GRAPHORMER-IR: GRAPH TRANSFORMERS CAN PREDICT EXPERIMENTAL IR AND IRMPD SPECTRA USING HIGHLY SPECIALIZED ATTENTION **Cailum M. K. Stienstra**[1], Liam Hebert[2], Teun van Wieringen[3], Patrick Thomas[1], Alexander Haack[1], Jason Guo[1], Jonathan Martens[3], W. Scott Hopkins[1,4,5] [1]Department of Chemistry, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada; [2]Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1; Canada [3]Radboud University, Institute for Molecules and Materials, FELIX Laboratory, 6525 ED, Nijmegen, The Netherlands; [4]Watermine Innovation, Waterloo, Ontario, N0B 2T0, Canada; [5]Centre for Eye and Vision Research, Hong Kong Science Park, New Territories, 999077, Hong Kong. (cmkstien@uwaterloo.ca)

Infrared (IR) spectroscopy is an important analytical tool in various chemical domains and a where in silico methods for predicting experimental spectra remain important. Here, we employ Graphormer, a graph neural network (GNN) transformer, to predict IR spectra using molecular structure using 53,528 high-quality spectra measured in five different experimental media. Using learned, augmented node embeddings generated from a novel feature encoder yields test scores to $SIS\mu$ = 0.8523 ± 0.0006, a total improvement of 19.7σ (t = 19) over the current state-of-the-art model Chemprop-IR ($SIS\mu$ =0.8409 ±0.0014, n = 5). These improved scores show how Graphormer-IR excels in capturing long-range interactions like hydrogen bonding, anharmonic peak positions in experimental spectra, and uncommon functional groups. Scaling our architecture to 210 attention heads demonstrates specialist-like behavior for distinct IR frequencies that improves performance. We further show that using transfer learning approaches and a small metabolite Infrared Multi-Photon Dissociation (IRMPD) spectral library from the FELIX Free Electron Laser (FEL) institute in the Netherlands, we perform the first ever accurate machine learning predictions of IRMPD spectra. This model will be used as a diagnostic pre-screening tool at WaterFEL, a future (open via application) user's facility at the University of Waterloo.