) 40 km/hr; (c) 60 km/hr

$$DFT_{20} = a + b \times \exp^{(-c \times N)} \quad (2)$$

where,

(a, b, and c) are the fitting parameters, N is the polishing cycles' number, and DFT_{20} is the friction coefficient determined by the DFT at 20 km/hr.

3.3. *British Pendulum Test*

3.3.1. *Analyzing the BPN Values*

Figures 16 and 18 exhibit aggregate BPN values recorded after and before 10-hour polishing cycles in the British wheel. Figure 16 depicts the average BPNs for aggregate sizes #6 – #8 and #4 – #6. To determine whether there was a significant difference between the means BPN

before polishing (first group) and after polishing (second group) at a significance level (α) of 0.05, a paired t-test was used. The p-value was 0.033, which is lower than the α value. Consequently, the two groups' means differed significantly from one another. Calcined Bauxite had the greatest average BPN value before polishing, followed by Meramec River Aggregate. Following polishing, Meramec River Aggregate had the greatest average BPN, followed by Calcined Bauxite. Before polishing, the average BPN values for Earthworks and Steel Slag were identical (77.5). After polishing, Earthworks had a BPN that was 0.2 greater than Steel Slag's. Rhyolite and Flint Chat exhibited similar average BPNs before polishing, while Flint Chat had lower average BPN values than Rhyolite after polishing. Figure 17 shows the percentage decrease in average BPN for aggregates after polishing. Meramec River Aggregate, Steel Slag, and Earthworks percentages decreased the least with less than −2% in the average BPN values, followed by Calcined Bauxite (−4.8%), Rhyolite (−6.7%), and then Flint Chat (−8.9%).

3.3.2. *Impact of Aggregate Size on BPN Values*

Figure 18 shows the influence of aggregate size (#6 – #8 and #4 – #6) on BPN values. To determine if there was a significant difference between the means BPN of the four groups—before polishing (#4 – #6), after polishing (#4 – #6), before polishing (#6 – #8), and after polishing (#6 – #8)—a two-way ANOVA test was employed at α = 0.05. Additionally, it was determined whether there was a significant difference between the means BPN of the four aggregate types shown in Figure 18. The four groups had a p-value of 0.031 based on the results of the ANOVA test,

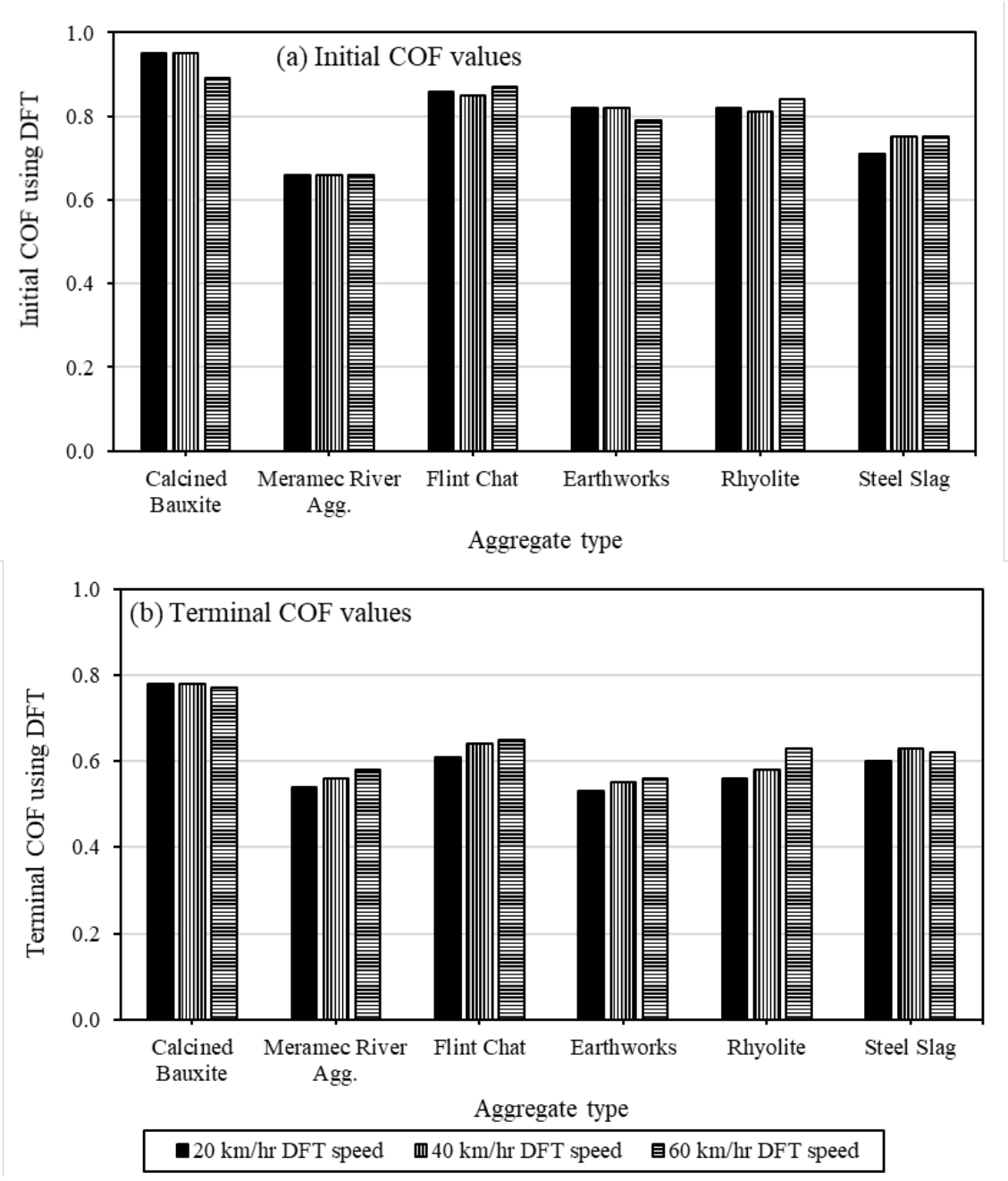


Figure 12: COF Values: (a) Initial; (b) Terminal

and the aggregate types also showed a p-value of 0.031. The significant difference between the means was indicated by both p-values being less than the α value. The results of Calcined Bauxite were compared to three alternatives: Meramec River Aggregate, Rhyolite, and Steel Slag. The aggregates sized #4 – #6 had greater BPN values after and before polishing than the aggregates size #6 – #8. Nevertheless, Meramec River Aggregate with #4 – #6 size had a lower BPN value after polishing than #6 – #8 size. The four aggregates showed lower BPN values after polishing than BPN values before polishing except for the Meramec River Aggregate with #6 – #8 size. This was related to the increase in the texture indices for AMD (105 or 180) when compared to the texture index for BMD. Figure 19 depicts the percentage increase or decrease in BPNs following the polishing procedure for aggregates of two sizes: #6 – #8 and #4 – #6. All aggregates' BPNs dropped after polishing, except for the Meramec River Aggregate of #6 – #8 size showed a 3.2% increase. Moreover, after polishing, the average BPN decreased at the lowest percentage in Steel Slag (−1.3% for the #6 – #8 size and −0.6% for the #4 – #6 size). This transpired because the texture indices for AMD (105 or 180) for both Meramec River Aggregate and Steel Slag increased when compared to the texture index for BMD. Micro-Deval abrasion of Steel Slag resulted in uneven and rougher surfaces as smoother ones were removed. This enhanced each particle's surface area, resulting in an increased texture index. Thus, polishing the Steel Slag particles with the TWPD and British wheel raised the texture index, resulting in the lowest percentage of losses in COF and BPN values.

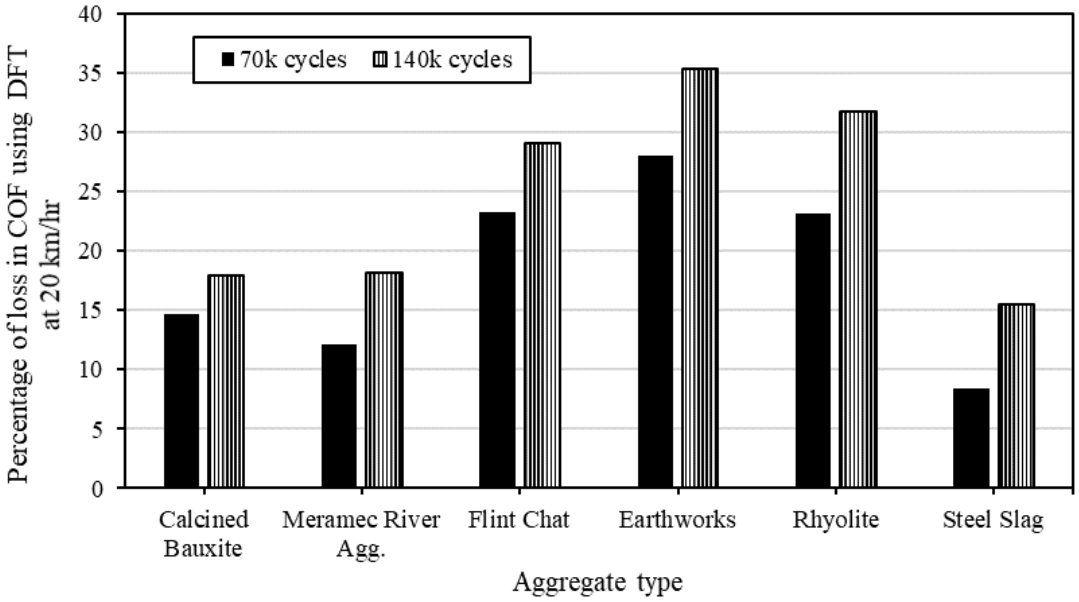


Figure 13: Percentages of Losses in COFs After Polishing

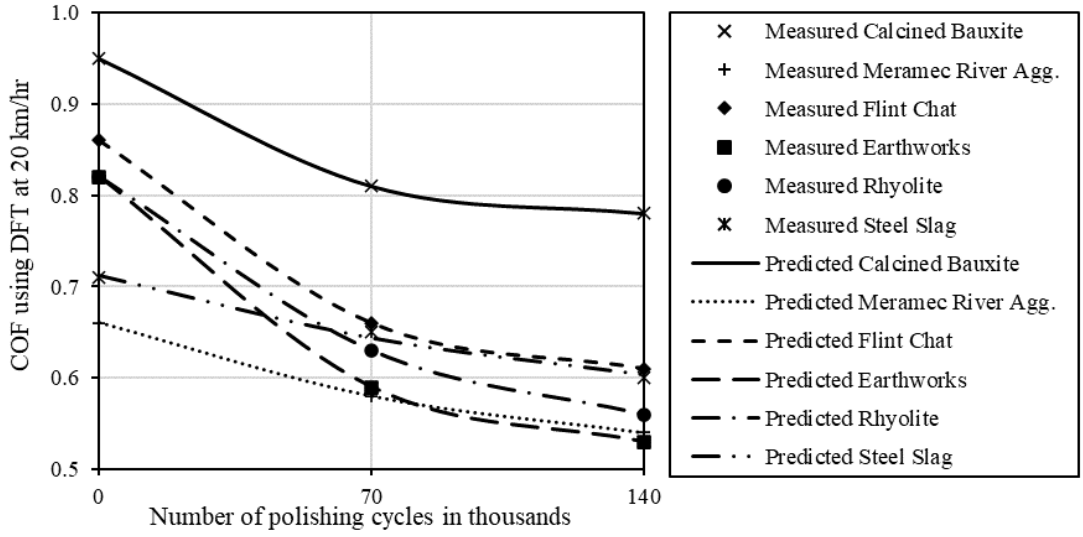


Figure 14: Relationship Between COFs and the Polishing Cycles' Number

Table 1: Fitting Parameters for the (DFT_{20}-N) Model

Aggregate	Fitting Parameters			*SSE*
	a	*b*	*c*	
Calcined Bauxite	7.718E−01	1.782E−01	2.20040E−02	1.174E−11
Meramec River Aggregate	5.000E−01	1.600E−01	9.90020E−03	2.058E−11
Flint Chat	5.933E−01	2.667E−01	1.98037E−02	1.954E−12
Earthworks	5.088E−01	3.112E−01	1.91958E−02	1.493E−12
Rhyolite	5.192E−01	3.008E−01	1.42638E−02	9.063E−12
Steel Slag	5.459E−01	1.663E−01	7.54360E−03	5.512E−05

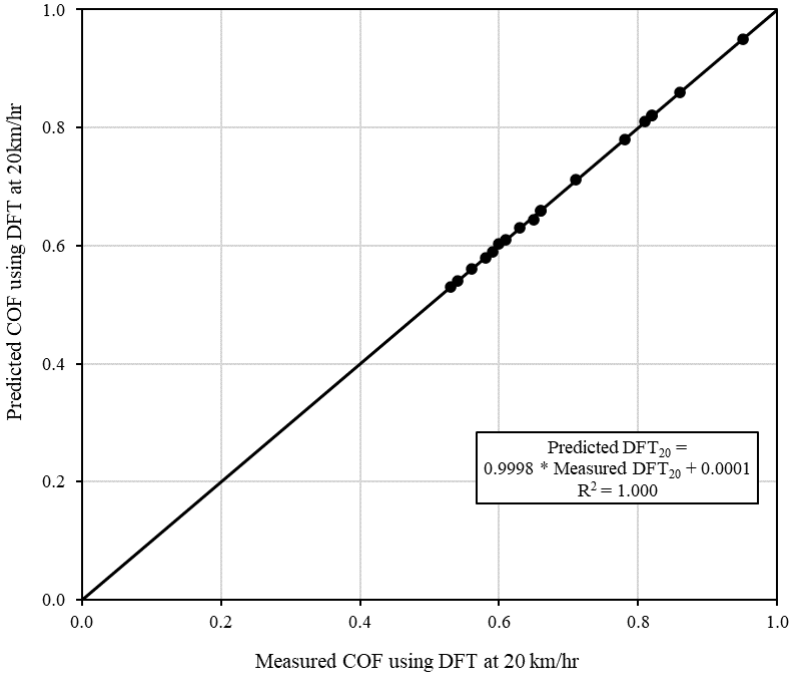


Figure 15: COF Values as Measured Versus Predicted

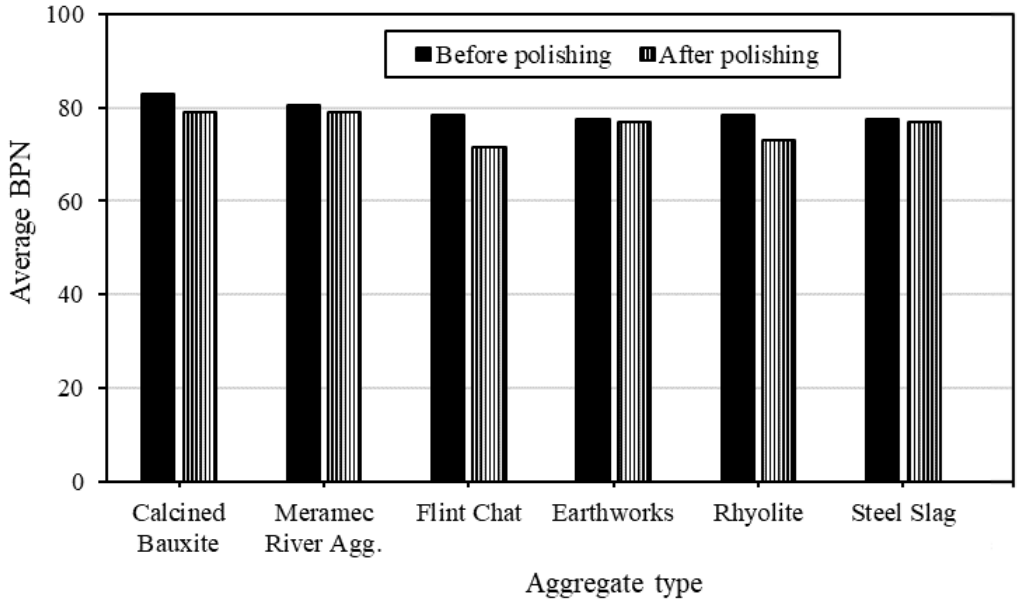


Figure 16: Average BPN Readings

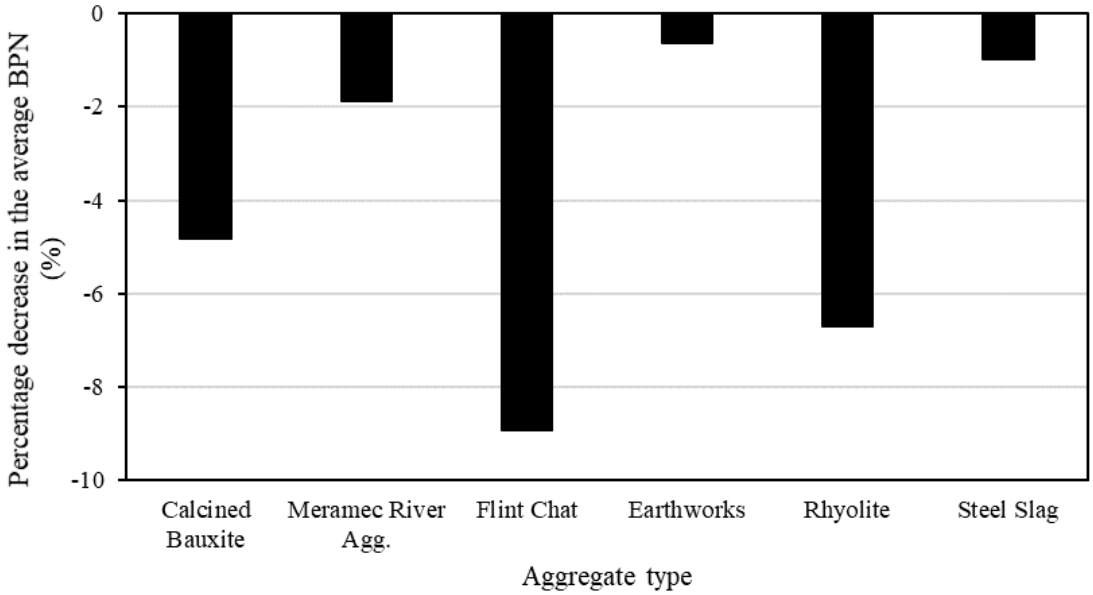


Figure 17: Percentage of Decrease in Average BPN Readings After Polishing

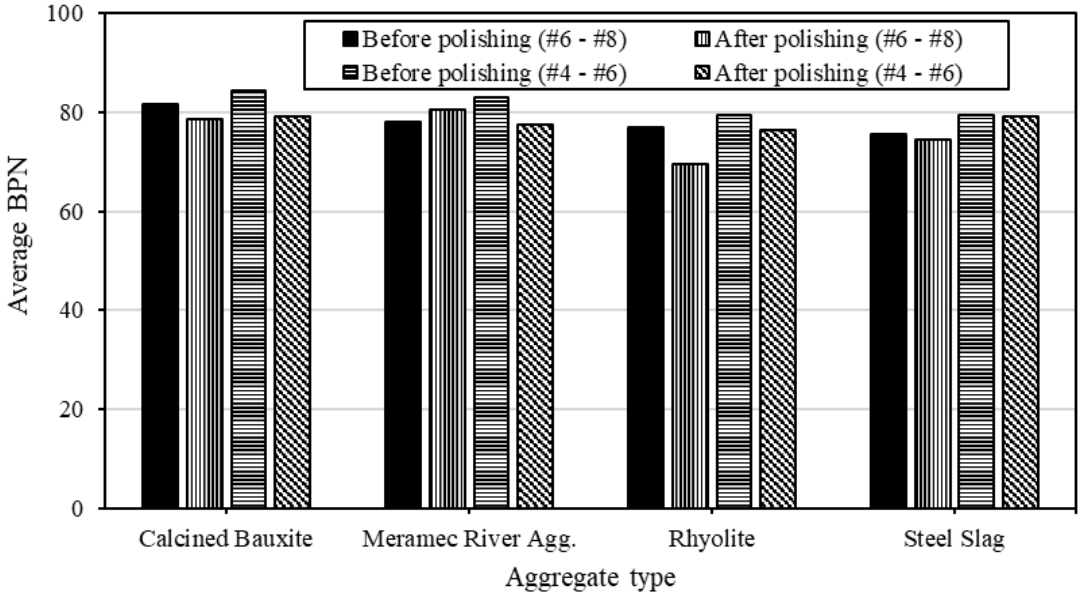


Figure 18: Average BPN Values with Two Sizes

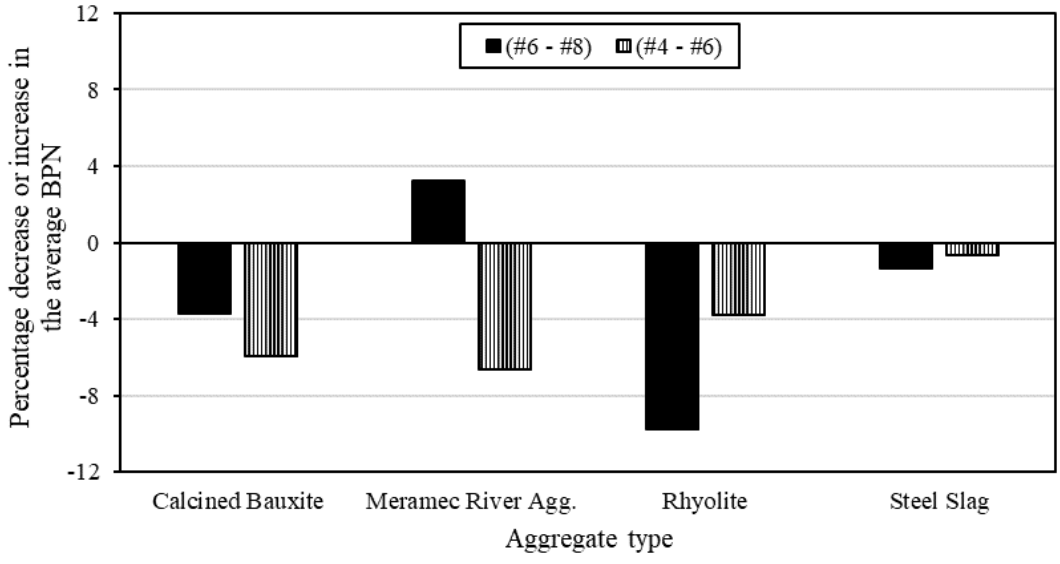


Figure 19: Percentages of Increase/Decrease in Average BPN Readings After Polishing

4. Conclusions

This study explored the frictional performance of Calcined Bauxite and five alternatives: Rhyolite, Earthworks, Meramec River Aggregate, Steel Slag, and Flint Chat. Additionally, among the alternatives, two byproducts recycled aggregates, Steel Slag and Flint Chat, were selected. The frictional performance tests comprised AIMS, accelerated friction testing (dynamic friction and sand patch tests), and BP testing. The effects of aggregates' sizes on the aggregates' frictional performances—AIMS and BP—were analyzed. The study led to the following conclusions:

1. Meramec River Aggregate, Flint Chat, and Steel Slag are considered to be substitutes for Calcined Bauxite.
2. The DFT results showed that Calcined Bauxite had the greatest COF values. Nonetheless, Meramec River Aggregate demonstrated, with values ranging from 1% to 5%, lower percentages of COF losses after polishing than Calcined Bauxite.
3. Percentages of losses in COF values with polishing cycles are good indicators for identifying Calcined Bauxite alternatives.
4. The polishing cycles' number and the COFs evaluated by DFT showed an exponential relationship. Increasing the polishing cycles' number lowered the COFs. After polishing, the COFs fell by 8% to 35%.
5. Flint Chat and Steel Slag had the greatest MTD values in the sand patch test, whereas Calcined bauxite had the lowest MTD value. Steel Slag and Flint Chat had MTDs that were 21% and 23% higher than those of Calcined Bauxite, respectively.
6. Calcined Bauxite had the highest average BPN before polishing, which was 3% greater than the average BPN for Meramec River Aggregate. However, Meramec River Aggregate had the highest average BPN after polishing, surpassing the value of Calcined Bauxite by 0.13%.
7. Larger aggregates had BPNs that were higher than those of smaller aggregates, with differences in values ranging from 0.8 to 7.
8. According to the AIMS data, Steel Slag had the highest average angularity and average texture indices AMD 180, outperforming Calcined Bauxite by 58.5% and 8%, respectively. Moreover, after polishing, Steel Slag showed 3.8% less BPN losses than Calcined Bauxite. This was related to the increase in texture indices AMD.
9. Percentages of change in texture and angularity indices after Micro-Deval abrasion can be used as screenings for alternatives to Calcined Bauxite.
10. Reducing aggregate size from 3/8" – 1/4" to 1/4" – #4 lowered texture indices by −7.3% to −16.5% and angularity indices by −1% to −6.9%, according to averaged AIMS indices based on BMD, AMD 105, and AMD 180.

5. Future Work

1. Evaluation of the frictional performance of HFST field sections with selected alternatives to Calcined Bauxite using DFT and BP.
2. The assessment of the frictional performance of other locally accessible aggregates with various aggregate sizes would allow the authors to validate their findings on a larger scale.

List of Acronyms

Acronym	Meaning
AIMS	Aggregate Image Measurement System
AMD	After Micro-Deval
BMD	Before Micro-Deval
BPN	British Pendulum Number
COF	Coefficient of Friction
CTM	Circular Texture Meter
DFT	Dynamic Friction Tester
HFST	High Friction Surface Treatment
HMA	Hot Mix Asphalt
MTD	Mean Texture Depth
TWPD	Three-Wheel Polishing Device

Conflict of Interest

The authors declare no conflict of interest.

Funding

This research was funded by MoDOT, project No. TR202005.

Acknowledgment

The authors thank MoDOT for technical and financial help in supplying equipment and information for this study. The DFT was carried out at the University of Idaho under the direction of Dr. Emad Kassem and Mr. Juan Pinto Ortiz.

Author Contributions

The authors confirm their contribution to the manuscript as follows: Conceptualization: Abdelrahman and Deef; methodology: Abdelrahman, Broaddus, and Deef; validation: Abdelrahman, Deef, and Broaddus; formal analysis: Abdelrahman and Deef; investigation: Abdelrahman, Broaddus, and Deef; resources: Abdelrahman; writing—original draft preparation: Abdelrahman, Deef, and Broaddus; writing—review and

editing: Abdelrahman and Deef; visualization: Abdelrahman and Deef; supervision, project administration, and funding acquisition: Abdelrahman; All authors reviewed the results and approved the final version of the manuscript.

References

[1] J. M. P. Mayora, R. J. Piña, "An assessment of the skid resistance effect on traffic safety under wet-pavement conditions," *Accident Analysis & Prevention*, vol. 41, pp. 881–886, 2009, doi:10.1016/j.aap.2009.05.004.

[2] H. Wang, Z. Wang, "Evaluation of pavement surface friction subject to various pavement preservation treatments," *Construction and Building Materials*, vol. 48, pp. 194–202, 2013, doi:10.1016/j.conbuildmat.2013.06.048.

[3] C. Chen, F. Gu, M. Heitzman, B. Powell, K. Kowalski, "Influences of alternative friction aggregates on texture and friction characteristics of high friction surface treatment," *Construction and Building Materials*, vol. 314, 2022, doi:10.1016/j.conbuildmat.2021.125643.

[4] G. Flintsch, I. L. Al-Qadi, R. Davis, K. K. McGhee, "Effect of HMA Properties on Pavement Surface Characteristics," *In Proceedings of the Pavement Evaluation Conference*, Roanoke, Virginia, U.S.A., 2002.

[5] Frequently asked questions about high friction surface treatments (HFST), FHWA-CAI-14-019, [Online]. Available online: https://www.fhwa.dot.gov/innovation/everydaycounts/edc-2/pdfs/fhwa-cai-14-019_faqs_hfst_mar2014_508.pdf (accessed on 4 August 2024).

[6] B. Wilson, A. Mukhopadhyay, "Alternative Aggregates and Materials for High Friction Surface Treatments," *Final Report, Project BDR74-977-05*, Texas A&M Transportation Institute, College Station, Texas, U.S.A., 2016.

[7] High friction surface treatments in Pennsylvania, Pennsylvania department of transportation, [Online]. Available online: https://www.tesc.psu.edu/assets/docs/high-friction-surface-treatments.pdf (accessed on 1 August 2024).

[8] High friction surface treatment (HFST) quick reference, DTFH61-13-D-00001, Task B9, FHWA, [Online]. Available online: https://www.fhwa.dot.gov/publications/research/safety/highfriction/High-Friction-Surface-Treatment-final.pdf (accessed 6 August 2024).

[9] D. Merritt, M. Moravec, M. Heitzman, High friction surface treatment aggregate durability study, pavement evaluation 2014, Blacksburg, Virginia, U.S.A., 2014, [Online]. Available online: https://vtechworks.lib.vt.edu/bitstream/handle/10919/54620/Merritt.pdf?sequence=1&isAllowed=y (accessed 24 July 2024).

[10] R. Milstead, X. Qin, B. Katz, J. Bonneson, M. Pratt, J. Miles, P. Carlson, "Procedures for Setting Advisory Speeds on Curves," *Final Report, FHWA-SA-11-22*, Federal Highway Administration, Office of Safety, Washington, D.C., U.S.A., 2011.

[11] D. L. Bloem, "Skid Resistance: The Role of Aggregates and Other Factors," *National Sand and Gravel Association Circular 109*, Silver Spring, MD, U.S.A., 1971.

[12] E. Deef-Allah, K. Broaddus, M. Abdelrahman, "Life cycle cost analysis of high friction surface treatment applications," *Transportation Research Record*, vol. 2676, pp. 512–526, 2022, doi:10.1177/03611981221079825.

[13] S. Li, R. Xiong, X. Dong, Y. Sheng, B. Guan, Y. Zong, C. Xie, J. Zhai, C. Li, "Effect of chemical composition of calcined bauxite aggregates on mechanical and physical properties for high friction surface course," *Construction and Building Materials*, vol. 302, 2021, doi:10.1016/j.conbuildmat.2021.124390.

[14] E. Deef-Allah, K. Broaddus, M. Abdelrahman, "Evaluation of Alternatives to Calcined Bauxite for Use in High Friction Surface Treatment (HFST) in Missouri," *Final Report, cmr 21-006*, Missouri Department of Transportation, Jefferson City, Missouri, U.S.A., 2021.

[15] E. Deef-Allah, K. Broaddus, M. Abdelrahman, "Physical properties and durability testing for calcined bauxite and its alternatives," *International Journal of New Technology and Research*, vol. 7, pp. 61–69, 2021, doi:10.31871/IJNTR.7.12.16.

[16] J. F. Bledsoe, H. S. Lee, "HFST Before and After Safety Analysis," *Final Report, cmr 21-003*, Missouri Department of Transportation, Jefferson City, Missouri, U.S.A., 2021.

[17] E. Kassem, A. Awed, E. A. Masad, D. N. Little, "Development of predictive model for skid loss of asphalt pavements," *Transportation Research Record*, vol. 2372, pp. 83–96, 2013, doi:10.3141/2372-10.

[18] P. S. Kandhal, Jr. F. Parker, "NCHRP Report 405: Aggregate Tests Related to Asphalt Concrete Performance in Pavements," *Final Report, Project D4-19 FY '94*, American Association of State Highway and Transportation Officials, Washington, D.C., U.S.A., 1998.

[19] L. K. Crouch, J. D. Gothard, G. Head, W. A. Goodwin, "Evaluation of textural retention of pavement surface aggregates," *Transportation Research Record*, vol. 1486, pp. 124–129, 1995.

[20] E. Deef-Allah, K. Broaddus, M. Abdelrahman, "Frictional performance correlations for calcined bauxite and alternative aggregates," *International Journal of New Technology and Research*, vol. 7, pp. 23–32, 2021, doi:10.31871/IJNTR.7.12.3.

[21] A. Roshan, M. Abdelrahman, "Evaluating friction characteristics of high friction surface treatment application under varied polishing and slippery conditions," *Transportation Research Record*, 2024, doi:10.1177/03611981241257505.

[22] E. Mahmoud, E. Masad, "Experimental methods for the evaluation of aggregate resistance to polishing, abrasion, and breakage," *Journal of Materials in Civil Engineering*, vol. 19, pp. 977–985, 2007, doi:10.1061/(ASCE)0899-1561(2007)19:11(977).

[23] E. Masad, A. Luce, E. Mahmoud, "Implementation of AIMS in Measuring Aggregate Resistance to Polishing, Abrasion, and Breakage," *Final Report, FHWA/TX-06/5-1707-03-1*, Texas Department of Transportation, Austin, Texas, U.S.A., 2006.

[24] E. Masad, A. Rezaei, A. Chowdhury, P. Harris, "Predicting Asphalt Mixture Skid Resistance Based on Aggregate Characteristics," *Final Report, FHWA/TX-09/0-5627-1*, Texas Department of Transportation, Austin, Texas, U.S.A., 2009.

[25] S. Li, R. Xiong, D. Yu, G. Zhao, P. Cong, Y. Jiang, "Friction Surface Treatment Selection: Aggregate Properties, Surface Characteristics, Alternative Treatments, and Safety Effects," *Final Report, FHWA/IN/JTRP-2017/09*, Indiana Department of Transportation, Indianapolis, Indiana, U.S.A., 2017, doi:10.5703/1288284316509.

[26] M. Heitzman, P. Turner, M. Greer, "High Friction Surface Treatment Alternative Aggregates Study," *Final NCAT Report 15--04*, Federal Highway Administration, Washington, D.C., U.S.A., 2015.

[27] ASTM D6928-17, "Standard Test Method for Resistance of Coarse Aggregate to Degradation by Abrasion in the Micro-Deval Apparatus," *ASTM International*, West Conshohocken, Pennsylvania, U.S.A., 2017, doi:10.1520/D6928-17.

[28] ASTM E965-15(2019), "Standard Test Method for Measuring Pavement Macrotexture Depth Using a Volumetric Technique," *ASTM International*, West Conshohocken, Pennsylvania, U.S.A., 2019, doi:10.1520/E0965-15R19.

[29] ASTM E1911-19, "Standard Test Method for Measuring Surface Frictional Properties Using the Dynamic Friction Tester," *ASTM International*, West Conshohocken, Pennsylvania, U.S.A., 2019, doi:10.1520/E1911-19.

[30] AASHTO T 278-90 (2017), "Standard Method of Test for Surface Frictional Properties Using the British Pendulum Tester," *AASHTO Provisional Standards*, Washington, D.C., U.S.A., 2017.

[31] AASHTO T 279-18, "Standard Method of Test for Accelerated Polishing of Aggregates Using the British Wheel," *AASHTO Provisional Standards*, Washington, D.C., U.S.A., 2018.

[32] E. Mahmoud, E. Ortiz, "Implementation of AIMS in Measuring Aggregate Resistance to Polishing, Abrasion, and Breakage," *Final Report, FHWA-ICT-14-014*, Illinois Department of Transportation, Springfield, Illinois, U.S.A., 2014.

Dr. Eslam Deef-Allah completed his bachelor's degree from Zagazig University in 2009. He obtained his master's degree from Zagazig University in 2015. He earned his Ph.D. from Missouri S&T in 2022. Deef-Allah is an assistant professor at Zagazig University's Faculty of Engineering, with over 25 publications in peer-reviewed journals, reports, and conference proceedings. He has more than 14 years of research and teaching experience. The areas of experience include characterization of modified asphalt binders with polymers or recycled materials, evaluation of extracted and recovered asphalt binders from mixes containing recycled materials, life cycle cost analysis, and assessing alternatives of Calcined Bauxite to be used in the high friction surface treatment applications.

Prof. Magdy Abdelrahman earned his bachelor's degree from Zagazig University in 1983. He obtained his master's degree from Zagazig University in 1988. He earned his Ph.D. from the University of Illinois at Urbana-Champaign in 1996.

Prof. Abdelrahman—Assoc. Chair for Graduate Studies and Research and Missouri Asphalt Pavement Association (MAPA) Endowed Professor—has over 100 publications in reputable international journals and conference proceedings. Research, industry, and teaching experiences are in the area of infrastructure sustainability with applications in pavement design, materials, and performance. Focus is on the interdisciplinarity aspects of infrastructure sustainability, making civil engineering more fundamental and more scientific. Integrating research into educational programs. National and international recognition. Service in national committees on construction/pavement areas. An expert in asphalt modifications including the use of recycled modifiers in civil/construction applications.

Received: 18 September, 2024, Revised: 17 October, 2024, Accepted: 18 October, 2024, Online: 08 November, 2024

DOI: https://doi.org/10.55708/js0311005

A Comparative Analysis of Interior Gateway Protocols in Large-Scale Enterprise Topologies

Saleh Hussein Al-Awami *, **Emad Awadh Ben Srity** , **Ali Tahir Abu Raas**

University of Benghazi, Department of Data Communications and Computer Networks, Benghazi, 00218, Libya

*Corresponding author: Saleh AL Awami, Department of Software Engineering Faculty of Information Technology, University of Benghazi ,Benghazi, Libya, & saleh.awami@uob.edu.ly

Author 2 Email: itstd.4074@uob.edu.ly , Author 3 Email: itstd.3522@uob.edu.ly

ABSTRACT: Interior gateway protocols (IGPs) have gained popularity in networking technologies due to their capacity to enable standardized and flexible communication among these algorithms. In autonomous systems (AS), network devices communicate with one another via IGPs. This work presents a fresh investigation into the performance of inner gateway protocols in large-scale enterprise topologies. Also, the experiment lab using GNS3 simulation has been conducted to evaluate and examine the performance of RIP, EIGRP, OSPF, and IS-IS, taking into account convergence time, latency, and jitter in large-scale network topology. The work has used a tri-connected architecture, with ten (10) routers connected via three serial connections and fifteen (15) network subnets, resulting in thirty (30) different paths for routing data packets for each tested routing technique. End-to-end delay, jitter, and convergence time are three measurement measures used to investigate network topologies. The experiment's outcomes have revealed that EIGRP has superior delay and convergence time performance. Furthermore, the results have been showed that IS-IS outperformed OSPF in terms of convergence time. Overall, this work improves the field by giving a grand average computation approach for measuring the jitter metric, which has been compared to a standard method. The method has been thoroughly explored utilizing derivate statistical equations and associated pseudo code.

KEYWORDS: RIP, EIGRP, OSPF, IS-IS, Interior gateway protocol, convergence time, In-depth analysis.

1. Introduction

Interior Gateway Protocols (IGPs) are routing protocols used to exchange routing information within an autonomous system (AS), such as a corporate network or a campus network. IGPs are essential for ensuring efficient and reliable communication among the routers and hosts within an AS. However, as the size and complexity of enterprise networks increase, so do the challenges and requirements for IGPs.

The process of routing involves determining the most efficient path for the flow of traffic within one or more networks. Routers utilize specific protocols that enable them to exchange relevant information on remote networks and update their routing tables regularly. These routing protocols determine the best path for each network, showcasing the interconnections among all routers within the network. This, in turn, facilitates communication primarily among neighboring routers and eventually throughout the entire network [1]. Routing can be classified into two categories: static and dynamic, as shown in Figure 1.

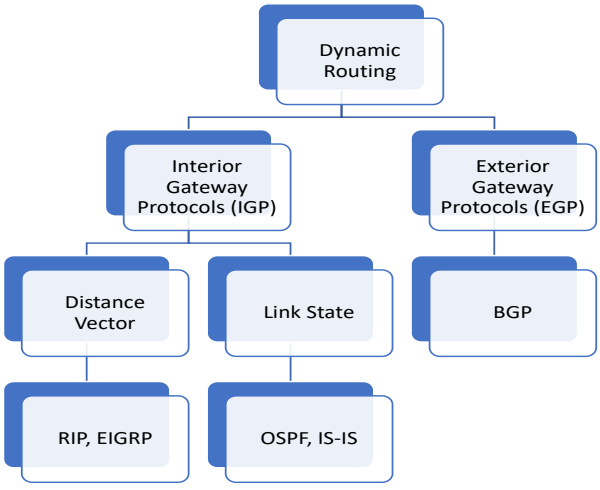


Figure. 1:Classified of Routing Protocols.

Static routing refers to the manual configuration of IP addresses and the manual entry of routes into the routing table by the network administrator before the routing process starts. Such routes can only be modified by the administrator and do not adapt to changes in the network, making them less suitable for large and unpredictable networks. They do not consume CPU memory or link

bandwidth, nor do they transmit information that can be intercepted by hackers:[1].

In contrast, dynamic routing utilizes routing protocols that allow the network administrator to input the protocol syntax instead of manually entering each IP address. These protocols are advantageous and practical in real-time as they can detect changes within the network and determine the shortest path. They read routing update messages and adjust to changing network conditions by recalculating the shortest path and transmitting the message:[2].

This article is organized into several sections. Section 2 covers Interior Gateway Protocols, while Section 3 is about related works. Section 4 discusses the problem statement, and Section 5 presents the proposed work. Section 6 describes the network topology model and measurement methodology. Section 7 presents the experiment results and proposed mathematical equations and comparing to the basic conventional one, and section 8 presents the discussion and Section 9 concludes the overall work of this article.

2. Interior Gateway Protocols

This section provides an overview of dynamic routing protocols in local area networks: Dynamic routing protocols can be classified into three main types: distance-vector, link-state, and hybrid. Distance-vector protocols rely on periodic exchanges of routing information between neighboring routers. Each router advertises the distance and direction to reach a destination network to its neighbors. Distance-vector protocols do not require routers to have a complete knowledge of the entire network topology. These protocols involve sending all or a portion of a router's routing table to adjacent routers. Examples of distance-vector protocols are RIP (version 1 and 2) and IGRP. Link-state protocols, on the other hand, require routers to collect link state information from all routers in the network and construct a network map based on that information. Then, each router calculates the best routes to each network using the network map. OSPF and IS-IS are examples of link-state protocols. Hybrid protocols combine the features of both distance-vector and link-state protocols.

2.1. RIP

Routing Information Protocol (RIP), a distance vector routing protocol, uses hops as the primary criterion for selecting a quality route. A maximum of 15 hops is allowed for RIP. RIP is also known as the Ford-Fulkerson or Bellman-Ford algorithm. There are three main versions of RIP: RIP v1, RIP v2, and RIPng. RIP information is carried in UDP packets. The port number for RIP V1 and V2 is 520 and the port number for RIPng is 521.The administrative distance assigned to the RIP protocol is 120. Routers that use RIP send their routing tables to their neighbors every 30 seconds. The updates are sent to the multicast address 224.0.0.9 for V1 and V2. The updates for RIPng are sent to the multicast address FF02 :: 9.

2.2. EIGRP

The Enhanced Interior Gateway Routing Protocol (EIGRP) is a hybrid routing protocol that combines the features of link-state and distance-vector protocols. It uses various criteria such as delay, bandwidth, reliability, and load to select the optimal path. It employs the Diffusion Update Algorithm (DUAL) for fast convergence and effective route optimization. It also uses the Reliable Transport Protocol (RTP) for delivering EIGRP packets. There are two versions of EIGRP: EIGRPv4 for IPV4 and EIGRPv6 for IPV6. The administrative distance of EIGRP is 90 by default, with a maximum hop count of 255. It sends packets to the multicast address 224.0.0.10 and uses the port number 88. Internal routes have an administrative distance (AD) of 90, while external routes have an AD of 170. The EIGRP metrics are computed using the constants K1 = 1, K2 = 0, K3 = 1, K4 = 0, and K5 = 0, as shown in (1), and in (2):.[3]

$$Metric = \left[K1 \times bandwith + \frac{(K2 \times bandwith)}{(256 - load)} + K3 \times delay\right] \times \left[\frac{K5}{reliability + K5}\right] \times 256 \quad (1)$$

$$\text{Metric} = 256 \times (Bandwith + delay) \quad (2)$$

2.3. OSPF

The Open Shortest Path First (OSPF) routing protocol utilizes Dijkstra's algorithm, a link-state routing algorithm. Dijkstra's algorithm calculates the shortest paths to all destinations from a source node, rather than a single source and destination pair, differing fundamentally from the Bellman-Ford algorithm and distance vector routing protocols. The routing metric in OSPF depends on the cost of a link. OSPF does not restrict the number of hops in a route. OSPF exists in two primary versions for IPv4(OSPFv2) and IPv6 (OSPFv3) networks, operating under the same principles. In OSPF, each router generates Link State Advertisements (LSAs) to establish and maintain a consistent representation of the routing topology across the entire domain. OSPF routers collect the network's link state data and store it in a Link State Database (LSDB). Each OSPF router determines the shortest path to each network segment using the Shortest Path First (SPF) algorithm. The well-known port for OSPF is 89. OSPF uses an administrative distance of 110. OSPFv2 utilizes multicast address 224.0.0.5, while OSPFv3 uses multicast address FF02::5. The OSPF metrics are illustrated in (3) [3].

$$Cost = \frac{10\text{^}8}{\text{bandwith}} \ (bps) \quad (3)$$

2.4. OSPF

The Intermediate System to Intermediate System (IS-IS) routing protocol operates similarly to OSPF. IS-IS is a link-state protocol developed by the International

Organization for Standardization (ISO) [4]. Like OSPF, IS-IS utilizes a link state database and Dijkstra's Shortest Path First (SPF) algorithm to determine the shortest path routes. IS-IS supports four distinct routing metrics: default, delay, expense, and error.

IS-IS does not limit the number of hops in routes and has an administrative distance of 115. IS-IS routers can be classified as level 1, level 2, or level 1-2 routers. Level 1 routers do not have direct connections to other areas, while level 2 routers connect multiple areas, similar to OSPF Area Border Routers (ABRs). Level 1-2 routers function as both level 1 and level 2 routers, connecting their area to the backbone.

3. Related Works

This study [2] focuses on the comparison of routing protocols used in computer networks and their performance metrics, including end-to-end delay, throughput, and convergence duration. The study utilizes the Optimized Network Engineering Tool (OPNET MODELER) to simulate different network topologies and compare the performance of three routing protocols: RIP, EIGRP, and OSPF. The simulation results show that OSPF has the fastest throughput among the three protocols, while EIGRP has the fastest convergence time. Additionally, EIGRP has faster throughput than RIP, but RIP has the highest queuing delay. This study provides valuable insights into the performance of different routing protocols and their suitability for different network topologies, but the methods of calculating performance metrics were not clarified, and the Riverbed Modeler they relied upon.

This study [5] compares the performance of EIGRP and OSPF routing protocols in terms of convergence time with link failures and the addition of new links in different network scenarios. The research uses a network simulator called GNS3 to simulate network topologies and validates the results using Cisco hardware equipment in the laboratory. The study finds that EIGRP has a faster convergence time than OSPF with a link failure or a new link added to the network. The experiment contributes to existing knowledge by identifying that the mesh topology has the best convergence time and that hardware implementations of routing protocols are better than using a network simulator. The study recommends further research in comparing BGP with EIGRP and OSPF and analyzing the protocol changes or adaptations in terms of convergence time with the versions of IPV4 and IPV6. Additionally, latency and quality of service are identified as vital areas of research in both EIGRP and OSPF routing protocols. The Wireshark results are used to check the network configuration and monitor accurate time responses for the various packets.

This study [6] compares the performance of two routing protocols, RIPv2 and EIGRP, in terms of data packet routing speed and convergence time. The research uses the Cisco Packet Tracer 7.10 software to simulate full mesh and half mesh topologies and conducts various tests, such as network connectivity, traceroute, and time testing. The study finds that EIGRP has a faster convergence time and routing speed than RIPv2 in both topologies. The results of the analysis suggest that EIGRP's hybrid routing characteristics (distance vector and link state) enable faster selection of vector-based paths and that EIGRP updates only the routing table of affected routers, while RIPv2 updates the routing table of all routers. The study concludes that EIGRP is a better routing protocol for an institution or college network that requires proper and safe network methods.

This study [7] aims to assess and compare the performance of different interior gateway routing protocols in an enterprise-level network that is complicated and operates in real-time. This is achieved by utilizing the GNS3 software to simulate the network, which consists of hosts, switches, and routers, and by putting each protocol into the topology. The protocols that have been examined are EIGRP, OSPF, and RIP. The authors guarantee secure data transmission at each node by offering authentication, and the assessment criteria include throughput, end-to-end delay, and convergence time. The study's findings indicate that EIGRP is the most practicable interior gateway routing protocol for a complicated enterprise-level network since it performs the best in terms of convergence time. In stable network conditions, there are no major differences observed in the end-to-end delay and throughput values of the three protocols; OSPF has a higher throughput than RIP and less significant delays than both RIP and EIGRP.

This study [8] aims to assess and contrast the effectiveness of many Interior Gateway Routing Protocols (IGRPs), more especially OSPF, RIPV2, and EIGRP. Using Graphical Network Simulator (GNS-3) simulations, the authors analyzed a number of characteristics, including throughput, jitter, convergence time, end-to-end delay, and packet depletion. EIGRP performs better than OSPF, according to the results. It was also mentioned that EIGRP uses greater processing power, which increases the system's power usage.

This study [6] the convergence times of the EIGRP and RIPv2 routing protocols were compared by the authors. Cisco Packet Tracer 7.10 was used to run the simulation on two different network topologies: full mesh and half mesh. The findings showed that the RIPv2 protocol took longer to converge—ranging from 0.01 to 0.19 seconds—while the EIGRP protocol had a faster average convergence time of 0.01-0.02 seconds.

4. Problem Statement

Network hardware performance is influenced by several factors, including convergence, delay, and jitter. Achieving optimal performance requires selecting the appropriate combination of routing protocols. Despite extensive research on interior gateway protocols, detailed computation of convergence, delay, and jitter has not been sufficiently addressed in these studies. In contrast, our paper employs equations to calculate these factors, providing a more comprehensive and accurate evaluation of network hardware performance, then the proposed analysis has .

5. Proposed Work

The primary aim of this work is to evaluate and analyze the performance of a complex enterprise network by considering convergence, delay, and jitter, and utilizing Interior Gateway Protocols such as OSPF, EIGRP, and IS-IS. Furthermore, the study seeks to determine the most suitable protocol for large-enterprise topologies. Notably, this research contributes to the field by introducing an enhancement analysis of the "delay, jitter, which is a factor in achieving optimal network performance. In order to obtain a higher accuracy measured value than the fundamental traditional method, the grand average statistical method has been utilized in the proposed method to quantify jitter, "delay", in a large-scale network topology. In order to determine the differences in the jitter, "delay", in the proposed Tri-connected network devices topology in this work, a comparison is made between the suggested approach and the conventional one using the interior gateway routing protocols.

6. Models and Measurement Parameters

This section provides an overview of the devices and media types employed in setting up the network topology, as well as an examination of the software used. Additionally, the measurement parameters used to analyze the network topologies are described.

Partial mesh (Tri-connected) topology is a network topology combining features of full mesh and star topologies, providing high fault tolerance and redundancy. It is commonly used in medium to large enterprises, offering a scalable traffic handling capability. Nodes are connected to a subset of other nodes or all nodes in the network. It is ideal for medium to large enterprises requiring a network that can handle a moderate to a high amount of traffic. Partial mesh topology is less expensive than a full mesh topology. It is an effective solution for enterprise networks that require fault tolerance, redundancy, and scalability.

The proposed network topologies comprise ten routers interconnected through three serial connections, with 15 network subnets forming 30 multiple paths for routing data packets for each assessed routing protocol. This configuration results in a complex network topology at the enterprise level, as illustrated in Figure 2 below.

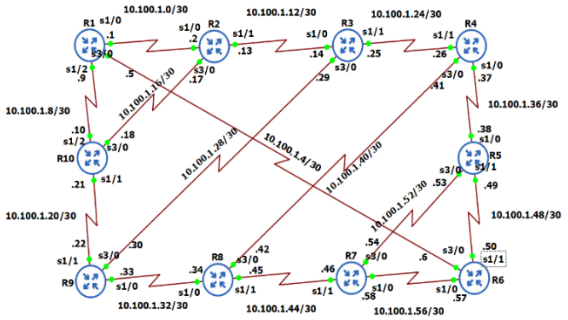


Figure. 2: Large-scale enterprise network topology.

The proposed large-scale infrastructure is built to ensure it ideal for enterprise applications and the complexity of network devices. The network configuration is built to meet the following:

A) **Network topology**: ten (10) routers connected via three (3) serial connections to interconnect using a single channel to routing packets.

B) **Subnets:** fifteen (15) different subnets to ensure the complexity of the topology.

C) **Routing path**: thirty (30) routing path for data packet routing (Tri-connected 10 routers).

D) **Routing protocol**: various interior routing protocols assessed for efficiency.

E) **Network topology Size**: Large-scale network topology

6.1. Hardware and software resources

This subsection provides details on the hardware and software resources utilized for the simulation and testing. The IP addressing configuration and network topology used in the experiments are illustrated in Figure 2 above. The network topology comprises ten Cisco 7200 routers, providing multiple paths between different sections of the topology. The routers used in the experiments are identical in model and specifications, as listed in Table 1. The software used to perform the experiments is GNS3 version 2.2.34, while the HP ProBook 450 5G was used as the testing device, with its specifications outlined in Table 2.

Table 1: Network devices employed in the network topology

Factor	Description
Device Model	Cisco 7200 router
Processor	NPE- 400
RAM Memory	512MB
Flash Memory	32 MB

Table 2: Laptop device specification for Personal Computer use.

Factor	Description
Model	HP ProBook 450 5G
Processor	Intel(R) Core (TM) i5-8250U
RAM Memory	8G
CPU Speed	CPU @ 1.60GHz
Operating System	Windows 11 Pro

From the standpoint of three initial parameters, network performance was examined using the same network topology and different protocols.

The measurement parameters would be used in this work can be defined as follows:

A) End to end delay

The total time required for a complete message to travel from its source to destination encompasses the duration from the initial transmission of the first bit of the message until the final delivery of the last bit of the message.

B) Jitter

The variation in packet delay, commonly referred to as packet delay variance, is an indication of the presence of jitter when data packets experience disparate delays across a network. Jitter is often measured in milliseconds (MS).

C) Convergence time

The metric measures the speed at which a group of routers restores a network to its normal state.

7. Experiment Results

In this section, each interior gateway protocol presented in this paper was evaluated to determine the most suitable protocol for the network topology under consideration. This evaluation was based on the calculation of end-to-end delay, convergence time, and jitter, in accordance with the study's objectives.

7.1. End-to-End Delay Testing Results

Table 3 presents the average end-to-end delay values for EIGRP, OSPF, and IS-IS, which were determined using the ping utility throughout the analysis.

Table 3: The average end-to-end delay values for the RIP, EIGRP, OSPF, and IS-IS protocols calculated for each subnet

Subnet #	IP Address	RIP	EIGRP	OSPF	IS-IS
1	10.100.1.1/30	170	40	48	42
2	10.100.1.5/30	196	42	24	50
3	10.100.1.9/30	152	42	40	48
4	10.100.1.13/30	83	24	26	32
5	10.100.1.17/30	26	20	22	20
6	10.100.1.21/30	21	20	22	25
7	10.100.1.25/30	40	39	32	30
8	10.100.1.29/30	48	42	57	27
9	10.100.1.33/30	35	31	28	42
10	10.100.1.37/30	57	56	66	64
11	10.100.1.41/30	77	68	86	52
12	10.100.1.45/30	46	58	65	60
13	10.100.1.49/30	40	40	38	42
14	10.100.1.53/30	44	38	36	37
15	10.100.1.57/30	26	21	24	19

Table 3 represents end to end delay for each protocol by taking the average of round-trip time (RTT) by using ping utility. Figures. 3, 4, and 5 represent samples of results of sixty (60) experiments for four (4) measured interior gateway routing protocol for fifteen (15) different subnets.

Figure 3: The Ping utility executed on 10 routers utilizing the RIP protocol.

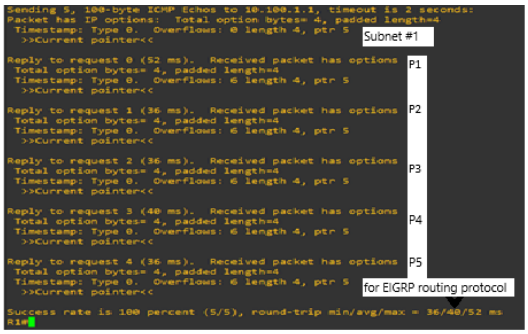


Figure. 4: The Ping utility executed on 10 routers utilizing the EIGRP protocol

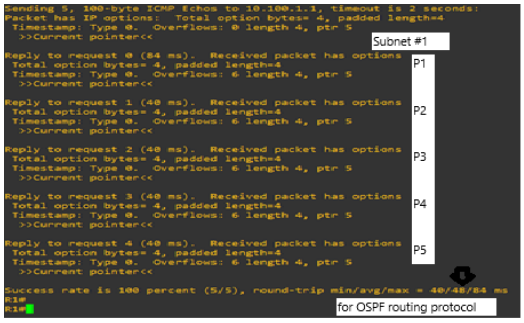


Figure 5: The Ping utility executed on 10 routers utilizing the OSPF protocol.

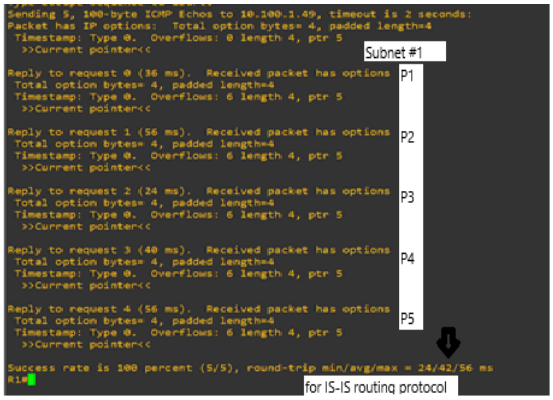


Figure 6: The Ping utility executed on 10 routers utilizing the IS-IS protocol.

The obtained results of the previous four Figures are measured using Ping utility using Internet Communication Message Protocol (ICMP) echoes. The round-trip is measured through packet traveling using different interior gateway routing protocol in each time using the same large-scale proposed network topology. As shown above, the results are different from one interior gateway routing protocol to other interior gateway routing protocol.

Figure 7 represents the delay variation, from the values in Table 3, so can see that for the 1st case EIGRP has the least delay, whereas the 2nd and the 3rd case OSPF has the least delay. For the 4th, 5th and 6th case EIGRP has the least delay. For the 7th and 8th case IS-IS has the least delay. For the 9th ,13th and 14th OSPF has the least value. So EIGRP has the least delay among the OSPF and IS-IS.

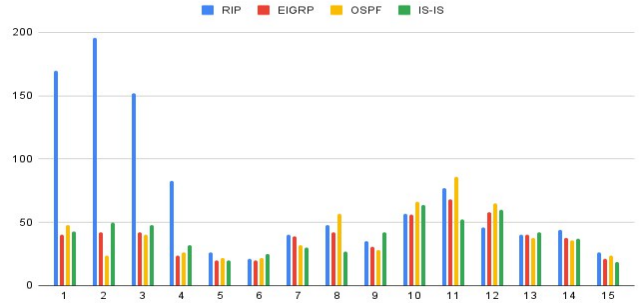


Figure 7: End-to-End Delay variation graph for RIP EIGRP, OSPF, and IS-IS.

The Table 4 below displays the average values of the end-to-end delay for RIP, EIGRP, OSPF, and IS-IS which has been computed by taking the average of the Table 3 averages for each protocol separately.

Table 4: Comparison of Average End to End Delay for RIP, EIGRP, OSPF and IS-IS.

S.No.	TYPE OF PROTOCOL	AVERAGE END TO END DELAY TIME (ms)
1	RIP	70.73
2	EIGRP	38.73
3	OSPF	40.93
4	IS-IS	39.33

The Figure 8 below represents the average end-to-end delay from the values in Table 4, and by analyzing Figure. 7, can see that EIGRP has the least average value of delay and does not differ much from IS-IS and OSPF. Figure 7 shows that OSPF has the worst average value of delay.

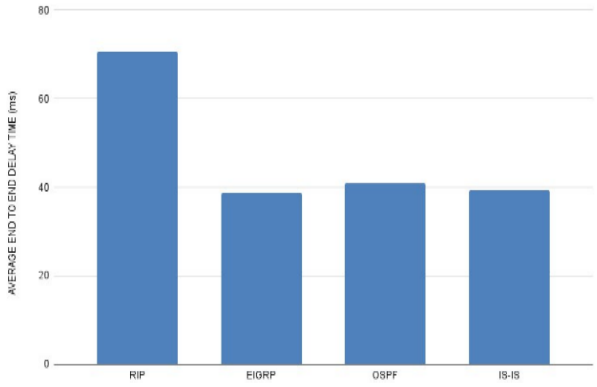


Figure 8: Delay Variation for RIP, EIGRP, OSPF and IS-IS.

7.2. *In depth analysis for Jitter Testing Results*

A) Proposed Method

In this section, to obtain the most accurate jitter result, a delay of 75 packets (fifteen (15) subnets multiplied by five (5) received messages using Ping utility using ICMP echoes) has been calculated for each protocol, and 60 jitter results were obtained for all interior gateway routing protocol By calculating the differences between the delays for the five packets, respectively, as shown in Tables 5, 6, 7, and 8 below.

An example of this is that the result of the first jitter is the difference between the delay for the first and second packets, and the result of the second jitter is the difference between the delay for the second and third packets, and so on. Let R represents the jitter result, i represents the number of received packet sequence within one subnet, and P represents packet delay as shown in (4).

$$R_i = |P_{i+1} - P_i| \tag{4}$$

The Table 9 below displays the average values of the in-depth analysis of jitter for RIP, EIGRP, OSPF, and IS-IS, which has been calculated by taking the average of 60 (the volume of the data packets is equal to fifteen (15) subnets multiply by four (4) different interior gateway routing protocols) values of jitter from the Tables 5, 6, 7, and 8 above for each protocol separately. Let R represent the Jitter result for each two send packets obtained in (4), i represent the iteration of measurement jitters in each subnet, and K represents the number of jitter results equal to 4 (i-th of 4) for each interior gateway routing protocol, and AVG is the average value of jitter measure calculated in each subnet iteration (i-th of 15).

$$AVG_{ith} = \frac{\sum_{i=1}^{K} R_i}{\mathrm{K}} \tag{5}$$

Table 5: In depth analysis JITTER result for RIP.

IP	End-to-End Delay					Jitter (Traditional Method)	Jitter (Ignored remained messages by traditional method of measuring Jitter calculation)			This Work Proposed Method
	Packet P1	Packet P2	Packet P3	Packet P4	Packet P5	Result R1=|P2-P1|	Result R2=|P3-P2|	Result R3=|P4-P3|	Result R4=|P5-P4|	AVG= (R1+R2+R3+R4)/4
10.100.1.1/30	176	172	168	172	164	4	4	4	8	5
10.100.1.5/30	184	308	144	168	176	124	164	24	8	80
10.100.1.9/30	139	136	144	172	172	3	8	28	0	9.75
10.100.1.13/30	72	88	84	84	88	16	4	0	4	6
10.100.1.17/30	40	16	28	12	36	24	12	16	24	19
10.100.1.21/30	40	24	20	16	8	16	4	4	8	8
10.100.1.25/30	44	40	36	48	36	4	4	12	12	8
10.100.1.29/30	40	80	24	72	28	40	56	48	44	47
10.100.1.33/30	36	44	36	36	24	8	8	0	12	7
10.100.1.37/30	40	80	56	52	60	40	24	4	8	19
10.100.1.41/30	72	96	56	76	88	24	40	20	12	24
10.100.1.45/30	76	52	56	16	32	24	4	40	16	21
10.100.1.49/30	40	36	44	40	40	4	8	4	0	4
10.100.1.53/30	48	76	20	36	40	28	56	16	4	26
10.100.1.57/30	40	32	20	20	20	8	12	0	0	5
Grand Average for overall large-scale enterprise network topology						24.46666667	All are used in new proposed method in this work.			19.25

Table 6: In depth analysis JITTER result for EIGRP

IP	End-to-End delay					Jitter (Traditional Method)	Jitter (Ignored remained messages by traditional method of measuring Jitter calculation)			This Work Proposed Method
	Packet P1	Packet P2	Packet P3	Packet P4	Packet P5	Result R1=|P2-P1|	Result R2=|P3-P2|	Result R3=|P4-P3|	Result R4=|P5-P4|	AVG= (R1+R2+R3+R4)/4
10.100.1.1/30	52	36	36	40	36	16	0	4	4	6
10.100.1.5/30	52	36	44	40	40	16	8	4	0	7
10.100.1.9/30	56	32	44	40	40	24	12	4	0	10
10.100.1.13/30	44	20	16	20	20	24	4	4	0	8
10.100.1.17/30	36	16	12	20	20	20	4	8	0	8
10.100.1.21/30	36	16	20	16	16	20	4	4	0	7
10.100.1.25/30	36	68	16	40	36	32	52	24	4	28
10.100.1.29/30	32	52	36	60	32	20	16	24	28	22
10.100.1.33/30	44	24	32	28	28	20	8	4	0	8
10.100.1.37/30	64	28	36	76	76	36	8	40	0	21
10.100.1.41/30	68	88	52	68	64	20	36	16	4	19
10.100.1.45/30	68	60	64	52	48	8	4	12	4	7
10.100.1.49/30	32	52	44	24	52	20	8	20	28	19
10.100.1.53/30	32	76	20	24	40	44	56	4	16	30
10.100.1.57/30	36	16	20	20	16	20	4	0	4	7
Grand Average for overall large-scale enterprise network topology						22.66666667	All are used in new proposed method in this work.			13.8

Table 7: In depth analysis JITTER result for OSPF

IP	End-to-End delay					Jitter (Traditional Method)	Jitter (Ignored remained messages by traditional method of measuring Jitter calculation)			This Work Proposed Method
	Packet P1	Packet P2	Packet P3	Packet P4	Packet P5	Result R1=|P2-P1|	Result R2=|P3-P2|	Result R3=|P4-P3|	Result R4=|P5-P4|	AVG= (R1+R2+R3+R4)/4
10.100.1.1/30	84	40	40	40	40	44	0	0	0	11
10.100.1.5/30	44	20	20	16	20	24	0	4	4	8
10.100.1.9/30	68	16	52	40	28	52	36	12	12	28
10.100.1.13/30	36	28	24	20	24	8	4	4	4	5
10.100.1.17/30	40	20	20	20	12	20	0	0	8	7
10.100.1.21/30	36	24	20	24	8	12	4	4	16	9
10.100.1.25/30	32	52	24	32	24	20	28	8	8	16
10.100.1.29/30	64	56	52	56	60	8	4	4	4	5
10.100.1.33/30	44	20	40	20	16	24	20	20	4	17
10.100.1.37/30	80	56	84	48	64	24	28	36	16	26
10.100.1.41/30	84	112	72	76	88	28	40	4	12	21
10.100.1.45/30	80	60	72	52	64	20	12	20	12	16
10.100.1.49/30	40	40	40	36	36	0	0	4	0	1
10.100.1.53/30	48	40	24	36	36	8	16	12	0	9
10.100.1.57/30	40	24	24	24	8	16	0	0	16	8
Grand Average for overall large-scale enterprise network topology						20.53333333	All are used in new proposed method in this work.			12.46666667

Table 8: In depth analysis JITTER result for IS-IS

IP	End-to-End delay					Jitter (Traditional Method)	Jitter (Ignored remained messages by traditional method of measuring Jitter calculation)			This Work Proposed Method
	Packet P1	Packet P2	Packet P3	Packet P4	Packet P5	Result R1=|P2-P1|	Result R2=|P3-P2|	Result R3=|P4-P3|	Result R4=|P5-P4|	AVG= (R1+R2+R3+R4)/4
10.100.1.1/30	76	40	32	40	28	36	8	8	12	16
10.100.1.5/30	92	36	40	40	44	56	4	0	4	16
10.100.1.9/30	88	28	40	44	40	60	12	4	4	20
10.100.1.13/30	40	68	16	16	20	28	52	0	4	21
10.100.1.17/30	36	32	24	8	4	4	8	16	4	8
10.100.1.21/30	40	20	20	24	24	20	0	4	0	6
10.100.1.25/30	12	56	44	16	24	44	12	28	8	23
10.100.1.29/30	24	12	44	36	20	12	32	8	16	17
10.100.1.33/30	68	24	40	44	36	44	16	4	8	18
10.100.1.37/30	80	76	68	8	92	4	8	60	84	39
10.100.1.41/30	32	60	40	76	52	28	20	36	24	27
10.100.1.45/30	72	72	56	52	48	0	16	4	4	6
10.100.1.49/30	36	56	24	40	56	20	32	16	16	21
10.100.1.53/30	56	16	80	20	16	40	64	60	4	42
10.100.1.57/30	1	40	16	8	32	39	24	8	24	23.75
Grand Average for overall large-scale enterprise network topology						29	All are used in new proposed method in this work.			20.25

In order to get the grand average (GAVG) of each jitter measured interior routing protocol in the proposed network topology that has fifteen (15) subnets the calculation in (6) has been used, using the obtained measured of average iteration values in (5) for each subnet (S of total of fifteen (15) in this experiment).

$$\text{GAVG} = \frac{\sum_{ith=1}^{S} AVG_{ith}}{S} \quad (6)$$

The Table 9 shows the calculated grand average of overall measured value for jitter in the proposed large-scale enterprise network topology, refer to Figure 2 above. These measured values are obtained using the proposed method that is presented in this work.

Table 9: Comparison of Average jitter for RIP, OSPF, EIGRP and IS-IS using proposed method

S.No.	TYPE OF PROTOCOL	JITTER GRAND AVERAGE (ms)
1	RIP	19.25
2	EIGRP	13.8
3	OSPF	12.46666
4	IS-IS	20.25

The Figure 9 below represents the grand average jitter from the values in Table IX, OSPF's routing protocol has the lowest average value of jitter compared to RIP protocols, EIGRP protocol, and IS-IS protocol. Especially IS-IS, the IS-IS protocol has the highest average value of jitter.

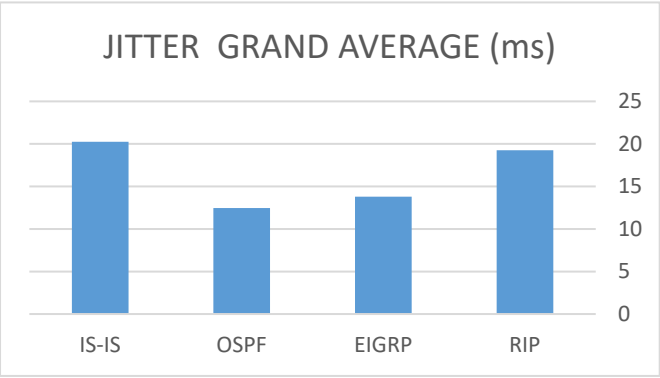


Figure. 9: Proposed Method for In Depth Jitter Variation Graph for RIP, EIGRP, OSPF and ISIS

B) Traditional Method to measure Jitter

This section also presents the classic technique of calculating Jitter, which is a well-known method for calculating the variance of only the first and second packets of replay to the requests messages using ICMP Echoes: [9], [10], [11] and [12].

Table 10 shows the calculated grand average of overall measured values for jitter in the proposed large-scale enterprise network topology, refer to Figure 2. These measured values are obtained using the traditional method of jitter metric, refer to Tables 5, 6, 7, and 8 above for each protocol separately.

Table 10: Comparison of Average jitter for RIP, OSPF, EIGRP and IS-IS using traditional method

S.No.	TYPE OF PROTOCOL	JITTER GRAND AVERAGE (ms)
1	RIP	24.46666667
2	EIGRP	22.66666667
3	OSPF	20.53333333
4	IS-IS	29

The Figure 10 below represents the grand average jitter from the values in Table 10, OSPF's routing protocol has the lowest average value of jitter compared to RIP protocols, EIGRP protocol, and IS-IS protocol. Especially IS-IS, the IS-IS protocol has the highest average value of jitter.

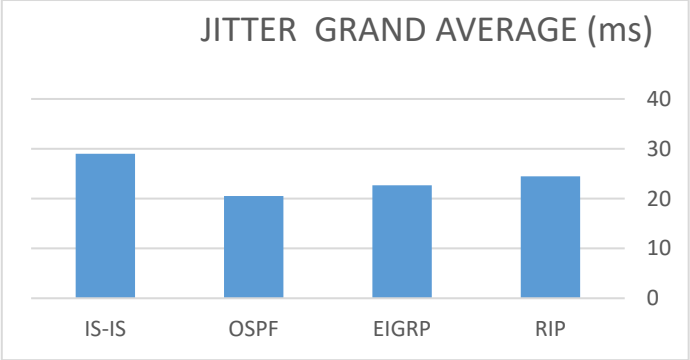


Figure 10: Traditional Method for Jitter Variation Graph for RIP, EIGRP, OSPF and ISIS

In order to get more clear of the results of both traditional mothed and proposed method, the comparing are shown in Figure 11 and in Table 11. The variance between both method is clear and indicate the difference that should be taken in account when measuring the Jitter metric especially in the large-scale enterprise network topology.

Table 11: Comparison of grand Average jitter for RIP, OSPF, EIGRP and IS-IS using traditional and proposed methods.

S. No.	Type of Protocol	Traditional Method (Jitter)	Proposed Method (Jitter)	The differences
1	RIP	24.46666667	19.25	5.216667
2	EIGRP	22.66666667	13.8	8.866667
3	OSPF	20.53333333	12.46666	8.066673
4	IS-IS	29	20.25	8.75

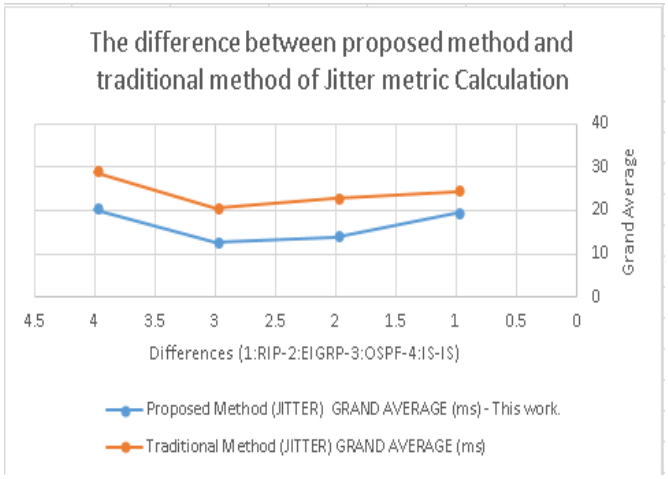


Figure 11: The veriation of Jitter metric using different measuring methods for In Depth Jitter Variation for RIP, EIGRP, OSPF and ISIS.

The Figure 11 above shows the differences that should be considered to pick the accuracy method to measure the Jitter in small, mid-sized, or large-scale network topology. The new method uses in-depth analysis technique using statistical method to improve the result of the measure using all reply messages to calculate the grand average when compare it to the traditional measuring method. The pseudo code for Jitter calculation using the proposed method in this work is illustrated program algorithm 1 below:

Algorithm 1: JITTER_PROPOSED_MEASURE.

Result: Grand_Average_Jitter_Measurement
Begin
REM %Initialization%;
REM SEND 100 Byte request ICMP Packet using PING Utility;
REM SEND 100 Byte request ICMP Packet using PING Utility.
REM % RECEIVE five (5) Messages by reply to requests%
REM %Let Packet represents as to receive messages time in milliseconds. %
REM % suppose the Received_Message_Time_ms array contains the time for the 5 received messages in milliseconds using ICMP Echo. %
double Traveled_Packet[1..5];
int Result [1..4] = 0;
double Received_Message_Time_ms[1..5] ← get time in milliseconds;
double SUM = 0.0;
double AVERAGE [1..15] = 0.0;
double AVERAGE_SUM = 0.0;
int K= 4,
int Number_of_Request_Messages = 5;
int Number_of_Subnets = 15;
For (**int** i:=1 to Number_of_Request_Messages)
    Traveled_Packet[i] =
    Received_Message_Time_ms[i];
Next i
REM %Calculate the variance between received messages for each subnet in the large-scale enterprise network topology. %
For (**int** S=1 to Number_of_Subnets)
For (**int** i=1 to Number_of_Request_Messages – 1)
    Result[i] = abs(Packet[i+1]-Packet[i]);
    SUM = SUM+Result[i];
Next i
  AVERAG[S] = SUM/K;
   AVERAGE_SUM = (AVERAGE_SUM +
   AVERAG[S]);
Next S
GRAND_AVERAG_JITTER_MEASURE = (AVERAGE_SUM/ Number_of_Subnets);
Print GRAND_AVERAG_JITTER_MEASURE;
END

7.3. *Convergence Time Testing Results*

When a route becomes unavailable, the router requires a certain amount of time to notify other routers of the failure and determine the best alternative path [13], [14], [15] and [16].

This process is commonly referred to as the convergence of time. In our experiment, packets of 100-datagram size and a 2-second timeout were transmitted from the source to the destination with a repeat count of 500. While sending these packets, we intentionally disrupted one of the routes and observed the subsequent loss of packets until the new path information was fully propagated to all routers. To determine the convergence time, we calculated the number of lost packets and multiplied it by the 2-second timeout. This process was repeated five times, and the average was calculated. Refer to Figures 12, 13, and 14 below.

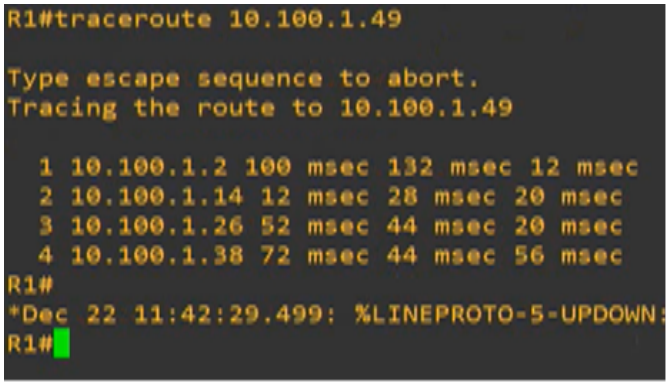


Figure 12: Traceroute for 10.100.1.49 Before Failure

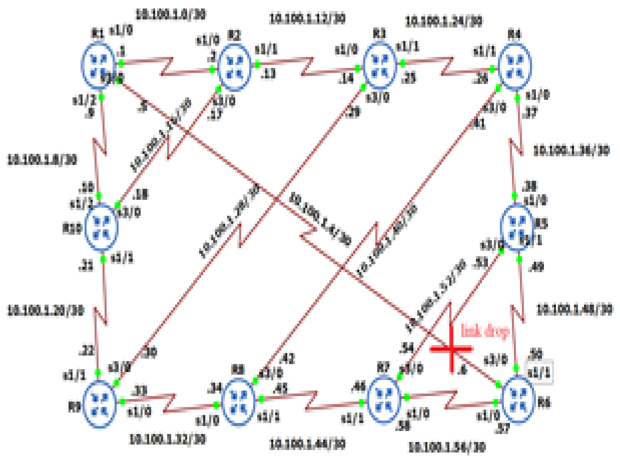


Figure 13: Link Failure for 10.100.1.49.

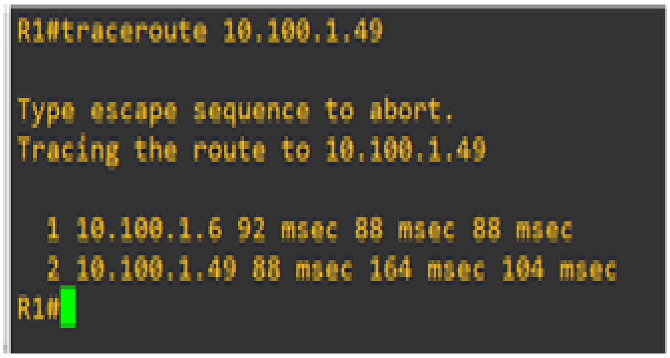


Figure. 14: Traceroute for 10.100.1.49 After Failure.

Table 12: Convergence Time of RIP.

S.No.	Packets received from 500 packets	Packet lost	Convergence time (s)	Average convergence time (s)
1	468	32	64	63.6
2	464	36	72	
3	471	29	58	
4	474	26	52	
5	464	36	72	

Table 13: Convergence Time of EIGRP

S.No.	Packets received from 500 packets	Packet lost	Convergence time (s)	Average convergence time (s)
1	489	11	22	20.8
2	488	12	24	
3	490	10	20	
4	489	11	22	
5	492	8	16	

Table 14: Convergence time of ospf

S.No.	Packets received from 500 packets	Packet lost	Convergence time (s)	Average convergence time (s)
1	481	19	38	40.8
2	480	20	40	
3	479	21	42	
4	479	21	42	
5	479	21	42	

Table 15: Convergence Time of IS-IS.

S.No.	Packets received from 500 packets	Packet lost	Convergence time (s)	Average convergence time (s)
1	486	14	28	30
2	484	16	32	
3	485	15	30	
4	485	15	30	
5	485	15	30	

Tables 12, 13, 14, and 15 above represent the convergence times and the average convergence times for each protocol. Figures 15, 16, 17, and 18 below represent samples of results.

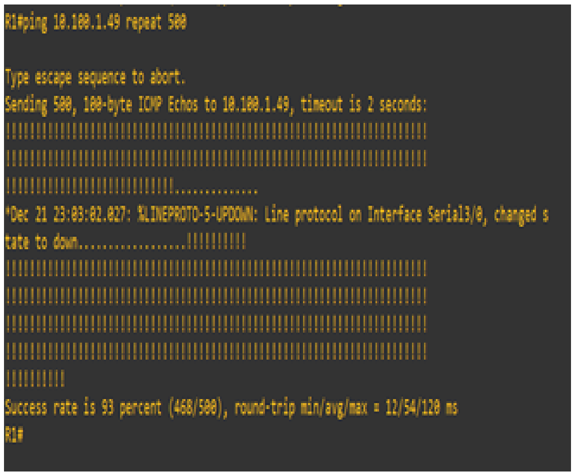


Figure 15: RIPConfiguration When R6's Port s3/0 is Blocked.

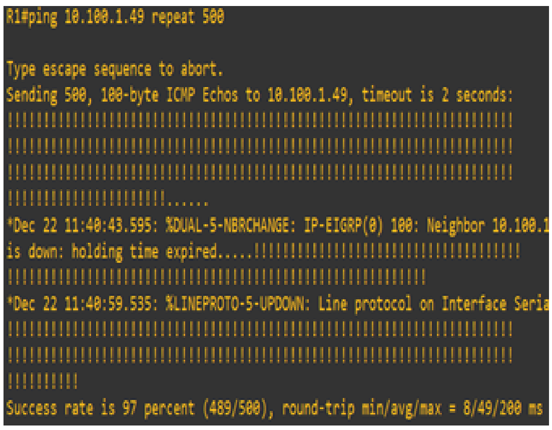


Figure 16: EIGRP Configuration When R6's Port s3/0 is Blocked.

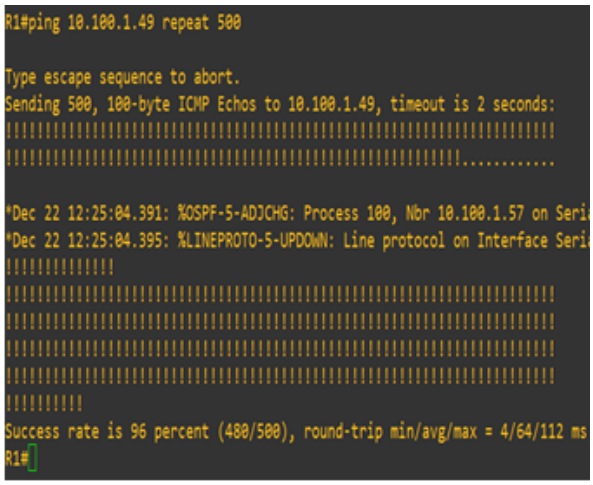


Figure 17: OSPF Configuration When R6's Port s3/0 is Blocked.

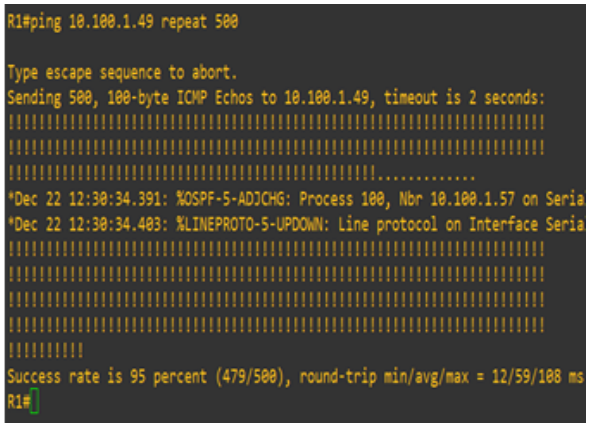


Figure 18 : IS-IS Configuration When R6's Port s3/0 is Blocked.

The Figure 19 below represents the average of convergence time from the values in Tables 12, 13, 14, and 15 above, and by analyzing Figure 19, so can see that EIGRP has the least average value of convergence time and the OSPF has the worst average value of convergence time.

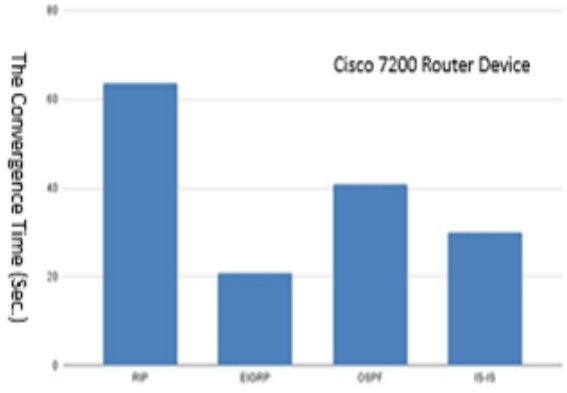


Figure 19: Convergence Time Variation Graph for RIP, EIGRP, OSPF and IS-IS.

As explained in the beginning of the sub-section "7.3", conversion time is calculated after the failure of one or more network router device in the network topology and the least measured value is the best recoverable protocol. The obtained result from the experiment indicate that the EIGP is more suitable for large-scale enterprise network topology.

8. Discussion

Many research papers have been presented the evaluation of the interior gateway routing protocols in different network topologies, each topology has different characteristics like the number of routing paths that used in their experiments, number of network devices, topology type as small; mid-sized; or large-scale network, number of subnets used in their experiments, internet protocol versions that used, default cost variation, route table size, dropped traffic and so on:[16][17][18] and [19].

Also there are different techniques have been used to measure the metric of these interior gateway routing protocols regardless of the network topology especially when measure the jitter to evaluate the performance of the network: [12] [19] and [20]. Here, in this work, the evaluation of these interior gateway protocols has been conducted in a large-scale network topology consist of 30-routers with tri-connected interfaces. A new method to calculate the Jitter has been presented in detail with its pseudo code and then compared with the simple technique method of measuring the Jitter in the computer networks. The result has shown there is a big different in term of the accuracy of both calculation methods, but both the jitter measurement methods have the same curve pattern.

9. Conclusion

This study compared the performance of RIP, EIGRP, OSPF, and IS-IS inner gateway protocols in terms of convergence time, end-to-end delay, and jitter for fifteen subnets in proposed network topologies. The results show that EIGRP outperformed other routing protocols in terms of latency and convergence time, although OSPF outperformed them in terms of jitter. In addition, IS-IS outperformed OSPF in terms of convergence time. In contrast, the results showed that RIP protocols had the poorest delay and convergence time performance. Finally in this work, it has introduced a new method for measuring jitter. The formula was developed to calculate network jitter with greater accuracy than the traditional, convention, method using the grand average calculations. Additionally, the suggested method may be replicated to model various network topology sizes, and the new approach can be contrasted with other methods to improve the speed and precision of jitter measurement in the field of routing networks. This would develop and improve new monitoring tools to measure jitter networks more accurately and quickly.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Usman, E.E. Oberafo, M.A. Abubakar, T. Aminu, A. Modibbo, S. Thomas, "Review of interior gateway routing protocols," in *2019*

15th International Conference on Electronics, Computer and Computation, ICECCO 2019, 2019, doi:10.1109/ICECCO48375.2019.9043204.

[2] (PDF) Performance Comparison of Routing Protocols | yousef aburawi - Academia.edu, Oct. 2024.

[3] A.G. Biradar, "A Comparative Study on Routing Protocols: RIP, OSPF and EIGRP and Their Analysis Using GNS-3," *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2020 - Proceeding*, 2020, doi:10.1109/ICRAIE51050.2020.9358327.

[4] C. Filsfils, A. Bashandy, H. Gredler, B. Decraene, "IS-IS Extensions for Segment Routing," 2019, doi:10.17487/RFC8667.

[5] I. Joseph Okonkwo, I. Douglas Emmanuel, "Comparative Study of EIGRP and OSPF Protocols based on Network Convergence," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, 2020.

[6] D.R. Prehanto, A.D. Indriyanti, G.S. Permadi, "Performance analysis routing protocol between RIPv2 and EIGRP with termination test on full mesh topology," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, 354–361, 2021, doi:10.11591/IJEECS.V23.I1.PP354-361.

[7] M. Athira, L. Abrahami, R.G. Sangeetha, "Study on network performance of interior gateway protocols — RIP, EIGRP and OSPF," *2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)*, 344–348, 2017, doi:10.1109/ICNETS2.2017.8067958.

[8] A. Manzoor, M. Hussain, S. Mehrban, "Performance Analysis and Route Optimization: Redistribution between EIGRP, OSPF & BGP Routing Protocols," *Computer Standards & Interfaces*, vol. 68, 103391, 2020, doi:10.1016/J.CSI.2019.103391.

[9] (PDF) Performance Analysis of Interior Gateway Protocols, https://www.researchgate.net/publication/303812574_Performance_Analysis_of_Interior_Gateway_Protocols, Oct. 2024.

[10] H. Karna, V. Baggan, A.K. Sahoo, P.K. Sarangi, "Performance Analysis of Interior Gateway Protocols (IGPs) using GNS-3," *Proceedings of the 2019 8th International Conference on System Modeling and Advancement in Research Trends, SMART 2019*, 204–209, 2020, doi:10.1109/SMART46866.2019.9117308.

[11] What Is Jitter, and How Does It Affect Your Internet Connection?, https://www.howtogeek.com/824032/what-is-jitter/, Oct. 2024.

[12] How to Measure Jitter - Obkio, https://obkio.com/blog/how-to-measure-jitter/, Oct. 2024.

[13] IMPROVING THE PERFORMANCE OF AD HOC NETWORK USING ENHANCED INTERIOR GATEWAY ROUTING TECHNIQUE - IJSER Journal Publication, https://www.ijser.org/onlineResearchPaperViewer.aspx?IMPROVING_THE_PERFORMANCE_OF_AD_HOC_NETWORK_USING_ENHANCED_INTERIOR_GATEWAY_ROUTING_TECHNIQUE.pdf, Oct. 2024.

[14] N. Kumari, D.T. Raj, "OSPF Metric Convergence and Variation Analysis During Redistribution with Routing Information Protocol," *International Research Journal on Advanced Engineering and Management (IRJAEM)*, vol. 2, no. 06, 1985–1991, 2024, doi:10.47392/IRJAEM.2024.0293.

[15] E.C. Obioma, E.C. Amadi, "Analysis of Optimal Routing Protocol for an Inter-Campus Private Cloud Network System," *International Journal of Modern Developments in Engineering and Science*, vol. 3, no. 2, 8–12, 2024, doi:10.5281/ZENODO.12643596.

[16] K. Shahid, S.N. Ahmad, S.T.H. Rizvi, "Optimizing Network Performance: A Comparative Analysis of EIGRP, OSPF, and BGP in IPv6-Based Load-Sharing and Link-Failover Systems," *Future Internet*, vol. 16, no. 9, 2024, doi:10.3390/FI16090339.

[17] N. Kumari, A. Sharma, "OSPF Default Cost Variation for Type-1 and Type-2 External Links during Redistribution," *International Journal of Computer Applications*, vol. 184, no. 28, 8–14, 2022, doi:10.5120/IJCA2022922345.

[18] M.E. Mustafa, ; Ahmed, A. Abaker, "International Journal of Computer Science and Mobile Computing Evaluate OSPF and EIGRP Based on Route Table Size and Dropped Traffic using Different Topologies (Star, Loop)," *International Journal of Computer Science and Mobile Computing*, vol. 13, 6–17, 2024, doi:10.47760/ijcsmc.2024.v13i04.002.

[19] D. Turlykozhayeva, S. Temesheva, N. Ussipov, A. Bolysbay, A. Akhmetali, S. Akhtanov, X. Tang, "Experimental Performance Comparison of Proactive Routing Protocols in Wireless Mesh Network Using Raspberry Pi 4," *Telecom 2024, Vol. 5, Pages 1008-1020*, vol. 5, no. 4, 1008–1020, 2024, doi:10.3390/TELECOM5040051.

[20] Performance Analysis of Routing Protocols RIP, EIGRP, OSPF and IGRP using Networks connector, 2020, doi:10.4108/eai.28-6-2020.2298167.

Saleh Hussein Al-Awami has done his bachelor's degree from University of Benghazi formerly named Garyounis University in 1996. He has done his master's degree in Computer Sciences - Artificial Intelligent from The Libyan Academy for Postgraduate Studies in Benghazi-Libya in 2010. His researches interested in Artificial Intelligent, Internet of Things (IoT), Computer Networks, Cloud Computing, Software Defined Networks, Computer Network Infrastructures Security, Fintech infrastructure and application services and System Design and Methodologies. He has many scientific papers and journals publications related to his researches. Saleh has published scientific papers in many conferences as in United States (Las-Vegas), India (Mumbai), Saudi-Arabia(Riyadh), Jordan, and in Libya. He has hold some attendance professional certificates from several international training centers in United Kingdom (UK-London) and Egypt (Cairo) holding CompTIA N+, Cisco Certified Network Associate (CCNA, 2002), Microsoft SQL databases, Microsoft Exchange Server (E-mail Systems administrator since 2007 edition), and Microsoft Server operating systems since 2003 edition. Also, he has hold some professional certificates as Microsoft network administrator, Novel network services administrator, and recently Oracle Cloud Infrastructure 2024 Generative AI certificated Professional.

Emad Awadh Ben Srity completed his Bachelor's degree in Computer Networks and Data Communications at the University of Benghazi in 2023. He is passionate about computer network research, network design, and analysis. Emad has published a scientific paper presented at a Libyan conference and received training in computer systems at Arabian Gulf Oil Company and ISP Smart Link Company. He worked successfully at an ISP company for about a year and also got an opportunity with one of the companies working on the maintenance of the Faculty of Science at the University of Benghazi. He worked as a supervisor of the Physics and Geology departments in terms of electricity and networks. He remained a site manager at the Faculty of Law in terms of electricity and networks, and worked on designing network maps, including the distribution of fire alarm points, motion sensors, access points, data, and telephone lines.

Ali Tahir Abu Raas has done his bachelor's degree from Computer Networks and Data Communications - University of Benghazi in 2023. He has interested in computer networks researches, network design and analysis. He received training in computer systems at Arabian Gulf Oil Company and ISP Company. Ali published a scientific paper in 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA).

JENRS

Received: 15 September, 2024, Revised: 10 October, 2024, Accepted: 03 November, 2024, Online: 08 November, 2024

DOI: https://doi.org/10.55708/js0311006

Smart Vehicle Safety System Using Arduino: An Experimental Study in Bahrain's Driving Conditions

Youmna Rabie Farag *, Bintu Jasson
University of Bahrain, Chemical Engineering, Hamad Town, Bahrain
*Corresponding author: Youmna Farag, Contact No: +973 37074411, Email: yrabie330@gmail.com

ABSTRACT: The high rate of vehicle accidents is increasingly linked to drivers failing to maintain adequate safety distances between their vehicles and those in front. This issue is exacerbated by varying weather conditions such as rain, sandstorms, and fog. To mitigate this problem, we propose an Arduino-based intelligent system designed to assist drivers in maintaining safe distances and adjusting their speed accordingly. This system integrates an ultrasonic distance sensor, a rain sensor, a liquid crystal display with I2C, four gear motors, a motor driver, and an Arduino UNO. The ultrasonic sensor measures the distance to the vehicle ahead, and based on this data, the system calculates the recommended speed using an equation derived from safe distance calculations for different speeds. In adverse weather conditions, detected by the rain sensor, the system adjusts the equation to account for reduced visibility and road traction. This ensures that the recommended speed remains safe under varying environmental conditions. Experimental results from multiple locations across Bahrain demonstrate the system's effectiveness in real-world scenarios. These tests reveal how the system can adapt to different environmental conditions and provide accurate speed recommendations to enhance road safety. By addressing the critical issue of maintaining safe distances and adjusting to changing conditions, this system significantly contributes to reducing accident rates and improving overall vehicle safety.

KEYWORDS: Accidents, Arduino Uno, HC-SR04, HW-038, Intelligent systems, L298N, Mathematical models, Vehicles

1. Introduction

By 2024, the global number of cars on the road is expected to reach approximately 1.5 billion, underscoring the growing need for robust road safety measures [1]. With the rising volume of vehicles in urban and rural areas, the risks associated with driving—such as collisions and traffic-related injuries—continue to escalate. Every year, around 1.19 million people lose their lives in traffic accidents, while another 20 to 50 million suffer non-fatal injuries, often leading to long-term disabilities [3]. Common factors contributing to these accidents include distracted driving, speeding, and adverse weather conditions like rain, fog, and strong winds [4].

One of the most effective strategies to mitigate these risks is maintaining a safe distance between vehicles, which can significantly reduce the likelihood of accidents. However, many drivers fail to consistently adhere to this safety measure, particularly in challenging weather conditions where visibility is reduced, or vehicle control becomes more difficult. This problem is especially pronounced in regions like Bahrain, where air quality fluctuations, rain, and sandstorms can affect not only driver visibility but also the overall performance of vehicle sensors and systems.

Vehicle safety is a crucial area of study, as road traffic accidents continue to be a leading cause of injury and death globally. By integrating affordable and effective components, this Arduino-based system can improve vehicle safety in real-world applications. While numerous systems exist for maintaining safe distances between vehicles, the flexibility and simplicity of the Arduino platform make this solution particularly useful for adapting to Bahrain's unique driving conditions, including sandstorms and rain.

Recent studies have explored technological solutions for improving vehicle safety, with many focusing on

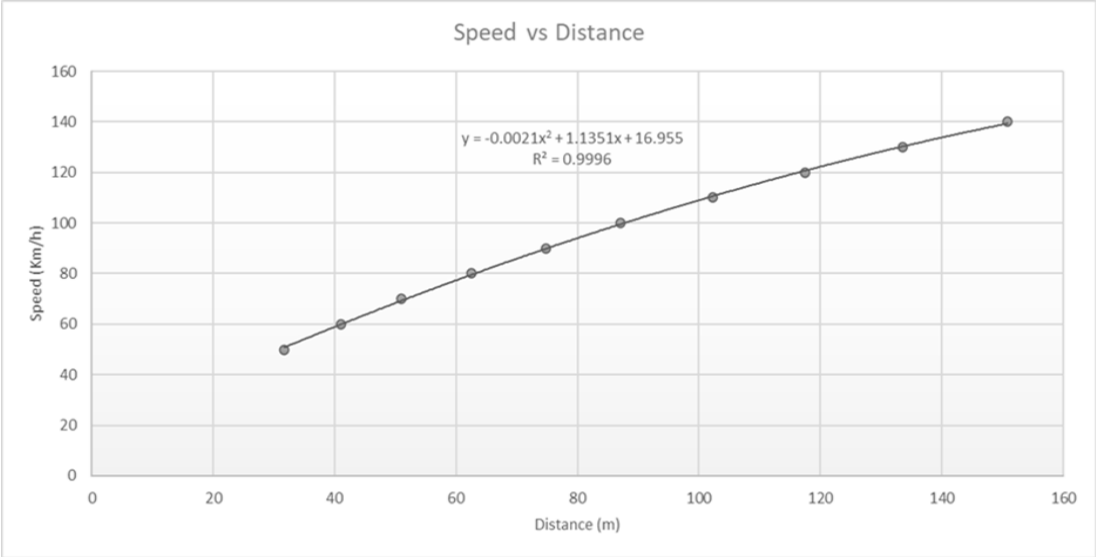


Figure 1: Safe speed versus measured distance

sensor-based safety systems. The use of Arduino in such applications has grown due to its affordability and flexibility in building customized systems. However, existing systems often fail to address region-specific challenges like Bahrain's adverse weather conditions. This research distinguishes itself by optimizing performance to address these challenges, improving on existing technologies by focusing on specific environmental issues, including rain and sandstorms.

2. Methods

2.1. Stopping Distance Calculation

The stopping distance of a vehicle consists of three parts:

1. Driver Reaction Distance (Drk) – the distance traveled while the driver reacts before applying the brakes.
2. Braking Force Build-up Distance (Dnh) – the distance covered as braking force increases from zero to its final value.
3. Braking Distance (Dh) – the distance traveled after the full braking force is applied.

The total stopping distance (Dz) is calculated using the following equation:

$$Dz = Drk + Dnh + Dh = Sp \cdot Trk + (Sp \cdot Tnh / 2) + (Sp^2 / 2 \cdot a) \quad (1)$$

Where:

- Dz is the total stopping distance,
- Drk is the reaction distance,
- Dnh is the distance during braking force buildup,
- Dh is the braking distance,
- Sp is the vehicle's speed,
- Trk is the driver's reaction time,
- Tnh is the braking force rise time,
- a is the vehicle's deceleration.

We assumed maximum values for Trk and Tnh of 1.2 seconds and 0.4 seconds, respectively. The deceleration value was taken as 8 m/s² under ideal weather conditions with an efficient braking system [5]. The speed ranged from 50 km/h to 140 km/h. Using the calculated safe distances as independent variables and corresponding speeds as dependent variables, we plotted the data and derived a polynomial equation:

$$v = -0.0021 \cdot d^2 + 1.1351 \cdot d + 16.955 \quad (2)$$

Where:

- v is the speed the driver should maintain in km/h,
- d is the distance between the driver's vehicle and the one ahead, in meters.

By converting (2) to fit our model:

$$v(cm/s) = -0.000021 \cdot d^2 + 0.011351 \cdot d + 472.82 \quad (3)$$

where d is in cm. Weather Condition Adjustments

In adverse weather conditions, such as rain, the safe stopping distance must increase due to reduced tire friction. For these conditions, a simpler equation is used:

$$v = d \quad (4)$$

Where the speed equals the distance between vehicles (e.g., a 60-meter gap for 60 km/h) [6]. This ensures that drivers are maintaining a sufficient safety buffer, particularly when road surfaces become slippery due to rain or dust accumulation during sandstorms. Equation (4) will be the same for the model.

2.2. Transition from Quadratic to Linear Modeling

In our original approach to developing a safety system for vehicle distance regulation, we employed a quadratic equation (3), to determine the recommended

speeds based on varying safe distances. This model was chosen with the assumption that it would accurately reflect the dynamic nature of vehicle behavior under different driving conditions. However, upon implementation and testing, the results produced by this equation were not only unrealistic but impractical for the gear motors used in our Arduino-based system. The speeds calculated often exceeded 5 meters per second, far beyond what is typically achievable or safe in the context of small-scale robotic projects or educational models.

The primary issue stemmed from the high speeds suggested by the quadratic model, which did not align with the mechanical limitations and operational capacities of the gear motors commonly utilized in Arduino projects. Such speeds would require motors and control systems far more robust and complex than those available for our application. Furthermore, the complexity of the quadratic equation added unnecessary complications in tuning and calibration during experiments.

Given these challenges, a decision was made to simplify the model to a linear equation. This transition to a linear model, represented by

$$v\ (cm/s) = 0.4*d\ (cm) + 12 \quad (5)$$

was guided by several critical factors:

1. Simplicity and Understandability: A linear equation is inherently simpler to implement and interpret, which is beneficial for both the development phase and for educational purposes where the system might be demonstrated.
2. Practicality in Application: The chosen linear model aligns well with the typical performance characteristics of small-scale gear motors. It offers a more manageable range of speeds from 20 cm/s to 50 cm/s, which are realistic targets for our hardware setup.
3. Ease of Calibration: Linear models are easier to calibrate and adjust in response to real-world testing feedback. They allow for straightforward scaling of input distances without the risk of producing extreme output values, which was a significant problem with the quadratic approach.

This recalibration to a linear model not only makes the system more practical but also enhances its reliability. It ensures that the vehicle safety system remains within the operational limits of the Arduino platform and associated hardware, thereby increasing the applicability and effectiveness of the model in real-world educational and hobbyist scenarios. This approach guarantees that the system can consistently provide accurate and safe guidance on vehicle distances, crucial for preventing collisions and enhancing road safety in controlled environments.

2.3. Hardware and System Design

This system is composed of several key components, each playing a vital role in ensuring the accurate detection of distance and the adjustment of vehicle speed. The integration of these components allows the system to function seamlessly, providing real-time information to the driver about safe driving speeds under both normal and adverse weather conditions.

2.3.1. Arduino Uno

At the core of the system is the Arduino Uno, which serves as the central processing unit. The Arduino Uno, based on the ATmega328P microcontroller, coordinates the actions of all the other components. It collects data from the sensors, processes it, and controls outputs such as the motors and display. The Uno has 14 digital input/output pins, 6 analog inputs, and supports easy integration with various sensors and actuators, making it ideal for this project. Its programmability allows for complex tasks like calculating safe driving distances based on sensor inputs and updating the driver via a display.

2.3.2. Ultrasonic Sensor

One of the essential sensors used is the HC-SR04 Ultrasonic Sensor, which measures the distance between the driver's vehicle and the one ahead. This sensor works by emitting ultrasonic waves and calculating the time it takes for the waves to bounce back after hitting an object, in this case, the car in front. The sensor is highly reliable for short-range distance measurements, with an operational range of 2 cm to 400 cm and an accuracy of ±3 mm. This data is crucial, as it feeds into a pre-defined algorithm that calculates the safe speed for the driver to maintain. The distance measured by the sensor is the primary variable used in these calculations, helping the system provide the most accurate speed recommendations for different conditions.

2.3.3. Rain Sensor

In addition to monitoring distance, the system uses a HW-038 Rain Sensor to detect rain, which can severely impact driving conditions. When rain is detected, the system adjusts the recommended safe distance and speed, accounting for the fact that wet roads reduce tire traction

and increase the likelihood of skidding. The rain sensor operates by detecting water droplets on its surface and is designed to switch between analog and digital modes, allowing for flexibility in its implementation. This feature makes it highly effective for detecting different levels of rain, thus ensuring that the safe speed calculations are adjusted accordingly.

2.3.4. *Liquid Crystal Display*

To provide real-time feedback to the driver, a 16x2 Liquid Crystal Display (LCD) with an I2C interface is used. The LCD shows the safe distance and speed based on the sensor readings, updating every second to reflect any changes in road conditions or vehicle position. The I2C interface reduces the number of pins required to control the display, simplifying the wiring and making it easier to integrate with the Arduino. This display ensures that the driver has a constant visual reference for maintaining safe driving conditions, enhancing the overall functionality of the system.

2.3.5. *Motor Driver*

The L298N Motor Driver is another crucial component, responsible for controlling the system's 4 gear motors. These motors simulate the vehicle's speed in the project model, allowing for real-time adjustments based on the safe speed calculated by the Arduino. The L298N motor driver uses pulse-width modulation (PWM) to control the motors' speed and direction, providing precise control over the system's mechanical movements. The ability to handle high-current motors makes the L298N ideal for this application, where reliable control is necessary to mimic the vehicle's response to different driving conditions.

2.3.6. *Gear Motors*

The 4 Gear Motors themselves play a significant role in the project, as they simulate the movement of the vehicle in response to the system's calculations. These motors provide high torque at low speeds, which is essential for accurately representing how a real vehicle would behave in similar conditions. The Arduino, working in conjunction with the motor driver, adjusts the motors' speed based on the real-time data from the sensors, ensuring that the system responds dynamically to changes in distance or weather.

Finally, the entire system is powered by an appropriate Power Supply, which ensures a steady flow of electricity to all components, including the Arduino, sensors, motor drivers, and display. A 9V power source is used. This consistent power supply is crucial for maintaining the accuracy and reliability of the system's real-time operations, as any interruptions could lead to delays in processing or incorrect data being displayed to the driver.

The system also relies heavily on the programming code uploaded to the Arduino. This code integrates all components, handling inputs from the ultrasonic and rain sensors, performing calculations to determine the safe speed based on environmental conditions, and updating the display to show the driver this information. The code is optimized for real-time processing, ensuring that the system adapts quickly to changes in distance or weather conditions. Algorithms within the code adjust the safe speed according to the unique environmental challenges the driver may face, such as rain, which affects road friction and stopping distance.

We conducted new experimental trials in various urban and rural areas of Bahrain, testing the system in diverse weather conditions such as dry, and rainy environments. These trials were designed to assess the impact of local weather and air quality on the system's performance. Adjustments were made to account for varying air quality, measured via real-time AQI data, which was integrated into the system's decision-making algorithms. These additions offer a more accurate and region-specific assessment of vehicle safety distance and speed under diverse conditions.

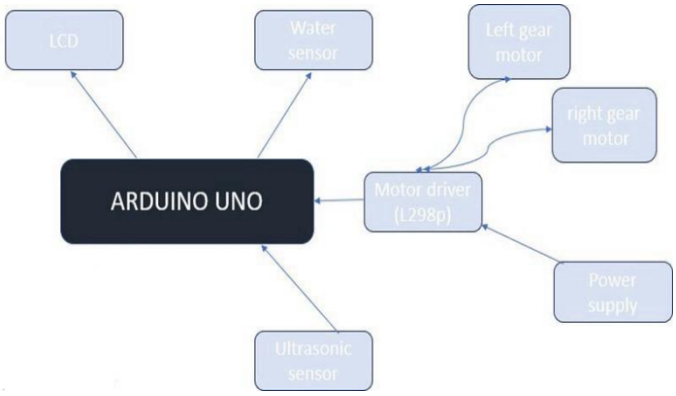


Figure 2: Block Diagram

3. Results and discussions

The enhanced Arduino-based intelligent vehicle safety system was tested on a model system designed to simulate real driving conditions. The trials were conducted in five different environments, representing a mix of typical weather conditions: Manama (dry and clear), Riffa (clear), Muharraq (sandstorm), Sitra (rainy), and Hamad Town (rainy). These locations were chosen to assess how well the system can adjust its functionality in response to environmental changes. Each test ran for a week, during

which the system's performance was analyzed in both normal and adverse weather conditions.

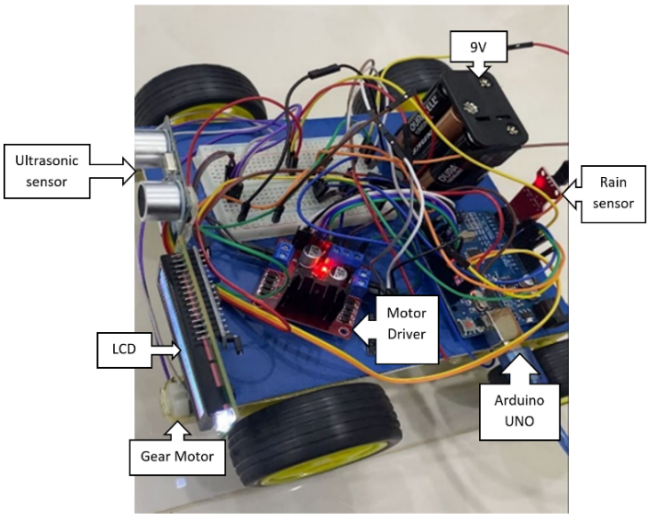


Figure 3: Model of the System

3.1. System Performance in Clear Weather

For the model system, the sensor accuracy under clear conditions was consistently high. In dry environments such as Manama, the ultrasonic sensor had an error margin of less than 1%, specifically around 0.5%. In this scenario, the recommended speed for the model vehicle was 10 cm/s, with a safe following distance of 6 cm. The system operated smoothly, demonstrating that under clear conditions, it provides reliable feedback for maintaining a safe distance between vehicles in a model environment.

3.2. System Performance in Rainy Conditions

During rainy conditions, which were simulated in Hamad Town and Sitra, the system exhibited slightly higher error margins of 1.5% to 2%. This increase in error occurred because the rain sensor detected water and adjusted the safe speed and distance accordingly, mimicking the reduced traction a real vehicle would experience on wet roads. For instance, in Hamad Town, where rainy conditions were simulated, the system recommended a speed of 14 cm/s with a safe distance of 10 cm. Similarly, in Sitra, the system recommended 13 cm/s with a distance of 9 cm. These tests confirm that the system can adjust to weather conditions appropriately, ensuring that the model vehicle maintains a safe distance in challenging weather.

3.3. System Performance in Sandstorm Conditions

The system faced its greatest challenge in Muharraq, where sandstorm conditions were simulated. During these tests, the system exhibited a higher error margin of 3% due to interference from particles simulating dust and sand, which affected the ultrasonic sensor's ability to detect accurate distances. Despite this, the system still functioned within acceptable limits, with the model vehicle moving at a recommended speed of 8 cm/s and maintaining a safe distance of 11 cm. This demonstrates that while the system adapts to sandstorm conditions, the sensors could benefit from further refinement for extreme weather accuracy.

Overall, the model system's performance was consistent across various conditions, with minimal errors in clear weather and acceptable accuracy in rainy or sandstorm conditions. These results demonstrate the system's effectiveness in adapting to different environmental factors and maintaining real-time, accurate feedback on safe vehicle distances.

Table 1: System Performance in Different Locations

Location	Condition	Error Margin (%)	Rec. speed (cm/s)	Safe Distance (cm)
Manama	Dry (Clear)	0.5	45	30
Hamad Town	Rainy	1.5	30	30
Riffa	Dry (Clear)	0.7	40	70
Muharraq	Sandstorm	3.0	50	95
Sitra	Rainy	2.0	25	25

Rec. = Recommended

3.4. Comparison with Recent Studies

Recent advancements in smart vehicle safety systems underscore the importance of integrating IoT technologies and real-time data processing to enhance vehicle safety and accident response. For instance, a study by (2024) explored an IoT-based accident detection and tracking system utilizing Arduino and GPS technology. This system effectively detects accidents in real time and provides accurate location tracking, facilitating prompt emergency responses [9]. Such developments highlight the increasing reliance on Arduino-based systems for safety monitoring in vehicles.

Additionally, in 2019 conducted a systematic review of smart vehicle technologies, discussing various vehicle communication systems that improve safety and traffic efficiency. Their findings suggest that integrating vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication can significantly enhance situational awareness for drivers and reduce accident rates [10].

In contrast to these studies, our project emphasizes not only accident detection but also the proactive notification of emergency services and the driver's family, offering a comprehensive approach to vehicle safety that incorporates real-time data communication and user-friendly interface features. This innovative aspect differentiates our system from existing solutions, demonstrating its potential impact on improving road safety in Bahrain and beyond.

4. Conclusion

The Arduino-based safety system for the model vehicle proved effective in helping maintain safe distances and adjusting the speed based on environmental conditions. In clear weather conditions, the system performed with minimal error, consistently providing accurate feedback to the user. In adverse weather such as rain and sandstorms, the system adjusted appropriately, though the sensors exhibited slightly higher error margins due to limitations in detecting distances during extreme conditions.

For the model system, the rain sensor effectively adjusted the speed and distance calculations to account for wet conditions, demonstrating its usefulness in environments where traction would be reduced. Sandstorm conditions were more challenging, causing some sensor interference, but the system was still able to provide reliable recommendations for safe speeds and distances.

To further improve the system's accuracy, particularly in extreme weather conditions like sandstorms, additional sensor technologies could be integrated to better handle dust and particulate interference. Overall, this model-based intelligent system shows great promise in simulating vehicle safety features, making it an effective tool for testing and demonstrating how similar systems could be implemented in real-world applications.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. S. K. Jalopnik, "The Number of Cars in the World," Jalopnik, 2023. Available: https://jalopnik.com/the-number-of-cars-in-the-world-brother-1850937293. [Accessed: Jun. 20, 2023].

[2] D. Bonnici, "How Many Cars Are There in the World," Which Car, 2023. Available: https://www.whichcar.com.au/news/how-many-cars-are-there-in-the-world. [Accessed: Jun. 20, 2023].

[3] World Health Organization, "Road Traffic Injuries," WHO, 2023. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed: Jun. 23, 2023].

[4] M. Steinger, "The 12 Most Common Causes of Car Accidents," Steinger, Greene & Feiner, 2023. Available: https://www.injurylawyers.com/blog/common-causes-car-accidents/. [Accessed: Jun. 23, 2023].

[5] K. Olejnik, "Analysis and Assessment of Defectiveness of Regulations for the Yellow Signal at the Intersection," Open Engineering, vol. 11, no. 1, pp. 189-194, 2021, doi: 10.1515/eng-2021-0020.

[6] B. Juga, "Important! Proper Safe Car Driving Distance," WULING, 2022. Available: https://wuling.id/en/blog/lifestyle/important-proper-safe-car-driving-distance. [Accessed: May 5, 2022].

[7] K. Olejnik, "Permissible Distance–Safety System of Vehicles in Use," Open Engineering, vol. 11, no. 1, pp. 303-309, 2021, doi: 10.1515/eng-2021-0021.

[8] Y. A. Badamasi, "The Working Principle of an Arduino," in 2014 11th International Conference on Electronics, Computer and Computation (ICECCO), pp. 1-4, 2014, doi: 10.1109/ICECCO.2014.7010214.

[9] M. Malik, "Smart Vehicle: A Systematic Review," International Journal of Research Publication and Reviews, vol. 4, no. 10, pp. 2728-2731, 2019.

[10] A. Abdullahi Omar, "IoT Based Smart Vehicle Accident Detection and Tracking Using Arduino Uno and GPS Modem," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 8, pp. 90-97, 2024, doi: 10.11591/ijece.v11i8.10683.

YOUMNA FARAG is an undergraduate student in Process Instrumentation and Control Engineering at the University of Bahrain, expected to graduate in 2025.

has one publication titled An Arduino Based Intelligent System to Reduce Accidents of Vehicles. Her academic focus is on control systems, process optimization, and green energy.

BINTU JASSON has done M.Tech in VLSI and Embedded Systems from Mahatma Gandhi University and B.Tech in Electronics and Communication from Calicut University, M.Tech 2013,B.tech 2010.

JENRS

Received: 13 September, 2024, Revised: 25 October, 2024, Accepted: 26 October, 2024, Online: 15 November, 2024

DOI: https://doi.org/10.55708/js0311007

A Bibliometric Analysis Review on Energy Optimisation while Designing Wireless Sensor Networks

Makwena Idah Masopoga, Mbuyu Sumbwanyambe, Zenghui Wang, Nthambeleni Reginald Netshikweta

Department of Electrical and Smart Systems Engineering, Unisa, Florida, 1709, South Africa

*Corresponding author: Makwena Idah Masopoga, 27 Galloway, 23 Stansted Street Highveld Centurion 0169, 0823049432 & masopmi@unisa.ac.za

ABSTRACT: Energy optimisation algorithms play an essential role in reducing energy usage. Hence, it is mandatory to identify current themes and predict future research study in energy optimisation algorithms (EOAs) in wireless sensor networks (WSNs). Several reviews have been conducted on energy optimisation algorithms (EOAs) in WSNs, although many focused on narrative reviews. This study focuses on the findings of a bibliometric analysis of the literature on the status of EOAs proposed to reduce energy consumption in WSNs. The main objective is to inform future research on the status of EOAs in WSNs. In this study, a systematic literature review and bibliometric analysis are performed to evaluate the depth and extent of the research done in EOAs in WSNs. The literature searched broadly and systematically concerning WSNs from the WOS database (WOS) from 2019 to 2023 articles. The analysis considers bibliometric data from bibliometric coupling, co-citation analysis, co-authorship analysis, and keyword co-occurrence units of analysis retrieved from VOSviewer software. Following the study of the bibliometric data, various themes on WSNs and their applications have emerged. The results of the bibliometric analysis revealed that many schemes are used for these WSNs. This article has projected a detailed bibliometric analysis of PSO and LEACH protocols in WSNs by searching the WOS Database. The study also revealed opportunities to improve research work in WSNs by using different routing protocols.

KEYWORDS: wireless sensor networks, optimisation algorithm, energy efficiency, clustering, routing, bibliometric analysis, literature review

1. Introduction

We have reviewed the previous articles from other authors involved in the same research. A few algorithms were also performed to correct the challenges we still face in WSNs. These algorithms have also contributed to improving the network performance. A few algorithms that have been performed that are very popular are the LEACH (Low Energy Adaptive Clustering Hierarchy) algorithm [1], the PSO (Particle Swarm) algorithm [2], EECA[3](Energy Efficiency Clustering Algorithm), Firefly optimisation algorithm[4], Ant Colony optimization algorithm [5] etc. In this paper we only focus on LEACH and PSO as they are the most popular as compared to the rest. LEACH [6]–[13] is an energy-efficient routing protocol that ensures that the selected cluster heads are regularly distributed over the network and that there is no possibility that all cluster heads concentrate on a single part of the network. It also balances the energy consumption of the sensor nodes and spreads it evenly in the entire network before the energy is depleted. One of the major drawbacks of LEACH is that if the cluster head dies for any reason, the cluster becomes useless because the data collected by the cluster nodes never reaches its destination. PSO, [14]–[21], [22]–[25] is an optimisation technique in which the social behavior of natural species is considered for the calculation. It is a population-based swarm intelligence technique that performs an optimisation process to optimize a fitness function. This approach uses a swarm to find each particle and records the fitness value of each particle. Then, the particles are linked with their appropriate speed. The best position of all the local intelligence particles optimizes the global best position to identify the cluster head position and minimize overall energy consumption. This algorithm has more efficiency and throughput than other mathematical and heuristic approaches; however, the distance between the sensor nodes is still unknown and reacts with environmental conditions.

Another big disadvantage of this algorithm is that it can converge prematurely and, especially for complex problems, be trapped in a local minimum. Poor battery performance directly impacts the WSNs performance and lifespan. This is the biggest challenge these days that affects the lifetime of WSNs [26]. The time used in WSNs should be long enough to keep them operating for an

extended period. A bibliometric analysis review was used to conduct this research. We implemented the exclusion/inclusion criteria to remain with the most relevant literature for analysis.

To address the research statement, the following objectives will be considered:

(1) To identify the current publication trends in WSNs.

(2) To identify the most influential authors, countries, and organizations in WSNs.

(3) To analyze the current state of collaboration involving WSNs.

(4) To recognize which area of study needs additional study or improvement in WSNs [27].

This article explores a bibliometric review of optimisation techniques, namely the PSO and LEACH Algorithms and their applications in WSNs. This article is written to understand the correlation between different research works carried out so far in using these algorithms. The application of bibliometric analysis [28] reviews is still a new topic for WSNs. This method is not only applied in WSNs but can be applicable in other areas of research such as in references [29] and [30], where the authors classically do a detailed bibliometric analysis, dividing the references into smaller groups related to the given topic, while in this paper, we do a different way of bibliometric analysis. Many of the publications have used the Scopus database. Hence, we selected to focus only on the WOS. The other reason is that the vast majority of cited materials are peer-reviewed articles based on other publications from different databases. These peer-reviewed articles have been analyzed and checked by experts more knowledgeable in the subject area. However, our future review proposed to extend our study to conference papers, book chapters, reviews, etc. So many papers have looked at the same algorithms this paper is reviewing. However, none of them has done a systematic literature [31] review coupled with the bibliometric analysis [32]. The significance of this study is summarised as follows. The overall energy consumption in WSNs will be reduced by implementing energy optimisation algorithm in the wireless network at each transmitter. Furthermore, it balances total energy consumption across the network by implementing an intelligent clustering approach to select cluster heads based on a well-defined energy gradient adaptively.

The following shows how the paper is organized: Section 2 focuses on the bibliometric analysis research design, followed by Section 3, which concentrates on the review report; Section 4 is the study's contributions; Section 5 is the study's discussions and recommendations; and Section 5 is the conclusions.

2. Bibliometric Analysis Research Design

The bibliometric analysis process followed is shown in Figure 1 below, showing briefly how the whole process was followed to conduct this research.

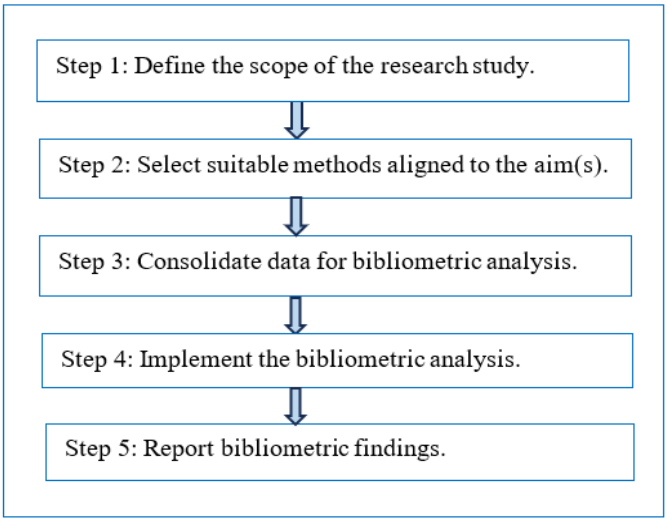


Figure 1: Flow diagram of bibliometric analysis adopted from [33]

Step 1: This first step is to define the aims and scope of the bibliometric study, which must occur before selecting bibliometric analysis techniques[34] and gathering the data. A bibliometric study's objectives ought to be connected to evaluating the effectiveness and science of a particular field of study. Regarding performance, bibliometric studies frequently aim to analyze the various research components in the subject, such as authors, institutions, nations, and journals. Scientifically speaking, bibliometric investigations are typically intended to highlight the bibliometric structure, which captures the networks among research components that contribute to the intellectual structure based on groups of relevant themes in the research field.

Step 2: The second step is to design the bibliometric study, wherein the techniques for bibliometric analysis are chosen to meet the aims and scope of the study in the previous step. One challenge that researchers often encounter at this stage is picking a technique based on the bibliometric data sought or choosing a method and then preparing the bibliometric data according to that selected technique. Bibliometric data is often retrieved in a raw format, as in this case. To conduct a comprehensive review of a research field utilizing an extensive bibliometric collection, a blend of co-citation analysis (past), bibliographic coupling (present), and co-word analysis (future) and future research directions of full texts can be chosen.

Step 3: The third step is to gather the data required for the selected bibliometric analysis techniques in the second step. To provide broad search results to justify bibliometric analysis and narrow to stay within the specific research subject. Researchers must define their search terms in this step. In this regard, researchers have

two options: the first is to consult the literature to identify a relevant combination of search terms, and the second is to brainstorm among themselves or with subject matter experts to curate suitable search terms. Following that, researchers will need to ascertain the bibliometric data that needs to be collected from the returned search results. Researchers should revisit the bibliometric analysis methods they selected for the investigation in this context. Researchers should concentrate on gathering the title, abstract, keywords, and full text of publications in the search results, for instance if they chose co-word analysis in the second step. However, in cases where the required data is unavailable, the first and second steps should be revisited. To reduce the need for the consolidation of different databases, the paper suggests choosing a single suitable database because reducing the number of pointless action items can reduce the possibility of human error. The researchers should then eliminate duplicates and incorrect entries. Researchers should clean the entry in such cases so that only one valid affiliation, "the affiliation of the author at the time of publication, "remains in the final dataset.

Step 4: Conducting the bibliometric analysis is the fourth step. Theoretically, conducting the bibliometric analysis and composing the bibliometric review are two distinct processes. But these action items frequently complement one another. This paper presents the bibliometric analysis, which produces bibliometric summaries, and the writing of the findings as a single step considering this feedback loop. The journal the researcher is writing for, and the field the study is conducting typically influence the writing style. For instance, one journal might prefer that researchers concentrate on the theoretical elements of the study. Still, another journal might prefer that researchers go directly to the summary of study findings. In this regard, this paper recommends researchers check targeted journals to see if they have a history of publishing review papers, and if yes, then to retrieve those papers, and if possible, those that use bibliometrics, so that a similar style of writing can be crafted. Lastly, the paper challenges academics to create perceptive debates that address pertinent trends and comparable arguments.

Step 5: The final step is to report the findings. The final phase involves interpreting the results of the bibliometric analysis. To do this, it is critical to comprehend the meaning associated with the topics of publications in each thematic cluster and the content within each one. To understand the content well, it is also important to examine its contextual meaning in relation to the entities or events that characterize that content. For example, the co-word analysis presents researchers with different clusters of words. Researchers may depend on terms that appear frequently in the cluster to comprehend its content; however, to interpret the context of each cluster, researchers also need to examine the relationships between the terms.

The following Table 1 shows how the research design was done. We propose the following research questions:

Table 1: Bibliometric analysis techniques to address the research questions.

RQN	Research Question	Technique	Usefulness
RQ1	What are the current publication trends in WSNs?	Bibliometric coupling	To evaluate the current evolution of topics in WSNs. In this study, we will evaluate it to identify current publications.
RQ2	Who are the most influential authors, countries, and organizations who researched WSNs?	Co-citation analysis	To provide information on the connections between publications and the most influential publications. Counting citations in significant research databases and internet sources keeps track of authors' words, the significance of articles, and the development of research ideas.
RQ3	What is the current state of collaboration involving WSNs?	Co-authorship	To analyze the social interactions and relationships among authors, their affiliations, and the number of publications they have collaborated on.
RQ4	Which areas need additional study or improvement in WSNs?	Co-word/ or co-occurrence analysis	To investigate the existing or future connections between research topics by examining the written content of a publication.

3. Search Criteria

Using the search query string in the WOS database, 442 research articles were found, and the publication language was English. The keywords used for searching this WOS database were WSN, LEACH, PSO[35], Energy optimisation, routing, bibliometric analysis, and literature review.

The following search criteria produced the results that were finally used in the entire analysis[36].

(1) Identification – in this category, we have identified records by searching through the currently available database, which, in this case, we only concentrated on the WOS.

(2) Screening – during this level, we check if there are any duplicates of papers/journals, so that if there are any, then we remove all the duplicates.

(3) Eligibility – when we check if all the articles match the research topic we are currently working on and

remove anything that is not in the same field. Furthermore, we only looked at the papers for the past five years and excluded anything before.

(4) Inclusion – this is where our refinement process was done, and we now know exactly how many articles we will use for the research.

Bibliometric information presented in this section is articulated using the following details by analyzing the .csv file obtained through WOS DB. The said information is as follows: Bibliometrics is used to understand the quantitative methods of analysis of scientific journals. The bibliometric analysis includes (1) Various parameters such as author information, publication year, journal in which articles are published, year publication, source of the articles, territory or geographic location of the authors, etc. (2) Understanding meaning insights such as the relationship with the authors and co-authors, frequency of publications, co-occurrences of keywords and articles, top researchers in the field of study, number of citations, etc. With all the quantitative information obtained from the bibliometric analysis, it is obvious that researchers will benefit from exploring future publication trends, the scope of journals that are focused on a specific domain of interest, etc. In this section, a detailed bibliometric analysis is conducted based on the" CSV" file obtained from the WOS database to understand the essence of PSO techniques in improving the LEACH[37] protocol of WSNs.

The inclusion and exclusion criteria were used and detailed below:
The criteria for inclusion are

- If the content of the article seems to discuss energy optimisation of WSN, we obtain its full reference
- Journal articles written in English
- Articles were chosen based on those that addressed energy optimisation in WSNs

The criteria for exclusion are

- Any study that is not about WSNs
- Studies before 2019, we believe someone would have adapted the recent methodology.
- Publications other than journal articles were excluded as they were not peer-reviewed by the top experts in the field
- Any articles addressing anything but not WSN were also excluded

The following PRISMA diagram shows our search criteria regarding the method and number of documents used for this literature review. The records found during the search were 442 articles. Only articles were selected in this case as we know that they are peer-reviewed by top researchers who are knowledgeable in the subject area. From this, we went through a screening process where all the duplicates were removed and ended up with 194 records.

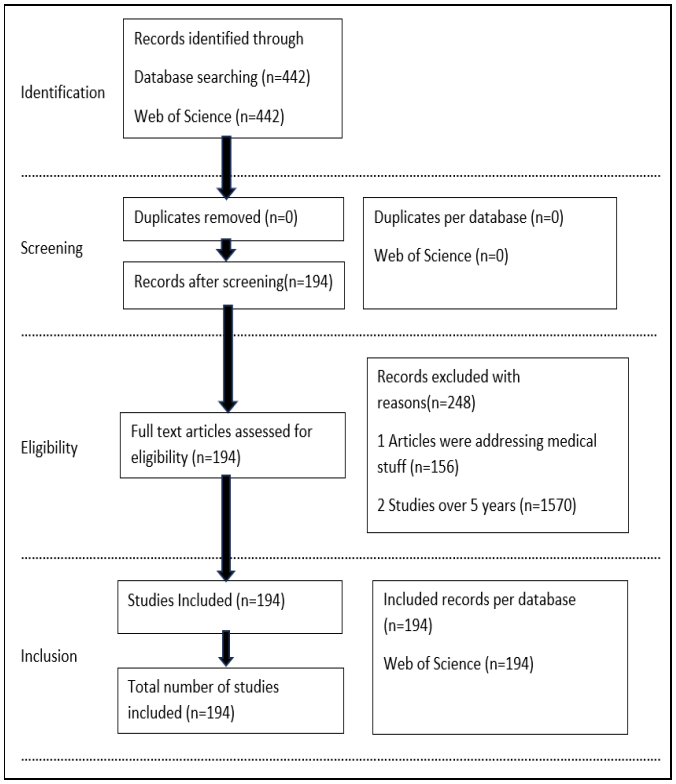


Figure 2: Prisma diagram based on our search criteria. Adapted from [38]

4. Review Report

A bibliometric analysis was performed, and the following results are based on the work done from 2018-2023 with about 442 articles downloaded from WOS. The topics used during the search string were WSNs, energy efficiency, optimisation algorithms, routing, and clustering. From these articles, a screening process was necessary to remove duplicates, which was done using Mendeley. Then, our records after screening remained at 442. The search was further refined under categories where only the Telecommunications category was selected. After this refinement, the articles published under Telecommunications went down to 194.

To produce all the figures from the Vosviewer database, the following step-by-step illustrations were followed: After the above refinement process,

- The first step was to save the text file from WOS and convert it to CSV.
- From there, we opened the VOSviewer software, went under file, and click on create.
- It will then lead to a screen where you select the data type which in this case we selected the an option to create map based on bibliographic data.
- It will to a screen where you choose the data source which we chose database file.
- From there the next screen will ask you to select the type of file you have saved from the database you are using.

- From here you choose the type of analysis and counting method per your preference

3.1 Volume of publications per year

Table 2 and Figure 3 show the number of publications done annually since 2019 from the WOS database. The results show that the overall number of authors who have done research in the subject matter that we are currently researching is 50, the number of times that their articles have been cited is more than 580 times, and we also see the overall reference count of about 1715 which is evident that researchers are still interested in the topic of energy optimization [39], [40] of WSNs.

Table 2: Summary of results from the WOS for the previous 5 years.

Publication Year	Data	Total
2019	Count of Authors	9
	Sum of Times Cited, All Databases	144
	Sum of Cited Reference Count	247
2020	Count of Authors	17
	Sum of Times Cited, All Databases	302
	Sum of Cited Reference Count	584
2021	Count of Authors	6
	Sum of Times Cited, All Databases	110
	Sum of Cited Reference Count	289
2022	Count of Authors	8
	Sum of Times Cited, All Databases	22
	Sum of Cited Reference Count	272
2023	Count of Authors	10
	Sum of Times Cited, All Databases	5
	Sum of Cited Reference Count	323
Total Count of Authors		50
Total Sum of Times Cited, All Databases		583
Total Sum of Cited Reference Count		1715

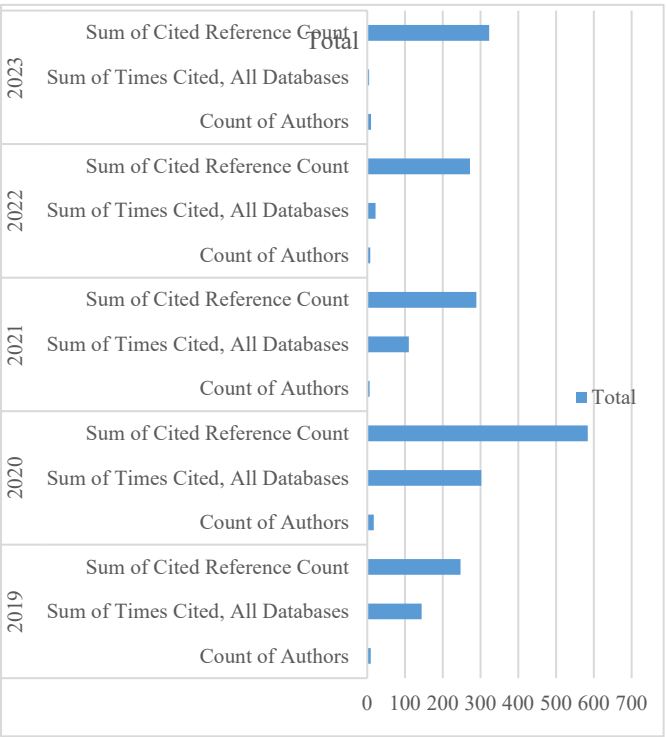


Figure 3: Results of author publications over the past 5 years.

Table 3 and Figure 4 from the WOS database show the number of authors and how many times they were cited in which issue.

Table 3: Author count and citations for the previous 5 years.

Issue	Count of Authors	Number of Times Cited
1	6	92
2	3	6
3	3	21
5	3	46
6	4	42
7	3	1
10	3	26
12	4	111
13	1	7
16	1	2
17	1	12
18	1	14
21	1	6
22	2	21
24	1	1
	13	164
Grand Total	50	572

3.2 Results from bibliometric coupling

The bibliometric coupling data are displayed in Table 4 below. It displays the most cited sources and the links' overall strength. The highest cited sources and the strongest overall link strength are displayed in Figure 5.

Table 4: The topmost sources with high citations per publication, Citations: TC, Total link strength: TLS

ID	Source	Documents	CT	TLS
1	ad hoc networks	8	196	926
2	annals of telecommunications	1	0	191
3	applied system innovation	1	0	152
4	computer communications	5	67	204
5	computer networks	5	55	464
6	eurasip journal on wireless communications and networking	6	128	475
7	ict express	1	0	21
8	ieee access	26	474	2164
9	ieee communications letters	2	78	180
10	ieee internet of things journal	6	100	457
11	ieee open journal of the communications society	1	0	0
12	ieee systems journal	2	45	359
13	ieee transactions on green communications and networking	2	62	129
14	ieee transactions on vehicular technology	1	47	137
15	iet wireless sensor systems	1	12	17
16	international journal of ad hoc and ubiquitous computing	1	0	20
17	international journal of communication systems	21	129	1587
18	international journal of distributed sensor networks	5	77	604
19	international journal of mobile computing and multimedia communications	1	2	84

20	international journal of network management	1	0	82
21	international journal of wireless information networks	3	13	272
22	internet of things	1	5	90
23	journal of ambient intelligence and humanized computing	9	231	518
24	journal of communications software and systems	1	0	98
25	journal of internet technology	1	0	93
26	ksii transactions on internet and information systems	3	22	244
27	mobile information systems	2	17	226
28	mobile networks & applications	2	50	141
29	peer-to-peer networking and applications	9	31	624
30	personal and ubiquitous computing	1	29	127
31	security and communication networks	1	14	5
32	telecommunication systems	3	88	420
33	transactions on emerging telecommunications technologies	3	12	78
34	wireless communications & mobile computing	2	224	111
35	wireless networks	15	320	1460
36	wireless personal communications	41	485	2962

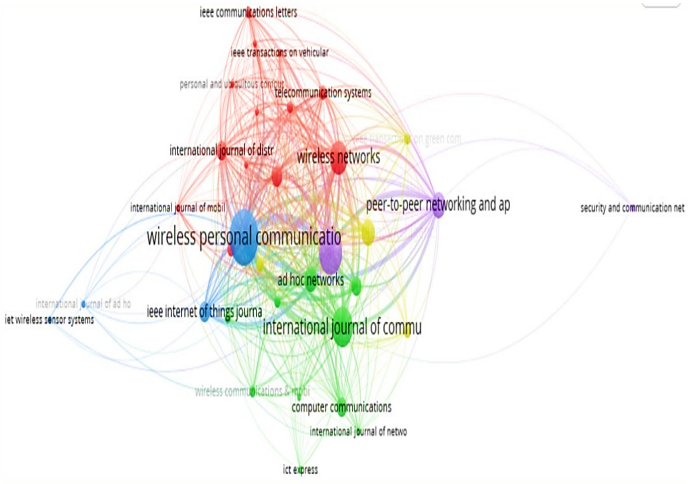


Figure 5: Sources showing from high to low citations and strong TLS.

3.3 *Publication impact of countries and journal articles*

3.3.1 *Publications per country*

The maximum number of authors per document selected was 25, with the minimum number of authors per document being 1 and the minimum number of citations of an author being 1. By looking at the results shown in Table 5 below, we can see that the topmost 3 countries are India leading with a total citation of 1722 and a link strength of 37 and followed by the Peoples's Republic of China with a total citation of 792 and a link strength of 28, the 3rd one being South Korea with a total citation of 348 and total link strength of 4.

Table 5: The topmost countries with high citations per publication, Citations: TC, Total link strength: TLS[41]

ID	Country	Documents	CT	TLS
1	australia	4	125	4
2	canada	2	43	6
hj3	egypt	4	76	9
4	england	6	157	8
5	finland	1	4	4
6	india	121	1722	37
7	iran	6	151	8
8	iraq	3	55	5
9	italy	2	23	4
10	japan	1	14	4
11	jordan	3	74	4
12	malaysia	5	72	12
13	pakistan	5	94	9
14	peoples r china	41	795	28
15	russia	1	14	4
16	saudi arabia	11	225	21
17	south korea	7	348	4
18	taiwan	4	51	8
19	tanzania	1	4	4
20	usa	10	76	13
21	yemen	1	4	4

Figure 6 below shows the topmost countries with high citations and the strongest total link strength. India is the first top country out of this category.

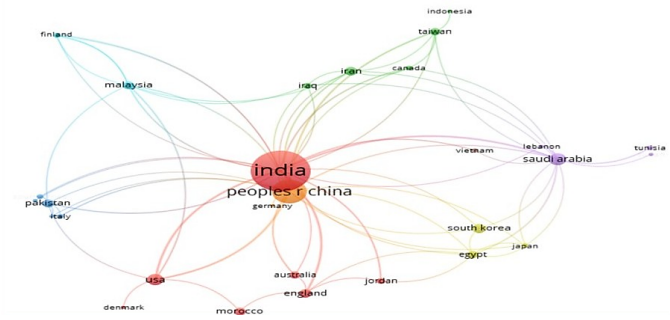


Figure 6: Network visualization of country collaboration in this research

3.3.2 *Publications per organizations*

The research went further, and we also checked the category of organizations. The results in Table 6 show the topmost organizations published in the study of WSNs with the most to the least total link strength. Figure 7 below illustrates the organizations with at least 3 documents per publication.

Table 6: Top organizations that published with the most to the least link strength, Citations: TC, Total link strength: TLS

ID	Organization	Documents	CT	TLS
1	al balqa appl univ	2	73	8
2	alagappa univ	2	79	9
3	anna univ	10	228	10
4	babol univ med sci	1	38	7
5	bonch bruevich st petersburg state univ telecommu	1	14	8

6	changchun univ sci & technol	1	14	8
7	chongqing univ posts & telecommun	3	89	9
8	guangzhou univ	1	38	7
9	harbin inst technol	1	14	8
10	howard univ	2	23	7
11	islamic azad univ	3	84	13
12	kalasalingam acad res & educ	2	79	9
13	king saud univ	1	38	7
14	manchester metropolitan univ	2	52	8
15	menoufia univ	1	14	8
16	natl yunlin univ sci & technol	1	38	7
17	prince sattam bin abdulaziz univ	3	66	18
18	princess nourah bint abdulrahman univ	4	27	12
19	rudn univ	1	14	8
20	sastra deemed univ	2	52	8
21	saudi elect univ	2	17	11
22	srm inst sci & technol	5	95	12
23	tokyo inst technol	1	14	8
24	univ regina	1	38	7
25	univ sistan & baluchestan	1	38	7
26	vellore inst technol	5	173	18

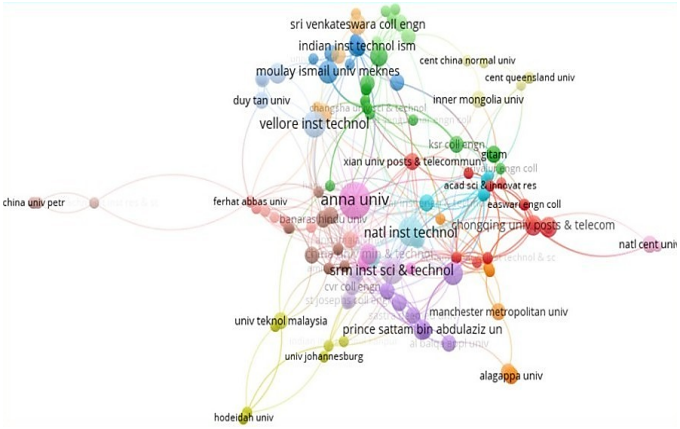


Figure 7: Graphical illustration of organizations with at least 3 documents per publication

3.3.3 Distribution of Publication Journals

The maximum number of authors per document selected was 25, with the minimum number of citations per document being 1 and the minimum number of citations of an author being 1. By looking at the results shown in Figure 8 below, it is evident that the topmost 3 journals are Wireless Personal Communications leading, with a total citation of 485 and a total link strength of 43, followed by IEEE Access, with a total citation of 474 and the total link strength of 42, the last one being Wireless Networks with a total citation of 320 and total link strength of 29.

3.4 Co-citation analysis

The maximum number of authors per document selected was 25, with the minimum number of authors per document being 1 and the minimum number of citations of an author being 1. Of the 598 authors found, only 501 meet the threshold, and 301 are connected through collaborations with the other authors.

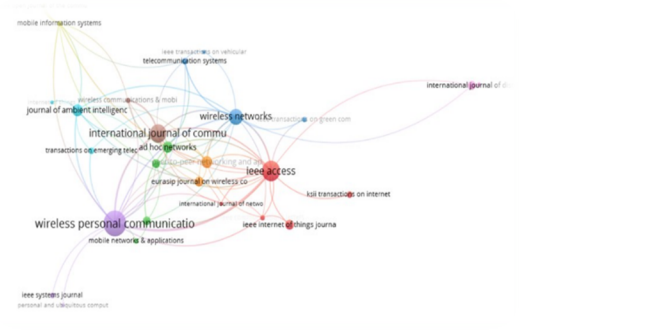


Figure 8: Demonstration of most to the least published journals

Figure 9 below shows they each worked with authors from two clusters in 2020, 2021, and 2022. These authors who worked so hard and collaborated more are K Shankar and Deepak Gupta. Shankar K has collaborated with all the authors in clusters 1 and 2, while Gupta Deepak has collaborated with authors in clusters 2 and 3. Refer to Table 7 and Figure 9 below to view the outcomes.

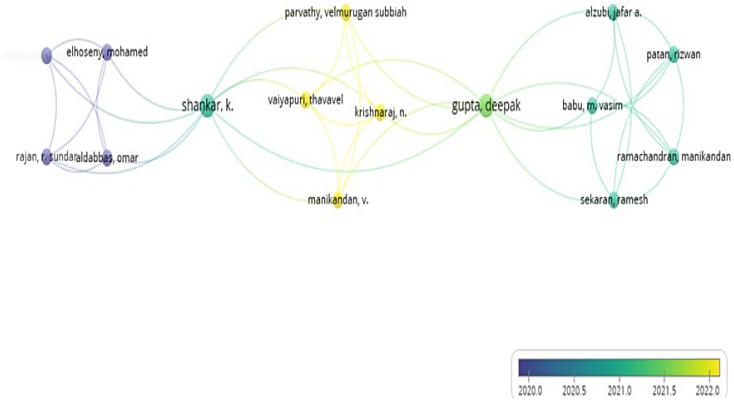


Figure 9: Network map of co-authorship collaboration based on the number of documents.

Table 7: Co-authorship collaboration based on the number of documents, Citations: TC, Total link strength: TLS.

ID	Author	Documents	Citations	TLS
1	abasikeles-turgut, ipek	1	218	0
2	abbas, sidra	1	218	5
3	abd el-latif, ahmed a.	1	218	6
4	abdelhaq, maha	1	218	5
5	abdennaceur, baghdad	1	218	2
6	abdul-qawy, antar shaddad hamed	1	143	5
7	abdulzahra, ali mohammed kadhim	1	143	2
8	abdulzahra, suha abdulhussein	1	122	2
9	abu bakar, kamalrulnizam	1	94	3
10	acken, john m.	1	85	2
11	adeliyi, timothy	1	85	2
12	adhikari, kabita	1	79	5

13	agrawal, sunil	1	72	2
14	ahn, seyoung	1	70	4
15	akilan, t.	1	68	3
16	al duhayyim, mesfer	1	68	7
17	al-otaibi, shaha	1	68	5
18	al-qurabat, ali kadhum m.	1	67	2
19	al-rasheed, amal	1	67	5
20	al-shdaifat, alaa	1	67	3

3.5 Co-occurrences of keywords

Table 8 below demonstrates unequivocally that WSNs remain the most explored keyword, appearing over 150 times more frequently than the others and having the strongest overall link strength of over 1500. Also, in both Figures 10 and 11, it is evident that "WSNs [42]" is the leading keyword being researched by authors.

3.5.1 Co-occurrences word count

Table 8: Co-occurrences of keywords in a publication, Total link strength: TLS.

ID	Keyword	Occurrences	TLS
1	wsn	26	291
2	WSNs	104	1071
3	wireless sensor network	34	324
4	scheme	26	307
5	routing protocols	25	320
6	routing protocol	36	419
7	routing algorithm	14	165
8	routing	49	500
9	protocols	10	135
10	protocol	47	460
11	particle swarm optimisation	9	102
12	optimisation	40	466
13	network lifetime	30	321
14	mobile sink	9	104
15	lifetime	31	362
16	leach	11	111
17	iot	11	150
18	internet of things	17	219
19	internet	23	288
20	fuzzy logic	12	114
21	energy-efficient	28	253
22	energy efficiency	81	814
23	energy consumption	30	353
24	energy	12	105
25	efficient	12	115
26	data aggregation	12	152
27	clustering-algorithm	11	113
28	clustering algorithms	20	273
29	clustering	69	590
30	cluster head selection	12	115
31	cluster head	10	103
32	aware	10	119
33	algorithm	64	663

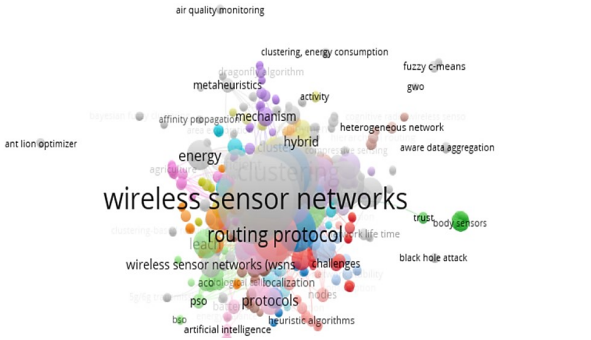


Figure 10: Network visualisation map of co-authorship based on all keywords.

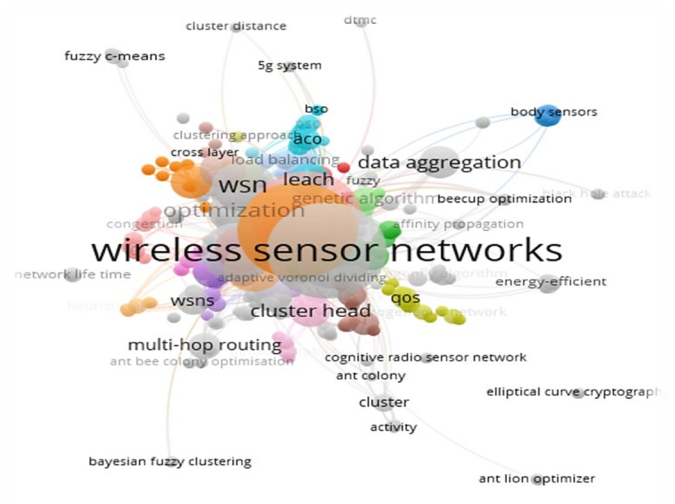


Figure 11: Network visualization map of co-authorship based on author keywords.

3.5.2 Keywords co-occurrence statistical analysis

The following statistical analysis was done on the keyword occurrences, and the results for each keyword were combined to get the total occurrences. The top 30 words were selected to complete this exercise. The results for the most keywords with 50 occurrences and above are shown in Table 8 below, while Table 9 and Figure 12 show the keywords with high occurrences:

The highest count being WSN; we found 6 occurrences of this word, the first word having 80, the second one having 32, the third one having 22, the fourth one having 8, followed by the fifth also having 8, and the last one having 6, which makes a total of 156. The lowest count is the Genetic Algorithm (GA)[43][44][45] with 4 occurrences, the first word having 5, followed by the second word with 4, while the third word has 3 and the fourth word having 2, which makes a total of 14.

Table 9: Author's keywords with high occurrences.

ID	Author's keywords	Acronyms	Occurrences
1	WSNs	WSN	156
2	clustering algorithm	CA	102
3	energy efficiency	EE	92
4	routing algorithm	RA	85
5	network lifetime	NL	40

6	cluster head	CH	34
7	energy consumption	EC	33
8	internet of things	IOT	25
9	particle swarm optimisation	PSO	25
10	leach	LEACH	19
11	fuzzy logic	FLG	18
12	genetic algorithm	GA	14
13	quality of service	QOS	9
14	delay	DLY	8
15	mobile sink	MS	7
16	multi-hop routing	MHR	6
17	heterogeneity	HTR	6
18	sensor node	SN	6
19	ant colony algorithm	ACA	6
20	grey wolf optimisation	GWO	6
21	monitoring	MNT	5
22	base stations	BS	5
23	metaheuristics	MTH	5
24	multi-objective optimisation	MOO	5
25	load balancing	LB	4
26	mobile nodes	MN	4
27	batteries	BTT	3
28	spread spectrum communication	SSC	3
29	artificial bee colony	ABC	3
30	optimal path	OP	2

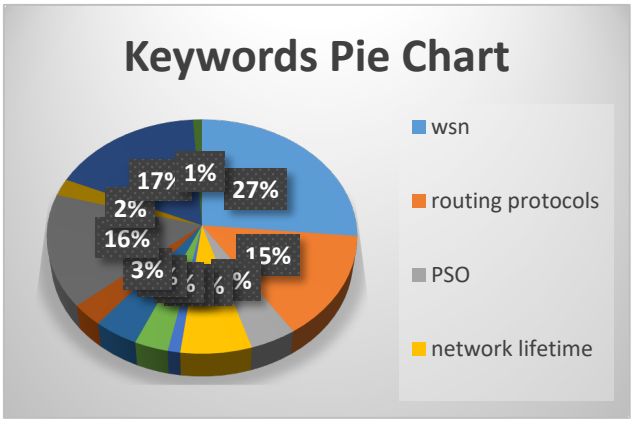


Figure 12: Pie graph of keywords with high occurrences.

3.5.2.1 *Summary statistics by groups - count, mean, standard*

Summary statistics are used to summarize a set of observations and to communicate the largest amount of information as simply as possible. Our results are shown in groups of 12 below, each group consisting of the keyword, count (N), mean (M), and standard (SD). The largest group from the results below is WSN with (N= 7, M= 22.6, SD=27.4), while we find the two lowest groups with interesting results. Groups 3 and 8 are the lowest group in this summary, having (N=3) while the mean and standard values differ. EC has (M=16, SD=19.8) while LEACH has (M=5.5, SD=4.95).

Table 10: Summary statistics by groups - count, mean, standard.

ID	Keywords	Count	Mean	SD
1	CA	3	30.7	34.3
2	CH	3	8.67	2.31
3	EC	2	16	19.8
4	EE	5	18.4	33.9
5	GA	3	3	1
6	ICR	3	2.67	0.577
7	IOT	3	8.33	7.77
8	LEACH	2	5.5	4.95
9	NL	5	8	12.3
10	PSO	3	8.33	6.81
11	RP	4	21.8	19.3
12	WSN	7	22.6	27.4

3.5.2.2. *Box and whisker plots*

Box plots are helpful as they provide a visual summary of the data, enabling researchers to quickly identify mean values, the dispersion/variability of the data set, and signs of skewness. Box plots divide the data into sections containing approximately 25% of the data in that set.

The median line of LEACH, EC, and NA box plots is the only one showing normal distribution of data sets, while the rest show skewness of data. The median line of the WSN, EE, NL, IOT [46] [47] [48], and PSO box plots shows that the distribution is positively skewed, while CH is the only one with a negatively skewed distribution. Then, when we compare the interquartile ranges (the box lengths) to examine how the data is dispersed between each sample, the longer the box, the more dispersed the data, and the smaller the box, the less dispersed the data. The longest box in our results is the CA box, which shows that the data is primarily dispersed in this sample, while the smallest box with the less dispersed is the GA[49].
Box and whisker plots also show the variability of data. In statistics, data variability is the extent to which a distribution is stretched or squeezed. We notice outliers when we look at our results in Figure 13 below. Outliers are data points located outside the whiskers of the box plots. In this case, we identified four outliers in our results.

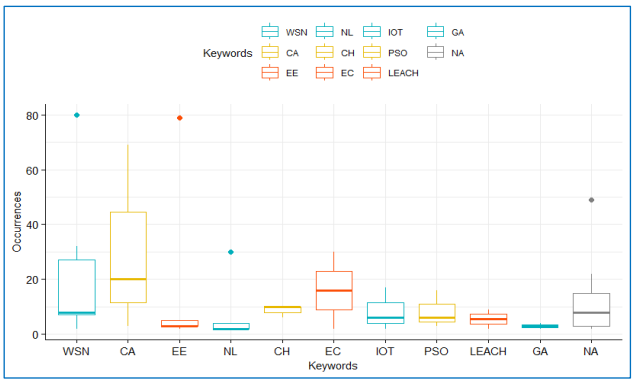


Figure 13: Keywords co-occurrence boxplot.

5. Contributions of the study

This review is worth carrying out as it seeks to make the following contributions:
We identified current publications by evaluating the evolution of energy optimisation topics in WSNs[50]. We achieved this by performing a systematic literature review and a bibliometric analysis review. Previously, researchers used the normal method and did not couple

the two. The results of the bibliometric review of research on energy optimisation algorithms of WSNs show that the topic is still highly interesting and can be explored more. The study also shows that the bibliometric methodology has become extremely popular recently because of the accessibility and utility of bibliometric databases and software, which make it easier to gather and evaluate vast amounts of scientific data for scientific research.

6. Discussions and Recommendations

All The co-citation analysis presents us with information on the connections between publications and the most influential publications. By counting citations in significant research databases and internet sources, we keep track of authors' keywords, the significance of articles, and the development of other research ideas.

We performed a co-authorship analysis to obtain relevant author's information and their affiliations. By analyzing this information, we identified the extent of the social interactions and relationships among authors, their affiliations, and the number of publications they have collaborated on.

Co-authorship analysis showed that India co-authored more WSNs with the People's Republic of China, South Korea, Iran, Australia, Saudi Arabia, and others in WSNs than with other countries. India, the People's Republic of China, and South Korea are the three most publishing countries that collaborate with other countries. India is the top publishing country in WSNs, and its collaboration with other countries is also high. The relationship between India and the People's Republic of China is stronger than that of other countries in terms of collaboration. We conclude that India is one of the most research centers in the world and can make a meaningful impact on WSNs.

The keyword occurrence analysis has revealed many optimisation techniques to address energy efficiency in WSNs. The densification of network devices in wireless networks has resulted in energy consumption challenges. Due to this growing challenge, research will continue to seek and propose new techniques to optimize energy in WSNs. Thus, research on energy optimisation is still a topic of interest to many researchers in WSNs [51]. This is also confirmed by the network visualization results obtained from the "co-occurrences of keywords" section

The review considered peer-reviewed journal articles. We propose considering other sources of information to widen the scope of the analysis. The potential sources of information include conference proceedings, books, and others. WOS is not the only database that can be analyzed using bibliometric software such as VOSviewer. Literature from other databases such as Scopus, IEEE, and Dimension can be analyzed, and comparisons made to evaluate their validity.

7. Conclusions

The results demonstrate the continued relevance and necessity of this study, as evidenced by the high total link strength and the continuous publication of research on WSNs in comparison to other keywords. It is evident that further investigation into the energy optimisation of WSNs is still required, and significant progress may be achieved in preserving the network's lifespan. In conclusion this paper also showed that bibliometric analysis is a scientific approach that can be helpful for researchers who want to look back on a wide range of rich and expansive topics in business studies, both seasoned and novice scholars.

The purpose of this research was to offer a review on the literature on optimisation algorithms to address energy efficiency on WSNs. The study considers literature obtained from WOS database from 2019 until 2023. We considered WOS database since it is one of the largest databases and most widely used for the analysis of scientific publications, including WSNs, a subfield of wireless networks. VOSviewer. By applying the bibliometric approach based on data from the WOS and using simple visualization tool, we were able to address all stated research questions.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge the University of South Africa and the Department of Electrical Engineering.

References

[1] I. Daanoune, A. Baghdad, and A. Ballouk, "Improved LEACH protocol for increasing the lifetime of WSNs," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, 3106–3113, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3106-3113.

[2] C. Wang, "A Distributed Particle-Swarm-Optimization-Based Fuzzy Clustering Protocol for Wireless Sensor Networks," *Sensors*, vol. 23, no. 15, 2023, doi: 10.3390/s23156699.

[3] Y. Gong, X. Guo, and G. Lai, "A Centralized Energy-Efficient Clustering Protocol for Wireless Sensor Networks," *IEEE Sens. J.*, vol. 23, no. 2, 2023, doi: 10.1109/JSEN.2022.3224180.

[4] H. Pakdel and R. Fotohi, "A firefly algorithm for power management in wireless sensor networks (WSNs)," *J. Supercomput.*, vol. 77, no. 9, 2021, doi: 10.1007/s11227-021-03639-1.

[5] M. Z. Abd Latif, K. G. Lim, M. K. Tan, H. S. Ee Chuo, T. Wang, and K. T. Kin Teo, "Energy-Efficient Ant Colony Based LEACH Routing Algorithm in Wireless Sensor Network," in *4th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2022*, 2022. doi: 10.1109/IICAIET55139.2022.9936851.

[6] S. Ramesh *et al.*, "Optimization of Leach Protocol in Wireless Sensor Network Using Machine Learning," *Comput. Intell.*

Neurosci., vol. 2022, 2022, doi: 10.1155/2022/5393251.

[7] S. Bharany *et al.*, "Energy-efficient clustering scheme for flying ad-hoc networks using an optimized leach protocol," *Energies*, vol. 14, no. 19, 2021, doi: 10.3390/en14196016.

[8] J. Mao, M. Gu, and Y. Huo, "Improved Routing Algorithm for Wireless Sensor Networks Based on LEACH," in *2022 International Communication Engineering and Cloud Computing Conference, CECCC 2022*, 2022. doi: 10.1109/CECCC56460.2022.10069668.

[9] M. Gamal, N. E. Mekky, H. H. Soliman, and N. A. Hikal, "Enhancing the Lifetime of Wireless Sensor Networks Using Fuzzy Logic LEACH Technique-Based Particle Swarm Optimization," *IEEE Access*, vol. 10, no. April, pp. 36935–36948, 2022, doi: 10.1109/ACCESS.2022.3163254.

[10] I. Daanoune, B. Abdennaceur, and A. Ballouk, "A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks," *Ad Hoc Networks*, vol. 114. 2021. doi: 10.1016/j.adhoc.2020.102409.

[11] L. Zhao, S. Qu, and Y. Yi, "A modified cluster-head selection algorithm in wireless sensor networks based on LEACH," *Eurasip J. Wirel. Commun. Netw.*, vol. 2018, no. 1, 2018, doi: 10.1186/s13638-018-1299-7.

[12] K. A. Raji, A. W. Asaju-Gbolagade, and K. A. Gbolagade, "EFFECTIVE ROUTING PROTOCOL FOR WIRELESS SENSOR NETWORKS USING CRT-LEACH ALGORITHM," *MALAYSIAN J. Comput.*, vol. 6, no. 2, 2021, doi: 10.24191/mjoc.v6i2.10396.

[13] L. Zhao, S. Qu, and Y. Yi, "A modified cluster-head selection algorithm in wireless sensor networks based on LEACH," *Eurasip J. Wirel. Commun. Netw.*, vol. 2018, no. 1, Dec. 2018, doi: 10.1186/s13638-018-1299-7.

[14] C. Vimalarani, R. Subramanian, and S. N. Sivanandam, "An Enhanced PSO-Based Clustering Energy Optimization Algorithm for Wireless Sensor Network," *Sci. World J.*, vol. 2016, 2016, doi: 10.1155/2016/8658760.

[15] O. Deepa and J. Suguna, "An optimized QoS-based clustering with multipath routing protocol for Wireless Sensor Networks," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 7, 763–774, Sep. 2020, doi: 10.1016/j.jksuci.2017.11.007.

[16] A. Yadav, S. Kumar, and S. Vijendra, "Network Life Time Analysis of WSNs Using Particle Swarm Optimization," *Procedia Comput. Sci.*, vol. 132, 805–815, 2018, doi: 10.1016/j.procs.2018.05.092.

[17] G. Srinivasalu and H. Umadevi, "Proposed energy efficient clustering and routing for wireless sensor network," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 4, 4127–4135, 2023, doi: 10.11591/ijece.v13i4.pp4127-4135.

[18] A. Singh, S. Rathkanthiwar, and S. Kakde, "Energy efficient routing of WSN using particle swarm optimization and V-LEACH protocol," in *International Conference on Communication and Signal Processing, ICCSP 2016*, Institute of Electrical and Electronics Engineers Inc., Nov. 2016, 2078–2082. doi: 10.1109/ICCSP.2016.7754544.

[19] A. Kaswan, V. Singh, and P. K. Jana, "A multi-objective and PSO based energy efficient path design for mobile sink in wireless sensor networks," *Pervasive Mob. Comput.*, vol. 46, 2018, doi: 10.1016/j.pmcj.2018.02.003.

[20] T. Shankar, S. Shanmugavel, and A. Karthikeyan, "Hybrid approach for energy optimization in wireless sensor networks using PSO," *Int. J. Commun. Antenna Propag.*, vol. 3, no. 4, 221–226, 2013.

[21] N. Karasekreter, M. A. Şahman, F. Başçiftçi, and U. Fidan, "PSO-based clustering for the optimization of energy consumption in wireless sensor network," *Emerg. Mater. Res.*, vol. 9, no. 3, 776–783, 2020, doi: 10.1680/jemmr.20.00107.

[22] R. L. P. Kumar and Vijayalakshmi, "Energy Efficient Routing in Wireless Sensor Network for Moving Nodes Using Genetic Algorithm Compared with PSO," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-25088-0_54.

[23] B. V. Divakar, D. Patil, and P. P. Subramanium, "Energy Optimization in Wireless Sensor Network using Clustering and PSO Algorithm," no. June, 5267–5270, 2020.

[24] P. Rawat and S. Chauhan, "Particle swarm optimization-based energy efficient clustering protocol in wireless sensor network," *Neural Comput. Appl.*, vol. 33, no. 21, 14147–14165, 2021, doi: 10.1007/s00521-021-06059-7.

[25] M. Z. Ghawy *et al.*, "An Effective Wireless Sensor Network Routing Protocol Based on Particle Swarm Optimization Algorithm," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/8455065.

[26] R. Nithya, R. Alroobaea, A. Binmahfoudh, and Z. I. Rizman, "Distributed Multi-hop Clustering Approach with Low Energy Consumption in WSN," *Comput. Syst. Sci. Eng.*, vol. 45, no. 1, 903–924, 2023, doi: 10.32604/csse.2023.030749.

[27] B. R. Al-Kaseem, Z. K. Taha, S. W. Abdulmajeed, and H. S. Al-Raweshidy, "Optimized Energy - Efficient Path Planning Strategy in WSN with Multiple Mobile Sinks," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3087086.

[28] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, no. May, 285–296, 2021, doi: 10.1016/j.jbusres.2021.04.070.

[29] D. Golubović, M. Erić, N. Vukmirović, and V. Orlić, "High-Resolution Sea Surface Target Detection Using Bi-Frequency High-Frequency Surface Wave Radar," *Remote Sens.*, vol. 16, no. 18, 2024, doi: 10.3390/rs16183476.

[30] D. Golubović, M. Erić, and N. Vukmirović, "High-Resolution Doppler and Azimuth Estimation and Target Detection in HFSWR: Experimental Study," *Sensors*, vol. 22, no. 9, 2022, doi: 10.3390/s22093558.

[31] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," *Journal of Planning Education and Research*, vol. 39, no. 1. SAGE Publications Inc., 93–112, Mar. 01, 2019. doi: 10.1177/0739456X17723971.

[32] V. Carneiro, A. Barata da Rocha, B. Rangel, and J. L. Alves, "Design Management and the SME Product Development Process: A Bibliometric Analysis and Review," *She Ji*, vol. 7, no. 2, 197–222, 2021, doi: 10.1016/j.sheji.2021.03.001.

[33] W. M. Lim, "Guidelines for interpreting the results of bibliometric analysis : A sensemaking approach," 17–26, 2024, doi: 10.1002/joe.22229.

[34] C. R. Bilder *et al.*, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, no. 3, 285–296, Mar. 2008, doi: 10.1198/jasa.2008.s251.

[35] A. Sun, S. Li, and Y. Zhang, "WSN clustering routing algorithm based on PSO optimized fuzzy C-means," *Tongxin Xuebao/Journal Commun.*, vol. 42, no. 3, 2021, doi: 10.11959/j.issn.1000-436x.2021053.

[36] J. R. Frank and L. A. Maggio, "Bibliometrics : Methods for studying academic publishing," 173–176, 2022, doi: 10.1007/s40037-021-00695-4.

[37] Saravana G, Jegadeeswari J S, NaveenKumar V, and PushpaValli M, "Energy Efficiency in Wireless Sensor Networks using Advanced LEACH Protocol," *Int. J. Eng. Technol. Manag. Sci.*, vol. 6, no. 6, 2022, doi: 10.46647/ijetms.2022.v06i06.011.

[38] N. C. Khin Khin Oo and S. Rakthin, "Integrative Review of Absorptive Capacity's Role in Fostering Organizational Resilience and Research Agenda," *Sustainability (Switzerland)*, vol. 14, no. 19.

MDPI, Oct. 01, 2022. doi: 10.3390/su141912570.

[39] A. Boudhir, M. Bouhorma, and M. Benahmed, "Energy Optimization Approaches In Wireless Sensor Networks: A Survey," *Int. J. Networks Syst.*, vol. 1, no. 1, 43–51, 2012.

[40] I. H. Osei and E. O. Osei, "Multipath Energy Optimization Broadcast Routing Protocol for Wireless Sensor Network Using Genetic Algorithm and Ant Colony Optimization," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4358067.

[41] B. A. Begum and S. V. Nandury, "Data aggregation protocols for WSN and IoT applications – A comprehensive survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2. King Saud bin Abdulaziz University, 651–681, Feb. 01, 2023. doi: 10.1016/j.jksuci.2023.01.008.

[42] C. Xu, Z. Xiong, G. Zhao, and S. Yu, "An energy-efficient region source routing protocol for lifetime maximization in WSN," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2942321.

[43] M. A. Aydin, B. Karabekir, and A. H. Zaim, "Energy Efficient Clustering-Based Mobile Routing Algorithm on WSNs," *IEEE Access*, vol. 9, 89593–89601, 2021, doi: 10.1109/ACCESS.2021.3090979.

[44] M. K. Singh, S. I. Amin, and A. Choudhary, "Genetic algorithm based sink mobility for energy efficient data routing in wireless sensor networks," *AEU - Int. J. Electron. Commun.*, vol. 131, 2021, doi: 10.1016/j.aeue.2021.153605.

[45] S. Wang, "Multipath Routing Based on Genetic Algorithm in Wireless Sensor Networks," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/4815711.

[46] B. R. Al-Kaseem, Z. K. Taha, S. W. Abdulmajeed, and H. S. Al-Raweshidy, "Optimized Energy - Efficient Path Planning Strategy in WSN with Multiple Mobile Sinks," *IEEE Access*, vol. 9, 82833–82847, 2021, doi: 10.1109/ACCESS.2021.3087086.

[47] M. Aljebreen *et al.*, "Binary Chimp Optimization Algorithm with ML Based Intrusion Detection for Secure IoT-Assisted Wireless Sensor Networks," *Sensors*, vol. 23, 1–17, Apr. 2023, doi: 10.3390/s23084073.

[48] H. Gunigari and S. Chitra, "Energy Efficient Networks Using Ant Colony Optimization with Game Theory Clustering," *Intell. Autom. Soft Comput.*, vol. 35, no. 3, 3557–3571, 2023, doi: 10.32604/iasc.2023.029155.

[49] L. Li and D. Li, "An Energy-Balanced Routing Protocol for a Wireless Sensor Network," *J. Sensors*, vol. 2018, 2018, doi: 10.1155/2018/8505616.

[50] Z. Al Aghbari, A. M. Khedr, W. Osamy, I. Arif, and D. P. Agrawal, "Routing in Wireless Sensor Networks Using Optimization Techniques: A Survey," *Wirel. Pers. Commun.*, vol. 111, no. 4, 2020, doi: 10.1007/s11277-019-06993-9.

[51] W. Xu, Y. Zhang, Q. Shi, and X. Wang, "Energy management and cross layer optimization for wireless sensor network powered by heterogeneous energy sources," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 5, 2814–2826, May 2015, doi: 10.1109/TWC.2015.2394799.

MAKWENA IDAH MASOPOGA completed her Bachelor's degree from Vaal University of Technology in 2013. She is a Junior Lecture with Unisa in the Department of Electrical and Smart Grid Engineering. She is currently busy with her Master's degree from the University of South Africa. She is currently researching on Energy Optimisation algorithms in WSNs. My research interests include more in robotics, IOT, Artificial Intelligence, and Smart Grid technologies.

PROF. ZENGHUI WANG graduated with Doctoral degree (2007) in Control Theory and Control Engineering, Nankai University, China. He is currently a Professor at the University of South Africa. His research interests include industry 4.0, control theory and control engineering, engineering optimisation, microgrid, image/video processing, artificial intelligence, and chaos. He was/is the principal investigator of six grants of South African national research foundation and one industrial project.

MBUYU SUMBWANYAMBE is a Professor and Head of Department for Electrical Engineering in the School of Engineering at the University of South Africa. He is also the cluster leader of telecommunications in the Department of Electrical Engineering. He is a Council and Senior member of SAIEE, a member of IEEE He has a MEng in Electrical and Electronic Engineering from the University of Johannesburg in South Africa and a PhD in Engineering Management from the same University. He has a wide range of industry experience, which he gained in South Africa and Zambia, working as consultant, project engineer and a researcher at the University of Johannesburg.

NTHAMBELENI REGINALD NETSHIKWETA received the Master of Technology in electrical engineering, from Tshwane University of Technology, in Pretoria, South Africa, in 2017. Currently, he is a PhD student in the College of Science, Engineering and Technology at the University of South Africa (UNISA), Pretoria. His main research interests are in the application of game theory to address resource management challenges in wireless networks. Specific current research topics include interference management in heterogeneous wireless networks. He is also a tenured Senior Lecturer with UNISA in the department of Electrical Engineering, where he is offering tuition in Telecommunication to undergraduate students.