On Industrial Strength Bio-design Automation

Wolf-Ekkehard Matzke[1(✉)], Chris P. Barnes[2], Eyck Jentzsch[1], Thorsten Mascher[3], and Michael Stumpf[4]

[1] MINRES Technologies GmbH,
Keltenhof 2, 85579 Neubiberg, Germany
{wolf,eyck}@minres.com
[2] University College London, Gower Street,
London WC1E 6BT, UK
christopher.barnes@ucl.ac.uk
[3] Ludwig-Maximilians-University Munich,
Grosshaderner Str. 2-4, 82152 Planegg-Martinsried, Germany
mascher@bio.lmu.de
[4] Theoretical Systems Biology, Imperial College London,
317 Sir Ernst Chain Building, London SW7 2AZ, UK
m.stumpf@imperial.ac.uk

Abstract. Bio-Design Automation (BDA) denotes the nascent domain-specific Information and Communication Technology (ICT) discipline for synthetic biology, which constitutes the core technology of the Knowledge-Based Bio-Economy (KBBE). Ultimately, the success or failure of synthetic biology and the emerging KBBE equates to the progress or lack of progress in establishing an industrial strength BDA discipline. In this paper, we seek answers to the question "What does it take for BDA to become an industrial strength discipline?" Our goal is to stimulate a broad community discussion including Business Managers, Computer Scientists, ICT professionals, Synthetic Biologists, etc. around this question. To jump-start the debate, we will provide four core hypotheses covering what we believe are the most important aspects to be considered. Given that industrial strength is a composite aggregate of several technical and managerial variables, we have chosen to take a holistic approach and not restrict ourselves a priori to any particular viewpoints. Last, but not least, we will apply our findings and provide a prototypical industrial implementation of a BDA platform.

Keywords: Bio-design Automation · Bio-Design System · Business models · Information and Communication Technology · Synthetic Biology

1 Introduction

The keynote given to the ICTERI 2013 conference at Kherson, Ukraine, [26], highlighted the defining role of Information and Communication Technology (ICT), Biotechnology and Synthetic Biology for the emerging Knowledge-Based Bio-Economy (KBBE). Bio-design Automation (BDA) was identified as the key domain-specific ICT for Synthetic Biology (and the KBBE). The paper at hand goes one step further from the keynote by providing a discussion framework for the industrial implementation of

V. Ermolayev et al. (Eds.): ICTERI 2014, CCIS 469, pp. 277–299, 2014.
DOI: 10.1007/978-3-319-13206-8_14

BDA that can perhaps serve as a guide for BDA practitioners and researchers. Some initial thoughts on industrial strength bio-design automation were presented at 5th International Workshop on Bio-Design Automation (IWBDA 2013) [28].

To shortly recapitulate: The KBBE aims at the "sustainable production and conversion of biomass, for a range of food, health, fiber and industrial products and energy", where "renewable biomass encompasses any biological material to be used as raw material."[1] Biotechnology is defined by the Organization for Economic Co-operation and Development (OECD) as "the application of science and technology to living organisms, as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services" [41]. Synthetic biology is defined as:

"(A) the design and construction of new biological parts, devices, and systems; and
(B) the re-design of existing, natural biological systems for useful purposes"[2].

ICT is the Economy-Defining Technology (EDT) of the Knowledge-Based Economy (KBE) and it will continue to be an EDT in the KBBE. This statement is undoubtedly true in terms of general purpose ICTs. Every year, Gartner, Inc.[3], a world-renowned information and technology research and advisory company, compiles a list of the top ten strategic general purpose ICT trends for the following year, thus providing an annual fresh perspective. For 2014 Gartner lists the following top ten ICTs[4]: (1) Mobile Device Diversity and Management; (2) Mobile Apps and Applications; (3) The Internet of Everything; (4) Hybrid Cloud and IT as Service Broker; (5) Cloud/Client Architecture; (6) The Era of Personal Cloud; (7) Software Defined Anything; (8) Web-Scale IT; (9) Smart Machines; (10) 3-D Printing. For preceding years Gartner included for example areas such as Strategic Big Data (2013)[5], Next-Generation Analytics (2012)[6] and Cloud Computing (2011)[7]. It is perhaps easy to imagine the application of these general purpose ICTs in the context of the KBBE as the advancement of the KBE. Clearly, the KBBE will benefit from extensive deployment and use of general purpose ICTs. However, the KBBE and in particular synthetic biology require a sophisticated domain-specific ICT solution. Contrary to general-purpose ICT, domain-specific ICT is created to solve problems within a particular area of concern. In the case at hand, the area of concern is synthetic biology. As noted above, the associated domain-specific ICT is comprised under the label BDA. Note that, unlike general purpose ICTs, domain-specific ICTs usually don't receive the public awareness - even among ICT professionals - they merit. For example: Most likely many ICT professionals are highly interested and knowledgeable in general purpose ICTs such as Mobile Apps and

[1] THE EUROPEAN BIOECONOMY IN 2030. Delivering Sustainable Growth by addressing the Grand Societal Challenges: http://www.epsoweb.org/file/560.

[2] http://syntheticbiology.org/

[3] http://www.gartner.com/technology/home.jsp

[4] http://www.gartner.com/newsroom/id/2603623

[5] http://www.gartner.com/newsroom/id/2209615

[6] http://www.gartner.com/newsroom/id/1826214

[7] http://www.gartner.com/newsroom/id/1454221

Applications, but only very few might have ever heard about Electronic Design Automation (EDA). EDA is the domain-specific ICT for electronics, which is the ICT enabling the design and development of electronic systems. What if there would be no EDA? There would be no electronics and no mobile apps and applications. It's as simple as this. Last, but not least, it should be mentioned that domain specific ICTs represent really fascinating research and engineering fields, provided one manages to overcome the domain-specific barrier to entry.

Exploitation of the industrial potential of synthetic biology has only recently begun. Currently, we witness the progression of synthetic biology from an emerging technology to an emerging industry. However, it will take many years for synthetic biology to evolve into a fully established industry. To which extent and when these transitions will be completed depends on a multitude of challenges and constraints facing synthetic biology. The siliconization of synthetic biology, that is the degree to which so-called *in silico*[8]-based design penetrates synthetic biology, is amongst the fundamental technical challenges to be addressed. Siliconization is driven by the necessity to increase the productivity of design, which is in turn a requirement for the successful industrialization of synthetic biology. Eventually, *in silico* design has to become the principal design approach of synthetic biology.

Siliconization of synthetic biology is the realm of the nascent field of BDA. It should be noted, however, that the application of *in silico* approaches in biology is by no means a new idea. Bioinformatics and computational biology are well-established *in silico* disciplines. In fact, both are cornerstones of BDA. Bioinformatics is a data-centric discipline. It focuses on the application of computational techniques to understand and organize biological data [23]. Computational biology on the other hand is concerned with computational models of biological phenomena [20]. It is therefore a model/hypotheses-centric discipline. The third cornerstone of BDA is formed by the broad spectrum of established (non-bio) design automation approaches, such as EDA, Architecture, Engineering and Construction (AEC), Mechanical Computer Aided Design (MCAD) etc. Note that, in general, design automations are the disciplines (and domain-specific ICTs) devoted to computerized design processes [7]. Like all domain-specific ICTs, design automations are design productivity-centric.

The positioning of BDA relative to bioinformatics, computational biology and (non-bio) design automation is depicted in Fig. 1. BDA builds on these three established *in silico* disciplines as cornerstones and reuses and integrates the underlying technologies whenever feasible and appropriate. Additionally, BDA provides its own set of solutions to address unique challenges of synthetic biology design.

BDA's ultimate promise is to increase the productivity of synthetic biology. To fulfill this promise, BDA needs to develop industrial strength solutions. Industrial strength is defined as a system's ability to work capably and dependably in an operational setting [36].

---

[8] *In silico* is a term popular among synthetic biologists. Wikipedia explains (see http://en.wikipedia.org/wiki/In_silico): "*In silico* is an expression used to mean 'performed on computer or via computer simulation.' The phrase was coined in 1989 as an analogy to the Latin phrases *in vivo*, *in vitro*, and *in situ*, which are commonly used in biology … and refer to experiments done in living organisms, outside of living organisms, and where they are found in nature, respectively."

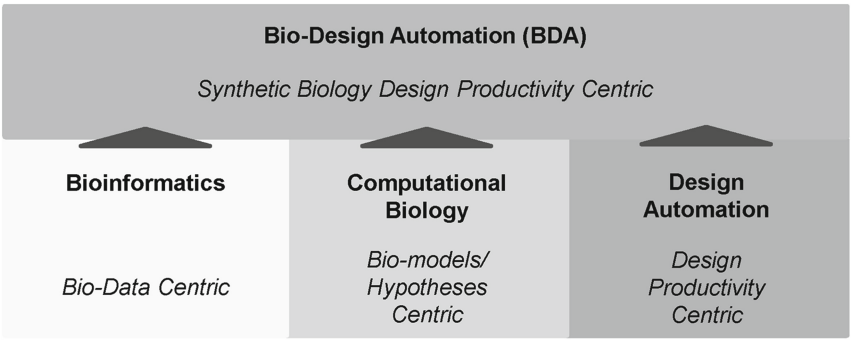


Fig. 1. Positioning of BDA. The figure details the position of *BDA* relative to *bioinformatics*, *computational biology*, and (non-bio) *design automation*. *BDA* reuses and integrates their underlying technologies whenever feasible and appropriate. Additionally, *BDA* provides its own set of solutions unique to synthetic biology design.

In the following we will present what we believe are the four most important aspects to reach industrial strength for BDA. We arrived at these four core hypotheses by examining the BDA industrial strength challenge from the perspectives of various engineering and management disciplines, namely engineering management [37], design theory [17], complexity engineering [6] and, business management theory [43]. In particular, the engineering management perspective led to The Bio-Design System Paradigm hypothesis. The design theory viewpoint revealed The Rational Design Fallacy conjecture. Taking a complexity engineering standpoint resulted in the Uncertainty Rules supposition. Utilizing a business management theory frame of reference led to the It's The Business Model (, Stupid)[9] hypothesis. Given the goal of

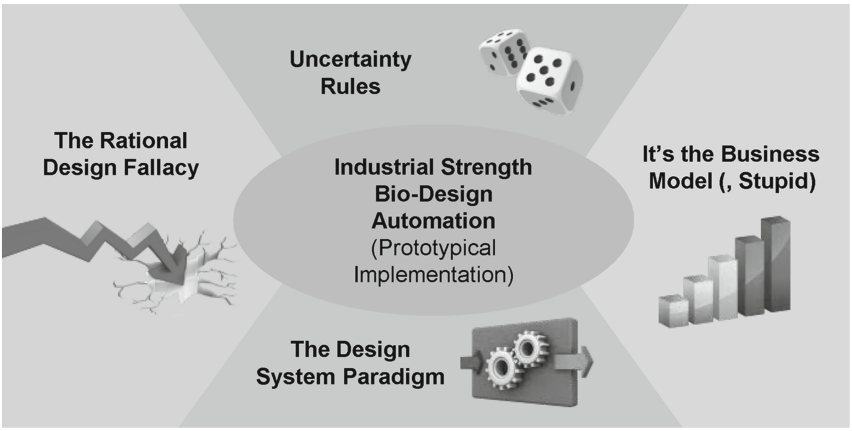


Fig. 2. Graphical abstract for the Table of Contents of this paper

[9] The labeling of the business related core hypothesis as The Business Model (, Stupid) hypothesis will be explained in Sect. 5 of this paper.

stimulating a community discussion, we will keep the following discussion at the conceptual level, the rational or the basics for what it takes to create an industrial strength BDA industry. Getting the concepts right for an emerging discipline is difficult, but also absolutely essential. Note that the authors come from four different perspectives: general purpose ICT, BDA industry, BDA research, and synthetic biology. The aim of the conceptual discussion is to integrate these different perspectives into a single framework for industrial strength BDA. For practical purposes this framework has to be translated into concrete methodological and implementation agendas. Therefore, we will present a prototypical implementation of a BDA platform in the final section of this paper.

Figure 2 provides a graphical representation of the table of contents of this paper.

2 The Bio-design System Paradigm

Let us define a bio-design system as the totality of resources required to perform synthetic biology design processes, with the goal to transform bio-design specifications into bio-production designs. Note that, in general, design systems are holonic systems [13]. Holonic systems are composed of distinct entities (called holons) which interact with each other in order to achieve global system goals. Self-organization is a key property of holonic systems. This allows a holonic system to evolve over time to optimize the achievement of its system goal. Let us recall that a paradigm is a general perspective or way of thinking that reflects the fundamental beliefs and assumptions about the nature (principles) of the subject under consideration [16]. The Bio-Design System Paradigm hypothesis is then the idea that a bio-design system is a holonic system that represents the frame of reference for all synthetic biology engineering activities. Thinking in bio-design systems terms is perhaps the key to industrial strength BDA. In fact, experiences – both good and bad - from established design automation industries (and beyond) underpin the imperative to take a system approach when developing design systems. Indeed, out of our four hypotheses, the Bio-Design Paradigm hypothesis might be the most important one.

Present-day synthetic biology is still largely devoted to solve fundamental (and persistent) problems of engineering biological artifacts [14]. This might lead to the impression that The Bio-Design System Paradigm hypothesis may be far beyond any practical relevance. But perhaps the opposite is true. In fact, the potential benefits of applying this paradigm can be numerous. To discuss this in more detail it is necessary to refine the definition of a bio-design system. As Martin and Odell stated [25]: “If we wish to build systems that perform correctly and consistently, we will want to be clear, concise, and unambiguous in our specifications” and concepts.

Figure 3 proposes a hierarchical definition of a bio-design system. At the highest level the system is treated as a whole, accepting design specifications and producing production designs (forward engineering) or vice versa (backward/reverse engineering). The next level is comprised of the design environment and synthetic biology engineering. Synthetic biology engineering denotes the team of synthetic biologists that perform the design processes. Synthetic biologists are human resources and are therefore carriers of highly valuable tacit knowledge. Tacit knowledge is defined by

Polanyi as knowledge that cannot be expressed explicitly in (whatever) codified form – "We know more then we can tell" [33]. It is work-related practical knowledge acquired informally through experience on the job [42]. Tacit knowledge can be passed between engineers "by personal contact but cannot be, or have not been, set out or passed on in formulae, diagrams, or verbal descriptions and instructions for action" [9]. Note that tacit knowledge constitutes often a company's decisive competitive advantage. Consequently, a bio-design system is not a purely technical system, but a socio-technical system where human, organizational and technical factors closely interact and evolve. Therefore the role of BDA is not solely one of providing automation tools. The BDA challenges faced in the context of bio-design systems are both technical and social in nature. In fact this holds true for any domain-specific ICT. This is important to acknowledge and understand.

Back to the bio-design system hierarchy, the next level down in the bio-design system hierarchy details the composition of the design environment. The function of the design environment is to enable and support the design of biological artifacts. It is composed of the design infrastructure and design environment engineering. The design infrastructure encompasses the *in silico* and the wet lab function. Note that the design infrastructure embodies the explicit knowledge of the bio-design system. This knowledge represents the antithesis to tacit knowledge. Explicit knowledge is highly codified and is easily transmittable [29]. It comes in a wide variety of codifications such as books, documents, policy manuals, standards, intellectual property, databases,

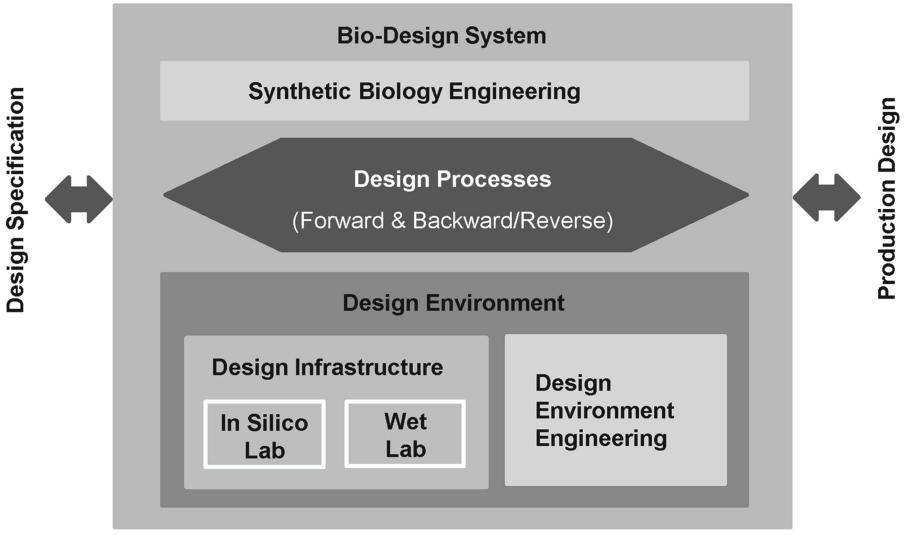


Fig. 3. The *bio-design system* hierarchy. *Synthetic biology engineering* and *design environment engineering* represent human resources who are the carriers of the *bio-design system*'s tacit knowledge. The *design infrastructure* represents the explicit/codified knowledge of the *bio-design system*. *Design processes* are supported bi-directionally (forward and backward/reverse). BDA technology is localized in the *in silico lab*. BDA engineering is part of *design environment engineering*.

software tools, wet lab equipment etc. Similarly like synthetic biology engineering, design environment engineering represents human resources that act to as carriers of tacit knowledge. All the synthetic biology engineering tacit knowledge remarks apply without limitation. Design environment engineering fulfills two major functions: first, it supports synthetic biology engineering in the application of the design infrastructure; second, it codifies its own and synthetic biology engineering's tacit knowledge into explicit concepts and therefore creates the basis for the future evolution of the bio-design system.

Based on the above we are ready to discuss some potential high-level practical benefits of applying The Bio-Design System Paradigm hypothesis:

(a) The Tower-of-Babel problem denotes the potential communication problem between experts that come from different disciplines and speak their own domain-specific languages. This impedes the process of establishing a shared understanding among these experts and hampers collaboration. Cross functional disciplines, such as synthetic biology and BDA, are especially prone to this problem. And indeed, as of now it seems that the BDA community started to build its own Tower-of-Babel. For example there is already a competing set of terms for BDA, namely bioCAD [18], BioCAD [4], Genetic Design Automation [19], etc. - more to come? This might be not an issue for the BDA community, but it is clearly a way to confuse synthetic biologists (as experienced by some of the authors multiple times). The Bio-Design System Paradigm provides the foundation to develop a shared, agreed upon and committed vocabulary that can be used in a coherent and consistent manner by all stakeholders, a prerequisite for true collaboration. One ongoing effort focusing on developing such a vocabulary is the BioParts Terms initiative[10];

(b) Performance management is a prime management concern in any industrial setting [27]. Note that productivity is the single most important performance measurement. The proposed bio-design system hierarchy provides a unifying framework within which the various resources (incl. and most importantly human resources) and processes of design can be situated and their roles and relations can be explicitly identified. Such a framework is pivotal for any performance management and measurement beyond the strategic level [12];

(c) Transforming tacit into explicit knowledge is key to the advancement of a bio-design system. This process needs to be actively managed. The proposed bio-design system hierarchy allows identifying the flow of knowledge and taking concrete actions to facilitate the transformation process. Although it might be a truism to some people, it needs to be stated that the knowledge transformation requires very tight collaboration between synthetic biology engineering and BDA engineering;

(d) Most of the literature defines BDA in the narrow sense that is purely technical. As such BDA is a technology situated in the *in silico* lab. However, given the central role assigned to be BDA in the process of siliconizing synthetic biology (the BDA mission) this is perhaps a too narrow definition. BDA in the broad sense

[10] http://www.minres.com/wiki/index.php/BioParts_Terms

encompasses technical and human resources. BDA is connected to every element of a bio-design system;

(e) The proposed bio-design system structure can serve as starting point for the development of a BDA roadmap. Such a roadmap is important to achieve industrial strength BDA for two reasons: first, it serves as an agreed-upon requirements document between the synthetic biology engineering community and the BDA community; second, it gives BDA researches and practitioners alike direction where to focus on to develop and deliver high-impact solutions. This will become even more important as the BDA community grows.

Some might consider The Design System Paradigm hypotheses still a "philosophical" issue. However, we believe that ignoring this paradigm will have the effect of making the collaboration between the communities of synthetic biology engineering and BDA engineering much more difficult, no matter how pragmatic it might seem to ignore this "philosophical" issue. It can be expected that the recently formed Bio-Design Automation Consortium (BDAC), a non-profit association which also organizes the IWBDA events, will take a leadership role in addressing and promoting the above benefits of bio-design systems[11].

3 The Rational Design Fallacy

Synthetic biology is founded on the notion of rational design, which implies the following propositions [32]:

(a) Design processes (what to do) and design methods (how to do it) are known explicitly and in detail;
(b) Given a well-defined goal, the design process progresses along the optimal path on the basis of logical principles.

Clearly, neither of these propositions will pass the reality test, which renders rational design a utopian vision. This is not a bad thing in itself, provided it is recognized for what it is: a guiding dream with positive transformative significance. However, if rational design is used as a mental model that directs practical actions [15], then it can turn straight into a fallacy. In fact, all design activities are rationally bounded [38].

The major pitfall for BDA caused by this fallacy is over-automation. The labeling of this pitfall is meant to express the respective BDA experience of synthetic biologists. The over-automation pitfall occurs when the level of abstraction for a design automation is raised beyond the essence horizon for the sake of automation: You cross the essence horizon if your abstractions alienate the substance of the system to be designed. Established classical engineering domains are as a rule relatively self-contained and mature. They offer defined representations of both the design artifact under development and the environment it is supposed to operate in. Therefore, it is possible to build comprehensive domain models that enable and support model-driven design approaches

[11] http://www.iwbdaconf.org/2014/

which proceed more or less smoothly from architectural to physical design. However, the situation in synthetic biology is completely different. Even so it is possible to create (to some extent) design artifacts following classical engineering approaches the overall synthetic biology experience is that there is no guarantee that the design artifact will operate as specified in the target environment. This is because bio-designs exhibit significant context-dependency and low predictability. Perhaps this is a prime challenge for BDA to focus on. Currently, the synthetic biology engineering way to cope with this challenge is to follow a minimum design approach.

Figure 4 details the above discussion visually. Every design process starts with a design specification which describes the initial design space including design metrics and constraints. The design progresses further along defined levels of abstractions where the design space is narrowed by proceeding from a higher level of abstraction to a lower one. Here, the biological system to be designed is a gene-regulatory network (GRN). Note that a GRN is a collection of deoxyribonucleic acid (DNA) segments in a cell which interact with each other indirectly - through their ribonucleic acid (RNA) and protein expression products - and with other substances in the cell[12]. Figure 4(a) shows an exemplary synthetic biology design process. The design proceeds from the design architecture, perhaps captured by a functional block diagram, to the detailed design, perhaps captured by the network topology, to the design implementation in form of abstract gene-regulatory network (AGRNs) to the physical design, represented by physical a GRN, and, finally to the assembly plan. Figure 4(b) captures the rational design (illusion): Rational design progresses straight from the initial design space captured in the design specification to the optimal production design along the optimal design path on the basis of purely logical decisions. That is, the designer is only concerned with the design specification. Everything else is done fully automatically and nontransparent (black-box) to the designer. The final production design would be optimal by definition. Of course, rational design does not pass the reality test. Figure 4 (c) represents the design process based on the bounded rational design methodology: All design activities are rationally bounded and transparent (no automations) or non-transparent (automations) to the designer. Design automations are only introduced when appropriate and their scope is fully described (to ensure that they are essence-preserving). All other activities in the design process are considered to be fraught with uncertainty and ultimately dependent on the tacit knowledge of the synthetic biologists. In general the bounded rational design process is iterative. The final production design will be acceptable (that is as a rule not optimal). Figure 4(d) depicts the concept of essence preserving design automations: A design automation introduced in the design process must comply with the entry essence horizon and the exit essence horizon of the design activity it automates. In practical terms an essence horizon is a validated design space at certain level of abstraction. Validated means that the design space complies with design metrics and constraints along the hierarchy of abstraction down to the final product. This way it is ensured that the essence of the system to be designed is maintained. A design space at a certain level of abstraction might very well fulfill the design metrics and constraints at this level of abstraction. However, if it fails to fulfill

[12] http://en.wikipedia.org/wiki/Gene_regulatory_network

the design metrics and constraints down the hierarchy, it does not constitute an essence horizon. Last, but not least, note that the exit essence horizon of an automation constitutes the entry essence horizon of the next design activity down the abstraction hierarchy. Figure 4(e) visualizes the over-automation pitfall: As explained above this pitfall occurs when the level of abstraction is raised beyond the essence horizon for the sake of automation. The scenario is as follows: Given a bounded rational design process, a design automation is introduced at an abstraction level which is not appropriate for automation. Such a design automation will prevent the synthetic biologist to make necessary engineering decisions based on his or her tacit knowledge. The necessity for engineering decisions is a consequence of having precisely not enough explicit knowledge to justify a design automation. The essence horizon compliance requirement is violated at the entry level and therefore produces results that constitute a

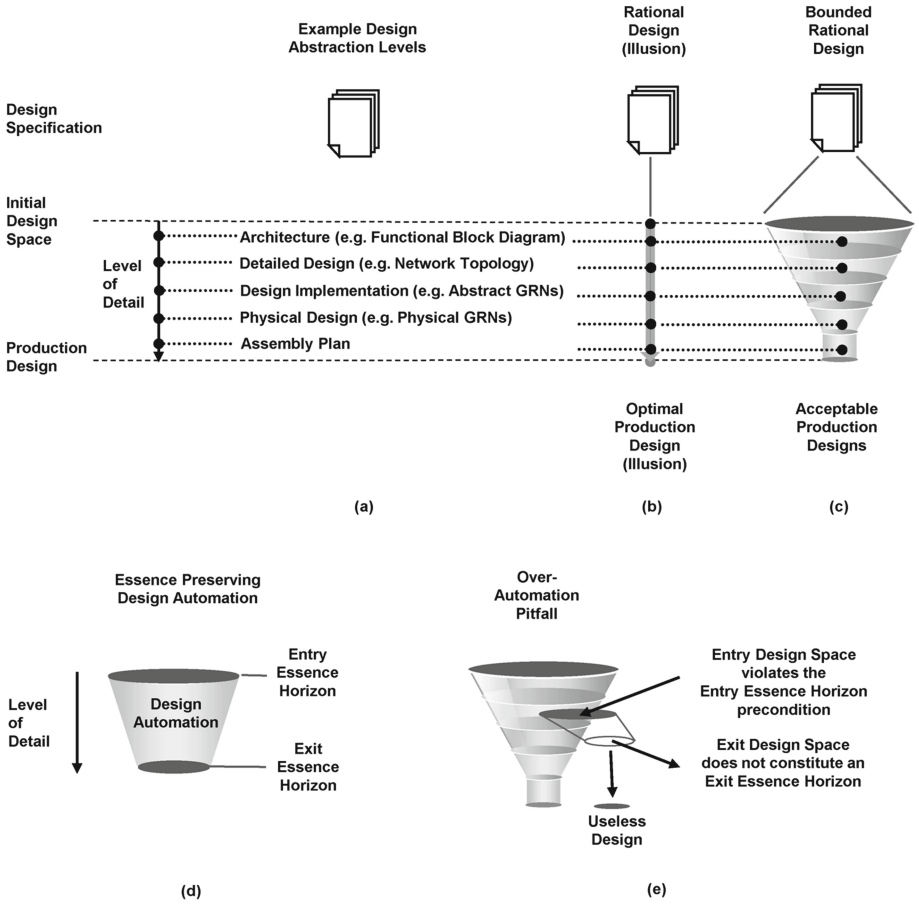


Fig. 4. Bio-design process. The system to be designed is gene regulatory network (GRN). (a) Example design abstraction levels. (b) Rational design illusion. (c) Bounded rational design. (d) Essence preserving design automation. (e) The over-automation pitfall.

design space which is off from the exit essence horizon. As a consequence the resulting "production designs" will be almost guaranteed to be useless.

To avoid the over-automation pitfall and others industrial strength BDA should adhere to the following principles and guidelines:

(a) Focus on the development of solutions for today's synthetic biology problems. Of course, there also need to be investments in longer-term R&D efforts that may not mature for years;
(b) Respect that the (baseline) design processes are established by synthetic biologists. This implies that synthetic biologists define the essence horizons. Indeed, in several instances this might bring some high-flying BDA aspirations down to earth;
(c) Support both forward and reverse engineering. Note, that in today's synthetic biology settings, reverse-engineering dominates;
(d) Given a set of essence horizons support the right levels of remaining computational abstractions along the design process;

Last, but not least, the perhaps most important rule is this:

(e) If your automations do not improve the BDA experience for synthetic biologists', start all over again!

4 Uncertainty Rules

This section emphasizes that BDA technology must be capable to cope with uncertainty. Generally, uncertainty is a second order effect in many classical engineering design settings, and is mostly related to concerns such as product yield, product lifetime, etc. The opposite is true, although often not explicitly acknowledged, for synthetic biology. Quantum physics aside, the world is in principle knowable and deterministic [11]. It follows that uncertainty is due to ignorance of the underlying causes. Ignorance is currently ubiquitous in synthetic biology, and it does not seem that this situation will change in the near future. Not surprisingly, the prevailing sentiment in the world of synthetic biology is that the inherent complexity of the field largely prevents the use of computational design approaches. This position caused a considerable debate between the BDA and the synthetic biology community [10]. Unfortunately, discussants often seem to talk at cross-purposes when they might well be seeing the same problem from different angles.

To arrive at some level of consensus among the communities consider the following:

(a) If we interpret complexity as a function of our ignorance about the reality's work principles, then synthetic biology is devilishly complex;
(b) Complex systems are inherently uncertain, that is stochastic, in nature.

Both statements should be viewed as consensual propositions. The immediate conclusion is that stochasticity is an essential property of biological systems. Indeed, more recent research has shown that stochasticity is a generic feature of many

biological systems, and may, in fact, be actively exploited by ensuring heterogeneity in a population of cells. Understanding the causes – in addition to the effects – of stochasticity is thus pivotal for applications in synthetic biology. Statistical inference provides us with the computational tools necessary to dissect noise and stochasticity in biological systems; such a statistical angle is essential, as we are no longer dealing with a single type of behavior but with a probability distribution over different (though related) types of behavior; this poses severe challenges to e.g. conventional optimization procedures.

Let us consider the design of biochemical reaction network (BRN) according to some stated design objectives. A BRN is defined by (e.g. [31, 40]):

(a) A set of variables which represent the amounts of biochemical species (molecules) in an reactor under consideration;
(b) A set of rules of temporal changes (reaction equations) of these variables.

A design objective might be for example the specified change of these variables over time. Let us further assume we have identified a set of potential designs, but we don't know the one that best meets the design objectives. Let us finally assume we have captured these designs by some appropriate modeling formalism. However, we face an additional challenge. In practice, it will be the rule rather than the exception that we have to cope with parameter uncertainty, often quite considerable (i.e. spanning orders of magnitude). The task is now to identify the design that best meets the design objectives (model selection) under conditions of parameter uncertainty (parameter inference). This requires an inference engine that is capable to handle model selection and parameter inference. Conceptually the inference engine needs to "sit" on top on whatever computational model we apply. Figure 5 depicts the positioning of the inference engine in the design process.

Unquestionably, it is clear that for any industrial strength BDA solution an inference engine is a prerequisite to cope with stochasticity. One superior inference engine (both in terms of conceptual and computational performance) is the Approximate Bayesian Computation – Sequential Monte-Carlo (ABC SMC) algorithm presented in [1]. The algorithm supports both forward- and reverse engineering and is flexible with respect to the underlying computational models. As long as appropriate computational models are available, it is possible to apply the algorithm at any design abstraction level (architecture, detailed design etc.). Perhaps the most important feature is that the algorithm can handle parameter uncertainty and model selection at once.

Figure 6 visualizes a typical design scenario in synthetic biology that is the design of a BRN, as discussed above. In this scenario the synthetic biologist has chosen to enter the design process at the network topology level. Biochemical adaptation is used as an example [24]. This example was investigated thoroughly in one of the authors' research group (see [1, 2]). The inference engine exploits the advantages of Bayesian statistics. The applied Approximate Bayesian Computation Sequential Monte-Carlo (ABC-SMC algorithm) is based on [2]. Parameter inference is accomplished by the efficient exploration of design and high-dimensional parameter spaces. Model selection, Fig. 6(a), is enabled by the ability to rank competing designs with respect to their ability to bring about the desired behavior. The synthetic biologist identified several design options (network realizations) which in principal will be able to implement the

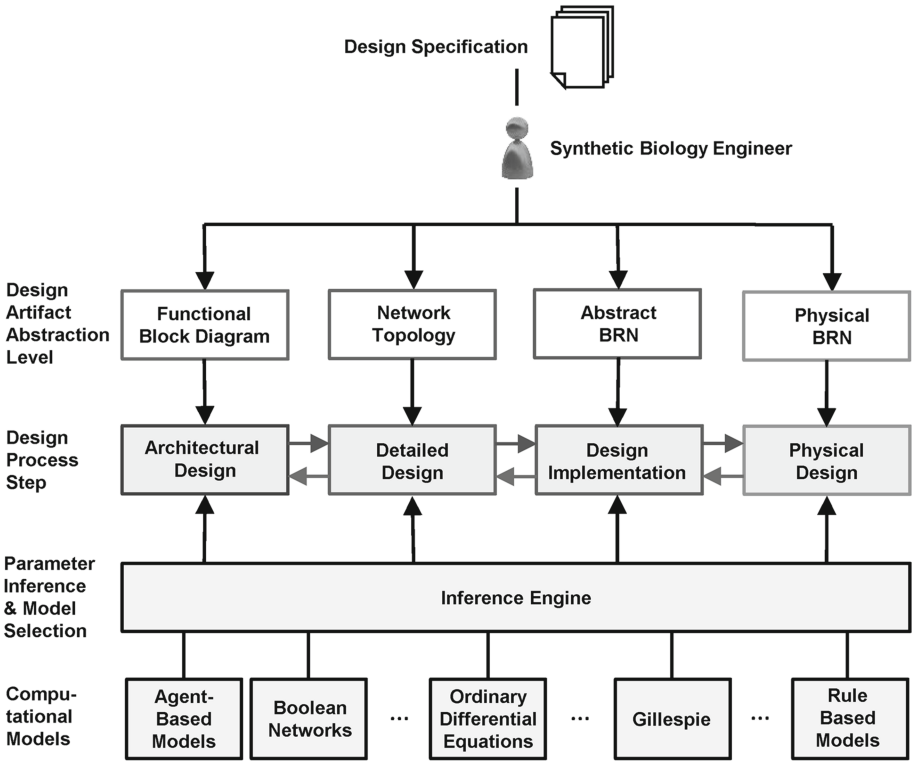


Fig. 5. Inference Engine. Given the *design specification*, the synthetic biologist chooses the appropriate design entry level among the available design artifact abstractions. In general it can be assumed that there are different design options available, which are represented by different design models. It is further assumed that these models can be represented in a computational form. For this, there exist several computational approaches, such as *agent based models*, *Boolean networks*, *ordinary differential equations*, etc. Furthermore, it is usually the case that these models will contain parameters with uncertain values. The *inference engine* allows to infer the parameter values and to select the model which has the highest probability to fulfill the design specification.

design objectives [24]. Computational models (deterministic or stochastic) are developed for each design model. Each model contains a set of uncertain kinetic parameters θ and associated prior distributions $\pi(\theta|\mathrm{M})$ on the parameters. A distance function $\rho(\mathrm{y}, O)$ relates model output, y, to the desired output characteristic O. The design space is explored using Sequential Monte Carlo (SMC). The desired behavior is more accurately approximated with each new population. The ability of each design to achieve the specified behavior is represented by the model posterior $p(\mathrm{M}|\mathrm{D})$. In the shown case, Model M3 encodes the specified behavior with the highest probability. Eventually, the parameter posterior $p(\theta|\mathrm{M}, \mathrm{D})$ allows to identify parameters that are sensitive or insensitive to the targeted behavior.

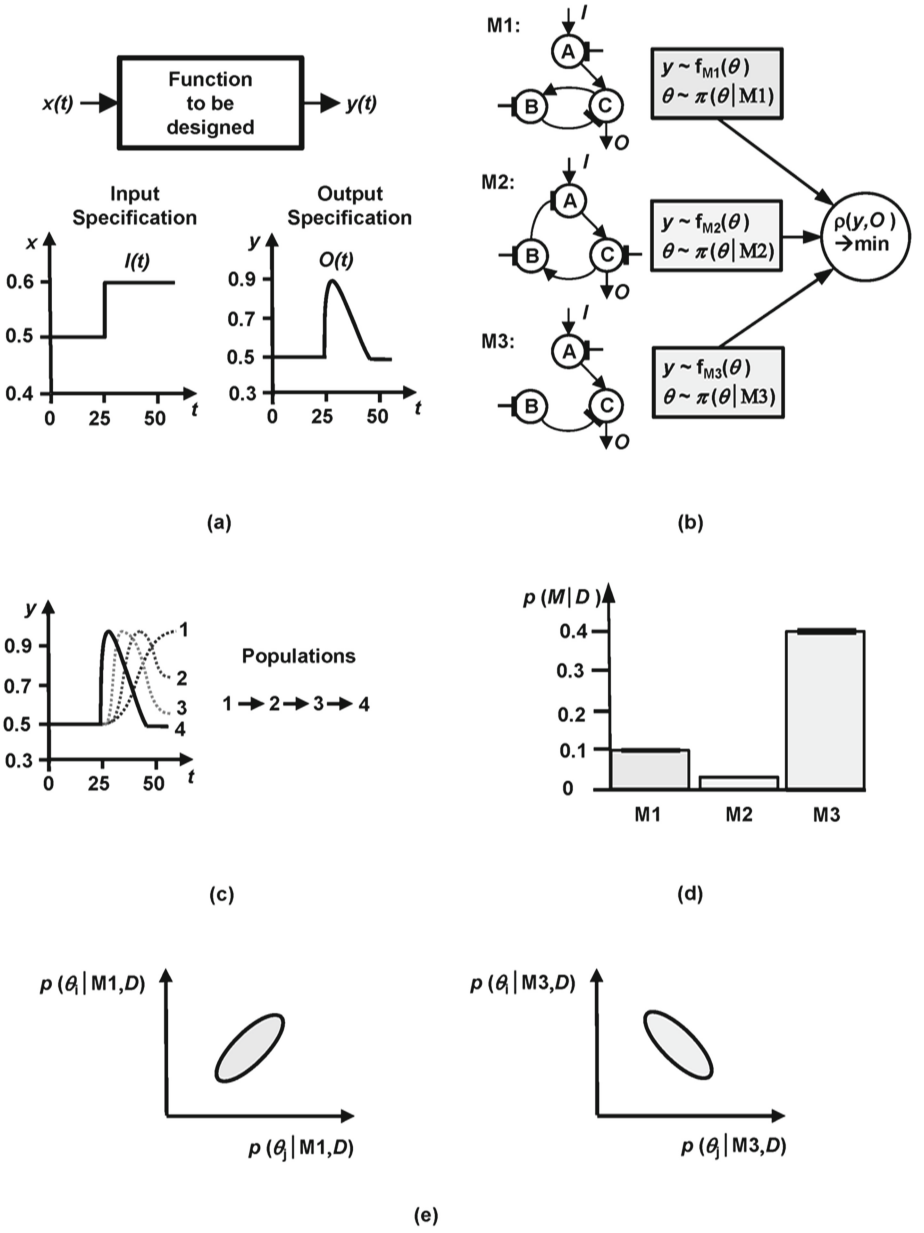


Fig. 6. Design of a biochemical adaptation function. Adopted from Fig. 1 in [1] with permission of the copyright owner (PNAS - Proceedings of the National Academy of Sciences of the United States of America). (a) Specification and encoding of the design objectives. (b) Design entry of competing designs at the network topology level. (c) Design space exploration. (d) The model posterior p (M|D). (e) The parameter posterior (sensitivity analysis).

5 It's the Business Model (, Stupid)

As clearly pointed out in [26], there ought to be no engineering discipline without economics. "It's the economy, stupid" was the famous 1992 U.S. presidential campaign slogan of Bill Clinton that helped him get elected. The slogan was meant to emphasize the economy as the number one issue to be addressed. Paraphrasing this motto, the title of this section aims to draw attention to the fact that business models are of prime concern for industrial strength. It is worth mentioning that the intrinsic relation between industrial strength and business models matches most people's basic intuition. However, it might not be obvious to everyone in the BDA community why he or she should care about business models, which is in most cases a distant concept for engineers and researchers alike. After all engineers develop technical artifacts and managers are supposed to develop business models. The straightforward answer is that businesses are the foundation of any professional activity. The more specific answer is that BDA is in its early formative stage, where everyone has an important role in defining the trajectories for tomorrow's business models, knowingly or unknowingly, willingly or unwillingly. Today's dominant BDA business model is public funding. Clearly, this is an unsustainable model. We already witness much of the "free" software developed in public funding settings end up as "Internet-Zombies". Nevertheless, the public funding model defines the starting point and direction of trajectories to future business models.

Let us recall the popular "There's no such thing as a free lunch" adage, which clarifies that is impossible to get something for nothing. That is, for BDA to become and stay industrial strength there needs to be adequate and sustained funding. This presupposes that BDA solutions are enabled, supported and, maintained by governance models which ensure the continuity of funding. A governance model is defined as a set of policies and practices that outline the responsibilities of the stakeholders (providers and users).

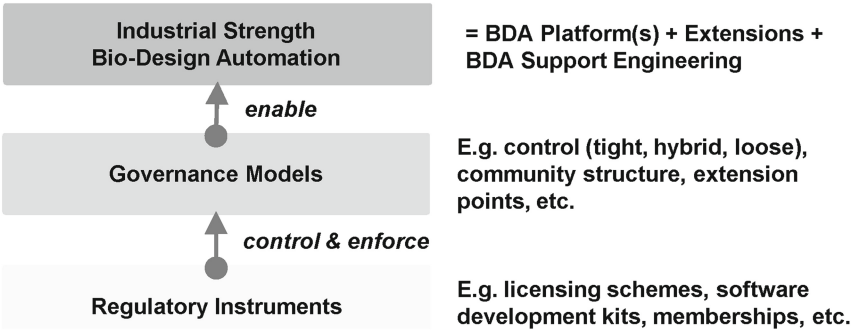


Fig. 7. BDA industrial strength business model framework. *Industrial strength BDA* is represented by the availability of *BDA platforms* with appropriate *extension* points for third party contributions and *BDA support engineering* function that endorses maintains and develops these platforms and their extensions. *Governance models* and *regulatory instruments* play a key role in ensuring the sustained availability of industrial strength BDA. This requires a well-orchestrated interplay between *governance models* and *regulatory instruments*.

It is probably safe to proceed on the assumption that governance models that enforce BDA solutions that are closed, highly complicated and, proprietary and which are controlled by a few vendors should be avoided by all means. That is perhaps a key lesson to learn from established design automation industries. With reference to Fig. 7, Industrial strength BDA needs to encompass (non-proprietary) BDA platforms with easily accessible extension points for third parties as well as an active BDA engineering function responsible for development and support of the community. Such a function will certainly be distributed over several companies, research groups etc. Governance models come in multiple forms, such as tight, loose and hybrid control. Other properties of a governance model are the community or stakeholder structure, the openness for extensions by third parties and so on. For more details the reader is referred to [30]. Governance models need to be enforced by regulatory instruments to be sustainable. Regulatory instruments are based on the use of some form coercive power. However, this does not mean that regulatory instruments are limited to legal enforceable instruments (such as license schemes). In fact, there is a wide range of managerial and technical regulatory instruments available, including mandatory development processes, software development kits, membership schemes, normative authority of the platform owner and many more. Note, that different types of regulation entail different models of governance. It is now the right time to start a discussion in the BDA community about sustainable innovative business models. The provision of open source solutions will probably have to be a key consideration in these discussions.

In any case, the BDA community should be open-minded about any and every possible business model proposed. For example, consider the single-vendor commercial open source business model. Prima vista, this sounds like an economic paradox. However, as explained by Riehle [34, 35] this turns out to be a superior business model in certain settings. If the BDA community does not manage to develop sustainably business models, we might very well end up in an unproductive open source jungle, but we will not arrive at industrial strength BDA. After all, the success or failure of BDA is a question of the right business models and their ability to adapt to changes.

6 Mendel – A Prototypical BDA Platform

The aim of this section is to present the prototypical BDA platform Mendel[13]. We take the position of a start-up aiming to develop and market a BDA solution. While reading this section, the reader is invited to consider himself or herself to be part of this start-up endeavor. In this way, it becomes perhaps more personal and relevant to the reader without losing the big picture. Note that a prototype is a small-scale working model of the final system, built to develop and test design ideas. To this end, we introduce the notion of minimum viable BDA product. In general, “a minimum viable product (MVP) is a strategy used for fast and quantitative market testing of a product or

[13] Mendel is the registered trademark of the BDA platform of MINRES Technologies GmbH. The name was chosen in honor of Gregor Mendel, the father of genetics.

product feature."[14] That is, our start-up follows the lean-start-up methodology. This methodology "favors experimentation over elaborate planning, customer feedback over intuition, and iterative design over traditional 'big design upfront' development" [5][15]. There are two important points to be made here. Ryan Dancey, the CEO of Goblinworks, Inc.[16], concisely expressed them recently as follows[17]:

(a) "The point of an MVP is not to make one and stop work. The point is to make one and use it as the very first point of interaction with real customers, and then start iterating on the design by tightly integrating the users' feedback into the development of new features and expansion of existing features."
(b) "A 'minimum viable product' is not a 'minimum' product. The key word is 'viable,' not 'minimum.'"

Clearly, the prototype must conform to the minimum requirements imposed by our four core hypotheses. It is beyond the scope of this paper to present and discuss the complete set of requirements. Here we will focus on some exemplary requirements from which we are able to develop our BDA prototype.

Let us start with a business and design system view. The backbone of the synthetic biology industry will be formed by small and medium enterprises (SMEs). Consequently the costs for introducing and maintaining BDA solution for a synthetic biology SME must be within the means of such a company. Our research shows that costs in the upper three digit, lower four digit price range seem to be acceptable for the combined hardware and software cost prices. In any case, a well-conceived product tiering[18] strategy is ab initio required. This in turn demands a highly scalable product architecture. Further, it needs to be highlighted that the confidentiality requirements of biotechnology companies are very strict. In fact, they prohibit to utilize recent ICT trends such as cloud computing, web-scale IT, mobile apps and application, etc. But also note that openness and confidentiality in a system context are not necessarily conflicting. Last, but not least, customers are increasingly concerned about the vendor lock-in problem. In this scenario, "the coupling between the customer and the provider become so high to a point that it is no longer economically viable to move from one provider to another" [3]. The imperative overarching requirement following from the above is openness. From a business point of view the openness requirement is represented by the requirement for open business models. Readers who are interested in more details and comprehensive descriptions of open business models are referred

[14] http://en.wikipedia.org/wiki/Minimum_viable_product

[15] Lately, the MVP approach and the lean-start-up methodology became very prominent among high-tech start-ups.

[16] https://goblinworks.com/, Goblinworks, Inc. is a start-up that develops Pathfinder, a massive-multiplayer online game.

[17] http://massively.joystiq.com/2014/03/03/pathfinder-onlines-ryan-dancey-on-crowdforging-a-minimum-viabl/

[18] "Product tiering is a pricing structure that is … used by producers, in which" customers "are segmented by willingness" and ability "to pay for specific (added) product benefits." See: Breetz C (2014) Product Packaging as Tool to Demand a Price Premium: Does Packaging Enhance Consumers 'Value Perception to Justify a Price Premium. Anchor Academic Publishing.

to [21]. From a design systems perspective the openness requirement translates into the requirement for open design systems (ODSs). ODSs are characterized by the reuse open source software whenever it is feasible (technically and legally), open standards, a platform approach, etc. Readers who are seeking a more thorough understanding of ODSs are referred to [8].

Next we look on The Rational Design Fallacy hypothesis. This hypothesis requires us to ensure that the design automations we introduce are essence preserving. This can only be accomplished by close collaboration of synthetic biology engineers and BDA engineers. The requirement is therefore to ensure that an appropriate collaboration process is in place and is continuously exercised. Such a process is highly iterative and interactive. It is instructive to note that the MVP approach and the lean-start-up methodology are perfect instruments to ensure the accomplishment of this requirement.

The key requirement of The Uncertainty Rules hypotheses is straightforward. Any BDA solution needs to provide a computational inference engine and the capability to deal with stochasticity.

Let us now discuss some implementation aspects of our prototype BDA platform Mendel. For this purpose we use the notion of a minimum viable design process based on the MVP concept. Here we consider the case of the design of BRNs. From a technological point of view, a BDA product is comprised of a hardware and a software part. The software part can be implemented straightforward based on a few design decisions. We decided to use Eclipse [39] as the basis for our BDA platform. Some of advantages of Eclipse are reusability, trustworthiness, confidentiality, quality, clarity, longevity, flexibility, and rapid development on the basis of a plug-in system. Figure 8 shows the minimum valuable design process for the design of BRNs and the Mendel GUI along the stages of this design process.

By their very nature stochastic computations are highly compute intensive tasks requiring significant computational resources. Practical cases will therefore principally require the use of massively parallel algorithms which in turn require an appropriate and affordable (!) high-performance computing (HPC) hardware. To this end we have analyzed the feasibility of various hardware acceleration approaches[19]. These days, there are three cost-effective different hardware acceleration options:

(a) GPU based HPC accelerators (NVidia Tesla, AMD Firepro),
(b) Many-Core based HPC accelerators (Intel Xeon Phi).
(c) Reprogrammable logic chips (FPGA - field programmable gate arrays).

Performance figures for the first (a) and (b) are known while for the latter approach no re-ports can be found. Recent developments allow now to transform a high-level behavioral description, e.g. in SystemC[20], into configuration information. Thus FPGAs can be used as highly configurable application accelerators executing simulation and analysis algorithms directly while consuming a magnitude less power. From a

[19] As the intended products are targeted mainly to SMEs, super computers are not a valid attempt for financial reasons.

[20] SystemC is a set of C++ classes and macros which provide an event-driven simulation interface in C++; http://en.wikipedia.org/wiki/SystemC.

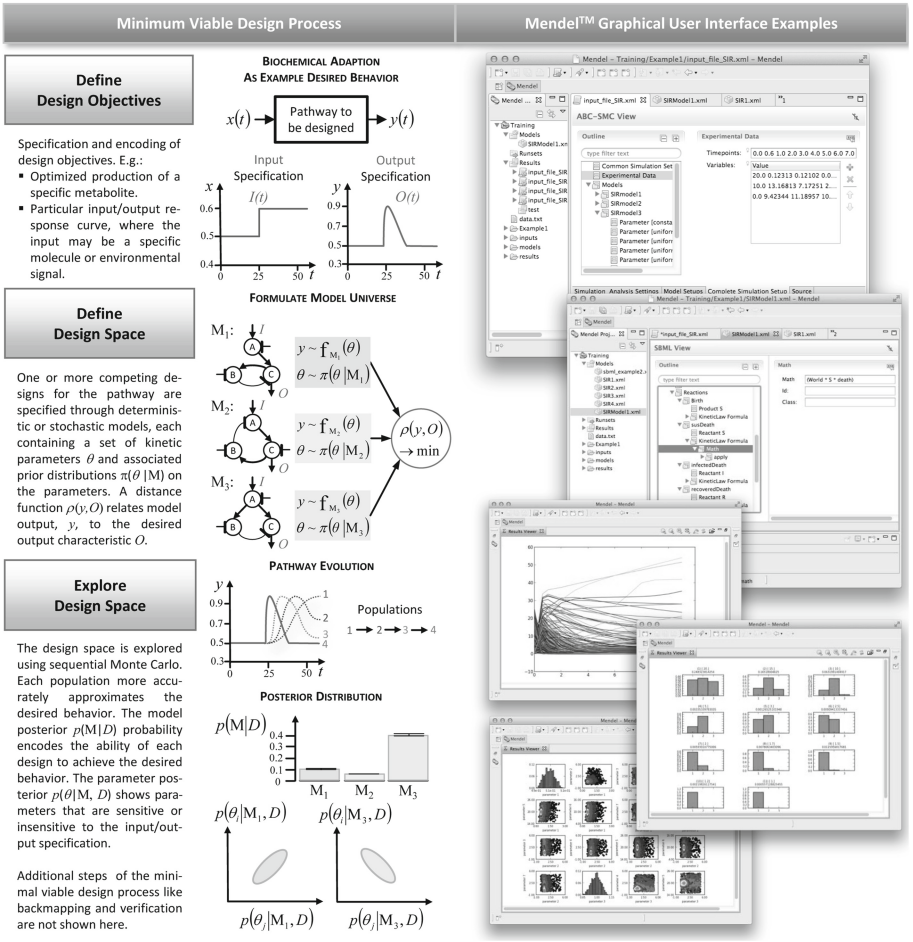
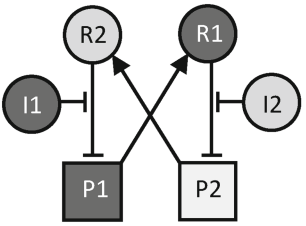


Fig. 8. Mendel GUI along the stages of a minimum viable BRN design process

Table 1. HLS Latencies. Duration of HLS as function of the unrolling factor[a]. HLS: High-Level Synthesis. ODES: First-order Ordinary Differential Equation System.

Unrolling factor	12	4	∞[a]	0
Duration of HLS	21 m47.418 s	2 m59.423 s	1 m52.367 s	1 m52.503 s
Time to solve a single ODES	868,19 µs	869,81 µs	868,54 µs	868,54 µs

[a]Unrolling, is a loop transformation technique that attempts to optimize a program's execution speed at the expense of its binary size. The unrolling factor can be controlled manually by the programmer or chosen by the software only (∞).



Toggle Switch Operation		
I1	I2	Action
0	0	No Change
0	1	R2=1
1	0	R1=1
1	1	Restricted Combination

⊣ Inhibition (Switch-Off)
→ Activation (Produce)

For i=1,2
Pi: Constitutive Promoter i
Ri: Repressor i
Ii: Inhibitor I

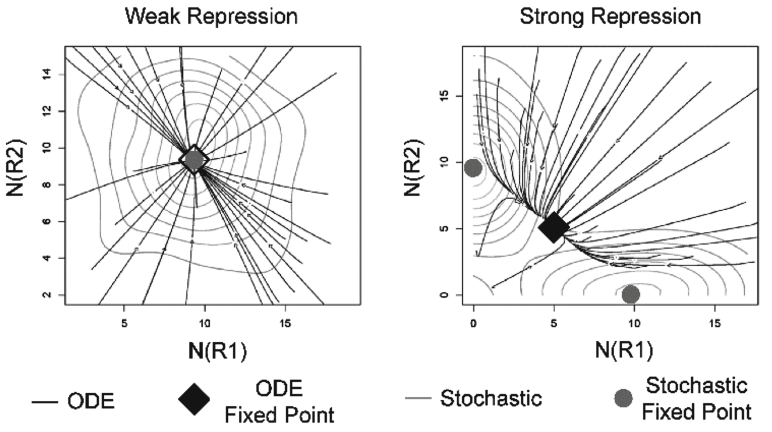
(a)

$$\frac{du}{dt} = \frac{\alpha_1}{1+v^{\beta}} - u$$
$$\frac{dv}{dt} = \frac{\alpha_2}{1+u^{\gamma}} - v$$

u: Concentration of repressor 1
v: Concentration of repressor 2
α_1: Effective rate of synthesis of repressor 1
α_2: Effective rate of synthesis of repressor 2
β: Cooperativity of repression of promoter 1
γ: Cooperativity of repression of promoter 2

$\beta, \gamma > 1$: Cooperativity; $\beta = \gamma = 1$: No Cooperativity

(b)



(c)

Fig. 9. Genetic toggle switch. (a) Schematic and Boolean operation table. (b) First-order differential equation system governing the dynamic behavior of the toggle switch. (c) Simulation results.

performance perspective our results show that FPGAs can compete with both GPU and Many-Core accelerators, while yielding higher power efficiency. However, one issue prevents the use of FPGAs in current solutions. Simulation and analysis are highly iterative and interactive tasks. Latency from starting an analysis until the first computation starts is governed by the runtime of the synthesis and mapping process. Currently this latency ranges from 2 to 20 min which is not acceptable in an interactive design process. Table 1 summarizes some results of the feasibility study. Recent announcements of FPGA vendors reveal a promising convergence towards OpenCL as lingua franca being supported by vendors of GPU as well as FPGA based solutions. This would allow to run the same algorithm description on different acceleration platforms utilizing their specific advantages (low latency, high efficiency) in a hybrid solution. However, currently we use a HPC-GPU solution.

Finally we want to present the results of analyzing a genetic toggle switch using our prototype platform Mendel. Toggle switches belong to the class of genetic circuits which received considerable interest in the synthetic biology community, e.g. [22]. Figure 9(a) presents a schematic of a toggle and the associated Boolean operation table.

The switch is composed of two repressors R1, R2 (with concentrations u, v) which negatively regulate each other's production. More details can be found in [22]. With I1 = S, I2 = R, R1 = Q, and R2 = $\overline{Q}$ it can be easily seen that the toggle switch represents a SR latch. Figure 9(b) presents the ODEs governing the dynamic behavior of the switch. Based on this, it was assumed that for the circuit to operate as a toggle switch strong repression and the cooperative binding $(\beta, \gamma > 1)$ is required. In the case of no cooperativity there should be no bistability (see Fig. 9(c)). However, if one uses a stochastic approach, the bifurcation diagram for strong repression and no cooperativity shows two fixed points, that is the switch is bi-stable. Without a stochastic approach this behavior would not have been revealed [22].

7 Conclusions

The four hypotheses of this paper aimed at highlighting what we believe are the most important current challenges for establishing industrial strength BDA, the domain-specific ICT for synthetic biology. Any BDA solution should comply with the requirements imposed by these hypotheses. There might be other important concerns that we have not touched. We ought to hear about them. We might have covered some aspects insufficiently or inaccurately to some reader. We want to learn about those. But most importantly, and in spite of these conceivable gaps, this paper will hopefully foster a broad community discussion about industrial strength BDA.

Acknowledgements. The HPC hardware related work, referred to in this paper, is supported in part by the Bavarian Ministry of Economic Affairs, Infrastructure, Transport and Technology, Grant No. TOU-1212-0002. We wish to thank particularly Vadim Ermolayev for encouraging us in writing this paper and for his stimulating discussions on the topic.

References

1. Barnes, C.P., Silk, D., Sheng, X., et al.: Bayesian design of synthetic biological systems. Proc. Natl. Acad. Sci. USA **108**, 15190–15195 (2011)
2. Barnes, C.P., Silk, D., Stumpf, M.P.H.: Bayesian design strategies for synthetic biology. Interface Focus **1**, 895–908 (2011)
3. Bessani, A., Correia, M., Quaresma, B., et al.: DepSky: dependable and secure storage in a cloud-of-clouds. Trans. Storage **9**, 1–33 (2013)
4. Bhatia, S., Densmore, D., Pigeon, A.: Design visualizer for synthetic biology. ACS Synth. Biol. **2**, 348–350 (2013)
5. Blank, S.: Why the lean start-up changes everything. Harvard Bus. Rev. **91**, 63–72 (2013)
6. Buchli, J., Santini, C.C.: Complexity engineering, harnessing emergent phenomena as opportunities for engineering. In: Reports of the Santa Fe Institute's Complex Systems Summer School (2005)
7. Cederfeldt, M., Elgh, F.: Design automation in SMEs-current state, potential, need and requirements. In: ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy, p. 1507. Engineers Australia (2005)
8. Christiansson, P., Svidt, K., Sørensen, K.B.: Future integrated design environments. J. Inf. Technol. Constr. **14**, 445–460 (2009)
9. Collins, H.M.: Tacit knowledge, trust and the Q of sapphire. Soc. Stud. Sci. **31**, 71–85 (2001)
10. Densmore, D.: Bio-design automation: nobody said it would be easy. ACS Synth. Biol. **1**, 296 (2012)
11. Earman, J.: A Primer on Determinism. Springer, Heidelberg (1986)
12. Ermolayev, V., Keberle, N., Matzke, W.-E.: An upper level ontological model for engineering design performance domain. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 98–113. Springer, Heidelberg (2008)
13. Ermolayev, V., Matzke, W.-E.: Towards industrial strength business performance management. In: Mařík, V., Vyatkin, V., Colombo, A.W. (eds.) HoloMAS 2007. LNCS (LNAI), vol. 4659, pp. 387–400. Springer, Heidelberg (2007)
14. Gardner, T.S.: Synthetic biology: from hype to impact. Trends Biotechnol. **31**, 123–125 (2013)
15. Gentner, D., Stevens, A.L.: Mental Models. Psychology Press, Hillsdale (1983)
16. Gioia, D.A., Pitre, E.: Multiparadigm perspectives on theory building. Acad. Manag. Rev. **15**, 584–602 (1990)
17. Gregor, S., Jones, D.: The anatomy of a design theory. J. Assoc. Inf. Syst. **8**, 312–335 (2007)
18. Habibi, N., Mohd Hashim, S.Z., Rodriguez, C.A., et al.: A review of CADs, languages and data models for synthetic biology. J. Teknologi **63** (2013)
19. Hassoun, S.: Genetic/bio design automation for (re-)engineering biological systems. In: Design, Automation and Test in Europe Conference & Exhibition (DATE 2012), pp. 242–247 (2012)
20. Huerta, M., Downing, G., Haseltine, F., et al.: NIH Working Definition of Bioinformatics and Computational Biology. US National Institute of Health (2000)
21. Lindgren, P., Rasmusssen, O.H., Poulsen, H., et al.: Open business model innovation in healthcare sector. J. Multi Bus. Model Innov. Technol. **1**, 23–52 (2012)
22. Lipshtat, A., Loinger, A., Balaban, N.Q., et al.: Genetic toggle switch without cooperative binding. Phys. Rev. Lett. **96**, 188101 (2006)
23. Luscombe, N.M., Greenbaum, D., Gerstein, M.: What is bioinformatics? A proposed definition and overview of the field. Methods Inf. Med. **40**, 346–358 (2001)

24. Ma, W., Trusina, A., El-Samad, H., et al.: Defining network topologies that can achieve biochemical adaptation. Cell **138**, 760–773 (2009)
25. Martin, J., Odell, J.J.: Object-Oriented Methods: A Foundation, UML edn., 2nd edn. Prentice-Hall, Upper Saddle River (1998)
26. Matzke, W.-E.: Biotechnology, synthetic biology, and ICT define the emerging knowledge-based bio-economy. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) ICTERI 2013. CCIS, vol. 412, pp. 1–19. Springer, Heidelberg (2013)
27. Matzke, W.-E.: Engineering design performance management–from alchemy to science through ISTa. In: Proceedings of the 4th International Conference on Information Systems Technology and Its Applications. LNI, GI, pp. 154–179 (2005)
28. Matzke, W.-E., Stumpf, M., Mascher, T.: Towards industrial strength BDA - a short essay. In: 5th International Workshop on Bio-Design Automation. Imperial College, London, UK (2013)
29. Nonaka, I.: The knowledge-creating company. Harv. Bus. Rev. **85**, 162–171 (2007)
30. Noori, N., Weiss, M.: Going open: does it mean giving away control? Technol. Innov. Manag. Rev. (2013)
31. Otero-Muras, I., Szederkényi, G., Hangos, K.M., et al.: Dynamic analysis and control of biochemical reaction networks. Math. Comput. Simul. **79**, 999–1009 (2008)
32. Parnas, D.L., Clements, P.C.: A rational design process: how and why to take it. IEEE Trans. Softw. Eng. **SE-12**, 251–257 (1986)
33. Polanyi, M.: The Tacit Dimension. Routledge & K. Paul, London (1967)
34. Riehle, D.: Erratum to: The single-vendor commercial open source business model. Inf. Syst. E-Bus. Manag. **10**, 427 (2012)
35. Riehle, D.: The single-vendor commercial open sourse business model. Inf. Syst. E-Bus. Manag. **10**, 5–17 (2012)
36. Rouse, M.: Industrial strength. In: WhatIs.com (2005)
37. Shaw, W.H.: Engineering management in our modern age. In: Proceedings of the Engineering Management Conference. IEMC '02, pp. 504–509. IEEE (2002)
38. Simon, H.A.: The Sciences of the Artificial. MIT Press, Cambridge (1996)
39. The Eclipse Foundation. Eclipse Foundation (2014)
40. Thiriet, M.: Conclusion. In: Intracellular Signaling Mediators in the Circulatory and Ventilatory Systems, pp. 911–918. Springer, New York (2013)
41. van Beuzekom, B., Arundel, A.: OECD Biotechnology Statistics 2009. OECD Publishing, Paris (2009)
42. Wagner, R.K., Sternberg, R.J.: Practical intelligence in real-world pursuits: the role of tacit knowledge. J. Pers. Soc. Psychol. **49**, 436 (1985)
43. Zott, C., Amit, R.: Business model design: an activity system perspective. Long Range Plan. **43**, 216–226 (2010)