Abstract

Modern engineered materials and technological systems exhibit strong couplings across multiple spatial and temporal scales, ranging from atomistic structure to device-level behaviour and system-level functionality. Traditional multiscale modelling approaches, while physically rigorous, increasingly struggle to address the combined challenges posed by high dimensionality, computational cost, heterogeneous data representations and expansive design spaces. In this context, artificial intelligence provides powerful tools for surrogate modelling, representation learning and generative exploration. However, its effective application requires principled integration with physical knowledge, structured data infrastructures and reproducible computational workflows.

This doctoral thesis develops and validates an AI-driven, data-centred multiscale modelling framework aimed at linking physical structure to functional prediction across heterogeneous domains. Rather than enforcing direct hierarchical coupling between physical models operating at different scales, the proposed approach treats data as the primary coupling layer. This enables information transfer through shared representations, automated workflows and machine-learning models grounded in physical constraints.

The methodological framework is articulated through two complementary case studies. At the atomistic scale, an automated and FAIR-compliant workflow is developed for the generation of a large dataset of defective graphene nanostructures, including electronic, energetic and transport properties, as well as virtual scanning tunnelling microscopy images. Gradient-boosted models and explainable AI techniques are employed to quantify structure-property correlations, assess model interpretability and identify the role of specific defect configurations. Generative models, including variational autoencoders and conditional diffusion processes, are further introduced to enable controlled exploration of the defect space aligned with target physical properties.

At the mesoscale, the thesis addresses device-level modelling of bulk heterojunction organic photovoltaic cells. Automated pipelines are constructed to generate three-dimensional morphologies and high-fidelity drift-diffusion simulations. Neural operator architectures are employed as surrogate models to learn the underlying physical operators governing charge transport and current-voltage characteristics, significantly reducing computational cost while preserving predictive accuracy. Morphology clustering and generative strategies are used to explore the relationship between mesoscale structure and device performance.

Beyond these specific applications, the thesis proposes a unifying conceptual framework in which artificial intelligence acts as an integrative layer across scales, supported by automation, structured data management and Key Enabling Technologies. The results demonstrate that data-driven models can capture complex physical mappings, support interpretability and enable generative design when embedded within disciplined computational workflows. The framework developed in this work is designed to be transferable across domains characterised by multiscale physical complexity, providing a methodological foundation for future AI-driven modelling in materials science, energy devices and emerging technological infrastructures.

Overall, this thesis addresses three central research questions: how heterogeneous multiscale physical systems can be coherently represented within a unified computational framework; to what extent data-driven models can approximate and abstract complex physical operators; and which methodological conditions are required to ensure reproducibility, interpretability and scalability in AI-driven multiscale modelling.

The results demonstrate that explicit hierarchical coupling between physical models is not a necessary condition for multiscale integration. Instead, a data-centred approach, supported by automation and structured data infrastructures, enables consistent information transfer across scales while preserving the integrity of domain-specific physical models. Learning-based surrogates and neural operators are shown to provide accurate and computationally efficient abstractions of high-fidelity simulations when embedded within disciplined workflows. At the same time, the work delineates clear epistemic limits of data-driven approaches, highlighting the dependence of interpretability and generalisation on data resolution, structure and physical consistency.

Within this framework, generative models emerge not as tools for unconstrained optimisation, but as mechanisms for guided exploration of high-dimensional design spaces, enabling the identification of families of structures consistent with target functional regimes. Taken together, these results position artificial intelligence as an integrative methodological layer for multiscale modelling, enabling physically informed exploration and reasoning across atomistic, device and system-level domains.