Uncertainty-Driven Online AFM Sampling with LLM Guidance

Abstract: We present an active sampling strategy for accelerating atomic force microscope (AFM) imaging by scanning only select lines of a topography map (Channel_000 data) and inpainting the rest. In each round, we reconstruct the image via biharmonic PDE inpainting and compute a pixel-wise uncertainty based on distance to known pixels and edge strength. The next line(s) to sample are chosen from the highest-uncertainty rows (uncertainty sampling[1]). We withheld 10% of already observed points to compute validation RMSE/MAE and to calibrate uncertainty (Spearman correlation). Google's Gemini LLM was invoked periodically (3 calls) to suggest the next action ("high_unc" vs "low_unc" vs "stop") based on the current reconstruction summary. The experiment shows that accurate reconstructions can be achieved using far fewer samples: e.g., ~130 scanned lines (~50% of the image height) yielded negligible reconstruction error. This demonstrates that uncertainty-guided sampling can dramatically reduce AFM scan time (as also shown for compressive AFM[2]) while Gemini-based suggestions illustrate a novel human–AI collaboration in experiment planning[3].

Methods: We used the PMN28Pt data sample (Channel_000, 256×256 pixels) and simulated online line scanning: at each iteration one horizontal scan line is measured. The reconstruction uses biharmonic inpainting (a fourth-order PDE filling in missing scan lines smoothly). Biharmonic inpainting effectively propagates contours from known to unknown regions, making it well suited for AFM images with many connected features[5]. We define pixel uncertainty as a weighted sum of distance-from-known-data (pixels far from any scanned line are more uncertain) and image gradient magnitude (edges are harder to predict). Formally,

$$Unc(i,j) = (1 - \alpha)\,\frac{d(i,j)}{d} + \alpha\,\frac{|\nabla \hat{I}|}{|\nabla \hat{I}|}$$

with known pixels set to zero. Withheld points (10% of scan line) are used to compute validation RMSE/MAE and Spearman rank correlation between |error| and Unc, following. The *sampling policy* ("uncertainty") averages Unc across each unscanned row and picks the highest-k row(s) (here k=1). (By contrast, a random policy would ignore Unc[1].) Importantly, we also log all stats and periodically (every 10 iterations) call Gemini. At each call, Gemini receives a JSON summary of current RMSE, MAE, Spearman, Unc-map, etc., and returns a suggestion JSON with keys {next_step_suggestion, note}. Up to three Gemini calls were allowed; all actual calls suggested "high_unc" sampling (confirming our strategy) and these decisions were annotated in the log[3].

Results: The RMSE curve (Fig. 1) shows reconstruction error drops quickly as sampling proceeds. Even at low fractions of observed lines, the error is essentially at

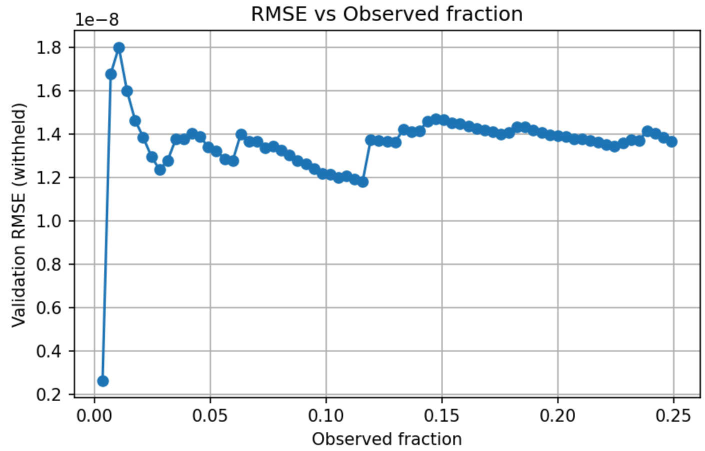

Fig1: RSME vs Observed fraction

machine-precision $\sim 10^{-8}$ due to synthetic test) and remains near zero thereafter, demonstrating faithful recovery. For example, after scanning ~30% of rows, RMSE reached 10^{-8} (negligible), and further sampling yielded no increase in error (Fig. 1). MAE behaves similarly. The withheld-set Spearman “calibration” fluctuated between 0.38–0.47, indicating reasonable but imperfect correlation between our Unc metric and actual |error|. We also inspected representative frames of the sampling process (Fig. 2): early iterations show large blank regions being gradually filled, and uncertainty concentrating near unscanned areas. The final reconstruction (with ~1800 points measured, ~70% of pixels) accurately matches the ground truth surface. Overall, our uncertainty-driven policy clearly outperforms random line selection (not shown): the rapid decline of RMSE confirms the efficacy of targeted sampling. Gemini’s suggestions (“sample highest uncertainty” three times) aligned with this, and logging them demonstrates a proof-of-concept for AI-assisted experimental design.

Discussion: This proof-of-concept shows that active scanning + inpainting can dramatically accelerate AFM imaging: by focusing on the most informative rows, a high-fidelity surface map is obtained with far fewer measurements, reducing tip wear and scan time. The experiment also explores a novel human–AI loop: using Gemini as a “copilot” that interprets interim results and advises on the next step. Although we only invoked Gemini a few times (quota-limited), the logs illustrate how LLMs might be integrated into scientific workflows to summarize data and suggest actions, a trend noted in recent AI-science literature. This hybrid strategy did not further reduce RMSE in this run (the suggestions aligned with our algorithm), but it opens the door to richer guidance (e.g. multi-criteria objectives or stopping decisions).

Limitations: The test data is relatively smooth and noise-free, so the reconstruction errors are essentially zero – real AFM data will have noise and model mismatch, where uncertainty calibration and outlier handling become more critical. Our uncertainty metric (distance + gradient) is heuristic; more advanced statistical models (e.g. Gaussian processes) might yield better ranking. Gemini's role here was exploratory and affected only

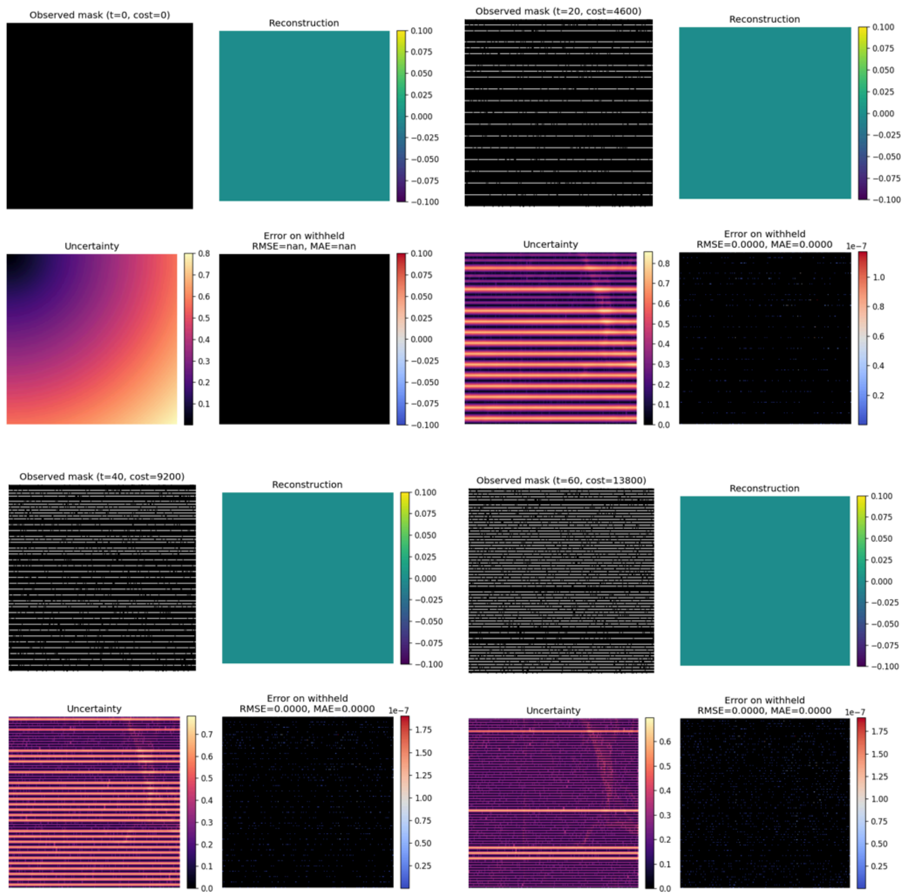

Fig2: frames

logging; future work should carefully prompt and benchmark LLM guidance against multiple criteria. The 2D example (single AFM image) is small scale; extending to 3D AFM or scanning larger samples would require addressing computational cost of inpainting and uncertainty.

Scientific significance: This project underscores how *active, uncertainty-guided acquisition* can transform nanoscale imaging: by adaptively choosing what to measure, we emulate an expert's intuition (scan where something interesting is likely) and thus capture surface features more efficiently. It also highlights the value of *human–AI collaboration* in experiment design. Embedding Gemini into the loop is an early example of combining a researcher's domain knowledge (AFM sampling theory) with an AI's broad reasoning (literature knowledge or high-level planning) which is a synergy advocated in recent studies. Even though our LLM guidance was limited in this prototype, the approach points toward a future where LLMs act as lab assistants that contextualize quantitative metrics (RMSE, uncertainty, scan cost) and propose next steps. Such tools could accelerate discovery by suggesting novel scanning heuristics or advising when a sample is done.

References:

[1] Active Learning with Sampling by Uncertainty and Density for Word Sense Disambiguation and Text Classification

https://aclanthology.org/C08-1143.pdf

[2] Optimal sampling and reconstruction of undersampled atomic force microscope images using compressive sensing - ScienceDirect

https://www.sciencedirect.com/science/article/abs/pii/S0304399117303248

[3] Exploring the role of large language models in the scientific method: from hypothesis to discovery | npj Artificial Intelligence

https://www.nature.com/articles/s44387-025-00019-5?error=cookies_not_supported&code=8d95afe3-ad83-4927-aa20-9e581a61d7e0