HAC Telecon

- 2nd Thursday (4pm EDT) every month from Sep-May (except summer)

- Projects led by the HAC team will be presented in a rolling fashion, each project one time.

- Next time: 4 pm-5pm EDT on Oct 12th

    Siqi is going to talk about sparse transformer for tracking.

HAC & MMA

▸ GNN for Icecube

▸ Ignacio Taboada & Pan Li initiated some discussion
  - New affiliates?

▸ Expected to have some prelim results in this semester

Algorithm-System-Hardware Co-Design for Efficient Point Cloud Processing

Team at MIT: **Zhijian Liu, Haotian Tang, Yujun Lin, Song Han**

- **Achievements in Q3 2023**:
  - TorchSparse++ is accepted to MICRO 2023. Prepare code release.
    - Achieved **2.9X** speedup over MinkowskiEngine (from NVIDIA).
    - Achieved **3.3X** and **1.8X** speedups over SpConv v1/v2 (from TuSimple).
  - Presented BEVFusion at ICRA 2023.
  - Presented FlatFormer, SparseViT, and BEVFusion-R at CVPR 2023.
  - Explored activation sparsity in other applications.
    - SparseRefine (submitted to ICLR 2023) reduces the latency of image segmentation by **1.5-3.9X**.
    - The input to video and diffusion models can be sparsified by 50% without loss of accuracy.

Scalable Geometric Deep Learning

Team: **Siqi Miao, Pan Li, Javier Duarte, Mia Liu, Kilian Lieret, Gage Dezoort**

- **Achievements in Q3 2023:**
  - Scalable geometric deep learning (GDL)
    - <u>For the tracking task</u>
      - Successfully replaced previous SOTA GNN with a Sparse Transformer
        - SOTA GNN performance:
          - 90.88% acc, 95.69% recall
          - In O(E), irregular graph operations
        - Our Sparse Transformer:
          - 93.86% acc, 99.17% recall
          - In O(NlogN), regular matrix multiplications, no irregular operations on graphs
        - 100x+ Speedup on GPUs
- **Plans in Oct:**
  - Summarize the lessons we learned from the tracking task.
  - Try our Sparse Trans to other GDL tasks (Pileup mitigation & MLPF)

Domain Adaptation on Graph Machine Learning & OOD

*Team: **Shikun Liu, Yongbin Feng (Fermi), Nhan Tran (Fermi), Han Zhao (UIUC), Pan Li***
*OOD team: **Deyu Zou, Siqi Miao, Shikun Liu, Pan Li, Shiyu Chang (UCSB), Victor Fung (GT)***

- **Achievements in Q3:**
  - Follow up on StruRW++
    - Finalize the pipeline StruRW++ that can handle both label shift and conditional shift on graphs
    - Implement the overall pipeline and achieve expected results on synthetic dataset in its effectiveness of handling conditional shift
    - Currently working on combining label shift and form a more stabilized training for real datasets
  - GDL-DS benchmark (geometric deep learning – distribution shift)
    - Finish submitting the paper to ICLR 2024

- **Plan in October:**
  - Organize the codes for GDL-DS benchmark and publish the arxiv version
  - Keep testing the StruRW++ on real datasets and get some prelimary results
  - (Long term) Apply StruRW++ to semi-supervised pileup mitigation

Hardware Accelerated Deep Learning for Sleep Spindle Detection

Team: **Rajeev Botadra, Xiaohan Liu, Scott Hauck**

- **Achievements in Sep** :
  - Present findings in thesis defense (Xiaohan)
    - Reported performance characteristics of quantized modified LFADs model deployed on Xilinx U55C.
  - Transfer of Knowledge (Rajeev)
    - Catch-up on LFADs, MRAE, and Transformers architecture.
    - Complete environment setup for model deployment from HLS to RTL.
- **Objectives for October**:
  - Test MRAE performance with and without Gaussian noise generation
  - MRAE implementation in HLS4ML, simulated performance and resource demands
  - Study Transformers implementation through HLS4ML