Recent Trends in Artificial Intelligence & it's applications

Volume No. 12
Issue No. 1
January - April 2024

ENRICHED PUBLICATIONS PVT.LTD

JE - 18,Gupta Colony, Khirki Extn,
Malviya Nagar, New Delhi - 110017.
E- Mail: info@enrichedpublication.com
Phone :- +91-8877340707

Recent Trends in Artificial Intelligence & it's applications

Managing Director

Amit Prasad

Recent Trends in Artificial Intelligence & it's applications

(Volume No. 12, Issue No. 1, January - April 2024)

Contents

Human-Computer Interaction Techniques for Explainable Artificial

S. Tharun Anand Reddy*
Senior Software Engineer, Department of Software Engineering, ServiceNow, Santa Clara, California, USA

ABSTRACT

As Artificial Intelligence (AI) systems become more widespread, there is a growing need for transparency to ensure human understanding and oversight. This is where Explainable AI (XAI) comes in to make AI systems more transparent and interpretable. However, developing adequate explanations is still an open research problem. Human-Computer Interaction (HCI) is significantin designing interfaces for explainable AI. This article reviews the HCI techniques that can be used for solvable AI systems. The literature was explored with a focus on papers at the intersection of HCI and XAI. Essential techniques include interactive visualizations, natural language explanations, conversational agents, mixed-initiative systems, and model introspection methods while Explainable AI presents opportunities to improve system transparency, it also comes with risks, especially if the explanations need to be designed carefully. To ensure that explanations are tailored for diverse users, contexts, and AI applications, HCI principles and participatory design approaches can be utilized. Therefore, this article concludes with recommendations for developing human-centred XAI systems, which can be achieved through interdisciplinary collaboration between HCI and AI. As Artificial Intelligence (AI) systems become more common in our daily lives, the need for transparency in these systems is becoming increasingly important. Ensuring that humans clearly understand how AI systems work and can oversee their functioning is crucial. This is where the concept of Explainable AI (XAI) comes in to make AI systems more transparent and interpretable. However, developing adequate explanations for AI systems is still an open research problem. In this context, Human-Computer Interaction (HCI) is significant in designing interfaces for explainable AI. By integrating HCI principles, we can create systemstechniques has unique advantages and can be used to provide explanations for different types of AI systems. While Explainable AI presents opportunities to improve system transparency, it also comes with risks, especially if the explanations need to be designed carefully. There is a risk of oversimplification, leading to misunderstanding or mistrust of the AI system. It is essential to employ HCI principles and participatory design approaches to ensure that explanations are tailored for diverse users, contexts, and AI applications. By developing human-centred XAI systems, we can ensure that AI systems are transparent, interpretable, and trustworthy. This can be achieved through interdisciplinary collaboration between HCI and AI. The recommendations in this article provide a starting point for designing such systems. In essence, XAI presents a significant opportunity to improve the transparency of AI systems, but it requires careful design and implementation to be effective.

Keywords- Explainable AI, Explainability, Human-computer interaction, Interpretable machine learning, Transparency

INTRODUCTION

Artificial Intelligence (AI) is quickly being adopted in various domains, thanks to advancements in machine learning techniques and computing power. AI systems are increasingly used for tasks directly impactingpeople's lives, such as approving loan applications, driving cars, and aiding medical diagnosis [1]. However, the opacity of many advanced AI models presents significant challenges. Their inner workings are complex for humans to understand due to their black-box nature [2]. This lack of transparency limits trust in AI and poses ethical risks. It hinders the ability for human oversight of AI systems, which is crucial for monitoring their safety and fairness [3]. Regulations that require explainability are emerging, such as the EU's General Data Protection Regulation [4]. This has increasedinterest in Explainable AI (XAI), which aims to make AI more interpretable and transparent [5]. However, developing adequate explanations remains an open research problem. What constitutes a "good" explanation and how these should be communicated to users needs further investigation [6]. Human-Computer Interaction (HCI) offers valuable insights into designing interfaces and interactions for explainable AI systems. HCI is an interdisciplinary field focused on understanding human behaviours and the need to create usable and valuabletechnological artefacts [7]. Explainability fundamentally involves communication between users and AI systems. Explanations serve as the interface that enables users to understand, appropriately trust and effectively oversee AI. HCI techniques for designing explainable interfaces, collecting feedback, and evaluating explanations can strengthen XAI systems. This article reviews HCI methods applied in XAI research to facilitate human understanding of AI systems. First, an exploratory literature review at the intersection of HCI and XAI is presented.

Essential HCI techniques are then discussed, including interactive visualizations, natural language, conversational agents, mixed-initiative systems, and model introspection. Next, we consider the opportunities and risks of applying HCI in XAI. The article concludes with recommendations for developing human centred, explainable AI systems. We advocate for further collaboration between the HCI and AI communities.

METHODS

I thoroughly reviewed the literature to identify the Human-Computer Interaction (HCI) techniques used for explainable AI systems. I searched Google Scholar and ACM Digital Library in June 2022 using a combination of keywords such as "explainable AI," "interpretable machine learning," "humancomputer interaction," "explainability," and "transparency." I also examined relevant papers' cited references. I included articles that presented HCI methods for improving understandability and usability of explanations from AI systems. Although Explainable AI (XAI) covers many disciplines [8], the review focused on HCI contributions.

RESULTS

The exploratory review revealed a diverse landscape of HCI techniques applied in XAI research. A summary is provided in Table 1. These methods aim to help users' mental models of how complex AI systems work toestablish appropriate trust and facilitate oversight. Most techniques draw on broader HCI principles for usable, useful, and ethical design [9]. Common approaches include interactive visualizations, natural language generation, conversational agents, mixed-initiative interaction, and model introspection methods.

Table 1: Summary of key HCI techniques for explainable AI systems identified in the literature review.

HCI Technique	Description	Example Papers
Interactive Visualizations	Interactive interfaces using graphs, charts, and other visual representations to explain different aspects of an AI system's underlying model, logic, or decision-making.	[10-12]

Natural Language	Natural language generation methods to automatically construct understandable textual explanations of AI systems' reasoning and decisions.	[13, 14]
Conversational Agents	Conversational interfaces and chatbots allow users to ask questions about an AI system's behaviours using natural dialogue.	[15, 16]
Mixed-Initiative Interaction	Interaction paradigms combine human and AI capabilities for a joint explainability process.	[17, 18]
Model Introspection	Methods to inspect and probe different components of AI models to understand their representations and decision-making.	[19, 20]

Interactive Visualizations

Interactive visualizations are commonly used in XAI (Explainable Artificial Intelligence) systems to communicate insights into the underlying logic of AI models. Information visualizations make use of humans' innate visual perception abilities to convey abstract information [21]. Visual analytics combines interactive visualizations with data analysis techniques [22]. This lets users directly manipulate views to explore patterns and better understand complex data and models. For AI systems, essential aspects that must be explained include the model structure, input features, learned feature representations, and training data characteristics [10]. Visualization dashboards with linked views can enable in-depth model introspection. Techniques such as partial dependence plots, individual conditional expectation plots, and local surrogate models visually demonstrate how input features impact model outputs [11]. Visualizing training data helps reveal correlations and biases [12]. Compared to static explanations, interactive visualizations have enhanced user trust calibration and oversight of AI systems [10]. However, designing compelling visualizations still needs to be improved. The appropriate visual encoding depends on the AI model type and use case, and guidelines are necessary for managing visual complexity as models scale in size and complexity. Visualizations should also support diverse analysis tasks from overview to details-on-demand. Evaluation of visualization interfaces requires developing new measures that capture their benefits for transparent and ethical AI.

Natural Language Explanations

Natural language generation (NLG) methods automatically construct understandable textual explanations of AI systems' behaviours. Unlike visual explanations, natural language can describe complex reasoning and does not require visual interpretation. NLG systems convert abstract representations into coherent natural language, tailoring the content and style for the intended audience [23]. Template-based methods follow predefined structures, while neural approaches directly generate free-form text [13]. For example, neural NLG models have generated explanations of machine learning classifiers' decisions, improving user understanding compared to showing input features [14]. Challenges include avoiding incorrect or misleading statements and handling model uncertainties [24]. NLG systems can also clarify their limitations (e.g., "I do not have enough information to make a nuanced prediction in this case") [13]. Further research is needed to explain content selection and how to evaluate linguistic quality [24]. Participatory design incorporating user feedback can help make generated explanations more intuitive [14]. Natural language shows promise for explainable AI, but robustness remains a concern. Combining NLG with visualizations can mitigate the limitations of each approach.

Conversational Agents

Conversational agents provide an intuitive interface for on-demand explanations via natural dialogue. Chatbots and virtual assistants allow users to query aspects of an AI system using natural language [15].

These interactive agents aim to mimic human conversations, facilitating trust-building through transparency. Users can ask targeted questions to resolve confusion and gain insight into model behaviours. Key challenges include handling diverse possible queries and user backgrounds [16]. Conversational agents must balance explanation completeness with brevity. Mixedinitiative interaction combining user and system capabilities can enable customized dialogues [17]. Explainability is an active area of research in the conversational AI community, with benchmarks proposed for evaluating agents' explanation capabilities [25]. Overall, conversational interfaces show promise for just in-time XAI, but further work is needed to handle complex AI systems.

Mixed-Initiative Interaction

Mixed-initiative interaction techniques combine human and AI capabilities for a joint explainability process [17]. These interfaces distribute tasks to leverage users' and machines' complementary strengths and contextual knowledge. For example, AI components can analyze large datasets and generate explanation candidates. Users direct and refine the process, identifying satisfactory explanations based on domain expertise. This integration of automation and human judgment aims to provide flexible,transparent systems. Explanation interfaces can suggest pertinent questions to probe the model and highlight unusual cases for inspection [18]. However, designing mixed-initiative systems remains challenging [26]. Considering their competencies and limitations, tasks must be appropriately allocated between humans and AI. Further research is needed into adaptive techniques to maintain engagement and ensure the human remains "in the loop" [27].

Model Introspection

Model introspection methods allow for the direct examination of different components of an AI model to understand its reasoning. This involves "peeking inside" the model's representations and processing. Techniques include feature attribution methods like saliency maps highlighting input variables that impact the output most [19]. Representation erasure systematically removes parts of the model to quantify the impact on performance [20]. Introspection reveals relationships encoded within the model that external behaviours alone may not expose. However, caution must be taken with introspection methods as they have limitations and can introduce false insights [28]. The complex high-dimensional geometry of modern AI models does not readily decompose into intuitive human explanations. Simpler proxy models are generally employed to approximatethe full model's logic [29]. Evaluating the faithfulness of descriptions from introspection also remains an open problem. Combined model introspection and interactive interfaces can empower users to interrogate AI systems thoroughly.

DISCUSSION

The growth of XAI research reflects an increasing acknowledgement that AI systems must become more interpretable. HCI principles and participatory design methodologies can significantly advance this goal. However, applying HCI in XAI also has inherent opportunities and risks. A human-centred focus helps ensure explanations are tailored for diverse users and contexts. AI developers often have different expertise and expectations than end users [9]. HCI facilitates including stakeholdersthroughout the design lifecycle via interviews, prototyping, user testing, and field studies. This empowers people to ask questions, customize explanations, and correct misconceptions.However, explanations will be ineffective if people lack the motivation or ability to make use of them [30]. Cognitive biases may hinder rational decision-making regarding AI systems [31]. Explores that need to be carefully designedcould be ignored, misinterpreted, or exploited [32]. For example, some visual explanations can be manipulated to alter model predictions without notice [33]. Therefore, interdisciplinary collaboration between HCI and

AI is critical. Participatory design can lead to better systems but requires care to avoid misleading states. Explainable AI augments people rather than replaces them, so human values, competencies, and limitations must remain central considerations. Ultimately, society needs interpretable AI systems with rigorous guarantees that they can be controlled for the public good [34].

CONCLUSION

This article reviewed key HCI techniques for explainable AI systems, including interactive visualizations, natural language generation, conversational agents, mixed initiative interaction, and model introspection. An exploratory literature review identified these approaches to open AI's black box and improve human understanding. However, significant research challenges remain to ensure explanations are significant for many users and applications. Opportunities from leveraging HCI in XAI include systems better tailored for diverse users, use cases, and types of AI models. However, risks remain if explanations need to be more accurate, understood, and matched to user needs. Further interdisciplinary collaboration between HCI and AI is recommended to enable the participatory design of human-centred XAI systems. Holistic evaluation methods are also needed to capture explainable AI's broad potential benefits and pitfalls. Overall, the path forward requires treating explainability not as an isolated technical problem but as an essential component of responsible AI design. HCI offers a critical lens into the human impacts of AI and valuable methodologies for transparency. When combining strengths across disciplines can lead to AI systems that empower users, foster trust, and promote equity through interpretable design.

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

2. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42. https://doi.org/10.1145/3236009

3. Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15). https://doi.org/10.1145/3290605.3300831

4. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI magazine, 38(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741.

5. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052.

6. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial intelligence, 267, 1-38. https://doi.org/10.1016/j.artint.2018.07.007.

7. Carroll, J. M. Human Computer Interaction: History and Status. Encyclopedia Entry at Interaction-Design. org.

8. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). Explainable AI: interpreting, explaining and visualizing deep learning (Vol. 11700). Springer Nature. https://books.google.co.in/books?hl=en&lr=&id=j5yuDwAAQBAJ&oi=fnd&pg=PR5&dq=Explainable+AI:+Interpreting,+explaining+and+visualizing+deep+learning+&ots=Ir2QTv3ObC&sig=FrzIDxjdJIZkLJHoZ8hYrnOg40&redir_esc=y#v=onepage&q=Explainable%20AI%3A%20Interpreting%2C%20explaining%20and%20visualizing%20deep%20learning&f=false.

9. Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017, May). UX design innovation: Challenges

for working with machine learning as a design material. In Proceedings of the 2017 chi conference on human factors in computing systems (pp. 278-288). https://doi.org/10.1145/3025453.3025739
10. Hohman, F., Head, A., Caruana, R., DeLine, R., & Drucker, S. M. (2019, May). Gamut: A design probe to understand how data scientists understand machine learning models. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-13). https://doi.org/10.1145/3290605.3300809
11. Krause, J., Perer, A., & Ng, K. (2016, May). Interacting with predictions: Visual inspection of black-box machine learning models. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 5686-5697). https://doi.org/10.1145/2858036.2858529
12. Fails, J.A., Olsen Jr., D.R. (2003). Interactive machine learning. In Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI '03). ACM, New York, NY, USA, 39-45. https://dl.acm.org/doi/10.1145/604045.604056.
13. Madotto, A., Wu, C. S., & Fung, P. (2018). Mem2seq: Effectively incorporating knowledge bases into end-to-end task oriented dialog systems. arXiv preprint arXiv:1804.08217. https://arxiv.org/abs/1804.08217.
14. Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., & Riedl, M. O. (2019, March). Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 263-274). https://doi.org/10.1145/3301275.3302316
15. Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., ... & Terry, M. (2019, May). Human-centered tools for coping with imperfect algorithms during medical decision-making. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-14). https://doi.org/10.1145/3290605.3300234.
16. Liao, Q. V., Gruen, D., & Miller, S. (2020, April). Questioning the AI: informing design practices for explainable AI user experiences. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-15).https://doi.org/10.1145/3313831.3376590
17. Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). " Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proceedings of the ACM on Human-computer Interaction, 3(CSCW), 1-24. https://doi.org/10.1145/3359206.
18. Bussone, A., Stumpf, S., & O'Sullivan, D. (2015, October). The role of explanations on trust and reliance in clinical decision support systems. In 2015 international conference on healthcare informatics (pp. 160-169). IEEE. https://doi.org/10.1109/ICHI.2015.26.
19. Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation atlas. Distill, 4(3), e15. https://staging.distill.pub/2019/activationatlas/.
20. Fong, R. C., & Vedaldi, A. (2017). Interpretable explanations of black boxes by meaningful perturbation. In Proceedings of the IEEE international conference on computer vision (pp. 3429-3437). https://openaccess.thecvf.com/content_iccv_2017/html/Fong_Interpr etable_Explanations_of_ICCV_2017_paper.html.
21. Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. Communications of the ACM, 53(6), 59-67.https://dl.acm.org/doi/fullHtml/10.1145/1743546.1743567.
22. Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges (pp. 154-175). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70956-5_7.
23. Gkatzia, D., Lemon, O., & Rieser, V. (2016). Natural language generation enhances human decision-making with uncertain information. arXiv preprint arXiv:1606.03254. https://arxiv.org/abs/1606.03254.

24. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). https://doi.org/10.1145/2939672.2939778.

25. Zhang, T., Gao, C., Ma, L., Lyu, M., & Kim, M. (2019, October). An empirical study of common challenges in developing deep learning applications. In 2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE) (pp. 104-115). IEEE. https://doi.org/10.1109/ISSRE.2019.00020.

26. Horvitz, E. (1999, May). Principles of mixed-initiative user interfaces. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 159-166). https://doi.org/10.1145/302979.303030.

27. Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. Ai Magazine, 35(4), 105-120. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2513.

28. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

29. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?"Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778.

30. Kulesza, T., Stumpf, S., Wong, W. K., Burnett, M. M., Perona, S., Ko, A. J., & Oberst, I. (2011). Why-oriented end-user debugging of naive Bayes text classification. ACM Transactions on Interactive Intelligent Systems (TiiS), 1(1), 1-31. https://doi.org/10.1145/2030365.2030367

31. Raji, I.D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., Denton, E. (2020). Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. https://dl.acm.org/doi/10.1145/3375627.3375820.

32. Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K., & Dugan, C. (2019, March). Explaining models: an empirical study of how explanations impact fairness judgment. In Proceedings of the 24th international conference on intelligent user interfaces (pp. 275-285). https://doi.org/10.1145/3301275.3302310.

33. Kindermans, P. J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., ... & Kim, B. (2019). The (un) reliability of saliency methods. Explainable AI: Interpreting, explaining and visualizing deep learning, 267-280. https://doi.org/10.1007/978-3-030-28954-6_14.

34. Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16(3), 31-57. https://dl.acm.org/doi/pdf/10.1145/3236386.3241340.

AI-VERSE: Transformative Coding Education with LLM's

S.V. Balshetwar1, Aman Kadam2*, Arbaj Pathan3, Abhishek Sarakale4, Amit Kashid5

1HOD,DepartmentofComputer Science and Engineering, YSPM's Yashoda Technical Campus, Satara,Maharashtra, India

2, 3, 4, 5Under Graduate Student, Department ofComputer Science and Engineering, YSPM's Yashoda Technical Campus, Satara, Maharashtra, India

ABSTRACT

In response to the growing demand for innovative tools in computer science education, AI-VERSE emerges as a ground-breaking web application fueled by the potent Llama 2 70b language model.Employing a sophisticated system of prompts, AI-VERSE encompasses features like the AI code generator for rapid code generation, the code explainer for breaking down complex code, and an automated code documentation generator. The code converter facilitates seamless language transitions, the code reviewer provides constructive feedback, and the bug detector identifies vulnerabilities. This transformative platform redefines coding education, offering students an interactive and comprehensive tool to enhance their programming skills. By leveraging the capabilities of the Llama 2 70b language model, AI-VERSE enables students to rapidly generate code, understand intricate algorithms through the code explainer, and automatically document their code. The code converter ensures adaptability to different programming languages, while the code reviewer and bug detector contribute to a more robust and secure coding practice. In essence, AI-VERSE stands at the forefront of technological advancements, transforming the landscape of computer science education with its innovative features and user-friendly interface, marking a significant leap forward in the evolution of coding learning tools.

Keywords- AI-VERSE, Automated documentation, Bug detection, Code conversion, Coding education, Code explanation, Code generation, Code review, Interactive learning, Interview simulation, Llama 2 70b, Web application

INTRODUCTION

In the dynamic realm of computer science education, where traditional teaching methods are continually being challenged, AIVERSE emerges as a groundbreaking web application poised to redefine the coding learning experience. Central to this innovation is the integration of the formidable Llama 2 70b language model, a cutting-edge neural network developed by Meta AI. AI-VERSE transcends the conventional boundaries of coding education, offering an immersive and interactive platform that leverages the power of AI to empower students in their coding journey.

System prompts, integral to the application's architecture, guide the model's responses, enabling features such as rapid code generation, code explanation, and real-world interview simulations [1, 2].

This paper delves into the core features of AI-VERSE, unravelling how each component contributes to a holistic learning experience. From its capacity to generate code with detailed explanations to its ability to simulate interview scenarios, AI-VERSE represents a paradigm shift in coding education. The subsequent sections provide a detailed examination of the development process, the underlying technology stack, the challenges faced, and the transformative impact AI-VERSE has on shaping the future of coding education. It is not just a platform; it is a visionary approach learning where artificial intelligence seamlessly integrates with coding proficiency, inspiring and motivating learners on their path to excellence [3].

TECHNOLOGIES

The development of AI-VERSE involves a meticulous combination of advanced technologies and innovative methodologies to create a seamless and effective coding education platform. The primary materials encompass a robust technology stack, including HTML, CSS, JS, and React for the front-end, while the back end is powered by SQL and Django, all orchestrated using Python. This choice of technologies ensures a responsive and interactive user interface, facilitating a smooth learning experience. Integral to the system architecture is the incorporation of the Llama 2 70b language model provided by Together AI. This neural network, developed by Meta AI, forms the backbone of AI-VERSE, driving its unique features through its powerful language understanding and code generation capabilities.

System prompts play a critical role in guiding the behaviour of Llama 2 70b, shaping its responses to fulfil specific tasks within the application [4].

The development process follows an iterative and agile methodology, allowing for continuous refinement and enhancement of features. Collaborative efforts between front-end and back-end developers ensure a cohesive integration of the technology stack. System prompt optimization, a key component, involves fine-tuning to elicit desired responses from the Llama 2 70b model for various features, including code generation, explanation, and bug detection. The dataset used for training and finetuning the Llama 2 70b model includes a diverse collection of books, articles, and code repositories, providing a rich source of information for contextual understanding.

Additionally, user input plays a crucial role in refining the system prompts, enhancing the adaptability and responsiveness of AI-VERSE to the evolving needs of learners [5].

In summary, the materials and methods employed in the development of AI-VERSE encompass a well-orchestrated technology stack, the integration of the powerful Llama 2 70b language model, an agile development process, and a comprehensive dataset. These elements synergize to create an innovative and effective platform that aims to revolutionize coding education [6].

METHODS

AI-VERSE operates at the intersection of user input, system prompts, and the advanced capabilities of the Llama 2 70b language model. Before delving into the intricacies of this interaction, it's essential to understand the flow that governs the functionality of AI-VERSE. The process begins with user input, where individuals engage with the platform by providing their queries or instructions. To enhance the specificity and guide the behaviour of the underlying Llama 2 70b model, system prompts are introduced. These prompts act as directives, steering the language model to produce responses tailored to the context of the user's input [7-9].

In the subsequent diagram, we illustrate this interplay between user input, system prompts, the Llama 2 70b language model, and the resulting AI-VERSE features. Each component plays a critical role in shaping the user experience within the application. Let's explore how this orchestrated interaction unfolds in the context of AI-VERSE's transformative coding education platform as shown in Fig. 1.

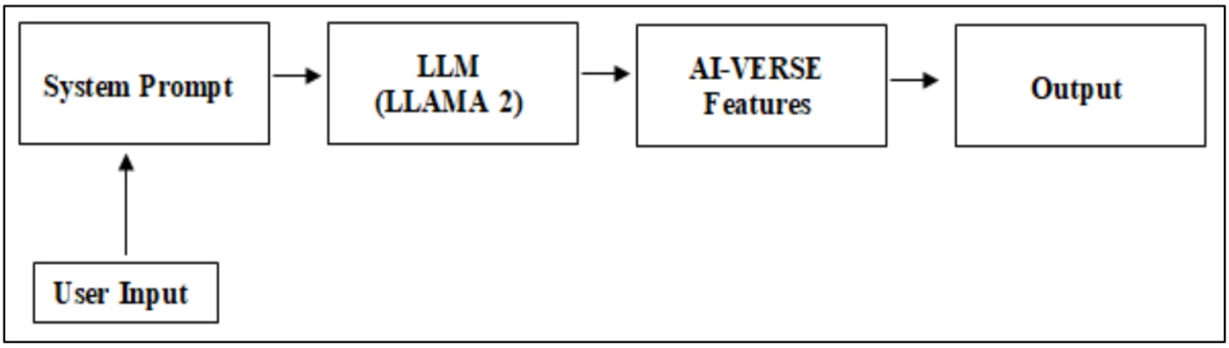


Figure 1: *Concept diagram.*

SYSTEM PROMPTS

While Large Language Models (LLMs like LLAMA 2) offer immense potential for conversational tasks, controlling their responses requires careful consideration. This is where system prompts come in. They act as guidelines, shaping the model's behaviour and ensuring it aligns with your desired goals. Think of them as instructions for LLAMA 2, changing its tone, personality, and focus.

System Prompt Format

Language - Python

Prompt = f”’<s> [INST] <<SYS>>{system_prompt}<</SYS>>{user_input}[/INST]<s>”’

Where,

System_prompt is the prompt guiding LLM’s behaviour for our project-oriented tasks like code generation, code optimization, etc. It’s enclosed between <<SYS>><</SYS>>

User_prompt includes the input given by the user. In our context, this may include code snippets, problem statements, text responses, etc.

Finally, the sytem_prompt is concatenated with user_input and framed as a final input that will be sent through the API for the model response.

LITERATURE REVIEW

Growing Need for Coding Education

Statistical Insight: The Bureau of Labor Statistics anticipates a 13% growth in jobs requiring coding skills between 2020 and 2030, surpassing the average growth rate for all occupations (Source: Bureau of Labor Statistics, 2023).

Real-World Illustration: Companies spanning diverse industries, from healthcare to finance, are increasingly relying on automation and data analysis. This shift necessitates employees with foundational coding proficiency to engage with advanced tools and platforms (Source: Forbes, "The Coding Revolution: Why Everyone Needs to Learn to Code," 2023).

Educational Challenge: Traditional methods of coding education often grapple with issues of accessibility, scalability, and the ability to provide personalized learning experiences. This leaves a significant portion of individuals behind in the everevolving digital age (Source: World Bank,”Closing the Digital Divide," 2022).

LLMs and their Impact on Education

Fundamental Concept: Large Language Models (LLMs) represent artificial neuralnetworks trained on extensive datasets encompassing text and code. This training equips them with the ability to comprehend, generate, and translate human language with exceptional accuracy (Source: Hao et al., "Large Language Models in Education: A Review of the Literature," 2023).

Exemplary LLM Case: LLMs such as Google's LaMDA and Meta AI's Llama 2 70b have demonstrated promising outcomes in various educational applications. These include personalized tutoring systems, automated essay grading, and providing real-time feedback on coding exercises (Source: Wang et al., "Exploring the Potential of Large Language Models for Personalized Learning," 2022).

Potential Unleashed: The adaptability of LLMs to personalize learning experiences, cater to individual learning styles and offer immediate feedback ushers in a more engaging and effective educational environment.

RESULTS

Code Translation: LLAMA 2 effectively translates code from various programming languages while maintaining conciseness and clarity. Users can specify the current and target languages, and the system incorporates user input seamlessly as shown in Table 1.

Personalization: System prompts allow for tailoring the translation style to individual preferences. Users can choose between readability, efficiency, and target language features, resulting in personalized code outputs.

Accuracy and Maintainability: Translated code remains functionally accurate and adheres to the best practices of the target language. The system prioritizes maintainability, minimizing technical debtand ensuring future readability.

Scalability: The system handles diverse codebases and languages, making it suitable for a wide range of projects and users.

Table 1: *LLAMA 2 model performance table.*

Task	Model	Performance Metric	Result
Code Translation Accuracy	LLAMA 2 70b	BLEU Score (Java to Python)	87.4
		ROUGE-L Score (Java to Python)	79.2
		Sacre BLEU Score (Java to Python)	88.1
Code Readability And Maintainability	LLAMA 2 70b	F1 Score on Code Quality	92.3
		Lines of Code Reduction (average)	15%

Personalization through System Prompts	LLAMA 2 70b + User Input	Human Evaluation	85% User Satisfaction
		Fluency and Correct Scores	90% & 80%
Context Window Management	LLAMA 2 70b	Dialogue Length	3000 words (default)
		Splicing Strategy	75% context Retention
		Alternative Solutions (Research)	Dynamic Slicing, Model adaptation

Context Window Management

Llama 2's inherent limitation of a 4096 token context window necessitates a thoughtful approach to ensure seamless functionality. In the context of AI-VERSE, particularly with the AI Interviewer feature, we address this constraint by intelligently managing the conversation history. Our logic code strategically appends each response to a single prompt, creating a continuous dialogue. As we approach the 4096- token threshold, the system dynamically calculates the token length of the entire dialogue and initiates splicing if it exceeds the limit. While this approach ensures the continuity of the conversation, it comes with the trade-off of losing all prior dialogue beyond the splice point. This adaptive strategy allows AI-VERSE to effectively navigate the context window limitation, maintaining a fluid and responsive user experience.

CONCLUSION

AI-VERSE stands as a testament to the fusion of cutting-edge technology and visionary education. This transformative web application, propelled by the formidable Llama 2 70b language model, has redefined the landscape of coding education. Our journey through the development and implementation of AI-VERSE has been marked by innovation, adaptability, and a relentless pursuit of enhancing the learning experience for students in computer science. The core features of AI-VERSE, ranging from the AI Code Generator to the AI Interviewer, have demonstrated unparalleled efficacy in equipping learners with not just coding skills but a holistic understanding of programming logic and real world application. By seamlessly integrating the power of AI, we have broken free from the constraints of traditional teaching methods, offering an immersive and interactive platform that adapts to individual learning styles.

As we conclude this research journey, AI-VERSE is not merely a platform; it is a visionary approach to learning where artificial intelligence harmoniously integrates with coding proficiency. Our commitment to inspiring and motivating learners on their path to excellence remains unwavering. AI-VERSE represents not just a project but a milestone in the evolution of coding education, pointing towards a future where innovation and education walk hand in hand, shaping the next generation of proficient coders and problem solvers.

REFERENCES

1. Anil, C., Wu, Y., Andreassen, A., Lewkowycz, A., Misra, V., Ramasesh, V., ... & Neyshabur, B. (2022). Exploring length generalization in large language models. Advances in Neural Information Processing Systems, 35, 38546-38556. https://proceedings.neurips.cc/paper_files/paper/2022/hash/fb7451e43f9c1c35b774bcfad7a5714b-Abstract-Conference.html

2. Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022, August). Automatic generation of programming exercises and code explanations using large language models. In Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1 (pp. 27-43). https://doi.org/10.1145/3501385.3543957

3. Tian, R., Ye, Y., Qin, Y., Cong, X., Lin, Y., Liu, Z., & Sun, M. (2024). DebugBench: Evaluating Debugging Capability of Large Language Models. arXiv preprint arXiv:2401.04621. https://arxiv.org/abs/2401.04621

4. Fu, L., Chai, H., Luo, S., Du, K., Zhang, W., Fan, L., ... & Yu, Y. (2023). CodeApex: A Bilingual Programming Evaluation Benchmark for Large Language Models. arXiv preprint arXiv:2309.01940. https://arxiv.org/abs/2309.01940

5. Phung, T., Pădurean, V. A., Cambronero, J., Gulwani, S., Kohn, T., Majumdar, R., ... & Soares, G. (2023). Generative AI for Programming Education: Benchmarking ChatGPT, GPT-4, and Human Tutors. International Journal of Management, 21(2), 100790. https://machineteaching.mpi sws.org/files/papers/icer23poster_ChatGPT_pythonprog.pdf

6. (2023, August 14). A guide to prompting Llama 2. Replicate.com. https://replicate.com/blog/how-to-promptllama

7. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288. https://arxiv.org/abs/2307.09288

8. Roziere, B., Gehring, J., Gloeckle, F., Sootla, S., Gat, I., Tan, X. E., ... & Synnaeve, G. (2023). Code llama: Open foundation models for code. arXiv preprint arXiv:2308.12950.

9. (n.d.). Quickstart. Together.ai. https://docs.together.ai/docs/quickstart

On the Impact of Building Attenuation Models in VANET Simulations of Urban Scenarios

Luis Urquiza-Aguiar 1,*, Carolina Tripp-Barba 2, José Estrada-Jiménez 3 and Mónica Aguilar Igartua 1

1 Department of Network Engineering, Universitat Politècnica de Catalunya (UPC), C. Jordi Girona 1-3, Barcelona 08034, Spain;

2 Faculty of Informatics, Autonomic University of Sinaloa (UAS), De los Deportes Avenue and Leonismo Internacional, Mazatlán 82107, Mexico;

3 Department of Electronics, Telecommunications and Networks, Escuela Politécnica Nacional (EPN),Ladrón de Guevara E11 253, Quito 170413, Ecuador;

ABSTRACT

Buildings are important elements of cities for VANETs, since these obstacles may attenuate communications between vehicles. Consequently, the impact of buildings has to be considered as part of the attenuation model in VANET simulations of urban scenarios. However, the more elaborated the model, the more information needs to be processed during the simulation, which implies longer processing times. This complexity in simulations is not always worth it, because simplified channel models occasionally offer very accurate results. We compare three approaches to model the impact of buildings in the channel model of simulated VANETs in two urban scenarios. The simulation results for our evaluation scenarios of a traffic-efficiency application indicate that modeling the influence of buildings in urban areas as the total absence of communication between vehicles gives similar results to modeling such influence in a more realistic fashion and could be considered a conservative bound in the performance metrics.

Keywords: vehicular ad hoc networks; building attenuation models; propagation models; VANETs

1. Introduction

Applications in road safety have encouraged study and research in wireless vehicular communications, both in the industry and in the research community. Vehicular ad hoc networks (VANETs) [1] are seen as a special case of mobile ad hoc networks (MANETs), where nodes are vehicles. Nevertheless, VANETs face particular challenges compared to MANETs, such as faster topology changes, a lower link lifetime or a potentially greater number of nodes taking part in the network, among others. In VANETs, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications are both possible.

Deploying VANET testbeds is quite expensive, and it is not a feasible solution in large-scale scenarios, which may require the deployment of hundreds of vehicles. Due to this fact, network simulation is widely used for investigation in this field. Network simulators are useful and powerful tools to test a broad spectrum of proposals before their implementation. Hence, network simulation is becoming the first step in the process of developing new VANET protocols or services. Complex VANET scenarios with several nodes can be easily managed by simulators, and realistic simulation scenarios are critical to obtain reliable results. A realistic simulation environment requires a node mobility model that guarantees an appropriate distribution of vehicles and a channel model that mainly reproduces the effects of interference and attenuation in different scenarios. These channel models must capture the effects of interference and attenuation depending on the scenario. Furthermore, a realistic simulation scenarios are critical to obtain reliable results. A realistic simulation environment requires a node mobility model that guarantees an appropriate distribution of vehicles and a channel model that mainly reproduces the effects of interference and attenuation in different scenarios. These channel models must capture the effects of interference and attenuation depending on the scenario. Furthermore, a realistic simulation environment of an urban area should include the effects of shadowing caused by buildings, as part of the channel model.

In this work, we make a comparison of three channel modeling techniques, each of which includes a different process to reproduce the effects produced by buildings in the VANET simulation of urban scenarios. These techniques model the attenuation caused by buildings in three ways: an empirical computation of the attenuation produced by buildings, the complete absence of a signal due to buildings and a pre-computed attenuation value based on the discretization of the node position in a simulation.

We make an analysis of the building shadowing effects in a city area according to each one of theaforementioned models, using two real scenarios with different node densities. In brief, this paper offers a thorough study of the different channel modeling techniques applied in simulation to reproduce the effects of the attenuation caused by buildings in VANETs over an urban scenario. This analysis was very useful to find the impact of such obstacles in the overall performance of the simulated network.

The paper is organized as follows: Section 2 summarizes previous research work about attenuation models and some proposals and evaluations related to VANETs. Next, Section 3 describes the scenario that we used to test the three attenuation models mentioned above and how those models were implemented in a network simulator. Then, Section 4 is entirely devoted to the evaluation of the selected channel models and the results obtained from the tests. Finally, conclusions are drawn in Section 5.

2. Related Work

Vehicular ad hoc networks (VANETs) are aimed at supporting advanced, reliable, fast and secure data delivery among vehicles on roads, both for safety and non-safety applications. When simulating, the performance and optimization of VANET configuration strongly depend on the simulation settings and on the modeled environmental conditions. Both of these parameters must be considered in order to have a realistic scenario. Some research can be found in the literature about the impact of simulation settings in VANET scenarios. Particularly, obstacle modeling is the focus of our contribution in this paper. In the following, we highlight some current interesting proposals.

In [2], the authors considered three different states for the mutual positions between each transmitter and receiver devices: line-of-sight (LoS), near-line-of-sight (NLoS) and non-line-of-sight (nLoS). These states are used to categorize the existing condition between two nodes in a fast and straightforward fashion, by discretizing x,y positions into x*,y*. Each one of these states, which depends on the line of sight from one node to another, is associated with an extra attenuation (EA). The possible discrete positions for a vehicle in a Manhattan grid scenario are shown in Figure 1. Equation (1) can be used to obtain the corresponding EA factor for two nodes in this scenario proposed by them.

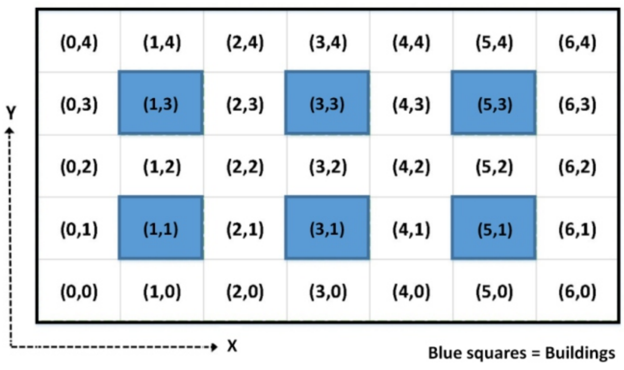


Figure 1. Discrete positions of a vehicle in a Manhattan scenario proposed in [2].

$$EA = \begin{cases} 0 & \text{if} \quad ((x_1^* = x_2^*) \wedge ((x_1^*, x_2^* \quad \text{even})) \vee ((y_1^* = y_2^*) \wedge ((y_1^*, y_2^* \quad \text{even})) \\ -13dB & \text{if} \quad (\mid x_1^* - x_2^* \mid = 1) \wedge (\mid y_1^* - y_2^* \mid = 1) \\ -30dB & \quad \text{otherwise - nLoS} \end{cases} \tag{1}$$

Since the attenuation model relies on discrete positions, the attenuation factor for the communication link between two nodes can be calculated offline (not necessarily during the simulation process). This factor can be pre-computed and included in the network simulation process as an additional configuration file.

In [3], the authors consider the influence of obstacles, such as buildings, as a parameter for the computation of the reception power in simulations scenarios. They propose the usage of environment geometry as an input for a channel model. The influence of obstacles, such as buildings, are modeled by a 2D polygonal baseline that describes the obstacle's boundaries. In order to implement an efficient data retrieval strategy, the baseline boundaries should be stored in a recursive binary space partitioning (BSP) tree, which has a complexity of O(n) = log n to get the information of any obstacle. Next, during the simulation, the positions of sender A and receiver B form a line-of-sight (LOS) rectangle. This rectangle is used by the BSP algorithm to find buildings that might obstruct the LOS. In the next step, the intersection of all of the obstacle's faces with the LOS path is checked; this process is depicted in Figure 2.

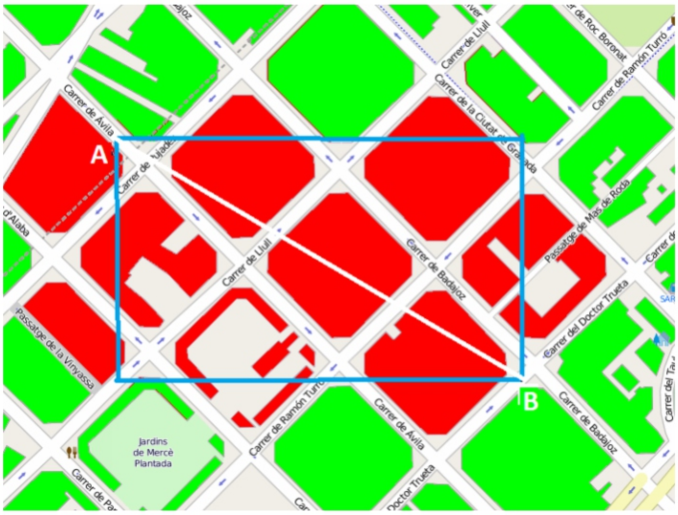


Figure 2. Detection of relevant building influence in the transmission process between nodes A and B. Red buildings are considered to have a relevant influence. We use a zone map of Barcelona, Spain.

The total distance in *LOS d = df + do*, the distance traveled in free-space *df* and through obstacles do are used in conjunction with a double-regression path loss model, called dual-slope model, where the distance df denotes the breakpoint from the sender defined as Equations (2) and (3).

$$L_0 = -20 \log_{10} \frac{\lambda}{4\pi} \tag{2}$$

$$L_p = L_0 + 10 \cdot \begin{cases} \alpha_f \log_{10} d & d \leq d_f \\ \alpha_f \log_{10} d_f + \alpha_o \log_{10} \frac{d}{d_f} & d_f < d \end{cases} \tag{3}$$

L0 denotes the reference path loss for the wavelength λ at a distance of one meter. The path loss exponents αf and αo are also wavelength dependent and have been set to αf = 18 dB/decade and α0 = 61 dB/decade [4].

In [5], the authors present an empirical and computationally inexpensive simulation model for IEEE 802.11p radio shadowing in urban environments. This model and its validation is based on real world measurements using IEEE 802.11p/DSRC devices, where they estimate the effects of building and other obstacle influence on radio communications between vehicles. The proposal considers building geometry and sender/receiver positions, and its model relies on building outlines, which are commonly available in modern geodatabases as Open Street Maps [6]. Furthermore, to keep the model computationally inexpensive, it only considers the line of sight between sender and receiver. By using the idea of [3] to detect the blocking effect of a building in the LOS between sender and receiver, the authors propose a generic model extension which is built on well-known propagation models, as shown in equation (4), where P represents the transmission power or receiver power, G represents the antenna gains, and L reflects the loss effects during transmission.

$$P_r[dBm] = P_t[dBm] + G_t[dB] + G_r[dB] - \sum L_x[dB] \tag{4}$$

In order to include the influence of the obstacles in the LOS between sender and receiver, the Equation (4) was extended to get Equation (5),

$$L_{obs}[dB] = \beta_n + \gamma d_m \tag{5}$$

where *Lobs* captures the additional attenuation caused by an obstacle in the transmission process, based on the number of times *n* that the border of the obstacle is intersected by the LOS, and the total length *dm* of this intersection. In Equation (5), *β* represents the attenuation caused by the outer wall of a building and γ is an approximation of the internal structure of a building. These parameters are used to adjust the model for managing the influence of different kinds of buildings when setting urban scenarios.

In order to improve VANET simulation results, the authors of [7] design and implement a more realistic radio propagation model, called the U.K. model (New University Kangaku) on NCTUns 6.0. This model was specifically proposed for VANET simulations in Tokyo, which represents a highly-populated environment. The New U.K. model considers both the line-of-sight (LOS) and non-line-of-sight (NLOS) conditions in its equations to compute the path loss. For the LOS condition, the distance between two vehicles, d, can be easily determined. In addition to the direct distance d, the computation of the LOS, shown in Equation (6), path loss involves several other parameters, such as the transmitter's height *(ht)*, the receiver's height *(hr)*, the widths of the streets in the scenario (Ws, W1, W2), the brake distance point *(db)*, and the frequency *f* in GHz. For NLOS path loss computation using Equation (7), *d1* represents the distance between a vehicle (which can be either a transmitter or a receiver) and the intersection, and *d2* represents the distance between another vehicle and such an intersection. The computation of the NLOS path loss also involves many parameters, such as the transmitter's height, the receiver's height, the brake point, and the frequency. Different from the LOS path loss computation, for the NLOS, the sum of *d1* and *d2 (d = d1 + d2)* is used as the distance between the transmitter and the receiver.

$$\begin{aligned} L_{LOS} &= \left\{7.2 + 7.1 \cdot \log\left(\frac{h_t \cdot h_r}{\lambda}\right)\right\} \cdot \log(d) + 28.3 \cdot \log\left(1 + \frac{d}{d_b}\right) \\ &-1.2 \cdot \log(f) - 19.6 \cdot \log(W_s) + 65.9 \end{aligned} \tag{6}$$

$$\begin{aligned} L_{NLOS} &= \left\{47.6 + 6.6 \cdot \log\left(\frac{h_t \cdot h_r}{\lambda}\right)\right\} \cdot \log(d) + \left\{89.1 - 33 \cdot \log\left(\frac{d_1}{\lambda}\right)\right\} \\ &\cdot \log\left(1 + \frac{d}{d_b}\right) + 19.9 \cdot \log(f) - 11.3 \cdot \log(W_1 \cdot W_2 + 2.8 \end{aligned} \tag{7}$$

The authors in [8] focus on the development of an adaptive algorithm to determine the condition of LOS between two vehicles, depending on the their position in the scenario's streets. Some characteristics of the transmitted signal may determine if the nodes can directly communicate with each other or not.

To this aim, three different cases were described:
• Vehicles on the same street: For two vehicles on the same street, there is an LOS between them,since no buildings interfere with the signal's path.
• Vehicles on different streets: If a couple of vehicles are located on different streets, it is necessary to check if there is an open area allowing communication between them (LOS). This involves identifying whether existing buildings completely interfere with the wireless signals. Success in communication, however, also depends on the distance between nodes and on the attenuation scheme used.
• Vehicles near junctions: Although there is no LOS between two vehicles, some electromagnetic phenomena of signals may help to obtain a successful communication. If the vehicles are on different streets, but near the corner where the streets meet, reflection, refraction or diffraction of signals over solid obstacles might sometimes produce such a positive effect. Some empirical results show that only vehicles close enough (< 20 m) to junctions are able to communicate with each other under NLOS conditions.

The following flowchart (Figure 3) shows the conditions used to determine if a packet is successfully received using the proposed model in [8].

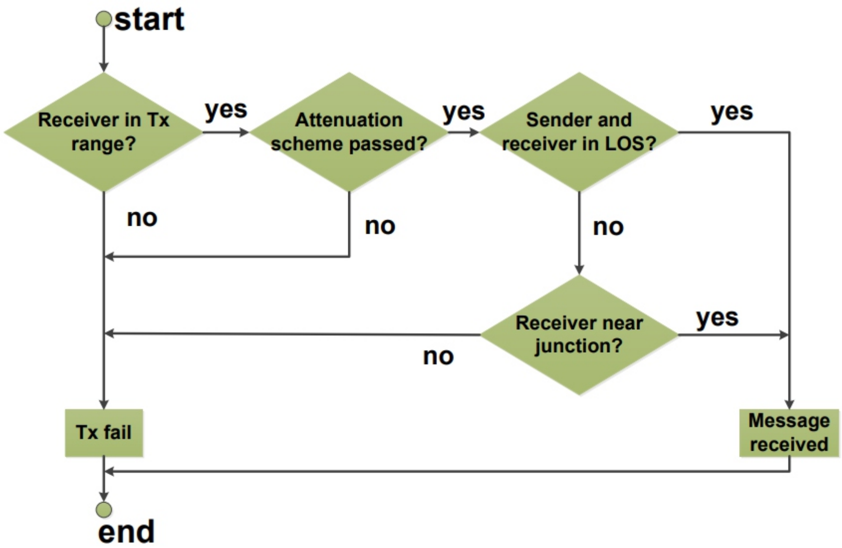


Figure 3. Flowchart of the visibility model proposed in [8].

As illustrated, the computation to determine if two vehicles are in the LOS is only done after two discarding steps. These steps are based both on the reception probability of packets and on the transmission range of nodes. This model tries to reduce the number of times that the LOS operations have to be done, since they are computationally expensive.

3. Evaluation of Building Attenuation Model

As stated in the Introduction, the aim of this paper is to provide a fair comparison among three different techniques to deal with the presence of obstacles in VANET simulation scenarios.Other research papers, such as [8,9], show the importance of considering the effect of obstacles in the overall performance of

VANET simulations. These works showed that using a channel propagation model that does not differentiate between LOS and NLOS situations leads to excessively optimistic performance evaluation. In this paper, we evaluate three techniques of VANET channel modeling for simulation to consider the influence of obstacles in the communication between vehicles, knows as the NLOS condition. The parameters that incorporate such an influence are: an empirically-computed attenuation factor, a complete blockage of the communication and a pre-computed attenuation factor.

From the models surveyed in the last section, we chose the empirical and inexpensive model of radio shadowing [5] as a realistic model that we use as a reference to compare the other two techniques, since this model is one of the most used for research in VANETs because of the following factors.

• It relies on real measurements taken with IEEE 802.11p devices.

• It is based on a simple modification of the free-space model, which captures the building attenuation with two easy-to-compute parameters.

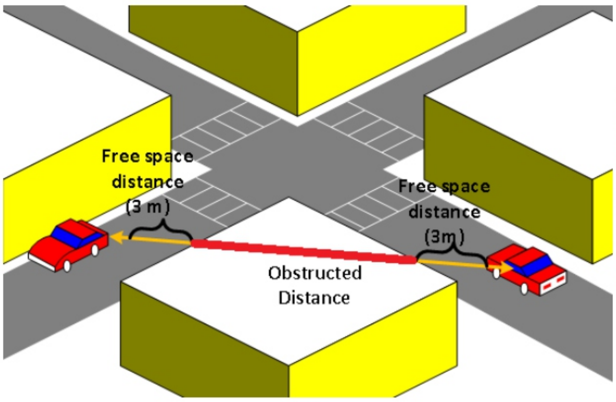
• Its implementation in a simulator is straightforward, and it is preloaded in VEINS, one of the most used VANET simulators.

As we showed in [10], any realistic propagation model can be used as a reference to compare other proposals, keeping the comparison results invariable. The reason is that specific propagation models designed for VANETs provide very close results with intermediate channel capacities.

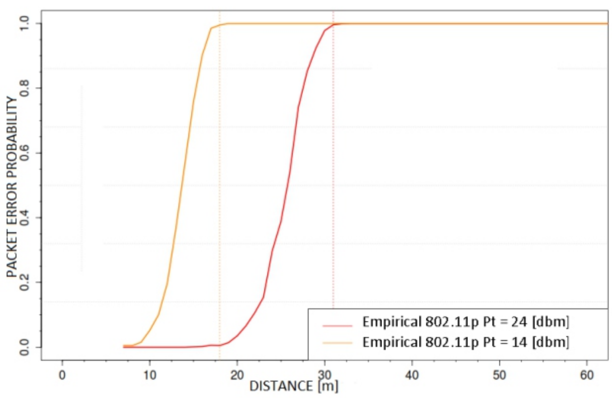
3.2. Total Blockage of Communications due to Obstacles

In this approach, considering the effects of obstacles in VANET simulations is done by assuming that communication between two nodes is completely attenuated. Some papers, such as [11] and [12], this focusing when evaluating their proposals.

To detect the presence of buildings in the LOS of a communication path, we used the same idea of [3], without having to compute all of the intersection points. However, sometimes, information about buildings is not available, and their influence is modeled in a conservative way, as is done in [13]. Figure 4b shows how the packet error probability (PEP) varies when the obstructed distance between two nodes increases, as shown in the scenario of Figure 4a



(a)



(b)

Figure 4. Packet error probability (PEP) behavior in an obstructed communication scenario between two vehicles near a corner. Empirical IEEE 802.11p channel model. Channel Capacity= 6 Mbps. Antenna sensitivity = −82 dbm. (a) Near-line-of-sight (NLoS) scenario at a corner; (b) PEP vs. distance.

Figure 4b shows that the value of the PEP reaches one at around an obstructed distance of 30 m,even when a robust modulation scheme is used (as happens in a low capacity channel). High power transmission and no interference are assumed. Thus, it would not be surprising that most of these types of communication fail. Assuming a total blockage of the signal due to obstacles, the overall performance would be close to that obtained using a more realistic channel model, as we will see in Section 4.

3.3. Pre-Computed Attenuation

The simulation of conventional channel models tends to be too slow, because of the great number of operations and searching algorithms implemented to determine an LOS or NLOS condition. As a result, some authors in [2] and [13] propose the use of an off-line file to store the values of the attenuation caused by the presence of buildings.

This approach requires the quantization of the movements on the streets in the simulation scenario.

For this, a quantization pace is used to map vehicle positions from continuous time to positions in discrete time. In the same way, continuous positions on a street can be replaced by discrete positions. The quantization error in this process depends on the size of the discretization step used. A small step will produce small errors, but also a large number of searching operations and, consequently, big files to store attenuation values. A more efficient strategy is performing discretization on the movements of vehicles when the simulation does not have to change the behavior of vehicles according to other events. Moreover, the map discretization is inevitable if the positions of the nodes change dynamically during the simulation. An issue with this map discretization is the need for vehicles to compute their discrete positions during the simulation. This might lengthen simulation times, since a searching algorithm is employed for this task.

For our work, we use a quantization process to determine the vehicles' positions in discrete moments. The pre-computed attenuation values should be stored in a fixed format to afford us an efficient retrieval of the nodes during simulation. We use an extended version of the output format proposed in [13], which includes the length of the quantization step.

The format for the output file is depicted in Figure 5a, where the first field, N, is the number of nodes; T is the simulation time; Step is the quantization step time; and Records is the number of discrete values for the two nodes in the simulation. The attenuation data are written in increasing order of source nodes, destination nodes and discrete time. In the example, at the top of Figure 5a, T is one and Step is equal to two. For instance, the first entry of the array $Att_{1,2,0}$ stores the attenuation that suffers a communication between Nodes 1 and 2 at Time 0, and the last entry $Att_{N-1,N,1}$ stores the attenuation factor between nodes $N-1$ and N at Time 1. The general procedure used to write the file can be found at the bottom of the same figure. The output of this straightforward mechanism is an array of pre-computed values whose length can be quickly calculated with the four first elements, N, T, Step and Records. If this file is written using a binary format to make the file lightweight, it is possible for the whole array of values to be read, by the simulator, with a single operation. Furthermore, due to the fact that the structure of the file is known, there is no necessity to implement a searching algorithm to find a specific value; it is only needed to know the time and the IDs of the nodes in the communication. The algorithm to compute the position in the array where the attenuation value (corresponding to the communication between nodes Src and Dst at time t) is located can be found in Figure 5b. This algorithm swaps the role of the source and destination nodes if their positions are not in increasing order. Then, the continuous time, t, is transformed to discrete time, based on the Step employed in this process (Lines 6 to 13). After that, the initial position, $prePos$, is computed for the recorded values corresponding to node Src (Lines 16 to 19). Finally (Line 20), the algorithm computes the offset value associated with the destination node Dst and the discrete time.

$N \quad T \quad Step \quad Records$
$Att_{1,2,0}$
$Att_{1,2,0.5}$
$Att_{1,2,1}$
...
$Att_{1,N-1,1}$
$Att_{1,N,0}$
$Att_{1,N,0.5}$
$Att_{1,N,1}$
$Att_{2,3,0}$
...
$Att_{N-1,N,0.5}$
$Att_{N-1,N,1}$

In general

...
$Att_{k,N,T}$
$Att_{i,j,t}$
$i > k, \quad j > i$
$t \in \{0, \frac{1}{Step}, .., 1 - \frac{1}{Step}, 1.., T - \frac{1}{Step}, T\}$

(a)

GETPOSITION(Src, Dst, t)

Require: Node IDs $1 \leq Src, Dst \leq N$ & $0 \leq t \leq T$
Ensure: Position Pos of pre-computed value for $Src \leftrightarrow Dst$ in t.

if $Src > Dst$ **then**
  SWAP(Src, Dst)
end if
$(t_{frac}, t_{int}) \leftarrow$ SEPARATE(t)
$i \leftarrow 1$
while DISCRETE $=$ **false do**
  if $t_{frac} \geq \frac{Step-i}{Step}$ **then**
    $t_{frac} = \frac{Step-i+1}{Step}$
    DISCRETE $\leftarrow$ **true**
  end if
  $i++$
end while
$t \leftarrow t_{frac} + t_{int}$
$prePos \leftarrow 0$
$preNode \leftarrow 1$
while $preNode < Src$ **do**
  $prePos \leftarrow prePos + (N - preNode) \cdot Records$
  $preNode++$
end while
$Pos \leftarrow prePos + (Dst - Src - 1) \cdot Records + tStep$
return Pos

(b)

Figure 5. Pre-computed attenuation file format with its corresponding localization value algorithm used in this work. (a) Output format; (b) Position computation algorithm.

4. Simulations and Results

In this section, we describe the most relevant aspects of the configuration in our simulations. We include a short description of the operation workflow of the channel and physical layers in a network simulator. Additionally, we describe the traffic application scenario in which we test the different building attenuation models and the simulation settings of the scenarios. This section ends up with an analysis of the results that we obtained from simulations using two different urban areas.

4.1. Operation of Channel and Physical Modules in a Network Simulator

A network simulator follows the logic of the protocol stack. This is, a set of modules defines the behavior of a node during the simulation. For a wireless node, the first two modules represent the wireless communication channel and the physical layer, which are the most relevant for the objective of our study.

Channel module: This is in charge of computing the packet signal attenuation caused by the distance.

This attenuation is typically called path loss or large-scale fading. It considers the attenuation caused by the presence of buildings. Then, the module introduces a variability effect in the computed power, which is called small-scale fading or just fading. Fading simulates effects, like reflectionand scattering.

If the power computed in the reception of a packet is lower than a minimum threshold, then the packet is discarded and it is not processed by the physical simulator's module. This threshold is typically obtained as a small fraction of the Nyquist noise associated with the channel (see Figure 6).

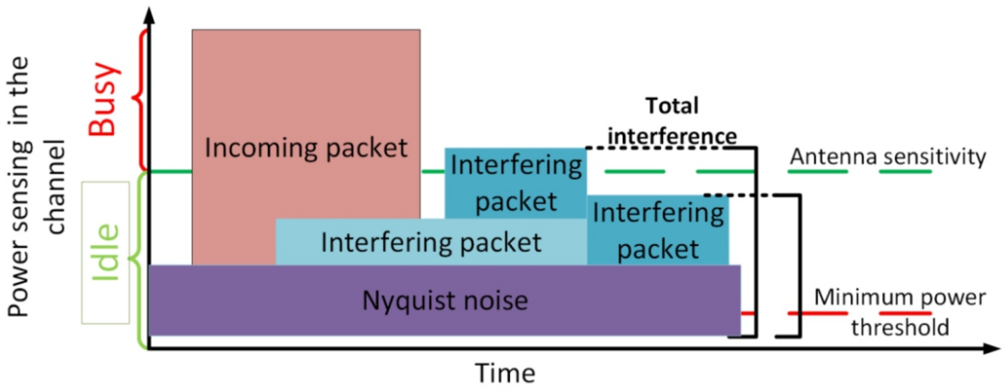


Figure 6. State of the channel and interference management of a wireless channel for a node in a network simulator [10].

Physical layer module: When a packet is received by this module, it is checked against the antenna sensitivity. If the power of the incoming packet is lower than the antenna sensitivity, then it is considered interference for other packets received during their duration. Otherwise, the packet is received, and its error probability is evaluated to determine if it was correctly demodulated. Multiple incoming packets with low power are grouped to get the total interference of the channel, as shown in Figure 6. In this work, we employ the analytical packet error model proposed in [14], which considers signal to interference and noise ratio (SINR), channel capacity and packet length to decide if a packet is erroneous. An additional task of the physical module is setting the state of the channel as busy or idle according to the sensed power compared to the antenna sensitivity level.

4.2. Characterization of the Application Scenario

The authors of [15] provide a classification of vehicular applications and their communication requirements. These categories are: safety, vehicular traffic efficiency and infotainment applications.

All of the efficiency-oriented applications (e.g., air pollution, noise level monitoring, etc.) require a continuous monitoring phase of the streets and city conditions. Our application scenario assumes traffic data generated by an efficiency application during the phase of collecting data only. The characteristics of the traffic are:

1. Vehicles obtain data from their sensors; they process such data and generate constant-length packets.
2. The packets are sent to the closest access point (AP). This is unicast and unidirectional traffic, since the information is useful only to the authority.
3. This kind of applications does not have important delay constraints as the safety related ones, so it is suitable to transport its traffic by delay-tolerant protocols.

4.3. Description of Simulation Scenarios

The simulation scenario consists of a multi-hop VANET, where we analyzed the performance of the channel modeling techniques described in Section 3. We used the Multi-Metric Map aware (MMMR)routing protocol [11], which is based on Greedy Perimeter Stateless Routing (GPSR) [16].

MMMR is a delay-tolerant protocol that improves the next forwarding node decision by employing four metrics; the distance to the destination, the vehicle density, the vehicle trajectory and the available bandwidth. This multi-metric parameter is obtained by each node, and it is used to find the neighboring node that is the candidate for being the next forwarding node. The scheme is self-configuring and able to adapt to the changing vehicle density in real time.

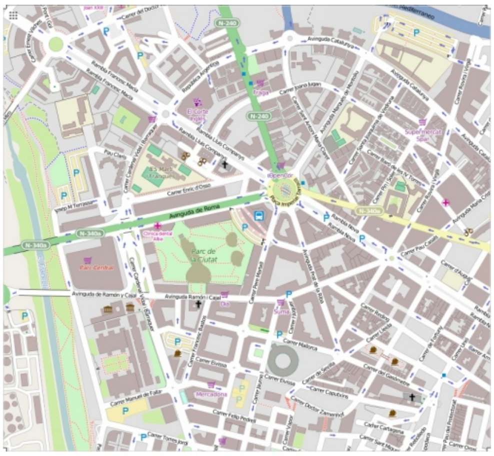
We carried out several simulations using the Estinet Network Simulator and Emulator [17]. Estinet includes the standard IEEE 802.11p and a simple and accurate way of designing VANET realistic scenarios. We used two real city areas of 1.5 and 2 km2 , obtained from the example district of Barcelona and the downtown district of Tarragona (see Figure 7), to evaluate two urban areas formed by streets with different densities of crossroads and buildings. In our scenarios, the Barcelona area has almost 1.3-times per square kilometer more streets and junctions than the Tarragona area. Furthermore, The example district of Barcelona has a high density of buildings compared to downtown Tarragona, specifically around four-times higher for the selected areas. Seeking to simulate a realistic scenario, the mobility model was obtained from CityMob for Roadmaps (C4R) [19], which is a mobility generator that uses the Simulation of Urban MObility (SUMO)engine [20]. C4R is able to import maps directly from OpenStreetMap [6] and to generate Network Simulator 2 (NS-2) compatible traces. C4R considers random origins and destinations for each vehicle. These points are located with higher probability in areas specified by the user. The path for a specific start and end point is computed through Dijkstra's algorithm in a directed graph (as a GPS-based navigation system computes a route). SUMO provides a realistic driver behavior in the route followed by a car during a movement simulation. We exported the NS-2 traces to Estinet, including the building information (orange lines, see Figure 7) using our own translating software, available at www.lfurquiza.com/research/estinet [18].

The scenarios also consist of fixed nodes, which are henceforth called access points or APs. The AP enables the connection, directly or by using multiple hops, to the services in the network. The Barcelona scenario only includes one AP (see Figure 7c), while the Tarragona area has six access points (see Figure 7d). We considered four vehicle densities of 67, 100, 133 and 167 cars per km2 . Each of these densities could represent different situations of a day; for instance, early morning, night, morning/afternoon and rush hour. The objective of using four arbitrary different densities is to test if the difference among the results coming from the models depends on the density of the scenario. A high vehicle density helps to avoid discarding packets, since a suitable next forwarding hop would surely be always available; however, data transmissions would be more prone to interference.

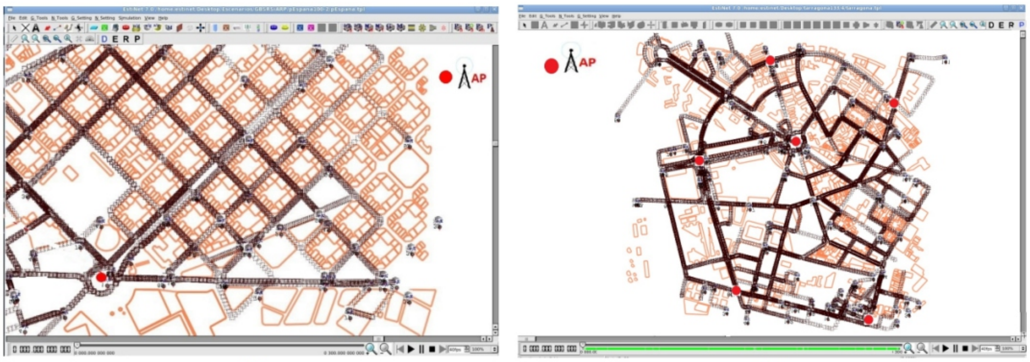
Each node during the simulations sends 1000-byte packets to the destination APs, during 300 s. In the case of Barcelona, the inter-packet time follows a uniform distribution between 2 and 6 s that has a mean of 4 s. On the other side, in the Tarragona scenario, we consider that this time is exponentially distributed with a mean of 4 s, but truncated between 1 and 10 s.

All of the results are presented with confidence intervals (CI) of 95%, obtained from ten simulations per each density value and attenuation model using ten different movement traces. This means that we generated ten different movement traces per density and scenario. We use each of them to evaluate the three attenuation models.

(a)



(b)



(c) (d)

Figure 7. Simulation scenarios from two cities of Catalonia, Spain. (a) Example district of Barcelona from OSM; (b) downtown area of Tarragona from OSM; (c) barcelona simulated scenario with an access point (AP); buildings from OpenStreetMap are included. (d) Tarragona simulated scenario with six APs; buildings from OpenStreetMap are included.

Table 1. Simulation settings.

Parameter	Value	
Map Zone	**Example District of Barcelona**	**Downtown of Tarragona**
Area	1.5 km × 1 km	1.6 km × 1.3 km
Number of junctions	712 (475 × km^2)	783 (373 × km^2)
Number of streets (edges)	920 (613 × km^2)	993 (472 × km^2)
Number of buildings (polygons)	2216 (1477 × km^2)	720 (346 × km^2)
Number of nodes	100, 150, 200 and 250 vehicles	133, 200, 266 and 333 vehicles
Inter-packet generation time	t∼U(2,6) s E(t) = 4 s	t∼Exp(mean = 4, min = 1, max = 10) s
Maximum packet size	1000 bytes	
Path loss model	Empirical model of IEEE 802.11p radio shadowing [5]	
Building models	• Realistic [5] $\alpha = 9db/wall \quad \beta = 0.4db/m$ • Total block of signal • Pre-computed attenuation (Figure 5b) $Step = 3$	
Fading model	Ricean (LOS) and Rayleigh (not in LOS)	
Power transmission	23 dbm	
Receiving sensing	−82 dbm (∼ 400 m in LOS)	
Mobility generator	SUMO [20] / C4R [19]	
MAC specification	IEEE 802.11p	
Bandwidth	6 Mbps	
Simulation time	300 s	
Routing protocol	MMMR [11]	
GPS precision	10 m	

Table 1 summarizes the main simulation settings.

4.4. Simulation Results

In this section, we present some results from comparing the simulations of the three aforementioned attenuation models: realistic, total blockage of signal and pre-computed attenuation. The evaluation is focused on four widely-used metrics applied to the performance analysis of VANET routing protocols. These metrics are the percentage of packet losses, average delay, average number of hops and average number of neighbors. Figures 8 and 9 illustrate these results for the four node densities in the two scenarios.

The mixed ANOVA statistical test [21] was employed to check if the performance metrics differences among attenuation models depends on the vehicle density in the evaluated area. There might be a relationship since a great number of nodes may generate more collisions and higher levels of interference.

We use mixed ANOVA, because for each vehicle density, the same ten vehicle movements were used test the attenuation models. For mixed ANOVA, our data are organized into twelve groups that are obtained by combining the three building attenuation models with the four vehicle densities in the simulation. All groups have the same number of elements, which allows one to have a balanced test. The outcome of the statistical test is a probability called the p-value, which is compared with a threshold named the significance level. If the p-value is lower than the significance level, then the performance results of the system are related to both the attenuation model and the vehicle density.

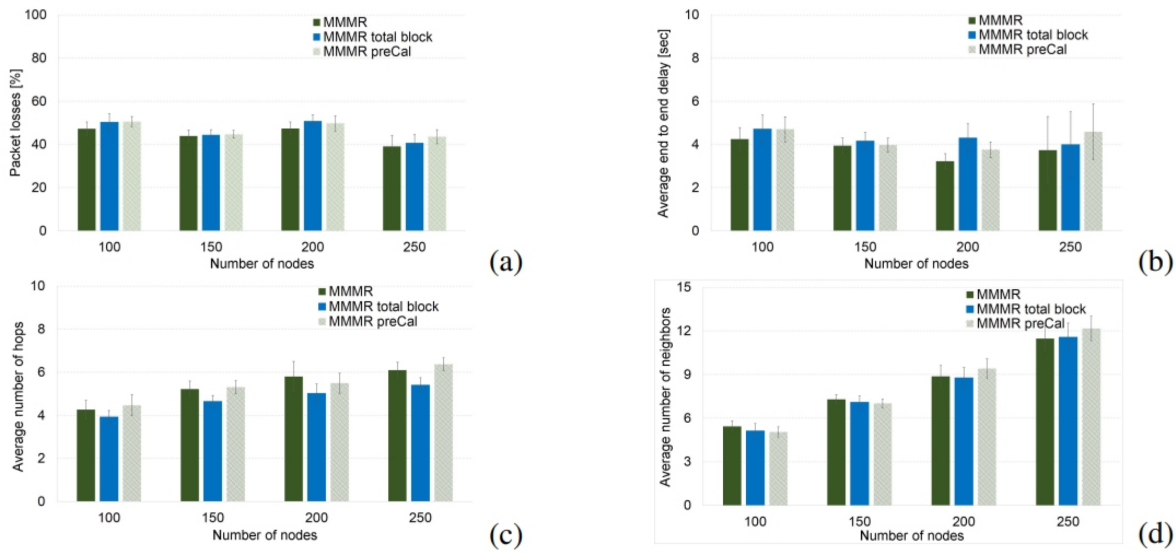


Figure 8. Building attenuation model comparison (Barcelona scenario) (CI 95%). (a) Percentage of packet losses; (b) average end-to-end delay; (c) average number of hops; (d) average number of neighbors.

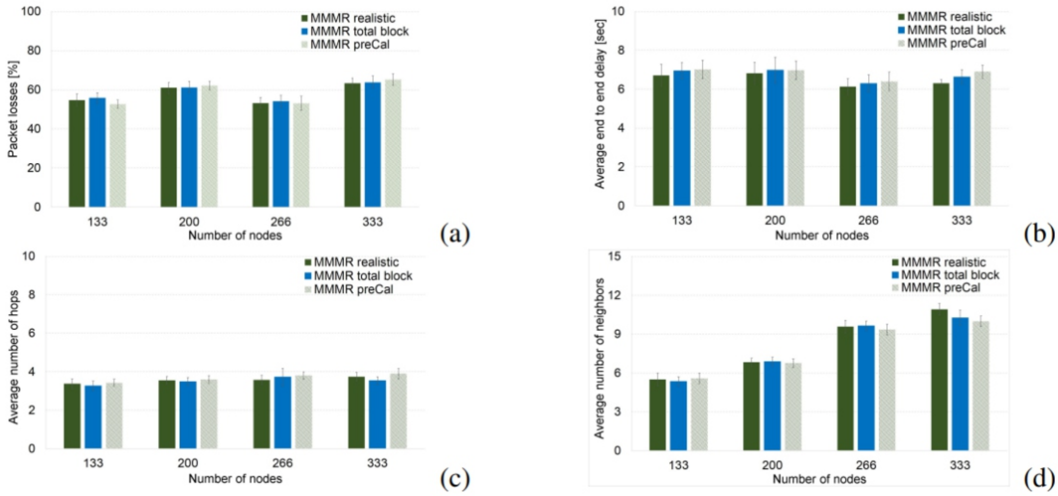


Figure 9. Building attenuation models comparison (Tarragona scenario) (CI 95%). (a) Percentage of packet losses; (b) average end-to-end delay; (c) average number of hops; (d) average number of neighbors.

The p-values, which we obtained from checking the dependency between the attenuation model and the vehicle density, are depicted in Table 2, where the F-ratio is the result of the F-test statistical test [22], df represents the degrees of freedom used to obtain the p-value and partial η 2 is a measure of the effect size. According to the results of the Mixed-ANOVA, we can conclude that there is no significant interaction between the attenuation models and the vehicle density, since the p-value is lower than 0.05 in none of the performance metrics for both scenarios, except for the average number of neighbors. Mixed ANOVA assumes that the data come from normally-distributed populations with a homogeneous variance and similar covariance and sphericity (differences between all possible pairs of groups are equal). Some of the twelve data groups of our performance results violate some of these assumptions, especially normal distribution and homogeneous variance. However, our test results can be considered reliable, because ANOVA is robust to these assumptions [23], especially under the balance number of elements in each group [24].

Table 2. p-Values from Mixed Analysis of Variance (ANOVA) test for the relation between the attenuation model and vehicle density. (a) Barcelona scenario. 1 AP. Uniformly distributed traffic; (b) Tarragona scenario. 6 APs. Exponentially distributed traffic. * Degrees of freedom correction applied, because of the sphericity assumption violation.

Performance Metric	F Value	df	*p*-Value	Partial η^2	Is It Significant (*p*-Value < 0.05)?
Packet losses *	1.204	4.2, 50.47	0.321	0.091	NO
Average delay	1.632	6, 72	0.151	0.120	NO
Average no. of hops	1.644	6, 72	0.147	0.121	NO
Average no. of neighbors	3.833	6, 72	0.002	0.242	YES

(a)

Performance Metric	F Value	df	p-Value	Partial η^2	Is It Significant (p-Value < 0.05)?
Packet losses	1.997	6, 72	0.077	0.143	No
Average delay	0.572	6, 72	0.751	0.041	No
Average No. of hops *	1.013	4.585, 55.02	0.415	0.078	No
Average No. of neighbors	2.657	6, 72	0.022	0.181	Yes

(b)

When there is not a significant dependency between the factors in the mixed ANOVA, as happens in our study between attenuation models and vehicle density for the percentage of packet losses, average end-to-end delay and average number of hops, the post hoc tests, i.e., repeated measures ANOVA [21], to study the attenuation model effects do not need to differentiate between the vehicle densities. On the other hand, the post hoc tests for the average number of neighbors metric have to be done for each vehicle density independently. According to mixed ANOVA results, the differences among these metrics depend on the vehicle density used in the simulation.

Due to the post hoc test (ANOVA with repeated measures), it is required to meet the same assumptions of mixed ANOVA, and even when this test is robust to violations of these assumptions, we decided to apply the equivalent non-parametric tests to assess the difference among the attenuation models, because our data conform to their requirements. For our data, both kinds of tests, parametric and non-parametric, agree with the same decision in all of the post hoc tests performed in this work. There is not a well-accepted non-parametric statistical test equivalent to mixed ANOVA. Table 3 shows the p-values obtained from the Friedman test [22], which is the non-parametric version of ANOVA for repeated measures. It is used to know if there are statistically significant differences among the building attenuation models.

Table 3. p-Values from the Friedman test to determine the effect of the building attenuation model. (a) Barcelona scenario: one AP; uniformly-distributed traffic. (b) Tarragona scenario: six APs; exponentially-distributed traffic.

Performance Metric	Vehicle Density	Test Statistic	df	p-Value 2 Sides	Is It Significant p-Value < 0.05?
Packet losses	all together	10.55	2	0.005	Yes
Average delay	all together	18.95	2	0.0001	Yes
Average No. of hops	all together	42.35	2	0.0001	Yes
Average No. of neighbors	67	9.8	2	0.007	Yes
	100	2.6	2	0.273	No
	133	9.614	2	0.008	Yes
	166	9.614	2	0.008	Yes

(a)

Performance Metric	Vehicle Density	Test Statistic	df	p-Value 2 Sides	Is It Significant p-Value < 0.05?
Packet losses	all together	1.55	2	0.461	No
Average delay	all together	10.4	2	0.006	Yes
Average No. of hops	all together	22.05	2	0.0001	Yes
Average No. of neighbors	67	5.6	2	0.061	No
	100	1.4	2	0.497	No
	133	4.66	2	0.097	No
	166	5.59	2	0.061	No
	all together	4.899	2	0.086	No

(b)

As the reader can notice, none of the p-values of Table 3a (Barcelona scenario) are higher than the significance threshold of 0.05, except the average number of neighbors for a vehicle density of 100 cars/km2 . Hence, this means that there is a statistically significant difference among the distribution of the results obtained when the simulation uses different building attenuation models, excluding the average number of neighbors for a vehicle density of 100 cars/km2 . All of these differences, i.e., the percentage of packet losses, average end-to-end delay and average number of neighbors, are not detected in the performance results of the Tarragona scenario. It can be seen in Table 3b that the p-values show significant differences in average delay and average number of hops. We would like to point out that even when the p-value of mixed ANOVA (to test the interaction between the attenuation models and the vehicle density in the Tarragona scenario; see Table 2b) is significant, the Friedman tests performed per vehicle density do not detect any difference in any of the cases. One reason that explains this result is that the effect size of the interaction is lower in the Barcelona scenario.

We used a pairwise comparison to determine the models among which there exists a difference in terms of the performance results. Table 4 shows the p-values for this comparison by using the Wilcoxon statistical test [22]. We employed the Bonferroni correction in these pairwise comparison. That is, the new significance threshold is 0.017, obtained by dividing 0.05 by the number of options compared (i.e., 0.05/3).

Table 4. p-Values of Wilcoxon signed rank test for a pairwise comparison of the building attenuation model effect. (a) Barcelona scenario. 1 AP. Uniformly distributed traffic; (b) Tarragona scenario 6 APs. Exponentially distributed traffic.

Performance Metric	Pairwise	Standardized Test Statistic	p-Value 2 Sides	Is the Difference Significant (p-Value < 0.017)?	Median of Differences
Packet losses	Realistic, total block	2.983	0.003	Yes	−2.3780%
	Realistic, pre-computed	3.602	0.0001	Yes	−3.0614%
	Total block, pre-computed	1.423	0.155	No	−0.91%
Average delay	Realistic, total block	4.019	0.0001	Yes	−0.5310 s
	Realistic, pre-computed	3.360	0.001	Yes	−0.4391 s
	Total block, pre-computed	-0.565	0.572	No	0.0136 s
Average number of hops	Realistic, total block	−5.309	0.0001	Yes	0.66 hops
	Realistic, pre-computed	1.559	0.119	No	−0.11 hops
	Total block, pre-computed	5.242	0.0001	Yes	−0.78 hops

Average number of neighbors 67	Realistic, total block	−2.091	0.037	No	0.285 nodes
	Realistic, pre-computed	−2.599	0.009	Yes	0.485 nodes
	Total block, pre-computed	−0.153	0.878	No	−0.075 nodes
Average number of neighbors 133	Realistic, total block	−0.816	0.415	No	0.06 nodes
	Realistic, pre-computed	1.734	0.83	No	−0.315 nodes
	Total block, pre-computed	2.499	0.012	Yes	−0.565 nodes
Average number of neighbors 166	Realistic, total block	−2.803	0.005	Yes	0.39 nodes
	Realistic, pre-computed	1.020	0.308	No	−0.185 nodes
	Total block, pre-computed	2.497	0.013	Yes	−0.525 nodes

(a)

Performance Metric	Pairwise	Standardized Test Statistic	*p*-Value 2 Sides	Is the Difference Significant	Median of Differences
Average delay	Realistic, total block	2.285	0.022	No	−0.3084 s
	Realistic, pre-computed	3.428	0.001	Yes	−0.2893 s
	Total block, pre-computed	0.82	0.412	No	−0.0919 s
Average number of hops	Realistic, total block	−2.393	0.016	Yes	0.836 hops
	Realistic, pre-computed	2.796	0.005	Yes	−0.1464 hops
	Total block, pre-computed	3.549	0.0001	Yes	−0.2651 hops

(b)

From Table 4a for the Barcelona scenario, we can see that the results of the performance metrics are significantly different (see Rows 1,4 and 7 and 16) if we model the effects of building presence as the total absence of communication, compared with the results obtained with a realistic channel model,specifically designed for VANETs. Comparison between these two models in the Tarragona scenario shows significant differences in the average number of hops (see the fourth row in Table 4b). Moreover, the total signal blocking always shows the most conservative behavior. This can be noticed in the values of the column of the median of the differences in both Table 4a and 4b). This behavior makes total sense, since this model involves nodes sensing fewer neighbors (vehicles cannot detect nodes behind obstacles in any case under the total-blockage attenuation model). Additionally, the absence of communications avoids the construction of paths that in a realistic approach may be feasible. Consequently, packets need to be stored in nodes for longer periods until finding a forwarding node, so the percentage of packet losses increases due to the timeouts. Notice that in the Barcelona scenario, the differences in the average number of neighbors are only detected in the high density case.

This behavior is explained by the communications between obstructed nodes not being the rule, and most of them entail high error probabilities. As a consequence, the differences in the performance metrics are small in our simulations scenarios, as is shown in the column of the reported medians.

Regarding the comparison between the realistic channel model and the pre-computed attenuation approach, it can be noticed (see Rows 8, 14 and 17 in Table 4a) that there is no statistically significant difference in the average number of hops or neighbors. Nevertheless, there are discrepancies in the percentage of packet losses and in the average end-to-end delay (see Rows 2, 5 of Table 4a). Notice that there is a difference in the average number of neighbors for a density of 67 vehicles/km2 . Similar results are obtained from Tarragona scenario (See Table 4b), but in this case, there are only significant differences in the average end-to-end delay and the average number of hops.

For both scenarios, the median of the pre-computed attenuation performance results are not so far from the medians in the realistic scenario. The presence of differences between the aforementioned models is a consequence of the discretization process done in the pre-computed attenuation approach.

Lastly, total blockage and pre-computed building attenuation models are compared in order to get an idea of the existing differences between these approaches and the realistic channel model. The reader can observe from Table 4a in the case of the Barcelona scenario and Table 4b for Tarragona that there are statistical differences in the average number of hops of a packet when traveling to reach the access point. Furthermore, differences in the average number of nodes between these two models appear only for the two highest densities applied in this work (133 and 166 nodes/km2) in the Barcelona scenario.

Hence, the results obtained with these two models could be similar, at least in the percentage, to the packet losses and delay.

To conclude this section, we summarize the major features and results drawn from our evaluation:

• We have simulated three building attenuation models used in the literature for VANETs (i.e., realistic, total blocking of the signal and the precomputed attenuation model) under two realistic urban scenarios that have different vehicle densities, junctions and buildings. The study considered four different vehicle densities with ten different mobility traces per density. Moreover, different probability distributions for traffic generation were used in each urban scenario.

• The results of statistical tests carried out with four performance metrics (percentage of packet losses, end-to-end delay, average number of hops and neighbors) show differences when employing different attenuation models.

• A complete attenuation of the signal due to the presence of buildings, and pre-computed attenuation models in our simulations can be considered for both scenarios as pessimistic bounds for all performance metrics.

• We did not find that the differences in the performance metrics are affected by the vehicle density employed during the simulation, except in the case of the average number of neighbors. This metric does not differ from the realistic one with intermediate vehicle density scenarios.

These are promissory results, because they lead to the idea of using the total signal attenuation model or pre-computed attenuation files when a realistic attenuation model is difficult to use; for instance, for map areas where building information is not available and a complete blockage outside the street has to be assumed. Pre-computed attenuation files would be a good alternative when scenarios involve a medium or high data traffic load, especially in preliminary studies. Such scenarios require that algorithms to calculate the level of attenuation caused by buildings are executed too many times. This can cause important time consumption due to the number of operations, even when such algorithms are efficient.

5. Conclusions

A statistical analysis has been performed in this work about the simulation of multi-hop vehicular ad hoc networks and, particularly, on how their performance metrics vary according to the attenuation effects obtained by modeling the presence of buildings in a VANET scenario. Our study compares three strategies to model the influence of buildings on the communication between vehicles. These strategies model this influence as the full attenuation of the communication signals, as a number of offline computed values of attenuation and as an inexpensive and accurate realistic propagation scheme [5].

The results we obtained support that the performance metric scores depend on the building attenuation model used in the simulations. Hence, the research community should use a realistic propagation model when possible.

The differences in the performance reached with the realistic model, compared with the other two models, are at the maximum 3% for the percentage of packet losses and 0.5 s for end-to-end delay in our two different simulation scenarios, for the traffic-efficiency application tested in this work. Furthermore, we could not find any statistical relationship between the vehicle density in the scenario (which may include a higher data traffic load) and the building attenuation model used, except by the average number of neighbors.

Obstructed communications with higher capacity channels are less probable, and consequently, the resulting gap between a realistic building attenuation model and a total blockage of the signal should decrease when using higher capacity channels. Future work may include testing the different channel capacities available in IEEE 802.11p in order to find out if the assumption of total attenuation differs from the realistic scenario for any channel capacity and average rate in the scenario. Furthermore, tests to empirically obtain the curve of the differences among the models in different types of scenarios is in the future work plan. Additionally, we are interested in testing different ways to map the continuous time into the discrete time for the offline attenuation files. Furthermore, the impact of the discretization step used to build the attenuation file (that currently has a size of 110 MB for a simulation of 300 second with 250 nodes) may be measured in a future study, since there would be a trade-off between accuracy and file size.

Another future work might involve assessing the differences among the three techniques explained in this paper, but applying them to delay sensitive applications, like dissemination of warning messages for safety purposes. This could be done by using more appropriate metrics, such as jitter, time to cover an area, number of notified vehicles, etc. We are planning to use real mobility traces, such as the ones provided in [25] or [26], to leverage a trustworthy evaluation of this kind of applications, especially in the analysis of how long and how fast an emergency message can be spread over a large area.

Acknowledgments

This work was partly supported by the Spanish Government through projects TEC2013-47665-C4-1-R Emergency Response In Smart Communities "EMRISCO" and TEC2010-20572-C02-02 Continuity of Service, Security and QoS for Transportation Systems "CONSEQUENCE". Luis Urquiza-Aguiar is the recipient of a full scholarship from the Secretaria Nacional de Educación Superior, Ciencia y Tecnología (SENESCYT) with the sponsorship of Escuela Politécnica Nacional (EPN) (Ecuador). Carolina Tripp-Barba would like to thank the support of the "Universidad Autónoma de Sinaloa (Mexico)".

Author Contributions

Luis Urquiza-Aguiar developed the proposal of the study, made the changes in the simulator code, carried out the statistical analysis of results and took care of most of the writing. Carolina Tripp-Barba carried out the simulations, helped in the analysis of the simulation results and wrote several sections. José Estrada-Jimenéz assisted with the Ât' review of the simulation results, statistical tests and made corrections throughout the manuscript Mónica Aguilar Igartua guided the entire process as well as the analysis of the proposal and made corrections. All authors contributed to the analysis of the simulation results, in order to provide a more comprehensible final version.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Hartenstein, H.; Laberteaux, K.; Ebrary, I. VANET: Vehicular Applications and Inter-Networking Technologies; John Wiley & Sons, Ltd.: Chichester, UK, 2010.

2. Scopigno, R.; Cozzetti, H. Signal Shadowing in Simulation of Urban Vehicular Communications. In Proceedings of the 2010 6th International Conference on Wireless and Mobile Communications (ICWMC), Valencia, Spain, 2010; pp. 131–138.

3. Nagel, R.; Eichler, S. Efficient and Realistic Mobility and Channel Modeling for VANET Scenarios Using OMNeT++ and INET-framework. In Proceedings of the 1st International Conference on

Simulation Tools and Techniques for Communications, Networks and Systems & Workshops (Simutools '08), Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, Brussels, Belgium, 2008; pp. 89:1–89:8.
4. Chia, S.T.S.; Snow, P. Characterising radio-wave propagation behaviour at 1700 MHz for urban and highway microcells. In Proceedings of the Micro-Cellular Propagation Modelling, IEEE Colloquium, London, UK, 1992; pp. 11:1–11:4.
5. Sommer, C.; Eckhoff, D.; German, R.; Dressler, F. A computationally inexpensive empirical model of IEEE 802.11p radio shadowing in urban environments. In Proceedings of the 2011 Eighth International Conference on Wireless On-Demand Network Systems and Services (WONS), Bardonecchia, Italy, 26–28 January 2011; pp. 84–90.
6. Open Street Maps. Available online: http://www.openstreetmap.org/ (accessed on 28 October 2014).
7. Wang, S.Y.; Wang, P.F.; Li, Y.W.; Lau, L.C. Design and implementation of a more realistic radio propagation model for wireless Vehicular Networks over the NCTUns network simulator. In Proceedings of the 2011 IEEE Wireless Communications and Networking Conference (WCNC),Cancun, Quintana Roo, Mexico, 28–31 March 2011; pp. 1937–1942.
8. Martinez, F.J.; Fogue, M.; Toh, C.K.; Cano, J.C.; Calafate, C.T.; Manzoni, P. Computer Simulations of VANETs Using Realistic City Topologies. Wirel. Pers. Commun. 2013, 69, 639–663.
9. Gozalvez, J.; Sepulcre, M.; Bauza, R. Impact of the radio channel modelling on the performance of VANET communication protocols. Telecommun. Syst. 2010, 50, 149–167.
10. Urquiza, L.; Tripp, C.; Martin, I.; Aguilar, M. Propagation and Packet Error models in VANET simulations. IEEE Latin Am. Trans. 2014, 12, 499–507.
11. Tripp-Barba, C.; Urquiza-Aguiar, L.; Igartua, M.A.; Rebollo-Monedero, D.; Mezher, A.M.; de la Cruz Llopis, L.J.; Aguilar-CalderÃ¸sn, J.A. A Multimetric, Map-Aware Routing Protocol for VANETs in Urban Areas. Sensors 2014, 14, 2199–2224.
12. Lochert, C.; Mauve, M. Geographic routing in city scenarios. ACM SIGMOBILE Mob. Comput. Commun. Rev. 2005, 9, 69–72.
13. Mezher, A.M.; Oltra, J.J.; Aguiar, L.U.; Paredes, C.I.; Barba, C.T.; Igartua, M.A. Realistic Environment for VANET Simulations to Detect the Presence of Obstacles in Vehicular Ad Hoc Networks. In Proceedings of the 11th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks PE-WASUN '14, ACM, Montreal, QB, Canada, 21–26 September 2014; pp. 77–84.
14. Abrate, F.; Vesco, A.; Scopigno, R. An Analytical Packet Error Rate Model for WAVE Receivers. In Proceedings of the 2011 IEEE Vehicular Technology Conference (VTC Fall), San Francisco, USA, 5–8 September 2011; pp. 1–5.
15. Caloca, C.; Garcia Macias, J.A. Adaptive Solutions in Multihop Communication Protocols for Vehicular Ad Hoc Networks. In Advances in Vehicular Ad-Hoc Networks: Developments and Challenges; Watfa, M., Ed.; IGI Global: Hershey, PA, USA, 2010; pp. 301–322.
16. Karp, B.; Kung, H.T. GPSR Greedy perimeter stateless routing for wireless networks. In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking—MobiCom '00, New York, NY, USA, 6–11 August 2000; pp. 243–254.
17. Estinet-Technologies. EstiNet Network Simulator and Emulator. Available online: http://www.estinet.com/products.php?lv1=1&sn=2 (accessed on 28 October 2014).
18. Software to VANET simulations. Available online: http://www.lfurquiza.com/research/estinet (accessed on 28 October 2014).
19. Fogue, M.; Garrido, P.; Martinez, F.J.; Cano, J.C.; Calafate, C.T.; Manzoni, P. A Realistic Simulation Framework for Vehicular Networks. In Proceedings of the 5th International ICST

Conference on Simulation Tools and Techniques Simutools '12, Brussels, Belgium, March 2012;pp. 37–46.
20. Krajzewicz, D.; Erdmann, J.; Behrisch, M.; Bieker, L. Recent Development and Applications of SUMO—Simulation of Urban MObility. Int. J. Ad. Syst. Meas. 2012, 5, 128–138.
21. Maxwell, S.E.; Delaney, H.D. Designing Experiments and Analyzing Data: A Model Comparison Perspective; Psychology Press: London, UK, 2004; Volume 1.
22. Sheskin, D. Handbook of Parametric and Nonparametric Statistical Procedures, 2nd ed.; Chapman & Hall CRC: Boca Raton, FL, USA, 2000.
23. Lewicki, P.; Hill, T. Statistics: Methods and Applications, online ed.; Statsoft: Tulsa, OK, USA, 2013.
24. Wilcox, R. Robustness in ANOVA. In Applied Analysis of Variance in Behavioral Science. Statistics: Textbooks and monographs; Marcel Dekker: New York, NY, USA, 1993; pp. 345–364.
25. Uppoor, S.; Trullols-Cruces, O.; Fiore, M.; Barcelo-Ordinas, J.M. Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset. IEEE Trans. Mob. Comput. 2014, 13,1061–1075.
26. Pigné, Y.; Danoy, G.; Bouvry, P. A Vehicular Mobility Model Based on Real Traffic Counting Data. In Communication Technologies for Vehicles of Lecture Notes in Computer Science series; Strang, T., Festag, A., Vinel, A., Mehmood, R., Rico Garcia, C., Röckl, M., Eds.; Springer: Berlin Heidelberg, Germany, 2011; Volume 6596, pp. 131–142.

Piezoelectric Polymer-Based Collision Detection Sensor for Robotic Applications†

J. Michael Wooten 1, David M. Bevly 2 and John Y. Hung 2,*

1 Applied Research Laboratories, University of Texas, Austin, TX 78758, USA;

2 College of Engineering, Auburn University, Auburn, AL 36849, USA;

ABSTRACT

The authors present a large area collision detection sensor utilizing the piezoelectric effect of polyvinylidene fluoride film. The proposed sensor system provides high dynamic range for touch sensation, as well as robust adaptability to achieve collision detection on complex-shaped surfaces. The design allows for cohabitation of humans and robots in cooperative environments that require advanced and robust collision detection systems. Data presented in the paper are from sensors successfully retrofitted onto an existing commercial robotic manipulator.

Keywords: piezoelectricity; polyvinylidene; polymer film; collision detection; collision sensor robotic manipulator

1. Introduction

Recent sensing work with polyvinylidene fluoride (PVDF) film based sensors includes tactile applications related to robotic skin for finger tips [1], large area coverage [2], stress sensing for shock wave measurements [3], deflection sensing [3], object identification [4], smart textiles [5] and power harvesting [6] to name a few. The use of PVDF for energy generation from nano generators [7] is particularly compelling. Advances in manufacturing processes have also increased viability of PVDF as a flexible and adaptable sensor solution for complex surfaces through MEMS based fabrications [8] as well as the use of organic transistors to create a highly sensitive pressure sensor [2]. More recently,nano structure work utilizing nano ribbons has been shown to drastically increase the charge coefficient characteristics of PVDF [9]. Further work with micro- and nano-structurization shows strong promise for flexible tactile sensation viability of PVDF [10].

The majority of PVDF sensing applications traditionally rely on either a film membrane based strain sensation [11] or film on rigid substrate pressure sensation [2]. Both of these methods present challenges when applied to collision detection against typical robotic manipulator structures, which require sensing over large surface areas, over complex shapes, and a large range of impact forces. The membrane approach has practical shortcomings for large-area and high-impact applications because of inherent physical limitations (e.g., puncture of membrane) and a high level of design complexity for large-area sensor applications and networks. Pressure based designs require more complex electronics and construction to achieve flexibility because of the low signal response and need for a stiff substrate to achieve pressure dynamics.

In this paper, a novel sensor design is suggested utilizing a flexible substrate that allows the PVDF film to operate in a pseudo-membrane configuration, as shown in Figure 1a. Complex environments associated with modern robotics require tactician for control feedback, safety concerns, and perception to name a few [12]. The primary concern of the presented sensor design is safety and risk mitigation for complex environment as such the new sensor design can achieve high dynamic range, uses simplified electronics, and can be robustly applied over a variety of surface shapes. Sensor robustness means that the sensor can operate in a wide variety of environments including, but not limited to high and low impacts, nonhuman inhabited environments. The proposed sensor construction allows for complex shape and non-planar surface applications. The design is intended for safety and control applications related to human-robotics interaction in cooperative environments, arm autonomy in high degree-of-freedom (DOF) arms in changing scenarios, and technology redundancy to minimize risk related to collision. Current safety standards limit the amount of force that a robot can impart to a human being as 150 N [13]. The proposed sensor provides a dynamic sensing range of 5 N to greater than 200 N in an effort to detect a state of collision before significant force has been imparted to the object.

2. Design Methodology

The sensor's mechanical structure, materials, physical properties, and electronic instrumentation are explained here. Additionally, the key design considerations are discussed.

2.1. Materials and Structures

PVDF is a piezoelectric and pyroelectric polymer commercially available in thin (<0.1 mm) sheets. Commercial uses of PVDF include, but are not limited to force sensors, accelerometer applications, high-frequency resonators, and deflection sensing [3]. Piezoelectric PVDF film is created from homopolymer PVDF sheets that are stretched, heated and simultaneously poled by application of a high electric field across the film [14]. The stretching and heat annealing processes align the polymer chains within the PVDF, and the high electric field orients the dipoles of the chains to create polarization in the film [14]. Poling the film enables the polymer to generate charge when stressed by heat or physical stress because tensile stress in the film causes the dipoles to flip, creating a charge gradient that generates an electrical displacement. The piezoelectric and pyroelectric effects of the polymer do not degrade over time (< 1% of original value) insuring longterm reproducibility of sensations as long as the material is kept below approximately 90 ◦C depending on PVDF construction. (At high temperature, the poles of the polymer become randomly oriented, eliminating the charge gradient [15].)

The proposed collision sensor is constructed of two PVDF film elements oriented with poles out of phase, adhered to a flexible elastic compressible substrate. The trilayer sensor is attached to the targeted surface, shown in Figure 1a. Element 1 and Element 2 represent the two PVDF film elements. In contrast, diagrams for traditional membrane and pressure based sensing are included and explained in the following subsection as Figure 1b,c, respectively.

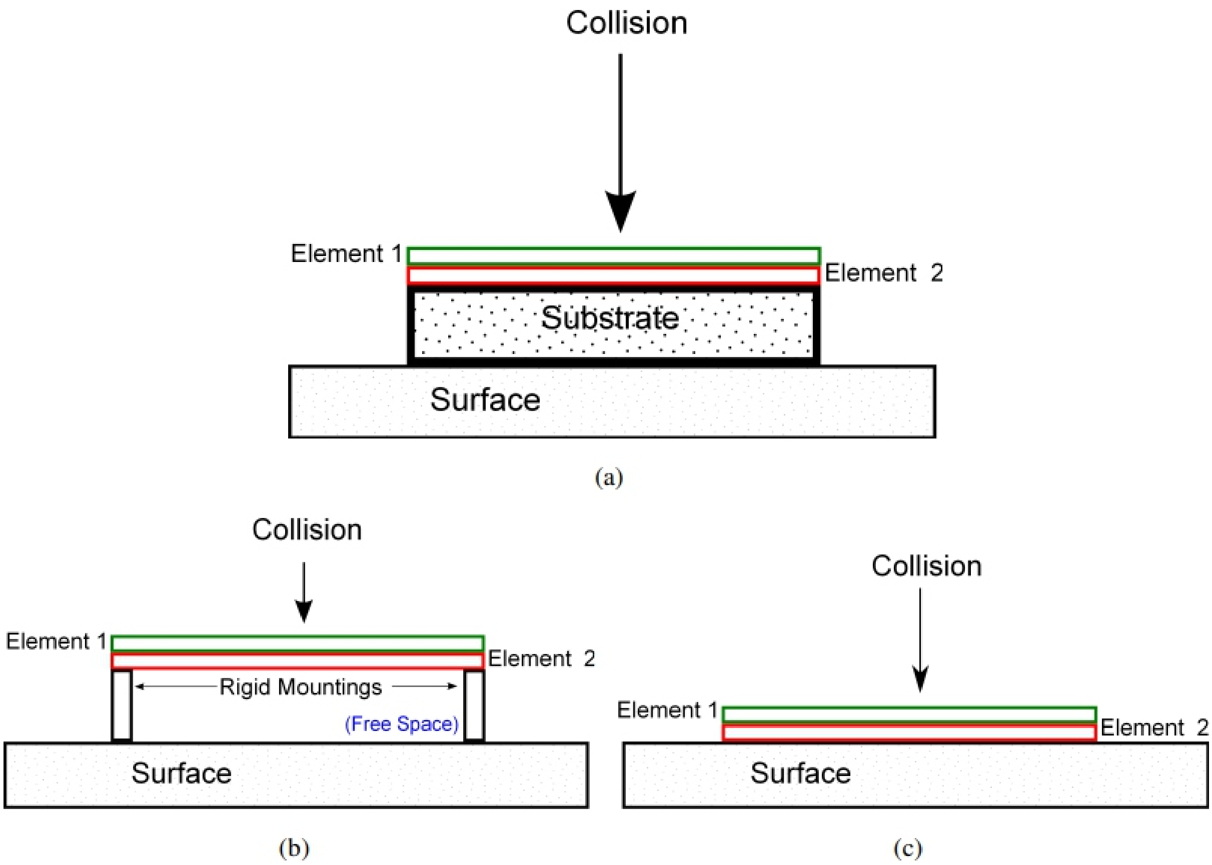


Figure 1. Sensor design diagrams. (a) Pseudo-membrane sensor construction; (b) Rigid construction; (c) Pressure construction

2.1.1. Sensor Structures Using PVDF Film

Physical structures of traditional rigidly mounted membrane stress sensation-based and pressure sensation-based polyvinylidene fluoride (PVDF) film sensors are illustrated in the following figures and descriptions. The Rigid Construction shown in Figure 1b requires free space for diaphragm deflection and rigid mountings in order to achieve dynamic sensing range. This creates issues requiring specialized construction for applications and limitation in sensor and surface curvature.

The pressure sensation-based construction, Figure 1c, eliminates the need for rigid mountings and free space of the membrane; however, the Pressure Construction sensor output of strain due to pressure is much lower yielding more complex and specialized electronics. The pressure based construction also requires a very rigid surface and/or substrate to generate pressure transduction.

In the proposed structure, a collision stimulus deforms, or compresses, the elastic substrate due to localized compression and creates a resulting mechanical strain on the PVDF film elements, similar to a rigid membrane. The elastic substrate should be chosen to maximize the linear stress strain response and also to minimize total sensor size for manufacturing and application concerns. The trilayer pseudo-membrane approach using PVDF and an elastic substrate is uniquely suited to large area coverage because large PVDF sensing elements are easily constructed, the elastic substrate can be made of polyurethane foams and other commercially available sufficiently compressible and elastic materials with a Young's Modulus lower than that of PVDF elements provided in Table 1, and the sensor is not limited to planar surfaces because the pseudo-membrane approach creates stress from localized substrate compression and not film strain as in the rigid membrane approach.

Table 1. Material properties of PVDF film.

Symbol	Property	Value	Units
E	Young's Modulus	2–4	nN/m^2
d_{31}	Transverse Coefficient	23	pC/N
d_{33}	Compressive Coefficient	−33	pC/N
p	Pyroelectric Coefficient	30	$\mu C/m^2K$

2.2. Sensing and Instrumentation

2.2.1. Piezoelectric Effect

The proposed sensor utilizes the piezoelectric effect of PVDF thin films to create sensation over a surface. The tactile element transduces experienced stress to an electrical displacement, D, which is the charge density of the film surface, Q/A. The electrical displacement has three additive components, consisting of pyroelectric, piezoelectric, and dielectric effects [16]:

$$D = p\Delta T + d_{jk}X_{jk} + \varepsilon E \tag{1}$$

The pyroelectric charge is a function of the change in temperature (ΔT) times pyroelectric charge coefficient *(p)*, the piezoelectric charge is a relation of stress applied in Cartesian direction *(Xjk)* with the corresponding piezoelectric charge coefficient *(djk)*, and the charge related to electric dipole moment is calculated by electric field *(E)* times the permittivity of the material *(ε)*. For collision sensing, the desire is for electrical displacement, *D*, to be a purely piezoelectric response, *djkXjk*. The electric field, E, can be minimized by proper design of the charge amplifier sensor interface, which is discussed below in Section 2.2.2. The pyroelectric component, ΔT, can be canceled because of phase orientation of the bilayer sensing element and common mode signal filtering; therefore, the pyroelectric and dielectric effects fall away reducing Equation (1) to the desired purely piezoelectric displacement in Equation (2).

$$D = d_{jk}X_{jk} = d_{31}X_{31} + d_{32}X_{32} + d_{33}X_{33} \tag{2}$$

The stress vector, Xjk, in Equation (2) represent tensile stress in length, width, and thickness directions respectively. Due to high compressibility of the substrate relative to the PVDF film, strain related to compression, X33 is approximately 0. The piezoelectric constants corresponding to tensile stress in the width and length, d31 and d32 respectively in Cartesian coordinate representation. The Cartesian coordinate system in this paper uses the following equivalent relations interchangeably, {x, y, z} = {1, 2, 3} = {width, length, thickness}. Where z+ is normal going away from the sensor and (x, y) are parallel. xx → x, yy → y, and zz → z, are equal. Therefore, the electrical displacement of the sensor is proportional to the total transverse and longitudinal stress in the film created by the collision reducing Equation (2) further to the reduced representation of the sensor electrical displacement in Equation (3)

$$D = d_{31}(X_{31} + X_{32}) \tag{3}$$

The applicable material properties of PVDF film are shown in Table 1 are provided by the film manufacturer, Measurement Specialties.

2.2.2. Electronics

The electronics interface for the PVDF film elements requires high signal gain, low output impedance, high input impedance, low time constant to capture 1 Hz collisions, and a minimization of the electric field effect of the sensor. A charge amplifier is used to minimize effects of sensor and line capacitance by minimizing input impedance, to minimize electric field by grounding sensor electrode, and because the elements act as a current source. The circuit, shown in Figure 2, acts as a single pole high-pass filter with a bleed resistor added in parallel to create a low enough cutoff frequency to properly detect physicalinteraction, in the 1 Hz to 1 kHz range [15]. The transfer function is given by:

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{sRC_s}{sRC + 1} \tag{4}$$

The system can be properly designed to yield a low enough corner frequency calculated from:

$$f_c = \frac{1}{2\pi RC} \tag{5}$$

which gives the desired low end frequency range. The signal is then low-pass filtered with a chosen high cut off frequency to attenuate unwanted high frequency noise. Finally, the signal is conditioned for input to the analog to digital converter (ADC). The charge amplifier allows for positive and negative voltage range of +/- Vcc to create the large dynamic range needed for collision detection.

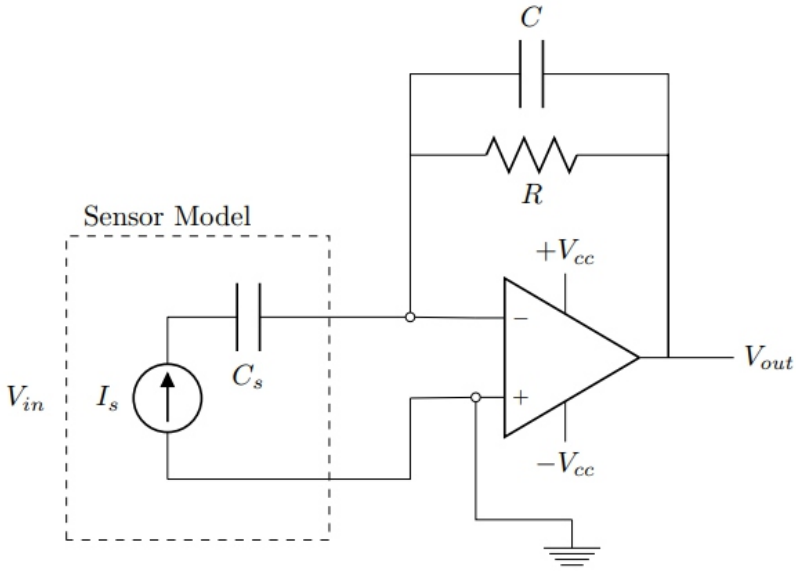


Figure 2. Charge amplifier schematic.

2.2.3. Strain Modeling

For modeling purpose, the area of concern is restricted to the local frame of the collision and the stress-strain response is approximately linear and the substrate will be treated as a continuum. The film stress from Equation (2) is equal to the stress of the surface of the substrate, assuming a perfect adhesive bond and negligible effects of film sensor on substrate stress characteristics,

$$\underline{\boldsymbol{\sigma}} = \begin{bmatrix} T^{e1} \\ T^{e2} \\ T^{e3} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_y & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_z \end{bmatrix} \tag{6}$$

By defining the Cauchy stress tensor (σ) of the substrate, Equation (6), in terms of the normal and shear stresses, {σx, σy, σz} and {τxy, τxz, τyz}, Xjk in Equation (2) can be replaced by the stress vector, T e3 , of the material surface resulting in Equation (7).

$$D = d_{31}(\tau_{xz} + \tau_{yz}) \tag{7}$$

From Equation (7), the stress contributing to the piezoelectric effect of the film is the orthogonal shear stress experienced by the substrate at the collision point. Therefore, the sensor dynamic range is dependent on the shear and normal stress characteristics of the chosen substrate. Using the shear modulus of elasticity (G) to relate shear strain to shear stress Equation (8), Young's Modulus of the substrate (E) to relate normal strain to normal stress Equation (8),

$$G = \frac{\tau_{xz}}{\gamma_{xz}} \qquad E = \frac{\sigma_z}{\epsilon_z} \tag{8}$$

along with the geometric representation of strain, Equation (9),

$$\epsilon_{ij} = \frac{1}{2}(\frac{\delta_i}{j_o} + \frac{\delta_j}{i_o}) \tag{9}$$

and Cauchy's strain tensor () Equation (10),

$$\underline{\underline{\epsilon}} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix} = \begin{bmatrix} \epsilon_x & \frac{\gamma_{xy}}{2} & \frac{\gamma_{xz}}{2} \\ \frac{\gamma_{yx}}{2} & \epsilon_y & \frac{\gamma_{yz}}{2} \\ \frac{\gamma_{xz}}{2} & \frac{\gamma_{zy}}{2} & \epsilon_z \end{bmatrix} \tag{10}$$

and assuming the substrate is under compression locally where δx = δy = 0, δz 6= 0, and xo, yo are known static quantities, Equation (7) is transformed to Equation (15), shown in Equations (11)–(14) using the relations Equation (8) through Equation (10) :

$$\begin{aligned} \tau_{xz} + \tau_{yz} &= G(\gamma_{xz} + \gamma_{yz}) = G(2\epsilon_{xz} + 2\epsilon_{yz}) && (11) \\ &= G(\frac{\delta_z}{x_o} + \frac{\delta_z}{y_o}) && (12) \\ &= G\frac{z_o(y_o + x_o)}{x_o y_o}(\epsilon_z) && (13) \\ &= \frac{G}{E}\frac{z_o(y_o + x_o)}{x_o y_o}(\sigma_z) = S && (14) \end{aligned}$$

Compressive stress, σz, is a monotonically increasing and directly proportionate function of the force of the collision normal to the sensor where force towards the sensor produces a positive response.

$$D = d_{31}(S) \quad (15)$$

Therefore a measured strain S in Equation (15) and the electrical displacement should also be monotonically increasing functions of the force. Modeling of strain was primarily accomplished with reference to [17].

3. Experimentation

3.1. Prototype I

The initial prototype sensors for testing were constructed from poled 28 μm thick PVDF film elements, each 171 mm by 19 mm (length and width) and a 12.7 mm (0.5 inch) polystyrene closed cell foam substrate. The polystyrene foam was chosen by commercial availability and to allow for large amounts of compression at collision. The trilayer sensor was constructed by adhering the two films, Element 1 and Element 2 (see Figure 1a), out of phase such that Element 1's top electrode is positive and Element 2's top electrode is negative, adhering the bilayer PVDF film to the polystyrene foam substrate, and then affixing to the sample robotic arm cover (mechanical shielding of the robot). Figure 3 is a photograph showing sensors mounted in both planar and non-planar configurations. Signal capture was performed using previously described amplifier circuit design interfaced to 12-b analog to digital converters on an Atmel Xmega microcontroller using a buffer and signal conditioning amplifier stage. The charge amplifier was designed with a 1.6 MΩ bleed resistor and 100 nF charge accumulating capacitor yielding the following corner frequency:

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi(100\,\mathbf{nF})(1.6\,\mathbf{M\Omega})} = 0.997\,\mathbf{Hz} \quad (16)$$

which gives the desired low end frequency range. The signal is then low-pass filtered with a 1 kHz cut off frequency to attenuate unwanted signals. The charge amplifier power supply (Vcc) range is ±15 V, which allows for high gain and large output dynamic range. However, the analog to digital converter (ADC) operates from 0–3 V. Therefore, the signal output of the charge amplifier is scaled down and level-shifted to the same range. The 12-b ADC uses a reference voltage of 1.5 V in differential mode resulting in a digital output range of −2048 to +2048 corresponding to a single bit resolution of 7.32 mV. The ADC's sampling frequency is 93.7 kHz, which is more than 40 times the bandwidth of the analog input. Data was logged using serial communication with signals down sampled to 10 kHz.

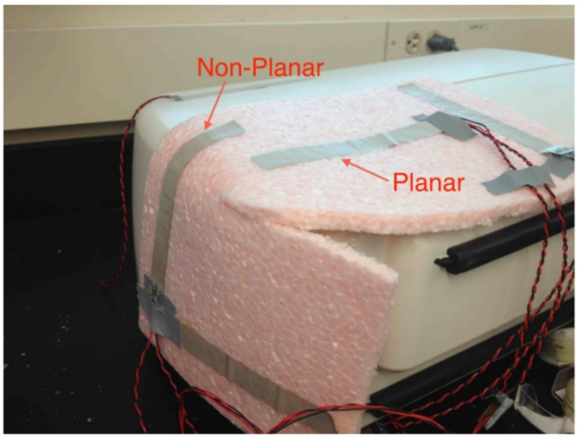


Figure 3. Sensor prototype used in testing showing planar (top of cover) and non-planar (rounded left

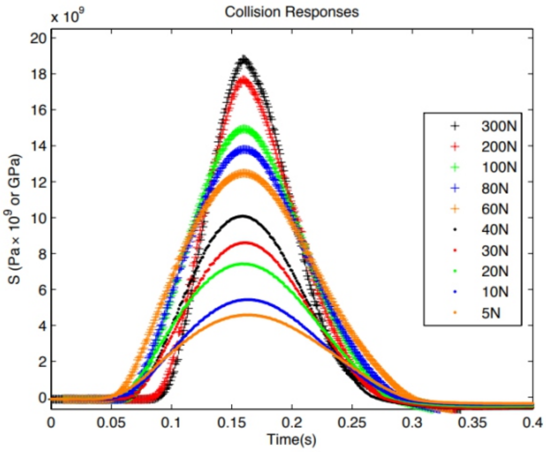
end) sensor applications on example robotic arm shielding. Wires in picture connect sensor electrodes to instrumentation.

3.1.1. Testing Method

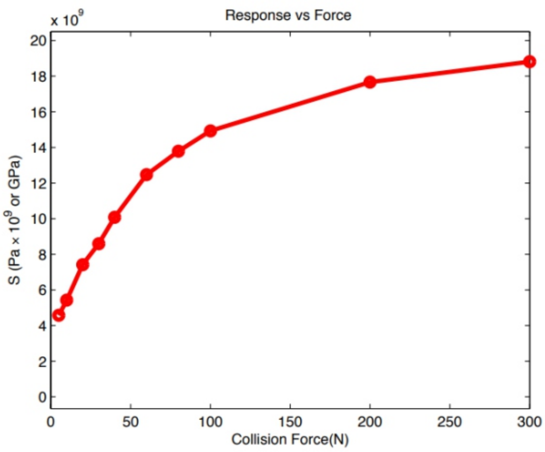
Collision stimuli for testing was generated by dropping an object of known weight and uniform contact area on the sensor from varied heights to produce controlled impact collisions. Force of the object at impact is taken from the velocity due to free fall and the relation of the work-energy principle where distance to slow down is compression of the substrate, defined as compressive strain multiplied by thickness. An approximation of distance for the object to slow down is a 50% compression of the substrate resulting in a slowdown distance of 0.635 cm.

3.1.2. Results and Discussion

The sampled mean results, presented in Figure 4a, show the wide dynamic sensor range and consistent response to collision. For collision forces starting at 80 N and above, the measured impacts show some attenuation and clipping which is most likely a result of elasticity in the cover and compression distance of the polystyrene substrate. At high levels of force impact, the rigidity of the testing cover and compression distance becomes a limiting factor and noise source. The relation of applied collision stimuli to measured stress peaks is plotted in Figure 4b. The sensor response is not perfectly linear, but does resemble the engineering stress strain curves of foam under uniaxial compression [18], which reinforces the previous assertion that stress measured by the sensor is related to the localized compressive strain at the impact point.



(a)



(b)

Figure 4. Results from sensor mounted on a planar surface. (a) Mean captured results for wide dynamic range of collision stimuli; (b) The relation of measured collision to force of object collision.

The results in Figure 5 show that the measured response from collision for a non-planar application strongly correlate with the results from planar application. Some deviation is expected, but the sensor provides detection over the desired dynamic range for the non-planar application. The consistency between planar and non-planar experiments demonstrates the robust application properties of the sensor.

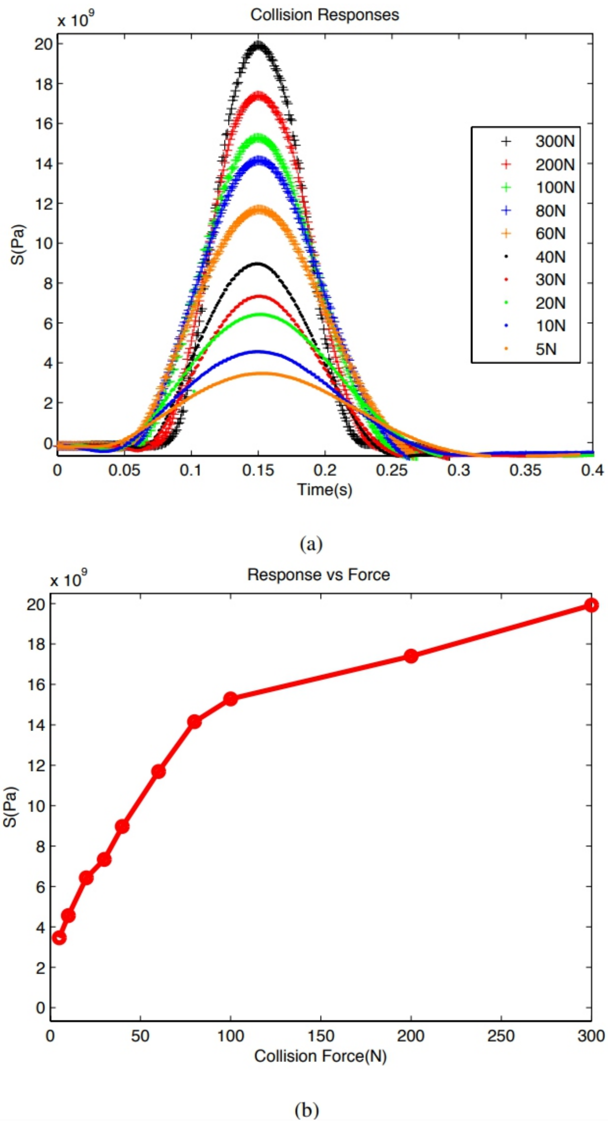


Figure 5. Results from sensor mounted on a non-planar surface. (a) Mean captured results for wide dynamic range of collision stimuli; (b) The relation of measured collision to force of object collision.

From Figure 6, there is evidence of time delay and difference in stress measured by the upper and lower elements. The authors believe the difference can be explained by delamination of the sensor elements over extended time; the initial testing prototype was constructed with double-sided tape not rated for repeated force impacts. Improved construction (adhesion) should eliminate the difference in sensors. The plots do show that the measured response of both elements is uniform. The digital response shown in Figure 6 is measured in ADC bits and shown in VLSB, which was previously discussed and calculated as 7.32 mV in Section 3.1.

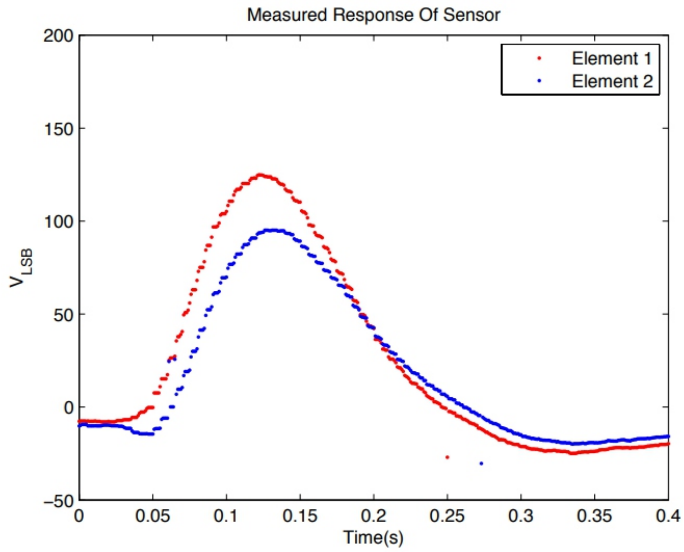


Figure 6. Mean digital response results of Element 1 and Element 2 for 5 N collisions.

3.2. Prototype II

The second prototype cover was designed to test the sensor functionality in commercial applications.The sensors were constructed from poled 28 µm thick PVDF film in the same form described in Figure 1awith Element 1 and Element 2 being opposite in polarity. The trilayer construction was implementedusing multiple substrates including 1 mm, 3 mm and 5 mm polyurethane foam and similar thicknesssilicon rubber. The sensor elements were cut to shape to achieve total coverage of a protective fasciafrom a commercial robotic arm. The electronics board was reused from Prototype I. The size of thesensors for this test varied but were approximately 20 cm by 10 cm with some being slightly largerand others smaller. Due to the significant increase in surface area vs. the sensors in Prototype I, theexpected sensation values in voltage should be smaller because of increased capacitive impedance ofthe element; additionally, the decreased depth of the substrate compression should lead to slightly lesslocalized stress in the sensors. The following testing and results show that despite these changes thesensors provide more than adequate sensation for commercial viability. The data shown in Figures 7–10is all displayed as the digital output of the ADC. The ADC bits correspond to VLSB steps previouslydiscussed in Section 3.1; however, data is shown as bits instead of volts in order to provide the readerwith a clear picture of the digital output of the sensor in a real-world environment.

3.2.1. Testing Method

The testing for commercial viability was accomplished with a multifaceted test routine. The prototype cover was attached to a commercial robotic arm and then put through a series of dynamics to gauge sensor response to stimuli. The following test states were used: normal operation (i.e., arm rotation with no collision) to characterize system noise and vibration detection, normal operation with simulated collisions to gauge signal to noise ratio of stimuli, normal operation with emergency stop (i.e., full speed arm rotation then emergency stop triggered) to gauge sensor false positive rejection, and dynamic movement with collisions and emergency stop to provide a full representation of sensor function. The sensor data captured and shown in the following sections corresponds to the 5 mm thick polyurethane foam substrate.

3.2.2. Normal Operation

Results for the prototype sensation during normal operational movements are shown in Figure 7. The protective cover is on the outside of an implement which is rotated at a constant speed of 10 ◦/s through a full rotation during the captured time window. The resulting data in Figure 7 shows the sensors perception of the noise due to arm vibration and electronics. The measurements have a bias with mean of 15.109 ADC counts and a standard deviation of ±0.826 count. The low deviation of the measurements, less than one least significant bit of the ADC, shows that during movement of the arm the sensor measurement maintains the steady state values; furthermore, the dynamics and vibrations of the manipulated implement do not affect the sensor.

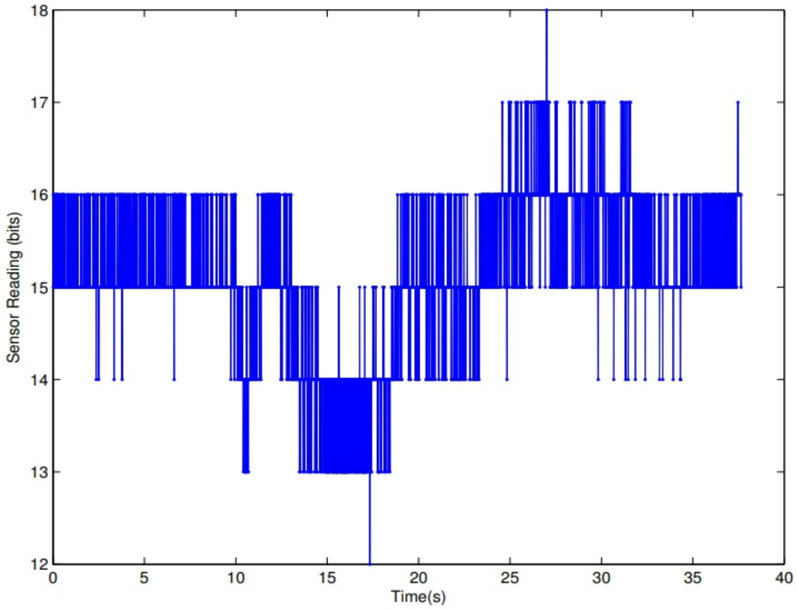


Figure 7. Sensor measurements from test using robotic arm.

3.2.3. Collision Perception

The results for a collision scenario are shown in Figure 8. The sensor measurement shows several distinct collisions generated by the stimuli. The testing scenario involved rotating the arm at normal operation speeds as in the previous test, see Figure 7, and applying a collision stimulus with the human hand, tapping or pressing the sensor. The estimated force generated by these light taps would be 5–10 N, or significantly below the threshold of pain or harm. The collisions are clearly detectable over the previously shown sensor noise in Figure 7. For reference, standard mechanical or fiber optic switch based sensors currently deployed with the arms have a sensation threshold of 60 N to 100 N. This is the derived force detection level necessary for arms to prevent hazards such as crushing, shearing, cutting or severing and entanglement [19]. Lower sensation levels are necessary in order to eliminate unintended movements, system overrun during collision and other potential hazards [20]. The data in Figure 8 shows the multiple collision event for which there is a positive sensation at initiation of collision, and a negative sensor response on release due to the sensors holding the electric charge generated by the piezoelectric effect. The area under the curve, or total charge, is approximately equal for the collision and release phase. In addition, near there end there is a stop event on the arm. The light collisions are clearly perceptible above the system noise shown previously and easily detected with implementable algorithms; however, the third collision event which is a press, release, press, and release sequence is initiated during the decay period of the second event causing the initial positive peak to be obscured. The final sensor oscillation of the sensor data starting at the 10 s mark is an emergency stop, which will be discussed further in the next section. For Figure 8, there are two tapping collisions at 2 s and 4 s and then a press and release collision sequence at the 6 s mark.

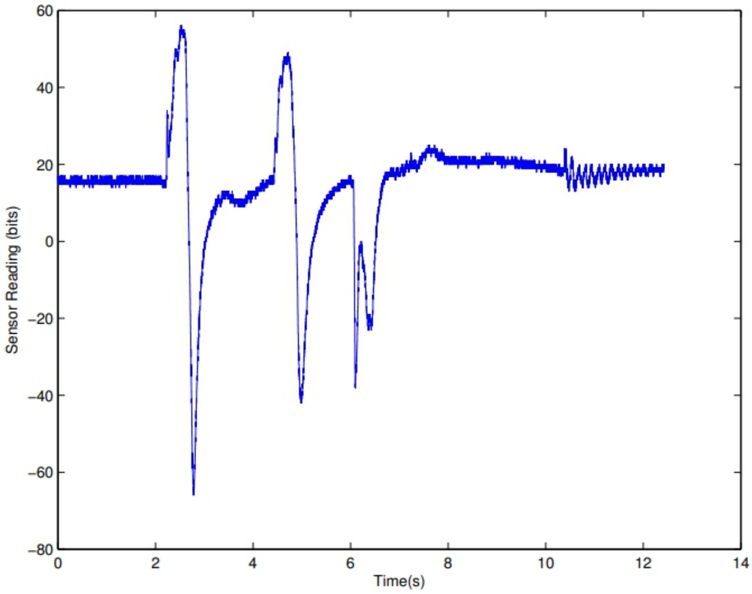


Figure 8. Sensor measurements from test using robotic arm with collision.

3.2.4. Emergency Stop Detection

Results for sensor perception of an emergency stop event are shown in Figure 9. The sensor is at the previously shown steady state from Figure 7; however, the vibrations created from the emergency stop show in the measurements of Figure 9. The mean line, shown in red in Figure 9, shows that the emergency stop sensation oscillates around the mean and does not generate a large sensor response. The sensation level, a couple of ADC counts above mean, is well below the perceived collisions in Figure 8. The sensed stress during emergency stop can be attributed to a couple of the following factors: weight of leads causing stress due to vibration and movement of cabling, cover flex during emergency stop event, high dynamics causing sensor to minutely move due to weight of the element.

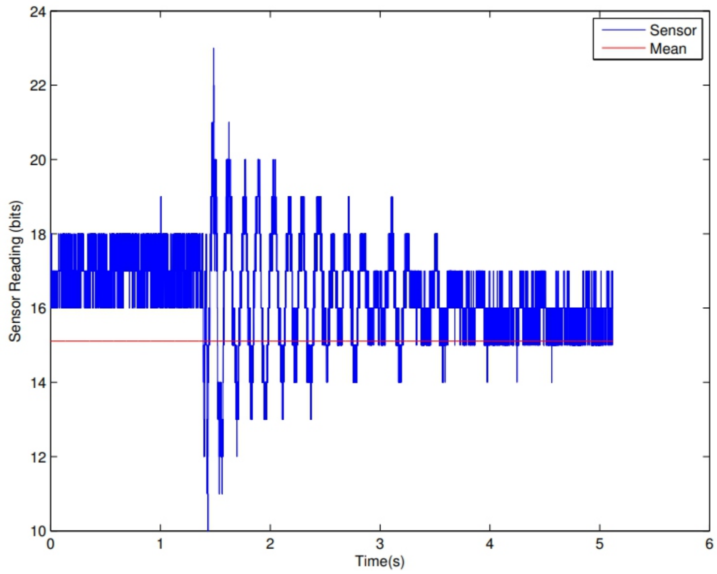


Figure 9. Sensor measurements from test using robotic arm with emergency stop.

3.2.5. Dynamic Movement with Collisions and Emergency Stop

Finally, the results from a test run in which multiple collision events occur followed by an emergency stop are shown in Figure 10. The dynamic run included rotations similar to those in the previous results, Figures 7–9; however, the collision stimuli purely consist of low frequency pressing events and releases. Due to the capacitive nature and filter tuning, very low frequency collision is difficult to detect. For the data in Figure 10, there are two press-and-release events at about 0 s and then again at 1 s. The resulting spikes from initiation of collision are hard to sense, about 10 ADC above the mean, but present in the data. The resulting opposite polarity spike from release of the pressure is clearly shown. Following the second rebound there is a slow press and slow release that occurs starting at the 1.5 s mark, this can be seen and in this case the release is slower than the capacitive time constant of the system so the negative rebound is not seen. Finally we see the emergency stop reading, the estop and press-and-release collisions are of significantly different curve structure and sensor level. As a result, the data can be easily processed to detect the collision events while rejecting the emergency stop sensation.

The collision events are clearly detectable over the mean. The reading generated from the emergency stop vibrations are significantly smaller than measured collision events, while both collision and emergency stop vibration events unique and differentiable from system noise.

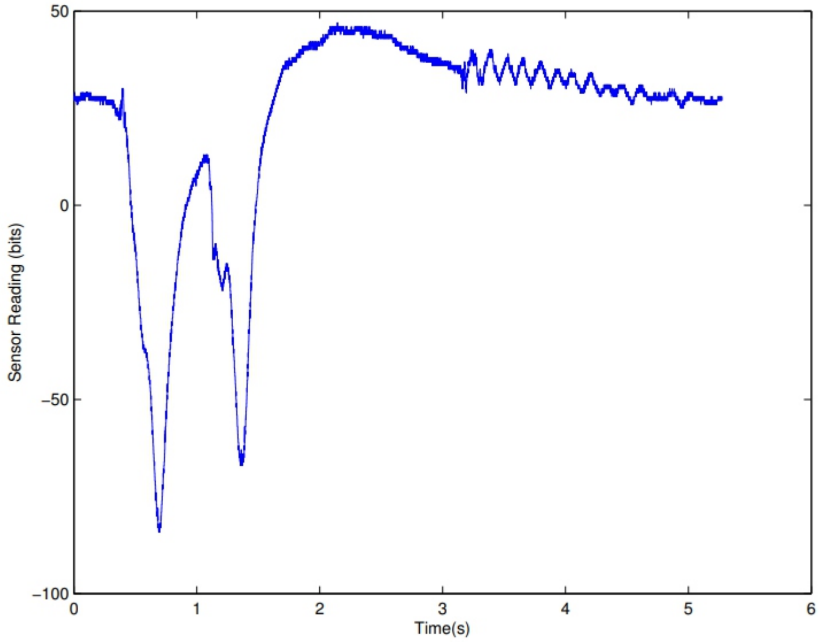


Figure 10. Sensor measurements from test using robotic arm with collision and emergency stop.

4. Concluding Remarks

In contrast to existing technologies, the proposed sensor design shows strong collision sensation for both planar and non-planar surfaces. The pseudo-membrane construction eliminates not only mechanical issues associated with pressure and membrane based sensation but also increases applicability of the PVDF sensing technology. Furthermore, the uniform and consistent response of planar and non-planar applications eliminates hardware specialization needs such that modular collision detection systems can be created. The interface electronics and sensor construction is accomplished with commercially producible parts such that retrofitting is easily accomplished. The results support the theoretical relation of compressive stress to measured response in the local frame and the sensor measurements are a monotonically increasing function of the force.

The commercial prototype, Prototype II, and testing with commercial arms show the viability of the design. The decreased foam substrate thickness used for Prototype II does not significantly degrade the performance and sensation of the system. The results from commercial testing clearly show collision detection above the level of system noise and false positive generating vibrations. In contrast to mechanical based collision switches, the sensation range starting as low as a few Newtons allows the system to determine what sensation is caused by collision vs. false positive generated by vibration. The additional sensation range vs. other sensors, robust application form and lack of mechanical parts increase the viability of retrofitting deployed systems with sensors to increase operational efficiency with autonomy, increased movement speeds, and lower safety risk.

The novel sensor design and resulting testing in this paper shows strong promise for a robustly applicable collision detection solution for complex robotic arms and non-standard operating environments.

Acknowledgments

The authors gratefully thank Siemens AG for sponsoring the research.

Author Contributions

J. M. Wooten conceived and designed the sensor; J. M. Wooten also designed and performed the experiments; J. M. Wooten and D. M. Bevly analyzed the data; J. M. Wooten and J. Y. Hung wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Fujimoto, I.; Yamada, Y.; Morizono, T.; Umetani, Y.; Maeno, T. Development of artificial finger skin to detect incipient slip for realization of static friction sensation. In the Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI-2003), Tokyo, Japan, 30 July–1 August 2003; pp. 15–20.

2. Seminara, L.; Pinna, L.; Valle, M.; Basirico, L.; Loi, A.; Cosseddu, P.; Bonfiglio, A.; Ascia, A.; Biso, M.; Ansaldo, A.; et al. Piezoelectric polymer transducer arrays for flexible tactile sensors. IEEE Sens. J. 2012, 12, 1–4.

3. Shapiro, Y.; Wolf, A.; Kósa, G. Piezoelectric Deflection Sensor for a Bi-Bellows Actuator. IEEE/ASME Trans. Mechatron. 2013, 18, 1226–1230.

4. Kolesar, E.S.; Dyson, C.S. Object imaging with a piezoelectric robotic tactile sensor. J. Microelectromech. Syst. 1995, 4, 87–96.

5. Lee, S.; Ahn, Y.; Prabu, A.; Kim, K. Piezoelectric Polymer and Piezocapacitive Nanoweb Based Sensors for Monitoring Vital Signals and Energy Expenditure in Smart Textiles. J. Fiber Bioeng. Inform. 2013, 6, 369–381.

6. Kymissis, J.; Kendall, C.; Paradiso, J.; Gershenfeld, N. Parasitic power harvesting in shoes. In the Proceedings of the Second International Symposium on Wearable Computers, Pittsburgh, PA, USA, 19–20 October 1998; pp. 132–139.

7. Pi, Z.; Zhang, J.; Wen, C.; Zhang, Z.; Wu, D. Flexible piezoelectric nanogenerator made of poly(vinylidenefluoride-co-trifluoroethylene) (PVDF-TrFE) thin film. Nano Energy 2014, 7, 33–41.

8. Han, H.; Nakagawa, Y.; Takai, Y.; Kikuchi, K.; Tsuchitani, S. PVDF film micro fabrication for the robotics skin sensor having flexibility and high sensitivity. In Proceedings of the 2011 Fifth International Conference on Sensing Technology (ICST), Palmerston North, New Zealand, 28 November–1

December 2011; pp. 603–606.
9. Kanik, M.; Aktas, O.; Sen, H.S.; Durgun, E.; Bayindir, M. Spontaneous High Piezoelectricity in Poly(vinylidene fluoride) Nanoribbons Produced by Iterative Thermal Size Reduction Technique.ACS Nano 2014, 8, 9311–9323.
10. Canavese, G.; Stassi, S.; Cauda, V.; Verna, A.; Chiodoni, A.; Marasso, S.; Cocuzza, M. Differentscale confinements of PVDF-TrFE as functional material of piezoelectric sensor devices. In Proceedings of the 11th IEEE Conference on Sensors (IEEE Sensors 2012), Taipei, Taiwan, 28–31 October 2012; pp. 1–4.
11. Mirbagheri, A.; Dargahi, J.; Aghili, F.; Parsa, K. Finite element analysis of a membrane-type piezoelectric tactile sensor with four sensing elements. In Proceedings of the 4th IEEE Conference on Sensors (IEEE Sensors 2005), Irvine, CA, USA, 30 October–3 November 2005; pp. 353–356.
12. Dahiya, R.; Metta, G.; Valle, M.; Sandini, G. Tactile Sensing - From Humans to Humanoids. IEEE Trans. Robot. 2010, 26, 1–20.
13. Haddadin, S.; Albu-Schäffer, A.; de Luca, A.; Hirzinger, G. Collision detection and reaction: A contribution to safe physical Human-Robot Interaction. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2008), Nice, France, 2008; pp. 3356–3363.
14. Roh, Y.; Varadan, V.; Varadan, V.K. Characterization of all the elastic, dielectric, and piezoelectric constants of uniaxially oriented poled PVDF films. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 2002, 49, 836–847.
15. Seminara, L.; Capurro, M.; Cirillo, P.; Cannata, G.; Valle, M. Electromechanical characterization of piezoelectric PVDF polymer films for tactile sensors in robotics applications. Sens. Actuators A: Phys. 2011, 169, 49–58.
16. Damjanovic, D. Ferroelectric, dielectric and piezoelectric properties of ferroelectric thin films and ceramics. Rep. Prog. Phys. 1998, 61, 1267.
17. Gere, J. Mechanics of Materials, 5th ed.; Brooks/Cole Publishing: Belmont, CA, USA, 2001; p. 926.
18. Bryson, J.A. Impact Response of Polyurethane. Ph.D. Thesis, Washington State University, 2009.
19. International Standard Organization. Robots and Robotic Devices—Safety Requirements for Industrial Robots: Part 2: Robot Systems and Integration; Number ISO 10218-2:E; International Standard Organization: Geneva, Switzerland, 2011.
20. International Standard Organization. Robots and Robotic Devices—Safety Requirements for Industrial Robots: Part 1: Robots; Number ISO 10218-1:E; International Standard Organization: Geneva, Switzerland, 2011.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.