2024-05-20
[Journal of Information Processing]
Title: Communication Overhead Description Schema for Multi Core Processor in Model-based Development
Authors: Yue Hou, Yutaro Kobayashi, Hiroshi Fujimoto, and Takuya Azumi

Dear Editor and Reviewer,

We are pleased to resubmit our manuscript entitled "Communication Overhead Description Schema for Multi Core Processor in Model-based Development" for consideration in your esteemed journal. This submission follows a previous round of review where it received mixed feedback, including one conditional acceptance and one reject decision.

Following the insightful comments from the reviewers, we have undertaken substantial revisions to address the concerns raised.

We have emphasized our manuscript's focus on the communication overhead schema, explaining how our proposed methodology allows for faster construction and reduced descriptive complexity of the schema. Additionally, we have addressed the source of the schema files by specifying the providers or creators of the XML used.

To further clarify the choices in our methodology, we have elaborated on why we employed a linear model to describe inter-core communication overheads. Considering factors such as spatial distance, data volume, network topology, and traffic load, a multidimensional approach could capture the complexity of inter-core communications more accurately. However, implementing more complex models, such as those involving cubic equations, would increase computational costs and the complexity of the model. For many practical applications, particularly those with limited computational resources or real-time processing capabilities, simpler models may be more appropriate.

Please note that all changes made in the manuscript are highlighted in red for your ease of review. We believe these revisions comprehensively address the previous concerns and significantly enhance the value of our work. We appreciate your time and consideration in reviewing our revised manuscript and look forward to your feedback.

Yours sincerely,

Yue Hou,
Saitama University
Email address: hou.y.556@ms.saitama-u.ac.jp

1 Response to Meta-reviewer

1.1 Response to Peer Review Feedback: Addressing Non-Acceptance Reason Based on Conditional Acceptance Requirement 1 of Meta-reviewer

Comment:

As the meta-reviewer in the previous submission mentioned, the accuracy of the overhead estimation results does not rely on the schema itself but on the models for the estimation.

Suppose the authors want to show the effectiveness of the proposed schema extension for the communication overhead estimation. In that case, it is required to provide the evaluation that the schema can describe various communication models in addition to the linear model.

The linear model is naive to describe/compute communication overhead among cores/nodes that it is insufficient to show the proposed extension's effectiveness. However, the proposed schema extension supports only the linear model. Evaluating how much the proposed extension is adjustable for various overhead models will be required.

Our reply:

Thank you for your insightful comments and suggestions regarding our manuscript. We appreciate your concerns about the simplicity of the linear model used in our study to estimate communication overhead among cores/nodes. We acknowledge that a multi-dimensional approach could more accurately capture the complexities of inter-core communication by considering factors such as spatial distance, data volume, and network topology and traffic load. Indeed, these elements significantly influence communication latency in real-world scenarios. However, our choice to use a linear model in this study was deliberate, based on several key considerations:

Simplicity and Understandability: The linear model provides a clear and straightforward approach, beneficial in the initial stages of research. Its simplicity facilitates quicker iterations and validations of the proposed schema, making it easier for both developers and users to adopt during the early phases.

Practical Implementation: Implementing more complex models, such as those involving cubic equations, would increase both computational overhead and schema complexity. For many practical applications, particularly where computational resources or real-time processing capabilities are limited, a simpler model may be more appropriate.

And we revised the article to clarify the problem points and specific solutions we hoped to address with our proposal approach. Also, in the evaluation section, it is made clear why linear equations are used to describe the inter-core communication delays, and the significance of each parameter is explained in detail. Meanwhile the experimental results show that using our proposal approach, not only can we achieve a higher accuracy of inter-core delay estimation, but also can greatly reduce the amount of description and make it easy to use. We hope this response clarifies our modeling choice. We remain committed to exploring more comprehensive models in future work to enhance the predictive accuracy and applicability of our schema.

Adjusted sentences in the "Introduction" section (Section 1)

The latest version of SHIM [14] currently still lacks an efficient method to describe inter-core communication latency, which poses significant challenges. This deficiency leads to substantial inaccuracies in time predictions within the Model-Based Parallelizer (MBP) due to the omission of inter-core communication latency calculations. Furthermore, the lack of a concise communication description necessitates extensive documentation of latency between cores, resulting in increased number of lines for the communication overhead description to more than ten thousand lines.

Adjusted sentences in the "Introduction" section (Section 1)

To address these issues, a revised schema is proposed that measures communication overhead in per-API units for more precise estimates and also simplifies the schema's structure to reduce its volume. This enhancement facilitates ease of use and construction, allowing for a more accurate and comprehensive representation of communication overhead. Additionally, it improves the reliability and efficiency of performance evaluations in multi-core environments.

Adjusted sentences in the "Evaluation" section (Section 5.2.2)

In the given scenario, two separate linear equations are utilized to approximate the communication overhead for sending and receiving operations, respectively. This differentiation is crucial because the processes involved in sending and receiving data often have distinct characteristics and resource requirements, which can significantly impact communication overhead. The first linear equation, with a $Coefficient$ of 48.5 and an $Intercept$ of 12,999, is used to estimate the overhead for receiving operations. The coefficient here likely represents the incremental overhead per unit of data received, while the intercept might account for fixed costs associated with initiating or terminating a receive operation. The second linear equation, with a $Coefficient$ of 20.0 and an $Intercept$ of 40,800, is designed for sending operations. In this case, the lower coefficient suggests that the variable overhead per unit of data sent is less than that of receiving. However, the higher intercept indicates a greater fixed cost, possibly due to the complexities and resources involved in preparing and dispatching data.

By using two different formulas, the model can more accurately reflect the differing nature and costs of sending and receiving data, leading to a more precise and realistic estimation of communication overhead in systems where these activities may vary in frequency, size, and complexity.

1.2 Response to Peer Review Feedback: Addressing Non-Acceptance Reason Based on Conditional Acceptance Requirement 1 of Meta-reviewer

Comment:

Regarding the usability and reusability for the users, I do not think the number of descriptions to be written in the XML files becomes a problem because processor vendors or software tools would provide the information/files. The authors should explain the SHIM use cases while clarifying by whom and how the XML files are provided to discuss usability/reusability.

Our reply:

Thank you for your insightful feedback on our manuscript. We appreciate the time and effort you have taken to review our work. Regarding your comment on the usability and reusability of the SHIM XML files, we have added explanations in the manuscript to address the concerns about the source of these schemas.

The creation of SHIM XML files by hardware providers is indeed essential for the effective usage of software tools. However, it is important to note that not all hardware providers supply these XML files, which can limit the adoption of certain hardware. To mitigate this issue, we propose the use of freely available reference authoring tools, along with detailed specifications. These tools enable users who have access to technical manuals and hardware (such as simulators or evaluation boards) to generate SHIM XML files for most multi-core and many-core systems. This approach significantly enhances the accessibility and integration of SHIM XML, making it easier for users to adopt and utilize the hardware effectively.
We hope this clarification addresses your concerns regarding the usability and reusability of the SHIM XML files.

Added sentences in the "SHIM" section (Section 2.1)

The creation of SHIM XML by hardware providers is essential for software tool usage, yet not all providers supply this XML. This limitation could hinder hardware adoption. To address this, freely available Reference authoring tools, accompanied by specifications, allow users with access to technical manuals and hardware (such as simulators or evaluation boards) to generate SHIM XML for most multi-many-core systems, thus enhancing accessibility and integration.

2 Response to 1st reviewer

2.1 Response to Peer Review Feedback: Addressing Non-Acceptance Reason Based on Conditional Acceptance Requirement 1 of reviewer 1

Comment:

> I have understood the differences from the existing SHIM 2.0 method sufficiently.
>
> However, it is believed that novelty and utility are not sufficiently demonstrated compared to general parallelization frameworks or parallelization methods.
>
> Please demonstrate the significance of methods that take into account this communication overhead, not limited to MBD methods.

Our reply:

> Thank you for your insightful feedback on our manuscript. We appreciate the time and effort you have taken to review our work.
>
> We understand your concerns regarding the novelty and utility of our methods compared to general parallelization frameworks or parallelization methods. Our research primarily focuses on optimizing the schema for describing inter-core communication latency, rather than on parallelization itself. In our paper, we have not discussed the utility of parallelization or its impact on parallelization performance. Instead, our study is dedicated to simplifying the description of communication latency within the schema. Traditional methods for describing latency can be extremely verbose, often requiring tens of thousands of lines. Our approach significantly reduces the complexity and volume of these descriptions, providing a more efficient and streamlined schema representation. We believe this clarification highlights the novelty and utility of our work. By reducing the descriptive burden and improving the accuracy of execution time estimates, our method offers substantial improvements in the field of schema optimization. We hope this addresses your concerns and underscores the significance of our contributions within the intended scope.

We have address the problems and proposed method in SHIM 2.0 in the "Introduction" section. We have also added references to the latest version of SHIM by citation [13], in the hope of showing that the problems we've mentioned haven't yet been solved in the latest version of SHIM.

Adjusted sentences in the "Introduction" section (Section 1)

The latest version of SHIM [14] currently still lacks an efficient method to describe inter-core communication latency, which poses significant challenges. This deficiency leads to substantial inaccuracies in time predictions within the Model-Based Parallelizer (MBP) due to the omission of inter-core communication latency calculations. Furthermore, the lack of a concise communication description necessitates extensive documentation of latency between cores, resulting in increased text volume that complicates data importation and editing.

[14] IEEE-Standard-Association, "SHIM 2.0," https://www.document-center.com/standards/show/IEC-63504-2804

Adjusted sentences in the “Introduction” section (Section 1)

To address these issues, a revised schema is proposed that measures communication overhead in per-API units for more precise estimates and also simplifies the schema’s structure to reduce its volume. This enhancement facilitates ease of use and construction, allowing for a more accurate and comprehensive representation of communication overhead. Additionally, it improves the reliability and efficiency of performance evaluations in multi-core environments.

2.2 Response to Peer Review Feedback: Addressing Non-Acceptance Reason Based on Conditional Acceptance Requirement 2 of reviewer 1

Comment:

> Despite not appearing to be related to papers on autonomous driving, I feel a sense of discomfort with the use of "autonomous driving system" at the beginning of the Abstract and Introduction.

Our reply:

> Thank you for your insightful feedback on our manuscript. We have carefully considered your comment regarding the use of "autonomous driving system" at the beginning of the Abstract and Introduction. We understand your concern and would like to clarify the relevance of this context to our research.
>
> In the field of autonomous driving, Simulink is frequently used for simulation and modeling. As autonomous driving systems become increasingly complex, Model-based Development (MBD) is being progressively adopted in system development. The use of Simulink in MBD development is also becoming a growing trend. Our research focuses on task parallelization within the context of MBD, which is highly relevant to the development of autonomous driving systems. Therefore, we believe that referencing autonomous driving systems at the beginning of the Abstract and Introduction helps to set the context and underline the applicability of our work.
>
> We hope this explanation addresses your concerns and clarifies the significance of our research in the context of autonomous driving system development.

Added sentences in the "Abstract"

Autonomous driving systems require a wider range of functionalities than traditional embedded systems, making Model-based Development (MBD) essential for efficiently managing the complexity. MBD with MATLAB/Simulink, allows for early simulation and validation of these functionalities, ensuring safety and reducing development time. At the same time, the real-time requirements of multi-tasking for autonomous driving cannot be ignored.

Added sentences in the "Introduction" section (Section 1)

In development of these systems, Model-Based Development (MBD) has raised for its functionalities such as early simulation, and validation [3,4]. Furthermore, low power consumption is also required to enable installation in automobiles. Thus, more comprehensive functions are required than those of conventional embedded systems [5]. To meet these requirements, many-core processors are being adopted. Many-core processors offer high computational performance and energy efficiency [6-8]. Efficient utilization of these processors, however, necessitates understanding their complex specifications, which can extend development times.

[3] MathWorks. Automated Driving Toolbox. https://jp.mathworks.com/products/automated-driving.html

[4] Miura, Keita and Tokunaga, Shota and Ota, Noriyuki and Tange, Yoshiharu and Azumi, Takuya, "Autoware toolbox: Matlab/simulink benchmark suite for ros-based self-driving software platform," In *Proceedings of the 30th International Workshop on Rapid System Prototyping (RSP)*, pp. 8–14, 2019

Adjusted sentences in the "Introduction" section (Section 1)

In MBD, MATLAB/Simulink is noted for its automatic C code generation from models, but it does not support parallelized code generation [9]. Addressing this, the Embedded Multicore Consortium's Model-Based Parallelizer (MBP) facilitates the automatic generation of parallelized C code. This integration streamlines simulation and code generation for multi-many-core processors, enhancing development efficiency. Utilizing SHIM (Software-Hardware Interface for Multi-Many-Core) to abstract essential processor information minimizes the reliance on physical hardware [10-13].

2.3 Response to Peer Review Feedback: Addressing Non-Acceptance Reason Based on Conditional Acceptance Requirement 3 of reviewer 1

Comment:

Is the proposed method only effective for Karlay MPPA3-80? Additionally, is there a specific reason for targeting Karlay MPPA3-80?

Our reply:

Thank you for your insightful feedback on our manuscript. We appreciate the time and effort you have taken to review our work.

We have carefully considered your comment regarding the effectiveness of the proposed method for the Kalray MPPA3-80 and the reason for targeting this specific processor. In the field of autonomous driving, our research focuses on task parallelization within Model-based Development (MBD). Autonomous driving systems have higher demands on device performance, power consumption, and other factors compared to traditional systems. The availability of devices equipped with multi-core processors that meet these stringent requirements is limited. The Kalray MPPA3-80 is one of the few processors that satisfy these criteria, making it an ideal target for our study.

Furthermore, we would like to emphasize that our proposed method is not dependent on the Kalray MPPA3-80 and can be applied to other devices as well. This flexibility underscores the broad applicability of our method across different hardware platforms. We have added a clarification on this point in the main text to ensure that the method's adaptability is clear.

We hope this explanation clarifies our choice and demonstrates the broader applicability of our method.

Added sentences in the "Evaluation" section (Section 5)

Additionally, it is important to note that the proposed method is not dependent on the Kalray MPPA3-80 and can be applied to other devices as well, including RISC-V multi-core processors such as the P870 and the P870-A provided by SiFive [21,22].

[21] Chishiro, Hiroyuki and Suito, Kazutoshi and Ito, Tsutomu and Maeda, Seiya and Azumi, Takuya and Funaoka, Kenji and Kato, Shinpei, "Towards Heterogeneous Computing Platforms for Autonomous Driving," In *Proceedings of IEEE International Conference on Embedded Software and Systems (ICESS)*, pp. 1–8, 2019

[22] SiFive. SiFive Performance P870/P870A. https://www.sifive.com/cores/performance-p870-p870a.

3 Response to 2$^{\text{nd}}$ reviewer

3.1 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 1 of Reviewer 2

Comment:

> I cannot catch the originality or novelty for the idea behind proposed method.

Our reply:

> Thank you for your detailed feedback and critical observations regarding our manuscript. We acknowledge the concerns raised about the originality and clarity of the proposed method and its significance compared to the existing schema such as SHIM 2.0. The novelty of the our proposed method is that we proposed a new schema to describe inter-core communication overhead.

Adjusted sentences in the "Introduction" section (Section 1)

To address these issues, a revised schema is proposed that measures communication overhead in per-API units for more precise estimates and also simplifies the schema's structure to reduce its volume. This enhancement facilitates ease of use and construction, allowing for a more accurate and comprehensive representation of communication overhead. Additionally, it improves the reliability and efficiency of performance evaluations in multi-core environments.

3.2 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 2 of Reviewer 2

Comment:

> I cannot understand what is the research problem to be solved in this paper. While the authors claim "a schema that can be expressed per API was proposed to solve the problem in the existing SHIM" in Conclusion, the problem itself is not clear at least for me.

Our reply:

> Thank you for your detailed feedback and critical observations regarding our manuscript. The problem is that the previous approach covers only the communication aspect and does not account for processes required for communication such as memory access and the acquisition of IDs for sending and receiving. Thus, the problem lies in the underestimation of communication overhead during estimation. And the lack of a concise description requires extensive documentation of the delay between cores, leading to a much larger description of the communication overhead. We have adjusted the sentences in the "Introduction" section to clarify the issue of SHIM 2.0.

Adjusted sentences in the "Introduction" section (Section 1)

The latest version of SHIM [14] currently still lacks an efficient method to describe inter-core communication latency, which poses significant challenges. This deficiency leads to substantial inaccuracies in time predictions within the Model-Based Parallelizer (MBP) due to the omission of inter-core communication latency calculations. Furthermore, the lack of a concise communication description necessitates extensive documentation of latency between cores, resulting in an increase of more than ten thousand lines for the communication overhead description.

3.3 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 3 of Reviewer 2

Comment:

In the first review, the meta reviewer mentioned that "Again, the performance improvement is due to the communication overhead estimation method (equation), not the schema representation. This paper does not have enough consistency among the objectives, methods, and evaluations." I feel this observation is very critical for this paper. The meta reviewer also mentioned that "This is not a problem with the schema but a problem with the overhead estimation method." However, the current form of the revised paper did not explain the reason why the proposed schema contributes to reducing the communication overhead.

Our reply:

Thank you for your detailed feedback and critical observations regarding our manuscript. First, we need to clarify that our proposed approach does not reduce the inter-core communication overhead. The contribution of our proposed method has two main aspects. On the one hand, it imports the inter-core communication overhead and thus reduces the error in the execution time prediction. On the other hand is the ability to greatly reduce the notation of the description file using our description method, thus facilitating import and editing.

In order to make our goal, methodology, and evaluation clearer, we have adjusted the composition and syntax of some parts of the article in the hope of better demonstrating the usefulness of our proposed methodology.

Adjusted sentences in the "Introduction" section (Section 1)

> The latest version of SHIM [14] currently still lacks an efficient method to describe inter-core communication latency, which poses significant challenges. This deficiency leads to substantial inaccuracies in time predictions within the Model-Based Parallelizer (MBP) due to the omission of inter-core communication latency calculations. Furthermore, the lack of a concise communication description necessitates extensive documentation of latency between cores, resulting in increased number of lines for the communication overhead description to more than ten thousand lines.

Adjusted sentences in the "Introduction" section (Section 1)

> To address these issues, a revised schema is proposed that measures communication overhead in per-API units for more precise estimates and also simplifies the schema's structure to reduce its volume. This enhancement facilitates ease of use and construction, allowing for a more accurate and comprehensive representation of communication overhead. Additionally, it improves the reliability and efficiency of performance evaluations in multi-core environments.

Adjusted sentences in the "Evaluation" section (Section 5.2.2)

In the given scenario, two separate linear equations are utilized to approximate the communication overhead for sending and receiving operations, respectively. This differentiation is crucial because the processes involved in sending and receiving data often have distinct characteristics and resource requirements, which can significantly impact communication overhead. The first linear equation, with a $Coefficient$ of 48.5 and an $Intercept$ of 12,999, is used to estimate the overhead for receiving operations. The coefficient here likely represents the incremental overhead per unit of data received, while the intercept might account for fixed costs associated with initiating or terminating a receive operation. The second linear equation, with a $Coefficient$ of 20.0 and an $Intercept$ of 40,800, is designed for sending operations. In this case, the lower coefficient suggests that the variable overhead per unit of data sent is less than that of receiving. However, the higher intercept indicates a greater fixed cost, possibly due to the complexities and resources involved in preparing and dispatching data.

By using two different formulas, the model can more accurately reflect the differing nature and costs of sending and receiving data, leading to a more precise and realistic estimation of communication overhead in systems where these activities may vary in frequency, size, and complexity.

3.4 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 4 of Reviewer 2

Comment:

Actually, the communication overhead is calculated by the formula $Coef_{i,j} * MS + Inte_{i,j}$ as shown in Line 17 of Algorithm 1. This is an estimation and there is no discussion why the extention of the exisitng schema is required. This is why I judge this paper did not fulfill the requirements of the 1st review by the meta reviewer.

Additionally, I cannot catch what is the essential point of the algorithm uniquely proopsed by the authors. What is the originality and novelty of this paper? To help readers understand the novelty, the authors should compare the algorithms for the case without proposed schema. This means that the authors should describe how the communication overhead is estimated in the existing schema (SHIM 2.0). Also, the authors should highlight the differences from the extended SHIM in [16] because Table 4 implies the different extension from this paper.

Our reply:

Thank you for your detailed feedback and critical observations regarding our manuscript. First of all, our goal is to achieve a higher precision estimation as well as a simpler way to describe it, so we propose a linear equation to describe the inter-core communication overhead. Existing models lack a simple way of describing the inter-core communication overhead, or require tens of thousands of lines to describe the inter-core communication overhead.

Adjusted sentences in the “Introduction” section (Section 1)

The latest version of SHIM [14] currently still lacks an efficient method to describe inter-core communication latency, which poses significant challenges. This deficiency leads to substantial inaccuracies in time predictions within the Model-Based Parallelizer (MBP) due to the omission of inter-core communication latency calculations. Furthermore, the lack of a concise communication description necessitates extensive documentation of latency between cores, resulting in increased text volume that complicates data importation and editing.

Adjusted sentences in the “Introduction” section (Section 1)

To address these issues, a revised schema is proposed that measures communication overhead in per-API units for more precise estimates and also simplifies the schema’s structure to reduce its volume. This enhancement facilitates ease of use and construction, allowing for a more accurate and comprehensive representation of communication overhead. Additionally, it improves the reliability and efficiency of performance evaluations in multi-core environments.

Adjusted sentences in the "Evaluation" section (Section 5.2.2)

In the given scenario, two separate linear equations are utilized to approximate the communication overhead for sending and receiving operations, respectively. This differentiation is crucial because the processes involved in sending and receiving data often have distinct characteristics and resource requirements, which can significantly impact communication overhead. The first linear equation, with a $Coefficient$ of 48.5 and an $Intercept$ of 12,999, is used to estimate the overhead for receiving operations. The coefficient here likely represents the incremental overhead per unit of data received, while the intercept might account for fixed costs associated with initiating or terminating a receive operation. The second linear equation, with a $Coefficient$ of 20.0 and an $Intercept$ of 40,800, is designed for sending operations. In this case, the lower coefficient suggests that the variable overhead per unit of data sent is less than that of receiving. However, the higher intercept indicates a greater fixed cost, possibly due to the complexities and resources involved in preparing and dispatching data.

By using two different formulas, the model can more accurately reflect the differing nature and costs of sending and receiving data, leading to a more precise and realistic estimation of communication overhead in systems where these activities may vary in frequency, size, and complexity.

3.5 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 5 of Reviewer 2

Comment:

Furthermore, the quality of English used in this paper is poor and it is very hard for reader to understand the body of contents. I think this paper is not valuable to readers.

Our reply:

Thank you for your feedback. I would appreciate it if you could specify multiple points regarding the aspects of the English language that need improvement.

Please note that I have previously used the advanced editing service from Enago for this manuscript, which includes checks by native English speakers in the field of computer science. The service covered:

- Language (expressions, grammar, etc.)
- Academic phrasing
- Formatting adjustments to meet submission guidelines
- Technical accuracy
- Academic writing style
- Paper structure
- Logical flow of arguments

For reference, here is the link to the service I used:
https://www.enago.jp/advanced-editing

Given this, I would be grateful for detailed feedback on how I can further improve the quality of the English in the paper.

3.6 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 6 of Reviewer 2

Comment:

I wonder why the Parts for ComponentSet and CommunicationSet, which could be seen in the title of the sub-sub-section, did not appear in Figure 1 as seen in the title of the other sub-sub-section.

Our reply:

Thank you for your insightful comments and suggestions regarding our manuscript. In response to the review, we have updated Figure 1 to include the parts for ComponentSet and CommunicationSet. We also made minor modifications to the original system model and have clarified once again that our proposal builds upon the existing SHIM framework. We appreciate your insightful comments which have guided these enhancements. Thank you for helping us improve our work.

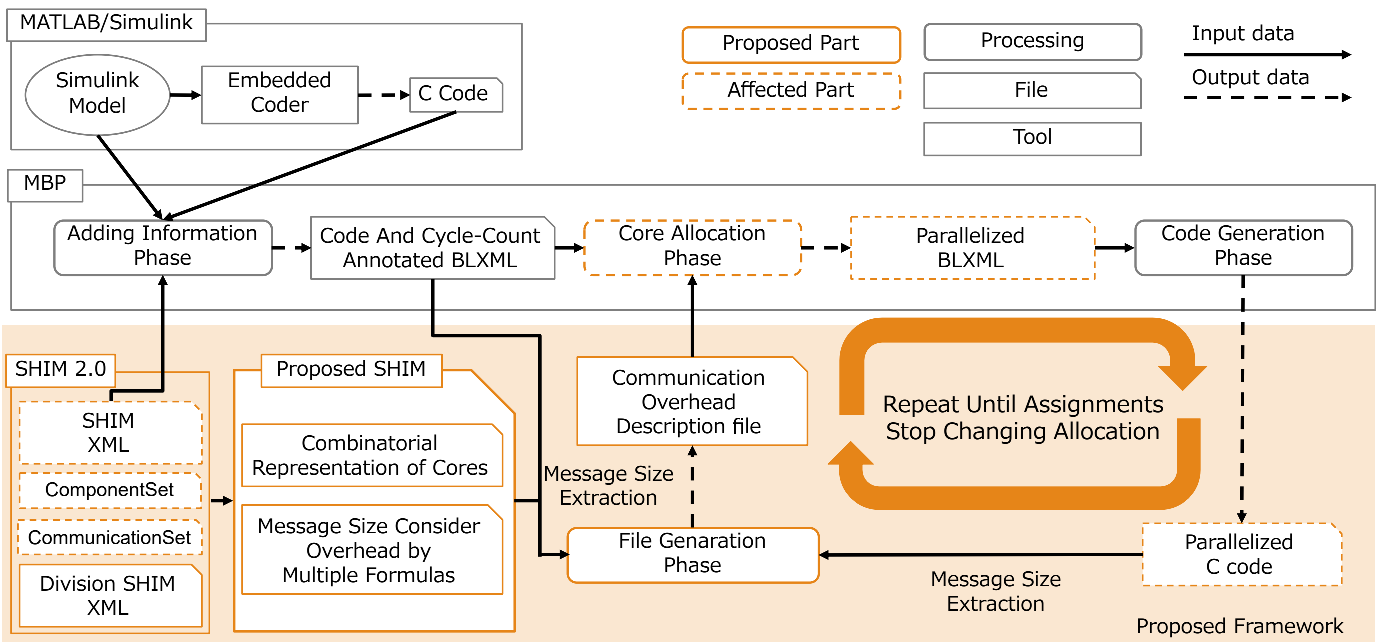

Fig.1 System model.

3.7 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 7 of Reviewer 2

Comment:

I cannot catch the discussion about multiple linear formulas in Section 5.2.2. In my understanding, all the authors show is a formula $Coef_{i,j} * MS + Inte_{i,j}$ of Algorithm 1 and this corresponds to the single linear formula. How about multiple linear formulas? The authors should present the actual equations to describe it.

Our reply:

Thank you for your insightful comments and suggestions regarding our manuscript. We apologize for any confusion caused by our initial lack of detailed description. To address this, we have expanded the discussion in Section 4.1 of our paper to better explain the choice of the model used in Algorithm 1 and the significance and importance of each parameter.

In Section 4.1, we clarify why this particular model was selected to describe the inter-core communication overhead. Furthermore, regarding your query about the multiple linear equations mentioned in Section 5.2.2, these are necessary because the inter-core communication involves both *send* and *receive* processes. Both processes are described using the linear models previously mentioned in Section 4.1. We hope this response addresses your concerns and clarifies our choice of modeling approach.

Adjusted sentences in the third paragraph of Section 4.1

In Line 3, initialize a $N \times N$-sized matrix A. Each entry $A_{i,j}$ in this matrix represents the overhead of communication from core i to core j. This matrix format allows for a structured representation of communication costs across all core pairs, facilitating efficient calculation and data management. In Line 8, there is no communication between the same cores ($i = j$). Therefore, the overhead is calculated as zero. This assumption simplifies the model by eliminating unnecessary calculations for non-existent self-communications.

Adjusted sentences in the forth paragraph of Section 4.1

Lines 9-16 obtain the parameters needed to calculate the communication overhead between the different cores ($i \neq j$); from the elements corresponding to Link and MS, the coefficients and intercepts are obtained. The coefficient in this context quantifies the incremental increase in communication overhead per unit increase in message size or other relevant metric. The intercept represents the baseline overhead associated with establishing a communication link, irrespective of the data quantity being transmitted.

Adjusted sentences in the fifth paragraph of Section 4.1

The coefficients $Coef_{i,j}$ represent the incremental delay or overhead added per unit of message size or other scaling factor, while the intercept $Inte_{i,j}$ represents the baseline overhead that exists even when no data is being transmitted. Calculate the communication overhead using the coefficients and intercept obtained in Line 17. It is particularly useful for quickly estimating communication overheads in complex multi-core systems where direct measurement for every possible communication scenario would be impractical.

3.8 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 8 of Reviewer 2

Comment:

> Additionally, what is the advantages for applying the proposed method to the actual applications? The authors should provide some evidences that make clear that the accuracy of estimation of communication overhead reflects to the gain of parallelization.

Our reply:

> Thank you very much for your insightful feedback on our manuscript. We sincerely appreciate the time and effort you have taken to review our work. We apologize for any confusion caused by the way our manuscript may have suggested improvements in parallelization precision. Our primary focus is on schema optimization, specifically aimed at reducing the complexity and volume of descriptions as well as improving the accuracy of execution time estimates.
>
> Accurate estimation of communication overhead is essential for determining core allocation, which is a crucial requirement in embedded systems. As demonstrated in our evaluation (Figure 8), our proposed method significantly enhances estimation accuracy. Regarding your concerns about commercial tools, we recognize that they are often black boxes that make it difficult for us to do more in-depth work or research. And enhanced parallelization tools are a separate area of research. We will consider this issue in our future work and incorporate it into our future research directions.

Adjusted sentences in the "Abstract"

Experimental results demonstrate that the average improvement rate in the error of the estimated communication overhead is at least 97.9%. Additionally, the schema description has been reduced by more than 90%, and the application of the proposed schema has further shortened execution time.

Adjusted sentences in the "Introduction" section (Section 1)

- This paper shows improvement in results by generating a communication overhead description file with an appropriate message size and using it in existing estimation methods. It demonstrates that the average improvement rate in the error of the estimated communication overhead is at least 97.9%.
- This paper proposes a new XML split specification to reduce the amount of description and the description of cores and combinations of cores by more than 90%, to reduce the burden on the users of SHIM.

Added sentences in the "Conclusion" section (Section 7)

In the future, based on the proposed communication description scheme, more accurate estimations are expected to be realized, which could be utilized in parallelization and contribute to the safety and real-time performance of autonomous driving.

3.9 Response to Peer Review Feedback: Addressing Non-Acceptance Reason 9 of Reviewer 2

Comment:

In Section 6: In the previous work [12] $->$ [16] It is unclear whether the this paper in the sentence "this paper shows a method to create an actual hardware SHIM based on our proposed approach" indicate [16] or the currently submitted paper.

Our reply:

Thank you for your diligent review and insightful feedback on our manuscript. We appreciate the time and effort you have taken to review our work. We apologize for the confusion regarding the citation in Section 6. Reviewer 2's comment is correct that there was an error in our reference. The current correct citation should be [19], indicating the previously published work. We have revised the manuscript to reflect this correction. Thank you once again for your valuable feedback. We hope this clarification resolves the issue.

Added sentences in the "Conclusion" section (Section 7)

In the previous work [19], we proposed a new schema for describing communication overhead per-API without relying on communication libraries.