A Survey on Engineering Approaches for Self-Adaptive Systems
(Extended Version)

Christian Krupitzer\(^a\), Martin Breitbach\(^a\), Felix Maximilian Roth\(^a\), Sebastian VanSyckel\(^a\), Gregor Schiele\(^b\), Christian Becker\(^a\)

\(^a\)University of Mannheim
Schloss, 68131 Mannheim, Germany
{christian.krupitzer, martin.breitbach, felix.maximilian.roth, sebastian.vansyckel, christian.becker}@uni-mannheim.de

\(^b\)Embedded Systems, Faculty of Engineering, Duisburg Essen University
Duisburg, Germany
gregorschiele@uni-due.de

Abstract
The complexity of information systems is increasing in recent years, leading to increased effort for maintenance and configuration. Self-adaptive systems (SASs) address this issue. Due to new computing trends, such as pervasive computing, miniaturization of IT leads to mobile devices with the emerging need for context adaptation. Therefore, it is beneficial that devices are able to adapt context. Hence, we propose to extend the definition of SASs and include context adaptation. This paper presents a taxonomy of self-adaptation and a survey on engineering SASs. Based on the taxonomy and the survey, we motivate a new perspective on SAS including context adaptation.

Keywords: Taxonomy, Self-Adaptation, Survey, Self-Adaptive Systems, Context Adaptation.

1. Introduction
The complexity of modern pervasive information systems is increasing. Due to the growing number of powerful mobile and embedded devices as well as the omnipresence of relatively high speed wireless networking, users today expect systems to operate whenever and wherever they want, while traveling, at home, at work, or during vacation. Systems are highly distributed and must integrate all available, highly specialized and heterogeneous devices (ranging from embedded sensor nodes to Cloud servers) and data streams (including web data and real time sensor data) that operate in an ever-changing environment with fluctuating network resources and availability. In additions, systems are no longer restricted to small, tightly controllable areas with single administrative responsibility, like smart rooms or buildings but are interconnected, leading to truly pervasive, global systems like Smart Cities or the Internet of Things.

Developing, configuring, and maintaining such systems is a very difficult, error prone, and time consuming task. One promising way to reduce this effort is self-adaptation. A self-adaptive system (SAS) is able to automatically modify itself in response to changes in its operating environment [1, 2]. The modification is done by adjusting attributes (parameters) or artifacts of the system in response to changes in the system itself or in its environment.

In recent years, SAS have seen an increasing level of interest in different research areas like Pervasive Computing, Autonomic Computing [2], and Nature-Inspired (Organic) Computing [3].

SAS provide so called self-* or self-management properties like self-configuration, self-healing in the presence of failures, self-optimization, and self-protection against threats [2, 4]. For achieving adaptive behavior, basic system properties are self-awareness and context-awareness [5]. Self-awareness describes the ability of a system, to be aware of itself, i.e., to be able to monitor its resources, state, and behavior [6]. Context-awareness means that the system is
aware of its operational environment, the so called context [7]. According to Dey, context is ”any information that can be used to [characterize] the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves” [8]. The system uses sensors to collect information about its context and reasons about the information.

In this paper, we provide a structured overview of self-adaptation and approaches for engineering SASs, analyze future research directions, and motivate the need for a new perspective on self-adaptation in pervasive computing systems. Our main contributions are as follows: First, we develop a taxonomy for self-adaptation that integrates existing views on self-adaptation and specifically context-adaptive systems, which are most relevant to pervasive computing. Second, we survey existing approaches for engineering SASs. Third, we discuss a new type of SASs.

These contributions are directly reflected in the structure of the remaining part of the paper. In the next section, our taxonomy for self-adaptation is presented. In Section 3, we present approaches for engineering SASs. Based on the taxonomy and the approaches, in Section 4 we describe a new perspective of a SAS. A conclusion closes the paper.

2. Self-Adaptation

In this section, we summarize different aspects and perspectives on self-adaptation in SAS research and adaptation in general, e.g., in pervasive systems, and present them in a comprehensive taxonomy for self-adaptation. Our taxonomy incorporates the results of an extensive literature review and integrates different existing taxonomies and works on (self-)adaptation. Figure 1 shows an overview of our taxonomy. In the remaining part of this section we discuss the different dimensions of our taxonomy.

Our taxonomy presents important characteristics of self-adaptation. These issues must be addressed by the implementation of the adaptation logic of such adaptive systems. They influence the reasoning about adaptation as well as monitoring. The presentation of techniques for the implementation of the adaptation logic follows in Section 3.2.

Before going into detail about our taxonomy, we first provide an overview of the main sources and influences for it. Different taxonomies for self-adaptation or adaptation have been developed over the years. As one of the first ones, Rohr et al. describe a classification scheme for self-adaptation research in 2006 [9]. In their work, they classify self-adaptation among the dimensions origin, activation, system layer, controller distribution, and operation. In 2009, Salehie and Tahvildari present an overview over the landscape of self-adaptive software and related research challenges, including their own taxonomy for self-adaptation [5]. A current overview of adaptation can be found in [10], in which Handte et al. classify the adaptation support for pervasive applications into the dimensions of time, level, control, and technique. Macías-Escrivá et al. describe in a recent survey current approaches, research challenges, and
applications for SASs [11]. All these surveys provide important insights into the field of SAS. However, none of them gives an integrated view, incorporating all different existing views and aspects. In contrast, the goal of our work is to present such a uniform taxonomy for self-adaptation.

Additionally, several works discuss aspects of self-adaptation. McKinley et al. highlight the difference regarding parameter vs. compositional adaptation [12]. In [1], the authors discuss the spectrum of adaptation from static activities to dynamic ones. In 2008 and 2010, two Dagstuhl seminars focused on research issues regarding the engineering of SASs [13, 14]. Furthermore, there are two surveys that focus on Autonomic Computing. In [4], Huebscher and McCan present an overview of Autonomic Computing and its applications, whereas Dobson et al. highlight autonomic communications in [15].

In [5], Salehie and Tahvildari introduce the 5W+1H questions for eliciting adaptation requirements:

- When to adapt?
- Why do we have to adapt?\(^1\)
- Where do we have to implement change?
- What kind of change is needed?
- Who has to perform the adaptation?
- How is the adaptation performed?

Other authors formulate similar questions (e.g., [16], [17], [12], [18]). According to Salehie and Tahvildari, the questions must be addressed during implementation of a SAS [5]. Therefore, it seems reasonable to answer these questions when speaking about adaptation. Our taxonomy shows, how we answer the six questions.

As mentioned above, the different aspects of the taxonomy answer the 5W+1H questions. However, our taxonomy has a different view on the dimension known as type of control. Automatic control is distinguished from manual control [10]. As a SAS should adapt without user involvement, we do not include this aspect in our taxonomy. In other words, the question: "Who has to perform the adaptation?" is not answered with our taxonomy, because the nature of a SAS leads to an automatic type of adaptation. Table 1 shows how the taxonomy answers the questions for adaptation. In the rest of the section, we present our taxonomy in more detail.

<table>
<thead>
<tr>
<th>Question</th>
<th>Dimension of the taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>When?</td>
<td>Time (Reactive vs. Proactive)</td>
</tr>
<tr>
<td>Why?</td>
<td>Reason (Context, Technical Resources, User)</td>
</tr>
<tr>
<td>What?</td>
<td>Technique (Parameter, Structure, Context)</td>
</tr>
<tr>
<td>Who?</td>
<td>N/A (Nature of a SAS leads to an automatic type of adaptation)</td>
</tr>
<tr>
<td>How?</td>
<td>Adaptation Control (Approach, Adaptation Decision Criteria, Degree of Decentralization)</td>
</tr>
</tbody>
</table>

Table 1: Relation of the Taxonomy Dimensions and the Questions.

2.1. Time

The time aspect is related to the when-question: "When should we adapt?". The "traditional" view is a reactive adaptation: After an event that causes the need for adaptation, e.g., a change in the resources or a drop in performance, an adaptation plan is worked out [10]. Rohr et al. identify two further perspectives that plan adaptation before an actual event happens: (i) predictive and (ii) proactive [9]. Here, the predictive dimension of time describes a situation, in which a system identifies the need of adaptation before a drop in performance occurs, whereas proactive is an adaptation to improve the performance without foregoing or anticipating drop in performance. Handte et al., however, divide the temporal aspects of adaptation in two dimensions: (i) reactive and (ii) proactive. The reactive dimension is in accordance with Rohr et al.’s description, however, proactive is defined as "modifications of an application performed before an application can no longer be executed" [10]. Therefore, Rohr et al.’s [9] predictive dimension is equal to the proactive dimension of Handte et al. [10], and the proactive dimension of Rohr et al. is not explicitly mentioned by Handte et al. [10]. In this work, we use the temporal definition of Handte et al. The split of time into three dimensions is not necessary. The distinction of adaptation before or after the need for adaptation is enough for describing the temporal aspect of adaptation. The third perspective introduced by Rohr et al. – continuous adaptation

\(^1\)The why question was changed for this work. In [5], it was understood as motivation for building SASs.
for performance improvements – is part of the self-optimization property [2]. Self-optimization is a subset of self-adaptation and, therefore, is implicitly included [5].

The time of adaptation is a central question. From the user’s point of view, proactive adaptation is preferable, since it avoids interruptions in the user’s workflow with the system. On the other hand, the prediction algorithms needed for proactive adaptation have several issues. They are complex to develop, their suitability is highly dependent on the specific prediction tasks, and faulty results can cause suboptimal or malicious adaptations. Therefore, many approaches focus on reactive adaptation [10]. However, the choice for proactive or reactive adaptation is not exclusive. Mapping the adaptation process to the Autonomic Computing MAPE cycle, the basic functionality for adaptation is monitoring the environment, analyzing for change, computing adaptation plans, and executing these plans [2]. Reactive and proactive adaptations involve similar activities regarding monitoring, planning, and executing, but strongly differ in the analyzing phase. With reactive adaptation, the monitored data is analyzed for abnormal patterns. With proactive adaptation, the monitored data is used to forecast system behavior or environmental state. It is possible to combine proactive and reactive adaptation such that proactive adaptation is the goal, and reactive adaptation is used as a back-up mechanism, i.e., if a change was not predicted (e.g., failing of a component).

An additional aspect related to time is the type of monitoring [5]. Continuous monitoring describes a constant monitoring effort regarding the resources and the environment. Adaptive monitoring refers to monitoring of selected features and, in case of anomalies, the monitoring process is intensified. The decision between continuous vs. adaptive monitoring influences the cost of monitoring and, hence, of the self-adaptation process. However, as it is an implementation detail and not related to the adaptation decision itself, it is not included in our taxonomy.

2.2. Reason

In general, adaptation is a reaction to a change. On the one hand, this reaction is costly. Accordingly, the type and impact of a change should be clearly identified in order to decide, if an adaptation is necessary. On the other hand, the parameters of an adequate adaptation are determined by the source of the change. Therefore, it is important to identify the reason for an adaptation: "Why do we have to adapt?". This is a central question influencing the reaction, because different reasons result in different adaptation activities.

In a SAS, the reason for an adaptation is a change in one or various system elements:
- (i) a change in the technical resources, e.g., a defect of a hardware component, software fault, or the availability of an alternative network connection,
- (ii) a change in the environment, e.g., the state of a context variable changed, or
- (iii) a change regarding the user, e.g., a change in the composition of the user group or the user preferences.

Therefore, in our taxonomy, adaptation can be triggered by changes in the technical resources, the context, or the user(s). The explicit inclusion of the user as a reason for adaptation is not mentioned in other taxonomies.

Other taxonomies for (self-)adaptation do not include the reasons for adaptation. An explanation could be that the reason for an adaptation is not seen as part of the decision, i.e., how the system should recover from a change or the adaptation process itself. In our taxonomy, we explicitly include the reason for an adaptation because it is the trigger of the adaptation process. In our understanding, it is fundamental for the adaptation decision to identify the reason for adaptation and to reason about it. Consequently, the question "Why do we have to adapt?" has to be included in the aspects of self-adaptation. Furthermore, the reasons for adaptation should be included in the aspects of self-adaptation, because they determine the elements that have to be monitored. Without an adequate monitoring process, the need for an adaptation can not be identified and, therefore, the adaptation process can not be started. Accordingly, it is important to monitor the state of the technical resources (including hardware and software), the environment, and the users interactions with sensors and interfaces for user interaction.

2.3. Level

Adaptation can be implemented on different levels in the system. In order to answer the question of "Where do we have to implement change?", one must be aware of the different levels of a SAS. In general, a SAS is composed of different elements: the managed element(s) and the adaption logic [19]. While the adaptation logic as the control unit of the technical resources often stays stable, the managed elements can be adapted. Furthermore, it is possible to adapt the environment and the user(s). This extends the common view on SASs, where the environment is only monitored, not changed. An adaptation of the adaptation logic itself would enable an improvement of the adaptation performance over time, but is out of scope of this work.
The managed elements are composed of various levels. The technical resources are hardware and other managed resources, such as computers, smartphones, robots, traffic signs, or production facilities. The hardware is controlled by system software, i.e., an operating system, and a middleware in case of distributed systems. On top of the system software, the application is set up. The application can be an application running on a single device, or a distributed application split in application parts running on various devices simultaneously. Both are seen as single application in the following. Furthermore, various different applications can run simultaneously, as well as interact and form ensembles of applications. This can lead to interferences, i.e., undesired interceptions and dependencies regarding the use of resources [20]. For the interaction between the managed resources as well as the adaptation logic elements, communication is needed. Communication is seen in two perspectives. The network infrastructure is the physical network connection, consisting of network cards, routers, WLAN, etc. Communication patterns are the logical communication, i.e., the style of interaction between the elements. Possible implementations can be event-based communication or pub/sub communication. An additional level besides the technical resources and software is the system’s context. Context-altering systems are able to adapt their context [7]. This extends the former view on SASs, where the environment is monitored but not explicitly altered. Implicitly, this may be done by actuators of the technical resources, but so far is not explicitly controlled by the adaptation logic. In current SAS literature, context adaptation is not included in the level of adaptation.

Adaptation can happen on all levels [9, 10]: Smartphone apps that switch to silent mode when the user is in a meeting, e.g., detected by using calendar information, offer adaptation on the application level. Adaptive middleware offers the possibility to exchange components at runtime [21, 22]. Autonomic communication techniques enable adaptation on the network level [15]. An example adapting the communication is switching the network connection, e.g., from 3G to WLAN as far as a WLAN connection is available. Self-healing capabilities enable the automatic start of back-up systems, e.g., in a data center, which alters the technical resources. Context-adaptive applications can adapt their behavior to the surrounding context, or adapt the context through actuators [7]. An example for context adaptation is a smart meeting room that automatically dims the light when a presentation starts.

The adaptation logic of a SAS must be aware of these different levels, possible adaptation alternatives, and define adaptation plans for the appropriate levels. Therefore, the question: “Where to adapt?” must be answered on the correct level(s) in order to achieve the system’s goals. In case the adaptation alternatives can target different levels, the adaptation logic must decide which plan to execute based on criteria like cost, e.g., in terms of time, or achievable utility for the system. Adaptation of the user is possible in theory, but not desirable in practice, as the application is for the user and not the other way round. Hence, we do not include adaptation of the user in our taxonomy.

2.4. Technique

It is not sufficient to only identify the levels, where the adaptation should take place. Additionally, the specific adaptation actions that should be carried out need to be identified, i.e., “What kind of change is needed?”. In literature, different techniques for adaptation can be found.

McKinley et al. distinguish between two approaches for adaptive software: (i) parameter adaptation and (ii) compositional adaptation [12]. Parameter adaptation achieves a modified system behavior by adjusting system parameters. This can be achieved quite easily, as the adaptation logic must only control and change parameters. On the other hand, changing parameters can involve high complexity, if the parameters are depended on each other. Furthermore, in case different algorithms for one component exist, it is possible to switch between them. However, the dynamic integration of new algorithms at runtime is not possible. The dynamic integration of new components is not possible as well. An example for parameter adaptation is a rule-based system, in which rules specify the necessary amount of servers running, depending on the current workload of the servers. In this setting, it is only possible to add new servers as long as configured servers are available. However, it is not possible to dynamically transfer the responsibilities of a faulty server instance to a new server instance, as such an exchange affects the system’s structure. Compositional adaptation enables the exchange of algorithms or system components dynamically at runtime. Therefore, it is possible to exchange defect components in order to prevent performance loses, improve the performance by adding new components, or adjust the system to new circumstances. To use the server example above, this means that during the execution of the application, additional servers can be integrated into the system, or that a new server instance can overtake the responsibilities and tasks of another server instance.

A third technique can be derived from Pervasive Computing. Here, systems are context-adaptive and can alter the context in which the systems are running. Therefore, we extend the range of techniques by context adaptation.
Context adaptation is not addressed by McKinley et al. [12]. Whereas the monitoring of the context and the detection of contextual changes is supported by many SAS, context alteration is often not integrated in SASs.

Handte et al. categorize adaptation techniques in behavior, composition and context [10]. Here, parameter adaptation is a form of behavior adaptation. As a change in composition can also lead to a modified behavior, this view overlaps. In this work, the techniques for adaptation are categorized in parameter, structure, and context. This is a combination of the approaches mentioned by McKinley et al. and Handte et al. Parameter refers to adaptation through the change of parameters. Structure subsumes change in the structure of the technical system, such as the exchange of components, a new composition of components, or the removal/addition of components. Further, changes in the relation between elements, technical resources or the environment/user(s), are structural adaptations as well. Context refers to any changes in the context, e.g., modifying the state of context variables via actuators. Combinations of techniques in one adaptation plan is possible, e.g., changing parameters of one component and adding further ones.

2.5. Adaptation Control

A SAS is composed of the adaptation logic and the managed resources. Responsible for controlling the adaptation is the adaptation logic. This involves monitoring the managed system as well as the environment and the user(s), analyzing the monitored data for anomalies, planning the adaptation, and executing the adaptation plans. Therefore, the adaptation logic is responsible for answering the question: "How to adapt?". By answering this question the adaptation logic determines how to perform adaptation.

Two approaches for implementing the adaptation logic can be found in literature. SASs following the internal approach intertwine the adaptation logic with the system resources. The external approach splits the system into adaptation logic and managed resources, which increases maintainability through modularization [5, 23].

The control unit needs a metric in order to decide how to adapt. Different metrics are present in the literature: models, rules and policies, goals, or utility functions [24]. The different adaptation possibilities must be analyzed with the help of the criteria and the best one must be chosen. Different criteria can be combined, e.g., goal model based planning where goals have additional utility values for solving conflicts between goals.

Another aspect of the adaptation logic is the degree of decentralization. A centralized adaptation logic can be a solution for systems with a small amount of resources to manage. When it comes to large systems with many components to manage, a decentralized approach for a split of the responsibilities can improve the system performance for adaptation [19]. Various degrees of decentralization are possible. In fully decentralized approaches, each sub-system has a complete adaptation logic and different patterns of communication are possible [19]. Hybrid approaches add central components to decentralized approaches or distribute the adaptation logic functionality to sub-systems.

In this section, we presented a taxonomy of self-adaptation. We combined different works on self-adaptation and adaptation in general. Self-adaptation can be described with the dimensions: Time, Reason, Level, Technique, and Adaptation Control. Developers of SASs must be aware of these dimensions and their systems must integrate these dimensions for monitoring and reasoning about adaptation. In Section 3, we give a detailed introduction to the adaptation logic. Further, we discuss different approaches for developing a SAS's adaptation logic.

3. Engineering Self-Adaptive Systems

In the previous section, we presented our understanding of self-adaptation by creating a taxonomy for self-adaptation. The dimensions of the taxonomy are important aspects for a SAS’s adaptation logic to reason about adaptation. In this section, we want to highlight the "How?" aspect: the construction of a SAS’s "brain" - the adaptation logic. The other dimensions of the taxonomy influence monitoring, reasoning, and executing, controlled by the adaptation logic. Therefore, we present the general structure of the adaptation logic of a SAS as control unit of the adaptation process and discuss general implementation issues. Subsequently, we discuss different approaches for constructing SASs and show which aspects of our taxonomy they address.

3.1. Adaptation Logic Issues

A SAS is composed of managed resources and the adaptation logic. This can be represented as a tuple $SAS = (AL, MR)$ with the adaptation logic $AL$ and the managed resources $MR$ [19]. Both, adaptation logic and managed resources, can be divided into various elements. The adaptation logic – as control unit for the adaptation [25] – is modeled as group of adaptation logic elements $AL = a_1,...,a_n$ and monitors the environment (M), analyzes the data
for change (A), plans adaptation (P), and controls the execution of the adaptation (E). These activities are known from Autonomic Computing as MAPE cycle or MAPE functionality [2]. Other authors propose similar feedback loops for SASs (e.g., [15], [26], [27]). Therefore, in this work, the MAPE cycle is used as basic feedback structure for the adaptation logic. The adaptation logic can be complemented by additional elements, such as a knowledge component responsible for managing content (e.g., monitoring values, rules, or policies) or a learning component. The managed resources are a group of resources $MR = mr_1, ..., mr_n$, such as hardware with software, smart phones, robotics, or unmanned vehicles. Figure 2 presents the basic layout of a SAS. The dashed line shows the system border.

The dimensions Time, Reason, Level, and Technique of our taxonomy can be mapped to the MAPE functionality. The Time dimension influences the decision of analyzing algorithms, as proactive recognition of the need for adaptation has other requirements – especially the need for predictions – as reactive detection of changes. Monitoring should be continuous no matter whether the adaptation is proactive or reactive. The Reason dimension influences monitoring, analyzing, and planning, as it describes the reasons for adaptation and, therefore, the aspects that should be monitored, where analyzing has to determine changes as well as the metrics, that must be addressed with the adaptation plans. The Level for adaptation is obviously important for planning and executing as these activities must be aware of the elements that should be adapted. Monitoring has to determine the elements for the levels that are present in the managed resources. The Technique dimension influences the planning and executing, as planning describes which adaptation techniques to use on which elements and executing controls the execution of the techniques. The fifth dimension Adaptation Control describes the structure of the adaptation logic and is not related to any specific MAPE functionality. Table 2 presents the mapping of the MAPE functionality to the dimensions of our taxonomy.

<table>
<thead>
<tr>
<th>Time</th>
<th>Reason</th>
<th>Level</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring</td>
<td>Continuous</td>
<td>What to monitor</td>
<td>Identification of the levels</td>
</tr>
<tr>
<td>Analyzing</td>
<td>Algorithms depend on reactive or proactive dimension</td>
<td>Where to analyze</td>
<td>—</td>
</tr>
<tr>
<td>Planning</td>
<td>—</td>
<td>What should be influenced by planning</td>
<td>Adaptation plans address these levels</td>
</tr>
<tr>
<td>Executing</td>
<td>—</td>
<td>Execution of the change on the levels</td>
<td>Execution of the techniques on different elements</td>
</tr>
</tbody>
</table>

Table 2: Relation of the MAPE Activities and the Dimensions of the Taxonomy.

Abstracted from the concrete implementation of the adaptation logic, general issues can be identified. The adaptation logic can be intertwined with the rest of the application or separated. For analyzing and planning of the adaptation, the adaptation logic can use different metrics. Another issue is the degree of decentralization of the logic and the distribution of the MAPE functionality on different sub-system parts. Following, these issues are presented in detail.
3.1.1. Implementation Approaches

In literature, two approaches regarding the interplay of adaptation logic and managed resources can be found [23]. The *internal* approach intertwines the application logic and the managed resources. Sensors, effectors, and adaptation logic elements are mixed with the managed resources. This approach has several drawbacks: issues in scalability and maintainability can arise, global information of the system is not guaranteed, and testing of the adaptation logic is complicated [23]. Nevertheless, the internal approach can be suitable for local adaptations, e.g., exception handling.

The *external* approach separates the adaptation logic and managed resources [23] and connects them via sensors and effectors [2]. Sensors are interfaces used by the adaptation logic for getting information from the managed resources, e.g., the system state or performance measurements. Effectors are used by the adaptation logic for enabling adaptation, e.g., through changing parameters or starting components. The external approach addresses the drawbacks of the internal one and offers scalability as one adaptation logic can manage various resources, maintainability as the responsibilities are divided and can be maintained separately, and eases the achievement of a global view. Furthermore, it offers reusability of the adaptation logic or at least of the processes and algorithms used in the adaptation logic [5, 23]. The external approach is superior in most cases and can be found more often in literature [23].

3.1.2. Adaptation Decision Criteria

Different approaches for analyzing and planning the adaptation are present. These approaches need a metric for identifying the need for adaptation and for choosing a suitable adaptation plans, respectively. These metrics are based on: *models*, *rules/policies*, *goals*, or *utility* [24].

In model-based approaches, models represent the actual and the desired situations. They include goals, the system architecture, the environment, or other circumstances. Through analysis of the models, suitable adaptation plans are worked out. For rule-based or policy-based approaches, rules or policies determine, how the system should react in different situations and how to adapt. Often, rules/policies are defined at design time, which leads to non-dynamic approaches. Goal-based approaches aims at fulfilling specific system goals. These goals influence, how the system should perform. During the planning process, the adaptation logic must define adaptation plans for achieving these goals. One has to mention, that the goals can be contradicting, which must be solved by the adaptation logic. In utility-based approaches, utility is a function of the system value for the user and involved costs. The goal is to maximize the overall system utility. The adaptation logic evaluates the utility values of adaption strategies and selects the one with the highest utility. Disadvantageous are the difficulty of defining utility functions, as well as the complexity and the uncertainty in calculating adaptation costs and utility values.

So far, most SASs monitor the context for detecting changes in the context. The explicit inclusion of context adaptation through actuators is not included in the decision criteria for the analyzing and planning. This could be done for reducing problems with plans that lead to unanticipated changes resulting of context adaptation.

3.1.3. Degree of Decentralization

Usually, SASs are systems-of-systems. SASs can be found in cloud computing, traffic control systems, production facility control, or pervasive systems. Central questions are, whether the adaptation logic should be decentralized or centralized and how the MAPE functionality should be distributed. In [28], the authors define self-adaptation as a top-down approach for system control. This means that a central unit has to control the system. Self-organization is seen as the opposite, a bottom-up approach. Dedicated units organize and coordinate themselves without a central instance. In this work, self-adaptation subsumes both. *Centralized self-adaptation* with a central instance for control is defined as self-adaptation in [28]. Self-organization with decentralized system control is in this work *decentralized self-adaptation*. In case the adaptation logic has to be distributed, developers have to define an interaction pattern.

Within a *centralized* adaptation logic (see Figure 3a), one sub-system implements the adaptation logic, is responsible for monitoring the context and all resources, and controls adaptation. Whereas a global maximum can be achieved because the central instance is aware of all information, this approach is not suitable for large systems, with a high amount of information, or resource-poor devices because of the calculation power needed [19, 29]. For large systems, a central approach is hard to achieve because of the size of the system and real time constraints.

One alternative is a *fully decentralized* approach (see Figure 3b), in which each sub-system has its own application logic with full MAPE functionality and adapts itself. Furthermore, the adaptation logic elements can communicate for achieving global goals. In the IBM reference model for Autonomic Computing [2, 30], additional orchestrating autonomic managers improve the coordination, but introduces a *hybrid approach*. Weyns et al. describe patterns for
Figure 3: Comparison of Adaptation Logic Structures (AL = Adaptation Logic, MR = Managed Resources, MAPE = MAPE functionality).

how to distribute the MAPE functionality on various sub-systems with different levels of interaction and decentralization [19]: (i) Coordinated Control Pattern, (ii) Information Sharing Pattern, (iii) Master/Slave Pattern, (iv) Regional Planning Pattern, and (v) Hierarchical Control Pattern. The patterns (i) and (ii) are fully decentralized with interaction, whereas the patterns (iii) - (v) have centralized elements and, therefore, are hybrid approaches. Figure 3c shows one concrete implementation of the Regional Planning Pattern with 2 groups as an example for a hybrid approach.

Designers of SAAs must be aware of the implications resulting from the distribution of their applications. It is important, to decide for a suitable pattern depending on the number of sub-systems, coordination needed between them for adaptation decision making, or relevance of achieving global goals.

These three issues – approach for the adaptation logic, adaptation decision metrics, and the degree of decentralization – influence the design of an adaptation logic. Furthermore, monitoring and reasoning (analyzing and planning), must include the dimensions presented in our taxonomy. In the following, we present different approaches for SASs.

3.2. Survey on Engineering Self-Adaptive Systems

So far, we presented different aspects of the adaptation logic’s implementation like the integration of adaptation logic, adaptation metrics, and decentralization of the logic. These are general issues. We have not answered the question, how a concrete adaptation logic is implemented. In literature, different approaches for SAS construction can be found. In this section, we categorize and present these approaches. Furthermore, we link the approaches to the dimensions of our taxonomy and the MAPE activities.

Our objective is to present the diversity of approaches that are present in literature and their appropriateness for SAS development. We do not want to discuss every category and approach in fully detail nor do we claim to present all available literature for each category as this would go beyond the scope of this paper. In literature, many works detail the use of the approaches, discuss their strengths and weaknesses, and point to similar works. The interested reader is referred to the cited works for more detailed information about the approaches.

3.2.1. Model-based Approaches

In Model-Driven Engineering (MDE), the focus is on models as first class entities for describing software and its environment [31]. Incomplete information at design time and changing conditions at runtime leads to a shift in the use of models from using models for design to the use of models at runtime [32]. Therefore, runtime models or models@run.time enable monitoring, e.g., by representing the system and its environment, and reasoning, e.g., comparing a model capturing the actual system state with a desired system state, at runtime and provide abstractions from code. In this section, we present different model-based approaches for SASs that use models for monitoring and reasoning. Therefore, we focus on runtime models and neglect the use of models at design time.

According to [4] and [33], three types of models are used for monitoring and reasoning in a SAS:

- System models represent the system state.
- Goal models, policies and rules, or utility functions model the adaptation decision criteria.
- Environment models capture the context.

In the following, we present the first two in more detail. More information on context modeling can be found in literature (e.g., in [34]). Three types of system models can be found in literature: architectural models, feature
Architectural models represent the system’s architecture (e.g., [35, 36, 23]). These models capture the architecture in different representations, e.g., as components [37] or layers [38]. In Section 3.2.2, we enlarge on the use of architectural models. Another type of system models are feature models that capture different features of software [39]. Defining all possible features and combinations of them with feature models offers a way for reasoning about adaptation by representing all possible configurations of a SAS and choosing the most suitable configuration [40]. Feature models for SASs are used by different authors (e.g., [41, 21, 42, 40, 43]). Often, feature models are used as variability models which capture commonalities and variability in a system family ([44, 40]). A third type of system models, behavioral models, describe the behavior of the system and the possible transitions between the different system behaviors [45]. Therefore, they are applicable for describing behavioral adaptation. Often, state automata [46] or UML state machines [47] are used as notation for behavioral models.

Goal models describe the goals that a system should fulfill [38], or its requirements [48], respectively. If the system does not fulfill its goals, adaptation is required. Goal models are used for reasoning as well as creation of adaptation plans (e.g., [48], [38], [49], [50]). Chen et al. combine goal models representing requirements and architectural models representing design decisions in order to achieve complex adaptations at runtime [51]. Mazzola Paluska et al. describe the behavior of an adaptive application with models of Goals and Techniques [52]. Using a Satisfaction score, the application chooses the most suitable Technique for the Goals at runtime.

In literature, different approaches using model-driven techniques for SASs can be found. The Software Engineering Institute defines a Software Product Line (SPL) as “a set of software-intensive systems that share a common, managed set of features satisfying the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way” [53]. Besides the first authors, Gomaa and Hussein presented the Reconfigurable Evolutionary Product Line Life Cycle based on SPLs [54]. Hallsteinse, Hinchev, and Schmid propose a dynamic SPL approach [55], which is used in [21] in combination with models@run.time. Further approaches that use dynamic software product lines can be found (e.g., [40], [43], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67]). All of these approaches use the SPL principle for modeling different system configurations, often with feature models [43] or variability models [40]. Additionally, [64] and [65] combine a model for the system with one for the system features. Using a SAT solver, the adaptation logic can use this information to choose a system configuration that satisfies the current context constraint (cf. [65]). The most suitable configuration is chosen and the system is adapted accordingly, e.g., with the help of component-based [54] or aspect-oriented programming techniques [21]. Therefore, the SPL approaches address mostly analyzing and planning. The MUSIC framework offers model-based development for SASs by integrating service-orientation and component-based development [23, 68, 69]. Design models are used in the design phase for specifying the system components and the initial system configuration. Component models represent the structure of the components. Runtime adaptation models are used for modeling system adaptation at runtime [69]. The MUSIC framework addresses all MAPE activities. MechatronicUML is a model-driven development approach tailored to self-adaptive or self-optimizing systems in mechanical engineering [70, 71]. It supports modeling based on a subset of UML and verification based on decomposition and compositional model checking. Similar, the Adapt Case Modeling Language [72] is a UML-based language which enables specification of self-adaptivity concerns.

Meta-modeling is essential for model-driven development [73]. Therefore, runtime models must be constructed on base of meta-models. Vogel, Seibel, and Giese propose to use megamodels for subsuming different runtime models and reasoning about adaptation on higher levels of abstraction [74]. They integrated their megamodels in a model-driven architecture with a model transformation engine between the adaptation logic and the managed system [75–77]. Lehmann et al. identify the need of additional runtime capabilities for meta-models and propose an approach for meta-modeling of runtime models [78].

Models are often used to support monitoring and reasoning in SASs. Different authors propose the use of models@run.time for SAS construction (e.g., [79], [80], [81], [82], [83], [32]). Vogel and Giese envision to unify development models and models@run.time. As even self-adaptive systems still require (manual) maintenance, an integration of the models could significantly facilitate the maintenance task [84]. Nevertheless, models are only a supporting tool for monitoring and reasoning that need to be accompanied by additional techniques for the change of the managed resources, such as aspect-oriented programming (e.g., [85], [41], [21]), component-based programming (e.g., [54], [86]), architecture-based approaches (e.g., [82], [35], [1]), service-oriented approaches (e.g., [69], [87]), or integration in frameworks, e.g., for context-aware systems [83]. In the further sections, we present these categories in detail. Furthermore, models can be used for checking and verifying system states [31, 85, 88].
3.2.2. Architecture-based Approaches

A software architecture is a set of software elements (or components), the relations between them, as well as the properties of both [89] and denotes the high-level structure of software [90]. Regarding adaptation purposes, software architectures are used for different activities. A SAS must be aware of its structure, which is called self-awareness [5]. Software architectures can be used for representing the system structure and reasoning about adaptation levels [15]. On the other hand, software architectures are used for the construction of SASs and defining responsibilities [38]. In this section, we present architecture-based approaches.

The Rainbow framework [35] is one of the most well-known frameworks for SASs. Driven by external control for self-adaptation, the SAS is divided into an architecture layer and a system layer with the managed resources. The architecture layer has different components, which define adaptation plans. A translation infrastructure translates and controls the deployment of these plans in the system layer. For supporting runtime adaptation, an architectural style notion of the system is used and extended with adaptation operators and strategies. This architectural notion offers the possibility to monitor the resources as a system model and reason about change. Rainbow supports structural and parameter adaptation on different system levels. The 3L Approach of Kramer and Magee [38] offers a three layer architecture model for self-management based on Gat’s three layer sense-plan-act architecture [91]. The Component Control layer controls the managed resources. On top of the Component Control layer, the Change Management layer monitors the system, signals changes to the uppermost layer, and implements adaptation plans on the components, e.g., compositional adaptation through switching components. The Goal Management layer on top of the Change Management layer, works out adaptation plans based on goals the SAS should fulfill. As the planning process is a reaction on changes, only reactive adaptation is supported.

Further approaches use agent-related techniques for implementing architectural change (e.g., [92], [93]; see Section 3.2.7) or dedicated architectural components for controlling the adaptation as Rainbow’s architecture layer [35], the Architectural Run-time Configuration Manager [94], or the Architecture evolution manager in Archstudio [1, 36]. Architecture-based approaches are accompanied by factors for controlling the adaptation as strategies (e.g., [35]), policies (e.g., [30], [2], [94], [95], [92]), goals (e.g., [96], [38], [97]), task characteristics (e.g., [98], [99]), or constraints (e.g., [35]). Architecture-based adaptation research is complemented with research on improvements in the use of architectures and architecture adaptation in a SAS, e.g., architectural patterns [19, 100–103], resource prediction for improving self-adaptation [104], or dynamic architectural styles [105]. The BASE framework defines four evaluation dimensions (behavior, synchrony of change, state, and execution context) which help to assess architectural styles and their suitability for runtime evolution [106].

Many architecture-based approaches represent the architecture in the form of models. Therefore, they use architectural models for monitoring the resources and/or reasoning about adaptation (e.g., [107], [108], [109], [35], [36], [1], [82], [86], [37], [110], [23], [111], [112]). Architecture Description Languages (ADL) describe architectural models. Different authors have developed ADLs or extended existing ones for SASs, e.g., DARWIN ([113] used, e.g., in [114]), C2/adl [36], Dynamic Wright [115], Gerel [116], CHAM [117], COMMUNITY [118], Rapide [119], LEDA [112], or ACME [107, 108]. A comparison of ADLs can be found in [120].

All architecture-based approaches implement the MAPE functionality (mostly implicitly, i.e., without dedicated components). For reasoning about the system’s structure, architectural models are used in combination with metrics as policies, goals, strategies, or constraints. Furthermore, all approaches are external, i.e., they have a dedicated adaptation logic, e.g., in the form of components (e.g., [36], [1], [94]) or layers (e.g., [35] or [38]).

3.2.3. Reflection Approaches

Reflection is the ability of software to examine and possibly modify its structure (structural reflection) or behavior (behavioral reflection) at runtime [121, 122]. The notion of computational reflection in programming languages was introduced by Brian Cantwell Smith [123] and is equivalent to the term reflection [124]. Reflection is divided into two activities: introspection refers to the observation of an application’s own behavior, intercession is the reaction on introspection’s results [12], i.e., structural, parameter, or context adaptation.

The ability of reflection is an underlying principle for SASs [12, 125, 126] that enables self-awareness. In [125], the authors describe a reflection reference model which explicitly includes a meta-level sub-system for reflection and a reflection prism describing the properties of reflection. Weyns, Malek, and Andersson describe a formal approach for self-adaptation based on computational reflection [33], which is presented in detail in Section 3.2.9.

Reflection can be used on different levels. Architectural reflection (or structural reflection) leads to reflection
Regarding the software architecture of an application, i.e., its components, interconnections, or data types [12, 127]. Approaches for architectural reflection are proposed by various authors (e.g., [128], [127], [129], [130], [131], [132], [133]). Behavioral reflection refers to reflection regarding the behavior of the software, e.g., algorithms for computation or communication mechanisms [12, 134]. Reflex is an approach for integrating partial behavioral reflection in Java [126]. Kava also adds behavioral reflection to Java by using byte code rewriting [134]. Additionally, approaches that use behavioral models for analyzing changes in the system fall into this category.

Reflective middleware, such as Dynamic TAO [135] and Open ORB [136], can support runtime reconfiguration of a component-based system [125, 137]. According to [138], reflective middleware can deal with highly dynamic environments and supports development of flexible, adaptive systems. CARISMA integrates the reflective middleware concept and context-awareness [139], which enables the construction of SASs.

In the approaches presented so far, reflection focuses on the software itself, e.g., on components, architecture, or communication behavior. Sawyer et al. propose requirements-awareness for SASs: the use of reflection techniques for introspection of requirements at runtime [140]. Introspection and intercession are basic functionalities for adaptation. With introspection, the reason for adaptation can be detected, i.e., introspection includes monitoring and analyzing. Since introspection classically only detects the cause for adaptation after it happened, only reactive adaptation is possible. Intercession determines and controls the adaptation, i.e., controls planning and execution.

3.2.4. Programming Paradigms

For parameter and structural adaptation, different programming paradigms can be used. This section presents different programming approaches. These are known from Software Engineering and are not specific to SASs.

In Component-Based Development (CBD), also known as Component-Based Software Engineering (CBSE) and Component-Based Programming (CBP), software components are encapsulated parts of software, that can independently from each other be developed, deployed, and composed [141]. Known examples are Enterprise JavaBeans, CORBA Component Model, or the CORBA Component Model [12]. In [12], the authors propose the use of components integrated in a framework that supports late binding for enabling structural adaptation. Therefore, developers must clearly define interfaces, and the adaptation logic must be able to handle the coexistence of components and control the exchange of components. For determining which components must be exchanged, some kind of metric is needed. In literature different metrics can be found, e.g., goal-based [38, 96] or model-based [37, 86]. Architectural models [37, 86] or component models – such as Fractal [142] (e.g., used in [143] or [144]), K-components [131], or OpenCom [137] – define self-contained components. Component-based approaches for SASs are used in many approaches (e.g., [12], [145], [131], [111], [146], [147], [96], [38], [37], [86], [148], [144], [149], [150], [151]). The use of CBP techniques at runtime for achieving structural adaptation of applications differs from using CBD at design time for developing adaptive software as in [47], where CBD is used for constructing autonomous component units.

In Aspect-Oriented Programming (AOP) the program is divided in distinct parts, called concerns. The goal is to use generic functionality in different classes (cross-cutting concerns). Therefore, logical concerns are separated from the concrete implementation [152]. One example of an AOP language is AspectJ, an extension to Java [153]. Whereas dynamic recomposition is often related to cross-cutting concerns as QoS, McKinley et al. propose AOP for enabling separation of concerns, which leads to simplified compositional adaptation [12]. Haesevoets et al. show different possibilities to use AOP techniques for self-adaptation [154]. Other authors highlight single aspects (e.g., [155], [143], [156], [157]). For SAS development, AOP is often used as an adaptation mechanism, e.g., for structural adaptation, that is controlled by the adaptation logic. Morin et al. present such an approach by combining MDE and aspect-oriented modeling to manage dynamic variability [21, 41].

Other authors propose the use of generative programming for SASs [12, 158, 159]. In generative programming, software is built with the help of high-level descriptions, that are mapped to generic classes, templates, aspects, and components [160]. Whereas the transformation from description to code is unproblematic, vice versa is not supported, what complicates the use of generative programming for SAS implementation [159]. Adaptive programming techniques can support the development of SASs [161, 162]. In [163], the authors propose an adaptive programming approach which integrates reinforcement learning for learning behavior at run-time, which can only roughly be determined by the developer at design-time. In Context-oriented Programming (COP) – known from Pervasive Computing and Ubiquitous Computing – the context is incorporated as a first-class construct in programming languages [164]. This way, the software is able to reason about changes in the context and to appropriately adapt. For a SAS, incorporating only the context is not sufficient but the managed resources must be integrated as well. COP approaches can be
found in [165], [164], [166], and [167]. Often, COP is combined with AOP [167].

The presented programming techniques, are mainly used for compositional, reactive adaptation. However, they could be used for all kinds of software adaptation because it only offers an adaptation mechanism that is controlled by the adaptation logic. They can be used for planning and executing of adaptation. Further procedures for monitoring and analyzing have to be included. Often, the techniques are supported by middleware and reasoning is model-based.

3.2.5. Control Theory

Control structures are an important part of the adaptation logic of SASs [27]. Control loop engineering poses different challenges: developing reference architectures for control loops, creating a catalog of control loop structures, middleware support for control loop integration, verification & validation techniques to test and evaluate control loops’ behavior, and integration of the user [168]. To address these issues, knowledge from control theory is used for the development of SASs [169]. Shaw states that control engineering methodologies should be considered for the software architecture “when the execution of a software system is affected by external disturbances” [170]. Furthermore, she separates the control system from the main process. There are two types of control systems: open loop and closed loop [170]. The difference is that a closed loop system uses information about monitored output variables to adjust process variables whereas an open loop system adjusts process variables without considering output variables. Closed loop systems can be further divided into feedback control systems and feedforward control systems. Whereas the feedback loop reacts on changes of the output variable, the feedforward loop adjusts process variables by anticipating the effects on the output variable. According to Abdelzaher et al. “control theory provides a formal approach to designing closed loop systems” [171].

Feedback control loops can be adaptive [27]. An adaptive control loop can adjust the controller to respond to changes of the controlled process. For this purpose, there is a second control loop on top of the main control loop. Two standard schemes of adaptive feedback control loops are Model Identification Adaptive Control (MIAC) [172] and Model Reference Adaptive Control (MRAC) [173]. However, adaptive control comes to its limitations when the control law, e.g., decision functions, stay fixed for their lifetime [174]. Then, the control law can only be updated in a limited way. To overcome this problem, reconfigurable control makes use of learning approaches in order to dynamically adapt the control law [174]. In [175], feedback control loops are represented using actor objects [176]. An actor encapsulates its state and behavior and can only receive, process, and send messages. Each actor is assigned one specific role (sensor, filter, controller, or effector). Adaptive behavior is achieved by connecting corresponding actors. Sensors are responsible for collecting data which is then analyzed by filters. Controllers decide on appropriate actions which are executed by effectors.

Kephart and Chess introduce in [2] the MAPE-K loop. There, the MAPE functionality is extended with a shared knowledge repository. Meanwhile, the MAPE-K loop concept is a widely used reference model for SASs (e.g., [19], [69], [29], [175], [87], [38], [177]). As the control loop is separated from the managed element, adaptation control is external. Adaptation time is reactive since the MAPE-K loop is a feedback loop. Similar to the MAPE-K loop, Dobson et al. presented the autonomic control loop [15]. Here, the MAPE functionality is called differently – collect, analyze, decide, and act – but has the same responsibilities [15].

In [178], Müller et al. propose to model feedback loops as first class design elements. Hebig et al. take up this issue and present a UML profile that allows for explicitly architecting control loops with component diagrams [179]. There, a control loop is built of four components: controller, process component, sensor, and actuator. Sensor and actuator are used to observe and adjust the managed element. Thus, they use an external approach. In [180], Diao et al. implement a deployable testbed for Autonomic Computing based on different control approaches. Filieri, Hoffmann, and Maggio formulated a methodology which generates a system model and then uses this model to set up a controller which is able to manage the non-functional requirements of the system [181].

Control engineering for SASs focuses on feedback loops which results in reactive adaptation. Proactive approaches might be feasible using feedforward loops. Adaptation control is mainly external and can take place decentralized. Furthermore, the approaches mentioned above address the whole MAPE functionality. The reader is referred to [182] for further information on designing SASs using control engineering.

3.2.6. Service-oriented Approaches

Services are small, encapsulated, and autonomous units of software that fulfill a specific task. In Service-Oriented Computing (SOC) such services are used to "support the development of rapid, low-cost, interoperable, evolvable, and
massively distributed applications” [183]. The key for SOC is a service-oriented architecture (SOA), which enables finding, use, and connection of services. The service approach can be transferred to SASs.

The naive approach for building a service-oriented SAS is to model the managed resources' functionality as services and use the SOA for communication. The adaptation logic decides, which services should run. Therefore, dynamic exchange of services offers structural adaptation capabilities [184]. The problem is, how to build the adaptation logic and how to enable the exchange of services.

Different frameworks for SASs use service-oriented techniques. In [185], the authors outline the principle of autonomic SOAs. The MUSIC framework offers model-driven development of service-based SASs [23, 68, 69]. The SASSY framework is a model-driven, self-architecting framework for SASs [87, 186]. Services in a directory, QoS-patterns, and adaptation patterns form the base for SAS construction. In MetaSelf, a SOA is controlled by dynamically modifying metadata, policies, and components [187]. The QoSMOS framework provides a full stack (MAPE) solution to service-based SAS with a focus on QoS [188]. It offers ways to formally specify QoS requirements and tool support for developers. Other approaches combine AOP and service-orientation for building SASs ([155], [157]), take advantage of SOA as technical basis for implementing the variability of dynamic SPLs [62], integrate multi-agent system approaches and SOA [189], use component models which are filled with service implementations at runtime ([190], [149]), apply requirements engineering techniques for building service-based SASs (CARE method) [191], adopt the Active Components concept for engineering decentralized control [192], or focus on QoS aspects within service-based self-adaptation, such as the MOSES framework [193].

Service-oriented approaches focus on structural adaptation through exchange of services or change in the composition of services. Therefore, the managed resources are represented as services and the adaptation logic, either implemented as services or as a layer above, controls the composition of services. Often, models are used for reasoning, which services need to be adjusted (e.g., [69], [87], [149]). These approaches are reactive, whereas proactive approaches would be feasible, too. The level of adaptation is mostly the application level, the services of the application. The re-composition or exchange of services concerns the planning and executing activities and needs to be accompanied by monitoring and analyzing procedures for building a SAS.

3.2.7. Agent-based Approaches

A software agent is a piece of software that fulfills a specific task autonomously and cooperate for common tasks. A multi-agent system (MAS) is a system of agents that share common goals and, therefore, communicate and cooperate. For MASs it is claimed, that "they are especially suited to develop software systems that are decentralized, can deal flexibly with dynamic conditions, and are open to system components that come and go” [93]. These properties make agent-based techniques valuable for the development of SASs.

De Wolf and Holvoet discussed several aspects in the construction of self-organizing MAS, such as dynamics and decentralized control in a MAS [194], emergence and self-organization for MASs [195], a proposition for a methodology for engineering self-organizing MASs [196], and design patterns for adaptable MAS [101]. Tesuaro et al. present Unity [197], a decentralized architecture for Autonomic Computing with autonomic elements. With Unity, autonomic system behavior, such as goal-driven self-assembly, self-healing, and real time self-optimization, can be achieved. Bernon et al. propose the Adaptive Multi-Agent Systems (AMAS) theory, in which the system is based on the MAS principle and the agents cooperate for achieving self-organization and adjustments to changes in the environment [198]. Furthermore, they offer tools for the development of SASs based on the AMAS theory. The TOTA middleware can support self-organization in MAS by integrating adaptivity and context-awareness [199].

Other authors integrate a MAS and SOA for building SASs [189], use CBD for designing autonomous agents as components [47], use adaptive object models for designing a MAS [132], or present design patterns for SASs based on agent structures ([200], [102], [101]). Further approaches use agent-related techniques for implementing architectural change ([92], [93]). The variety of agent-based approaches offers (in theory) both temporal aspects of self-adaptation. Adaptation happens mainly on the application level, but other levels are feasible, too. Most approaches rely on a decentralized adaptation logic, as the MAS should not be controlled by one central unit. The main focus in this category of the MAPE cycle is the planning component.

3.2.8. Nature-inspired Approaches

Natural systems are composed of a large number of decentrally organized interacting components [201]. Each component has only limited information based on which they adapt. The resulting overall behavior of the system is then different from the behavior of the individual components, called emergence [27, 195]. Nature-inspired systems
Biological approaches in computer science have emerged with the study of collective behavior in natural MAS by Parunak [203]. Biological mechanisms, such as flocking [204–206], foraging [204, 207, 208], nest building [204], molding [204], local inhibition [201, 209, 210], lateral inhibition [211], chemotaxis [201, 212, 213], embryogenesis [204], morphogen gradient [201], local monitoring [201], quorum sensing [201, 204, 214], consensus [201, 213], firefly synchronization [215], stigmergy [203, 204, 212, 213], web weaving [204, 216], brood sorting [204], and the human immune system [213, 217] or human autonomous nervous system [2, 204], respectively, have been applied in self-organizing systems and can be transferred to SASs. Fernandez-Marquez et al. specified bio-inspired design patterns for self-organizing systems [218].

Physical approaches, so far, mainly focus on the metaphor of potential fields pioneered by Kathib who employed them for obstacle avoidance in path planning in [219]. Reif and Wang propose a distributed method to control autonomous robots based on potential fields [220]. In [221], the authors apply potential fields for exploration and foraging tasks of autonomous robots. Weyns et al. employ potential fields for adaptive task assignment in MAS [222]. In [199], Mamei and Zambonelli present the TOTA approach, a framework for pervasive and mobile computing approaches. Here, communication takes place with the help of potential fields. The SAPERE approach also allows for the creation of potential fields [213].

Chemical approaches focus on chemical reactions. The Higher Order Chemical Language (HOCL) serves as a mean for programming service composition based on chemical reactions and enables late binding of components at runtime [223]. Virol and Casadei propose biochemical tuple spaces, a coordination model for self-organizing systems based on chemical reactions [224]. SAPERE contains a spatial substrate which acts as a coordination space for agents which interact via tuples called live semantic annotations (LSAs) [213].

Social approaches concentrate on market and auction mechanisms, as well as social norms (e.g., [225], [226], [227]). As an example, in [227], coordination in MAS is based on social conventions.

Most investigated approaches focus on the planning component. Thus, adequate procedures for monitoring, analyzing, and executing need to be included.

3.2.9. Formal Modeling and Verification Approaches

Self-* properties and dynamics make it hard to prove the correctness of SASs [228, 229]. However, a SAS requires behavioral and structural guarantees [26], especially in safety-critical domains, e.g., traffic light systems [230]. In contrast to regular software, self-adaptive software further needs verification at runtime to match requirements and adaptation decisions [231]. A comprehensive overview of existing techniques to assure that self-adaptation maintains system goals and corresponding challenges is given in [232].

As an approach to reduce the complexity of the verification task, Güdemann, Ortmeyer, and Reif introduced the Restore Invariant Approach (RIA) [228, 229, 233] based on transition automata for representing system components and temporal logic to express functional properties. In this context, an invariant is a formula that divides all possible allocations of functional properties into two sets: the functionality can be provided, i.e., the invariant holds, and the functionality cannot be provided, i.e., the invariant is violated [233]. The former set is also called corridor [229]. As long as the invariant holds, the system finds itself in the production phase. As soon as it gets violated, the system changes to the reconfiguration phase, in which it restores the invariant and then switches back to production. So far, the adaptation control of the RIA has been centralized and external, but a decentralized control unit is possible, too [229]. RIA is reactive and adapts only to changes in the (technical) resources by altering the structure.

Weyns, Malek and Andersson suggest the formal reference model for self-adaptation (FORMS) [33]. They use the Z notation in order to define a SAS, its relationship with the environment, and self-adaptation. Adaptation control takes place centralized and external. King et al. propose a self-testing framework for Autonomic Computing systems that validates structural and behavioral change requests at runtime in order to avoid costly system failures [234, 235]. When a change request comes in, it is forwarded to a special testing autonomic manager which validates the request. This approach uses an external adaptation control.

Other approaches [236–239] explicitly address the domain of adaptation in cyber-physical systems. The DEECo model and framework allows to simulate such systems on a large scale to examine the effects of adaptation strategies [236, 237]. In [238], the authors present IRM-SA, a design method and model for cyber-physical systems that focuses on
on dependability. It is combinable with DEECo. Gabor et al. apply the concept of ultra-high fidelity simulations (digital twin) to online planning in cyber-physical systems [239].

Franco et al. formulate a stochastic approach to improve adaptation [240]. Based on a formal architecture description of a system in an ADL, the approach generates Discrete Time-based Markov Chains for each adaptation strategy. By solving this model, the system determines the most promising strategy. Smith and Sanders [241] present an incremental top-down approach to formally develop self-organizing systems, which is built on the three "scales of observation" introduced in [242]. The Autonomic System Specification Language is a framework for formally specifying and generating autonomic systems [243]. Cordy et al. use a model checking technique based on transition systems in order to verify dynamic software product lines [60]. Another approach focuses on probabilistic runtime model checking using Discrete Time Markov Chains [244, 245]. Arcaini, Riccobene, and Scandurra specify MAPE-K loops in Abstract State Machines to verify the correctness of the AL and to discover unwanted interferences between multiple loops at design time [246]. Priesterjahn et al. present a formal model for the timed hazard analysis of self-healing systems [247]. De la Iglesia and Weyns use timed computational tree logic (TCTL) and timed automata (TA) to formally specify and verify behaviors of self-adaptive systems [248]. Mongiello, Pelliccione, and Sciancalepore combine the Design by contract concept and reflection to verify context-aware applications at runtime [249]. Nguyen et al. developed a notation for verification of time-based constraints through simulation with colored Petri nets [151]. Zhang, Goldsby, and Cheng propose to model SASs as a collection of steady-state programs and a set of adaptations which switch between these steady-state programs [250]. They use Linear Temporal Logic (LTL) to describe the non-adaptive parts of the system and A-LTL (an extension to LTL) for the adaptive parts. A modular model checking approach then verifies requirements satisfaction. Weyns and Ifikhar introduce a modular decision making process for SASs which includes runtime verification based on ActiveFORMS [251, 252]. The system permanently maintains models (system model, environment model, and quality models) and simulates them to verify whether it still meets requirements or requires adaptation [252]. The DECIDE approach uses formal verification based on continuous-time Markov chains and continuous stochastic logic to decentralize control loops of SAS and to guarantee QoS compliance [253].

This category mainly addresses the analyzing part of the MAPE-loop. Further information on formal methods in SASs can be found in [254].

3.2.10. Learning Approaches

Learning in a SAS is tightly coupled to self-optimization. A SAS continually optimizes its structure, parameters, or algorithms in order to become more efficient with regard to performance or cost [2]. Therefore, learning in a SAS focuses on structural optimization (e.g., [197], [255], [42], [256], [257], [258], [259]). In Unity [197], self-optimization works with the help of a central resource arbiter that computes optimal allocations of resources based on a utility function, which is refined by learning. Elkhodary et al. present a feature-oriented architecture for SASs [42, 255]. Each goal has a utility function that is used for optimization. Fisch et al. propose a collaborative learning approach by exchanging knowledge [256]. Rules represent knowledge. When an agent discovers a novel useful rule, it broadcasts the rule, so that other agents can make use of it.

Several approaches use reinforcement learning for structural optimization. Dowling et al. present collaborative reinforcement learning for decentralized coordinated self-adaptive components [257]. Collaborative reinforcement learning makes collective adaptation possible by collaborative feedback [260]. Kim and Park rely on Q-learning to improve the mapping between environmental change and software configuration in a SAS at runtime [261]. Other approaches apply reinforcement learning for decentralized self-adaptive service assembly [262], generation and evolution of adaptation rules [263], or learning policies for the decision phase of a decision engine [264].

Parra et al. employ a constraint-based approach to optimize context-aware adaptations [258]. Pandey et al. integrate fast, deterministic planning and slower, but more optimal Markov Decision Process planning to a hybrid planning approach which is able to handle the trade-off between timeliness and optimality [259]. Prothmann et al. use evolutionary programming for learning [230]. In an organic traffic control scenario, they introduce an observer/controller architecture with on- and offline learning. A learning classifier system, responsible for online learning, is combined with an offline learning evolutionary algorithm. Tomforde et al. extend the observer/controller architecture with a decentralized collaboration mechanism [265] and discuss several distribution possibilities of the architecture [266]. Cakar et al. use a generic observer/controller architecture, proposed in [177], as basis for their learning approach [267]. Brockmann et al. developed a framework for controlled self-optimization in modular system architectures
based on goals with an online learning approach based on machine learning [268]. Rodrigues Filho and Porter sketch a framework for unsupervised online learning which considers the requirement, implementation, and deployment stage of software development [269, 270]. Tanabe et al. propose an online learning technique which uses stochastic gradient descent. It updates the environmental model of a SAS based on execution traces collected at runtime [271]. The Fossa framework provides a simulation-based offline learner to automatically create rules for a rules-based planner [272]. To achieve this, the framework’s learning engine specifies new potential rules with an exploration strategy such as genetic programming and evaluates them offline in a test environment afterwards.

Other works address parameter optimization. [273] gives an overview on evolutionary algorithms with regard to parameter optimization. Abdelwahed, Kandasamy, and Neema present a control framework for self-managing distributed systems [274] and apply model learning procedures online for managing varying environmental and operating conditions.

Algorithmic optimization aims at optimizing how a goal is achieved, hence altering the algorithms. Oreizy et al. mention evolutionary programming and AI-based learning techniques in order to generate new algorithms to encounter uncertain changes [1]. Kafaf and Kim propose a hybrid planning approach that combines a local and a global knowledge base. Here, they use supervised machine learning with kNN and unsupervised learning with k-means clustering for online optimization of an object recognition algorithm [275].

The field of search-based software engineering (SBSE) [276] aims at applying search-based metaheuristic techniques to software engineering. Search techniques such as genetic programming examine large search spaces of candidate solutions to find a (near) optimal solution to problems concerning requirements, design, or testing [277]. Traditionally, SBSE is used at design time. Contrary to the traditional approach, dynamic SBSE applies the principle of SBSE at runtime to determine the most suitable system configuration during the planning phase of self-adaptation [278, 279]. The SASSY framework uses SBSE (hill climbing) for reconfiguration [87]. SA:DuSE is a domain-specific instance of the DuSE approach for design space representation and automated architecture design [280]. It applies SBSE to search the design space of a SAS. The Hermes approach [100] automatically evolves adaptation paths at design or runtime with the help of genetic programming. It is an extension of Plato [281], which generates the target system configurations for the adaptation paths. Coker, Garlan, and Le Goues outline challenges of using SBSE in the SAS domain [282]. Further approaches that use dynamic SBSE in SASs can be found, e.g., [283–288].

Learning approaches mainly focus on the planning component. Adequate procedures for monitoring, analyzing and executing have to be added. Further, learning can be used for meta-adaptation of the adaptation logic.

3.2.11. Requirements Engineering for Self-Adaptive Systems

For SASs, a specialized form of requirements engineering (RE) is needed. Berry, Cheng, and Zhang defined four levels of RE for dynamic adaptive systems: (i) Level 1: general definition of the system and its reaction by developers, (ii) Level 2: the system does RE at runtime for achieving adaptation, (iii) Level 3: decision of developers about adaptation mechanisms, and (iv) Level 4: research regarding adaptation mechanisms [289]. In this work, we focus on the dynamic aspects of RE at runtime (level 2), as it enables adaptation of the SAS.

The dynamic nature of a SAS potentially results in uncertainty at runtime [140], due to a lack of information at design time [13]. According to Whittle et al., changes in a system’s environment, such as “sensor failures, noisy networks, malicious threats, and unexpected (human) input” [290] are the main reasons for uncertainty. They further propose to relax non-critical system goals in emergency situations in order to achieve high-level goals. Different authors propose to include runtime capabilities for RE. Bencomo, Whittle, and Sawyer propose to model requirements as runtime entities and to integrate reflection capabilities for requirements [291]. FLAGS is a goal model based approach for modeling requirements at runtime [292]. In [293], the authors present a runtime infrastructure that is able to manage requirements at runtime based on modeling the requirements with FLAGS. Souza proposes two new classes of requirements for SASs: awareness requirements prescribe indicators of convergence in requirements and evolution requirements represent strategies for adaptation addressing changes in the requirements models themselves [294]. Furthermore, he describes an approach for requirements-based system adaptation. Morse et al. introduce a technique based on fuzzy logic that characterizes and ranks different system configurations depending on their conformance with (possibly conflicting) non-functional requirements [295].

Different requirement languages for SAAs can be found in literature. In [296], the authors present RELAX, which explicitly allows to specify and deal with uncertainty. As RELAX, still most notations for specifying requirements are based on natural language prose [13]. However, there exist also other approaches based on goal models, agents,
or scenarios. Goal-based approaches are for example KAOS [297], i* [298], LoREM [48], FLAGS [292], and CARE [299]. However, only CARE explicitly supports uncertainty. Tropos4AS [300] is an agent-based design framework for modeling SAS requirements based on Tropos [301]. No explicit support for uncertainty is given. Scenario-based notations, such as live sequence charts [302], also do not explicitly support uncertainty.

Since requirements can influence all dimensions of our taxonomy, runtime capabilities for managing requirements are important for reasoning about adaptation. Requirements Engineering approaches usually address the whole MAPE cycle. As requirements are a kind of goal the SAS should fulfill, RE approaches are mainly goal-based.

3.2.12. Further Approaches

There are further approaches, which cannot be categorized into one of the categories presented above.

In task-based adaptation, the system determines suitable adaptation policies based on the users’ task characteristics [98, 99]. The adaptation mechanism is not defined, it only affects reasoning. Therefore, the approach cannot be classified into one of the mentioned categories. It can be seen as utility-based adaptation with the objective to maximize the system utility for the users in supporting their tasks. Task-based self-adaptation is used within the Aura project [303].

Some approaches use middleware-centric adaptation for achieving structural adaptation. Sadjadi and McKinley present an overview on adaptive middleware [22]. In [21], the authors build their approach on top of adaptive middleware. Hallsteinsen, Floch, and Stav proposed an approach for building SASs by using generic middleware for handling the adaptation [130].

Moreover, several approaches introduce a Domain Specific Language (DSL) tailored to describe self-adaptation. Stitch [304] is a language which is able to express adaptation strategies for architecture-based self-adaptation. It empowers developers to represent potential strategies in a decision tree. A utility-based algorithm then chooses the best adaptation strategy at runtime. Alvares, Ruten, and Seinturier propose Ctrl-F [305]. This high-level DSL consists of a static description of architectural elements, similar to common ADLs, and a dynamic part which is able to specify behaviors and policies. Sun and Satoh [306] as well as Anthony [307] further introduce languages capable of specifying adaptation policies.

Additionally, some authors propose processes and modeling dimensions for simplifying the engineering of SASs. In the FESAS project, we present a framework for using reusable components of a component library for adaptation logic development [308]. As outcomes of Dagstuhl Seminars about software engineering (SE) for SASs [13, 14], modeling dimensions (goals, change, mechanisms, and effects) [309], a design space for SASs with five principal clusters of design decisions (observation, representation, control, identification, and enacting adaptation) [25], and SE processes for moving activities from design to runtime when engineering SASs [310] are proposed. Zhang and Cheng proposed a model-based process for the construction of adaptation models, automatic code generation from the models for building adaptive programs, and verification and validation of the models [88]. In [311], the authors propose a guideline with 14 tasks for the construction of self-organizing systems. Lightstone presents a guideline for Autonomic Computing application development focusing on integration of the user [312]. Tomforde and Mülle-Schloer introduce a meta-design process for the development of adaptive systems with five design decisions (observation model, configuration model, similarity metric, performance metric, and validation method) and three phases (preparation, evaluation, utilization) [313]. Different authors propose the use of design patterns for developing SASs (e.g., [19, 100–103, 200, 218, 308, 314–317]).

3.3. Conclusion

In this section, we highlighted implementation issues for the adaptation logic of a SAS. The adaptation logic can be integrated into the resources or dedicated. Different metrics such as models, goals, rules/policies, or utility influence the adaptation decision. Furthermore, developers must specify the degree of decentralization, distribution of the adaptation logic elements, and interaction patterns.

Additionally, we discussed different approaches that can be used for building a SAS’s adaptation logic and we classified the approaches into model-based, architecture-based, reflection, programming paradigms, control theory, service-oriented, agent-based, nature-inspired, formal modeling & verification, learning, and requirements-oriented. In contrast, Macías-Escrivá et al. differentiate approaches for building SASs into main approaches (external control mechanisms, CBSE, model-driven, nature-inspired computing, MAS, feedback systems), global tools & methods (models, simulation, architecture, frameworks), and specific tools & methods (feedback control loops, decision-making, RE) [11]. Whereas the categories are similar as simulation is part of the learning approaches we presented, we
did not divide the approaches in their application for SAS development. Nevertheless, we acknowledge that our classification is not interception-free, e.g., in some architecture-based approaches, models are used for representing the architecture and CBSE for the structural adaptation of the architecture. Furthermore, the categories can be combined for developing a SAS, e.g., model-based analyzing with component-based adaptation.

We mapped the approaches to our taxonomy of self-adaptation as well as the MAPE activities. By doing this, we identified that most approaches for development of SAS are reactive and do not explicitly include the influence of actuators of the managed resources on the context, i.e., context adaptation. This can lead to serious problems and, in the worst case, end up in a cycle where the adaptation logic decides to adapt, the adaptation leads to an unforeseen change in the context, which results in further adaptation.

For that reason, in the next section, we extend the definition of a SAS to context-altering SAS based on the mapping of our taxonomy for self-adaptation to the survey of SAS approaches.

4. A new Perspective on Self-Adaptive Systems and Research Challenges

The survey of approaches showed that the inclusion of context in most approaches is not sufficient. Whereas most approaches monitor the context, explicit alteration of context is not included in many approaches and the environment remains uncontrollable for the adaptation logic [310]. This can lead to undesired adaptation results. Therefore, we propose explicit integration of context alteration into the reasoning process. In this section, we include context adaptation for SASs and extend the definition of a SAS to a context-altering SAS. Furthermore, we will present different research issues related to the implementation of SASs.

A SAS is composed of managed resources and the adaption logic [19]. The surrounding context is used as input for the analyzing process. For a SAS, the context is not controlled by the adaptation logic [310], i.e., it is not included in the planning activity. So, context alteration is not explicitly included in the reasoning process. Furthermore, the missing control of the context can lead to undesired change in the context through the managed resources’ actuators and result in interferences [20]. In our understanding of SASs, the possibility to change the context is explicitly integrated. Derived from the taxonomy presented in Section 2 – there, context is a level where adaptation can happen – we propose the explicit integration of context for planning. Modeling context and context-altering capabilities as a construct enables reasoning about it and enables the control of context adaptation by the adaptation logic. So, the adaptation logic explicitly manages the context. Therefore, we extend the definition of a SAS and integrate context-adapting and context-altering, respectively:

“A Self-Adaptive System is able to modify its behavior or its environment in response to changes in its operating environment. The operating environment includes anything observable, such as end-user input, hardware devices, surrounding context, or program instrumentation.”

This new perspective on SASs leads to a new definition of the system border of a SAS (compared to Section 3.1). For a context-altering SAS, the system is able to change either its behavior through parameter or structural adaptation of the managed resources or the context. This can result in additional dynamics for the adaptation compared to a common SAS. Therefore, we include control of the context into the SAS and extend the SAS to a triple $SAS = (AL, MR, C)$ with the technical system – composed of the adaptation logic $AL$ and the managed resources $MR$ – and the context $C$. Different context variables, e.g., temperature, noise, or light level, define the context. They are influenced via actuators of the managed resources, controlled by the adaptation logic. Therefore, the context is modeled as $C = [c_1, ..., c_n]$ where each $c_i$ symbolizes a context variable, e.g., the temperature. The user is not integrated because the user should not be adapted. Figure 4 shows the principle layout of our view on SASs. The dashed line shows the system border. This new view on SASs leads to different issues.

**Context adaptation.** The inclusion of context adaptation controlled by the adaptation logic leads to research issues that are not relevant in the traditional view of SAS, where the environment is not controlled by the adaptation logic [310]. Different issues for context reasoning, e.g., unreliability of sensor information, are present [34]. These issues are valid for context-altering SASs, as well as for analyzing activities of common SASs. So far, context is mainly included for analysis purposes in SASs. For a context-altering SAS, the context must be integrated into planning, i.e., information about the context-altering capabilities of the technical resources is necessary for planning. The reasoning part of the adaptation logic must know all adaptation possibilities [318]. This implies that the actuators of managed resources for changing context variables must be integrated in the reasoning process and the context
variables must be modeled accordingly [318]. Therefore, existing context modeling and reasoning techniques (e.g., see [34]) must be integrated in approaches for SAS for building context-altering SAS.

Decentralization of the adaptation logic. As already mentioned in Section 3.1.3 and in the taxonomy in Section 2.5, one issue of implementing SASs is the decentralization of the adaptation logic [26]. Different authors already addressed issues of a decentralized adaptation logic (e.g., [29], [19], [101], [194], [179]). Nevertheless, there are open challenges: partial shared knowledge, coordination in the MAPE functions, uncertainty, conflicting goals, system-wide assurance & verification, overhead for coordination, and missing standardization [19, 29]. Research needs to tackle these challenges. Furthermore, we propose systematic research on factors that influence the degree of decentralization of the adaptation logic, as well as which interaction pattern should be used [308].

The inclusion of context alteration introduces further challenges. In a distributed system, different elements share the context. Having a decentralized adaptation logic, the problem of uncontrollable context adaptation can still happen, even if context adaptation is included in the reasoning process, e.g., if two subsystems A and B both have an adaptation logic that do not interact, but an adaptation of subsystem A influences the context in a negative way for subsystem B. Therefore, decentralization of the adaptation logic can lead to a fragmentation of context(-altering) information. This must be addressed, e.g., by exchange of information about context-altering capabilities between the different adaptation logic elements or context needs, respectively or with frameworks, such as the COMITY framework [20]. Furthermore, distributed monitoring of the context can be an issue as different parts of a distributed system can have different types of sensors and gain different information. An approach can be the implementation of a context broker, that distributes the context information within the adaptation logic, e.g., implemented similar to the information sharing pattern [19].

Proactive adaptation. Proactive adaptation is one of the temporal dimensions for adaptation (see Section 2.1). The aim is to adapt before it becomes necessary based on prediction. From the users point of view, this is preferable, as it reduces interruptions and adaptations can be optimized for a sequence of events [10, 318]. Further, proactive adaptation includes context adaptation via actuators in order to avoid unwanted situations. However, proactive adaptation has several remaining challenges, especially in scenarios of multiple applications or systems that share context. First, it is very complex to develop without suitable frameworks, which do not exist at this point. Second, it is highly dependent on the correctness of the predictions, as faulty predictions can cause suboptimal adaptations. The major challenges here are predicting the time of an event with high enough accuracy, as well as predicting user behavior and rare events. Third, as mentioned before, adaptations from independent sources can cause oscillating adaptations. This problem is especially prevalent for proactive adaptation, as additional adaptations would cancel its benefits. Hence, coordination is very important for proactive adaptation. As the survey has shown, so far most approaches for adaptive systems focus on reactive adaptation. Integration of proactive adaptation into SASs is a challenging task that needs further research.

Further challenges for SAS development. Further challenges can be found in literature. Integration of the user in the adaptation process, i.e. human-in-the-loop integration, gets increasing importance within the community. Central issues for runtime models are the integration of different model representations, how to use known techniques from design models at runtime, reversible model transformations, or modeling the relation between runtime models and
the represented architecture [82, 319]. Due to uncertainty, aspects of requirements engineering must be performed at run-time ([140], [320]). Verification & validation (V&V) techniques need to handle the dynamics of requirements as well as uncertainty ([168], [26]). SE processes have to be adapted for SAS development. Especially the shift of activities from development time to runtime is challenging [310]. This has been addressed in the second Dagstuhl seminar on SE for SASs [13]. Reusability of processes as well as components is addressed in the FESAS project [308]. A detailed presentation of these challenges would go beyond the scope of this paper.

5. Conclusion

In this paper, we presented a taxonomy for self-adaptation, a survey on engineering approaches for SASs, and a new perspective on SASs, the context-altering SAS. Based on literature research and combination of existing surveys, the taxonomy describes self-adaptation in the dimensions time, reason, technique, level, and adaptation control. The adaptation logic, which controls the adaptation, must be appropriately designed. Developers need to define the approach, adaptation decision criteria, and degree of decentralization, as well as the integration of the self-adaptation dimensions for monitoring and reasoning. For building SASs, different approaches can be found in literature. We classified the approaches in different categories: model-based, architecture-based, reflection, programming paradigms, control theory, service-oriented, agent-based, nature-inspired, formal modeling & verification, learning, and requirements-oriented. These categories can overlap, e.g., some architecture-based approaches use architectural models for reasoning and CBD for structural adaptation. The taxonomy showed that most approaches focus on reactive adaptation without the integration of context adaptation. This motivates our new view on SASs. Whereas in a common SAS the context is only monitored, for a context-altering SAS we assume that the adaptation logic is aware of actuators and uses them for adapting the context. The new context-altering capabilities introduce additional challenges, such as the integration of context alteration into the reasoning process. These challenges, such as integration of the context for analyzing and planning, decentralization of context monitoring and reasoning on context, and proactivity of adaptation should be addressed within the SEAMS community for easier development of SAS that uses the potential of context adaptation.

Appendix A. Overview of the Approaches Presented in the Survey

Table A.3 shows an overview of the most relevant approaches that we presented in Section 3.2 and shows their relation to MAPE activities, the taxonomy of Section 2, and the issues for implementing an adaptation logic of Section 3.1.

Within the table, we used the following abbreviations:

- M/A/P/E = Monitoring, Analyzing, Planning, Executing
- React = Reactive; Proact = Proactive;
- Ctx = Context; TR = Technical resources; U = User(s);
- App = Application; Comm = Communication; Sys = System Software;
- Tec = Technique; Par = Parameter; Str = Structure;
- Appr = Approach; Ext = External; Int = Internal;
- DC = Decision Criteria;
- DDec = Degree of Decentralization; Hyb = Hybrid; Dec = Decentralized; Cen = Centralized

If a cell is marked with "-", this means that the approach does not have a specific requirement or that no further information is given.
<table>
<thead>
<tr>
<th>Approach</th>
<th>MAPE</th>
<th>Time</th>
<th>Reason</th>
<th>Level</th>
<th>Tec</th>
<th>Adaptation Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUSIC [23, 68, 69]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>App/</td>
<td>Par/</td>
<td>Goals, - Policies,</td>
</tr>
<tr>
<td>Meta-models/Megamodels [74–76, 84]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Tra/</td>
<td>Tr</td>
<td>Utility</td>
</tr>
<tr>
<td>MechatronicUML [70, 71]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>TR/</td>
<td>Par/</td>
<td>Models, Hyb/ Utility</td>
</tr>
<tr>
<td>Rainbow Framework [35]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>App/</td>
<td>Par/</td>
<td>Goals, - Models</td>
</tr>
<tr>
<td>3L Approach [38]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Tr/</td>
<td>Par/</td>
<td>Policies</td>
</tr>
<tr>
<td>Architectural Run-time Configuration Manager [94]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>App/</td>
<td>Par/</td>
<td>Policies, - Models</td>
</tr>
<tr>
<td>Archstudio [1, 36]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Tr/</td>
<td>Par/</td>
<td>Models</td>
</tr>
<tr>
<td>Introspection [12]</td>
<td>M/A</td>
<td>-</td>
<td>Ctx/</td>
<td>TR/</td>
<td>U</td>
<td>Both</td>
</tr>
<tr>
<td>Intercession [12]</td>
<td>P/E</td>
<td>-</td>
<td>-</td>
<td>All</td>
<td>All</td>
<td>-</td>
</tr>
<tr>
<td>Reflection Reference Model [125]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Tr/</td>
<td>Par/</td>
<td>Ext (Meta) - Models</td>
</tr>
<tr>
<td>FORMS</td>
<td>See category &quot;Formal Modelling and Verification Approaches&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflex [126]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>App/</td>
<td>Par/</td>
<td>Ext (Meta) - Models</td>
</tr>
<tr>
<td>Kava [134]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>App/</td>
<td>Par/</td>
<td>Ext (Meta) - Models</td>
</tr>
<tr>
<td>Reflective Middleware [135, 136, 139]</td>
<td>P/E</td>
<td>-</td>
<td>-</td>
<td>Sys/</td>
<td>Par/</td>
<td>Ext (Meta) - Models</td>
</tr>
<tr>
<td>CARISMA [139]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Tr/</td>
<td>Par/</td>
<td>Ext Policies Dec/ Hyb</td>
</tr>
<tr>
<td>Component-Based SE [12, 37, 38, 86, 96, 111, 131, 144–151]</td>
<td>P/E</td>
<td>-</td>
<td>-</td>
<td>App/</td>
<td>Par/</td>
<td>Ext All All</td>
</tr>
<tr>
<td>Aspect-Oriented Programming [21, 41, 153, 154]</td>
<td>P/E</td>
<td>-</td>
<td>-</td>
<td>App/</td>
<td>Par/</td>
<td>Ext All All</td>
</tr>
<tr>
<td>Generative Programming [12, 158, 159]</td>
<td>P/E</td>
<td>-</td>
<td>-</td>
<td>App/</td>
<td>Par/</td>
<td>Ext All All</td>
</tr>
<tr>
<td>Autonomic Computing [2]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>Ctx/</td>
<td>Par/</td>
<td></td>
</tr>
<tr>
<td>Autonomic Communication [15]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>Ctx/</td>
<td>Par/</td>
<td></td>
</tr>
<tr>
<td>Control Loop Patterns [178]</td>
<td>All</td>
<td>React</td>
<td>Ctx/</td>
<td>Sys/</td>
<td>Par/</td>
<td></td>
</tr>
<tr>
<td>Control Loop UML Profile [179]</td>
<td>All</td>
<td>React</td>
<td>TR/</td>
<td>Comm/</td>
<td>Par/</td>
<td></td>
</tr>
</tbody>
</table>

22
<table>
<thead>
<tr>
<th>Approach</th>
<th>MAPE</th>
<th>Time</th>
<th>Reason</th>
<th>Level</th>
<th>Tec</th>
<th>Adaptation Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Theory Foundation [180]</td>
<td>All</td>
<td>React TR/ Ctx</td>
<td>Sys Par</td>
<td>Ext</td>
<td>Models -</td>
<td></td>
</tr>
<tr>
<td><strong>Service-oriented</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUSIC</td>
<td>All</td>
<td>React Ctx/ TR</td>
<td>App Str</td>
<td>Ext</td>
<td>Goals, Dec/ Utility Hyb</td>
<td></td>
</tr>
<tr>
<td>SASSY Framework [87, 186]</td>
<td>All</td>
<td>React Ctx/ TR</td>
<td>App Str</td>
<td>Ext</td>
<td>Policies, All Rules</td>
<td></td>
</tr>
<tr>
<td>MetaSelf [187]</td>
<td>All</td>
<td>React Ctx/ TR</td>
<td>App Str</td>
<td>Ext</td>
<td>Models Cen</td>
<td></td>
</tr>
<tr>
<td>QoSMOS [188]</td>
<td>All</td>
<td>React Ctx/ TR</td>
<td>App Str</td>
<td>Ext</td>
<td>- - -</td>
<td></td>
</tr>
<tr>
<td>Aspect-oriented and Service-oriented Computing [155, 157]</td>
<td>P/E</td>
<td>- -</td>
<td>App Str</td>
<td>Ext</td>
<td>- - -</td>
<td></td>
</tr>
<tr>
<td>Agent systems and SOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component models and services [149, 190]</td>
<td>All</td>
<td>React TR App Str</td>
<td>Ext</td>
<td>- - -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CARE method [191]</td>
<td>A/P/E</td>
<td>React Ctx/ TR</td>
<td>App Str</td>
<td>Ext</td>
<td>Goals, Cen Utility</td>
<td></td>
</tr>
<tr>
<td>MOSES Framework [193]</td>
<td>All</td>
<td>React TR App Str</td>
<td>Ext</td>
<td>Goals, Cen Utility</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Agent-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOCAS [47]</td>
<td>All</td>
<td>React Ctx Sys Par</td>
<td>-</td>
<td>Policies Dec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design Patterns [101]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>Models Dec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent-Based Modeling, Dynamical Systems Analysis, and Decentralized Control [194]</td>
<td>All</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>Models Dec/ Hyb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nature-inspired</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimization [205, 206, 208]</td>
<td>P</td>
<td>React TR/ Ctx</td>
<td>Sys Par/ Str</td>
<td>-</td>
<td>Utility Dec</td>
<td></td>
</tr>
<tr>
<td>Autonomic Computing for Pervasive ICT [209]</td>
<td>P</td>
<td>React TR App/ Sys Str</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>Frequency Planning [210, 211]</td>
<td>P</td>
<td>React TR App Par</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>Data Harvesting [212]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>Utility Dec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Synchronicity [215]</td>
<td>P</td>
<td>React TR Sys Par</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>Region Detection [216]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>Immune System [217]</td>
<td>P</td>
<td>React Ctx Sys Par</td>
<td>-</td>
<td>-</td>
<td>Dec/ Hyb</td>
<td></td>
</tr>
<tr>
<td>Obstacle Avoidance [219]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>Int Utility</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential Fields [213, 220, 221]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>-</td>
<td>Dec/ Hyb</td>
<td></td>
</tr>
<tr>
<td>Task Assignment [222]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>Utility Dec</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecosystem Framework [225]</td>
<td>-</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>Utility Dec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAPPID [226]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>Social Conventions [227]</td>
<td>P</td>
<td>React Ctx App Par</td>
<td>-</td>
<td>-</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td><strong>Formal Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FORMS [33]</td>
<td>All</td>
<td>React TR/ Ctx TR</td>
<td>Sys Par/ Str</td>
<td>Ext</td>
<td>Models -</td>
<td></td>
</tr>
<tr>
<td>Restore-Invariant Approach [228, 229, 233]</td>
<td>A</td>
<td>React TR TR Str</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Markov Chains [244, 245]</td>
<td>A, P</td>
<td>React TR/ Ctx TR Par Str</td>
<td>Ext</td>
<td>Models Cen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach</td>
<td>MAPE</td>
<td>Time</td>
<td>Reason Level</td>
<td>Tec</td>
<td>Adaptation Control</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------</td>
<td>-------</td>
<td>--------------</td>
<td>------</td>
<td>-------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Abstract State Machines [246]</td>
<td>All</td>
<td>React</td>
<td>TR / TR</td>
<td>Par / Str</td>
<td>Ext Models All</td>
<td></td>
</tr>
<tr>
<td>Timed Hazard Analysis [247]</td>
<td>A/P</td>
<td>React</td>
<td>TR / TR</td>
<td>Str</td>
<td>Ext Models -</td>
<td></td>
</tr>
<tr>
<td>MAPE-K Formal Templates [248]</td>
<td>All</td>
<td>React</td>
<td>TR / TR</td>
<td>Par / Str</td>
<td>Ext Models -</td>
<td></td>
</tr>
<tr>
<td>ActiveFORMS [251, 252]</td>
<td>All</td>
<td>React</td>
<td>TR / App / U</td>
<td>Par / Str</td>
<td>Ext Models Cen</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unity [197]</td>
<td>P</td>
<td>React</td>
<td>TR / Sys / U</td>
<td>Par</td>
<td>- Goals, Dec</td>
<td></td>
</tr>
<tr>
<td>FUSION [42, 255]</td>
<td>P</td>
<td>React</td>
<td>TR / App / Ctx / Sys / U</td>
<td>Par / Str</td>
<td>- Utility Dec</td>
<td></td>
</tr>
<tr>
<td>Reinforcement Learning [163, 257, 260–264]</td>
<td>P</td>
<td>React</td>
<td>TR / App / Ctx / Sys / U</td>
<td>Par / Str</td>
<td>Ext Utility All</td>
<td></td>
</tr>
<tr>
<td>Control-Based Framework [274]</td>
<td>P, A, P</td>
<td>React</td>
<td>Ctx / App / Ctx / Sys / U</td>
<td>Par / Str</td>
<td>Ext Utility All</td>
<td></td>
</tr>
<tr>
<td>Fossa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBSE [87, 100, 278–280, 282–288]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LoREM [48]</td>
<td>All</td>
<td>React</td>
<td>- / App / Ctx / TR / U</td>
<td>Par / Str</td>
<td>- Goals, Dec</td>
<td></td>
</tr>
<tr>
<td>Requirements@Runtime [293]</td>
<td>A/P</td>
<td>React</td>
<td>- / App / Ctx / TR / U</td>
<td>Par / Str</td>
<td>- Goals, Dec</td>
<td></td>
</tr>
<tr>
<td>Zanshin [294]</td>
<td>All</td>
<td>React</td>
<td>Ctx / App / Ctx / TR / U</td>
<td>Par / Str</td>
<td>- Goals, Dec</td>
<td></td>
</tr>
<tr>
<td>Further</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task-based adaptation [98, 99, 303]</td>
<td>A/P</td>
<td>React</td>
<td>- / App / Ctx / TR / U</td>
<td>Par / Str</td>
<td>Ext All Dec / Hyb</td>
<td></td>
</tr>
<tr>
<td>Middleware-centric adaptation [21, 22, 130]</td>
<td>All</td>
<td>React</td>
<td>Ctx / App / Ctx / TR / U</td>
<td>Par / Str</td>
<td>Ext All Dec / Hyb</td>
<td></td>
</tr>
<tr>
<td>Adaptation languages [304–307]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processes [88, 308, 310–313]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling Dimensions [25, 309]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design patterns [19, 100–103, 200, 218, 308, 314–317]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.3: Overview of the Approaches of the Survey.

References

URL: http://www.sei.cmu.edu/productlines/


[118] M. Wermelinger, A. Lopes, J. L. Fiadeiro, A Graph Based Architectural (Re)configuration Language, ACM SIGSOFT Software Engineering