GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

# Schlussbericht

Verbund: 05M2020 - MaGriDo

| | |
|---|---|
| Zuwendungsempfänger: | Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung eingetragener Verein |
| Projektleitung: | Dr. rer. nat. Jan Hamaekers |
| E-Mail: | jan.hamaekers@scai.fraunhofer.de |
| Förderkennzeichen: | 05M20AAA |
| Förderzeitraum: | 01.04.2020 - 31.03.2023 |
| Zuwendung: | 358.944,20 € |
| Projektträger: | Projektträger DESY |
| Zusätzlicher Kontakt: | jan.hamaekers@scai.fraunhofer.de |
| Zusätzlicher Name: | Jan Hamaekers |

| Genutzte Großgeräte: | **Labor** | **Gerät** | **Experiment** |
|---|---|---|---|

| | |
|---|---|
| Diplomarbeiten: | 0 |
| Dissertationen: | 0 |
| Habilitationen: | 0 |
| Referierte Publikationen: | 3 |
| Andere Veröffentlichungen: | 0 |
| Patente: | 0 |
| Bachelorarbeiten: | 0 |
| Masterarbeiten: | 2 |
| Staatsexamen: | 0 |

Dieser Bericht wurde beim Projektträger über einen individuellen Online-Zugang vom Projektleiter eingereicht und am 29.09.2023 16:33 für eine Veröffentlichung freigegeben.

# Schlussbericht

| | |
|---|---|
| Zuwendungsempfänger: | Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. Postfach 20 07 33 80007 München<br>Ausführende Stelle: Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI) |
| Projektleitung: | Dr., Jan Hamaekers |
| Verbund: | MaGriDo: Mathematik für maschinelle Lernmethothoden für Graph-basierte Daten mit integriertem Domänenwissen. |
| Thema: | Teilprojekt 4: Entwicklung und Umsetzung von tiefen Graphnetzwerken für Anwendungen in der Materialentwicklung |

# Zusammenfassung

Methoden des Deep Learning sind in den letzten Jahren erfolgreich für verschiedene Problemstellungen angewendet worden (Bilderkennung etc.). Hier wurden bisher meist sogenannte „end-to-end“ Lernansätze verwendet. Zu deren Umsetzung sind in der Regel sehr große Mengen von strukturierten Daten notwendig sind, welches dazu führt, dass diese in vielen möglichen Anwendungsfällen aus den Naturwissenschaften, Medizin und Industrie nur bedingt einsetzbar sind.
Ziel des Verbundvorhabens MaGriDo war es daher, neue Ansätze zu entwickeln, zu analysieren und auf Problemstellungen anzuwenden, die es erlauben existierendes Wissen in die Architektur der Netzwerke einzubauen und somit ermöglichen von den komplementären jeweiligen Stärken von „end-to-end“ Lernansätze und „a priori Modellen/Regeln“ zu profitieren. Solch ein Vorgehen verspricht substantiell effizientere Lösungen für viele der genannten Anwendungsfelder zu ermöglichen.
Da üblicherweise komplexe Systeme sehr gut als Zusammensetzungen von Entitäten
und deren Wechselwirkungen repräsentiert werden können, lag der Schwerpunkt der Forschung und Entwicklung in MaGriDo auf sogenannten Graphnetzwerken. Diese enthalten zum Beispiel konventionelle Fully-Connected-NN, Convolution-NN und Recurrent-NN als Spezialfall, können insbesondere auf relationalen Strukturen angewendet werden und ermöglichen eine hierarchische Prozessierung der Eingabedaten.
In dem Teilprojekt 4 lag der Schwerpunkt auf der Entwicklung und Umsetzung von Graphnetzwerken für Anwendungen in der Materialentwicklung, wobei der Fokus auf entsprechenden Fragestellungen bezüglich Molekülen, Polymeren und Gläsern lag. Dazu wurden insbesondere drei typische Problemstellungen der rechnergestützten Chemie und der Materialwissenschaften betrachtet:

- Problemstellungen der ersten Generation: Prädiktive Modelle zur Vorhersage von Materialeigenschaften basierend auf der atomistischen Struktur.
- Problemstellungen der zweiten Genration: Prädiktive Modelle zur Vorhersage von Materialeigenschaften basierend auf deren chemischen Zusammensetzung.
- Problemstellungen der dritten Generation: Generative Modelle zur Vorhersage von Strukturen oder Kompositionen von Materialien mit gewünschten Eigenschaften.

Zu diesen Problemstellungen wurden jeweils geeignete Modelle mit Hilfe von praxisrelevanten Problemstellungen entwickelt und untersucht.



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

# Bericht

## 1 Aufgabenstellung und Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

*Aufgabenstellung Verbundvorhaben:*

Ziel des Verbundprojektes war es, tiefe neuronale Netzwerke (TNNs) für Problemstellungen aus der Industrie (weiter) zu entwickeln und zu analysieren, die es erlauben, existierendes Domänenwissen in die Architektur der Netzwerke einzubauen, wobei der Schwerpunkt der Forschung und Entwicklung in auf sogenannte Graphnetzwerke (GNNs) gelegt wurde.

Die mathematischen Untersuchungen im Vorhaben sollen letztendlich zum besseren Verständnis der Architektur von tiefen neuronalen Netzen beisteuern. Die mathematischen Beiträge des Vorhabens umfassen

- die Untersuchung und Erarbeitung einer Expressivitätsanalyse für Graphnetzwerken,
- die Erweiterung von Interpretierbarkeitsalgorithmen und deren (neue) Theorie auf Graphnetzwerke,
- die Untersuchung effizienter Lernverfahren für das Training von Graphnetzwerken,
- die Betrachtung von Graphnetzwerkarchitektur, Regularisierung und Optimierungsverfahren,
- die strukturierte Integration von Domänenwissen durch geeignete mathematisch Formalismen,
- die Erarbeitung von mathematischen Konzepten zum Transferlernen und aktiven Lernen,
- die Erarbeitung und Untersuchung von systematisch verbesserbaren Wechselwirkungspotentialen,
- die Untersuchung generativer Netzwerke.

*Aufgabenstellung Teilprojekt 4:*

Die wesentlichen Aufgaben des Teilprojekt 4 waren:

- Entwicklung von Graphnetzwerken und Workflows zur Vorhersage von Materialeigenschaften basierend auf deren atomaren Struktur.
- Entwicklung von Graphnetzwerken und Workflows zur Vorhersage von Materialeigenschaften basierend auf deren Komposition.
- Entwicklung von Graphnetzwerken und Workflows zur Vorhersage von Materialzusammensetzungen hinsichtlich vorgegebener Materialeigenschaften.

Anwendung der Verfahren für Problemstellungen aus dem Bereich der Materialentwicklung für Moleküle/Polymere und Gläser.

*Voraussetzungen:*

Das Verbundkonsortium besteht aus den Gruppen:

- (Verbundkoordinator): Arbeitsgruppe Garcke (Universität Bonn),
- Arbeitsgruppe Kutyniok (LMU München),
- Arbeitsgruppe Lorenz (TU Braunschweig),
- Arbeitsgruppe Hamaekers (Fraunhofer SCAI).

In dem Vorhaben arbeiteten universitäre Arbeitsgruppen mit Expertise in den Disziplinen numerische Mathematik, Analysis, Optimierung, und Angewandte Funktionalanalysis mit dem Fraunhofer SCAI zusammen. Die universitären Arbeitsgruppen haben komplementäre Schwerpunkte und adressierten im Vorhaben unterschiedliche Aspekte der betrachteten Lernverfahren, wobei das verbindende Element das Design der Architektur war, sei es aus Sicht der Integration des Domänenwissens und des Transferlernens (Bonn), der Expressivität und Interpretierbarkeit (München), des Trainierens (Braun-



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

schweig) und der konkreten Anforderungen aus der Anwendung (SCAI mit den assoziierten Projektpartnern). Das Fraunhofer SCAI hatte hier schon besondere Expertise hinsichtlich Softwareengineering-Aspekten, ingenieurmäßiger Anbindung und der späteren Verwertung der Projektergebnisse. Zusätzlich sind die beteiligten assoziierten Partner Covestro und Schott, jeweils Experten für konkrete Fragestellungen in den jeweiligen Anwendungsbereichen Moleküle/Polymere und Gläser.

## 2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Der Stand der Forschung zeigt, dass das maschinelle Lernen, insbesondere das tiefe Lernen, große Erfolge bei der Mustererkennung in verschiedenen Bereichen erzielt hat, wie z.B. Computer Vision, Spracherkennung, Textverständnis und Spiele-KIs. Diese Erfolge basieren auf dem datenbasierten Ansatz des Lernens aus einer Vielzahl von Beispielen. Allerdings stoßen rein datengetriebene Ansätze in einigen Fällen an ihre Grenzen oder führen zu unbefriedigenden Ergebnissen. Zum Beispiel, wenn nicht genügend Daten vorhanden sind, um komplexe und genaue Modelle zu trainieren, oder wenn das Ziel darin besteht, die Vorhersagequalität zu verbessern oder die Trainings- und Inferenzzeit zu reduzieren. Darüber hinaus besteht ein Bedarf an interpretierbaren Modellen und an Transfer-Lernen-Ansätze, um die Abstraktionsfähigkeiten von maschinellen Lernmodellen zu erhöhen. Diese Aspekte werden weitgehend in den Arbeitspaketen 2, 3 und 4 von den jeweiligen universitären Arbeitsgruppen bearbeitet. In Teilprojekt 4 wurden im Wesentlichen die Arbeitspakete 5 und 6 des Verbundvorhabens bearbeitet, daher werden im Folgenden für die Arbeitspakete 5 und 6 die wissenschaftlichen und technischen Startbedingungen zum Projektstart dargelegt:

In der rechnergestützten Chemie und den Materialwissenschaften gibt es drei typische Ansätze. Der Standardansatz der ersten Generation ist die Berechnung der physikalischen Eigenschaften einer gegeben atomaren Eingangsstruktur, die oft über eine Annäherung an die Schrödinger-Gleichung (meist mittels Dichtefunktionaltheorie(DFT)-Methoden) erfolgt. Hier werden maschinelle Lernmethoden eingesetzt, um rechenintensive quantenmechanische Berechnungen mittels trainiertem Ersatzmodel zu umgehen. Zum Beispiel konnte SCAI bereits erfolgreich mit ML-Verfahren teure quantenmechanische Simulationen für Moleküle beliebiger Größe ersetzen [1]. Außerdem hat SCAI tiefgehende Erfahrung im Bereich der Entwicklung und Implementierung von Kraftfeldern(Ersatzmodellen) für die Molekulardynamik und vertreibt ein entsprechendes Softwaremodul Tremolo-X/ATK-ForceField [2]. Im Ansatz der zweiten Generation, dienen als Eingangsdaten nur die chemischen Zusammensetzungen, zu denen eine Vorhersage über die Struktur oder das Ensemble von Strukturen und deren Eigenschaften berechnet werden soll. Hier können auch Methoden des maschinellen Lernens verwendet werden um Eigenschaften, die gegebenenfalls nur mit geringer Genauigkeit und rechenintensiv mit Hilfe von Modellierung und Simulation berechnet werden können, mit experimentellen Daten zu kombinieren, um Genauigkeit und Effizienz substantiell zu erhöhen [3]. Der neue Ansatz der dritten Generation ist es, maschinelle Lerntechniken zur Vorhersage von Zusammensetzung, Struktur und Eigenschaften anzuwenden. Hier hatte SCAI zum Beispiel erste Arbeiten Molekül-Generatoren [4]. Insgesamt unterscheiden sich die Voraussetzungen und geeigneten Methoden insbesondere hinsichtlich der Datenlage.

[1] Barker, J., Bulin, J., Hamaekers, J., & Mathias, S. LC-GAP: Localized Coulomb descriptors for the Gaussian approximation potential. In Scient. Comp. & Algor. in Industrial Simulations, 25-42, Springer, 2017.[BDC+18] Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O. & Walsh, A.. Machine learning for molecular and materials science. Nature, 559(7715), 547–555, 2018.

[2] Schneider, J., Hamaekers, J., Chill, S. T., Smidstrup, S., Bulin, J., Thesen, R., & Stokbro, K. ATK-ForceField: a new generation molecular dynamics software package. Modelling and Simulation in Materials Science and Engineering , 25 (8), 085007, 2017.

[3] Garcke, J., Hamaekers, J. & Iza-Teran, R. (2018). Maschinelles Lernen für die virtuelle Produktentwicklung. In: Neugebauer R. (eds) Digitalisierung. Springer Vieweg, Berlin, Heidelberg, 255-260.

[4] Israels, R., Maaß, A., & Hamaekers, J. (2017). The octet rule in chemical space: generating virtual molecules. Molecular diversity , 21 (4), 769-778.



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

## 3 Planung und Ablauf des Vorhabens sowie Kooperation mit Dritten

Die Planung und Durchführung des Projektes gliederte sich in folgende Arbeitspakete, wobei der jeweilige Teilprojektkoordinator in Klammern vermerkt ist:

- *Arbeitspaket AP0 (UB): Koordination:*
  Das Projektmanagement hat die Aufgabe die verfügbaren Ressourcen in kontrollierter und strukturierter Weise zu verwalten. Nicht nur der reibungslose Ablauf des Projekts sollte sichergestellt werden, auch die dafür notwendigen Grundlagen sollten geschaffen werden, wie die Kommunikation innerhalb des Konsortiums und mit dem Projekträger.
- *Arbeitspaket AP1 (alle, je 2 PM): Demonstratoranwendung:*
  Spezifikation einer gemeinsamen Demostratoranwendung und notwendiger Datenformate, sowie Realisierung von Schnittstellen, gemeinsamer Datenrepositorien und Coderepositorien.
- *Arbeitspaket AP2 (LMU): Expressivität und Interpretierbarkeit*
- *Arbeitspaket AP3 (TUBS): Training und Architektur*
- *Arbeitspaket AP4 (UB): Integration von Domänenwissen, Erklärbarkeit und Transferlernen*
- *Arbeitspaket AP5 (SCAI): Anwendung in der Materialentwicklung*
  In AP5 sollen Graphnetzwerk-basierte Methoden mit spezieller Anwendung in der Materialentwicklung entwickelt und implementiert werden. Dazu sollen generelle Workflows und Frameworks für die Fragestellungen der erwähnten drei Generationen und entsprechende Prototypen entwickelt werden. Dazu wurde AP5 in wie folgt untergliedert:
  - *AP 5.1: Graphnetzwerke und Workflows für Probleme der 1. Generation*
  - *AP 5.2: Graphnetzwerke und Workflows für Probleme der 2. Generation*
  - *AP 5.3: Graphnetzwerke und Workflows für Probleme der 3. Generation*
  - *AP 5.4: Generierung von Daten*
- *Arbeitspaket AP6 (SCAI): Praxisrelevante Anwendung*
  In AP6 sollen die Problemstellung der Praxispartner detailliert untersucht werden und erste Prototypen mit Hilfe der Ergebnisse von AP1-AP4 und insbesondere AP5 zu deren Lösung entwickelt und validiert werden. Hier sollen insbesondere die assoziierten Partner Covestro und Schott einbezogen werden.
  - AP 6.1: Anwendungsfall Polymere (+ assoziierter Partner Covestro)
  - AP 6.2: Anwendungsfall Gläser (+ assoziierter Partner Schott)

## 4 Verwendung der Zuwendung (wichtigste Positionen des zahlenmäßigen Nachweises, z. B. Investitionen, Personalmittel)

Im Teilprojekt 4 des Verbundvorhabens MaGriDo waren die wesentlichen Kosten die Personalkosten. Der zahlenmäßige Nachweis über die Verwendung der Zuwendungen mit Angabe der wichtigsten Positionen liegt diesem Schreiben gesondert bei.

## 5 Erzielte Ergebnisse mit Gegenüberstellung der vereinbarten Ziele

Die folgende Darstellung umfasst im Detail nur die Aufgaben und Ergebnisse des Teilprojektes 4 des Verbundprojektes MaGriDo. Hinsichtlich der Ziele und Ergebnisse der Arbeitspakete:

- Arbeitspaket AP2 (TUB): Expressivität und Interpretierbarkeit,
- Arbeitspaket AP3 (TUBS): Training und Architektur,
- Arbeitspaket AP4 (UB): Integration von Domänenwissen, Erklärbarkeit und Transferlernen,

wird auf die Berichte der anderen Teilprojekte verwiesen.

### 5.1 Arbeitspaket AP0: Koordination

Das Projektmanagement hatte die Aufgabe die verfügbaren Ressourcen in kontrollierter und strukturierter Weise zu verwalten. Nicht nur der reibungslose Ablauf des Projekts sollte sichergestellt werden,



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

sondern es sollten auch die dafür notwendigen Grundlagen geschaffen werden, wie die Kommunikation innerhalb des Konsortiums und mit dem Projektträger.

### 5.1.1 Ergebnisse AP0 (SCAI)

Erstellung von E-Mail Listen (über SCAI) zur vereinfachten Kommunikation und aufsetzen von geteilten über das Internet erreichbaren Verzeichnissen (Owncloud Server bei SCAI) und eines gitlab Repository zum vereinfachten Austausch von Daten und Software (gitlab Server bei SCAI), sowie Erstellung einer Webseite [5].

## 5.2 Arbeitspaket AP1: Demonstratoranwendung

Um eine gemeinsame Sicht auf die Aufgabenstellungen und eine Synergie der verschiedenen mathematischen Forschungsarbeiten zu ermöglichen, sollte zum Beginn des Projekts eine geeignete Demonstratoranwendung identifiziert werden. Diese sollte zum einen nah an den Anforderungen der Anwender sein, aber andererseits die (auch gemeinsamen) mathematischen Untersuchungen ermöglichen.

### 5.2.1 Ergebnisse AP1 (SCAI)

Hier wurde die Vorhersage von Moleküleigenschaften auf dem Benchmarkdatensatz QM9 [6] als Demonstratoranwendung identifiziert und auch die notwendigen Datenformate, Schnittstellen, gemeinsame Datenrepositorien und Coderepositorien realisiert

## 5.3 Arbeitspaket AP5: Anwendung in der Materialentwicklung

In AP5 sollten Graphnetzwerk-basierte Methoden mit spezieller Anwendung in der Materialentwicklung entwickelt und implementiert werden. Dazu sollten generelle Workflows und Frameworks für die Fragestellungen der erwähnten drei Generationen und entsprechende Prototypen entwickelt werden.

### 5.3.1 Ergebnisse AP5 (SCAI)

*AP 5.1 - Prototypen von Graphnetzwerken und Workflows für Problemstellungen der 1. Generation*

- Hinsichtlich der Demonstratoranwednung wurden hier Graph-Convolution-Netzwerke (GCNs) zur Vorhersage von Eigenschaften von Molekülen basierend auf deren 3D Struktur entwickelt. Außerdem wurden verschiedene GCN basierend auf SMILES (2D Topologie des Molekülbindungsgraphen) entwickelt und untersucht. Hier hat sich gezeigt, dass GCN für diese Problemstellung flexibel anwendbar sind. Zum Beispiel konnte für verschiedene Eigenschaften die Genauigkeit im Bereich der gewünschten chemischen Genauigkeit erreicht werden, vgl. Tabelle 1. Zu beachten ist hier auch die Wahl von geeigneten Eingangsfeatures. Weiter Details hinsichtlich GCNs für Moleküle wurden in [BBH+21] veröffentlicht.

[5] https://www.scai.fraunhofer.de/de/projekte/MaGriDo.html

[6] Rupp, Matthias, et al. "Fast and accurate modeling of molecular atomization energies with machine learning." Physical review letters 108.5 (2012): 058301.

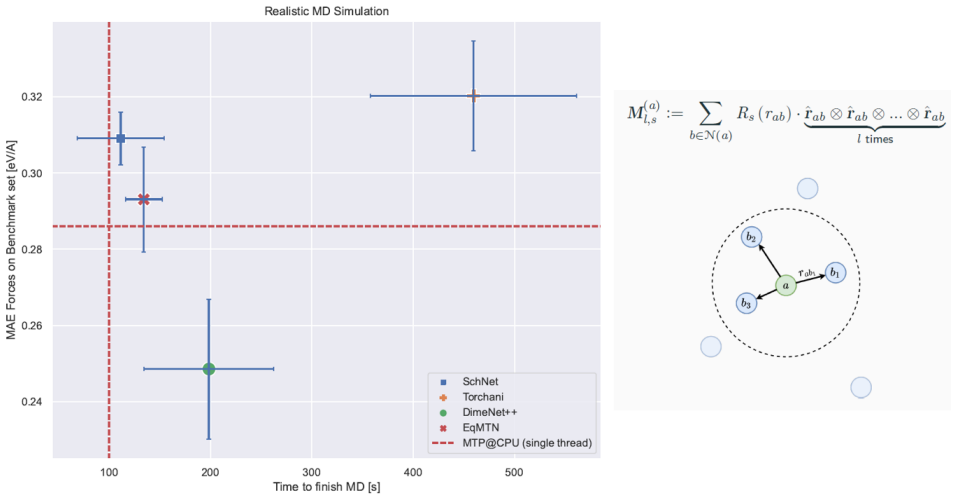


Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

| Property | Number of data points | Training MAE | Prediction MAE | Chemical Accuracy |
|---|---|---|---|---|
| H@298K | ~134 k | ~0.02 eV | ~0.05 eV | ~0.04 eV |
| $\Delta$ H | ~45 k | ~0.02 eV | ~0.06 eV | ~0.04 eV |
| $\mu$ | ~134 k | ~0.37 Db | ~0.51 Db | 0.2-0.5 Db |
| HOMO | ~134 k | 0.06 eV | 0.11 eV | 0.03 eV |
| $T_{crit}$ | ~0.9 k | - | 11 K | 2-5 K |

**Tabelle 1: Vorhersagegenauigkeit für Moleküleigenschaften von einem GCN mit zwei Convolution-Layern welches Atom- und Bondfeatures nutzt.**

- Es wurden GCN zur Approximation der Born-Oppenheimer Energiefläche und somit zur Darstellung von Wechselwirkunspotentialen (WP) für atomistische Systeme entwickelt und untersucht. Dabei wurden insbesondere verschiedene Architekturen analysiert, wobei sich ResNet-artige Architekturen als vorteilhaft erwiesen haben, vgl. Masterarbeit [Waldrof21]. Konventionelle GCN mit invarianten Layern bezüglich der Atomumgebung, haben den Nachteil, dass tensoriale Groessen (z.B. Dipolmoment) nicht gut vorhergesagt werden können, da die Information über die Beziehung zwischen verschiedenen Atomumgebungen verloren geht. Daher wurden hier neuartigere euivariante GCNs (EGCN) entwickelt und untersucht. Hier hat sich gezeigt, dass diese ECGN WPs flexibel für Problemstellungen für atomistische Systeme, zum Beispiel für Molekulardynamische Simulationen, genutzt werden können. Der entwickelte equivariante Ansatz beruht auf Moment-Tensoren (MT) und der Clebsch-Gordan Transformation. Im Vergleich zu anderen existierenden Methoden weist unser Ansatz ein kompetitives Kosten-Nutzen-Verhältnis für praxisrelevante Molekulardynamiksimulationen auf, vgl. Figur 1. Weitere Details zu den Grundlagen des neu entwickelten EGCN finden sich auch in der von SCAI betreuten Masterarbeit [Oerder]. Dieses Modell wurde auch bzgl. des Einflusses von Trainingsmengenauswahlverfahren untersucht, wobei sich gezeigt hat, dass geeignete Auswahlverfahren insbesondere den maximalen Fehler für kleinere Trainingsmengen reduzieren kann, vgl. [BCG+23].



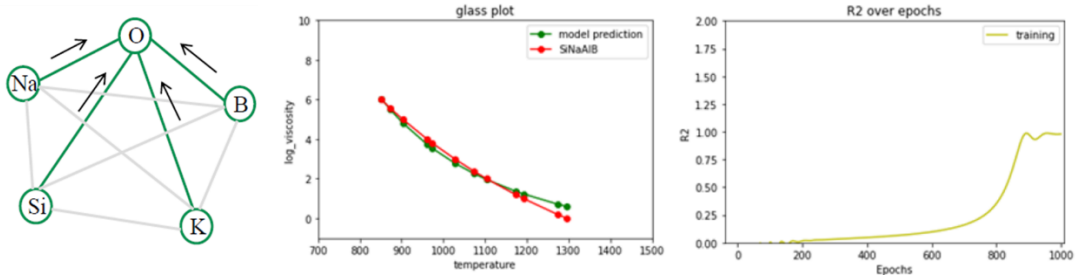

**Figur 1: Vergleich in Kosten und Genauigkeit verschiedener GCNs für eine praxisrelevante Molekulardynamische Simulation und Darstellung von Momenttensoren für Atomumgebungen. EqMTN ist das in MaGriDo entwickelte EGCN und MTP ein von SCAI entwickeltes lineares MT basiertes WP.**

AP 5.2 - Prototypen von Graphnetzwerken und Workflows für Problemstellungen der 2. Generation



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

- Hinsichtlich der Vorhersage von Materialeigenschaften basierend auf deren Komposition wurden verschiedenen Graphnetzwerk Architekturen entwickelt und analysiert. Außerdem, wurde die Integration von Domaenenwissen mittels physikalischer Features untersucht. Zusätzlich wurden Modelle zur Vorhersage von mehreren Eigenschaften basierend auf erweiterte Architekturen entwickelt und untersucht. Hier hat sich gezeigt, dass falls Eigenschaften korreliert sind, Ansätze des sogenannten Multi-Task-Lernen die Vorhersagegenauigkeit gegenüber Modellen jeweils nur für eine Eigenschaft verbessern können, vgl. [MHZ23]. Außerdem hat sich gezeigt, dass Ansätze des Informierten-Lernens die Exptrapolationsfähigkeit eines Modells verbessern können, vgl. [MHM+23]. Dies hat sich insbesondere auch mittels der Integration bekannter physikalischer Zusammenhänge für die Vorhersage von Messkurven erwiesen, wie zum Beispiel Temperatur-Viskositätskurven, vgl. Figur 2.

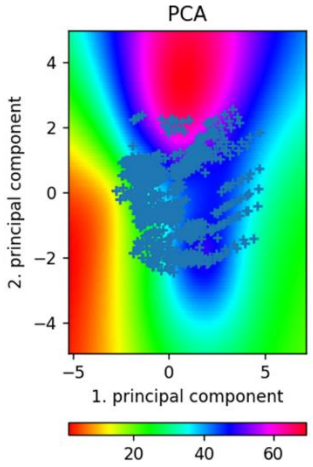



**Figur 2: Schematische Darstellung eines Graphnetzwerkes für Materialkompositionen und Ergebnisse für Temperatur-Viskositätskurven.**

Als vorteilhaft haben sich hier Graphnetzwerke mit einem Janossy Pooling mit k=1 und k=2 erwiesen. Zum Beispiel sind die bekannten Self-Attention Netzwerk ein Spezialfall von diesen[7].

AP 5.3 - Prototypen von Graphnetzwerken und Workflows für Problemstellungen der 3. Generation

- Hier wurde ein Ansatz des Inversen Designs mittels Bayesscher Optimierung und der Modellansätze aus AP5.2 für Rezepturen entwickelt und untersucht. Dieser wurde zum Beispiel für einen Datensatz für Beton umgesetzt, vgl. Figur 3.




| | Acq | CS [MPa] | stdev [MPa] | Cement | Blast Furnace Slag | Fly Ash | Water | Superplasticizer | Coarse Aggregate | Fine Aggregate | Age [days] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 90.670161 | 90.670161 | 11.759876 | 0.1574 | 0.179835 | 0.012581 | 0.051436 | 0.004258 | 0.328417 | 0.266074 | 137.030197 |
| 1 | 90.670160 | 90.670160 | 11.757422 | 0.1574 | 0.179785 | 0.012553 | 0.051437 | 0.004258 | 0.328479 | 0.266089 | 137.006296 |
| 2 | 90.670161 | 90.670161 | 11.759194 | 0.1574 | 0.179836 | 0.012564 | 0.051434 | 0.004257 | 0.328440 | 0.266068 | 137.023982 |
| 3 | 90.670161 | 90.670161 | 11.759283 | 0.1574 | 0.179817 | 0.012577 | 0.051435 | 0.004258 | 0.328437 | 0.266076 | 137.020264 |
| 4 | 90.670161 | 90.670161 | 11.759939 | 0.1574 | 0.179844 | 0.012574 | 0.051434 | 0.004258 | 0.328421 | 0.266068 | 137.036528 |

**Figur 3: Inverse-Design Beispiel zur Druckfestigkeit von Beton. Die Druckfestigkeit im Rezeptraum ist hier farbkodiert und die Punkte entsprechen bekannten Rezepturen. In der Tabelle sind die berechneten vielversprechende Rezepturen angegeben.**

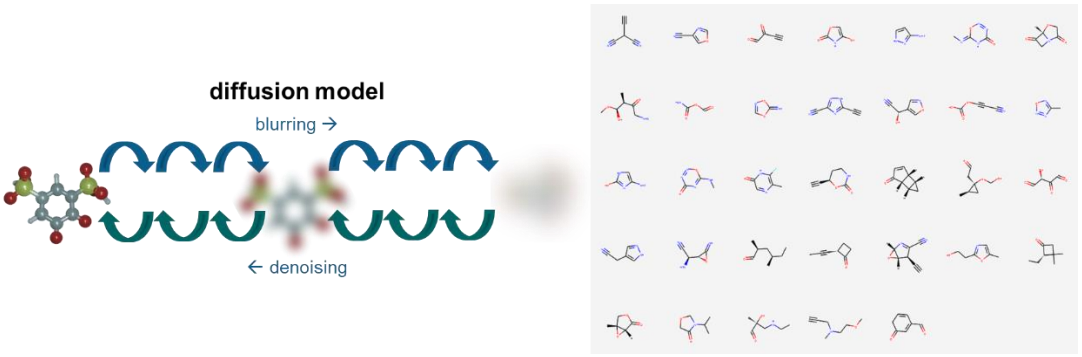
- Zur Generierung von Molekülen wurden zunächst auf equivarianten Graphnetzwerken basierte Diffusionsmodelle entwickelt und untersucht. Diese haben sich zur Generierung von 3D

[7] https://fabianfuchsml.github.io/learningonsets/



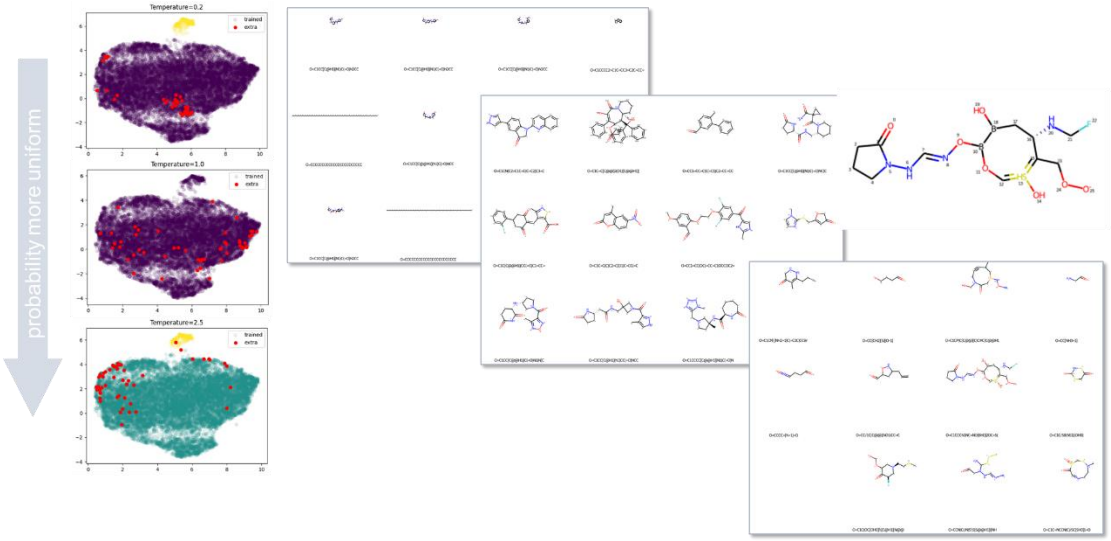
Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

Strukturen ohne Nebenbedingungen als flexibel anwendbar erwiesen. Jedoch konnte die Generierung unter Nebenbedingungen mit den bisherigen untersuchten Ansätzen nicht ausreichend zufriedenstellende Ergebnisse liefern.




**Figur 4: Equivariantes Diffusionsmodell zur Generierung von Molekülen.**

- Zusätzlich wurden Transformerarchitektur basierte (ChatGPT-artige) generative Modelle für SMILES von Molekülen entwickelt und untersucht. Diese haben sich als sehr vielversprechend auch zur Generierung unter Nebenbedingung erwiesen, vgl. [DMH23].




**Figur 5: Beispiele von Molekülen die mittels Transformermodell mit Graph-Self-Attention-Layer generiert wurden, wobei der sogenannte Temperaturparameter den Grad der *Halluzinierung* bestimmt.**

*AP 5.4 – Generation geeigneter Daten*

- Hinsichtlich der lokalen Strukturdaten von Gläsern wurden während der Projektlaufzeit geeignete Daten veröffentlich [8] [9]. Außerdem wurde auch eine Lizenz des Interglad Datensatzes[10] erworben. Diese Datensätze konnten für die Aufgaben in AP5 und AP6.2 genutzt werden.
- Notwendige Datensätze zu Polymeren wurden einerseits mittels Literaturrecherche zusammengestellt und andererseits wurden Workflows zu Generierung von neuen Daten mittels Molekulardynamiksimulation erarbeitet. und zur Datengenerierung durchgeführt. Zwei entsprechende Publikationen sind dazu in Vorbereitung und sind geplant 2023 zur Veröffentlichung in wissenschaftlichen Zeitschriften einzureichen [BH23, BH23b].

[8] Lu, Xiaonan, et al. "Predicting boron coordination in multicomponent borate and borosilicate glasses using analytical models and machine learning." Journal of Non-Crystalline Solids 553 (2021): 120490.
[9] Yu, Zheng, et al. "Structural signatures for thermodynamic stability in vitreous silica: Insight from machine learning and molecular dynamics simulations." Physical Review Materials 5.1 (2021): 015602.
[10] https://www.newglass.jp/interglad_n/gaiyo/info_e.html



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

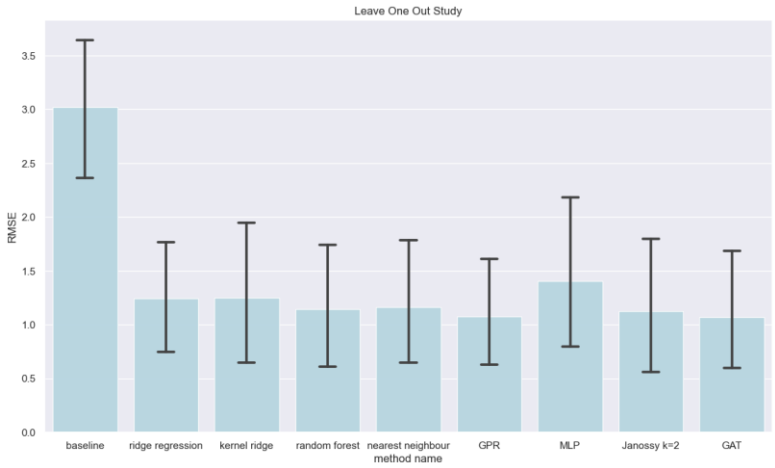
## 5.4 Arbeitspaket AP6: Praxisrelevante Anwendung

In AP6 sollten die Problemstellung der Praxispartner detailliert untersucht werden und erste Prototypen mit Hilfe der Ergebnisse von AP5 zu deren Lösung entwickelt werden.

### 5.4.1 Ergebnisse AP6 (SCAI)

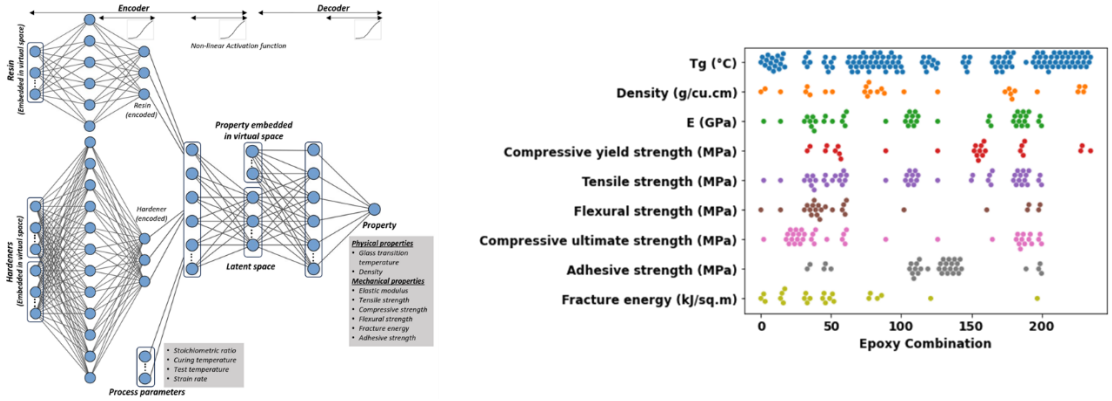
*AP 6.1: Anwendungsfall Polymere*

- Es wurden verschiedene Graphnetzwerkarchitekturen entwickelt und mit konventionellen Methoden des Maschinellen Lernens zur Vorhersage der Eigenschaften von Polyurethane verglichen. Zusätzlich wurden hier auch Ansätze des Informierten-Lernens umgesetzt. Mit Hilfe der im Projekt zur Verfügung stehenden geringen Datenlage konnte bisher jedoch noch nicht abschließend geklärt werden, ob die neuen GCN Ansätze hier einen substantiellen Vorteil aufweisen. Eine wissenschaftliche Veröffentlichung dazu ist in Planung [OHT24].




**Figur 6: Vergleich von konventionellen Verfahren des maschinellen Lernens mit einem Graph-Attention basiertem Modell (GAT) für die Vorhersage der Bruchdehnung.**

- Es wurden Encoder-Decoder basierte Architekturen zur Vorhersage von Eigenschaften von Epoxid-Polymeren entwickelt und untersucht. Hier wurden der in AP 5.4 zusammengestellte Datensatz genutzt. Zusätzlich wurden hier auch Ansätze des Informierten-Lernens umgesetzt, welche zu Verbesserungen der Vorhersage geführt haben, vgl. [BH23].

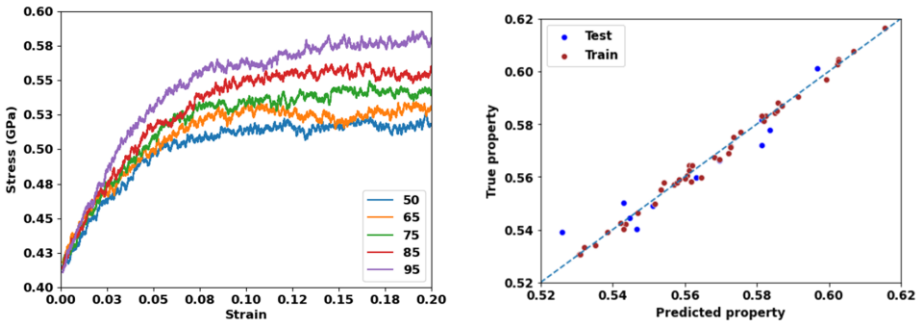



**Figur 7: Schematische Darstellung des Encoder-Decoder Netzwerkes für Epoxid-Polymere und eine Übersicht zu den zusammengestellten Datensatz, vgl. [BH23].**



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

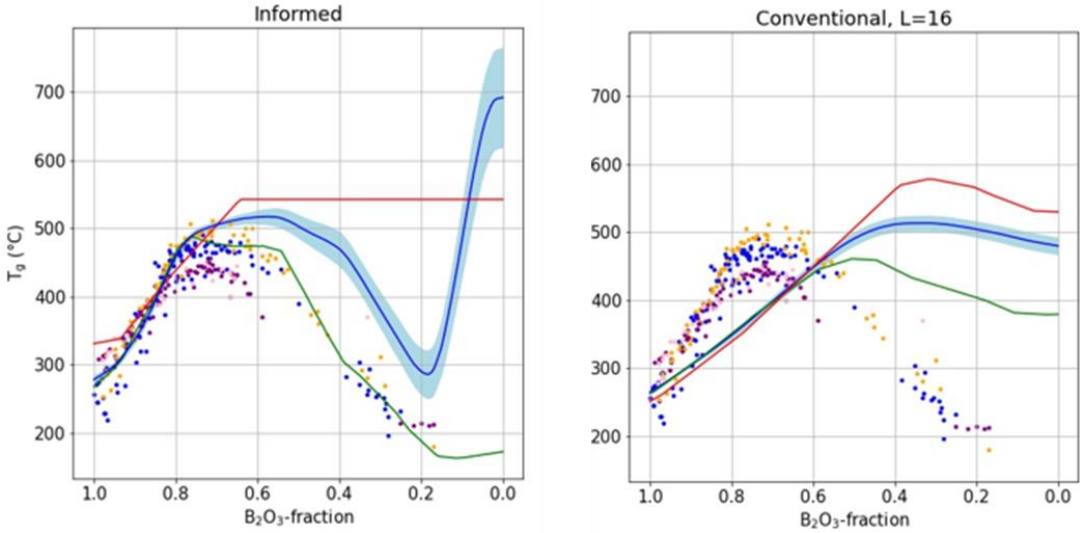
- Die in AP5.4 mittels Molekulardynamiksimulationen generierten Datensätze wurden genutzt, um ein Vorhersagemodell für die mechanischen Eigenschaften von Epoxid-Polymeren zu entwickeln, wobei auch ein Ansatz zum Informierten-Lernen genutzt wurde, vgl. [BH23b].




**Figur 8: Verschiedene simulierte Zug-Dehnungs-Kurven und Vergleich der Vorhersage der Zugfestigkeit., vgl. [BH23b].**

*AP 6.2: Anwendungsfall Gläser*

- Die zugänglichen Datensätze wurden bereinigt und aufbereitet. Auf der Grundlage von AP5 wurden dann damit Modelle zur Vorhersage der Eigenschaften von Gläsern entwickelt und untersucht. Hier wurde insbesondere ein Ansatz zum (Physik-)Informierten-Lernen umgesetzt und analysiert. Dieses Modell ermöglicht insbesondere auch Vorhersagen für Kompositionen deren beinhalteten Elemente nicht alle in der Trainingsmenge vorhandenen sind, vgl. Figur 9 und [MHM+23]. Außerdem wurden auch Physik-Informierte Ansätze zu Vorhersage von Temperatur Viskositätskurven umgesetzt.



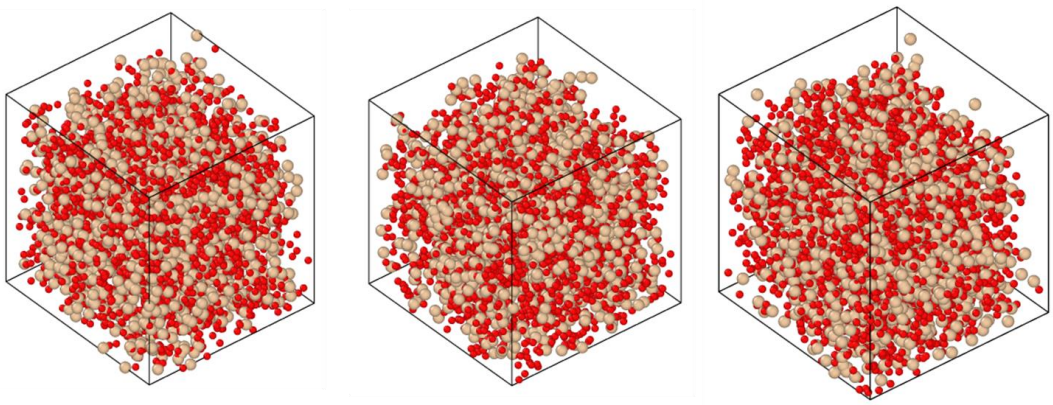

**Figur 9: Vergleich von der Vorhersage des Tg eines binären SiO2-B2O3 Glases in Abhängigkeit des B2O3 Anteiles von einem Physik-Informierten-Ensemble-Modell und einem konventionellem Modell, wenn in der Trainingsmenge kein Glass mit der Komponente B2O3 vorkommt, vgl. [MHM+23].**

- Zusätzlich wurden Ansätze des Transfer-Lernens zur Entwicklung von Graphnetzwerken zur Multi-Task Vorhersage entwickelt und untersucht. Hier hat sich gezeigt, dass dieser Ansatz in bestimmten Fällen die Genauigkeit der Vorhersage gegenüber der Vorhersage mit getrennten Modellen verbessern kann. Zum Beispiel konnte eine gewisse Verbesserung im Falle der Vorhersage der Tg und der Viskosität beobachtet werden, vgl. [MHZ23].
- Zur Generation von atomistischen Strukturen von Gläsern wurden verschiedene generative Verfahren analog zu den generativen Verfahren für die 3D Struktur von Moleküle umgesetzt.



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

Unter anderem wurde ein Diffusions-Transformer-Modell entwickelt. Der Ansatz nutzt den Diffusionsansatz, um Atome in 3D zu generieren. Im Training werden Koordinaten und Eigenschaften mit Gaussian Noise verrauscht. Ein Transformermodell lernt dann, diese Koordinaten zu entrauschen und den globalen Zusammenhang zwischen den Koordinaten zu erfassen. Hier hat sich zum Beispiel gezeigt, dass diese Verfahren potentiell eingesetzt werden können, um zum Beispiel geeignete Startkonfigurationen für weitere simulationsbasierte Verfahren zu erzeugen. Für deren direkter Verwendung muss die Genauigkeit der Verfahren noch verbessert werden.



**Figur 10: Mittels Diffusions-Transformer Ansatz generierte atomistische SIO2 Glasstrukturen.**

## 6 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Der Verlauf der Arbeit im Teilprojekt 4 folgte im Wesentlichen der im Projektantrag formulierten Planung. Die Aufgaben wurden erfolgreich bearbeitet, es waren keine zusätzlichen Ressourcen für das Projekt nötig. Die Notwendigkeit und Angemessenheit der geleisteten Arbeit sind in den Arbeitsergebnissen fachlich schon untersetzt dargestellt.

## 7 Voraussichtlicher Nutzen, insbesondere Verwertbarkeit der Ergebnisse

Die in Teilprojekt 4 erarbeiteten Workflows und entwickelten prädiktiven und generativen Modelle können im Rahmen von Forschungs- und Entwicklungsprojekten im Bereich der Materialwissenschaften genutzt werden, denn diese ermöglichen es die Anzahl er notwendigen Experimente gegenüber dem Versuch-und Irrtum Vorgehen substantiell zu reduzieren und somit auch insbesondere neue Industriekunden zu gewinnen. Insbesondere die neu entwickelten Methoden zur Generation von Molekülen unter Nebenbedingungen versprechen hier ein großes Verwertungspotential vor allem im Bereich der Energiematerialien.

Die neu entwickelten Kraftfelder, welche auf equivarianten Graphnetzwerk basieren, werden durch die Integration in das SCAI Softwaremodul TremoloX/ATK-ForceField mittels Softwarelizenzen verwertet.

Die neuen wissenschaftlichen Erkenntnisse wurden teilweise bereits mittels wissenschaftlicher Fachzeitschriften verbreitet [BBH+21, BCG+23, MHM+23] und es sind weitere Veröffentlichungen in der Vorbereitung [DMH23, MHZ23, OHT24]. Diese sollen auch genutzt werden, um neue Kunden anzusprechen.



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

## 8 Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Durchführung des Vorhabens wurden einige wissenschaftliche Arbeiten zu verschiedenen Kraftfeldern basierend auf equivarianten Graphnetzwerken veröffentlicht [11] [12] [13]. Die damit verbundenen Erkenntnisse sind auch bei der Entwicklung in Teilprojekt 4 berücksichtigt worden, aber keine der veröffentlichen Methoden entsprach der unseren, bzw. hat unsere Entwicklung beeinträchtigt.

## 9 Erfolgte und geplante Veröffentlichungen der Ergebnisse

### 9.1 Referierte Publikationen (z. B. in Fachzeitschriften oder -büchern und referierte Konferenzproceedings)

- [BBH+21] Barker, J., Berg, L. S., Hamaekers, J., & Maass, A. (2021). Rapid prescreening of organic compounds for redox flow batteries: a graph convolutional network for predicting reaction enthalpies from SMILES. Batteries & Supercaps, 4(9), 1482-1490.
- [BCG+23] Breustedt, N., Climaco, P., Garcke, J., Hamaekers, J., Kutyniok, G., Lorenz, D. A., ... & Shukla, C. V. (2023). On the Interplay of Subset Selection and Informed Graph Neural Networks. arXiv preprint arXiv:2306.10066. *(Zur Veröffentlichung akzeptiert)*
- [MHM+23] Maier, G., Hamaekers, J., Martilotti, D. S., & Ziebarth, B. (2023). Predicting Properties of Oxide Glasses Using Informed Neural Networks. arXiv preprint arXiv:2308.09492*. (Zur Veröffentlichung akzeptiert)*

*In Vorbereitung:*

- [DMH23] Niklas Dobberstein, Astrid Maaß, Jan Hamaekers (2023). *LLamol: A Dynamic Multi-Conditional Generative Transformer for De Novo Molecular Design*.
- [MHZ23] Gregor Maier, Jan Hamaekers, Benedikt Ziebarth (2023)- *Predicting Properties of Oxide Glasses Using Self-Attenion Mechanism*
- [BH23] Sindu BS and Jan Hamaekers (2023) - *Encoder-decoder model for simultaneous prediction of physical and mechanical properties of epoxy polymers*
- [BH23b] Sindu BS and Jan Hamaekers (2023) - *Feature-property correlation of cross-linked epoxy polymers through molecular dynamics and machine learning techniques*
- [OHT24] Rick Benedikt Oerder, Jan Hamaekers, Jim Thompson (2024) – *Predicting Properties of Polyurethane Using Informed Neural Networks*

### 9.2 Andere Veröffentlichungen (z. B. Konferenzbeiträge wie Vorträge und Poster, unreferierte Proceedings, Conference Notes)

### 9.3 Abschlussarbeiten (Bachelor, Master, Diplom, Staatsexamen, Promotion, Habilitation)

- [Walldorf21] Waldorf, Konstantin: *Graph-convolutional neural networks for chemical applications*, Masterarbeit, Institut für Numerische Simulation, Universität Bonn, 2021. (Haemaekers: Co-Betreuer – Garcke: Erstgutachter)

[11] Satorras, V. G., Hoogeboom, E., & Welling, M. (2021, July). E (n) equivariant graph neural networks. In International conference on machine learning (pp. 9323-9332). PMLR.
[12] Schütt, K., Unke, O., & Gastegger, M. (2021, July). Equivariant message passing for the prediction of tensorial properties and molecular spectra. In International Conference on Machine Learning (pp. 9377-9388). PMLR.
[13] Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., ... & Kozinsky, B. (2022). E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nature communications, 13(1), 2453.



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

- [Oerder22] Oerder, Rick: *Equivariant Machine Learning on Quantum Chemistry Data*,.Masterarbeit, Heinrich-Heine-Universität Düsseldorf, 2022. (Haemaekers: Co-Betreuer und Zweitgutachter)




Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

# Kurzbericht

- öffentlich -

| | |
|---|---|
| Zuwendungsempfänger: | Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. Postfach 20 07 33 80007 München<br>Ausführende Stelle: Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI) |
| Projektleitung: | Dr. Jan Hamaekers |
| Verbund: | MaGriDo: Mathematik für maschinelle Lernmethothoden für Graph-basierte Daten mit integriertem Domänenwissen. |
| Thema: | Teilprojekt 4: Entwicklung und Umsetzung von tiefen Graphnetzwerken für Anwendungen in der Materialentwicklung |

## 1. Ziel und Inhalt des Projektes

In den letzten Jahren wurden Deep Learning-Methoden erfolgreich für verschiedene Problemstellungen wie Bilderkennung eingesetzt. Bisher wurden dabei meist sogenannte "end-to-end" Lernansätze verwendet. Diese erfordern in der Regel große Mengen an strukturierten Daten, wodurch sie in vielen Anwendungsfällen aus den Naturwissenschaften, der Medizin und der Industrie nur begrenzt einsetzbar sind.

Das Ziel des Verbundvorhabens MaGriDo war es daher, neue Ansätze zu entwickeln, zu analysieren und auf Problemstellungen anzuwenden, die es ermöglichen, vorhandenes Wissen in die Netzwerkarchitektur einzubauen. Dadurch können die jeweiligen Stärken von "end-to-end" Lernansätzen und "a priori Modellen/Regeln" kombiniert werden. Dieses Vorgehen verspricht effizientere Lösungen für viele Anwendungsfelder zu ermöglichen.

Da komplexe Systeme oft gut als Zusammensetzungen von Entitäten und deren Wechselwirkungen repräsentiert werden können, lag der Schwerpunkt der Forschung und Entwicklung in MaGriDo auf sogenannten Graphnetzwerken. Diese können verschiedene Arten von neuronalen Netzen wie Fully-Connected-NN, Convolution-NN und Recurrent-NN als Spezialfall enthalten. Sie können insbesondere auf relationalen Strukturen angewendet werden und ermöglichen eine hierarchische Verarbeitung der Eingabedaten.

In dem Teilprojekt 4 lag der Schwerpunkt auf der Entwicklung und Umsetzung von Graphnetzwerken für Anwendungen in der Materialentwicklung, wobei der Fokus auf entsprechenden Fragestellungen bezüglich Molekülen, Polymeren und Gläsern lag. Dazu wurden insbesondere drei typische Problemstellungen der rechnergestützten Chemie und der Materialwissenschaften betrachtet:

- Problemstellungen der ersten Generation: Prädiktive Modelle zur Vorhersage von Materialeigenschaften basierend auf der atomistischen Struktur.
- Problemstellungen der zweiten Genration: Prädiktive Modelle zur Vorhersage von Materialeigenschaften basierend auf deren chemischen Zusammensetzung.

- Problemstellungen der dritten Generation: Generative Modelle zur Vorhersage von Strukturen oder Kompositionen von Materialien mit gewünschten Eigenschaften (das auch sogenannte Inverse Design).

Zu diesen Problemstellungen wurden jeweils geeignete Modelle mit Hilfe von praxisrelevanten Problemstellungen aus den Anwendungsbereichen Moleküle, Polymere und Glaeser entwickelt und untersucht.

## 2. Ablauf und Ergebnisse des Vorhabens

Das Vorhaben umfasste die Zusammenarbeit von universitären Arbeitsgruppen mit Expertise in numerischer Mathematik, Analysis, Optimierung und angewandter Funktionalanalysis in Kooperation mit dem Fraunhofer SCAI. Die universitären Arbeitsgruppen hatten unterschiedliche Schwerpunkte und behandelten verschiedene Aspekte der betrachteten Lernverfahren. Das Design der Architektur diente als verbindendes Element, sei es durch die Integration von Domänenwissen und Transferlernen (Bonn), die Betrachtung von Expressivität und Interpretierbarkeit (München), das Training der Modelle (Braunschweig) oder die Anforderungen aus der Anwendungsperspektive (SCAI mit den assoziierten Projektpartnern). Das Fraunhofer SCAI brachte seine Expertise in den Bereichen Softwareengineering, ingenieurmäßige Anbindung und Verwertung der Projektergebnisse ein. Darüber hinaus waren die assoziierten Partner Covestro und Schott, Experten für konkrete Fragestellungen in den Anwendungsbereichen Moleküle/Polymere und Gläser, beteiligt.

Die Planung und Durchführung des Vorhabens umfasste folgende Arbeitspakete, wobei hier nur die SCAI betreffenden Aufgaben detaillierter dargestellt sind (der jeweilige Teilprojektkoordinator ist in Klammern vermerkt):

- Arbeitspaket AP0 (UB): Koordination:
- Arbeitspaket AP1: Demonstratoranwendung:
- Arbeitspaket AP2 (LMU): Expressivität und Interpretierbarkeit
- Arbeitspaket AP3 (TUBS): Training und Architektur
- Arbeitspaket AP4 (UB): Integration von Domänenwissen, Erklärbarkeit und Transferlernen
- Arbeitspaket AP5 (SCAI): Anwendung in der Materialentwicklung
  In diesem Arbeitspaket wurden auf Graphnetzwerken basierte Methoden für die spezielle Anwendung in der Materialentwicklung entwickelt und implementiert. Dazu wurde es wie folgt untergliedert:
  - AP 5.1: Graphnetzwerke und Workflows für Probleme der 1. Generation
  - AP 5.2: Graphnetzwerke und Workflows für Probleme der 2. Generation
  - AP 5.3: Graphnetzwerke und Workflows für Probleme der 3. Generation
  - AP 5.4: Generierung von Daten
- Arbeitspaket AP6 (SCAI): Praxisrelevante Anwendung
  In diesem Arbeitspaket wurden die neuen Methoden auf die Problemstellungen der Praxispartner angewendet und detailliert untersucht.
  - AP 6.1: Anwendungsfall Polymere (+ assoziierter Partner Covestro)
  - AP 6.2: Anwendungsfall Gläser (+ assoziierter Partner Schott)

Das Vorhaben hat bedeutende Beiträge zur Entwicklung von innovativen datengetriebenen Materialdesignmethoden geliefert. Es wurden insbesondere die Optimierung von Algorithmen des maschinellen Lernens sowie die Bewertung von Grenzen datengetriebener Modelle adressiert. Die Anwendung in den Materialwissenschaften hat eine Verknüpfung modellbasierter und daten-getriebener Ansätze ermöglicht, indem sowohl reale Messdaten

als auch Daten aus numerischen Simulationen als simulierte Trainingsdaten eine Rolle spielten. Der Transfer von Grundlagen und Methoden der Mathematischen Modellierung, Simulation und Optimierung (MMSO) in die industrielle Anwendung wurde erfolgreich umgesetzt.

## 3. Darstellung der wesentlichen Ergebnisse und deren konkreter Nutzen sowie ggf. die Zusammenarbeit mit anderen Forschungseinrichtungen

Die wesentlichen Ergebnisse (und deren Nutzen) von Teilprojekt 4 sind:

- Neuartige Vorhersagemodell fuer Molekül- und Materialeigenschaften
    - Diese können zur Unterstützung der Entwicklung neuer Substanzen und Materialen genutzt werden.
- Neuartige Wechselwirkungspotentiale für atomistischer Systeme
    - Diese können zur Molekuardynamiksimulation verwendet werden und somit zum Beispiel innerhalb der Entwicklung von Energiematerialien.
- Neuartige generative Methoden zur Generation von Molekülen und Materialien unter Nebenbedingungen.
    - Diese können dazu genutzt werden Vorschläge für Moleküle/Materialen mit spezifischen Eigenschaften zu generieren, dem sogenannten Inversen Design. Dies verspricht die Anzahl der notwendigen kosten- und zeitintensiven Experimente in der Molekül- und Materialentwicklung substantiell zu reduzieren.