# Integrating AFM and STEM Digital Twins with LLMs for Automated Data Interpretation

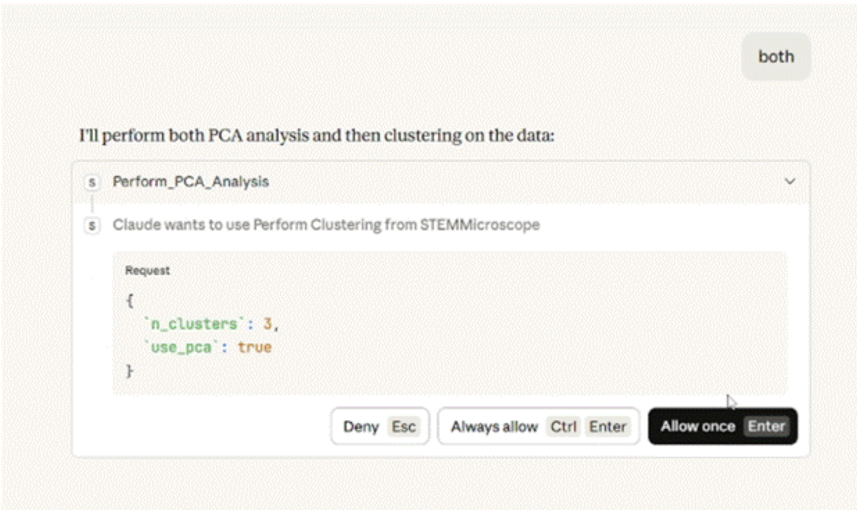
**Josep Cruañes** **Fanzhi Su**

Modern materials characterization via STEM and AFM generates high-dimensional datasets that require significant domain expertise to process, creating a technical bottleneck for non-expert researchers. This work introduces Model Context Protocol (MCP) servers that bridge Claude AI with microscopy digital twins, allowing users to execute complex analytical workflows—such as automated image segmentation and HDF5 data parsing—through natural language. By functioning as a conversational "AI co-pilot", this framework democratizes access to sophisticated microscopy, reducing cognitive load and accelerating the transition from raw binary data to actionable scientific insight.

## 1.STEM Microscopy MCP Server

**Architecture:** The server creates a bridge between Claude AI and a digital twin of a Scanning Transmission Electron Microscope using locally calls. Built on the FastMCP framework, it provides a conversational interface to complex microscopy workflows.
**Data Management:** Implements tools for initializing microscope connections, downloading remote data files, and registering H5 format datasets with comprehensive error handling and path validation.
**Spectroscopic Capabilities:** Two primary modes enable flexible data collection: point spectroscopy extracts spectra from specific (x,y) coordinates, while grid spectroscopy collects data across customizable grids for comprehensive sample mapping.


**Analysis Pipeline:** Integrates scikit-learn algorithms for Principal Component Analysis (dimensional reduction while preserving variance) and K-means clustering (automatic identification of material phases). All results are automatically visualized using matplotlib and saved with timestamped filenames.



**Workflow:** Users control the entire analysis through natural language: initialize connection, retrieve overview image, collect grid spectra, perform PCA, execute clustering, and generate spatial cluster maps overlaying the sample image. Claude orchestrates all tool calls and provides contextual interpretation.

## 2. **Segmentation MCP Server**

**Model Development:** A custom Convolutional Neural Network with U-Net inspired encoder-decoder architecture was trained using PyTorch for segmentation. The training pipeline includes a data augmentation strategy generating 100 random 32×32 pixel patches from original microscopy images, with paired cropping ensuring ground truth labels remain aligned.

**Integration:** HyperSpy integration ensures compatibility with standard STEM data formats. The model includes integrated peak detection functionality to identify particle centers from segmentation output.

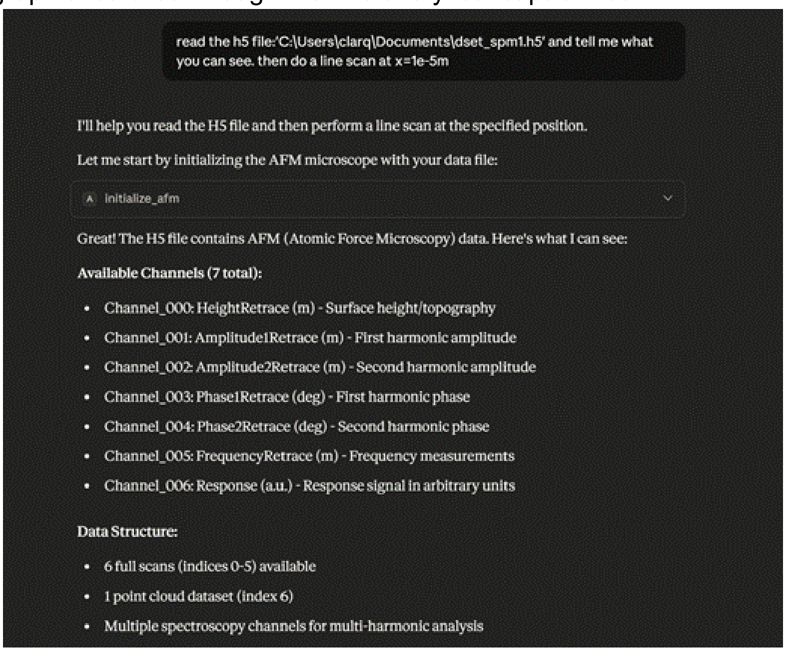
**Server Functionality:** Provides model status verification and image segmentation with side-by-side visualizations. Implements lazy loading to optimize startup time and memory. The model performs excellently in standalone testing; integration challenges with MCP protocol remain under investigation.

## Technical Implementation

Both servers utilize FastMCP for standardized tool registration with descriptive metadata. Comprehensive error handling includes file existence checks, connection verification, and input validation. The visualization pipeline uses matplotlib with professional styling, saving outputs to a dedicated directory with absolute paths for Claude reference.

## 3. AFM Microscopy MCP Server

Parallel to the STEM microscopy integration, the Atomic Force Microscopy (AFM) digital twin is bridged to Claude AI through a dedicated MCP server. This architecture enables the AI to process local file system calls, allowing it to ingest HF5-formatted datasets provided by the user. Once connected, the AI can programmatically inspect the file metadata, interpret the underlying data structure, and identify available signal channels—tasks that are inherently infeasible for standard LLMs due to their inability to interact with binary file systems. This framework not only enables automated data discovery but also provides a foundation for the high-level interpretation of AFM spectra and topographic features through the AI's analytical capabilities.




## Conclusion

These MCP servers demonstrate the potential of AI-assisted materials characterization by creating intuitive, conversational interfaces for complex analytical workflows. This implementation serves as a robust proof of concept, confirming that the MCP framework effectively mitigates "data silos" by allowing researchers to interact with specialized binary formats via natural language. The STEM and AFM servers are now fully operational and ready for research applications, enabling non-expert users to perform sophisticated microscopy analysis through streamlined dialogue. Ultimately, this integration proves that AI-driven digital twins can significantly lower the barrier to complex data characterization, providing a scalable blueprint for future autonomous laboratories and accelerated scientific discovery.