

Deep Learning for Scientific

A. J. Martinez

A dissertation submitted in partial fulfilment of the requirements
for the degree of *Doctor of Philosophy*

Department of Computer Science
University of Cambridge

2025

Contents

1	Introduction	1
2	Background	3

Introduction

The automatic understanding of scientific documents presents unique challenges at the intersection of natural language processing and mathematical reasoning. Modern transformer-based architectures have demonstrated remarkable ability to process structured text, yet the rich symbolic language of mathematics — with its equations, theorems, and logical dependencies — remains an open frontier.

This dissertation makes three primary contributions:

1. A novel multimodal encoder for joint text and formula understanding.
2. A large-scale benchmark of 100 000 annotated scientific papers.
3. State-of-the-art results on three downstream reasoning tasks.

Background

Definition 2.1: Transformer Architecture

A *transformer* is a sequence-to-sequence model built entirely from self-attention and feed-forward layers, without recurrence. Given input embeddings $\mathbf{X} \in \mathbb{R}^{n \times d}$, the multi-head attention output is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V.$$

Theorem 2.2: Universal Approximation for Transformers

Transformers with sufficient depth and width can approximate any continuous sequence-to-sequence function on compact domains to arbitrary precision.