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

Preprint ISSN 2198-5855

Model pathway diagrams for the representation of mathematical models

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

Florian Rabe[3]

submitted: September 27, 2017

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

[2] Informatik, FAU Erlangen-Nürnberg
Martenstr. 3
91058 Erlangen
Germany

[3] Jacobs University Bremen
Campus Ring 1
28759 Bremen
Germany

No. 2431
Berlin 2017

2010 *Mathematics Subject Classification.* 86T30, 00A71, 68Q55.

Key words and phrases. Mathematical models, research data, model pathway diagrams, drift-diffusion equations.

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/

Model pathway diagrams for the representation of mathematical models

Thomas Koprucki, Michael Kohlhase, Karsten Tabelow, Dennis Müller, Florian Rabe

Abstract

Mathematical models are the foundation of numerical simulation of optoelectronic devices. We present a concept for a machine-actionable as well as human-understandable representation of the mathematical knowledge they contain and the domain-specific knowledge they are based on. We propose to use theory graphs to formalize mathematical models and model pathway diagrams to visualize them. We illustrate our approach by application to the van Roosbroeck system describing the carrier transport in semiconductors by drift and diffusion. We introduce an approach for the block-based composition of models from simpler components.

1 Introduction

The numerical simulation of optoelectronic devices has become a very successful research area involving different disciplines like physics, mathematics and electrical engineering. This is proven by the availability of mature methods and software, a growing number of publications and a broad range of applications (Piprek, 2017a). However, there are growing concerns how the quality of the results submitted for peer-review can be guaranteed. To address this problem in the field of optoelectronic device simulation, e.g. Piprek (2017b) suggested guidelines for simulation papers that focus on improving the reproducibility of presented research findings. In more general terms, this can been seen as part of the ”reproducibility crisis” in science as discussed, e.g., in Stodden et al (2013). In a survey (Baker, 2016), more than 80% of the scientists blame the unavailability of methods and code as a factor for irreproducible research. The NUSOD 2017 special issue in the journal *Optical and Organic Electronics* accounts for this by implementing general quality guidelines[1] for its contributions.

In general, numerical simulations of optoelectronic devices are characterized by (possibly huge amounts of) data and software used for its generation. In order to ensure reproducibility as well as re-usability of the scientific results appropriate options for their storage and long-term accessibility of the involved research data are required, e.g. to adopt the above mentioned guidelines.

The *numerical data* is generally recognized as research data that should be findable, accessible, interoperable and reusable (FAIR), cf. Wilkinson et al (2016). This stimulated the setup of data repositories and related information services such as DataCite (Brase, 2009), or RADAR (Razum et al, 2014). However, the reproducibility of the scientific results requires the corresponding *software* to be available. Hence, software is increasingly recognized as research data by scientific communities and funding agencies, so information services for mathematical software such as swMath[2] (Greuel and Sperber, 2014) emerge. Accompanying principles for data and software citation are developed, e.g., by the FORCE11 initiative[3] (Bourne et al, 2012).

[1] See `http://www.nusod.org/2017/conf_oqe.html`
[2] See `http://www.swmath.org`
[3] See `https://www.force11.org`

DOI 10.20347/WIAS.PREPRINT.2431 Berlin 2017

The aim of this paper is to discuss a holistic concept for research data in the simulation of optoelectronic devices or more general in mathematical modeling and simulation (MMS) by including mathematical models further addressing the problem of reproducibility. Moreover, we will present representations for mathematical models that enable their handling as research data and part of the digital science practice.

2 Mathematical models as research data

Numerical data and the corresponding software alone are not enough to fully characterize the research data that has been utilized to achieve the scientific results: they can only be correctly interpreted and used if the corresponding *mathematical models* are explicitly linked to both. Therefore Koprucki and Tabelow (2016) and Koprucki et al (2016) proposed to categorize mathematical models as the third pillar of research data in MMS beside numerical data and software.

However, finding an appropriate representation for models is far less obvious than for numerical data and software. The current practice is a mixture of mathematical formulae and natural language in scientific publications. This (rigorous, but) informal approach creates ambiguity, potential incompleteness of the presentation, less reproducibility and often “re-invention of the wheel”.

In particular, this representation is not suited for the creation of a “model repository” in analogy to those for data and software. To remedy this Kohlhase et al (2017) proposed a new machine-actionable, but human-understandable representation of mathematical models based on **Model Pathway Diagrams** (MPD). MPDs specify the physical quantities that are described in the model as well as the relations between them (laws, constitutive equations) and informal documentation about them. These specifications are represented in a special machine-readable description language, namely OMDoc/MMT (Kohlhase, 2006; Rabe and Kohlhase, 2013).

3 Model Pathway Diagrams

Kohlhase et al (2017) introduced MPDs as a diagrammatic representation of mathematical models that captures their inner (physical) structure. In an MPD the physical quantities are depicted as circles with their physical notations as labels connected by the physical laws in rectangles labeled with the respective equations.

As an example, we consider the stationary van Roosbroeck model describing the semi-classical transport of electrons and holes in a self-consistent electric field using a drift-diffusion approximation (Selberherr, 1984). The van Roosbroeck model is the standard model to describe the current flow in semiconductor devices and is widely employed for the numerical simulation of optoelectronic devices covering LEDs, lasers and solar cells.

The MPD of a unipolar version of the van Roosbroeck model is shown in Fig. 1, see also Kohlhase et al (2017). In this MPD we can directly get an overview over the structure of the model. We observe the (nonlinear) Poisson equation complex on the top and the carrier transport complex at the bottom. In the Poisson equation we see a sub-MPD for the displacement field describing isotropic materials.

In the context of perturbation theory in quantum field theory the usefulness of diagrammatic representations, notably Feynman diagrams, for complex physical phenomena is well-established. We hope that MPDs can similarly provide an easy access to the inner structure of complex mathematical mod-

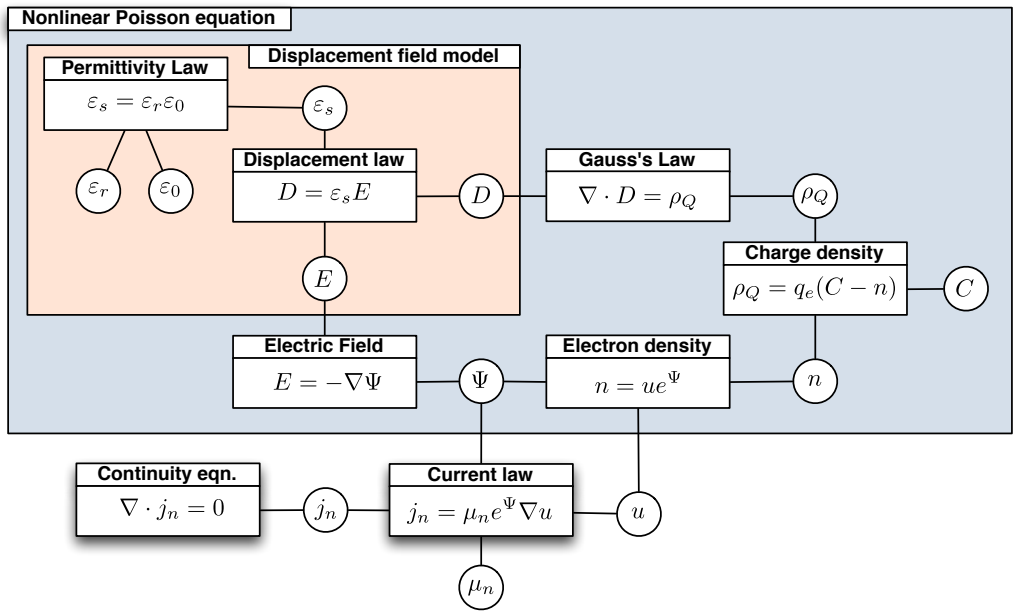


Figure 1: Model Pathway Diagram (MPD) of the unipolar van Roosbroeck system using Slotboom variables. One observes the nonlinear Poisson equation complex on the top (highlighted with blue color) and the carrier transport complex below. The loop structure of the nonlinear Poisson equation reflects the self-consistency of the electric field, whereas the transport complex reveals a tree-like structure. (Material) parameters of the van Roosbroeck model appear as leafs of the MPD such as the doping profile C, the relative permittivity ε_r or the electron mobility μ_n. For simplicity the representation of the boundary conditions have been omitted as well as some quantities, e.g. the temperature T.

els because their topology elegantly captures important properties of the subsystems. For instance, the loop of the nonlinear Poisson equation in Fig. 1 reflects the self-consistency of the electric field.

4 Block-based model composition and iterative methods

Complex models can be understood as a coupled system of sub-models. For the van Roosbroeck system these are for example the Poisson equation for the electrostatic potential and the drift-diffusion equation describing the carrier transport. In this case the coupling between the sub-models is defined by two quantities appearing in both: the electrostatic potential Ψ and the charge density represented by the Slotboom variable u. This property can be used to construct the van Roosbroeck system by a *block-based composition* of two sub-models as illustrated by the MPD in Figure 2: the upper part of the Figure depicts the nonlinear Poisson block, the lower part the drift-diffusion block. In each model block, the coupling variables are represented as input-output ports. The blocks are now linked by the identification of the corresponding quantities indicated by ”double bonds”. A flattening of this MPD can be performed by the elimination of the double bonds which results in the MPD in Figure 1.

Moreover, using the block-based representation one can explain an important concept for the solution of the fully coupled model, namely the *Gummel mapping*. There one solves the nonlinear Poisson equation starting with an initial guess for the density $u = u_0$. The resulting electrostatic potential Ψ is then inserted into the drift-diffusion equation, from which an updated value for u is calculated. This cycle is indicated in Figure 2 by red arrows and defines the Gummel mapping $G : u \to u$. Using G one can consider the solution of the van Roosbroeck system a fixed point problem (Jerome,

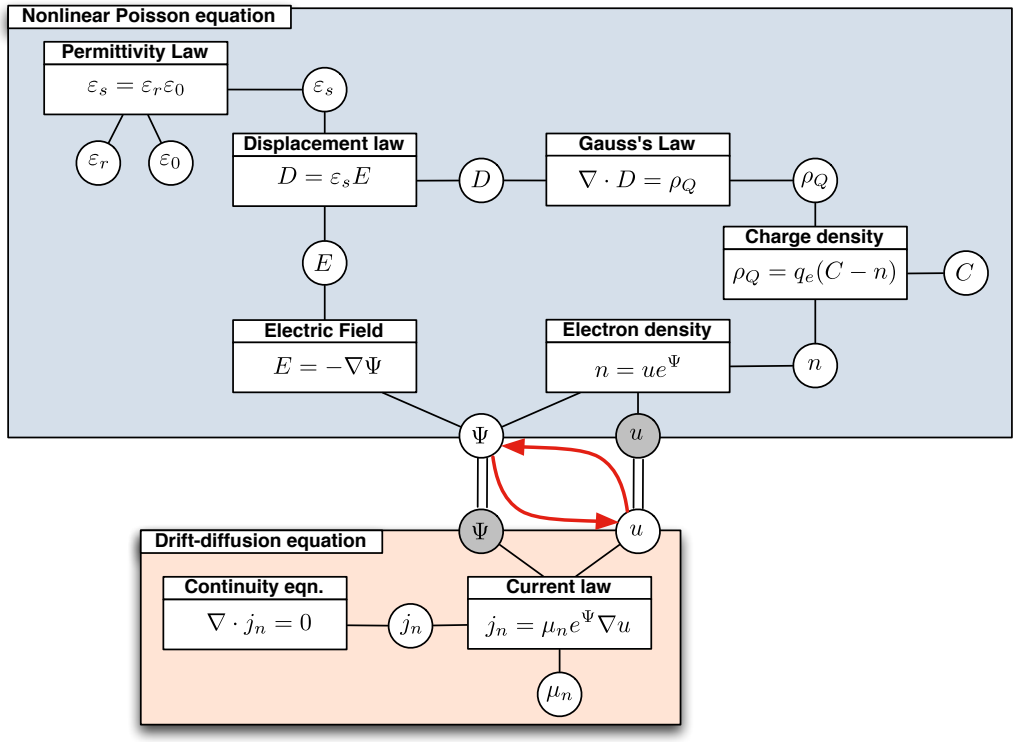


Figure 2: Block-based composition of the van Roosbroeck system by interlinking the model blocks for the nonlinear Poisson equation and the drift-diffusion equation. The coupling between model blocks is represented by ”double bonds” connecting the corresponding input/output ports: the electrostatic potentials Ψ and the carrier density given by the Slotboom variable u.

1996). This leads to an iterative method for the numerical solution of this fixed point problem, known as Gummel's (decoupling) method (Selberherr, 1984). In terms of the corresponding MPD (Figure 2), the nonlinear Poisson block is then replaced by a linearized version. Our experience has shown that such visualizations, e.g., of the Gummel mapping, are a valuable tool for the scholarly communication of the solution strategies of coupled problems.

The block-based composition seems to be well-suited for the representation of complex models like electro-optical models for lasers coupling carrier transport of electrons and holes with optical fields or electro-thermal models for self-heating effects in organic LEDs. Furthermore, we expect that multi-scale modeling and simulation approaches from the atomistic to device level such as considered within the EU funded project DEEPEN[4], can be represented by a hierarchy of interlinked MPD blocks.

Further research will focus on the understanding of numerical schemes and discretizations as MPDs providing the connection between the continuum formulation of a model (partial differential equations) on the one hand and the software implementing its solution of the other hand. In particular, we will study the representation of the finite volume discretization in combination with the Scharfetter-Gummel scheme for the drift-diffusion equation as a guiding example.

[4]See `http://www.nmp-deepen.eu`

5 Formalization in OMDoc/MMT

MPDs were presented as visual aid for understanding structural properties of models and algorithms for numerical simulations. Indeed, these properties are directly induced by the structure of the underlying mathematical knowledge. This shows that the MPDs represent an inherent quality of the models.

For a formalized representation of mathematical models and their underlying physics, we use the OMDoc/MMT format (see Kohlhase (2006); Rabe and Kohlhase (2013) for details) that allows specifying the mathematical vocabulary of models, i.e. the concepts, their properties, and relations. Like an object-oriented programming language, OMDoc/MMT allows organizing knowledge into small, reusable modules, called **theories**, that inherit functionality from each other. There are various inheritance mechanisms between theories available in OMDoc/MMT, the most important for us here are **inclusions**, see below. Unlike a programming language, OMDoc/MMT also formalizes conceptual dependencies and properties (e.g., physical laws in our case).

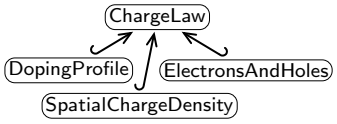
```
theory ChargeLaw : top:?Base =
  include ?DeviceGeometry
  include ?DopingProfile
  include ?PhysicalConstants
  include ?ElectronsAndHoles
  include ?SpatialChargeDensity

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

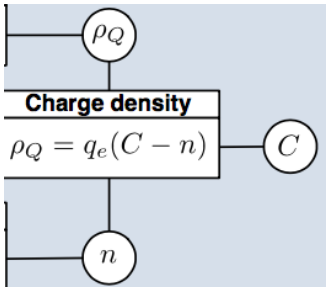
Figure 3: Excerpt from the OMDoc/MMT Formalization (Total Charge Density Law)

As an example Figure 3 shows the OMDoc/MMT code for the formalization of the charge law of the MPD from Figure 1 in a bipolar version. Please note, that the usual expression for the total charge density $Q(x) = q_e \cdot \big(C(x) + p(x) - n(x)\big)$ with the electron density n and the hole density p has been rewritten to $Q(x) = q_e \cdot \big(C(x) + z_h \cdot \rho_h(x) + z_e \cdot \rho_e(x)\big)$ by using generic expressions for the carrier densities involving densities ρ_i and charge numbers z_i for each species ($z_e = -1$ for electrons and $z_h = 1$ for holes). The theory ChargeLaw inherits from various other theories:

1 DeviceGeometry defines the spatial domain Ω with $x \in \Omega$,
2 SpatialChargeDensity defines the charge density $Q(x)$,
3 DopingProfile defines the doping profile $C(x)$,
4 PhysicalConstants defines the elementary charge q_e, and
5 ElectronsAndHoles defines two instances electrons and holes of a theory Species, which defines the charge number z and the density $\rho(x)$.



These five *theories* provide the quantities needed to formally state the charge density law via *inclusions*. For example, electrons/density x is the formal expression corresponding to $\rho_e(x)$ – the value at x of the density of the instance of Species called electrons. The introduction of a generic carrier Species in the formalization allows to express that electrons and holes share structural properties and components.



The theory graph corresponding to the OMDoc/MMT snippet in Figure 3 is shown on the right above omitting the theories DeviceGeometry and PhysicalConstants. Note that it is structurally equivalent to the corresponding part of the MPD in Figure 1.

We make use of this correspondence in an experimental extension of TGView (Rupprecht et al, 2017) that generates MPDs automatically from OMDoc/MMT representations. Figure 4 shows such a generated MPD, which is – up to layout – equivalent to the manually crafted MPD in Figure 1. All nodes and edges are clickable and produce interactive HTML5 renderings of the underlying formalizations. Note that the graph layout in Figure 4 is automatically computed by TGView and can be hand-tweaked in the system for more intuitive results.

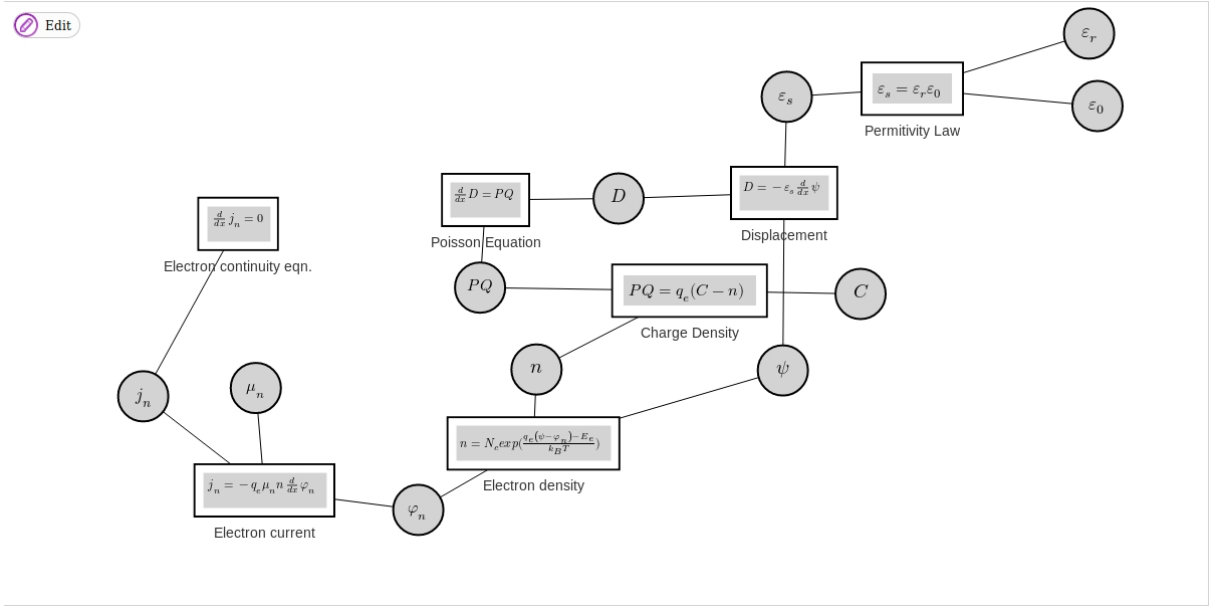


Figure 4: Screenshot of the TGView displaying the van Roosbroeck Theory Graph, cf. Figure 1

The block-based model composition detailed in Section 4 fits very well with that of OMDoc/MMT: MPD blocks correspond to connected theory-subgraphs and coupling via quantity theories is just multiple inheritance via inclusions. The added functionality of the OMDoc/MMT formalization is that internal consistency and coupling constraints that remain implicit in both the MPDs and the published papers can be checked by the MMT system (Rabe, 2013), which implements OMDoc/MMT.

With this system support, we can handle theory graphs with tens of thousands of theories, which equates to thousands of model components. Note that due to the modular structure of OMDoc/MMT, even competing models can co-exist in a single theory graph and thus share structural components. This suggests that we can formalize and curate more and more MPD graphs, share the underlying physics and mathematics, use the OMDoc/MMT tools to manage – i.e. develop, check, and display – them. We started a collection effort for MPDs on MPD Hub[5].

We are also extending the MMT system with functionality to directly represent MPD structures for theory graphs containing models and to directly implement MPD-specific structures (e.g., model blocks), management processes (e.g., block coupling or solution strategies). We anticipate that this will make the interaction with large MPD graphs much more powerful and intuitive.

[5]MPD Hub `https://github.com/WIAS-BERLIN/MPDHub/wiki/`

6 Conclusion

We presented a diagrammatic representation of mathematical models based on MPDs together with a formalization based on OMDoc/MMT. MPDs provide a visual tool for understanding the structural properties of models as well as algorithms for numerical simulations. We discussed the block-based composition of models from sub-models which seems to be well-suited for the representation of complex multi-physics models or multi-scale models.

The formal representation of mathematical models in OMDoc/MMT as outlined in the Section 5 makes them machine-actionable. This enables the unique identification, the automatic derivation of relationships between them, and the modular creation of new models from existing ones, and semantic services for them. As a first concrete application, we have created a visualization tool that creates interactive HTML-based layouts for MPD from OMDoc/MMT formalizations.

Additionally, OMDoc/MMT have a well-established interface to a semantically enhanced Version of TEX. This can be used to write semantically enhanced articles and papers that directly refer to the database of mathematical models. From these we can generate interactive HTML documents, which are interlinked with the MPD formalizations and can be instrumented with semantic services like the ones above.

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 Wolfram Sperber (general math background).

References

Baker M (2016) Is there a reproducibility crisis? Nature 533(7604):452–455

Bourne PE, Clark TW, Dale R, de Waard A, Herman I, Hovy EH, Shotton D (2012) Improving The Future of Research Communications and e-Scholarship (Dagstuhl Perspectives Workshop 11331). Dagstuhl Manifestos 1(1):41–60, DOI 10.4230/DagMan.1.1.41

Brase J (2009) Datacite – A global registration agency for research data. In: Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO'09. Fourth International Conference on, IEEE, pp 257–261

Greuel GM, Sperber W (2014) swmath – An Information Service for Mathematical Software. In: Hong H, Yap C (eds) Mathematical Software – ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 691–701

Jerome JW (1996) Analysis of Charge Transport. A Mathematical Study of Semiconductor Devices. Springer, Berlin Heidelberg

Kohlhase M (2006) OMDoc – An open markup format for mathematical documents [Version 1.2]. No. 4180 in LNAI, Springer Verlag, URL http://omdoc.org/pubs/omdoc1.2.pdf

Kohlhase M, Koprucki T, Müller D, Tabelow K (2017) Mathematical models as research data via flexiformal theory graphs. In: Geuvers H, England M, Hasan O, Rabe F, Teschke O (eds) Intelligent Computer Mathematics: 10th International Conference, CICM 2017, Edinburgh, UK, July 17-21, 2017, Proceedings, Lecture Notes in Artificial Intelligence, vol 10383, Springer International Publishing, Cham, pp 224–238, DOI 10.1007/978-3-319-62075-6_16

Koprucki T, Tabelow K (2016) Mathematical models: A research data category? In: Greuel GM, Koch T, Paule P, Sommese A (eds) Mathematical Software – ICMS 2016: 5th International Conference, Berlin, Germany, July 11-14, 2016, Proceedings, Lecture Notes in Computer Science, vol 9725, Springer International Publishing, Cham, pp 423–428, DOI 10.1007/978-3-319-42432-3_53

Koprucki T, Tabelow K, Kleinod I (2016) Mathematical research data. PAMM 16(1):959–960, DOI 10.1002/pamm.201610458, URL http://dx.doi.org/10.1002/pamm.201610458

Piprek J (ed) (2017a) Handbook of Optoelectronic Device Modeling and Simulation. Series in Optics and Optoelectronics, CRC Press, Taylor & Francis

Piprek J (2017b) NUSOD Blog: How to get your simulation paper accepted. https://nusod.wordpress.com/2017/01/04/how-to-get-your-simulation-paper-accepted/, accessed 2017-03-30

Rabe F (2013) The MMT API: A Generic MKM System. In: Carette J, Aspinall D, Lange C, Sojka P, Windsteiger W (eds) Intelligent Computer Mathematics, Springer, no. 7961 in Lecture Notes in Computer Science, pp 339–343, DOI 10.1007/978-3-642-39320-4

Rabe F, Kohlhase M (2013) A scalable module system. Information & Computation 230:1–54, URL http://kwarc.info/frabe/Research/mmt.pdf

Razum M, Neumann J, Hahn M (2014) Radar–Ein Forschungsdaten-Repositorium als Dienstleistung für die Wissenschaft. Zeitschrift für Bibliothekswesen und Bibliographie 61(1):18–27

Rupprecht M, Kohlhase M, Müller D (2017) A flexible, interactive theory-graph viewer. In: Kohlhase A, Pollanen M (eds) MathUI 2017: The 12th Workshop on Mathematical User Interfaces, URL http://kwarc.info/kohlhase/papers/mathui17-tgview.pdf

Selberherr S (1984) Analysis and simulation of semiconductor devices. Springer, Wien, New York

Stodden V, Borwein J, Bailey D (2013) Setting the default to reproducible: Reproducibility in computational and experimental mathematics. SIAM News 46:4–6

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al (2016) The FAIR guiding principles for scientific data management and stewardship. Scientific data 3(160018)