CAAP Quarterly Report

04/06/2023

Project Name: " Accelerating Transition towards Sustainable, Precise, Reliable Hydrogen Infrastructure (Super-H2): Holistic Risk Assessment, Mitigation Measures, and Decision Support Platforms"

Contract Number: 693JK32250007CAAP

Prime University: North Dakota State University

Prepared By: Zhibin Lin (PI), Kevin Wang, Jace Anderson, Hong Pan, and Xingyu Wang

Contact Information: Zhibin Lin, Email: zhibin.lin@ndsu.edu, Phone: 701.231.7204

Reporting Period: 01/01/2023 – 04/01/2023

Project Activities for Reporting Period:

In the Quarter 1 report, Task 1 was completed. In this quarter (Quarter 2), the research team has worked on Tasks 2.1, 3.1, and 4.1. The summaries for the major activities that were completed during this reporting period are detailed below:

Task 2.1 Build an Automatic text extraction model to extract information and integration this information. During this reporting period, the research team (Dr. Zhibin Lin and Dr. Hong Pan from NDSU) built a natural language-based automatic feature extractor to transfer different information to the knowledge graph as summarized below.

Figure 1. The feature embedding process for XAI-based models.

1) Pipeline risk-related information is typically stored in various forms of text format, such as inspection reports, incident reports, regulatory codes, and other relevant documents. This information is critical for identifying potential hazards and risks associated with pipeline operations, and for evaluating the suitability of transporting hydrogen. Specifically, we have created an automatic model to transform text to a knowledge graph.

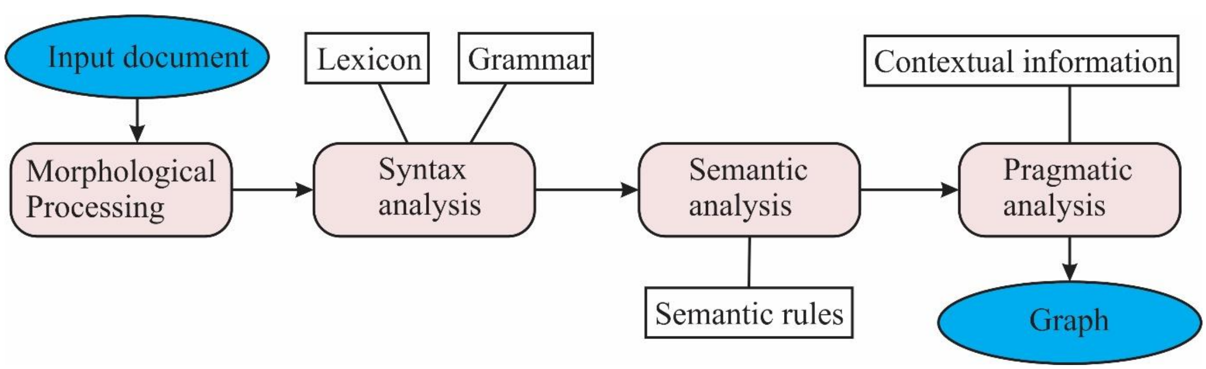


Figure 2. the logical steps in natural language processing

In Natural Language Processing (NLP), there are typically four phases: morphological processing, syntax analysis, semantic analysis, and pragmatic analysis as shown in Figure 2. Morphological processing involves breaking down language input into smaller units, while syntax analysis involves analyzing the grammatical structure of sentences. The semantic analysis focuses on understanding the meaning of the text, and pragmatic analysis considers the context in which the text is used to determine the most likely interpretation. Currently, deep learning, particularly fine-tuned language models, has become a dominant approach in natural language processing (NLP), showing remarkable success in various tasks such as text classification and sentiment analysis. Leveraging this technology, we have utilized different LLMs to construct our risk-based graph. Like risk-based bow tie models as show in Figure 3, each graph incorporates a scheme to connect different concepts for evaluating the target properties.

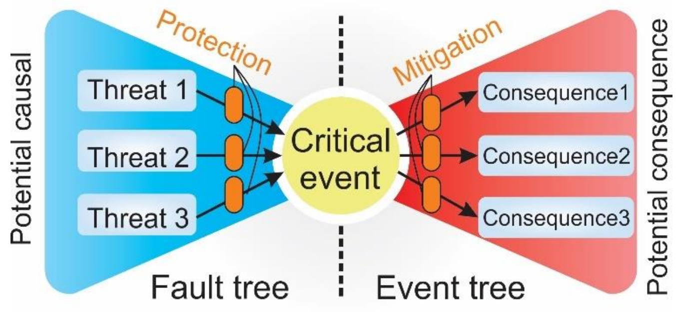


Figure 3. the Bow-tie risk assessment model scheme

2) As shown in Figure 4, We have designed a deep learning NLP process to extract any task specifical graph for evaluating the suitability of our existing pipeline. Given that our risk-associated documents are usually lengthy, a summary process is crucial to simplify and condense our target graph without losing important information. However, for certain graphs based on the Swiss cheese model, we use NLP to directly identify the entities and relations in the original document without the need for summarization. Subsequently, the task-specific scheme determines the entities and relations in our graph, which serve different purposes. For instance, in the context of hydrogen repurposing criteria, we emphasize different types of pipelines and the criteria that must be satisfied or unsatisfied for repurposing. On the other hand, for risk assessment, we focus on threats, protection, and mitigations. It is worth noting that this process can automatically acquire the target knowledge for us to evaluate the suitability of the existing pipeline without human intervention. That means an unprecedented, structured dataset size will be created, enabling us to make better decisions.

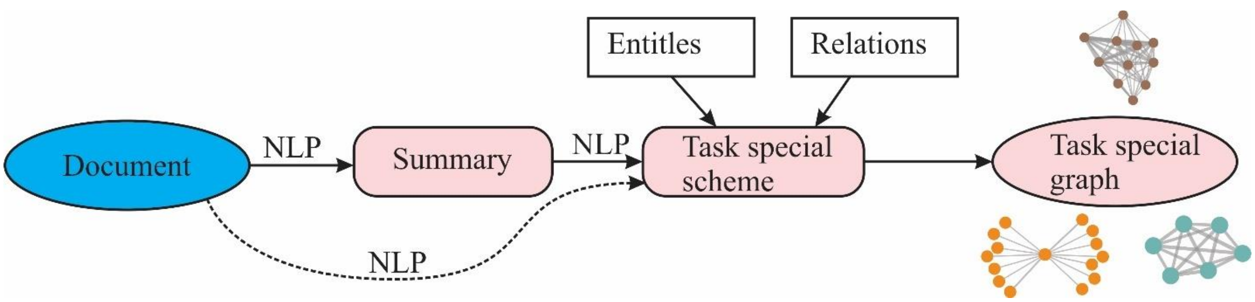


Figure 4. The natural language process (NLP) based domain specific graph process.

3) To demonstrate the correctness of our automatic knowledge graph creation. We download the incident report from PHMSA. This accident is happened in Oklahoma, Apr/08/2014. The main

cause of this accident is the internal corrosion. The contributing factor in this report is described as:

"Determining the cause of the failure relied on visual observations of the failed section when exposed in the field, review of operational records, interviews with operator personnel, and the results of the metallurgical testing.

When the failure point was observed in the ditch (prior to cut-out) by the PHMSA inspector, it was clear that the hole had formed at the 6 o'clock position on the lowest point of that portion of the line. This low point was further confirmed when the cut-out was removed, and both cut ends continued to drain their respective portions of remaining product.

Interviews with station operations personnel revealed that while the line did see turbulent flow when in use (up to 17,000 barrels per hour), it did not receive any cleaning pig operations (undiggable at time of failure) or other activities to ensure the hygiene of these lines against internal corrosion.

Ultimately, the cause of the failure was a result of internal corrosion. When determining the cause of the failure, discussions between PHMSA and the operator prompted a full line replacement within the station from the supply manifold to the tankage, rather than repairing the failed section with a pup. This line replacement included upgrading this portion of the line to allow maintenance pigging activities as a preventative measure against a similar type of failure in the future." (information from https://www.phmsa.dot.gov/safety-reports/failure-report-enterprise-crude-pipeline-llc-4812*)*

The NLP process outlined in Figure 4 is used to extract a graph from accident descriptions, as shown in Figure 5. The graph includes four root nodes: failure cause, factors contributing to failure, evidence, and actions taken to prevent future failures. This structured format is extracted automatically from the accident report text using NLP techniques. After this automatic extraction, downstream information integration processes can organize all accident reports in a uniform and knowledgeable manner.

Figure 5. The graph of the accident in Oklahoma, 2014

Similarly, we extracted a decision-related graph on hydrogen repurposing from the ACER document "Transporting Pure Hydrogen by Repurposing Existing Gas Infrastructure" using the summary paragraph from the document as input. Figure 6 shows the resulting graph, which indicates that the transportation of hydrogen is influenced by factors such as cost, flow rate, compression power, and more.

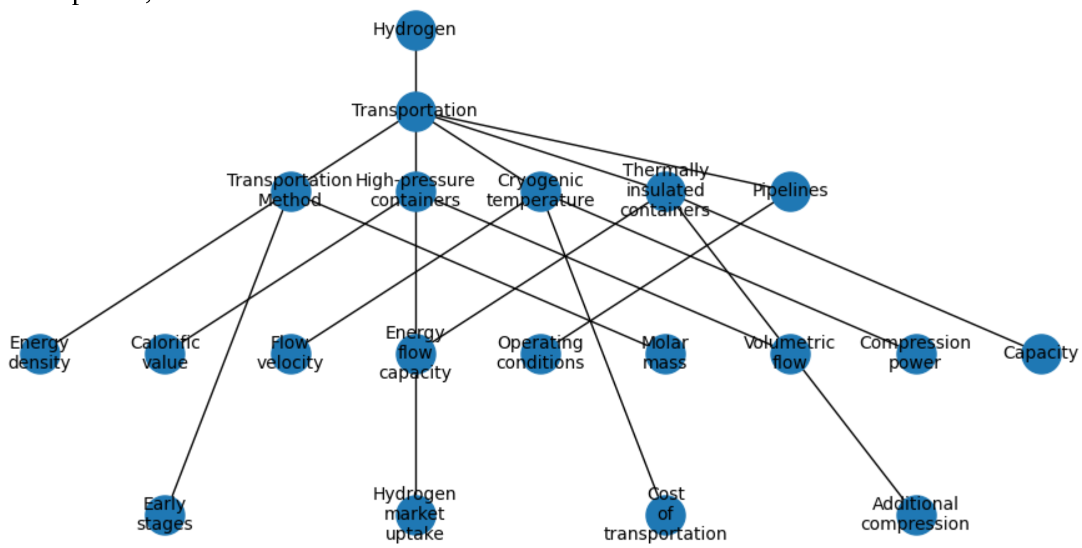


Figure 6. The decision graph for pipeline transport hydrogen decisions

Task 3.1 Preparing the near real-world testbed for hydrogen testing: During this reporting period, the research team (Mr. J. Anderson from EERC) have meted with a teammate and prepared his testbed according to the bi-weekly meeting. The progress on EERC for this reporting period is summarized below:

1) The EERC contracts department has successfully completed the contract with NDSU and the EERC can begin moving forward with their task.
2) The preliminary design from the early proposal phase has begun review with the EERC's Director of Design and Research to begin coordinating where/when to buy materials needed to construct the test stand and what changes/adjustments can be made to the design to improve efficiency while minimizing cost.
3) A kickoff meeting for a small group of EERC employees who will assist with the design and fabrication of the test stand will be scheduled for early April.

Task 4.1 Gaining an understanding of long-term hydrogen impacts: In the reporting period, the Virginia Tech team (Dr. K. Wang from Virginia Tech) focused on developing a multiscale computational model for predicting the long-term impact of hydrogen gas to pipeline materials and structures as detailed below.

1) Hydrogen-induced material degradation (e.g., embrittlement) and damage (e.g., fracture) is a long-term issue that often occurs after years of exposure to hydrogen gas. It is also sensitive to material defects, pre-stress, and the pressure of the hydrogen gas. Due to these complexities, it is difficult to collect test data and build reliable empirical models. Physics-based models, in comparison, allow predictions to be made based on fundamental laws (i.e. first principles) and verified causal relations. It requires less data to be collected on the application scale over a long period of time.

The fundamental physical process that leads to material damage is the dissociation of hydrogen molecules on the material's surface (i.e., adsorption) and the motion of hydrogen atoms within the material's lattice and microstructure (i.e. absorption). The dynamics of individual hydrogen molecules and atoms has a length scale of Angstroms (10^{-10} m) and a time scale of picoseconds (10^{-12} s) or less. These scales are many orders of magnitude lower than the application length and time scales, namely meters and years. A key research goal of the Virginia Tech team is to connect different physical models at atomistic and continuum scales using the diffusive molecular dynamics (DMD) method (Figure 7). Previously, the team has implemented the DMD solver in a parallel C++ code and applied it to analyze the absorption the hydrogen by palladium (Pd) nanomaterials. In the first year of the project, the team will focus on expanding DMD to model and analyze the adsorption and absorption of hydrogen in different types of pipeline steels (alloys) with pre-stress and microscopic defects.

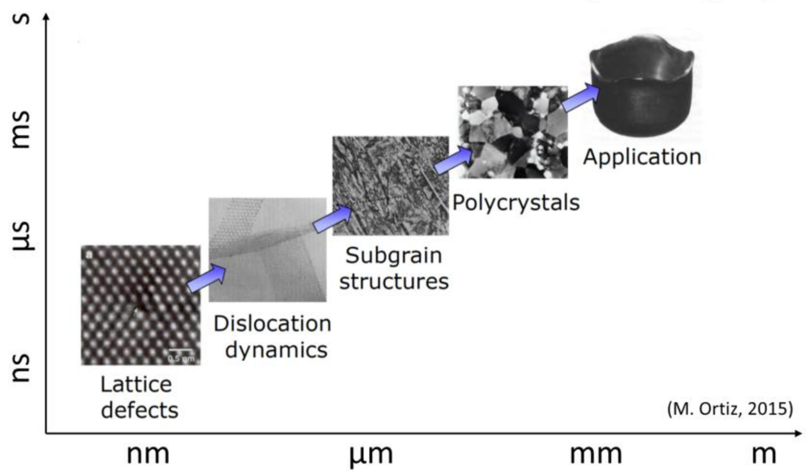
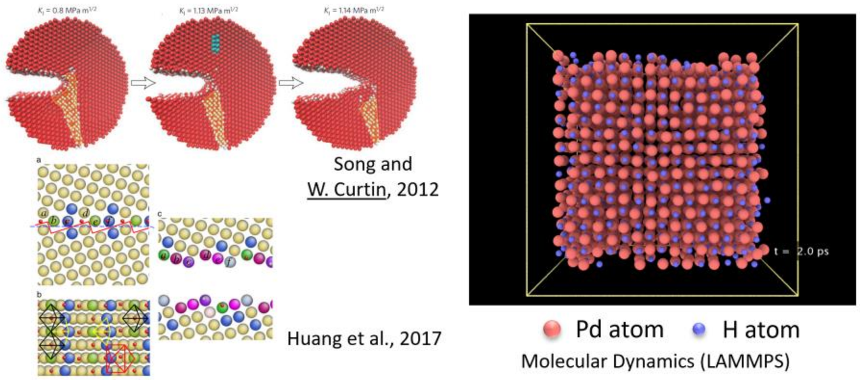
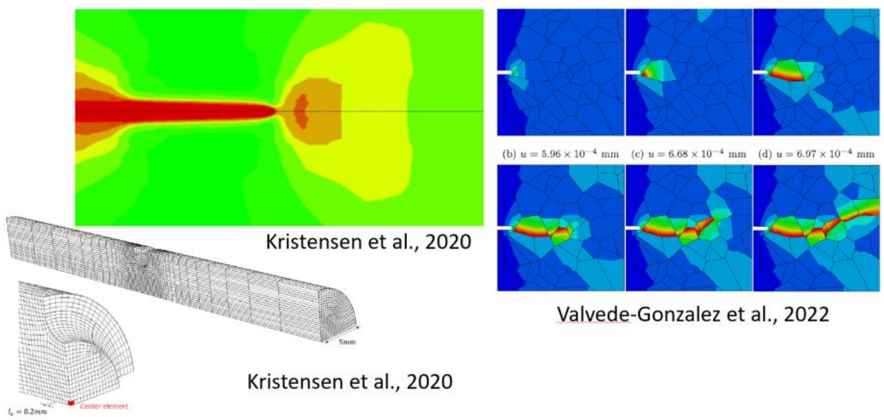


Figure 7. Solid material at different lengths and time scales

2) According to a comprehensive literature review, we found that most studies in the past have focused on either the atomistic scale or the continuum scale. A major limitation is that the connections between the findings obtained at these different scales are still unclear. At the atomistic scale (Figure 8(a)), widely used methods include molecular dynamics, molecular statics (i.e., neglecting thermal vibration, but retaining interatomic and chemical potentials), and crystallography. The main advantage of these methods is that they explicitly resolve individual atoms of the solute (Hydrogen) and the solvent (e.g., Iron). The fundamental questions addressed in these studies include (1) How does H influence the propagation of a crack tip at atomic scale? (2) What are the effects of atomic-scale features (e.g., impurities, defects, and grain boundaries)? The main issue of the atomistic scale methods is that the length and time scales are highly limited. As a result, direct comparison between simulation and experimental results is challenging, and rarely conducted.
   At the continuum scale, popular methods include hydrogen diffusion models, material degradation models, and static mechanical equilibrium analysis (e.g., using finite element method) (Figure 8(b)). These methods do not resolve individual atoms or molecules. But they can represent grain boundaries as internal boundaries in the computational domain. The fundamental questions addressed in these studies include (1) How does hydrogen influence the propagation of a crack in a single- or poly-crystalline material? (2) What are the effects of hydrogen pressure, material type, grain boundary, and impurities? A weakness of these continuum-scale methods is that they rely on empirical models to account for the adsorption and absorption of hydrogen. Also, comparison between simulation and experiment is rare, due to the long-time scale (years) and difficulties in collecting experimental data.



(a)



(b)

Figure 8. Literature review of computational models and analysis of hydrogen embrittlement

3) We have uploaded the source code of our DMD solver to GitHub (www.github.com/kevinwgy). The DMD solver was developed to simulate the transport of hydrogen in palladium nanomaterials over a long period of time. It couples an atomistic non-equilibrium thermodynamics model with an empirical diffusion law. To model hydrogen diffusion in pipeline material (e.g., steels), the solver needs to be generalized in several aspects. First, palladium has an FCC (face-centered cubic) lattice structure, whereas steels have either the FCC (face-centered cubic) or the BCC (body-centered cubic) lattice. Also, the DMD model requires an interatomic potential of the alloy as an input. Previously, we have implemented two EAM (embedded atom method) potentials for the Pd-H system. To model hydrogen diffusion in steels, new interatomic potentials need to be added to the solver.

   Currently, we are modifying and extending the DMD solver to account for different lattice structures. In the next step, we will identify and implement interatomic potentials for pipeline steel materials into the solver. Then, deformation-diffusion coupled simulations can be performed to predict hydrogen absorption and potential material damage.

Project Financial Activities Incurred during the Reporting Period:

The cost breakdown during the reporting period in each category according to the budget proposal is shown in Table 1.

Table 1 Cost breakdown during the reporting period (Q2)

Category	Amount spent during Q2
Personnel	
Faculty	$0
Postdoc	$9,000
Students (RA and UR)	$500
Benefits	$6,000
Operating Expenses	
Travel	$0
Materials and Supplies	$0
Recharge Center Fee	$0
Consultant Fee	$0
Subcontracts	Subawards issued
Indirect Costs	$13,341

Project Activities with Cost Share Partners:

The Match fund from NDSU for this project is coming from faculty academy hours of NDSU and Virginia Tech. The graduate students who are working on this project during the reporting period (Q2) receive teaching assistance support, so there is no RA tuition wavier during this period used for the match fund. More match fund will be generated during the summer at Q3 when we will have graduate student RA tuition wavier. The match fund from Dr. Lin (NDSU) and Dr. Wang (Virginia Tech) during Q2 is estimated to be $13,341.

Project Activities with External Partners:

During this reporting period, the subawards to the sub-universities were established. The research team meets on a regular basis bi-weekly and the sub-universities have conducted research as planned.

Potential Project Risks:

No potential risks were noticed during this reporting period.

Future Project Work:

In the next quarter, the research team will continue working on Tasks 2.2, 3.1, 4.1, and expand the work onto Task 2.1 as planned.

Potential Impacts to Pipeline Safety:

The graph resulting from text extracted by automatic NLP shows the potential of the proposed framework to accelerate the analysis of vast amounts of risk and incident data. With these uniform and structured datasets, decision-makers will be able to effectively find better statistical evidence to support safety-related decisions.