Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Mathematical models as research data via flexiformal theory graphs

Michael Kohlhase[1], Thomas Koprucki[2], Dennis Müller[1], Karsten Tabelow[2]

submitted: March 31, 2017

[1] Informatik, FAU Erlangen-Nürnberg
Martensstr. 3
91058 Erlangen
Germany

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: thomas.koprucki@wias-berlin.de
karsten.tabelow@wias-berlin.de

No. 2385
Berlin 2017

2010 *Mathematics Subject Classification.* 86T30, 35Q68, 68Q55.

Key words and phrases. Mathematical models, OMDoc/MMT, Research data, Mathematical knowledge management.

MK and DM gratefully acknowledge EU funding for the OpenDreamKit project in the Horizon 2020 framework under grant 676541.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Mathematical models as research data via flexiformal theory graphs

Michael Kohlhase, Thomas Koprucki, Dennis Müller, Karsten Tabelow

Abstract

Mathematical modeling and simulation (MMS) has now been established as an essential part of the scientific work in many disciplines. It is common to categorize the involved numerical data and to some extent the corresponding scientific software as research data. But both have their origin in mathematical models, therefore any holistic approach to research data in MMS should cover all three aspects: data, software, and models. While the problems of classifying, archiving and making accessible are largely solved for data and first frameworks and systems are emerging for software, the question of how to deal with mathematical models is completely open.

In this paper we propose a solution – to cover all aspects of mathematical models: the underlying mathematical knowledge, the equations, boundary conditions, numeric approximations, and documents in a flexiformal framework, which has enough structure to support the various uses of models in scientific and technology workflows.

Concretely we propose to use the OMDoc/MMT framework to formalize mathematical models and show the adequacy of this approach by modeling a simple, but non-trivial model: van Roosbroeck's drift-diffusion model for one-dimensional devices. This formalization – and future extensions – allows us to support the modeler by e.g. flexibly composing models, visualizing Model Pathway Diagrams, and annotating model equations in documents as induced from the formalized documents by flattening. This directly solves some of the problems in treating MMS as "research data" and opens the way towards more MKM services for models.

1 Introduction

Mathematics is a common ground for science and technology: research problems are described using mathematical models, which are then solved either by symbolic derivation or numerical simulation. In the last decade *mathematical modeling and simulation* (MMS) has been established as a primary scientific research method alongside the classical methods of experiment and theory. It is now an essential part of the scientific work in many disciplines and application areas. Research in the area of MMS is characterized by *mathematical models*, *scientific software* for their treatment, and *numerical data* related to computations (input, output, parameters), see [1]. There, it was proposed to categorize these three parts as the research data in MMS as they are jointly required to understand and verify research results, or to build upon them.

Specifically, *numerical data* is generally regarded as research data in usual sense and data repositories and information services such as DataCite [2] or RADAR [3, 4] exist or are emerging. Increasingly, *software* is also categorized as research data [5] and an information service on mathematical software, swMath [6], has already been developed.

The representation of mathematical models themselves – the mathematical knowledge they contain and the discipline-specific knowledge they are based on – is far less clear. Current practice is to publish them as mathematical papers with a mixture of mathematical formulae and natural language.

DOI 10.20347/WIAS.PREPRINT.2385 Berlin 2017

This leads to ambiguity, duplication, and incompleteness in presentation and makes the treatment of models as research data impossible. To allow "data repositories" for models we would need a way of automatically identifying and classifying them. Analogously, connecting models to input/output data or to software systems is impossible without.

Classically the computer-actionable representation of "models" is the domain of "modelling languages like the Universal Modelling Language (UML [7]), Systems Biology Markup Language (SBML [8, 9]), or MODELICA [10], a "modelling language for physical systems". These allow to describe complex software/biological/physical systems in terms of their components and the connections/interactions between them in a machine-actionable way. These languages usually support visualizations of the respective "models" diagrammatically for communication with/among humans and code generation for the computational systems of the domain. Many of them come with large libraries of standardized components that make the assembly of models less tedious.

We claim that these languages only solve part of the description problem for MMS models: They describe large technical or biological systems composed of elementary units and their effective behavior by parametrized and often empirical *compact models*. However, they do not provide a detailed description of the physical (biological or chemical) spatio-temporal processes governing their behavior on the level of fundamental laws and constitutive relations, typically expressed by partial differential equations. This type of mathematical models is important for many disciplines such as the natural and engineering sciences as well as life and environmental sciences, but a machine-understandable representation is missing.

In this paper restrict our attention to such models: complex models of simple devices. If we can formalize and machine-support them, we can scale them up to complex devices with classical modeling languages. We will perform a case study for the representation of such type of models using the van Roosbroeck system [11]. It is the standard model [12, 13] to describe the current flow in semiconductor devices ranging from diodes and transistors to LEDs, solar cells, lasers and novel materials such as organic semiconductors. Even its relatively simple one-dimensional stationary version has more than ten non-trivially connected equations.

In this paper we propose a solution to the "models as research data" problem by flexiformalization of all aspects of mathematical models: the underlying mathematical knowledge, the equations, boundary conditions, numeric approximations, and documents in a framework that has enough structure to support the various uses of models in scientific and technology workflows.

Concretely we propose to the OMDoc/MMT framework to model mathematical models and show the adequacy of this approach by modeling a simple, but non-trivial model: van Roosbroeck's drift-diffusion model in one-dimensional devices. To make this paper self-contained we introduce OMDoc/MMT theory graphs in the next section and the mathematics of the van Roosbroeck model in Section 3. Then we discuss the flexiformalization in Section 5 and the services that can be built on this in Section 6. Section 7 concludes the paper.

2 Flexiformal Theory Graphs

OMDoc [14] is a wide-coverage representation language for mathematical knowledge (formal) and documents (informal/narrative). In the last decade development has focused on the formal aspect leading to the OMDoc/MMT instance (Meta-Meta-Theories [15, 16, 17]), which increases expressivity, clarifies the representational primitives and formally defines the semantics of this fragment.

OMDoc/MMT is designed to be foundation-independent and introduces several concepts to maximize modularity and to abstract from and mediate between different foundations, to reuse concepts, tools, and formalizations. The OMDoc/MMT language *integrates successful representational paradigms*

- the logics-as-theories representation from logical frameworks,
- theories and the reuse along theory morphisms from the heterogeneous method,
- the Curry-Howard correspondence from type theoretical foundations,
- URIs as globally unique logical identifiers from OpenMath,
- the standardized XML-based interchange syntax of OMDoc,

and makes them available in a single, coherent representational system for the first time. The combination of these features is based on a small set of carefully chosen, orthogonal primitives in order to obtain a simple and extensible language design.

OMDoc/MMT offers very few primitives, which have turned out to be sufficient for most practical settings. These are

1. *constants* with optional types and definitions,
2. types and definitions of constants are *objects*, which are syntax trees with binding, using previously defined constants as leaves,
3. *theories*, which are lists of constant declarations and
4. *theory morphisms*, that map declarations in a domain theory to expressions built up from declarations in a target theory.

Using these primitives, logical frameworks, logics and theories *within* some logic are all uniformly represented as OMDoc/MMT theories, rendering all of those equally accessible, reusable and extendable. Constants, functions, symbols, theorems, axioms, proof rules etc. are all represented as constant declarations, and all terms which are built up from those are represented as objects.

Theory morphisms represent truth-preserving maps between theories. Examples include theory inclusions, translations/isomorphisms between (sub)theories and models/instantiations (by mapping axioms to theorems that hold within a model), as well as a particular theory inclusion called *meta-theory*, that relates a theory on some meta level to a theory on a higher level on which it depends. This includes the relation between some low level theory (such as the theory of groups) to its underlying foundation (such as first-order logic), and the latter's relation to the logical framework used to define it – e.g. LF; see [18] for an overview.

All of this naturally gives us the notion of a *theory graph*, which relates theories (represented as nodes) via edges representing theory morphisms (as in Figure 1), being right at the design core of the OMDoc/MMT language. It is a central advantage of the OMDoc/MMT system that theory morphisms "transport axioms, definitions, theorems, . . . " to new contexts and thus induce knowledge that is not explicitly represented in the graph. Therefore it is a central design invariant of the system that we can name all induced objects with canonical URIs, the MMT URIs, which contain enough information to reconstruct the induced objects themselves – given the graph.

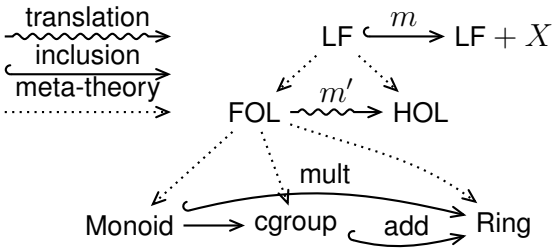

Figure 1: A Theory Graph with Meta-Theories

Recently, OMDoc/MMT has been extended to enable handling content of *flexible formality* [19] in a bid to reach full OMDoc coverage. In a nutshell, *informal* parts are modeled as opaque constants,

objects or theories [20]. While they can obviously not be formally analyzed with respect to their formal structure, they can still be used in (and be subject to) the various knowledge management services provided by MMT, in particular they can be connected to formal content via theory morphisms. As a result, we believe we can use OMDoc/MMT to represent all kinds of mathematical models and their domains of application in a unified manner, whether they can be fully formalized in some logic or need to be represented informally.

This approach seems to be feasible in view of the the general L-concept of physical theories [21], which relies on a formalization of statements on physical objects and their relations also using a formalized mathematical theory.

3 Van Roosbroeck Model

As guiding example for the formalization of models we consider a simplified variant of the van Roosbroeck system [11, 12, 13]. It describes the flow of charge carriers (electrons and holes) in a self-consistent electric field in a semiconductor device using a drift-diffusion approximation. Therefore it is also frequently called the *drift-diffusion model*.

In its unipolar version, that means only considering one charge carrier species, it is suited for the simulation of many devices ranging from simple layered n-doped/intrinsic/n-doped (nin) structures, see Figure 2, to organic transistors [23]. Specifically, we will focus in this paper on the one-dimensional unipolar van Roosbroeck system assuming a homogeneous material on an interval $\Omega = [0, L]$ with Ohmic contacts at each end. It consists of two nonlinear ordinary differential equations for the unknown electrostatic potential $\psi(x)$ and the quasi Fermi potentials for electrons $\varphi_n(x)$: follows:

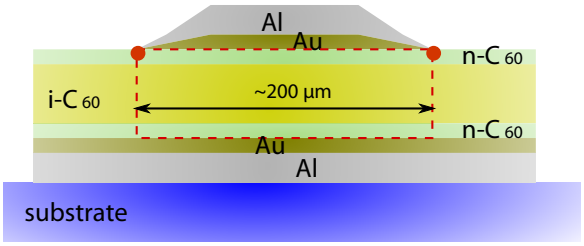

Figure 2: Schematic structure of a device consisting of n-doped/intrinsic/n-doped layers of the organic semiconductor C_{60}. Reprinted from [22] with permission from Elsevier.

$$-\frac{d}{dx}\left(\varepsilon_s \frac{d}{dx}\psi\right) = q\left[C - N_c \exp\left(\frac{q(\psi-\varphi_n)-E_c}{k_B T}\right)\right], \tag{1a}$$

$$\frac{d}{dx}\left[-q\mu_n N_c \exp\left(\frac{q(\psi-\varphi_n)-E_c}{k_B T}\right)\frac{d}{dx}\varphi_n\right] = 0. \tag{1b}$$

It links Poisson's equation (1a) for the electrostatic potential to the continuity equation for electron density

$$n(\psi,\varphi_n) = N_c \exp\left(\frac{q(\psi-\varphi_n)-E_c}{k_B T}\right). \tag{2}$$

The term in brackets in continuity equation (1b) represent the flux of the electron current density j_n. We stress here we have used the so-called Boltzmann approximation. In general, the exponentials will be replaced by some monotonically increasing *statistical distribution function* $\mathcal{F}$ [13].

The elementary charge q and the Boltzmann constant k_B are universal physical constants. The (absolute) dielectric permittivity $\varepsilon_s = \varepsilon_0\varepsilon_r$ is given as the product of the vacuum dielectric permittivity ε_0 and the relative permittivity of the semiconductor material ε_r. The electron mobility μ_n, the effective

conduction band density of states N_c as well as the conduction band-edge energy E_c are assumed to be constant. The temperature T is also assumed to be constant. The doping profile $C = C(x)$ describes material properties. For a more detailed discussion on the physics behind these different quantities we refer to [12, 13].

The system (1) needs to be supplied with boundary conditions at $x = 0$ and $x = L$. For applied external voltages U_1 and U_2, we require the Dirichlet boundary conditions at the *Ohmic contacts*, that is

$$\psi(0) = \psi_0(0) + U_1, \qquad \psi(L) = \psi_0(L) + U_2, \tag{3a}$$

$$\varphi_n(0) = U_1, \qquad \varphi_n(L) = U_2, \tag{3b}$$

For a discussion on the meaning and choice of the build-in potentials $\psi_0(0)$, $\psi_0(L)$ see [13].

In the literature, authors frequently use a different set of unknowns for the van Roosbroeck system by replacing the quasi Fermi potential with the electron density n. A formulation based on the electron density involves an alternative expression for electron current entering the continuity equation, namely the classical drift-diffusion form, see [13]. Consequently, the choice of the unknowns introduces an ambiguity of the model equations which a mathematical knowledge management system for its formal representation has to reflect.

Here, we start with the more general formulation using the quasi Fermi potential, see (1), as it naturally appears in the thermodynamic description of the van Roosbroeck system since the negative gradient of the quasi Fermi potential is the driving force of the current [24]. Moreover, it is even mathematically more beautiful as it makes it possible to write the whole van Roosbroeck system in a gradient form [25]. In the following section we will develop a formal representation of this model and demonstrate how these different aspects of the coupled system can be explained therein.

4 Model Pathways Diagrams

As we have seen in the last section, even relatively simple mathematical models – the model only covers the stationary (time-independent) case in one dimension – can be quite intimidating. The central intuition that helps understand them is a model makes assertions about measurable quantities in the real world – in our case in the physical world. Concretely, models employ a system of physical laws that assert relations between physical quantities. For instance a “displacement law” asserts a relation between the electric field E, the dielectric permittivity ε_s of a material and the resulting displacement field D accounting for induced polarization due to the bound charges, and can therefore be used to compute the latter from the former. Similarly, a permittivity law can be used to compute the (absolute) permittivity ε_s from the vacuum dielectric permittivity ε_0 and a material parameter ε_r.

To help understand the inner (physical) structure of mathematical models (and guide formalization) we have developed a special kind of diagrams: **Model Pathway Diagrams** (MPD). These depict physical quantities – see [26] for a list – as circles labeled with their physical notations and connect them by physical laws, which we draw as named rectangles which contain the respective equation. In a MDP each law node must be connected to all quantities that appear in the equation by a path. We use undirected edges in MPDs, since mathematically, physical laws are relations only and therefore have no prescribed input/output directions. MPDs may be cyclic, but should be transitively non-redundant. The usefulness of diagrammatic representations for physical phenomena has been proven for example by Feynman diagrams [27] in the perturbation theory for quantum field theory.

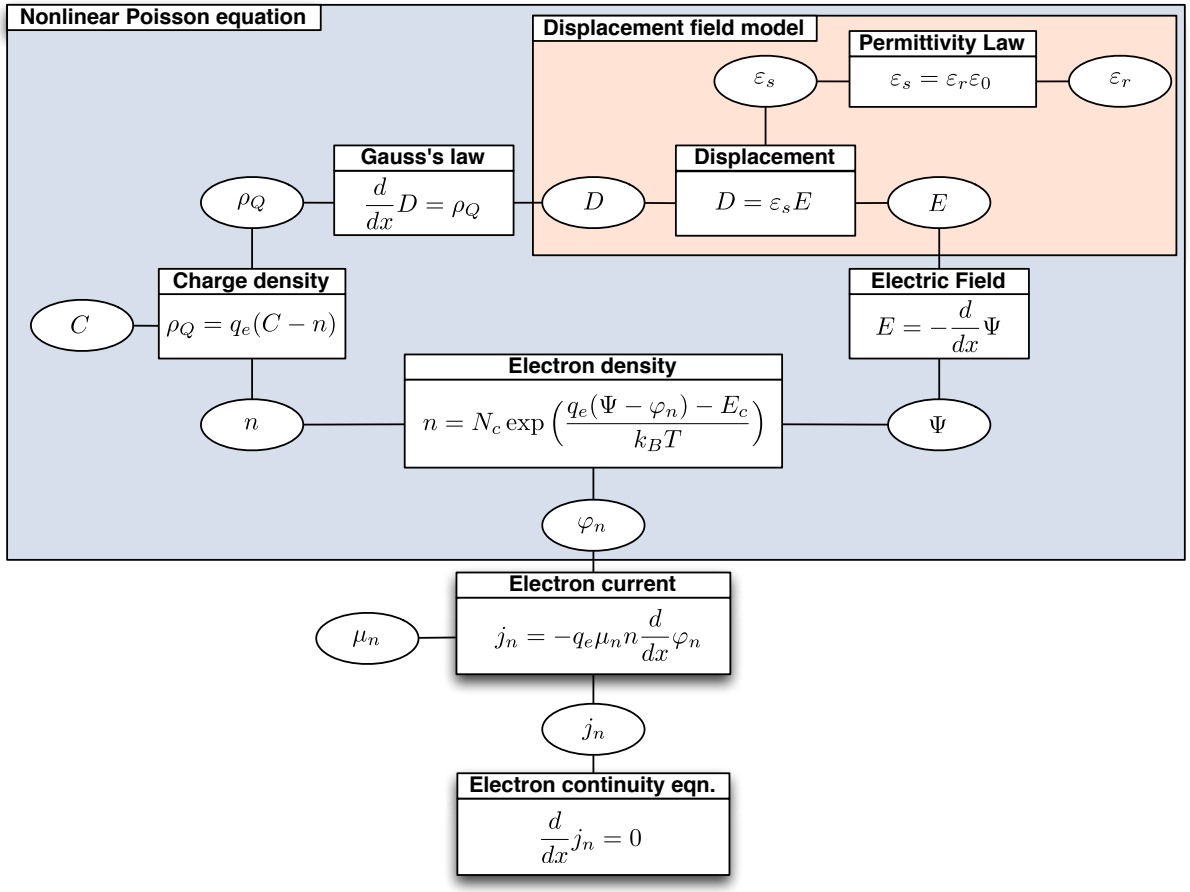

Figure 3: The Model Pathway Diagram for the van Roosbroeck model, cf. Eqs.(1). Two sub MPDs are highlighted using colored boxes: the displacement field model (red) and the nonlinear Poisson equation (blue). For simplicity the handling of the boundary conditions has been dropped as well as some quantities, e.g. conduction band-edge energy E_c or the temperature T.

Figure 3 shows a MPD for the van Roosbroeck system that teases apart the respective contributions of the various laws and shows the quantities they relate.

Definition 1 (MPD Model) Let $\mathcal{Q}$ denote the set of quantities in an MPD M and $\mathcal{L} \subseteq \mathcal{Q}$ be the set of leaves in the quantities of M. The triple $\mathcal{M} = (M, \mathcal{U}, \mathcal{P})$, is a **MPD model for** a set of **unknown quantities** $\mathcal{U} \subseteq \mathcal{Q}$ **given** the parameters $\mathcal{P} = \mathcal{L} \backslash \mathcal{U}$, if the MPD has at least one quantity leaf, i.e., $\mathcal{L}$ is not empty. An MPD model for $\mathcal{U}$ given $\mathcal{P}$ is **underdetermined** if $|\mathcal{Q}| > |\mathcal{R}| + |\mathcal{U}| + |\mathcal{P}|$, where $\mathcal{R}$ is the set of relations (laws, constitutive relations) in the MPD.

Example 1 (van Roosbroeck model) *The MPD in Figure 3 is an MPD model, the van Roosbroeck model, for the electrostatic potential ψ and the quasi-Fermi potential φ_n given the doping profile C, the relative permittivity ε_r, and the electron mobility μ_n. Thus, $\mathcal{U} = \{\psi, \varphi_n\}$ and $\mathcal{P} = \mathcal{L} = \{C, \varepsilon_r, \mu_n\}$.*

Example 2 (Displacement field model) *The sub MPD in Figure 3 highlighted by the red color is a model for the displacement field D given the electric field E and the relative permittivity ε_r. $\mathcal{U} = \{D\}$ and $\mathcal{P} = \mathcal{L} \backslash \mathcal{U} = \{E, \varepsilon_r\}$. Its flexi-formal representation in OMDoc/*MMT *will be discussed in Section 5.*

Example 3 (Nonlinear Poisson equation) *The sub MPD in Figure 3 highlighted by the blue color is a model, the nonlinear Poisson equation, for the electrostatic potential ψ given the quasi-Fermi potential φ_n, the doping profile C, and the relative permittivity ε_r. Thus $\mathcal{U} = \{\psi\}$ and $\mathcal{P} = \mathcal{L} = \{C, \varepsilon_r, \varphi_n\}$.*

Remark 1 (Physical equivalent sets of unknown quantities) *The choice of unknown quantities $\mathcal{U}$, e.g., in the nonlinear Poisson equation, involves some ambiguity: Instead of $\mathcal{U} = \{\psi\}$, one could as well use, e.g., the electric field E or the electron density n. These choices are physically equivalent and the particular selection depends on the specific aspects to be modeled.*

Remark 2 (Inverse problems) *Furthermore, if we replace $\mathcal{U} = \{\psi\}$ by the doping profile $\mathcal{U} = \{C\}$ the resulting MPD model for the doping profile C is underdetermined and constitutes an inverse problem. In order to enable its solution an additional leaf for a quantity ψ_0 can be added connected to the MPD by a target law $\psi = \psi_0$. Then the solution determines a doping profile, when the target potential ψ_0 is attained.*

In the MPD in Figure 3, we can directly get an overview over the structure of the van Roosbroeck model. We observe the (nonlinear) Poisson equation complex on the top and the carrier transport complex on the bottom, which are coupled by the quantities n (electron density), and Ψ. N.B. the electron density n is physically equivalent to the quasi-Fermi potential in the sense of Remark 1. The corresponding sub-MPDs are directly related to the sub equations (1a) and (1b) of the van Roosbroeck model. Both sub-MPDs posses a distinct topological structure: The nonlinear Poisson equation is characterized by the loop (diamond structure) of the density n, the charge density ρ_Q, the displacement D, the electric field E and the electrostatic potential Ψ. Contrary the sub-MPD for the carrier transport has a tree-like structure with the root being the quasi-Fermi potential φ_n. These topological structures can be utilized for the mathematical theory of the respective equations. Furthermore, it paves a way for the development of iterative schemes for the numerical solution of the fully coupled system, e.g. the Gummel's (decoupling) method [28, 12, 13].

5 Flexi-Formalizing a non-trivial Model, a Case Study

We are currently studying the model introduced in the last section, formalizing the inherent knowledge in OMDoc and augmenting (parts of) [13] into an active document, see [29] for first results.

We base the development on a higher-order logic with records, (predicate) subtypes, and literals [30] for basic mathematical objects like real numbers and arithmetics. We use this theory as the meta-language for all theories, but do not show it in our diagrams. Instead we start the development with the Math-in-the-Middle development of elementary maths [31] and let the formalization be guided by the MPD in Figure 3.

$\mathcal{Q}_{DielectricPermittivity}$
permittivity : type $= \frac{\text{charge}}{\text{voltage}\cdot\text{length}}$

We model physical quantities as special theories that introduce a (type) constant which is a SI dimension. For instance the theory for the dielectric permittivity on the right. We call these theories **quantity theories**; they directly implement the quantity nodes (circles) in MPDs. Given these basic theories, (and more mathematical and physics background; e.g. SI units), we can implement the physical laws as OMDoc/MMT theories. Figure 4 has e.g. the theory DispLaw, which includes three quantity theories $\mathcal{Q}_{DielectricPermittivity}$, $\mathcal{Q}_{EField}$, and $\mathcal{Q}_{Displ}$ (and some more background material not shown in Figure 4) and states their relation in a simple equation. Similarly, the theory DPMat corresponds

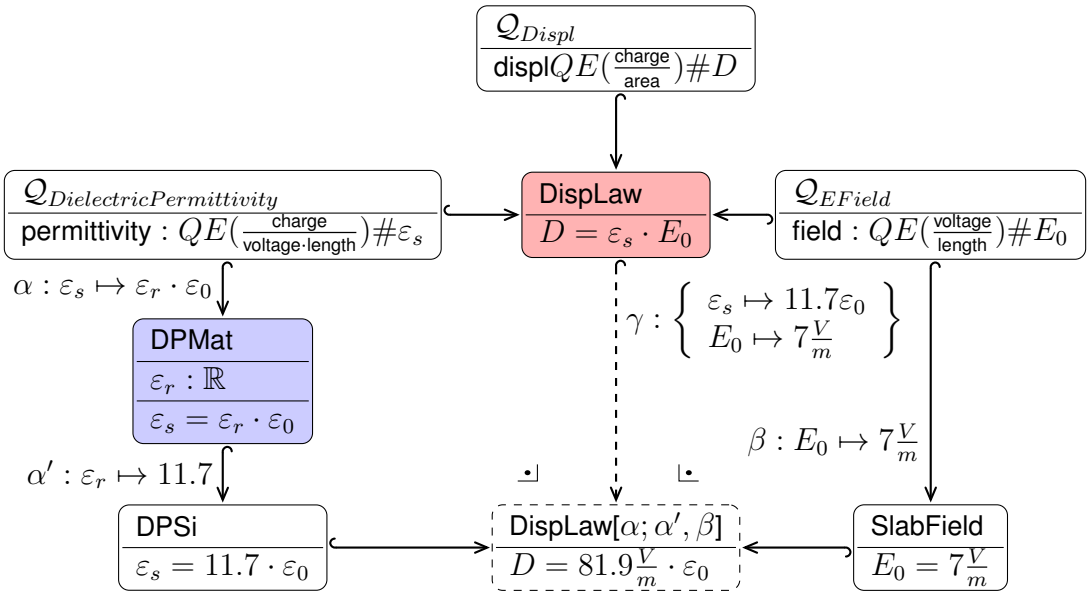

Figure 4: Application of Models

to the Permittivity Law theory in Figure 4. Figure 5 has the OMDoc/MMT source for the charge law (called "electron current" in Figure 3).

Note that a theory graph $\mathcal{T}$ as the one in Figure 4 induces a MPD $\mathcal{M}$ in the sense of the last section. In particular it corresponds the MPD Displacement field model in Figure 3, see also Example 2. This motivates the following

Definition 2 (Model in a Theory Graph) Let $\mathcal{T}$ be an OMDoc/MMT theory graph, $\mathcal{M}$ a subgraph in $\mathcal{T}$ consisting only of quantity/law theories and inclusions between them, and $\mathcal{D}_{\mathcal{M}} := (N, E)$ where N and is the set of nodes of $\mathcal{T}$ labeled with the a) notation of the quantity symbol for quantity theories and b) the content of the axioms for the law theories and E is the set of inclusions in $\mathcal{M}$. Then we call $\mathcal{M}$ a **model** in $\mathcal{T}$, iff $\mathcal{D}_{\mathcal{M}}$ is a MPD model. We define a "model for $\mathcal{U}$ given $\mathcal{P}$" as in Definition 1.

The lower part of Figure 4 shows the process of instantiating the model in the upper part to concrete values via theory extensions (via the inclusions α' and β) to concrete values – here $\varepsilon_r = 11.7$ for the relative dielectric permittivity of Silicon (in theory DPSi) and $E_0 = 7\frac{F}{m}$ for the electric field applied to the silicon slab (in theory SlabField). These together allow to compute the theory DispLaw[$\alpha; \alpha', \beta$] and inclusion γ (the dashed part of Figure 4) as the co-limit of DPSi and Slabfield modulo DispLaw. Technically, the co-limit DispLaw[$\alpha; \alpha', \beta$], which represents the result of the model instantiated to device that is a Silicon Slab with an external field of $7\frac{V}{m}$ can be computed as the double pushout along $\alpha; \alpha'$ and β as all the theory morphisms involved are injective. Note that α is a model-internal instantiation that connects the two laws into a model, whereas α' and β are application morphisms that instantiate it to a particular situation description.

In Figure 4 we have only shown a small part of the van Roosbroeck model, but we have developed the whole OMDoc/MMT theory graph corresponding to Figure 3.

Figure 5 shows a snippet of our formalization in its original OMDoc/MMT syntax. It shows the law for charge densities as occurring in the in the Van Roosbroeck model as a constant total_charge_law. The equation is stated in the type using the Curry-Howard Isomorphism: $\vdash E$ is the type of proofs of E. It relates the total charge $Q(x)$ to various other quantities (such as the doping profile $C(x)$ and the

```
theory ChargeLaw : top:?Base =
  include device?DeviceGeometry
  include device?DopingProfile
  include ?PoissonParameters
  include top:/Species?ElectronsAndHoles
  include top:?SpatialChargeDensity

  /T The total charge denoted by $Q$ composed of doping profile $C$ and
   electron and hole densities $n$ and $p$, defined by
   $Q$ = $q$($C$+$p$ - $n$).
  total_charge_density : {x : Ω} ⊢ Q x ≐
    -q · ((C x) + (holes/z · (holes/density x)) + (electrons/z · (electrons/density x))
```

Figure 5: Excerpt of the OMDoc/MMT Formalization (Total Charge Density Law)

respective densities of the electrons and holes) that are imported via theory inclusions. This modular structure allows for selectively substituting the imported quantities with specific values.

The complete theory graph (not counting the background knowledge in the MitM ontology) contains 38 theories and 63 inclusions. As these theories are exclusively physical quantities and laws, we expect them to be highly reusable.

6 Knowledge Management Services for Models

The most immediate consequence of Definition 2 is that given an overall theory graph $\mathcal{T}_{MMS}$ of quantities, laws, and background knowledge formalizations, we can represent models as sets of MMT URIs – of the theories and inclusions in the model. This allows to build model repositories as envisioned in the introduction.

Another consequence of the correspondence between theory graphs and MPD established by Definition 2 is that we can build a MPD-viewer for models (again given $\mathcal{T}_{MMS}$). All we need is to build a graph viewer that distinguishes between quantity and law theories – we have encoded the necessary information in special MMT metadata in our formalization. Figure 6 shows our formalization of the van Roosbroeck model in our prototype MPD viewer.

The representation of the physical relations between the quantities in the MDP and the corresponding theory graph allows for a quick and easy creation of a zoo of related mathematical models. For example, in the MPD of the van Roosbroeck system in Figure 3 one can easily replace the exponential Boltzmann law for the electron density by a Fermi-Dirac or Gauss-Fermi statistics for the description of organic semiconductors. The relationship between these models is given by the sharing of the same structure using different theories for specific relations between quantities. In total this enables the derivation of a classification of mathematical models based on the representing theory graphs or the corresponding MPDs.

The fresh view on mathematical models by MPDs allows us to highlight the complex structure of coupled systems of partial differential equations and supports their development from smaller building blocks. The classical van Roosbroeck system for electrons *and* holes can be constructed from the unipolar version for electrons only, cf. Figure 3 by duplicating the transport complex and introducing

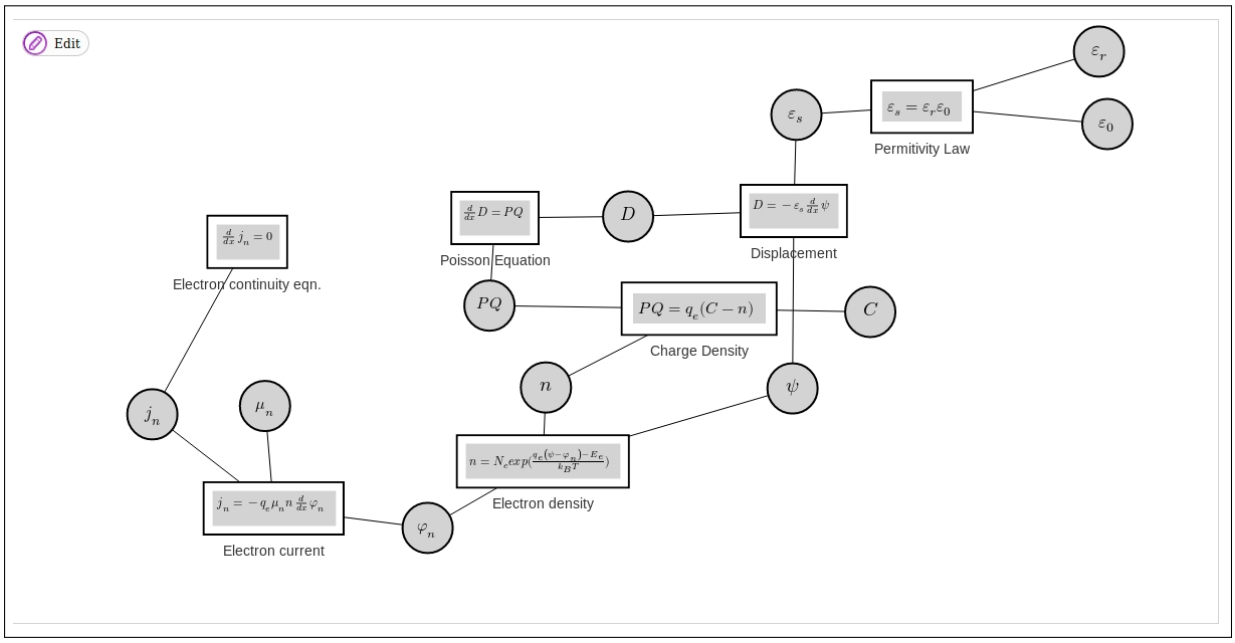

Figure 6: Screenshot of the MPD Viewer displaying the van Roosbroeck Theory Graph, cf. Figure 3

a hole density quantity p and a corresponding quasi-Fermi potential φ_p. The expression for the total charge density ρ_Q has to be adjusted accordingly. Furthermore, recombination processes can be added to the model combining electron and hole density on the level of the transport complex. By doing so, the two topological separate sub-MPDs for the electrons and holes couple, see [32]. This requires a modular structure by parametrizing and thus sharing the formalization of density and current laws for both species, electrons and holes. This can be achieved by standard features (structures; named imports) in OMDoc/MMT .

The elimination of quantities from the MPD or theory graph by incremental inserting – suitably rewritten versions of the – relations generates a large variety of derived MPDs using fewer quantities and relations increasing their individual complexity and losing semantic details represented by the full graph. This process can be meta-modeled in OMDoc/MMT by adding induced views to the theory graph (outside of the MPD subset) so that the relation reduction corresponds directly to view application. The main advantage for MKM purposes is that the flattened version can be directly written in the form of a MMT URI. Even though it is not explicitly represented in the graph it can be computed by the MMT system given only the represented graph and the MMT URI.

For example, the nonlinear Poisson equation in (1a) represents the fully flattened version of the depicted MPD, cf. Example 3. These more compact versions of the model might have advantages for specific analysis or objectives even though they are semantically less vivid.

7 Conclusion and Future Work

We have developed a knowledge-based meta-modeling approach for mathematical models and a first set of added-value services that make use of the involved representations.

For discovering structure in mathematical models we have introduced *model pathway diagrams* (MPDs). As these constitute an important didactical and structural representation of mathematical models, we have started a collection effort for MPDs at [32]. On the one hand this can serve as a seed for a future

model repository that complements existing research data repositories, on other hand, MPDs guide formalization, and so provide a valuable first step towards scaling the model formalization effort started in this paper.

We have further formalized a simple, but non-trivial model in a modular OMDoc/MMT theory graph $\mathcal{T}_{vR}$ and have used this representation for associating the equations by their MMT URI. These can be used explaining the mathematical provenance of the "flattened" equations in papers from the physics knowledge in $\mathcal{T}_{vR}$. Future services could include "formula search modulo flattening" via our $\flat$-search system [33] to discover instances of models in the literature or plug-and-play model composition services, where models can be assembled from theories in a "physics theory graph" based on on MPD viewer.

The next steps of our further research will focus on the understanding of finite-volume discretization as MPDs, which will provide the connection to the numerical methods for the determination of a specific solution for given parameters and the software implementing those. We will investigate the relation between the MPDs of the continuous model (system of PDEs) and its discretized counterpart. Moreover, a modular concept for spatially multi-dimensional models and its theory graphs has to be developed. Having established a collection of models with deep MPD graphs we will study their refactoring into smaller components. Here, we gained first experience by refactoring of the MPD of the bipolar van Roosbroeck system for electrons and holes using parameterized MPDs describing a generic carrier species, see Section 6.

The main difference of our approach here to classical modeling languages like MODELICA is that this is mostly interested in modeling *physical systems* that are built up from connected components, whereas we are interested in modeling *the physics of a system* (similar words, but very different meaning). Consequently our models consist of quantities connected by physical laws (or dually: physical laws connected by quantities), not component systems by physical connections. The actual model descriptions are similar, since the physical connections are often governed by equations as well, e.g. for transport of some material flow or electrical current. SBML is the same, only that it model cells or sub-cellular structures as systems of reactants, connected by reactions (which are governed by rules). Even UML is similar, there we have software components which are connected by ports.

Acknowledgements

We gratefully acknowledge EU funding for the OpenDreamKit project in the Horizon 2020 framework under grant 676541. Our discussions have particularly profited from contributions by Florian Rabe (MMT advice) and Wolfram Sperber (general math background). Finally, Marcel Rupprecht has developed the MPD viewer mentioned in Section 4.

Literatur

[1] Thomas Koprucki und Karsten Tabelow. "Mathematical Models: A Research Data Category?" In: *Mathematical Software – ICMS 2016: 5th International Conference, Berlin, Germany, July 11-14, 2016, Proceedings*. Hrsg. von Gert-Martin Greuel u. a. Cham: Springer International Publishing, 2016, S. 423–428. ISBN: 978-3-319-42432-3. DOI: 10.1007/978-3-319-42432-3_53.

[2] Jan Brase. *DataCite – A global registration agency for research data*. IEEE, 2009.

[3] M. Razum, J. Neumann und M. Hahn. "RADAR–Ein Forschungsdaten-Repositorium als Dienstleistung für die Wissenschaft". In: *Zeitschrift für Bibliothekswesen und Bibliographie* 61.1 (2014), S. 18–27.

[4] Angelina Kraft. “RADAR-a repository for long tail data”. In: *Proceedings of the IATUL Conferences. Paper 1.* 2015.

[5] Deutsche Forschungsgemeinschaft. *DFG Guidelines on the Handling of Research Data*. Adopted by the Senate of the DFG at September 30. 2015.

[6] Gert-Martin Greuel und Wolfram Sperber. “swMATH – An Information Service for Mathematical Software”. In: *Mathematical Software – ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings*. Hrsg. von Hoon Hong und Chee Yap. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, S. 691–701.

[7] *Unified Modeling Language*. URL: http://www.uml.org/ (besucht am 13. 09. 2016).

[8] M. Hucka u. a. “The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models”. In: *Bioinformatics* 19.4 (2003), S. 524.

[9] *The Systems Biology Markup Language*. URL: http://sbml.org (besucht am 17. 03. 2017).

[10] ModelicaAssociation. *Modelica-A Unified Object-Oriented Language for Physical Systems Modeling-Language Specification Version 3.3 Revision 1, 2014*. 2014. URL: http://www.modelica.org.

[11] W. Van Roosbroeck. “Theory of the Flow of Electrons and Holes in Germanium and Other Semiconductors”. In: *Bell System Technical Journal* 29.4 (1950), S. 560–607.

[12] S. Selberherr. *Analysis and simulation of semiconductor devices*. Wien, New York: Springer, 1984.

[13] Patricio Farrell u. a. *Numerical methods for drift-diffusion models*. WIAS Preprint No. 2263. To appear in: Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Models — Volume Two. Ed. J. Piprek, Taylor & Francis, 2017. Berlin, 2016.

[14] Michael Kohlhase. *OMDoc – An open markup format for mathematical documents [Version 1.2]*. LNAI 4180. Springer Verlag, Aug. 2006. URL: http://omdoc.org/pubs/omdoc1.2.pdf.

[15] Florian Rabe und Michael Kohlhase. “A Scalable Module System”. In: *Information & Computation* 0.230 (2013), S. 1–54. URL: http://kwarc.info/frabe/Research/mmt.pdf.

[16] Fulya Horozal, Michael Kohlhase und Florian Rabe. “Extending MKM Formats at the Statement Level”. In: *Intelligent Computer Mathematics*. Hrsg. von Johan Jeuring u. a. LNAI 7362. Berlin und Heidelberg: Springer Verlag, 2012, S. 65–80. ISBN: 978-3-642-31373-8. URL: http://kwarc.info/kohlhase/papers/mkm12-p2s.pdf.

[17] Florian Rabe. “How to Identify, Translate, and Combine Logics?” In: *Journal of Logic and Computation* (2014). DOI: 10.1093/logcom/exu079.

[18] Frank Pfenning. “Logical Frameworks”. In: *Handbook of Automated Reasoning*. Hrsg. von Alan Robinson und Andrei Voronkov. Bd. I and II. Elsevier Science und MIT Press, 2001.

[19] Michael Kohlhase. “The Flexiformalist Manifesto”. In: *14th International Workshop on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012)*. Hrsg. von Andrei Voronkov u. a. Timisoara, Romania: IEEE Press, 2013, S. 30–36. ISBN: 978-1-4673-5026-6. URL: http://kwarc.info/kohlhase/papers/synasc13.pdf.

[20] Mihnea Iancu. “Towards Flexiformal Mathematics”. Diss. Bremen, Germany: Jacobs University, 2017.

[21] Joachim Schröter. *Zur Meta-Theorie der Physik*. Walter de Gruyter GmbH & Co KG, 1996.

[22] Axel Fischer u. a. “Self-heating effects in organic semiconductor crossbar structures with small active area”. In: *Organic Electronics* 13.11 (2012), S. 2461 –2468. ISSN: 1566-1199.

[23] Felix Kaschura u. a. “Operation mechanism of high performance organic permeable base transistors with an insulated and perforated base electrode”. In: *Journal of Applied Physics* 120.9 (2016), S. 094501.

[24] U. Bandelow, H. Gajewski und R. Hünlich. “Fabry–Perot Lasers: Thermodynamics-Based Modeling”. In: *Optoelectronic Devices*. Hrsg. von Joachim Piprek. Springer, 2005.

[25] A. Mielke. “A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems”. In: *Nonlinearity* 24.4 (2011), S. 1329.

[26] Wikipedia. *List of physical quantities — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/w/index.php?title=List_of_physical_quantities (besucht am 22. 03. 2017).

[27] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. In: *Phys. Rev.* 76 (6 1949), S. 769–789. DOI: 10.1103/PhysRev.76.769.

[28] H. K. Gummel. “A self-consistent iterative scheme for one-dimensional steady state transistor calculations”. In: *Electron Devices, IEEE Transactions on* 11.10 (1964), S. 455–465.

[29] Michael Kohlhase u. a. *A Case study for Active Documents and Formalization in Math Models: The van Roosbroeck Model*. URL: https://mathhub.info/MitM/models (besucht am 05. 02. 2017).

[30] Florian Rabe. “Generic Literals”. URL: http://kwarc.info/frabe/Research/rabe_literals_14.pdf.

[31] *MitM: The Math-in-the-Middle Ontology*. URL: https://mathhub.info/MitM (besucht am 05. 02. 2017).

[32] *MPDHub wiki*. URL: https://github.com/WIAS-BERLIN/MPDHub/wiki/ (besucht am 22. 03. 2017).

[33] Michael Kohlhase und Mihnea Iancu. “Searching the Space of Mathematical Knowledge”. In: *DML and MIR 2012*. Hrsg. von Petr Sojka und Michael Kohlhase. Masaryk University, Brno, 2012. ISBN: 978-80-210-5542-1. URL: http://kwarc.info/kohlhase/papers/mir12.pdf.