SPRING / SUMMER 2025
VOLUME 17, NUMBER 1

Print ISSN: 2152-4157
Online ISSN: 2152-4165

WWW.IJERI.ORG

International Journal of Engineering Research & Innovation

Editor-in-Chief: Mark Rajai, Ph.D.
California State University Northridge

Published by the
International Association of Journals & Conferences

International Journal of Engineering Research and Innovation

The INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION (IJERI) is an independent and not-for-profit publication, which aims to provide the engineering community with a resource and forum for scholarly expression and reflection.

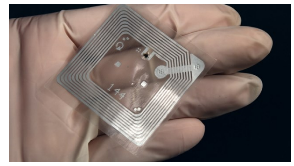
IJERI is published twice annually (fall and spring issues) and includes peer-reviewed research articles, editorials, and commentary that contribute to our understanding of the issues, problems, and research associated with engineering and related fields. The journal encourages the submission of manuscripts from private, public, and academic sectors. The views expressed are those of the authors and do not necessarily reflect the opinions of the IJERI editors.

EDITORIAL OFFICE:

Philip D. Weinsier, Ed.D.
Manuscript Editor
Office: 419.372.0628
Email: philipw@bgsu.edu
Department of Applied Sciences
Bowling Green State University-Firelands
One University Dr.
Huron, OH 44839

The International Journal of Engineering Research and Innovation Editors

TABLE OF CONTENTS



In This Issue (p.5) RAIN RFID

Philip Weinsier, IJERI Manuscript Editor

"We started NextNav with an audacious goal: build the future of geolocation. The country's principal positioning, navigation, & timing (PNT) system, GPS, is foundational to national security and the US economy (and powering critical infrastructure), but coverage is limited indoors and in urban canyons, and GPS is vulnerable to jamming and spoofing. In a recent FCC filing, NextNav proposed rearranging the Lower 900 MHz band to enable its use for terrestrial PNT services, as well as mobile broadband. A terrestrial 3D PNT complement and backup to GPS will mitigate the risks to the country, help public safety by providing location information and situational awareness indoors and in multi-story buildings, and unleash commercial opportunities" (NextNav.com).

Let's assume for the moment that you, like most people, have heard of GPS and understand that it helps us navigate from place to place. Let's further assume that you've heard of RFID tags, devices that can be embedded in products and shipping containers to allow for electronic verification of the contents of the containers without having to open and visually inspect them. With that as our starting point, what can we glean from NextNav's statement of its goal and our featured article in this issue in which the authors present their findings about RAIN RFID and the impact of noise?

RFID technology has been evolving since its inception and is now the most-used tracking system across industry. But what does RAIN have to do with RFID? The term RAIN is derived from RAdio frequency IdentificatioN; it is a cloud-based infrastructure where RFID data can be stored, managed, and shared via the internet. RAIN is the brand name for passive Ultra-High Frequency (UHF) RFID technology. It is an industry alliance to promote the widespread use of passive UHF RFID. The companies interested in this cause formed a RAIN Alliance whose objective is to promote the widespread use of passive UHF RFID technology. RAIN is not a product but an authority that certifies certain passive UHF RFIDs based on their qualities. RFIDs that are RAIN certified are called RAIN RFIDs. As billions of tags are deployed on products and assets that might move through multiple RFID reader infrastructures, the guideline makes it easier for each application to identify tags of interest, and to automatically disregard those the system does not recognize. Only RAIN Alliance members can get their products RAIN certified. RAIN is the fastest growing segment of the RFID market and uses a single, global standard: UHF Gen 2 (ISO/IEC 18000-63).

"RAIN is different from many other RF systems, and the signal power of a tag signal into a reader rarely, if ever, exceeds -30dBm (1000nW) and infrequently exceeds -40dBm (100nW). Based on input from RAIN industry experts who have worked with this technology for many years, the RAIN Alliance preliminary analysis assumes a signal power of -70dBm (0.1nW)." This is 1,000-10,000 times weaker than NextNav's "wildly inaccurate statement" (Aileen Ryan, RAIN Alliance President & CEO).

According to my own research, some are noting that the technology coming out of NextNav Inc., while once promising, is largely noncompetitive and grossly overstated. Its Pinnacle Network—which would provide vertical location (z-axis) positioning—is not only competing with Apple and Google but requires a third-party chip in phones that is unlikely to gain traction. Its TerraPoint Network requires a massive infrastructure buildout, estimated at ~$1 billion, that is unlikely to occur in the short term. The goal of this Network is to replace GPS with a terrestrial network that enables tech-like drone delivery and eVTOLs (electric vertical takeoff and landing aircraft). "NextNav is a classic example of a SPAC that should not exist" (https://www.bleeckerstreetresearch.com/research/nn). SPAC is a special purpose acquisition company formed to raise money through an initial public offering so it can later purchase or merge with an existing company. SPACs have no commercial operations and are formed solely to raise capital through an IPO and then be used to acquire or merge with another company. NextNav, founded in 2007, aligned itself with real estate in the Metaverse, self-driving cars, and flying cars. Despite its lofty-sounding goals and promise of a better alternative to GPS—able to locate people inside buildings and in dense city environments—it found itself on the brink of bankruptcy in 2021.

But let's focus on our featured article in which the authors investigated the ability of a RAIN RFID system to read various tag population sizes in the presence of a noise source in the ISM bands that exceed current FCC limits. The authors found that their data clearly showed that, during the first few seconds when RAIN RFID tags are read, they are highly susceptible to a certain noise profile when it is transmitted at power levels that are higher than Part 15 currently allows.

Editorial Review Board Members

Potential Impacts of NextNav on RAIN RFID

Kevin Berisso, University of Memphis; Rajesh Balasubramanian, University of Memphis

Abstract

NextNav Inc. has generated a significant buzz in the wireless signals world by petitioning to have the current industrial, scientific, and medical (ISM) bands between 902 MHz and 928 MHz modified. The comments and responses that have been collected by the United States Federal Communications Commission (FCC) as a part of their review process for the NextNav petition have speculated on the impacts of the changes for the RAIN RFID community. What has been missing is physical testing to back up the various perspectives that both sides have adopted. In this current study, the authors investigated the ability of a RAIN RFID system to read various tag population sizes in the presence of a noise source in the ISM bands that exceeds current FCC limits. The authors report on testing done at the University of Memphis AutoID Lab and discuss the impacts of radio frequency noise.

Introduction

On April 16, 2024, NextNav Inc. submitted a petition to the United States Federal Communications Commission (FCC) requesting a modification to the industrial, scientific, and medical (ISM) radio frequency bands. The RAIN Alliance and AIM Inc., along with numerous other organizations, submitted comments to the FCC requesting that the NextNav petition be denied. A primary concern from these groups was the potential transmission levels of between 1000W and 4000W, as indicated in the NextNav petition documentation (Federal Communications Commission, 2024). In their reply to the initial comment period, NextNav (2025) stated that "commenter(s) assert or assume, generally without empirical support or study, that NextNav's proposal would worsen the interference environment for the Part 15 devices in the band" (p. 25). In furtherance of their request, and in response to some of the feedback that the FCC received during the open comments period, NextNav filed an engineering study on February 27, 2025.

That study reported on initial independent research done in response to NextNav's concern within their reply, as well as the NextNav engineering report, which specifically stated that the proposed NextNav reallocation of the 902-928MHz ISM band would result in minimal interference to the RAIN radio frequency identification (RFID) community (NextNav Inc., 2025). Since filings from NextNav and groups opposed to the NextNav petition only contain theoretical, calculated, or simulated data supporting the various positions, the authors of this current study attempted to generate empirical data from which the FCC could extrapolate the potential impacts of the NextNav petition on the RAIN RFID industry. RAIN RFID is the ultra-high frequency RFID technology that has been adopted by much of the retail industry (e.g., Wal-Mart, Dick's Sporting Goods, Macy's, etc.) and their suppliers (Delen, Hardgrave & Sharda, 2007; Thiesse & Michahelles, 2006). More specifically, it is the technology behind the electronic product code (EPC) that is found in many of these stores (Aguirre, 2007; Hardgrave, Patton, Periaswamy, Payne, Chambers & Gulley, 2021). The research conducted in this current study was not directly funded. Instead, indirect funding by various RAIN Alliance members by way of the various in-kind donations that have been made over the years provided the means to conduct the study.

Background

The ISM bands within the United States allow for unlicensed usage, so long as various limits are complied with by the system. FCC 47 CFR § 15 ("part 15") include the current rules for governing the use of unlicensed devices that transmit radio frequency signals via frequency hopping, such as the RAIN RFID solution. According to FCR 47 § 15.247, systems are required to hop between at least 50 channels and can dwell on a channel for no longer than 400 milliseconds if the bandwidth of the hopping channel is less than 250 kHz. Furthermore, systems are constrained to 1W of transmission power with 6 dBi of directional antenna gain for 50-channel solutions (Federal Communications Commission, n.d.).

As a part of the ISM band, the advantage of working within these rules is that a dedicated radio license is not required, allowing for the ease of mass production of the equipment for general use. Since the late 1990s, various RFID solutions have been leveraging this option for use in supply chain tracking and optimization, with a significant boom in usage beginning in the early 2000s because of the efforts of Wal-Mart, the U.S. Department of Defense, and the AutoID Lab at the Massachusetts Institute of Technology (Sarma, 2005). The resulting solutions, currently governed by GS1's EPCGlobal initiative, have been internationally adopted as ISO/IEC 18000-63 Information technology—Radio frequency identification for item management.

NextNav Inc. is "a leader in next generation positioning, navigation and timing (PNT), enabling a whole new ecosystem of applications and services that rely upon 3D geolocation and PNT technology" (NextNav Inc, n.d.). As a part of their petition, NextNav requested that the FCC "…reconfigure the 902-928 MHz band (Lower 900 MHz Band) to enable a high-quality, terrestrial complement and backup to the U.S. Global Positioning System (GPS) on which the nation relies for the positioning, navigation, and timing (PNT) services" (NextNav Inc., 2024, p.i). As a part of their original petition, NextNav indicated that the lower

900 MHz band is underutilized (NextNav Inc., 2024) and that the reallocation will support PNT efforts that will provide a terrestrial backup for GPS.

In their April 28, 2025, response, NextNav modified their position, removing their contention that the space is underutilized, but instead stated that the "[d]ecades-old command-and-control Multilateration Location and Monitoring Service (M-LMS) service rules in the Lower 900 MHz band stand in the way of 5G-powered 3D PNT" (NextNav Inc., 2025, p.7). Specifically, NextNav stated that "private- and public-sector experts have concluded that the United States needs robust terrestrial PNT to complement and back up GPS because space-based PNT systems' coverage gaps and vulnerabilities pose significant risks to U.S. national security, economic, and public safety interests" (NextNav Inc., 2025, p.ii). Figure 1 shows that NextNav, in support of their request, requested that the upper third (918-928 MHz) of the IMS part 15 frequency allocation be modified such that they are allowed to transmit at higher levels.

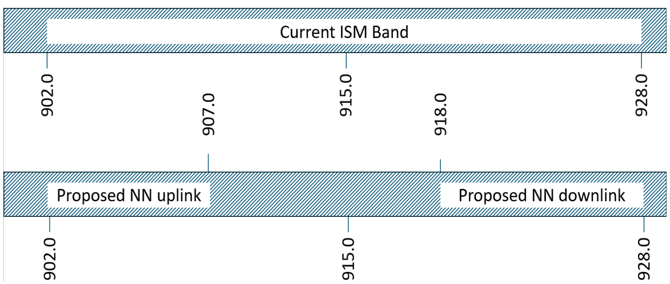


Figure 1. NextNav frequency reallocation request: top—current ISM allocation; bottom—blocks NextNav is requesting.

RAIN RFID is the name for ultra-high frequency RFID solutions that comply with ISO/IEC 18000-63. The RAIN Alliance is "a consortium of companies that together want to create a smarter and more sustainable world by using RAIN technology to connect trillions of everyday items across their entire lifecycle, simply and inexpensively" (RAIN Alliance, n.d.). The RAIN Alliance supports its members by creating awareness and adoption efforts across multiple industries. Used for everything from access control to retail solutions and even tolling solutions, RAIN RFID has reported that 52.8 billion tags/ICs were shipped in 2024 (RAIN Alliance, 2025). Common use cases center around the supply chain, supporting the tracking of pallets and products as well as individual items. At the individual item level, RAIN RFID is allowing retailers to better understand what is in their stores, as well as the supply chain, enabling consumers to do things such as buy online and then pick up their items in the store—often on the same day (Hardgrave et al., 2021; Latini, 2019; Patton, Periaswamy, Dunn & McDaniel, 2023). Other use cases include vehicle tagging for toll roads, the tracking of automobile tires, and even passive sensing solutions (e.g., temperature, humidity, strain, etc.).

Methodology

The test plan supporting this research was intentionally simple and limited in terms of scope and scale to ensure a timely publication of the results and to ensure that it fits within the functional limitations discussed below. Additionally, the unfunded nature of the research resulted in additional delimitations that will be discussed. The primary limitation of this study lies in the fact that NextNav has not publicly published any sort of specification on the characterization or power level of the transmissions they plan for the 902-928 MHz frequencies. As a result of this, the best that can be done is to test one of the two potential transmission methods that are anticipated: a narrow-band-frequency hopping solution or a broadband "white-noise" type of solution. Due to equipment accessibility limitations, this study was limited to testing the narrow-band-frequency hopping solution with the following assumptions.

- A transmission profile generated by a RAIN RFID reader, but without data on the carrier wave that would have been a sufficient interference simulant.
- A transmission power of 4W with a 6 dBi gain antenna that was sufficient to provide a minimal simulated level of interference.
- No testing of band interference that would have minimally altered the results.

After numerous discussions with others in the RFID field, it was determined that using a RAIN RFID reader that had the capability of transmitting the carrier wave at a higher than allowed power level, but without any transmitted data, would be sufficient for simulating a worst-case scenario (from the RAIN RFID perspective) in terms of interference from NextNav. The rationale is that if NextNav were to follow a similar narrow-band-frequency hopping approach to what RAIN RFID already uses then the presence of an RF carrier wave following a similar frequency-hopping pattern would have the greatest chance of confusing the RAIN RFID tags, since the tags would be predisposed to trying to communicate with the signal. Alternatively, a broadband noise source that covered up to 10 MHz of bandwidth at greater than allowed levels would be expected to blanket the 918-928 MHz range, resulting in either the tag's inability to "hear" the RAIN RFID reader or the tag's coupling with the broadband signal resulting in an inability of the RAIN RFID reader to communicate with the tag. However, a lack of access to the necessary equipment for generating this type of signal precluded it from the study, reserving such a test for future efforts.

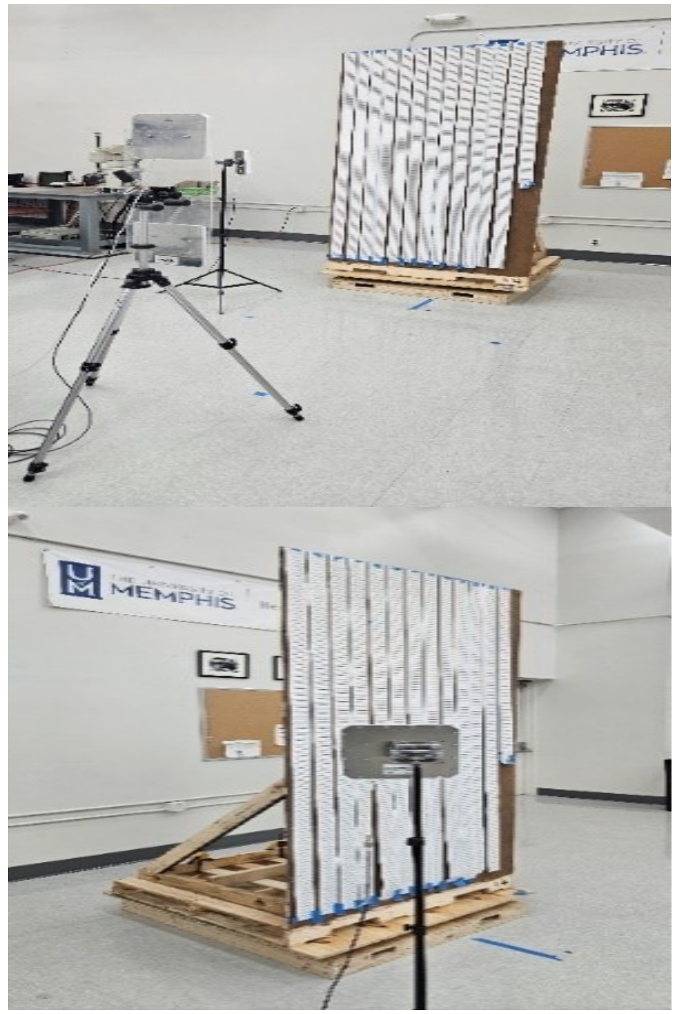
The selection of 4W of power with a 6 dBi antenna was a purely practical choice, as the RAIN RFID reader selected to generate the "noise" had a 4W maximum power limit. Therefore, the obtained FCC special temporary authority that was received requested a 4W maximum transmission limit with a 6 dBi gain antenna. Since the actual signal level experienced by the RAIN RFID tag from the NextNav transmitter was dependent upon the distance to the NextNav transmitter and the environmental conditions (e.g., buildings, dense trees, etc.), 4W of power at a distance of 3.25 feet was assumed to be a reasonable simulation of a potential signal strength for the purposes of testing. It was determined that, while a more powerful noise source could potentially result in a decrease in the RAIN RFID reader's ability to read tags, the 4W provided by the available equipment would be sufficient.

The RAIN RFID tags used for the test were Avery Dennison Belt tags with the NXP UCode 8 and NXP UCode 9 chips. The NXP UCode 8 tag is the older generation of tag relative to the UCode 9 tag, with the primary difference being that the UCode 9 is slightly more sensitive. The other primary differences (write speeds, memory size, etc.) were not assumed to have had an impact on this study. The RAIN RFID tags were combined with a 73 mm x 17 mm label and were encoded with unique values in the EPC memory so that each tag was uniquely identifiable, both within and between groups. Because of the default spacing of the UCode 8 tags as they came from Avery Dennison, the labels were placed on heavy printer paper to make mounting and moving of the tags as similar as possible between the two sets. The equipment used for the tests was as follows.

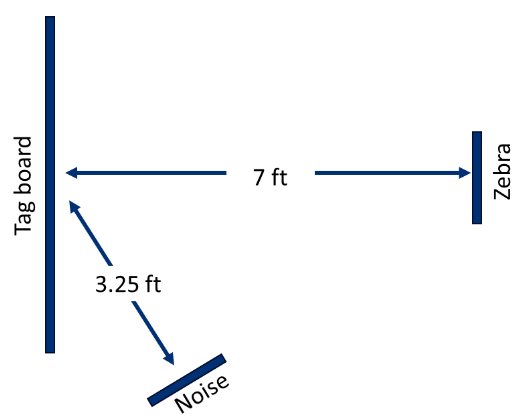
- CSL CS463 four-port reader – RAIN RFID reader
- Impinj R700 four-port reader – RAIN RFID reader
- Zebra FX9600 eight-port reader – RAIN RFID reader
- Two 8 dBi circular polarized antenna (attached to RAIN RFID readers)
- 6 dBi circular polarized antenna attached to noise source
- Feig LRU3500 four-port reader (noise source)

Figure 2 shows the test environment. As can be seen, this is not an anechoic chamber but instead was conducted inside the AutoID Lab at the University of Memphis. Both the reader antennas and the noise antenna were placed at a height such that they were aligned with the middle of the tag board in the vertical plane, and the noise antenna was at a 33° angle relative to the tag board with the indicated 3.25 feet of separation to the center of the tag board.

The area around the test board was cleared of all existing tags and the RAIN RFID reader was run for approximately 10 minutes to determine if the space was sufficiently clear of tags. And while initial tests showed that no tags were within range, up to 10 tags did occasionally show up in the recorded data, indicating that the space was not as empty as originally believed. However, continued attempts to find the wayward tags failed to locate them, so it was determined that they would just be purged from the dataset during post-processing. The noise was generated by connecting the 6 dBi circular polarized antenna to the LRU3500 and then by setting the Feig LRU3500's antenna port to 4W and the channel range from 917.25 MHz to 927.25 MHz. The LRU3500's radio was turned on by setting the RF Output OnOff value to "On" for the connected antenna.



(a) Pictures of the testing area.



(b) Distances to the tag board.

Figure 2. Test setup and lab configuration.

Tests were run without noise and with noise for 1, 50, 100, 250, 500, 750, 1000, and 1500 tags mounted to the tag board. The selection for the population sizes of tags was done based on the assumption that there was a need to oversaturate the reader in the sense that the true maximum number of tags being read was achieved. By ensuring that the reader was unable to read the full maximum tag population, any reductions in the number of tags read could be directly and proportionately assigned to the current test and not to limitations with the reader, its configuration, or the antenna selection. Each tag population was then interrogated by the RAIN RFID reader under test for 120 seconds, with the resulting tag reads being individually captured and recorded. The data were then cleaned to capture the number of unique tags and total number of reads per second, resulting in 120 data sets per combination of reader (CSL, Impinj, Zebra), tag (NXP UCode 8, NXP UCode 9), and for the two test conditions (without noise, with noise).

Results

The number of times each RAIN RFID reader reported the reading of a tag using the demonstration software provided by each reader manufacturer and custom software was developed to clean the data for plotting. Stray tags that were not part of the specific test set were removed, and all data sets were trimmed to a total of 120 entries. Figure 3 shows the average cumulative number of unique tags read for the 1500-tag population across all three readers over the 120-second time period for both tag models and with and without the noise. The data clearly show that there was both a slower accumulation of unique tags and a reduction in the total number of unique tags read when the 4W noise was present. Additionally, it must be noted that while there was a significant difference in general performance between the two tags without noise, when the noise was introduced, both tags experienced a similar reduction in their ability to be read by the RAIN RFID readers.

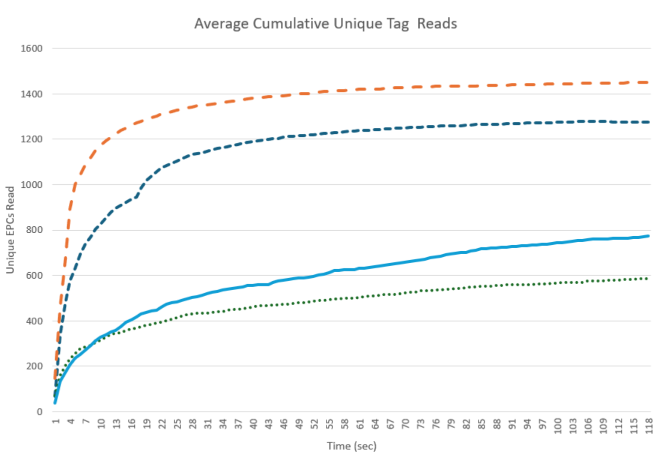


Figure 3. Plot of the cumulative number of unique tags read over time for the 1500-tag population.

Figure 4 is a zoomed in version of Figure 3, showing just the first 10 seconds of time. As can be seen in Figure 4, the first three to 10 seconds of time makes a significant difference in the number of tags seen; but, when noise was introduced, both tag models suffered, although the UCode 9 suffered more during the first 10 seconds. This difference is most likely attributed to the increased sensitivity that the UCode 9 has been designed with, resulting in a higher level of susceptibility to external factors.

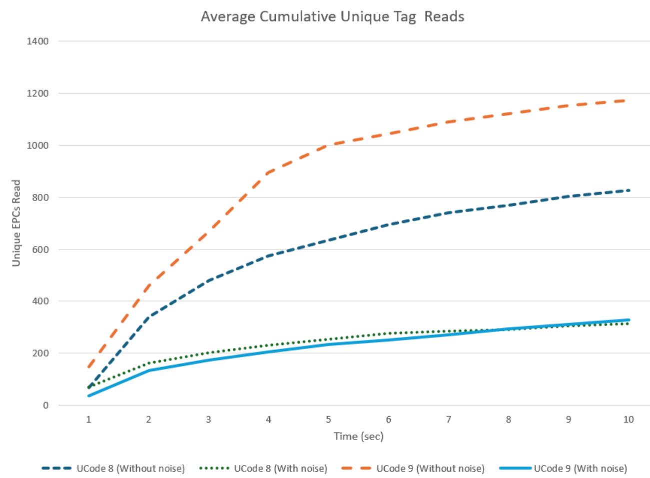


Figure 4. Plot of the cumulative number of unique tags read in the first 10 seconds for the 1500-tag population.

Figures 5-7 show the percentage of the overall number of unique tags read at 3, 5, and 10 seconds, respectively. As can be clearly seen, even for the smaller tag populations, there was a drop in the ability of the reader to read the tags when the noise source was enabled. Even after 10 seconds, for tag populations over 250, at best only one third of the tags were seen when the noise was enabled.

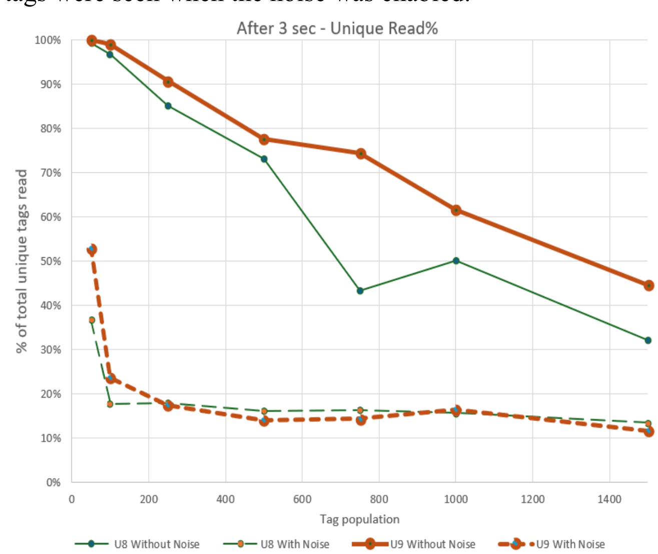


Figure 5. Plot of the cumulative number of unique tags read in the first 3 seconds.

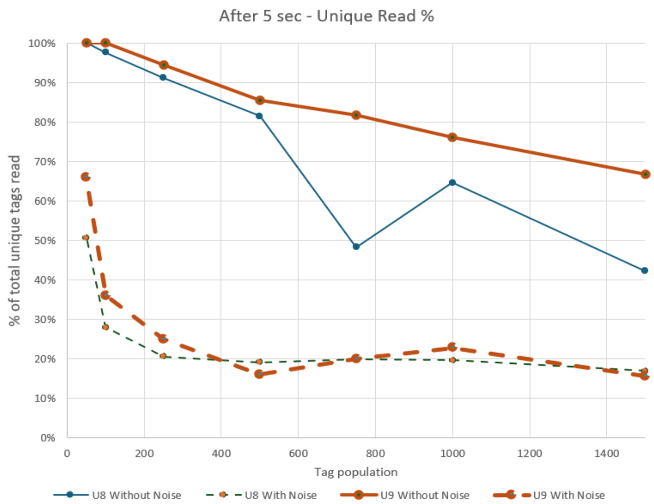


Figure 6. Plot of the cumulative number of unique tags read in the first 5 seconds.

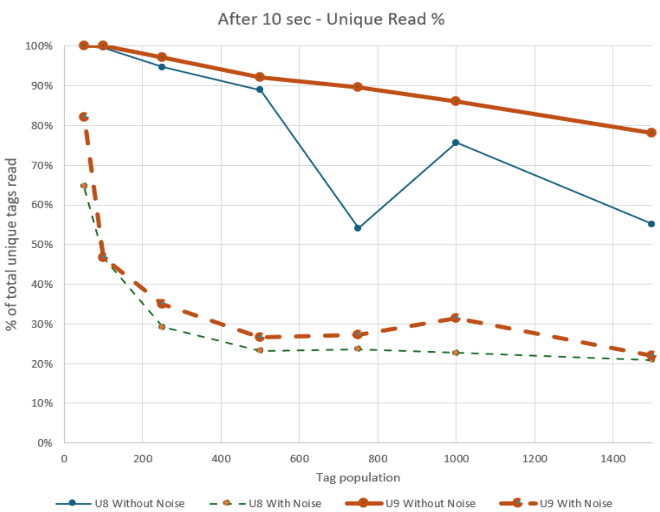


Figure 7. Plot of the cumulative number of unique tags read in the first 10 seconds.

One additional item of note is that, even after 10 seconds, the 50-tag population test still was not able to read 100% of the tags when the noise was enabled, whereas it was able to read all 50 tags without the noise. This clearly indicates that, despite the lack of a data signal, when other devices were transmitting at greater than currently allowed levels on the same frequencies that RAIN RFID used, the tags attempted to communicate with the louder signal, resulting in an inability of the tags to communicate with the RAIN RFID reader.

Conclusions

The data clearly show that, while the RAIN RFID readers were still able to read tags in the presence of the tested noise profile, there was a significant loss of information (tag reads) when there were more than 100 tags in range of the reader. For cases where the tags remain static and within the reader's range, it can be hoped that all the tags might eventually be read. However, many of the cases for RAIN RFID included tags, or readers, that were in motion (e.g., dock doors, conveyor systems, retail store floors, etc.). In these cases, there could be a limited amount of time—sometimes as little as one second—where the read event must occur. In these events, when the number of tags to be read exceeds 50 (which is typical), the introduction of additional noise can result in a significant reduction in the number of unique tags read during the first few seconds, resulting in missed or lost tags. For solutions where high-dollar values are tied to the objects and to which the RAIN RFID tags are affixed, this could have severe financial results.

These tests showed that additional testing, specifically with a NextNav transmission profile, is needed before a conclusive opinion can be developed in terms of the impacts of the NextNav petition. While it is acknowledged that the inferences that can be derived from these data are limited to a noise profile that is similar to what was tested, it clearly shows that during the first few seconds when RAIN RFID tags are read, they are highly susceptible to this noise profile when it is transmitting at power levels that are higher than part 15 currently allows. And if the NextNav transmission profile ends up with a similar design, there is a high probability that RAIN RFID operations will experience issues. Other areas that could be pursued relative to this study include the previously mentioned testing of a broadband signal, testing against an actual NextNav signal, the optimization of the tag population selection, and an improved reader optimization routine. NOTE: the raw data files can be made available, should the reader wish access.

Acknowledgements

The authors would like to thank Mr. Josef Preishuber-Pflügl, Head of Regulatory Affairs for the RAIN Alliance, for his advice in the design of the testing protocol. Tags were provided thanks to Mr. Jason Ivy—Logistics Solution Consultant Director for Avery Dennison. The authors would also like to thank Mr. Rod Saunders of Convergence System Limited (CSL) for the loan of the CS463 RAIN RFID reader, Ms. Debbie Greenway—EVP of Feig Electronics, Inc., and Mr. Jason Warschauer—Director of Technical Solutions of Feig Electronics, Inc., for the loan of the Feig RAIN RFID reader that was used as the noise source. The Impinj R700 reader was donated, thanks to Dr. Megan Brewster, VP Advanced Technology at Impinj. The RFID antenna selection was assisted by Times-7. The Zebra RAIN RFID reader was donated to the AutoID Lab prior to the testing by Mr. Michael Fein and Mr. Anthony Tortorello of Zebra. Finally, the authors would like to thank Dr. John Kim of NextNav Inc., for his time and explanations of how the NextNav system would work and for his clarification of some of the technical aspects of their petition.

References

Aguirre, J. I. (2007). *EPCglobal: a universal standard* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Delen, D., Hardgrave, B. C., & Sharda, R. (2007). RFID for better supply-chain management through enhanced information visibility. *Production and operations management*, *16*(5), 613-624.

Federal Communications Commission. (n.d.). *47 C.F.R. Part 15 – Radio frequency devices*. Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-47/chapter-I/subchapter-A/part-15

Federal Communications Commission. (2024, September 5). *Comment submitted by RAIN Alliance & AIM Inc*. FCC Electronic Comment Filing System. https://www.fcc.gov/ecfs/search/search-filings/filing/10906412205140

Hardgrave, B., Patton, J., Periaswamy, S., Payne, J. C., Chambers, J., & Gulley, A. (2021, March 16). *AUBIE: BOPIS index evaluation*. Auburn University RFID Lab. https://rfid.auburn.edu

Latini, M. A. R. T. I. N. A. (2019). *Factors that influence the adoption of an online-to-offline strategy in retail* (Unpublished doctoral dissertation). Thesis, Maynooth University.

NextNav, Inc. (n.d.). *Investor relations overview*. https://ir.nextnav.com/overview/default.aspx

NextNav, Inc. (2024, April). *Petition for rulemaking of NextNav, Inc.* https://nextnav.com/wp-content/uploads/2024/04/Petition-for-Rulemaking-of-NextNav-Inc.pdf

NextNav, Inc. (2025, February 27). *5G NR and unlicensed Part 15 technologies in the lower 900 MHz band: A coexistence analysis* [Technical report]. https://nextnav.com/wp-content/uploads/2025/02/NextNav-Technical-Study-Feb.-27-2025-2.pdf

NextNav, Inc. (2025, April 28). *Comments of NextNav Inc. in the matter of promoting the development of positioning, navigation, and timing technologies and solutions (WT Docket No. 25-110)*. Federal Communications Commission. https://www.fcc.gov/ecfs/document/10428748204313/1

Patton, J., Periaswamy, S., Dunn, L., & McDaniel, A. L. (2023). *Retail case studies: 2023 survey of retail industry innovation*. Auburn University RFID Lab.

RAIN Alliance. (n.d.). *Mission and vision*. https://therainalliance.org/mission-vision/

RAIN Alliance. (2025, February 26). *RAIN Alliance report shipments of 52.8bn RAIN tag chips globally in 2024.* https://therainalliance.org/rain-alliance-report-shipments-of-52-8bn-rain-tag-chips-globally-in-2024/

Sarma, S. (2005). A History of the EPC. In S. Garfinkle & B. Rosenberg (Eds.), *RFID: Applications, Security, And Privacy* (pp. 37-55). Addison-Wesley Professional.

Thiesse, F., & Michahelles, F. (2006). An overview of EPC technology. *Sensor review*, *26*(2), 101-105.

Biographies

KEVIN BERISSO is the Director of the AutoID Lab at the University of Memphis. He has been teaching within the AutoID industry for over 20 years, is a member of the AIDC 100, and has been awarded the Ted Williams award by AIM Inc. for his work in the industry. He is a regular presenter at RFID Journal Live and has worked with GS1 on barcode research as well as having been involved in RFID research for returnable transport items in the automotive industry and tagging items in the aerospace industry. Dr. Berisso may be reached at kberisso@memphis.edu

RAJESH BALASUBRAMANIAN is an assistant professor at the University of Memphis. He worked in industry for over 27 years before accepting a teaching position. He has been teaching quality control, facility design, and project planning to undergraduate students. Given his industrial background, he works on statistical analysis and data presentation in the RFID space. Mr. Balasubramanian may be reached at rbalasub@memphis.edu

Design and Implementation of a Modular Integrated Stackable Instrumentation System for CubeSat Research Experiments

Gang Sun, Northern Kentucky University; Avery Devore, Northern Kentucky University; Jaycie Bussell, Northern Kentucky University

Abstract

Microgravity experiments offer significant potential for advancing both life on Earth and deep-space exploration. CubeSats have become essential platforms for such research, with projections indicating between 300 and 500 launches annually. However, many CubeSat missions face reliability challenges due to several key issues: the use of non-aerospace-grade electronics as well as limited system adaptability and functionality. To address these issues, the authors of this current project studied the design and implementation of a cost-effective, modular, and scalable instrumentation system, based on NASA-standard modular integrated stackable layers (MISL) architecture.

The new instrumentation system supports CubeSat experiments requiring multi-channel data acquisition and storage by integrating space-qualified MISL hardware, reliable electronic components, and modern communication protocols. The development process included a series of structured engineering stages: conceptual block diagram (CBD), functional block diagram (FBD), detailed schematics, printed circuit board (PCB) layout, and bill of materials (BOM) design. Hardware and software testing, conducted using the MISL-ASE (analog system environment) platform, verified system functionality in power management, sensor data sampling, conversion, processing, and storage. Testing results confirmed that the instrumentation system met functional requirements. Two areas for future improvement were identified: the incorporation of signal conditioning modules and refinement of hardware layout. This work contributes a robust instrumentation framework for enhancing the reliability and adaptability of CubeSat experiments and supports the broader development of space-qualified systems for emerging scientific and commercial missions.

Introduction

Experiments conducted in microgravity environments hold significant potential for discoveries that could improve life on Earth and advance future deep-space exploration (NASA, 2024). Recently, miniaturized nanosatellites, known as CubeSats, have been increasingly used in low-earth orbit (LEO) to carry out these in-space experiments. Compared to conventional satellites, CubeSats offer a remarkable reduction in design and launch costs. Additionally, CubeSats can be networked together in orbit to form satellite constellations, providing inherent redundancy and potentially greater coverage. As a result, CubeSats have opened up new opportunities for microgravity scientific research. Since 2018, NASA has transitioned from government-led to private sector-led human spaceflight activity in LEO, aiming to promote technological innovations and foster a thriving LEO economy (Besha & MacDonald, 2024). This shift has attracted significant private enterprise investment in space, more than ever before. The commercialization of LEO has primarily focused on the design and launch of low-cost CubeSats, along with the development of electronics for data acquisition and communications in microgravity. Over the next few years, it is anticipated that between 300 and 500 CubeSats will be launched annually, with the majority of these missions transitioning from academic projects to commercial ventures.

However, it is worth noting that nearly 50% of university-led CubeSat missions failed between 2003 and 2012 (Swartwout, 2013). Several factors contribute to this high failure rate, but one key issue is that CubeSats often rely on off-the-shelf electronic components (e.g., Arduino, Raspberry Pi, etc.), which are not always capable of meeting the rigorous demands of the aerospace environment (Morgan & Porter, 2015). These consumer-grade components and platforms are not designed to withstand extreme temperatures, radiation exposure, and launch-induced vibrations encountered in space missions. Moreover, power efficiency and electromagnetic compatibility are often suboptimal in such systems, further increasing the risk of mission failure. Furthermore, many CubeSat electronic systems lack the reconfigurability needed to support the diverse requirements of microgravity research experiments, which limits their applicability across different missions.

To improve mission success rates, there is a clear need for electronic systems that are both reliable and adaptable to space conditions. In this current project, the authors adopted NASA-standard, space-qualified modular integrated stackable layers (MISL) hardware architecture to develop a cost-effective, modular, and scalable instrumentation system for data acquisition and storage in CubeSat experiments. The MISL architecture (Yim, 2012), originally designed by NASA's C&DH Branch at Johnson Space Center (JSC), consists of a compact, modular computer system capable of withstanding harsh environments, including temperature extremes, vibrations, and radiation. Figure 1 shows how the MISL stack is built from industry-standard layers that can be easily assembled to create an embedded system. Developers can select from existing MISL layers, such as those for power, intelligence, ethernet, Wi-Fi, and sensors, or they can design custom layers to meet specific needs.

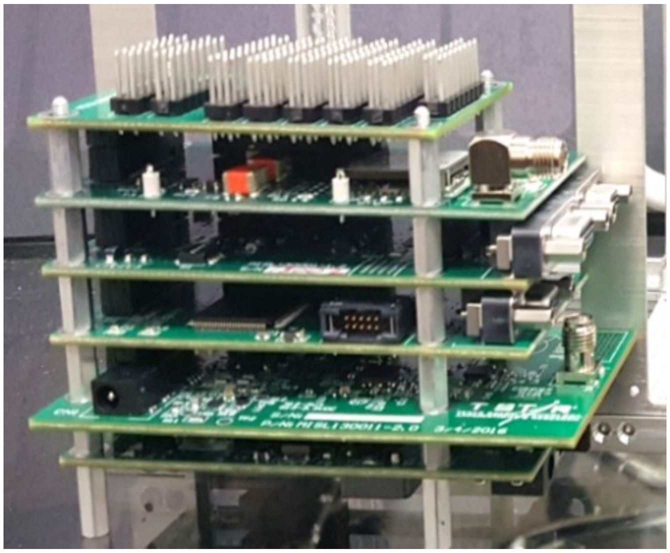


Figure 1. The MISL stack.

Currently, there is no available MISL instrumentation layer that supports experimental data sampling, conversion, and storage. The objective of this study, then, was to design a cost-effective, modular, and scalable instrumentation system for data acquisition and storage in CubeSat experiments by integrating state-of-the-art embedded MISL architecture with modern electronic components and communication protocols.

Materials and Methods

The flowchart of Figure 2 shows how the entire design process of the MISL-based instrumentation system followed a series of carefully planned development stages.

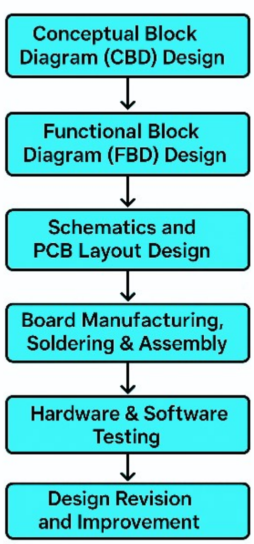


Figure 2. Flowchart of the entire design and development process.

The first stage involved the conceptual design during which a conceptual block diagram (CBD) was created to outline the basic functions and functional requirements of the system. In the second stage, the functional requirements were translated into specific, measurable performance specifications for each requirement (e.g., power, data bus, interface protocols, etc.). In the third stage, a comprehensive functional block diagram (FBD) was developed to depict the components, pin configurations, signals, and all necessary interfaces to ensure proper system functionality. Following this, detailed schematics and a printed circuit board (PCB) layout were created using Altium professional design software (Altium, 2024). The next stage involved the manufacturing of the instrumentation layer during which electronic components were populated onto the boards using automated assembly machines. Finally, the hardware unit was tested and the software was evaluated to ensure that the system's overall functionality met the defined requirements.

Design of a MISL-based Instrumentation System: The Conceptual Block Diagram Design

Figure 3 illustrates the CBD of the MISL-based CubeSat instrumentation system, which included a new MISL instrumentation layer for multi-channel, high-resolution data sampling, conversion, processing, and storage, alongside existing MISL power and MSP430F5438A intelligence layers. Figure 3 also shows the layout of the MISL architecture, including the positions of the 40-pin power bus and 100-pin data bus connectors. The data bus connector was always positioned horizontally at the top of the layer, while the power bus connector was positioned vertically on the left side of the layer. This standardized placement allowed the board to be any size, while ensuring the correct alignment of the data and power bus connectors.

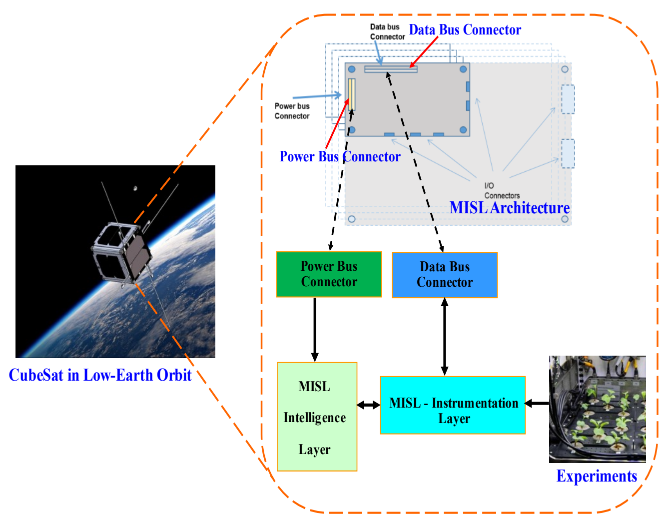


Figure 3. The CBD of the MISL-based CubeSat instrumentation system.

The Functional Block Diagram Design

Figure 4 illustrates the overall FBD of this CubeSat instrumentation system that supports simultaneous design and development of both hardware and software. The proposed MISL instrumentation system integrates an 8-channel, 16-bit, high-speed external analog-to-digital converter (ADC) (the Texas Instruments ADS8345) that was designed to facilitate precise data sampling, conversion, and processing for CubeSat experiments. The ADS8345 is particularly well-suited for such applications, due to its high resolution and speed, as it provides the accuracy needed for sensitive measurements in space.

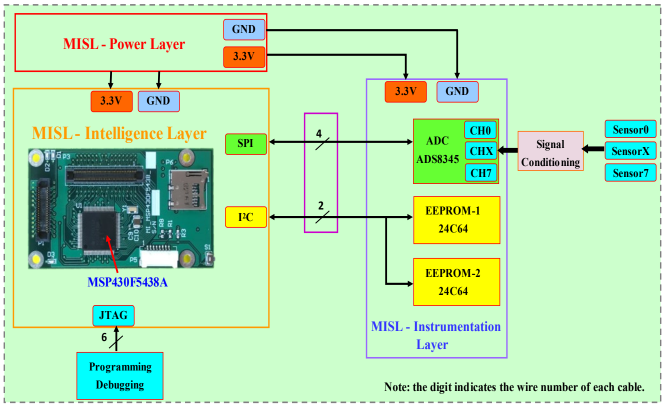


Figure 4. The overall FBD of the MISL-based CubeSat instrumentation system.

A key feature of the ADS8345 is its high-resolution conversion, which significantly enhances the accuracy of data by converting analog signals into a digital format with 16-bit precision. This higher resolution ensures that even the smallest variations in the analog signal from a sensor are captured, providing more accurate and reliable results. The ADS8345 operates through a synchronous serial interface, the serial peripheral interface (SPI), utilizing pins (MOSI, MISO, SCLK, and CS) from the P9 port of the MSP430F5438A microcontroller, which itself resides on the MISL intelligence layer. The SPI interface is widely used for its simplicity and speed, making it ideal for high-performance data acquisition tasks.

The combination of the ADS8345's high-speed conversion and low power consumption makes it an excellent choice for portable and battery-operated systems, such as multi-channel data loggers and remote measurement equipment. Additionally, the inclusion of a low-power, high -speed multiplexer allows the ADS8345 to handle multiple input channels, which is particularly beneficial for applications requiring the simultaneous measurement of several sensors. This feature enhances the versatility and scalability of the MISL instrumentation layer, enabling it to support a broad range of experiments in space research, where efficiency and precision are paramount.

Although the MISL intelligence layer includes an SD card socket for data storage, SD cards require significant space and power, which may not be ideal for low-power, compact systems like CubeSats. To address these limitations, multiple 2-wire serial electrically erasable programmable read-only memory (EEPROM) modules (24C64-64K) were integrated into the MISL instrumentation layer for experimental data storage. The 24C64 EEPROM offers a low-power alternative and allows data to be written and retained for extended periods without draining excessive power. This memory type is particularly suitable for space applications, where minimizing power consumption is crucial for prolonged mission durations.

The EEPROM 24C64 modules communicate with the MSP430F5438A microcontroller on the MISL intelligence layer via the inter-integrated circuit (I^2C) bus. The I^2C interface enables half-duplex communication and allows the microcontroller to read from and write data to the EEPROM devices efficiently. This setup provides a reliable and energy-efficient solution for storing experimental data in space research, where data logging and preservation are vital for the success of long-duration space missions. The use of multiple 24C64 EEPROMs further enhances the system's capacity to handle up to eight data channels simultaneously from the A/D converter, while maintaining a compact and low-power design.

The Schematic and Printed Circuit Board (PCB) Design

Once the circuit and peripheral interface designs for the MISL instrumentation layer were finalized, Altium design software was utilized to create both the schematic and the PCB layout. Figure 5 depicts the main circuit and peripheral interface designs, including the wiring connections between the ADS8345 ADC, the 24C64 memory, and the MISL data and power connectors. For the PCB layout design, design rules were strictly followed to ensure manufacturability and to meet performance standards. The final output of the layout was a set of Gerber files, which were sent to PCB manufacturers for fabrication. Moreover, a BOM was compiled and a cost analysis was performed to optimize quantity discounts for commonly used components. For the version I design, the main issue was that the data bus connector was slightly mispositioned on the layer, preventing it from properly connecting to the MISL intelligence layer. Also, two easily resolvable issues were identified. First, an incorrectly placed C3 capacitor disrupted the ADC's reference voltage. This capacitor was mistakenly positioned in series with the resistor voltage divider, causing the voltage reference to be equal to the supply voltage. Second, neither of the EEPROM chips were receiving supply voltage, due to an incorrectly placed C5 capacitor that was in series with Vcc (the power supply). This error was corrected by removing the capacitor entirely. Figure 6 displays the version II overall PCB layout of the MISL instrumentation layer.

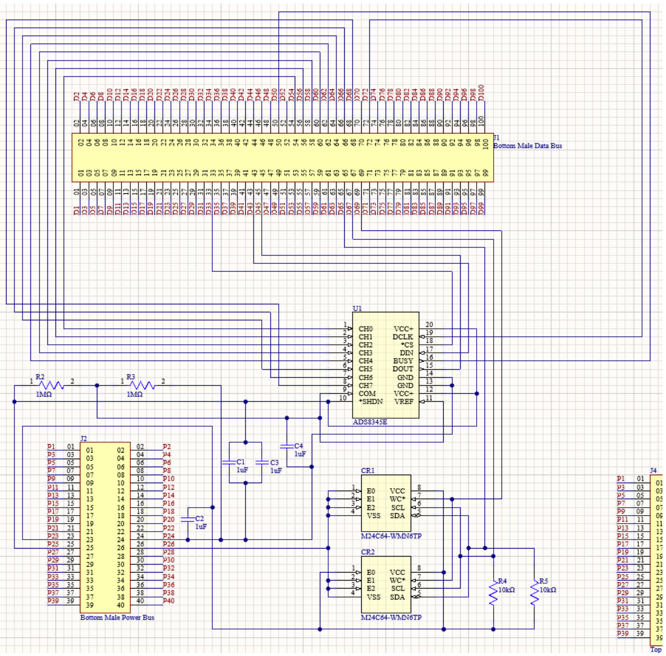


Figure 5. The main circuit and peripheral interface design diagram.

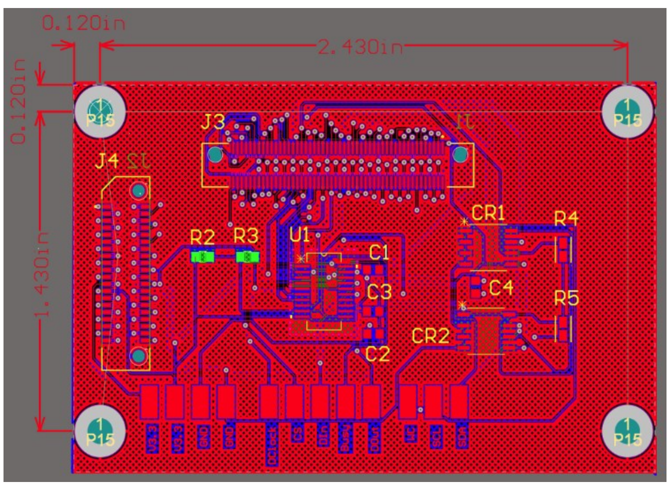


Figure 6. The version II overall PCB layout of the MISL instrumentation layer.

Debugging and Evaluation

After the version II instrumentation layer was manufactured and populated, it was integrated with the MISL power and intelligence layers to form a cost-effective, modular, and scalable instrumentation system; Figure 7 shows the construction of the system for data acquisition and storage in CubeSat experiments. The hardware and software tests of the MISL-based instrumentation system were conducted to verify whether the system's overall functionality met the defined requirements. All traces on the PCB layer were correctly routed; for example, data lines were properly routed to and from the necessary pins, and the power supply (3.3V) and ground (GND) were correctly configured. Power and data were successfully transferred between MISL layers, and the data bus was used to control the electronic components on the instrumentation layer. All passive components had the correct values and were positioned properly on the PCB.

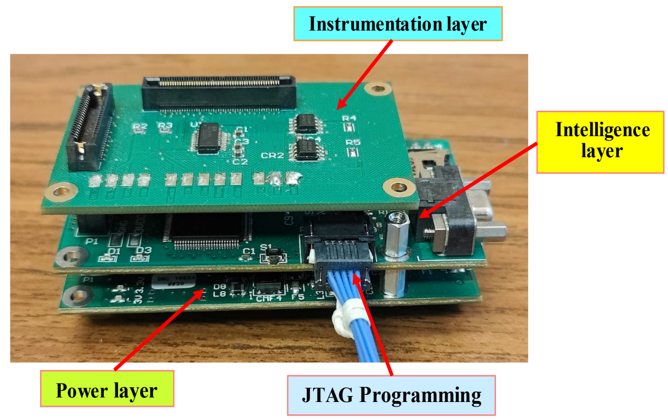


Figure 7. The MISL-based instrumentation system.

Testing the MISL layers was a challenge, as there were no standard output devices or HMIs available, aside from debugging tools such as logic analyzers, oscillators, and multimeters. For this project, the MISL-ASE (modular integrated stackable layer - analog system environment) platform (Sun, Morgan & Porter, 2021) was used for both hardware debugging and software evaluation. Figure 8 shows the MISL-ASE board that incorporated a wide range of fundamental and advanced analog and digital circuits and modules. Fundamental circuits included LEDs, 7-segment displays, a 1602 LCD, switches, keypads, A/D conversion, multiple analog signal generators, and RS-232/485 serial communication, among others. In addition to these basic features, the ASE board included advanced capabilities not typically found in other embedded system development tools, such as a TFT LCD with touchscreen, Ethernet LAN network support, battery life density measurement, a 3-axis accelerometer, a high-resolution ADC converter, motor drive, and remote-control functionality. The MISL-ASE board also supports new communication interfaces and protocols, including UART (USB, RS-232/485, Bluetooth, and Zigbee), SPI (Ethernet, 2.4G Wi-Fi, Micro SD card, and flash memory), I^2C (DAC and EEPROM), and 1-wire communication devices.

As illustrated in the left picture of Figure 8, the stackable hardware structure (circled by blue dash lines) was designed to allow the MISL-MSP430 intelligence layer to be directly interfaced with the ASE board, enabling control of all on-board peripherals and devices. This unique architecture also facilitates the connection of the ASE board to various other embedded intelligence layers (e.g., ATMEL and PIC microcontroller layers) as well as other MISL layers. Thus, in this research project, the MISL-based instrumentation layer, along with the MSP430 intelligence layer, was integrated into the ASE board for hardware troubleshooting, error correction, and programming code inspection.

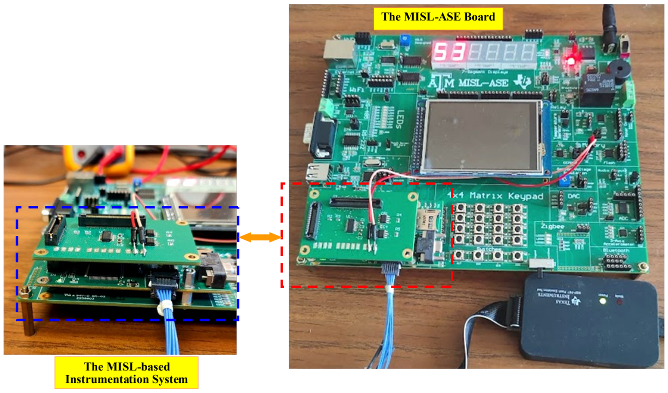


Figure 8. Testing the MISL-based instrumentation system using the MISL-ASE board.

For verifying the data storage function, the TA0 timer of the MSP430 microcontroller was programmed to measure elapsed time in seconds. The timer value, updated every second, was then stored in the EEPROM (24C64). At the same time, the timer value, ranging from 00 to 59 seconds, was displayed on the 7-segment displays of the ASE board. The 24C64 is a 64 Kbit EEPROM with an I²C interface, allowing the storage of sensor measurements while retaining them during power loss. The EEPROM's I²C address is determined by the A0, A1, and A2 pins, which are grounded on the instrumentation layer, thereby setting the device address to 0x50. The MSP430F5438A microcontroller features USCI (Universal Serial Communication Interface) or eUSCI modules that support I²C communication. The SCL (clock) and SDA (data) lines were connected with 10 KΩ pull-up resistors.

After initializing the MSP430 I²C interface, the real-time timer value was written to the 24C64 by sequentially sending the following items: START condition, EEPROM address (0x50), memory address byte (sub_address), the timer value, and STOP condition. When power was restored, the MSP430 microcontroller read the last stored timer value from the EEPROM, and the 7-segment displays continued to show the timer value from where it left off. As depicted on the ASE board in Figure 8, the real-time timer value (53 seconds) was successfully read from the 24C64 and displayed on the 7-segment displays by sending the following sequence: START condition, EEPROM address, memory address byte (sub_address), repeated START, the timer value, and STOP condition, followed by an acknowledgment (ACK). For verifying the sensor data sampling, conversion, and processing functions, a potentiometer was employed to simulate an analog signal input, ranging from 0-3.3V, typically received from a sensor. The potentiometer was connected to one of the eight available input channels on the system.

This analog voltage signal was then fed into the ADS8345, a 16-bit, 8-channel ADC. The ADS8345 is equipped with a SPI that allows for efficient communication with external microcontrollers. The MSP430 microcontroller's UCA2 interface, specifically designed for universal serial communication with SPI peripherals, was chosen for the communication between the MSP430 and the ADS8345. This communication took place over four wires: chip select (CS), serial clock (SCLK), data input (DIN), and data output (DOUT). The SPI interface enables data transmission between the MSP430 and the high-resolution ADC to ensure accurate conversion of the analog signal into a digital value. To establish communication, the SPI settings on the MSP430 were configured, including selecting the appropriate communication mode, setting the clock speed (baud rate), and determining the data order—most significant bit (MSB) first or least significant bit (LSB) first.

Once the configuration was complete, the MSP430 could begin reading the channel connected to the potentiometer. This was achieved by sending an 8-bit control byte that specified the channel and mode. Upon receiving the control byte, the ADS8345 processed the input signal and returned the corresponding 16-bit digital result, which represented the converted value of the analog signal. The converted sensor data value could then be stored in the EEPROM for further processing or analysis. This setup ensures reliable analog-to-digital conversion from CubeSat research experiments, which serves as an essential step in the experimental data processing chain. Figure 9 visually presents the ADS8354 conversion results; sensor measurements were varied by adjusting the potentiometer then sampled and converted by the ADS8354. The converted data were transmitted via UART serial communication and displayed on a computer screen.

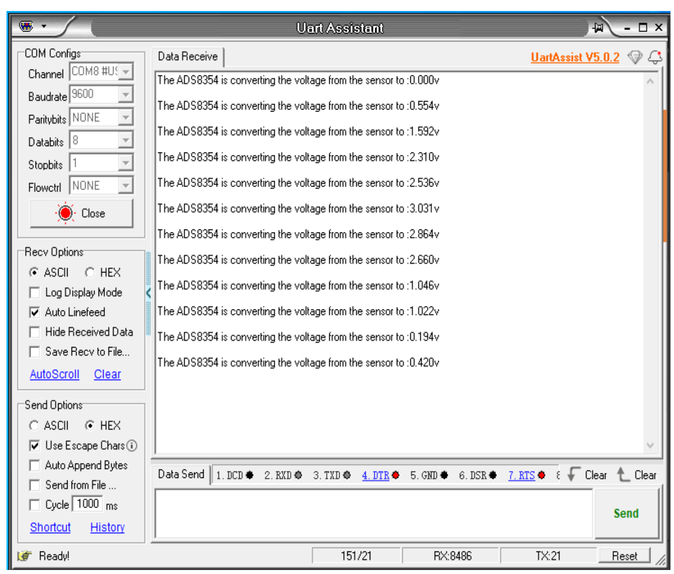


Figure 9. The ADS8354 conversion results shown on the computer screen.

Table 1 presents a test summary matrix to capture all tests conducted to validate the functionality of the MISL-based instrumentation system. This matrix was designed to confirm that the prototype met all specified functional requirements. Functional requirements are listed along the X-axis, while the corresponding tests are listed along the Y-axis. An 'X' in a cell indicates that the respective test successfully verified the corresponding requirement. Rows containing multiple 'X' marks represent tests that addressed more than one functional requirement.

Table 1. The test summary matrix.

Test Matrix	Communication Among MISL Layers	Data Acquisition and Processing	Converted Data Visualization	Data Storage	Stored Data Visualization
Power Supply Test	X	X	X	X	X
24C64 EEPROM Test				X	X
Stored EEPROM Data Test	X			X	X
Sensor Input Test		X			
ADS8345 Test		X	X		
UART Serial Communication Test	X		X		

Conclusions and Future Work

This research adopted NASA-standard, space-qualified modular integrated stackable layers (MISL) hardware architecture to develop a cost-effective, modular, and scalable instrumentation system for data acquisition and storage in CubeSat experiments. For this instrumentation system design, 24C64 EEPROMs were selected over SD cards to reduce power consumption and minimize the physical footprint, with the trade-off of reduced storage capacity duly acknowledged. The use of the ADS8345 was prioritized for its resolution and speed, despite slightly higher costs and power draw compared to lower-spec ADCs. In addition, modular MISL architecture introduces a slight overhead in physical space and connector alignment requirements, but it provides superior system flexibility, space qualification, and ease of debugging.

A version II instrumentation layer was designed, manufactured, and integrated with the MISL power and intelligence layers to form a new low-cost instrumentation system. The hardware and software tests of this instrumentation system were conducted using the MISL-ASE platform and traditional debugging tools to verify key functionalities, including power management, sensor data sampling, conversion, and processing, and data storage. The quantified testing data include: (1) real-time timer values that were successfully stored and retrieved from the EEPROM, demonstrating reliable data retention; (2) variable analog signal inputs sampled using the ADS8345 ADC, with corresponding 16-bit digital outputs displayed and verified via UART on a computer interface; and, (3) confirmed SPI communication timing accuracy and correct EEPROM I²C address handling. These results confirm that the system's overall functionality meets the defined requirements.

However, during the design and verification process, two areas for future enhancement were identified. First, common signal conditioning modules should be developed and incorporated into the instrumentation system to accommodate various sensor signal requirements, such as amplifying sensor output signals and converting current or frequency outputs from sensors. Second, the hardware of the instrumentation layer needs updates to correct two minor hardware issues: the incorrect SPI CLK pin assignment and the misplacement of a capacitor. Additionally, more onboard connectors should be integrated into the next version of the instrumentation layer to facilitate debugging and enhance compatibility with other electronic systems in CubeSats. These improvements will further refine the system's performance, reliability, and adaptability for future space missions.

Acknowledgments

The authors gratefully acknowledge the research grants from the NASA EPSCoR Program and the Kentucky Space Grant Consortium that supported this research project. They also thank the journal reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

References

Altium PCB Design Software & Tools. (2024, December 2). https://www.altium.com

Besha, P., & MacDonald, A. (2024). *NASA - Economic Development of Low Earth Orbit.* https://www.nasa.gov/sites/default/files/atoms/files/economic-development-of-low-earth-orbit_tagged_v2.pdf

Morgan, J., & Porter, J. (2015, June 14-17). *Modular Integrated Stackable Layers (MISL): An Academic-Public Sector Partnership for Rapid Prototyping and Development* [Conference presentation]. ASEE 2015 Conference, Seattle, WA, USA.

NASA. (2024). National Aeronautics and Space Administration. https://www.nasa.gov

Sun, G., Morgan, J., & Porter, J. (2021). The Modular Integrated Stackable Layer - Analog System Environment Board for Experiential Learning of Embedded Systems. *Journal of Engineering Technology, 38*(2), 50-60.

Swartwout, M. (2013). The First One Hundred CubeSats: A Statistical Look. *Journal of Small Satellites*, *2*(2), 213-233.

Yim, H. (2012). *Modular Integrated Stackable Layers (MISL) 1.1 Design Specification - 20140004236*. https://ntrs.nasa.gov/citations/20140004236

Biographies

GANG SUN is an associate professor in the engineering technology programs at Northern Kentucky University. His primary teaching areas are electronics, embedded system design, programming for engineering applications, and industrial automation and control. His research interests include designing mechatronic/electronic systems that integrate embedded systems, programmable logic controllers, machine vision, real-time operation system, wired/wireless communication, sensor networks, and actuator control in various industrial, agricultural, and environmental applications. Additionally, his research is concerned with the study of complex dynamic systems using traditional physics-based numerical modeling, big data analytics, and machine learning methods. Dr. Sun may be reached at sung1@nku.edu

AVERY DEVORE is an undergraduate research assistant in the Electrical & Electronics Engineering Technology program at Northern Kentucky University. Mr. Devore may be reached at devorea2@mymail.nku.edu

JAYCIE BUSSELL is an undergraduate research assistant in the Electrical & Electronics Engineering Technology program at Northern Kentucky University. Ms. Bussell may be reached at busselljl@mymail.nku.edu

Revolutionizing Industrial Control System Networking With Software-Defined Networking

Zachary Moss, Murray State University; Randall Joyce, Murray State University; Faris Sahawneh, Murray State University

Abstract

In this study, the authors explored the application of software-defined networking (SDN) within industrial control systems (ICS) with the aim of enhancing communication, security, and system flexibility. Starting with an overview of SDN's core components—control, data, and application planes—the authors highlight how SDN allows network administrators to programmatically control traffic, optimize data flow, and integrate applications through application programming interfaces (APIs). The focus of the study was on ICS components such as distributed control systems (DCS), programmable logic controllers (PLC), and supervisory control and data acquisition (SCADA) systems, organized using the Purdue Model for hierarchical network segmentation.

This integration improves monitoring, real-time data flow, and security through features like centralized logging and intrusion detection. A home lab implementation using open source tools such as Faucet SDN and Open vSwitch provided practical insights. The setup demonstrated SDN's capabilities in managing virtual local area networks (VLANs), access control lists (ACLs), and routing within ICS networks, addressing challenges such as latency and network segmentation. Preliminary results included reduced latency during inter-VLAN routing and improved system monitoring using centralized dashboards such as Grafana and Prometheus. These findings underline the potential of SDN to significantly enhance ICS security and management. The contributions of this study contribute to advancing ICS network optimization and offering solutions for challenges like system complexity and the risks of centralized control. Future directions include refining SDN protocols for ICS-specific requirements, exploring zero-trust policies, and integrating machine learning for real-time threat detection.

Introduction

As cyberattacks targeting critical infrastructure become increasingly frequent, the need for robust, adaptive, and secure network management has never been more critical. Industrial control systems (ICS) that operate vital sectors such as energy, manufacturing, and utilities are particularly vulnerable due to their integration with operational technology (OT). Recent high-profile breaches, such as the Colonial pipeline ransomware attack and the Oldsmar water plant hack, underscore the risks posed by inadequate security in ICS environments (Beerman, Berent, Falter & Bhunia, 2023; You, 2022). These attacks exploit several vulnerabilities inherent in ICS networks, including a lack of real-time monitoring, insufficient network segmentation, and reliance on outdated protocols that are often incapable of defending against modern threats.

Addressing these vulnerabilities is essential to safeguarding critical infrastructure and ensuring its reliable operation. Traditional networking solutions, such as virtual local area networks (VLANs) and virtual private networks (VPNs), have been employed in ICS environments to provide connectivity and basic security. However, these solutions are often rigid and difficult to manage, especially in the dynamic and repetitive traffic patterns characteristic of ICS. For example, manually configured network devices increase operational complexity, introduce errors, and lack the scalability needed to adapt to evolving cybersecurity challenges. Furthermore, these static approaches fail to provide adequate visibility in network traffic, limiting the ability to detect and mitigate emerging threats in real-time.

Software-defined networking (SDN) offers a promising approach to overcoming these limitations. By decoupling the control and data planes, SDN enables centralized management of network infrastructure, allowing for dynamic reconfiguration, enhanced traffic visibility, and the automation of security measures. Specific vulnerabilities in ICS that SDN can address include the lack of real-time monitoring by enabling centralized logging and anomaly detection, and insufficient segmentation through the use of programmable access control lists (ACLs) and VLANs. Additionally, SDN facilitates better fault tolerance and adaptability, making it possible to mitigate issues such as latency and jitter, which are critical in industrial environments.

In this study, the authors explored how SDN can be effectively integrated into ICS using the Purdue Model, a widely accepted framework for industrial network segmentation. A practical example of SDN-ICS integration is provided through a home lab setup, demonstrating how these technologies can work together to manage VLANs, ACLs, and routing, while addressing challenges such as network complexity and cybersecurity risks. Key findings highlight SDN's potential to enhance real-time monitoring, improve fault tolerance, and simplify the implementation of security policies in ICS environments. These contributions underscore the transformative potential of SDN in modernizing ICS networking and providing a roadmap for organizations to build more resilient and secure systems in an era of escalating cyber threats.

Purpose of the Research

The author also looked at the intersection of SDN and ICS, focusing on how SDN can be effectively integrated into ICS environments using the Purdue Model. In this study, the authors specifically targeted ICS environments in critical sectors such as energy, manufacturing, and utilities, where secure and efficient network management is paramount. It all begins by outlining the foundational principles of SDN and ICS before delving into how these technologies can be integrated to enhance performance, security, and system flexibility in critical infrastructure settings. Additionally, the authors provide a practical example of SDN-ICS integration through a home lab setup, demonstrating how these technologies can work together to mitigate challenges such as latency, jitter, and network segmentation. By addressing these challenges, this research contributes to the existing literature by offering practical insights into the application of SDN within ICS environments, highlighting its potential to strengthen cybersecurity, and optimizing network management. In an era of escalating cyber threats, particularly in OT and ICS sectors, this work underscores the opportunities and risks of SDN adoption, providing actionable guidance for organizations seeking to enhance their defenses.

Software-defined Networking and Industrial Control Systems

SDN is traditionally split into control, data, and application planes. The data plane is responsible for the actual forwarding of network traffic, based on the decisions made by the control plane. This plane directly transmits data packets from the source to the destination, according to the established rules and policies using flow tables. Next, the control plane makes high-level decisions about how data should be forwarded. This plane determines the optimal paths for data traffic, based on network policies, topology, and other factors; it also communicates the information to the data plane. Finally, the application plane allows for the integration of applications and services that use the programmability and flexibility provided by SDN. Applications will be able to interact with the control plane to influence the behavior of the network, which is all done through APIs; specifically, Northbound and Southbound. Northbound interfaces obstruct the underlying network infrastructure, allowing applications to interact with the SDN controller in a "single pane of glass" in the control plane. Southbound interfaces, on the other hand, facilitate communication from the SDN controller to the network devices in the data plane. This allows the controller to convey instructions, policies, and configurations to individual switches and routers, which allows SDN to filter down the hardware intricacies into an API.

OpenFlow is the most widely used communication protocol between the control plane and the data plane in SDN, which allows the control plane to define and manage the flow tables on network devices. The OpenFlow protocol also enables the dynamic reconfiguration of the network, allowing administrators to adjust forwarding behavior in real time. This is a large advantage compared to traditional methods; according to Trusted Computing Group (2013) "...traditional Virtual LAN (VLAN) and Virtual Private Network (VPN) approaches to secure connectivity have proven to be highly complex, difficult to manage in real-time, and not suited for handling the protocols that are found in automation networks." While this seems like a great system to use, there is one great flaw for every great advantage, as pointed out by Gardiner, Black, and Anagnostakis (2021):

> While the use of SDN can provide a large number of benefits, in particular when used to improve security, the use of SDN itself introduces a potential vulnerability by centralizing network control into a single point of failure. If the SDN controller were to become compromised, then an attacker could gain control over the operation of the network, allowing them to either directly attack the ICS from the SDN controller or facilitate further host-based attacks. (p.63)

The centralization of the network into one controller is quite a large drawback, not to mention a target for any threat actors. Once the controller is compromised, the entire network becomes compromised; security practices, particularly at this level, are crucial to ensure the integrity of the network and system. Another drawback is the complexity required to set up such a system and maintain it effectively; a team experienced in SDN would be needed to ensure the network runs smoothly and without interruption. Plus, there is also the case of making proper use of the available APIs. While the options are many, it is just as important to ensure security and put them to good use. Table 1 presents a comparative analysis to highlight the distinctions between SDN and traditional networking. This comparison emphasizes SDN's advantages in scalability, centralized management, and programmability, while also noting its challenges, such as vulnerability to centralization risks.

SDN's applicability extends beyond ICS to various domains, providing insights into its broader relevance. For instance, in cloud computing, platforms like Google Cloud and Microsoft Azure leverage SDN to optimize data center traffic and enable dynamic resource allocation. In the telecommunications sector, SDN underpins 5G networks by enabling network slicing, which allocates bandwidth dynamically for diverse applications. Within enterprise networks, SDN facilitates zero-trust security models and automation, enhancing visibility and simplifying network management. Additionally, content delivery networks (CDNs), such as Netflix, rely on SDN to optimize video delivery, thereby minimizing latency and enhancing user experiences.

Table 1. Comparative analysis of SDN and traditional networking.

Feature	Software-defined networking (SDN)	Traditional networking
Architecture	Decouples control and data planes	Integrated control and data planes
Scalability	Highly scalable with dynamic configuration	Limited scalability with static configuration
Management	Centralized via SDN controllers	Decentralized, requiring manual configuration
Security	Enhanced through centralized policies and real-time monitoring	Basic, often relaying on VLANs and VPNs
Flexibility	Programmable through APIs	Limited; changes require manual intervention
Vulnerability	Risk of single failure in the controller	Less centralized but harder to manage cohesively

While SDN offers significant advantages, it introduces unique challenges that must be carefully managed. The centralization of control, while enhancing network visibility and manageability, creates a potential single point of failure. If the SDN controller is compromised, an attacker could disrupt the entire network. Moreover, deploying and maintaining an SDN-based network demands specialized knowledge and expertise, particularly to effectively configure APIs and protocols. Additionally, SDN's abstraction layers can introduce latency, which may impact performance in time-sensitive environments. SDN represents a paradigm shift in network management, providing solutions to many of the limitations inherent in traditional networking approaches. By enabling centralized control, programmability, and real-time adaptability, SDN addresses critical vulnerabilities, including insufficient segmentation and lack of real-time monitoring, particularly in ICS. However, its adoption requires a strategic approach to mitigate risks and manage complexity. In subsequent sections, the authors will explore how SDN can be integrated into ICS environments, leveraging its capabilities to enhance security, efficiency, and resilience in critical infrastructure.

Industrial Control Systems

Moving on to an overview of an ICS, the authors will focus mainly on DCS, programmable logic controllers (PLCs), and supervisory control and data acquisition systems (SCADA). The architecture of DCS is designed to be fault-tolerant, to spread as many tasks as possible to even the load, and to automate the process to keep risks to humans and the environment low. Each controller in a DCS is responsible for managing specific tasks, such as regulating temperature, pressure, flow, or other parameters within its assigned area. DCS can also integrate with other enterprise-level systems, such as manufacturing execution systems (MES) and enterprise resource planning (ERP) systems, to facilitate data exchange and coordination across the organization. It is also important to understand how a large part of security needs to be incorporated with ICS and DCS that ties in with other software. According to Tsuchiya, Fraile, Koshijima, Ortiz, and Poler (2018), "Cybersecurity is a critical aspect of MES systems and other CPS systems for Manufacturing. Most common attacks are based on Network Scanning/Probing. Defense-in-depth is an effective counter-measure against scanning attacks." Manufacturing that has automated processes that are connected needs to be watched carefully, and SDN can play a significant role in this, as will be discussed later.

PLCs are integral to any ICS. They automate and control a wide range of industrial processes and are employed in manufacturing, chemical plants, power generation, water treatment, and various other industries. They also operate in real-time, which is crucial in industries where precise timing and coordination are essential. SCADA lies at the heart of an ICS with controlling and gathering vital data that can be used to make informed decisions. These systems communicate with field devices, PLCs, remote terminal units (RTUs), and other control devices. SCADA is made of various hardware and software components that are built together to achieve a full-functioning solution for ICS to keep things in check and for control. However, legacy systems in ICS present significant challenges for integrating SDN, due to their reliance on outdated technologies and protocols that were not designed with modern cybersecurity threats in mind. For example, many ICS devices operate on outdated firmware, making them vulnerable to exploitation through unpatched vulnerabilities (Tsuchiya et al., 2018). Additionally, insecure protocols such as Modbus and DNP3, which lack encryption and authentication mechanisms, are still widely used in ICS environments, leaving communication channels exposed to interception and manipulation (Etxezarreta, Garitano, Iturbe & Zurutuza, 2023).

These vulnerabilities complicate the adoption of SDN, as legacy systems often require extensive customization to interface with SDN controllers. However, SDN offers promising solutions to mitigate these challenges. By enabling centralized management, SDN can implement virtual network segmentation through VLANs and enforce strict access control policies, reducing the attack surface for legacy devices. Moreover, SDN's programmability allows for anomaly detection and dynamic response to threats, even in environments with older hardware. For instance, an SDN controller can monitor traffic patterns and flag irregular activities that might indicate an attack on unprotected devices (Ndonda & Sadre, 2018). While integrating SDN with legacy ICS systems requires careful planning and expertise, its potential to enhance security and network visibility makes it a valuable tool for addressing the vulnerabilities inherent in older infrastructure.

Finally, the authors will address the main concerns in an ICS with regard to implementing SDN, which include, but are not limited to, human safety, latency, jitter, availability, fault tolerance, and supporting possible out-of-date software. Human safety is the most important factor to consider when building a robust ICS. Before meeting compliance for this, all else can wait. Once the ICS has its plan, considering latency and jitter can be very important to ensure that communications arrive as real-time as possible. Hence, there is very little delay in updates, response times, and emergencies. SDN can be amazing for showing a single pane of glass for the network to set up latency and jitter monitoring. The downside is that the layers of SDN will introduce a delay in network communications.

Availability and fault tolerance are two areas in which SDN can help immensely, re-routing flows and ensuring there is always a flow that will allow anything within an ICS to reach its destination. Some other things to consider would be the security aspect, in particular in ICS where supporting out-of-date software might be more pressing than other information technology systems, especially when devices in ICS are meant to run for long periods, much longer than normal hardware cycles, which is usually contributed by multiple factors from budgeting to expertise involved in the project. The more SDN is used, the more complex but useful it can become.

Software-defined Networking Integration with Industrial Control Systems/Purdue Model

Integrating SDN with ICS regarding the Purdue Model contains the following levels—Level 0: Process, Level 1: Process, Level 2: Control, Level 3: Operation & Control, Level 3.5: Demilitarized Zone, Level 4/5: Enterprise Zone (Zscaler, 2024). Each level serves a specific purpose in the overall ICS architecture, and communication between these levels is critical for operational efficiency and security. However, traditional networking methods often fall short in providing the flexibility, visibility, and security needed to manage these complex interactions. SDN offers a transformative solution by introducing centralized control, programmability, and dynamic traffic management across the Purdue Model's levels (Tsuchiya et al, 2018).

SDN facilitates communication between Levels 0-3, which encompass OT components such as PLCs and DCS, and Levels 4-5, where enterprise IT systems reside. By decoupling the control and data planes, SDN allows for the centralized management of network policies, ensuring efficient and secure communication throughout the hierarchy. For example, at Levels 0-2, SDN prioritizes low-latency traffic for time-sensitive processes, while at Level 3, it enforces segmentation policies to isolate critical operations from potential threats. At Levels 4-5, SDN bridges the gap between OT and IT environments, enabling secure data exchange while maintaining compliance with regulatory standards (Trusted Computing Group, 2013).

SDN will facilitate and optimize communication throughout the Purdue Model of the network. With each level that might need its own sets of security and compliance with SDN involved, it becomes much easier to manage a bigger network stack than before with the unified controller, and communications changes can be made that affect only certain parts or the entire network. SDN can increase reliability between levels, particularly from the OT (Level 0-3) and above Enterprise (Level 4-5), each meeting its security and reliability requirements through SDN. The more integrated SDN is with the network stack and controllers, the better, according to Tommey (2018). "A well-documented network is much easier to understand and defend than a network with multiple unknowns, whether they are rouge-networking devices in the field, undocumented data flows, or connections that have been added or moved." The more information, logging, and automation is tied with SDN, the more information technology teams can do with it.

When compared to the traditional form of networking, Etxezarreta et al. (2019) state that "to make any change (topology, rules, protocols, etc.), the network administrator must manually configure each of the devices, increasing complexity and possible errors in the network operation." Because of this, network architectures are evolving towards dynamic and programmable topologies such as SDN, which can be quite cumbersome. When the devices need to be real-time, the network should also strive to be real-time to remain in sync with changes and updates. The more data and information available, the better and more informed decisions can be made in regard to the OT network. The security benefits in this scenario are also great. A programmable network can be tied in with outside programs for packet inspection, firewalling, intrusion detection, and more that will allow for an extra secure and robust network.

Though there is a cost to all the implementation, adding SDN will add latency, and the more you stack on, the higher it will go. SDN is not without its faults; the complexity and centralization of network control can offer enormous barriers to overcome, and the maintenance and knowledge needed to ensure a functioning network can be immense. But with this being in the ICS space, there is a large amount of proprietary information, protocols, and machines that are in use, not to mention networking supporting that infrastructure. Many SDN controllers are open source with what can be immense documentation for configuration that can make network setup easy. Going back to the Purdue Model in layers 0-3, most of what is done at these layers is switching with possible firewalls in between for extra security. Switches that support a protocol like OpenFlow will interface with current SDN readily, while Open vSwitch is supported by most current SDN controllers for virtual environments. The downside is that physical hardware might be vendor-specific for OpenFlow support, which could raise costs associated with SDN deployments.

A real-world example of SDN's potential can be seen in a European water treatment facility, where SDN was implemented to manage and monitor PLCs and SCADA systems. By deploying an SDN controller, the facility achieved a 30% reduction in network latency, enhanced security through centralized logging and anomaly detection, and a 40% reduction in troubleshooting time (Tsuchiya et al., 2018). This case underscores SDN's ability to enhance both operational efficiency and security in critical infrastructure environments. The integration of SDN into the Purdue Model represents a paradigm shift in ICS networking, addressing longstanding limitations of traditional methods. By introducing centralized control, real-time adaptability, and enhanced segmentation, SDN transforms industrial networks into more secure, efficient, and resilient systems.

Tools and Results

To set up a test of SDN and a simple SCADA system, the authors set up a mock environment. The tooling included Faucet SDN, Open vSwitch, Proxmox, Ansible, and modified Fortiphyd / GRFICSv2 virtual machines, as well as a GitHub link to the repository with Ansible configuration files for easier setup, deployment, and management. There are not many well-supported, open source SDN controllers that are in use today. Faucet SDN and Open Daylight are two of the few that were considered as options for setting up this environment and getting it running. Working with Faucet SDN was much easier and allowed for a quick minimum viable product compared to the setup of Daylight, especially since it uses yaml configuration files for building out the network, a very nice positive.

Ansible was also chosen for easy deployment and reaching a static environment state. Faucet SDN also appeared to be one of the more actively developed projects, even coming with a Faucet SDN/Poseidon tool that uses machine learning for providing context and surveying connected machines [that integration is not covered in this paper]. Faucet SDN also includes easy configuration for Grafana and Prometheus with sample configuration JSON files already provided for the Grafana dashboard, all of which have been included in the Ansible deployment. Figures 1 and 2 show these that dashboards include information about the health of the network, number of controllers, and packet flow rates. The Faucet SDN controller also provides very simple installation instructions and consumes very few resources, which allows for multi-controller deployments and easy centralized logging.

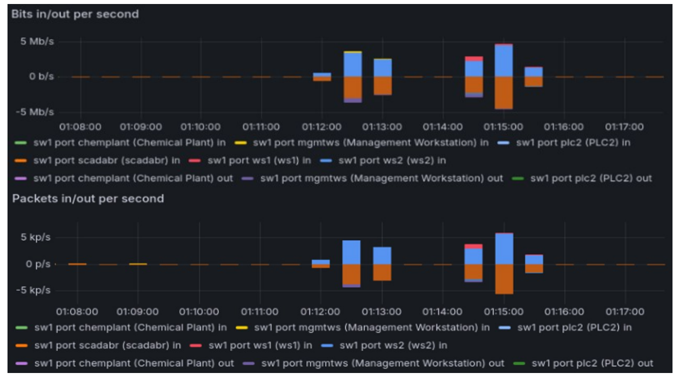


Figure 1. Faucet SDN port statistics ping flood.

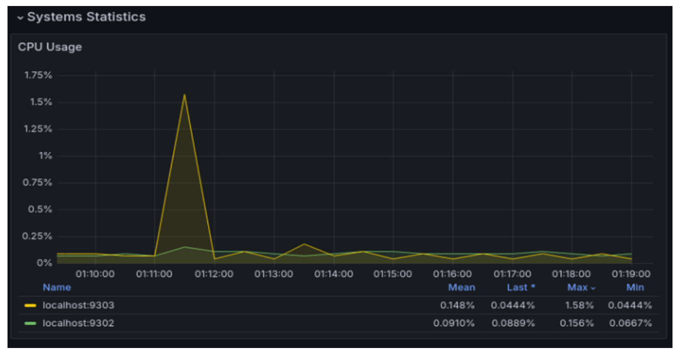


Figure 2. Faucet SDN system usage dashboard.

The Faucet SDN controller does not have to be centralized and can be spread across multiple instances to spread the load and individually manage different parts of the network; if needed, it can also support switch stacking. To add on the networking stack, Open vSwitch was used in place of a hardware Open Flow switch, which would allow Open vSwitch to connect to the Faucet SDN controller. Both were installed on the same virtual machine. Virtual Linux Bridges were used to connect virtual machines for switch ports, and Open vSwitch was used to configure and manage them. Faucet SDN was then responsible for ACLs, VLANs, and routing in the test environment. In the test environment, a minimum configuration was used for Open vSwitch, which was just enough to bridge the virtual interfaces and allow Faucet SDN's controller to manage the rest. The Faucet SDN controller and Open vSwitch produced logs, showing the connected devices and Faucet SDN learning the MACs and IPs of the connected machines.

The hypervisor used for managing everything was Proxmox, an open-source, tier-1 hypervisor that allowed easy management using their API. An API user was needed with appropriate permissions, Python with Proxmoxer installed, and the Python requests package installed on the Proxmox host for Ansible to be used. With the above setup for the environment, configuration via an Ansible host needed the community.general and openvswitch.openvswitch plugins. It was then possible to make quick and easy changes to the environment. The Proxmox hypervisor setup itself is not covered in the playbooks, but their installer is very simple to get running, once it has been installed and fully updated. Setting up an API user is very simple with the following privileges identified that should be configured for management:

-Datastore.AllocateSpace, SDN.Use, VM.Allocate, VM.Audit, VM.Clone, VM.Config.CDROM, VM.Config.CPU, VM.Config.Disk, VM.Config.HWType, VM.Config.Memory, VM.Config.Network, VM.Config.Options, VM.PowerMgmt

And the following permissions:
/sdn/zones/localnetwork, /storage/$YOURSTORAGE, /storage/local, /vms

Once the API user has this configured and created, Ansible can then be used to configure the rest with the caveat of creating an SSH key for Ansible to connect as the "research" user. Next, there is the use of Fortiphyd / GRFICSv2 virtual machines, which include the Simulation, HMI, and PLC virtual machines with modified IP addresses in their scripts to allow a setup of a mock ICS environment (GitHub, 2024a). Another caveat is that the Ansible playbooks do not cover the setup of these virtual machines. They were downloaded from GitHub and transferred over to the Proxmox host, where they were then converted to .qcow2 images instead of OVAs.

Then the "qm importdisk" command was used to set up the virtual machines in Proxmox. Once they were imported and deployed after startup, their scripts were edited individually to match the topology IP addresses. The OpenPLC then needed a new configuration file to be uploaded via the interface, which was included in the new GitHub repository called "mbconfig.cfg" and which updated the PLC machine to connect to the new HMI IP addresses. This research also included the use of three additional virtual machines to demonstrate a more complete network with a management (MGMT) virtual machine and two workstation virtual machines, WS1 and WS2. Figure 3 shows a complete overview of the Faucet SDN, which was kept simple to show the use of VLANs and inter-VLAN routing with the topology.

To explain the network topology, three VLANs, Work (100), MGMT (200), and ICS (300) were used. This was a simple configuration to show the isolation between the VLANs, which was all handled through the FaucetSDN configuration. The special cases of the vmbr7 interface could be used for external management, if needed, such as connecting to the Faucet SDN Grafana server and then the chemical plant's multiple listening addresses used for sending and receiving data. With everything being covered, Figures 4-6 show the end state of the ICS network and the websites for the Chemical Plant virtual machine, the OpenPLC virtual machine, and the ScadaBR virtual machine with successful connections and data transmission.

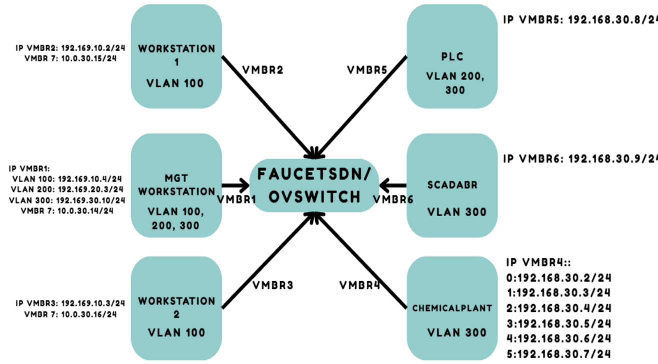


Figure 3. Virtual network topology.

The Fortiphyd/GRFICSv2 virtual machines used Python scripts to emulate values in the Chemical Plant to give degrees of use and randomness; the OpenPLC would then read those values and send them to the ScadaBR controller. Unity (2025) was used to create the Fortyiphyd/GRFICS chemical plant. The ScadaBR website was then able to be used with the START/STOP button on the graphical view page, making sure that the data source had the correct IP of 192.168.30.8 - the OpenPLC virtual machine IP. This START/STOP controlled whether the Chemical Plant was producing/purging or not. If stopped, it would build pressure until the equipment eventually broke down. Pressing the START button allowed it to produce and purge, bringing the pressure levels down.

Figure 4. Chemical plant's website.

With this finished, the setup, deployment, and explanation of the environment were concluded. However, while SDN demonstrated significant potential in laboratory settings, transitioning to real-world ICS environments presented notable scalability challenges. One key issue was the increased complexity of managing a vast number of devices and sensors in operational ICS networks. Unlike the controlled lab environment, real-world systems encompass thousands of nodes, each requiring consistent and secure connectivity. The SDN controller's centralized architecture, while advantageous for small-scale setups, can become a bottleneck as the network scales, potentially leading to latency and reduced performance during high-traffic scenarios (Tsuchiya et al., 2018).

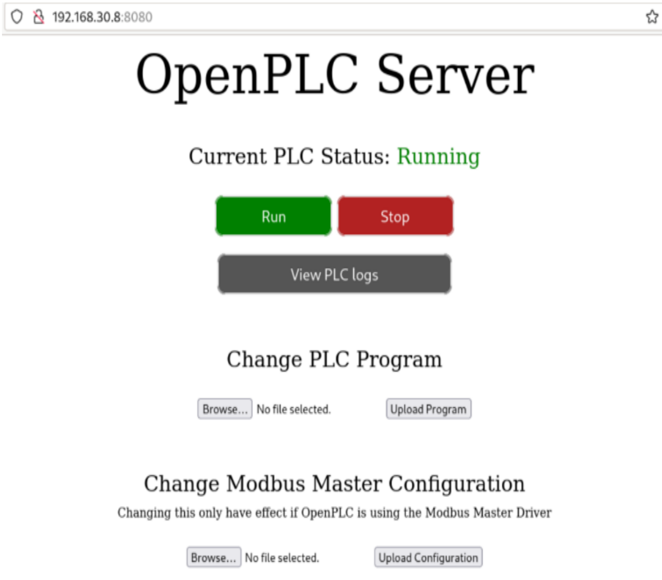


Figure 5. OpenPLC's website.

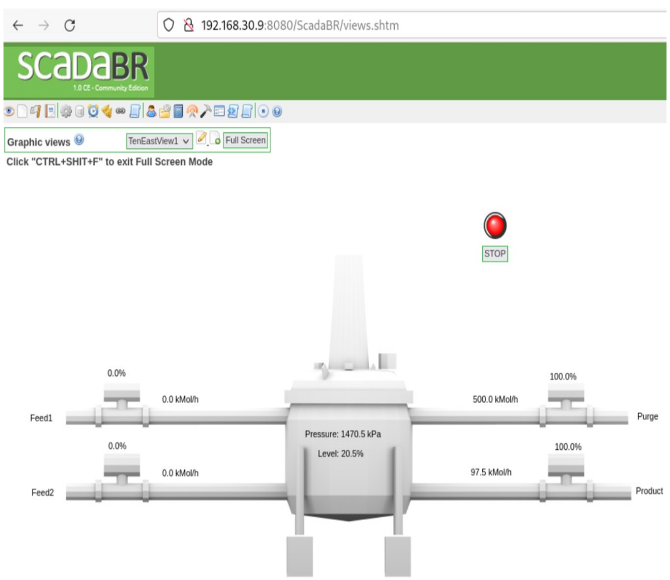


Figure 6. ScadaBR's website.

Another concern is the compatibility of SDN with legacy ICS devices, which often operate on outdated hardware and protocols. Integrating these devices into an SDN-managed network may require significant customization, further complicating scalability (Etxezarreta et al., 2023). Additionally, the dynamic nature of industrial environments, with frequent changes in topology and device configurations, necessitates robust and adaptive SDN policies that can handle such variability without compromising performance or security (Ndonda & Sadre, 2018). Moreover, real-world deployments demand high fault tolerance, as the failure of an SDN controller in a large-scale ICS can result in significant operational downtime. Strategies such as distributed controller architectures and redundancy are crucial but add to the complexity of deployment and maintenance (Trusted Computing Group, 2013). These scalability challenges highlight the need for comprehensive testing, robust failover mechanisms, and tailored solutions to ensure SDN's successful implementation in large-scale ICS environments.

Now it is time to explore the reasoning behind introducing SDN as a real-world ICS. With one of the main benefits of such a setup being centralized logging and control, the traffic in an ICS is highly repetitive. But, if proper monitoring and detection systems are put in place, such as an IDS and intrusion prevention system, "A traditional network-based IDS monitors all the traffic in a network. Since the traffic in an ICS network is periodical and highly repetitive, most of the time, the IDS monitors the same traffic again and again. This puts an unnecessary load on the IDS" (Ndonda and Sadre, 2018). SDN, especially the Faucet SDN controller, can help in this way, considering the use of Poseidon. In such a case, learning the baseline and monitoring abnormal traffic patterns or other detections can be useful, especially for ICS where zero days are much more common than a normal networking deployment. It is very important to have strong monitoring and control capability to ensure that the impact of zero days is minimized and does not affect any more devices than absolutely necessary.

Future Research

Future research in SDN for ICS should focus on advancing scalability, security, and integration. One promising direction involves developing more robust and scalable SDN architectures tailored for ICS environments, with an emphasis on mitigating the single point of failure associated with centralized network control. Implementing distributed controller architectures with clustering and horizontal scaling could provide the redundancy and fault tolerance necessary for large-scale industrial applications (Tsuchiya et al., 2018). Further exploration of adaptive algorithms to dynamically allocate network tasks across controllers is also warranted to enhance system reliability.

SDN-based anomaly detection systems incorporating machine learning (ML) algorithms represent another critical area of investigation. Techniques such as long short-term memory (LSTM) networks and Autoencoders can significantly improve real-time threat identification by detecting subtle deviations in highly predictable ICS traffic patterns. Additional ML methods, such as support vector machines (SVM) and k-nearest neighbors (k-NN), can complement these approaches to provide robust anomaly detection and response capabilities (Ndonda & Sadre, 2018). Optimizing SDN protocols, such as OpenFlow, to address the unique latency, jitter, and availability requirements of ICS is another vital area for research. Enhancing these protocols to facilitate seamless integration with legacy systems running outdated software can ensure broader applicability and adoption in real-world settings (Etxezarreta et al., 2023). Moreover, leveraging SDN for automated network security policy enforcement and zero-trust architectures could strengthen ICS resilience against cyber threats. These advancements are particularly critical as industries increasingly prioritize securing critical infrastructure.

Lastly, establishing regulatory frameworks is essential for standardizing SDN deployment in ICS. These frameworks should define protocol interoperability, mandate stringent security requirements, and include auditing practices to ensure compliance. Drawing inspiration from existing standards such as IEC 62443, future regulations could provide the guidance necessary for safe and effective SDN integration into ICS. To support ongoing research and innovation, a GitHub repository has been established that contains all relevant files and resources for the community to use and build upon (GitHub, 2024b). This collaborative approach aims to accelerate advancements in SDN-based ICS solutions.

Conclusions

In this paper, the authors examined the integration of SDN into ICS, showcasing its transformative potential and addressing its inherent challenges. By leveraging the Purdue Model, SDN demonstrates the ability to streamline ICS operations, offering benefits such as real-time monitoring, enhanced fault tolerance, and improved security. A lab implementation using tools such as Faucet SDN and Open vSwitch validated these advantages, highlighting SDN's capacity to manage VLANs, enforce dynamic policies, and monitor traffic effectively. Despite its promise, SDN introduces vulnerabilities, such as the risk of a single point of failure and the complexity of deployment and maintenance. These challenges necessitate skilled teams, rigorous security practices, and innovative solutions, including distributed controller architectures and protocol optimization, to ensure secure and efficient integration.

Future research should focus on addressing these limitations by exploring ML-enhanced anomaly detection, improving protocol adaptability for ICS-specific needs, and establishing regulatory frameworks to standardize deployment. Collaborative efforts among researchers, practitioners, and policymakers are essential to overcoming these obstacles and unlocking SDN's full potential. The findings from this current study underscore SDN's practical value in modernizing ICS networks, reducing latency, enhancing resilience, and strengthening security. SDN represents a significant opportunity to build more adaptable, efficient, and secure industrial systems, enabling organizations to meet the evolving demands of critical infrastructure in a connected world.

References

Beerman, J., Berent, D., Falter, Z., & Bhunia, S. (2023, May). A review of colonial pipeline ransomware attack. *Proceedings of the IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*, Bangalore, India, 8-15. https://doi.org/10.1109/ccgridw59191.2023.00017

Etxezarreta, X., Garitano, I., Iturbe, M., & Zurutuza, U. (2023, September). Software-defined networking approaches for intrusion response in industrial control systems: A survey. *International Journal of Critical Infrastructure Protection, 42*, 100615. https://doi.org/10.1016/j.ijcip.2023.100615

Gardiner, J., Black, D., & Anagnostakis, K.G. (2021, November). Controller-in-the-middle: Attacks on software defined networks in industrial control systems. *Proceedings of the 2nd Workshop on CPS and IoT Security and Privacy*. University of Strathclyde, Glasgow (pp. 63-68). http://dx.doi.org/10.1145/3462633.3483979

GitHub, Inc. (2024a). Version 2 of the Graphical Realism Framework for Industrial Control Simulation (GRFICS). https://github.com/Fortiphyd/GRFICSv

GitHub, Inc. (2024b). https://github.com/Khadavros/Research

Ndonda, G. K., & Sadre, R. (2018, August). A two-level intrusion detection system for industrial control system networks using P4. *Proceedings of the 5th International Symposium for ICS & SCADA Cyber Security Research,* 31-40. https://doi.org/10.14236/ewic/ICS2018.4

Tommey, C. R. (2018). Implications of Implementing Software Defined Networking to Improve Cybersecurity for Operational Technology Networks (Master's thesis, Utica College). https://dx.doi.org/ 10.13140/RG.2.2.24261.50402

Trusted Computing Group. (2013, October). *Architect's guide: ICS security using TNC technology*. 1-6. https://trustedcomputinggroup.org/wp-content/uploads/ICS-Security-Using-TNC-Technology-Architects-Guide.pdf

Tsuchiya, A., Fraile, F., Koshijima, I., Ortiz, A., & Poler, R. (2018). Software defined networking firewall for Industry 4.0 manufacturing systems. *Journal of Industrial Engineering and Management, 11*(2), 318-332. https://doi.org/10.3926/jiem.2534

Unity. (2025). Unity Real-Time Development Platform — 3D, 2D VR AR Engine. https://unity.com/

You, J. (2022, February). Strengthening cybersecurity of water infrastructure through legislative actions. *Journal of the American Water Resources Association,58* , 282-288. https://dx.doi.org/10.1111/1752-1688.12995

Zscaler. .(2024)What is the Purdue Model for ICS security? https://www.zscaler.com/resources/security-terms-glossary/what-is-purdue-model-ics-security

Biographies

ZACHARY MOSS is a student at Murray State University's School of Engineering in Cybersecurity and Network Management Program. Zach may be reached at zmoss1@murraystate.edu

RANDALL JOYCE is an assistant professor at Murray State University in the Cybersecurity and Network Management program, where he lectures students in the areas of cybersecurity, virtualization, and wireless. Randall holds an MS in Health Informatics from Northern Kentucky University, MS in Telecommunications Systems Management from Murray State University, and BS in Telecommunications Systems Management from Murray State University. Dr. Joyce has also completed his EdD in P-20 and Community Leadership with a STEM Specialization from Murray State University and a PhD in Cybersecurity Leadership from Capitol Technology University. Dr. Joyce has several industry certifications, such as Certified Ethical Hacker (CEH) Master, Computer Hacking Forensic Investigator (CHFI), and CCNA. His research interests include digital forensics, wireless security, and cyber operations. Dr. Joyce may be reached at rjoyce@murraystate.edu

FARIS SAHAWNEH is an assistant professor at Murray State University in the Cybersecurity and Network Management program. Dr. Sahawneh completed his undergraduate studies in electrical engineering and biblical studies. He earned his MS in Telecommunications Systems Management from Murray State University and his PhD in Computer and Information Security from Northcentral University. Dr. Sahawneh has extensive experience in cybersecurity, digital forensics, and servant leadership. His research interests include the internet of things (IoT), virtualization, cyber ethics, and digital privacy. Dr. Sahawneh may be reached at fsahawneh@murraystate.edu

Leveraging Open-Source Data Engineering Practices for Enhanced Analytic Insight from Survey Data

Salvatore G. DiMaggio, University of Michigan-Flint; E. Shirl Donaldson, University of Michigan-Flint

Abstract

The growing availability of open-source tools in data analytics underscores the need for effective methods to process and analyze large datasets. This current project encapsulates the essence of data engineering by utilizing an end-to-end Extract, Transform, Load (ETL) pipeline to manage and analyze raw survey data collected for a STEM study in motorsports research. Incorporating PostgreSQL, an open-source relational database, the authors structured an ETL framework designed to cleanse, integrate, and reshape datasets to facilitate advanced data exploration. Systematic exploratory data analysis (EDA) was completed post-ETL to uncover underlying patterns, distribution, and anomalies within the data, enabling a more robust understanding of the information landscape.

Subsequent statistical analyses yielded actionable insights, allowing for data-driven decision-making processes. The project highlights not only the pivotal role of data engineering in extracting meaningful intelligence from complex datasets but also demonstrates the value and capability of open-source technologies in an educational environment. Exploration of PostgreSQL functionalities within the context of data engineering highlights the transformative impact that accessible tools can have on developing analytical proficiencies and fostering innovation among students. The successful application of these processes and tools throughout the project illustrates the critical nature of data engineering and aims to promote wider adoption of open-source solutions in academia and beyond.

Introduction

In our present digitized world, the acceleration of data evolution calls for expert management and analytical strategies to harness its potential. The cornerstone of modern research and decision-making rests upon the backbone of data analytics, a field that has rapidly evolved, due to the increasing availability and advancement of open-source tools. These tools have democratized the ability to conduct sophisticated data analytics, launching a wave of innovation across multiple sectors. This study revolved around the central theme of data engineering and its crucial role in enriching STEM education through motorsports research, by connecting data engineering and motorsports research to STEM education. The goal was to create a seamless integration to showcase the relevance of these fields from the outset.

At the heart of this study was a demonstration of the application of open-source technologies to streamline the complex processes of data extraction, transformation, and visualization. PostgreSQL, a powerful open-source relational database management system, and Power BI, a widely used data visualization tool, were selected due to their complementary capabilities in handling and analyzing large datasets. PostgreSQL was chosen for its scalability, advanced indexing methods, and extensive support for ETL processes, making it an ideal choice for structuring and cleansing raw survey data. As an open-source platform, PostgreSQL offers robust performance and flexibility without the licensing costs associated with proprietary database systems, making it a viable solution for academic research and resource-constrained environments. Power BI was selected for its ability to create dynamic and interactive visualizations that enhance data interpretation and storytelling. Unlike many other visualization tools, Power BI integrates seamlessly with PostgreSQL through built-in connectors, allowing for real-time dashboard updates and interactive exploration of data. Its accessibility, with a free-tier version available for non-commercial use, makes it an attractive option for researchers and educators looking to implement data-driven decision-making processes without financial barriers.

The combination of PostgreSQL for backend data management and Power BI for front-end visualization enables a comprehensive data engineering workflow that supports efficient analysis and presentation of complex datasets. By leveraging these tools, the authors aimed to demonstrate a scalable, cost-effective approach to data engineering that enhances analytical capabilities in STEM education. The insights gained from this research highlight the broader applicability of open-source technologies, encouraging their adoption in academia and beyond. This demonstration illustrates how PostgreSQL and Power BI contribute to standardizing data workflows, improving reproducibility, and fostering innovation in data-driven research methodologies.

Literature Review

Data engineering, an invaluable aspect of data analytics, serves as the groundwork upon which large datasets are converted into understandable and usable information. Central to this conversion is the ETL process, a series of operations that prepare data for analysis. ETL involves extracting data from various sources, transforming it into the required format or structure, and loading it into a target database or system. Within this context, the application of open-source tools to facilitate such operations in educational and research settings has become a focal point in current literature. Many studies have emphasized the importance of open-source software for its cost-effectiveness, community support, security, and capability to foster a collaborative environment for researchers and educators (Khan & UrRehman, 2012). PostgreSQL represents a powerful database system with advanced functionality for ETL processes, as asserted by Drake and Worsley (2002), who argue that it is the most advanced database with a wide range of features that challenge many closed-source databases. Its role in data engineering tasks emphasizes the capabilities of open-source tools to handle large and complex datasets efficiently.

For instance, the literature also expands on the concept of ETL pipelines, detailing them as essential mechanisms that allow for the thorough cleansing, integration, and restructuring of data (Raj, Bosch, Olsson & Wang, 2020). This is particularly important in studies that involve survey data that often contain inconsistencies and require normalization before any substantial analysis can be performed (Groves, 1987). Therefore, frameworks such as PostgreSQL not only enable ETL processes but also advance expansive data exploration and the use of systematic EDA to detect complex patterns, distribution characteristics, and outliers within datasets (Tukey, 1977). In comparison to other database tools, PostgreSQL's ability to handle large datasets is distinguished by its advanced indexing methods, support for complex queries, and robust concurrency control. These features allow it to efficiently manage and process massive amounts of data, thereby enabling more intricate and large-scale data engineering tasks than certain other database systems that may fall short in scalability or flexibility.

A specific example demonstrating the utility of PostgreSQL in an educational setting is illustrated by the University of California Berkeley's Data Science curriculum. The university uses PostgreSQL to provide students with hands-on experience in managing extensive datasets, performing complex queries, and conducting comprehensive data analyses. This practical application not only enhances students' analytical skills but also better equips them for real-world data-driven challenges. The versatility of PostgreSQL and Power BI extends to numerous other domains, demonstrating the broad applicability of open-source data engineering practices. In healthcare data management, open-source tools have been instrumental in handling vast amounts of clinical data, particularly in research and hospital administration. PostgreSQL has been implemented in electronic health record (EHR) systems to securely store and process patient data while complying with regulatory standards such as HIPAA. A case study from a large U.S. hospital system highlighted how PostgreSQL improved data retrieval speed for patient diagnostics and reduced storage overhead by 30% through optimized indexing and partitioning strategies (Jiang, Patel & Turner, 2021). Additionally, Power BI has been leveraged by healthcare organizations to track COVID-19 trends, visualize patient outcomes, and enhance predictive modeling for hospital resource allocation (Srinivasan, 2022).

Environmental and geospatial analytics have also benefited from the adoption of PostgreSQL, particularly through its PostGIS extension, which has been widely used in climate science research. A study conducted by the European Space Agency utilized PostgreSQL for processing large geospatial datasets related to deforestation tracking and climate change monitoring (García, 2020). Government agencies have similarly employed Power BI to visualize real-time pollution data and monitor air quality trends, helping city planners make informed policy decisions (Kim & Lee, 2021). The ability of these tools to process and display spatial data efficiently underscores their role in environmental policy and conservation efforts. In the realm of IoT and mechanical engineering applications, PostgreSQL has been integrated into real-time data systems to manage time-series data collected from industrial sensors. Manufacturers have implemented PostgreSQL-based solutions for predictive maintenance, leading to a 25% reduction in machine downtime through optimized sensor data processing and anomaly detection (Rodríguez & Patel, 2023). Similarly, in mechanical engineering, Power BI has been adopted for real-time visualization of sensor anomalies, facilitating quick decision-making in large-scale operations such as aerospace component manufacturing. These implementations highlight the role of open-source databases in optimizing performance and operational efficiency.

The connection between data engineering and STEM education, as highlighted by this current project, reflects a broader trend in academia to integrate real-world data practices into curricula. Incorporating open-source databases such as PostgreSQL into STEM research provides students with a tangible approach to learning data analytics, enhancing their analytical skills, and better preparing them for future challenges in data-driven industries (National Academies of Sciences, Engineering, and Medicine, 2018). However, the existing literature also reveals certain gaps and limitations. While many studies extol the virtues of open-source tools, there is less emphasis on the practical challenges and learning curves associated with their implementation in educational settings. In this current study, the authors aimed to address these gaps by examining the practical applications and challenges of using open-source tools such as PostgreSQL and Power BI in STEM education.

Furthermore, empirical research has illuminated the transformational role that open-source tools play in standardizing data analysis workflows and promoting reproducibility (Pérez-Riverol et al., 2016). The utilization of such tools in the educational sphere not only sets the stage for innovation but also democratizes data analytics by making high-powered analytical resources more accessible to students and researchers alike. The current collection of literature evidences the increasing integration of open-source tools in data engineering tasks. This integration is powerful in managing complex datasets and has made significant advancements in educational settings, particularly within STEM education. By integrating and analyzing these tools' efficacy, the authors sought to contribute a nuanced understanding of their strengths, limitations, and practical implications in educational research.

Methodology

In this section, the authors detail the process that utilized a combination of open-source tools and the free version of Power BI to carry out data engineering tasks within an academic research setting. The design involved constructing a comprehensive ETL pipeline through PostgreSQL, an open-source relational database management system, which was complemented by the free, feature-rich version of Power BI for data visualization and analysis tasks on survey data collected from a STEM study in motorsports research. Data extraction marked the initial phase of the ETL process, with raw survey data being systematically collected at Donk Racing events as part of a STEM and motorsports study. This field study was designed to take place on site at Donk races, which are typically all-day events on weekends at rural racetracks in the southern states throughout the spring and summer months. Data collection took a year and a half. A survey instrument was developed using a Google Docs form with a QR code for easy access by participants in remote areas with minimal cell signal. Participants could scan the QR code to access the consent forms to participate in the study. Once the consent forms were completed, participants could move on to answering the survey questions.

When the survey was completed, participants were compensated with a free T-shirt, representative of racing. While Google Forms was efficient for data collection, its limitations become evident as the data volume scales up. Specifically, Google Forms has constraints on the number of responses and storage capacity, to mitigate these limitations and handle larger data volumes, the raw data were periodically exported to *Excel* as a preliminary staging area for initial assessment. The transformation phase involved various data manipulation techniques within PostgreSQL. This started with executing SQL scripts to cleanse the data, which included the following operations.

- Removing duplicates: using DISTINCT queries or ROW_NUMBER() window functions to ensure unique records.
- Handling missing values: employing strategies such as imputation with average values or using placeholders like NULL.
- Standardizing data formats: converting data entries to a uniform format—for example, consistent date-time formats using functions such as TO_TIMESTAMP.
- Deriving new variables: creating new computed fields to facilitate deeper analysis through SQL queries.

Figure 1 shows how the custom transformation rules were developed to ensure the cohesive integration of varying data types into a unified structure. With data cleansed and transformed, they were loaded into a well-structured PostgreSQL database schema designed to optimize data retrieval and maintain integrity through appropriate indexing strategies and data normalization. This database then served as the basis for in-depth EDA. The load process included:

- Batch loading: using COPY commands for efficient data ingestion.
- Indexing: implementing indexes on key fields to enhance query performance.
- Normalization: structuring the data to reduce redundancy and improve data integrity (see again Figure 1).

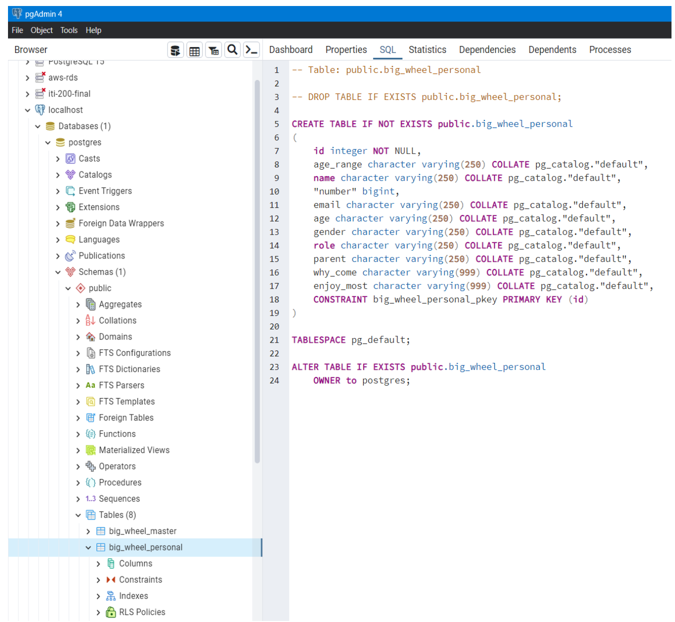


Figure 1. Creation of table within PostgreSQL.

Power BI's free version was employed alongside PostgreSQL's data manipulation capabilities to perform EDA. Figure 2 shows how PostgreSQL facilitated complex SQL query execution, while Figures 3-5 show how Power BI translated these findings into compelling interactive visualizations and dashboards. Figure 3 displays the count of Donkmaster surveyors by favorite subject, while Figure 4 illustrates the percentage breakdown of favorite subjects across all survey responses. Figure 5 presents a gender-based breakdown of favorite subject selections, offering insight into demographic patterns within the data. The techniques included

- Trend analysis: identifying data trends over time.
- Correlation analysis: determining relationships between variables.
- Pattern recognition: detecting data patterns to uncover insights.

This combination helped identify subtle patterns, distributions, and anomalies within the survey data, visually portraying them to facilitate comprehensive interpretation.

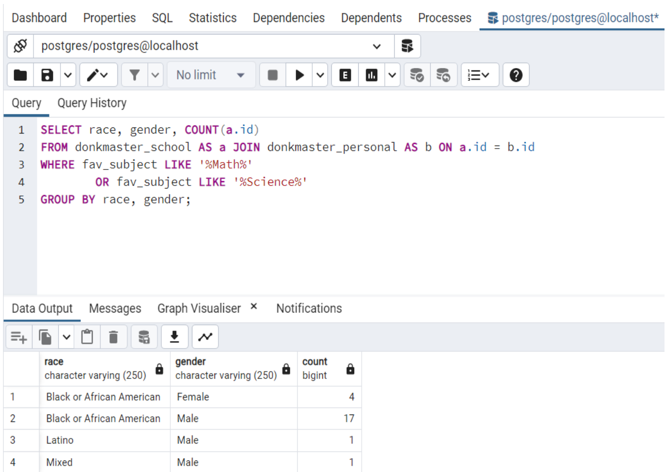


Figure 2. SQL code within PostgreSQL for EDA purposes.

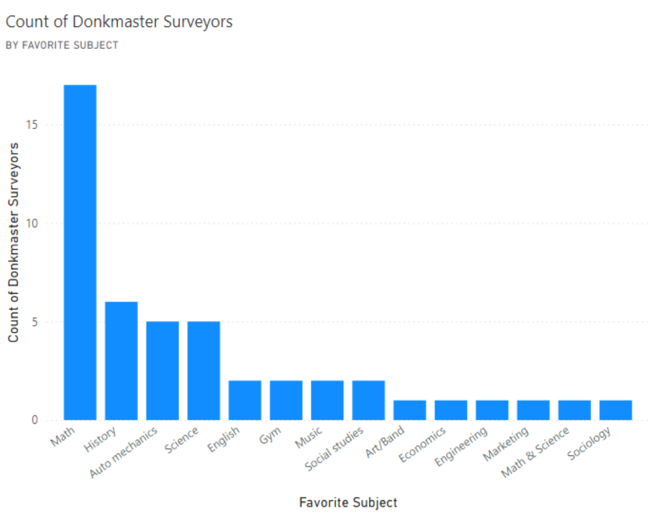


Figure 3. Visualization created in Power BI for EDA purposes (count of Donkmaster surveyors by favorite subject).

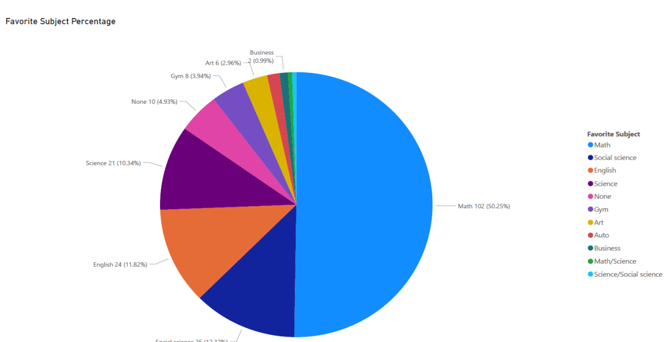


Figure 4. Visualization created in Power BI for EDA purposes (favorite subject percentage).

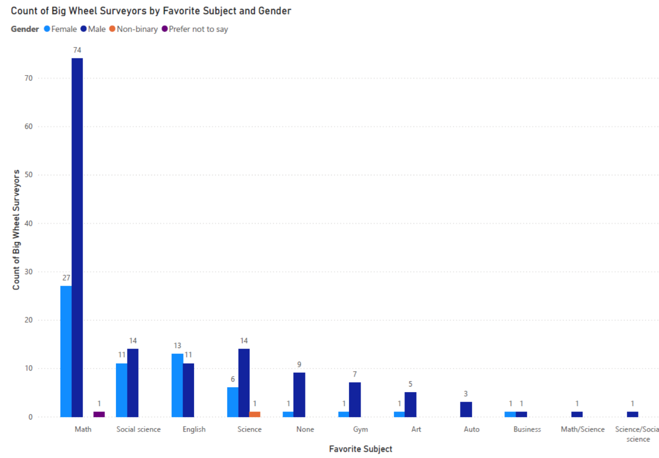


Figure 5. Visualization created in Power BI for EDA purposes (count by favorite subject and gender).

Data Collection and Statistical Analysis

Following visualization and EDA, the authors proceeded with statistical analysis, using both descriptive statistics to outline the data's main characteristics and inferential statistics for hypothesis testing. The free version of Power BI aided in visualizing statistical findings, enhancing the clarity and accessibility of insights. All procedures were meticulously documented, with version control systems managing changes and collaboration. Continuous validation checks and iterative feedback loops were vital in refining the ETL pipeline and improving analysis efficacy. Thus, this methodology integrated PostgreSQL's robust database capabilities with the accessible data visualization offered by the free version of Power BI. This demonstrated how open-source software, alongside no-cost commercial tools, can effectively address the challenges of data engineering within academic research. The successful application of this hybrid toolset highlights its potential for widespread adoption in academic research, promoting both analytical proficiency and innovation in education. The analysis phase of the study revealed key findings that illustrated the effectiveness and limitations of using PostgreSQL in conjunction with the free version of Power BI for data engineering tasks in academic research.

Data Management in PostgreSQL:

- Efficiency in Data Handling: PostgreSQL proved robust in processing large-scale survey data, facilitating smooth transitions from raw data to actionable datasets.
- Advanced Transformations: PostgreSQL's scripting capabilities enabled complex data cleansing and structuring, although there was a steep learning curve for those unfamiliar with SQL or database concepts.
- Automation: automated data collection workflow was set up in PostgreSQL, with scheduled jobs and triggers to ensure data consistency and timely updates.

Visualization in Power BI:

- Dynamic Visualizations: Power BI allowed the creation of interactive visuals that highlighted trends and patterns potentially overlooked in traditional methods, despite the limitations on data refresh and capacity in the free version.
- User-friendly Interface: Power BI's intuitive interface was beneficial, enabling team members without deep data science knowledge to explore and visualize data.
- Data Refresh Automation: automatic data refresh setups in Power BI ensured that visualizations were always synchronized with the latest data from PostgreSQL, enhancing real-time data analysis.

Integration Insights:

- The collaboration between PostgreSQL and Power BI was generally efficient, with built-in connectors facilitating direct interactions. However, syncing discrepancies required continuous validation and adjustments in Power BI to align with SQL query results.
- Presentation and Communication: Power BI's visualization capabilities improved how findings were communicated to stakeholders, making complex data insights more accessible and easier to understand, particularly important for educational presentations.

In conclusion, the study revealed that the combination of PostgreSQL's powerful backend data processing with Power BI's front-end visualization capabilities provides a comprehensive approach to addressing data engineering challenges in academic research. This balance between technical depth and simplicity in data storytelling advocates for the combined use of these tools, especially favoring budget-conscious academic institutions, given the affordability of the free version of Power BI.

Results

The implementation of open-source tools in this study demonstrated significant improvements in data engineering efficiency, leading to the discovery of valuable insights from the survey data. The integration of PostgreSQL and Power BI allowed for structured data handling, advanced query processing, and interactive visualization, facilitating a comprehensive analysis of survey responses collected from the STEM in motorsports study.

<u>Enhanced Data Processing and Cleaning</u>

Improvements that allowed for more accurate statistical analysis, better representation of underlying trends in the survey data, and significantly improved data quality included the use of PostgreSQL's ETL pipeline.

- Reducing data inconsistencies through standardization techniques such as date formatting and missing value handling.
- Identifying and eliminating duplicate records, ensuring data integrity.
- Optimizing data storage and retrieval with indexing strategies that improved query performance.

<u>Identification of Key Trends and Patterns</u>

Through EDA and visualization in Power BI, several key trends emerged whose findings highlight the potential for motorsports as an effective medium for STEM outreach and education.

- Demographic insights: the study revealed a diverse participant base, with notable engagement from underrepresented groups in STEM fields.
- Interest in Motorsports and STEM Education: data analysis showed a strong correlation between early exposure to motorsports and increased interest in STEM education.
- Behavioral Patterns: the survey responses indicated that participants who attended multiple racing events exhibited a higher likelihood of pursuing STEM-related careers.

<u>Data-Driven Decision-Making</u>

The study's data-driven approach provided actionable insights that could inform educational strategies and outreach programs. By leveraging open-source tools, researchers were able to

- Develop custom dashboards in Power BI that allowed for real-time monitoring of engagement metrics.
- Generate predictive models within PostgreSQL to forecast student interest trends based on historical data.
- Identify areas for improvement in survey design by analyzing response patterns and dropout rates.

<u>Open-Source Tools in Educational Applications</u>

The results underscore the growing significance of open-source tools in academic research and education. The affordability and flexibility of PostgreSQL and Power BI make them viable solutions for institutions with limited resources. The ability to process large-scale datasets efficiently enhances student learning experiences, enabling hands-on exposure to real-world data engineering tasks.

Discussion

The insights obtained from the study on the use of PostgreSQL in tandem with the free version of Power BI have significant implications for data engineering, particularly within the context of academic research and resource-constrained environments. The discoveries underscore the utility of combining open-source database management systems with accessible data visualization tools to effectively manage and represent large datasets. This hybrid approach to data engineering could act as a blueprint for educational institutions and researchers seeking cost-effective solutions without compromising analytical capabilities. The successful use of PostgreSQL highlights its status as a viable backbone for demanding data-engineering tasks. Given the volume and complexity of data generated in research studies, PostgreSQL's capabilities in managing ETL processes showcase that advanced data transformations, database integrity maintenance, and complex query execution are all achievable without the need for costly commercial software. Researchers can leverage these tools to perform sophisticated data analyses and maintain complex datasets efficiently.

For instance, an educational institution running longitudinal studies on student performance data can utilize PostgreSQL to handle data from multiple semesters, performing automated ETL tasks to cleanse and transform the data for deeper insights. Additionally, research projects requiring extensive data manipulation, such as genomic studies, can depend on PostgreSQL's robustness for managing and analyzing voluminous datasets. Power BI's free version as a visualization tool within the study broadens the scope for business intelligence (BI) applications in academia. Crafting interactive dashboards and visually appealing data presentations without incurring significant costs is invaluable for institutions with tight budgets. For example, university departments can use Power BI to create and share real-time visualizations of enrollment statistics or funding allocations, enhancing strategic decision-making processes.

Moreover, the seamless integration of Power BI with PostgreSQL demonstrates that flexible, scalable toolkits are feasible even with minimal investment. This ease of integration means that educational institutions can combine various data sources, such as student information systems, financial databases, and research data, into comprehensive, cohesive descriptions. Institutions promoting interdisciplinary research and collaboration can greatly benefit from this capability, as it facilitates a unified approach to data management. The study signifies that combining open-source and free commercial software can foster innovative research methodologies and a culture of data literacy. For instance, a multi-disciplinary research initiative analyzing social determinants of health could employ PostgreSQL for rigorous data management and Power BI for accessible visualization to effectively communicate findings to policymakers. These insights encourage further investigation into integrating diverse analytical tools and may inspire new strategies for data management and visualization. The technical robustness and economic viability of these tools affirm the growing trend towards more open, integrated, and accessible data toolsets in the data engineering community.

Field-Specific Data Engineering Applications

Beyond motorsports, open-source data engineering tools have applications across various academic disciplines. Fields such as healthcare, mechanical engineering, and environmental science present unique data challenges that PostgreSQL and Power BI can effectively address.

- Healthcare Data Management: structured handling of patient records is essential, requiring compliance with privacy regulations such as HIPAA. PostgreSQL's robust indexing and transaction control ensures data integrity, while Power BI allows for real-time monitoring of clinical and research data.
- Mechanical and Sensor Data Engineering: IoT and engineering applications require real-time sensor data processing. PostgreSQL supports time-series data and JSON-based structures, making it ideal for handling high-frequency measurements and integrating diverse data formats.
- Environmental and Geospatial Data Processing: climate research relies on large-scale geospatial datasets. PostgreSQL's PostGIS extension enables advanced geospatial queries, while Power BI allows visualization of climate patterns and geographic trends.
- Social Sciences and Policy Analysis: survey-based research in sociology and political science often requires text analytics, data normalization, and trend analysis. PostgreSQL enables efficient processing of qualitative responses, and Power BI provides interactive, data-driven storytelling for researchers.

By acknowledging these diverse applications, this study highlights the scalability and adaptability of open-source data engineering solutions across disciplines. Future research could explore how these tools perform under different data conditions and institutional settings. Ultimately, this study serves as a call to action for educational institutions and researchers to adopt and implement these hybrid tools. They offer practical approaches to managing and visualizing data that can significantly enhance academic research outcomes and foster a deeper understanding of data-driven insights.

Comparative Analysis of Methodologies and Tools

To further evaluate the adaptability of PostgreSQL and Power BI, a comparative analysis was conducted to explore their optimization for different dataset types and research contexts. This comparative discussion demonstrates the flexibility of these open-source tools and how they can be optimized for various domains. Future work could explore how different configurations impact performance in large-scale data engineering projects.

- Healthcare vs. Engineering: while healthcare datasets require structured, relational databases with strong compliance measures, engineering datasets often involve real-time processing of unstructured sensor data. PostgreSQL's indexing and concurrency control support both cases, making it flexible for diverse applications.
- Geospatial vs. Text Analytics: environmental research benefits from spatial indexing (PostGIS), while social sciences require advanced text querying. PostgreSQL supports both use cases, making it adaptable across disciplines.

- Static vs. Dynamic Visualizations: business intelligence dashboards in Power BI work well for static, predefined reports, whereas dynamic, exploratory analysis may require additional scripting tools like Python or R for deeper insight extraction.

Educational Implications of Open-Source Data Engineering

The adoption of open-source tools and data engineering practices holds significant potential for interdisciplinary education and collaboration across academic domains. As universities and research institutions increasingly rely on data-driven decision-making, equipping students with hands-on experience in open-source data tools is essential for workforce readiness. PostgreSQL, Power BI, and similar platforms provide a practical foundation for students and researchers across various fields, including STEM, social sciences, healthcare, and business analytics. Figure 6 illustrates the multifaceted advantages of using open-source tools are multifaceted and summarizes the key benefits of open-source platforms in academic and engineering contexts.

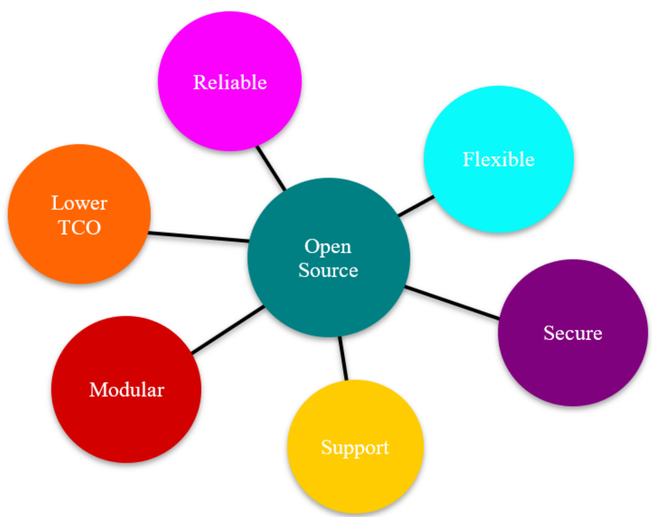


Figure 6. Benefits of open source.

One of the most notable educational benefits of open-source tools is their accessibility. With limited financial barriers, institutions can integrate PostgreSQL into coursework across disciplines, allowing students to engage in real-world data-engineering practices without incurring high software licensing costs. For example, in data science and computer science programs, students can develop ETL pipelines and practice database optimization techniques, while in business and economics courses, Power BI can be used for financial modeling and market trend analysis. The interdisciplinary nature of data engineering fosters collaboration among different academic fields. Open-source tools enable students from diverse backgrounds to work on shared datasets, breaking down silos between disciplines. For instance, a joint research project between environmental science and public policy students could use PostgreSQL to store and analyze climate data, while Power BI aids in visualizing the potential impacts of policy changes on environmental conditions. Similarly, healthcare students working alongside data analysts can leverage open-source tools to evaluate patient trends and make data-driven recommendations for public health initiatives.

Another advantage of incorporating open-source tools into education is their alignment with industry practices. Many organizations rely on PostgreSQL for large-scale data management and Power BI for real-time reporting, making these skills highly relevant for students entering the job market. By exposing students to these tools early, academic programs better prepare graduates for data-centric careers, whether in engineering, finance, healthcare, or research. Furthermore, open-source platforms encourage innovation and experimentation. Since PostgreSQL allows for extensive customization through extensions and scripting, students can explore advanced analytical techniques such as machine learning integration, geospatial analysis, and real-time data processing. Faculty members can also develop new curriculum components centered around open-source data engineering, enhancing the depth and breadth of data education.

Ultimately, the integration of open-source data tools into academic settings promotes digital literacy, enhances problem-solving skills, and encourages collaboration between students and faculty from diverse disciplines. As data continue to shape decision making across industries, ensuring that students have proficiency in open-source platforms like PostgreSQL and Power BI will be instrumental in fostering a data-literate workforce prepared to tackle complex challenges in a rapidly evolving technological landscape.

Applicability of PostgreSQL and Power BI in Engineering Disciplines

Beyond motorsports research, PostgreSQL and Power BI have practical applications in multiple engineering fields. In mechanical engineering, PostgreSQL can be utilized for predictive maintenance by managing large volumes of sensor data collected from manufacturing processes and IoT-enabled devices. Engineers can analyze machine performance, detect anomalies, and optimize maintenance schedules to reduce downtime. Power BI enhances this process by visualizing real-time performance metrics and failure trends, making it easier to identify inefficiencies in production systems. In civil engineering, PostgreSQL is valuable for managing infrastructure data, such as traffic modeling and structural integrity analysis. Large datasets collected from sensors embedded in bridges, roads, and buildings can be efficiently stored and analyzed using PostgreSQL's advanced indexing and query optimization. Power BI can then be used to create interactive dashboards that display structural health trends, helping engineers make data-driven decisions about infrastructure maintenance and expansion.

In electrical engineering, PostgreSQL plays a crucial role in managing energy grid data and optimizing power distribution. Utilities can use PostgreSQL to store and analyze vast datasets related to energy consumption, grid failures, and load balancing. With Power BI's visualization capabilities, engineers can track power demand in real time, identify inefficiencies, and develop strategies to enhance energy distribution and sustainability. In environmental engineering, PostgreSQL's PostGIS extension is particularly useful for geospatial data analysis related to climate change, pollution control, and water resource management. Researchers can store and process large datasets from satellite imagery, weather stations, and remote sensing devices, conducting spatial queries to assess environmental changes over time. Power BI provides an intuitive way to visualize these insights, enabling policymakers and environmental scientists to make informed decisions about conservation efforts and urban planning.

Comparing Alternative Tools

While PostgreSQL and Power BI offer robust data engineering and visualization capabilities, other tools are also commonly used for similar tasks. MySQL and SQLite are widely used relational database management systems, but they lack some of PostgreSQL's advanced features such as support for complex queries, full-text search, and geospatial analytics. MySQL is optimized for web applications and transactional databases, while SQLite is lightweight and best suited for embedded applications rather than large-scale data engineering. For data visualization, Tableau and Google Data Studio are alternatives to Power BI. Tableau is known for its advanced analytics and visualization capabilities, particularly for complex datasets. However, it is a premium tool that requires licensing fees, making it less accessible for academic institutions compared to Power BI's free version. Google Data Studio, on the other hand, is a free visualization tool that integrates well with Google's ecosystem but has limited functionality for handling large datasets compared to Power BI's broader range of data connectors and analytical features. Ultimately, the choice of tools depends on the specific needs of a project. PostgreSQL and Power BI were selected for this study, due to their open-source accessibility, scalability, and strong integration capabilities. However, researchers and engineers should evaluate different tools based on their data complexity, budget constraints, and analytical requirements.

Conclusions

The results of this study underscore the transformative potential of open-source tools and free versions of commercial software such as Power BI in the field of data engineering. The study provided concrete evidence of the capabilities of these tools in handling complex, large-scale data sets effectively, which is particularly valuable in academic research settings. The use of PostgreSQL for ETL processes in the study demonstrated that open-source database systems are robust enough to manage extensive data transformations with high accuracy and consistency. This makes advanced data engineering practices accessible to researchers who might lack significant financial resources, thus democratizing the field. Moreover, the integration of the free version of Power BI for data analysis and visualization proved to be both practical and effective. Despite its limitations, Power BI enabled clear, interactive, and visually appealing data presentations, which were crucial for identifying correlations, trends, and outliers that traditional methods might miss.

The implications of these findings are significant for future research. The combined use of PostgreSQL and Power BI offers a practical approach to communicate complex data insights to diverse audiences, an often-required capability in academia and beyond. In broader terms, the study highlighted the promise of a more inclusive approach to data analytics—bridging powerful analytical capabilities with financial and technical accessibility. These outcomes advocate for integrating open-source and free software tools into academic curricula and research activities, fostering innovation and preparing students for a data-centric world. The aim of this study was to encourage further exploration of cost-effective tools in data engineering and analytics. Building on the findings, subsequent research can embrace the vast resources available through open-source communities and free software offerings, promoting a balanced approach to data science. This study also contributes to ongoing discussions about resource optimization in higher education. As universities aim to prepare future professionals effectively, the cost of learning materials and software remains a concern. Evaluating the feasibility of open-source materials can significantly reduce these costs, while achieving desired educational outcomes and meeting industry demands.

Future Research

The findings from this study not only address the stated research questions but also open avenues for future exploration. Given the dynamic nature of information technology and data engineering, continuous advancements will present more opportunities for research. A significant extension of this study could involve collecting and analyzing a larger dataset. This would allow for a more comprehensive evaluation of the functionality and performance of various peer software products available on the market. Such a broader study could enhance the robustness of the findings and provide deeper insights into the comparative efficiency and usability of these tools in different contexts. Additionally, future research could explore the application of these tools in other fields beyond STEM education and motorsports.

For instance, investigating their use in healthcare, social sciences, or environmental studies could demonstrate the versatility and broader applicability of PostgreSQL and Power BI.

Further experimentation with other open-source and low-cost commercial tools in combination with PostgreSQL and Power BI could also be valuable. This would help in identifying the best tool sets for specific data engineering tasks and in developing refined methodologies that optimize both functionality and cost-efficiency. The potential for integrating machine learning algorithms within the PostgreSQL-ETL pipeline and visualizing predictive analytics in Power BI is another promising avenue. Such integration could revolutionize how educational institutions and researchers approach complex data problems, providing predictive insights in a visually engaging manner.

Finally, investigating the pedagogical impact of incorporating these tools into academic curricula can provide insights into how effectively they prepare students for real-world data engineering challenges. This line of research would not only contribute to academic literature but also offer practical recommendations for educational institutions aiming to foster a data-savvy culture. In summary, this study sets the stage for further research aimed at enhancing the understanding and application of open-source and low-cost commercial tools in data engineering. The potential for future research is vast, promising ongoing contributions to the field and broadening the scope of cost-effective, high-quality data analysis and visualization.

Acknowledgments

This work was supported in part by the College of Innovation and Technology at the University of Michigan-Flint as a faculty-guided student investigation project to support community engaged research.

References

BriteWire. (2018, August 11). Why open source? [Blog post]. https://britewire.com/why-open-source/

Drake, J. D., & Worsley, J. C. (2002). *Practical PostgreSQL*. O'Reilly Media, Inc.

García, M., López, J., & Ruiz, F. (2020). Geospatial analytics in climate monitoring: PostgreSQL-based solutions for environmental data management. *European Space Agency Journal, 32*(4), 115-130.

Groves, R. M. (1987). Research on survey data quality. *The Public Opinion Quarterly, 51*, S156-S172.

Jiang, X., Patel, D., & Turner, R. (2021). Optimizing electronic health records using PostgreSQL: A case study in large-scale hospital data management. *Journal of Medical Informatics, 18*(2), 67-89.

Khan, M. A., & UrRehman, F. (2012). Free and open-source software: Evolution, benefits and characteristics. *International Journal of Emerging Trends & technology in Computer Science, 1*(3), 1-7.

Kim, H., & Lee, J. (2021). Air quality monitoring and policy implementation using Power BI visualization. *Environmental Data Science, 29*(3), 54-72.

Microsoft. (2015). Power BI. https://app.powerbi.com/

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press.

Pérez-Riverol, Y., Gatto, L., Wang, R., Sachsenberg, T., Uszkoreit, J., Leprevost, F. D. V. ...Vizcaíno, J. A. (2016). Ten simple rules for taking advantage of Git and GitHub. *PLOS Computational Biology, 12*(7), e1004947. https://doi.org/10.1371/journal.pcbi.1004947

PostgreSQL Global Development Group. (2024). PostgreSQL official website. https://www.postgresql.org/

Raj, A., Bosch, J., Olsson, H. H., & Wang, T. J. (2020, August). Modelling data pipelines. *Proceedings of the 2020 46th Euromicro conference on software engineering and advanced applications (SEAA)*, 13-20. IEEE.

Rodríguez, A., & Patel, S. (2023). IoT data management in industrial automation: Leveraging PostgreSQL for predictive maintenance. *International Journal of Smart Manufacturing, 20*(5), 112-128.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Biographies

SALVATORE DIMAGGIO is a senior student at the University of Michigan-Flint pursuing a joint bachelor's degree in information technology & informatics and minoring in data analytics, and master's degree in software engineering with a concentration in data engineering. At the University of Michigan-Flint, Salvatore is the co-founder/ president of a student organization called the Innovators & Makers Club. Salvatore is also a member of the Dean of Students Think Tank along with being a student ambassador for the College of Innovation and Technology. Salvatore is also a Data Engineer Intern at Stellantis. Salvatore may be reached at saldim@umich.edu

E. SHIRL DONALDSON, as a PMP Academician, is her second career. Dr. Donaldson worked as a partner in a family-owned manufacturing firm resulting in exceptional insight into entrepreneurship, manufacturing, operations management, project management, quality systems, technology management, and STEM education. In August of 2021, Dr. Donaldson joined the College of Innovation and Technology as inaugural faculty at the University of Michigan-Flint. Previously, Dr. Donaldson taught operations management, project management, supply chain, and quality systems at The University of Texas and Purdue University. Problem-solving is Shirl's passion. She is a certified project management professional (PMP) and an active member of Project Management International (PMI). Her current research agenda is focused on several community engagement projects, such as modifying vehicles for handicapped children, racial equity and STEM, STEM education and Donk racing, motorsports and applied STEM, leveraging project management for student success, and entrepreneurship in underserved communities. Dr. Donaldson may be reached at shirld@umich.edu

Instructions for Authors: Manuscript Formatting Requirements

The INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION is an online/print publication designed for Engineering, Engineering Technology, and Industrial Technology professionals. All submissions to this journal, submission of manuscripts, peer-reviews of submitted documents, requested editing changes, notification of acceptance or rejection, and final publication of accepted manuscripts will be handled electronically. The only exception is the submission of separate high-quality image files that are too large to send electronically.

All manuscript submissions must be prepared in Microsoft Word (.doc or .docx) and contain all figures, images and/or pictures embedded where you want them and appropriately captioned. Also included here is a summary of the formatting instructions. You should, however, review the sample Word document on our website (http://ijeri.org/formatting-guidelines) for details on how to correctly format your manuscript. The editorial staff reserves the right to edit and reformat any submitted document in order to meet publication standards of the journal.

The references included in the References section of your manuscript must follow APA-formatting guidelines. In order to help you, the sample Word document also includes numerous examples of how to format a variety of scenarios. Keep in mind that an incorrectly formatted manuscript will be returned to you, a delay that may cause it (if accepted) to be moved to a subsequent issue of the journal.

1. **Word Document Page Setup:** Two columns with ¼" spacing between columns; top of page = ¾"; bottom of page = 1" (from the top of the footer to bottom of page); left margin = ¾"; right margin = ¾".

2. **Paper Title:** Centered at the top of the first page with a 22-point Times New Roman (Bold), small-caps font.

3. **Page Breaks:** Do not use page breaks.

4. **Figures, Tables, and Equations:** All figures, tables, and equations must be placed immediately after the first paragraph in which they are introduced. And, each must be introduced. For example: "Figure 1 shows the operation of supercapacitors." "The speed of light can be determined using Equation 4:"

5. **More on Tables and Figures:** Center table captions above each table; center figure captions below each figure. Use 9-point Times New Roman (TNR) font. Italicize the words for table and figure, as well as their respective numbers; the remaining information in the caption is not italicized and followed by a period—e.g., "*Table 1*. Number of research universities in the state." or "*Figure 5*. Cross-sectional aerial map of the forested area."

6. **Figures with Multiple Images:** If any given figure includes multiple images, do NOT group them; they must be placed individually and have individual minor captions using, "(a)" "(b)" etc. Again, use 9-point TNR.

7. **Equations:** Each equation must be numbered, placed in numerical order within the document, and introduced—as noted in item #4.

8. **Tables, Graphs, and Flowcharts:** All tables, graphs, and flowcharts must be created directly in Word; tables must be enclosed on all sides. The use of color and/or highlighting is acceptable and encouraged, if it provides clarity for the reader.

9. **Textboxes:** Do not use text boxes anywhere in the document. For example, table/figure captions must be regular text and not attached in any way to their tables or images.

10. **Body Fonts:** Use 10-point TNR for body text throughout (1/8" paragraph indention); indent all new paragraphs as per the images shown below; do not use tabs anywhere in the document; 9-point TNR for author names/affiliations under the paper title; 16-point TNR for major section titles; 14-point TNR for minor section titles.


Incorrect


Correct

11. **Personal Pronouns:** Do not use personal pronouns (e.g., "we" "our" etc.).

12. **Section Numbering:** Do not use section numbering of any kind.

13. **Headers and Footers:** Do not use either.

14. **References in the Abstract:** Do NOT include any references in the Abstract.

15. **In-Text Referencing:** For the first occurrence of a given reference, list all authors—last names only—up to seven (7); if more than seven, use "et al." after the seventh author. For a second citation of the same reference—assuming that it has three or more authors—add "et al." after the third author. Again, see the *sample Word document* and the *formatting guide for references* for specifics.

16. **More on In-Text References:** If you include a reference on any table, figure, or equation that was not created or originally published by one or more authors on your manuscript, you may not republish it without the expressed, written consent of the publishing author(s). The same holds true for name-brand products.

17. **End-of-Document References Section:** List all references in alphabetical order using the last name of the first author—last name first, followed by a comma and the author's initials. Do not use retrieval dates for websites.

18. **Author Biographies:** Include biographies and current email addresses for each author at the end of the document.

19. **Page Limit:** Manuscripts should not be more than 15 pages (single-spaced, 2-column format, 10-point TNR font).

20. **Page Numbering:** Do not use page numbers.

21. **Publication Charges:** Manuscripts accepted for publication are subject to mandatory publication charges.

22. **Copyright Agreement:** A copyright transfer agreement form must be signed by all authors on a given manuscript and submitted by the corresponding author before that manuscript will be published. Two versions of the form will be sent with your manuscript's acceptance email.

    Only one form is required. Do not submit both forms!

    The form named "paper" must be hand-signed by each author. The other form, "electronic," does not require hand signatures and may be filled out by the corresponding author, as long as he/she receives written permission from all authors to have him/her sign on their behalf.

23. **Submissions:** All manuscripts and required files and forms must be submitted electronically to Dr. Philip D. Weinsier, manuscript editor, at philipw@bgsu.edu.

24. **Published Deadlines:** Manuscripts may be submitted at any time during the year, irrespective of published deadlines, and the editor will automatically have your manuscript reviewed for the next-available issue of the journal. Published deadlines are intended as "target" dates for submitting new manuscripts as well as revised documents. Assuming that all other submission conditions have been met, and that there is space available in the associated issue, your manuscript will be published in that issue if the submission process—including payment of publication fees—has been completed by the posted deadline for that issue.

    Missing a deadline generally only means that your manuscript may be held for a subsequent issue of the journal. However, conditions exist under which a given manuscript may be rejected. Always check with the editor to be sure. Also, if you do not complete the submission process (including all required revisions) within 12 months of the original submission of your manuscript, your manuscript may be rejected or it may have to begin the entire review process anew.

THE LEADING JOURNAL OF ENGINEERING, APPLIED SCIENCE, AND TECHNOLOGY

International Journal of Modern Engineering

www.ijme.us

Print ISSN: 2157-8052
Online ISSN: 1930-6628

International Association of Journals and Conferences

www.iajc.org

INTERNATIONAL JOURNAL OF MODERN ENGINEERING

ABOUT IJME:

- IJME was established in 2000 and is the first and official flagship journal of the International Association of Journal and Conferences (IAJC).
- IJME is a high-quality, independent journal steered by a distinguished board of directors and supported by an international review board representing many well-known universities, colleges and corporations in the U.S. and abroad.
- IJME has an impact factor of **3.00**, placing it among the top 100 engineering journals worldwide, and is the #1 visited engineering journal website (according to the National Science Digital Library).

OTHER IAJC JOURNALS:

- The International Journal of Engineering Research and Innovation (IJERI)
  For more information visit www.ijeri.org
- The Technology Interface International Journal (TIIJ).
  For more information visit www.tiij.org

IJME SUBMISSIONS:

- Manuscripts should be sent electronically to the manuscript editor, Dr. Philip Weinsier, at philipw@bgsu.edu.

  For submission guidelines visit
  www.ijme.us/submissions

TO JOIN THE REVIEW BOARD:

- Contact the chair of the International Review Board, Dr. Philip Weinsier, at philipw@bgsu.edu.

  For more information visit
  www.ijme.us/ijme_editorial.htm

INDEXING ORGANIZATIONS:

- IJME is indexed by numerous agencies. For a complete listing, please visit us at www.ijme.us.

Contact us:

Mark Rajai, Ph.D.

Editor-in-Chief
California State University-Northridge
College of Engineering and Computer Science
Room: JD 4510
Northridge, CA 91330
Office: (818) 677-5003
Email: mrajai@csun.edu

www.tiij.org

www.ijeri.org

IJERI

The International Journal of Engineering Research & Innovation (IJERI) is the second official journal of the International Association of Journals and Conferences (IAJC). IJERI is a highly-selective, peer-reviewed print journal which publishes top-level work from all areas of engineering research, innovation and entrepreneurship.

IJERI Contact Information

General questions or inquiry about sponsorship of the journal should be directed to:

Mark Rajai, Ph.D.

Founding and Editor-In-Chief

Office: (818) 677-5003

Email: editor@ijeri.org

Department of Manufacturing Systems Engineering & Management

California State University-Northridge

18111 Nordhoff St.

Room: JD3317

Northridge, CA 91330