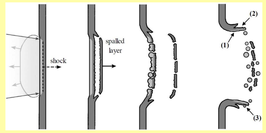
OpenMSI: A Materials Semantic Infrastructure for Streaming Data Integration in Accelerated Materials Development

Margaret Eminizer*, Jacob Diamond**, Christopher DiMarco**, Samuel Tabrisky***, Angela Torrejon****, Somanth Ghosh†, Todd Hufnagel††, K.T. Ramesh†††, David Elbert**

*JHU IDIES, **JHU Hopkins Extreme Materials Institute, ***Dartmouth College, ****Duke University, †JHU Dept of Civil and Systems Engineering, ††JHU Department of Materials Science and Engineering, †††JHU Department of Mechanical Engineering

elbert@jhu.edu

Award #: 1921959



Rességuier et al., 2014
https://doi.org/10.3390/met4040490

Science Driver: Spall fracture is shock-induced, dynamic fracture that is of practical and scientific importance. Spall fracture occurs during rapid loading by explosives, impact, or direct energy deposition. Spall experiments make it possible to reach stress levels approaching the theoretical strength of a material and probe limiting mechanisms of material failure.

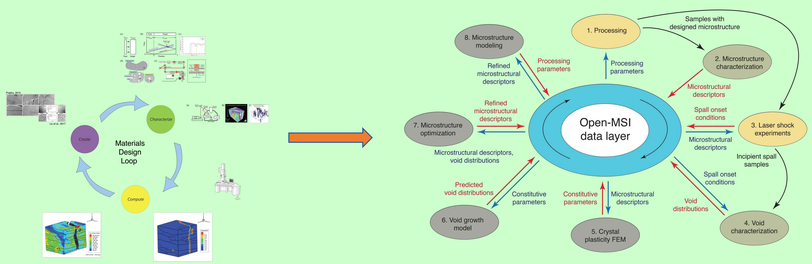
Data Drivers:
- Integration of multiple experimental and computational sources of information
- Acceleration of iterative process to understand mechanisms and refine material processing to create better alloy

Solutions:
- Streaming data integration (Apache Kafka/Confluent Cloud)
- Integrated semantics with metadata capture
- **OpenMSI python package to extend Apache Kafka to use case https://github.com/openmsi/openmsipython/**

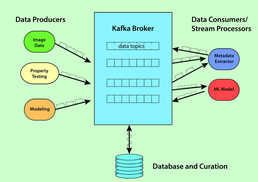
Abstract: The OpenMSI project establishes a new paradigm for the materials discovery and design loop centered on flow of data rather than individual modeling or experimental tasks. Integrated data flow accelerates the iterative materials research process and provides a framework for Materials Genome Initiative science. OpenMSI centers on an open semantic infrastructure and streaming data platform based on Apache Kafka to integrate the processing, experimental, and modeling components of materials design. The openmsipython package (https://github.com/openmsi/openmsipython/) extends Apache Kafka for serialization/deserialization and chunking of large files and efficient streaming. Asynchronous data producers and consumers run as services or daemons on laboratory control nodes providing a loosely coupled architecture that is efficient to deploy, maintain, and extend.

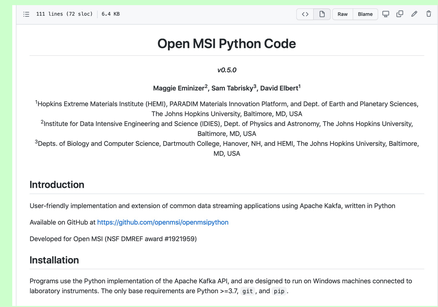
Infrastructure development is embedded in a science program focused on creating aluminum alloys resistant to spall failure in high-energy environments. Such alloys have high value in aircraft and spacecraft while understanding the underlying mechanism of failure has broad scientific importance in understanding ultimate material strength. Experimental characterization utilizes laser-induced shock waves to drive foil micro-flyers to impact samples and effect spall failure. Micro-flyer velocities are captured with photon doppler velocimetry (PDV) in one-dimensional FFT traces of interferometry and include our first data streamed through OpenMSI services. PDV data streams are asynchronously consumed to automate analysis and provide rapid feedback during repeated shock experiments.

1. Map Materials Design Loop onto Data Driven Streaming Infrastructure



2. Extend Apache Kafka for Laboratory Use

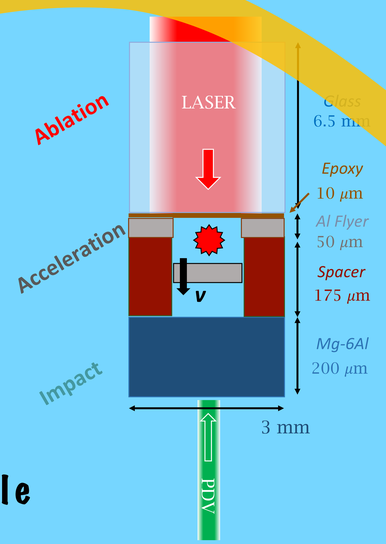




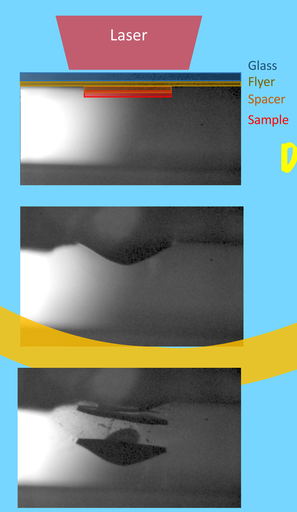
3. Laser Shock Experimental Lab

Experiment Schematic

Laser, Optics, Control, Micro-Flyer, and Sample

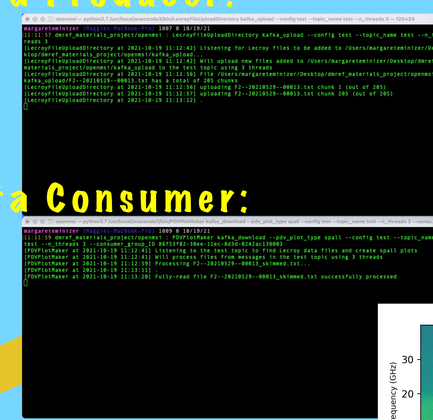


Photon Doppler Velocimetry of Flyer

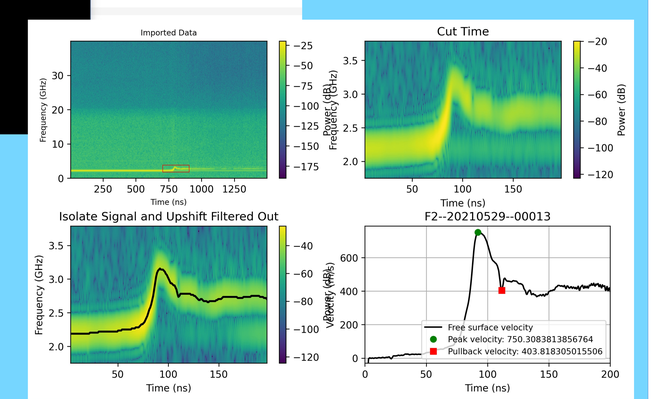


High-Speed Imagery

Data Producer:



Data Consumer:



Consumer PDV Analysis