APEX INSTITUTE OF TECHNOLOGY

Weekly Lab Meeting – Progress Update

Sparse-attention forecasting for high-frequency sensor streams

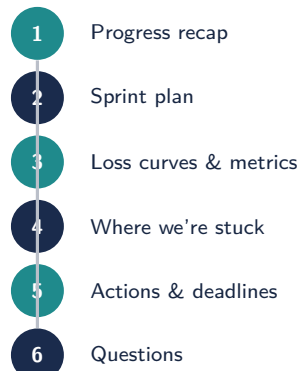
MC

Maya Chen
PhD Candidate, Vertex Lab

Week 31 ▪ August 3, 2026 ▪ PI: Prof. David Osei

Today's Agenda

1. Last week: what shipped
2. This week: the plan
3. Results: baseline vs. Meridian
4. Blockers needing help
5. Next steps & owners
6. Open discussion



Last Week – What Shipped

✔ Completed

Finished the sparse-attention ablation across 4 window sizes (64 / 128 / 256 / 512) on the Nexa cluster.

Rebuilt the Nimbus-Weather ingestion pipeline; resampling bug that duplicated the 03:00 UTC tick is fixed.

Wired the new Synapse quantile loss into the training loop and confirmed gradients are finite in FP32.

Wrote up the Week-30 ablation notes for the shared lab wiki.

18	412	0.71

This Week – The Plan

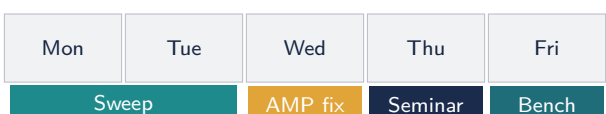
Planned Work

Run the full hyperparameter sweep (learning rate × window size, 20 configs) on the held-out validation split.

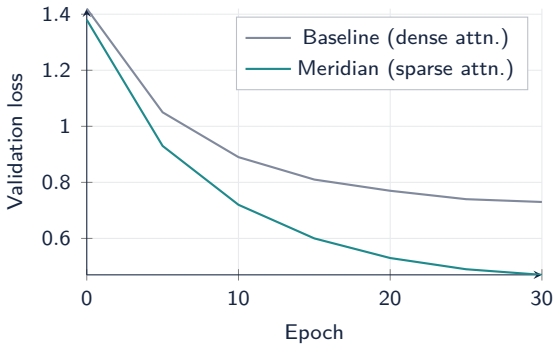
Integrate the Synapse quantile loss with mixed precision once the NaN issue below is resolved.

Draft the mid-cycle slide for the Apex Institute cross-lab seminar (due Friday).

Pair with Ravi on the streaming inference benchmark for the Nimbus-Weather live feed.



Results – Baseline vs. Meridian



Reading the curve:

Meridian converges to a lower loss by epoch 15.

Gap widens with longer context windows (512 tokens).

Sparse attention cuts memory 2.3× at equal batch size.

Held-out MASE:

Baseline: 0.73 Meridian: **0.47**

Blockers – Where I Need Help

⚠ GPU queue contention on the Nexa cluster

The hyperparameter sweep needs 20 concurrent jobs; shared queue caps us at 6. **Ask:** can Prof. Osei request a priority allocation from the Vertex compute committee for this week?

⚠ Nimbus-Weather licence renewal pending

Legal review of the extended data agreement has not closed; live-feed access is frozen. **Mitigation:** working from the already-licensed 2025 offline snapshot in the meantime.

⚠ Synapse loss $\times$ mixed precision $\rightarrow$ NaN gradients

The quantile term overflows under FP16 when residuals exceed ± 40. **Mitigation:** pinned to Synapse v2.3 and disabled AMP for this term; full fix targeted for Thursday.

Next Steps – Actions & Owners

Task	Owner	Deadline	Status
Full hyperparameter sweep (20 configs)	Maya Chen	Aug 6	In progress
Escalate Nexa priority queue request	D. Osei	Aug 4	Blocked
Fix Synapse loss under mixed precision	Elena Popescu	Aug 7	In progress
Streaming inference benchmark	Ravi Patel	Aug 8	Not started
Nimbus data licence follow-up	D. Osei	Aug 6	Blocked
Cross-lab seminar slide draft	Maya Chen	Aug 7	On track

Legend: On track · In progress · Blocked · Not started

Open Discussion

Is a 2.3× memory reduction worth pursuing a 512-token window on next month's Nexa allocation, or should we cap at 256 and bank the compute?

Should the quantile loss ship with a hard clip on residuals, or is a soft Huber-style transition safer for the Nimbus-Weather tail events?

Do we present the Week-30 ablation at the Apex cross-lab seminar, or hold for the sweep results first?

Prior work this sweep builds on:

[Vaswani et al., 2017] for the base attention mechanism,
[Zhou et al., 2021] and [Nie et al., 2023] for long-horizon sparse variants,
[Oreshkin et al., 2020] as the pure-forecasting baseline family.

References I

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115, 2021.