Int J Softw Tools Technol Transfer (2018) 20:211–235
https://doi.org/10.1007/s10009-017-0448-3

REGULAR PAPER

Automated translation of VDM to JML-annotated Java

Peter W. V. Tran-Jørgensen[1] · **Peter Gorm Larsen**[1] · **Gary T. Leavens**[2]

Published online: 11 February 2017

Abstract When a system specified using the Vienna Development Method (VDM) is realised using code-generation, no guarantees are currently made about the correctness of the generated code. In this paper, we improve code-generation of VDM models by taking contract-based elements such as invariants and pre- and postconditions into account during the code-generation process. The contract-based elements of the Vienna Development Method Specification Language (VDM-SL) are translated into corresponding constructs in the Java Modelling Language (JML) and used to validate the generated code against the properties of the VDM model. VDM-SL and JML are both Design-by-Contract (DbC) languages, with the difference that VDM-SL supports abstract modelling and system specification, while JML is used for detailed specification of Java classes and interfaces. We describe the semantic differences between the contract-based elements of VDM-SL and JML and formulate the translation as a set of rules. We further demonstrate how dynamic JML assertion checks can be used to ensure the consistency of VDM's subtypes when a model is code-generated. The translator is fully automated and produces JML-annotated Java programs that can be checked for correctness using JML tools.

Keywords Design-by-Contract · Formal methods · VDM · Java · JML · Code-generation

✉ Peter W. V. Tran-Jørgensen
peter.w.v.jorgensen@gmail.com; pvj@eng.au.dk

Peter Gorm Larsen
pgl@eng.au.dk

Gary T. Leavens
leavens@cs.ucf.edu

[1] Aarhus University, Aarhus, Denmark

[2] University of Central Florida, Orlando, FL, USA

1 Introduction

Design-by-Contract (DbC) is an approach for designing software based on concepts such as preconditions, postconditions and invariants [1]. These concepts are referred to as "*contracts*", according to a conceptual metaphor for the conditions and obligations of a business contract. An example of a formal method that uses DbC elements is the Vienna Development Method (VDM), which was originally developed at IBM in Vienna for the development of a compiler for PL/1 [2–4]. One way to realise a VDM specification in a programming language is through refinement [5]. This is a stepwise process by which one can transform a formal model into a program that can be verified to semantically satisfy its contracts [6].

Another way to realise a VDM specification is using code-generation. The idea is for the generated code to be a refinement of the specification, but which is not achieved through stepwise refinement, but rather in one step through code-generation translation rules. Code-generation aims to reduce the resources needed to realise the model as well as to avoid introducing problems in the implementation due to manual translation of model into code. However, current VDM code-generators do not make any guarantees about the correctness of the generated code, nor do they provide the necessary means to help check that the code meets the specification. Naturally, this casts doubt on the value of code-generation as a way to realise a VDM model, since the goal is to develop software that meets the specification.

In this paper, we improve code-generation of VDM models by allowing the generated code to be checked against the system properties described by the VDM contracts. This helps ensure that the generated code meets the VDM specification, and is achieved as described in this paper.

Some DbC technologies are tailored to specify detailed designs of programming interfaces for a particular

programming language [7]. An example of one such technology is the Java Modelling Language (JML) [8]—a formal specification language that uses DbC elements, written as specialised comments, to specify the behaviour of Java classes and interfaces. JML annotations can be analysed statically or checked dynamically using JML tools. Therefore, JML can be seen as a technology that serves to bridge the gap between an abstract system specification and its Java implementation.

In this paper, we attempt to bridge this gap even further by proposing a way to automatically translate a specification written in the Vienna Development Method Specification Language (VDM-SL) to a JML-annotated Java implementation. Current VDM code-generators either ignore or provide limited code-generation support for the contract-based elements and type constraints of VDM. Ideally we should be able to preserve the contracts and type constraints when the system specification is implemented, since (1) they serve to document the intention and properties of the system and (2) they can be used to check the system realisation for correctness. Ensuring that the contracts and type constraints, as originally specified in VDM, hold for the system implementation potentially requires many extra checks to be added to the code. Adding these checks to the code manually is tedious and prone to errors. Instead, these checks could be generated automatically. Representing contracts and type constraints in JML also has the advantage that these checks may be ignored by the Java compiler. This allows the system realisation to be executed without the overhead of checking the contracts and type constraints, if desired.

The two main contributions of our work are (1) a collection of semantics-preserving rules for translating a VDM-SL specification to a JML-annotated Java program and (2) an implementation of these rules as an extension to Overture's [9,10] VDM-to-Java code-generator [11].

The rules propose ways to translate the DbC elements of VDM-SL to JML annotations; these annotations are added to the Java code produced by Overture's Java code-generator. The rules cover checking of preconditions, postconditions and invariants, but the translator also produces JML checks to ensure that no type constraints are violated across the translation. We present the rules one by one and demonstrate, using a case study model of an Automated Teller Machine (ATM), how the code-generator extension translates a VDM-SL specification to JML-annotated Java code.

Since the translation is not formally defined we have used the OpenJML [12] runtime assertion checker to validate our work—in particular by generating JML constructs supported by this tool. More specifically, the JML translator has been tested by running examples through the tool in order to validate each of the translation rules (see Sect. 8 for more details).

Following this section, we describe DbC with VDM-SL and JML in Sect. 2. We continue by presenting the implementation of the JML translator in Sect. 3. Then we describe the rules used to translate a VDM-SL specification to a JML-annotated Java program in Sects. 5, 6 and 7. Next, we assess the correctness of the translation in Sect. 8. Finally, we describe related work in Sect. 9 and present future plans and conclude in Sect. 10.

2 DbC with VDM-SL and JML

In this section, we describe VDM-SL and JML. We cover different types and all the contract-based elements of VDM-SL, focusing specifically on the VDM-10 release, which we are targeting in our work. The JML constructs described in this section cover those that are used to implement the translation rules.

2.1 VDM-SL

VDM-SL is an ISO standardised sequential modelling language that supports description of data and functionality. The ISO standard has later been informally extended with modules to allow type definitions, values (constants) and functionality to be imported and exported between modules. A module may define a single state component, which can be constrained by a state invariant. State is modified by assigning a new value to a state designator, which can be either a name, a field reference or a map or sequence reference, as described in the VDM language reference manual [13].

Module state, if specified, implicitly defines a record type, which is tagged with the state name and also defines the type of the state component. The state type can be used like any other record type explicitly defined by the modeller—the difference being that the state invariant [14] constrains the state type and thus every instance of this record type.

Data are defined by means of built-in basic types covering, for instance, numbers, booleans, quote types and characters. A quote type corresponds to an enumerated type in a language such as Pascal. The basic types can be used to form new structured data types using built-in type constructors that support creation of union types, tuple types and record types. A type may also be declared optional, which allows **`nil`** to be used to represent the absence of a value. For collections of values, VDM-SL supports sets, sequences and maps. The built-in data types, type constructors and collections can be used to form named user-defined types, which can be constrained by invariants. We refer to these types as *named invariant types*. As an example, Listing 1 shows the definition of the named invariant type `Amount`, which is used to represent an amount of money deposited or withdrawn by an account holder. This type is defined based on natural numbers (excluding zero), i.e. the built-in basic type **`nat1`** in VDM-SL. For this particular example, we say that **`nat1`** is the *domain type* of

Amount. We further constrain Amount using an invariant, by requiring a value of this type to be <2000. Specifically, for the invariant shown in Listing 1, the a on the left-hand side of the equality is a pattern that matches values of type Amount. This pattern is used to express the invariant predicate for this named invariant type.

```
types
Amount = nat1
inv a == a < 2000;
```

Listing 1 Example of a VDM-SL named invariant type.

2.1.1 Functional descriptions

In VDM, functionality can be defined in terms of functions and operations over data types with a traditional call-by-value semantics. Functions are referentially transparent, and therefore, they are not allowed to access or manipulate state directly, whereas operations are. Therefore, a function cannot call an operation.[1] In addition to accessing module state, operations may also use the **dcl** statement to declare local state designators which can be assigned to. Subsequently the term *functional description* will be used to refer to both functions and operations. As an example, a function that uses the MATH`sqrt library function to calculate the square root of a real number is shown in Listing 2.

```
sqrt : real -> real
sqrt (x) == MATH`sqrt(x)
pre x >= 0
post RESULT * RESULT = x;
```

Listing 2 VDM-SL function for calculating the square root of a number.

Functional descriptions can be implicitly defined in terms of pre- and postconditions, which specify conditions that must hold before and after invoking the functional description. Alternatively, a functional description can be *explicitly* defined by means of an algorithm, as shown in Listing 2. The JML translator supports both implicitly and explicitly defined functional descriptions. However, only methods that originate from explicitly defined functional descriptions can be executed.

The precondition of a function can refer to all the arguments of the function it guards. The same applies to the postcondition of a function, which can also refer to the result of the execution using the reserved word **RESULT**. For the square root function in Listing 2, we require that the input is a positive number (the precondition) and that the square of the function result equals the input value (the postcondition).

Function definitions are derived for the pre- and postconditions of sqrt from sqrt's **pre** and **post** clauses. These function definitions do not appear in the model, but they are used internally by the Overture interpreter to check for contract violations. However, to clarify, the pre- and postcondition functions of sqrt are shown in Listing 3. In this listing, +> specifies that pre_sqrt and post_sqrt are total functions, and not partial functions, which use the -> type constructor.

```
pre_sqrt:real +> bool
pre_sqrt(x) == x >= 0;

post_sqrt:real*real +> bool
post_sqrt(x,RESULT) == RESULT *
    RESULT = x;
```

Listing 3 Pre- and postcondition functions for the sqrt function shown in Listing 2.

Similarly, the pre- and postcondition functions of an operation are also derived. To demonstrate this, consider the inc operation in Listing 4. This operation takes a real number as input, adds it to a counter (defined using a state designator), and returns the new counter value. In this listing, counter~ and counter refer to the counter values before and after the operation has been invoked, respectively.

```
inc : real ==> real
inc (i) == (
 counter := counter + i;
 return counter;
)
pre i > 0
post counter = counter~ + i and
     RESULT = counter;
```

Listing 4 VDM-SL operation for incrementing a counter.

A precondition of an operation can refer to the state, s, before executing the operation, whereas the postcondition of an operation can read both the before and after states. State access is achieved by passing copies of the state to the pre- and postcondition functions. The corresponding pre- and postcondition functions for inc are shown in Listing 5 where the parameters s~ and s of post_inc refer to the state (that contains the counter value) before and after execution of inc. We further use S to denote the record type that represents the module's state.

[1] With the recent introduction of **pure** operations into VDM-10 (not to be confused with **pure** methods in JML), it has become possible to invoke operations, albeit **pure** ones, from a function. This feature was introduced to address issues with the object-oriented dialect of VDM, called VDM++, but was made available in every VDM-10 dialect (including VDM-SL).

```
pre_inc:real*S +> bool
pre_inc(i,s) == i > 0;

post_inc:real*real*S*S +> bool
post_inc(i,RESULT,s~,s) ==
s.counter = s~.counter + i and
RESULT = s.counter;
```

Listing 5 Pre- and postcondition functions for the `inc` operation shown in Listing 4.

The function descriptions in Listing 5 assume that the pre- and postconditions are defined (using the **`pre`** and **`post`** clauses) and that the state of the module enclosing the functional description exists. For the cases where pre- and postconditions are not defined, they can be thought of as functions that yield **`true`** for every input. Furthermore, when no state component is defined, the pre- and postcondition functions simply omit the state parameters. Similarly, when an operation does not return a result (it specifies void as the return type), the postcondition function omits the **`RESULT`** parameter.

For each type definition constrained by an invariant, such as `Amount` shown in Listing 1, a function is implicitly created to represent the invariant—see Listing 6. The Overture tool uses this function internally to check whether a value is consistent with respect to a given type (e.g. `Amount`) [15]. Note that since all invariants are functions, they are not allowed to depend on state of other modules. Specifically, invariants can only invoke functions and access global constants (possibly defined in other modules).

```
inv_Amount : Amount +> bool
inv_Amount (a) == a < 2000;
```

Listing 6 Invariant function for type definition `Amount`.

2.1.2 Atomic execution

Multiple consecutive statements are sometimes needed to update the state designators to make them consistent with the system's invariants. For example, assume that we have a system that uses two state designators called $\texttt{evenID}_1$ and $\texttt{evenID}_2$ to store even and different numbers. For this example, we will assume that these state designators are of type `Even`—a type that constrains these state designators to store even numbers. To help ensure that the uniqueness constraint (a state invariant) is not violated during an update, multiple assignments can be grouped in an **`atomic`** statement block as shown in Listing 7. Given the type `Even` of the state designators $\texttt{evenID}_1$ and $\texttt{evenID}_2$, it is as if the atomic statement is evaluated as shown in Listing 8.

```
atomic (
 evenID1 := exp1;
 evenID2 := exp2;
)
```

Listing 7 Atomic update in VDM.

```
let t1 : Even = exp1,
    t2 : Even = exp2
in (
 -- Turn off invariants
 evenID1 := t1;
 evenID2 := t2;
 -- Turn on invariants
 -- Check invariants hold
);
```

Listing 8 The execution semantics of the **`atomic`** statement.

Executing the **`atomic`** statement block is semantically equivalent to first evaluating the right-hand sides of all the assignments before turning off invariant checks, and then binding the results to the corresponding state designators. After all the assignments have been executed, it must be ensured that all invariants hold.

There are three properties that follow from the evaluation semantics of the **`atomic`** statement block that are worth mentioning:

1. When evaluating the right-hand sides of the assignment statements, potential contract violations will be reported.
2. Temporary identifiers, used to store the right-hand side results, are explicitly typed and therefore violations of named invariant types for these variables will be reported. The explicit type annotations thus ensure that the right-hand side of a state designator assignment is checked to be consistent with the type of said state designator.
3. Assignment statements cannot see intermediate values of state designators.

2.2 JML

Although JML [16] is designed to specify arbitrary sequential Java programs, in this subsection we only describe the features needed for the translation from VDM-SL.

A method specified with the **`pure`** modifier in JML is not permitted to have write effects; such methods are allowed to be used in specifications. Pure methods are used to translate VDM-SL functions.

A class invariant in JML should hold whenever the non-helper methods of that class are not being executed; thus invariants must hold in each method's before and after states. However, a method declared with the **`helper`** annotation in

a type `T` does not have its pre- and postconditions augmented with `T`'s invariants. Helper methods (and constructors) must either be pure or private [16], so that the invariant will hold at the beginning and end of all client-visible methods [17]. The before and after states of non-helper methods and constructors are said to be *visible states*; thus invariants must hold in all visible states. JML distinguishes between instance and static invariants. An *instance* invariant can refer to the non-static (i.e. instance) fields of an object. A *static* invariant cannot refer to an object's non-static fields; thus static invariants are used to specify properties of static fields.

An assertion can reference the invariant for an object explicitly using a predicate of the form `\invariant_for(e)`, which is equivalent to the invariant for `e`'s static type [11, section 12.4.22].

In JML, pre- and postconditions are written using the keywords **`requires`** and **`ensures`**, respectively. In the specification of a postcondition, one writes `\old(e)` to refer to the before state value of an expression `e`. For example, an `increment` method that writes a field `count` could be specified as shown in Listing 9.

```
1 //@ requires count < Integer.MAX_VALUE;
2 //@ modifies count;
3 //@ ensures count == \old(count)+1;
4 void increment() {
5   count++;
6 }
```

Listing 9 Example of a JML specification for a Java method.

Method postconditions may also use the keyword **`\result`** to refer to the value returned by the method.

Specification expressions in JML can use Java expressions that are pure (have no write effects), and also some logical operators, such as implication `==>`, and quantifiers such as **`\forall`** and **`\exists`**.

In addition to method pre- and postconditions, one can also write assertions anywhere a Java statement can appear, using JML's **`assert`** keyword. Such assertions must hold whenever they are executed.

One way to specify the abstract state of a class is to use JML's **`ghost`** variables. Ghost variables are specification-only variables and fields of objects that can only be used in JML specifications and in JML **`set`** statements. A set statement is an assignment statement whose target is a ghost variable.

By default, JML variables and fields may not hold the **`null`** value. However, should one wish to specify that all fields of a class may hold **`null`**, then one can annotate the class's declaration with **`nullable_by_default`**.

3 The implementation of the JML translator

The JML translator is implemented as an extension to Overture's VDM-SL-to-Java code-generator, which provides code-generation support for a large executable subset of VDM. This section describes how the JML translator has been implemented, and explains the details of the Java code-generator that are needed in order to understand how the JML translator works.

3.1 The implementation

The Java code-generator is developed using Overture's code-generation platform—a framework for constructing code-generators for VDM [11]. This platform is used by the Java code-generator to parse the VDM-SL model sources and to construct an Intermediate Representation (IR) of the model—an Abstract Syntax Tree (AST) that constitutes an internal representation of the generated code. The Java code-generator uses the code-generation platform to *transform* the IR into a tree structure that eventually is translated directly into Java code. The translation of the IR into Java is handled by the code-generation platform's code emission framework, which uses the Apache Velocity template engine [18].

The Java code-generator exposes the IR during the code-generation process, which allows the JML translator to intercept the code-generation process and further transform the IR. These additional transformations are used to decorate the IR with nodes that contain the JML annotations. Using the code emission framework, the final version of the IR is translated into a JML-annotated Java program.

The JML translator is publicly available in Overture version 2.3.8 (as of July 2016) onwards [10]. Furthermore, the JML translator's source code is available via the Overture tool's open-source code repository [19].

3.2 Overview of the translation

In the generated code, a module is represented using a **`final`** Java class with a **`private`** constructor, since VDM-SL does not support inheritance and a module cannot be instantiated. Due to the latter, both operations and functions are code-generated as **`static`** Java methods.

Module state is represented using a **`static`** class field in the module class to ensure that only a single state component exists at any given time. The state component is represented using a record value, and as a consequence, an additional record type is generated to represent it.

Each variable in VDM-SL is passed by value, i.e. as a *deep copy*, when it is passed as an argument, appears on the right-hand side of an assignment or is returned as a result. As a consequence, aliasing can never occur in a VDM-SL model. Types are different in Java, where objects are modified via

object references or pointers. Therefore different object references can be used to modify the same object. To avoid such aliasing in the generated code, data types are code-generated with functionality to support value type behaviour.

Every record definition code-generates to a class definition with accessor methods for reading and manipulating the fields. This class implements `equals` and `copy` methods to support comparison based on structural equivalence and deep copying, respectively. In this way the call-by-value semantics of VDM-SL can be preserved in the generated code by invoking the `copy` method, which helps to prevent aliasing. Similarly the `equals` method can be invoked to compare code-generated records based on structural equivalence rather than comparing addresses of object references. A record object can then be obtained by invoking the constructor of the record class or by invoking the `copy` method of an existing record object.

Java does not support the definition of aliases of existing types, such as the `Amount` named invariant type in Listing 1. Therefore, the Java code-generator chooses not to code-generate class definitions for these types. Instead, a use of a named invariant type is replaced with its domain type (described in Sect. 2.1). Since the named invariant type is an alias of an existing type this is fine, as long as we make sure to check that the type invariant holds.

To assist the translation of VDM to Java, the existing Java code-generator uses a runtime library, which, among other things, includes Java implementations for some of the different VDM types and operators. The `Tuple` class, for example, is used to represent tuple types and enables construction of tuple values. Sets, sequences and maps are represented using the `VDMSet`, `VDMSeq` and `VDMMap` classes, which themselves are based on Java collections and so on. The runtime library's collection classes are used as raw types (e.g. `VDMSet`) in the generated code, and therefore they are never passed a generic type argument. Raw types provide a convenient way to represent VDM collections that store elements of some union type—a kind of type that Java does not support.

In addition to using the existing runtime library, the JML translator also contributes a small runtime library to aid the generation of JML checks. This runtime library, which we subsequently refer to as `V2J`, is an extension of the existing Java code-generator runtime library. As we see in Sect. 6.3, the `V2J` runtime is mostly used in the generated JML checks to ensure that instances of collections respect the VDM types that produce them.

4 Case study example

Throughout the paper we will demonstrate the translation rules using a case study model of an ATM. The model consists of a single module, `ATM` (shown in Listing 10), which uses a state definition to record information about

- The debit cards considered valid by the system (named `validCards`).
- The debit card currently inserted into the ATM, if any (`currentCard`).
- If a valid PIN code has been entered (`pinOk`) for the debit card currently inserted into the ATM and,
- all the bank accounts known to the system (`accounts`).

```
1  module ATM
2  definitions
3  state St of
4   validCards : set of Card
5   currentCard : [Card]
6   pinOk : bool
7   accounts : map AccountId to
        Account
8   init St == St = mk_St({},nil,false
        ,{|->})
9   inv mk_St(v,c,p,a) ==
10   (p or c <> nil => c in set v)
11   and
12   forall id1, id2 in set dom a &
13    id1 <> id2 =>
14    a(id1).cards inter a(id2).cards
        = {}
15 end
16  ...
17 operations
18 GetStatus : () ==> bool * seq of
       char
19 GetStatus () == ...
20
21 OpenAccount : set of Card *
       AccountId ==> ()
22 OpenAccount (cards,id) == ...
23
24 AddCard : Card ==> ()
25 AddCard (c) == ...
26
27 RemoveCard : Card ==> ()
28 RemoveCard (c) == ...
29
30 InsertCard : Card ==>
31   <Accept>|<Busy>|<Reject>
32 InsertCard (c) == ...
33
34 EnterPin : Pin ==> ()
35 EnterPin (pin) == ...
36
37 ReturnCard : () ==> ()
38 ReturnCard () == ...
39
40 Withdraw : AccountId * Amount ==>
       real
41 Withdraw (id, amount) == ...
42
43 Deposit : AccountId * Amount ==>
       real
44 Deposit (id, amount) == ...
45 end
```

Listing 10 VDM-SL module representing an ATM.

For simplicity, Listing 10 omits type definitions and only shows the state definition (including the state invariant) and the signatures for some of the operations. The state invariant, shown in Listing 10, requires that at all times the following two conditions must be met: a debit card must at most be associated with a single account and secondly, for a PIN code to be considered valid, the debit card currently inserted into the ATM must itself be a valid debit card.

When the ATM model is translated to a JML-annotated Java program, it can be checked for correctness using JML tools. To demonstrate this, consider the example in Listing 11, which creates a debit card, inserts it into the ATM and performs a transaction scenario.

```
1 Card c = new Card(5,1234);
2 // atm.ATM.AddCard(c); (missing statement)
3 atm.ATM.InsertCard(c);
4 atm.ATM.EnterPin(1234);
5 System.out.println(atm.ATM.GetStatus());
6 /* Transaction related code omitted */
7 atm.ATM.ReturnCard();
```

Listing 11 Java code demonstrating use of the implementation of the ATM model.

If this program is executed using the OpenJML runtime assertion checker, the output in Listing 12 is reported.

```
Exception in thread "main" java.lang.
    AssertionError: Main.java:12: JML
    precondition is false
         atm.ATM.EnterPin(1234);
                          ^
atm/ATM.java:276: Associated
    declaration: Main.java:12:
  //@ requires pre_EnterPin(pin,St);
      ^
     at Main.main(Main.java:17)
```

Listing 12 Inconsistent use of the system detected using the OpenJML runtime assertion checker.

For this particular example, this error is reported because the debit card `c` is not recognised as a valid debit card by the system. More specifically, the scenario did not invoke `atm.ATM.AddCard(c)` immediately after creating the debit card. The return value of the `Insert` method did indicate that the debit card was rejected, but this value was mistakenly discarded in Listing 11. The error is reported by the runtime assertion checker because entering a PIN code when no debit card is inserted into the ATM is considered an error. After changing the example in Listing 11 to add `c` as a valid debit card, no problems are detected by the runtime assertion checker, as expected. Therefore, the code executes as if it was compiled using a standard Java compiler and executed on a regular Java virtual machine. More specifically, the system will report the status as shown in Listing 13 to indicate that the ATM is not awaiting a debit card and that a transaction is in progress.

```
mk_(false, "transaction in progress.")
```

Listing 13 System output after fixing the problem in Listing 11.

As we proceed, in Sects. 5 and 6 we elaborate on the specifics of each VDM definition in the case study model and demonstrate the translation to JML-annotated Java.

5 Translating VDM-SL contracts to JML

In this section, we present the rules used to translate the DbC elements of VDM-SL to JML annotations that are added to the generated Java code. For each of the elements, we describe the approach used to translate the element to JML. This is afterwards generalised as a rule, which appears in a grey box.

5.1 Allowing null values by default

Overture's Java code-generator may sometimes introduce auxiliary variables that are initialised to **`null`** when it code-generates some of the constructs of VDM. To avoid having errors reported when checking the generated code with a JML tool, we allow **`null`** as a legal value by default for all references in the generated code.

1. Allowing null values by default
Annotate every class output by the Java code-generator with the **`nullable_by_default`** modifier to allow all references to use **`null`** as a legal value.

As a consequence, we also have to guard against **`null`** values for variables that originate from VDM variables or patterns[2] that do not allow **`nil`**.

5.2 Translating functional descriptions to JML

Recall that a VDM-SL function code-generates to a **`static`** Java method. In addition, a VDM-SL function does not have side effects and therefore the code-generated version of the method can be annotated as JML **`pure`**.

2. Translation of functions
Any function – whether it is defined by the user or derived, e.g. from a **`pre`** or **`post`** condition clause – code-generates to a **`static`** Java method that is annotated with the **`pure`** modifier.

Operations, on the other hand, can read and manipulate the state of the enclosing module, or invoke other operations that may have side effects. Therefore, the method that

[2] The generated code uses variables to represent the patterns (record pattern, tuple pattern, identifier pattern, etc.) introduced by use of pattern matching in VDM.

the operation code-generates to cannot be annotated as JML **pure**.

When a VDM-SL definition (e.g. a functional description) is code-generated to Java, the visibility of the corresponding Java definition can, in principle, be set according to whether the VDM-SL definition is exported (**public**) or not (**private**). In the presentation of the translation rules following this section, we omit explicit use of access specifiers in the rule formulation as we do not consider it crucial to our work.

5.3 Translating preconditions to JML

In terms of semantics, there is no difference between a precondition in VDM-SL and JML. There are, however, interesting issues worth mentioning regarding how the JML translator implements the translation. We start by covering preconditions of operations, and we end this subsection by describing how they differ from those of functions. As an example of how a VDM-SL precondition is translated, consider the operation in Listing 14. This operation models withdrawal from a bank account identified by the parameter `id`.

```
1  Withdraw : AccountId * Amount ==>
       real
2  Withdraw (id, amount) ==
3  let newBalance =
4      accounts(id).balance - amount
5  in (
6   accounts(id).balance := newBalance
       ;
7   return newBalance;
8  )
9  pre
10 currentCard in set validCards and
       pinOk and
11 currentCard in set accounts(id).
       cards and
12 id in set dom accounts
```

Listing 14 VDM-SL operation for bank account withdrawal guarded by a precondition.

In order to withdraw money from the account, we require that a valid card has been inserted, the PIN code is accepted and the bank account exists. Note that since `currentCard` is of the optional type `[Card]`, it can be **nil**, which is not a valid member of `validCards`. Therefore, the precondition is **false** when no debit card has been inserted into the ATM. The `pre_Withdraw` function, which is not a visible part of the model, is derived from the **pre** clause of the `Withdraw` operation. In the generated code, this function is represented using a **pure** method according to rule 2—see Listing 15. Note that for the method in Listing 15, the Java code-generator uses extra variables to perform the equivalent VDM computation. These extra variables are also type-checked using JML (although they are only used to store intermediate results).

The `Withdraw` operation is translated to the method shown in Listing 16. This method introduces several JML assertions that are described in Sect. 6. Note that the `pre_Withdraw` method is invoked from the **requires** clause of the `Withdraw` method to check whether the precondition is met. In addition to the input parameters of the `Withdraw` method, the `pre_Withdraw` method is also passed the state `St`.

3. Translating the precondition of an operation

Let `op` be a method code-generated from a VDM-SL user-defined operation and let the signature of `op` be:
static `R op(`I_1 i_1`,...,`I_n i_n`)`
Then `op` has a code-generated precondition method `pre_op` that is **pure** and which in addition to the parameters of `op` also takes the state component `s` as an argument, i.e.
`/*@` **pure** `@*/` **static boolean**
`pre_op(`I_1 i_1`,...,`I_n i_n`,S s)`
To ensure that the precondition is evaluated, we annotate `op` with the following **requires** annotation:
`//@` **requires** `pre_op(`i_1`,...,`i_n`,s);`

Rule 3 assumes the existence of a state component `s`. However, when the state of the module enclosing `op` is not defined, rule 3 changes to not include the state parameter in the definition of `pre_op`.

The example above considers the case where the precondition is guarding an operation (i.e. `Withdraw`). As described in Sect. 2, a precondition is defined differently for a function than it is for an operation. In particular, the precondition of a function is not passed the state, so neither is the code-generated version of it. We also note that the visibility of the precondition function must be the same as that of the functional description it guards. Otherwise, it cannot be invoked from the corresponding **requires** clause.

4. Translating the precondition of a function

Let `f` be a method code-generated from a VDM-SL user-defined function and let the signature of `f` be:
static `R f(`I_1 i_1`,...,`I_n i_n`)`
Then `f` has a code-generated precondition method `pre_f` that is **pure** and which accepts the same parameters as `f`, i.e.
`/*@` **pure** `@*/` **static boolean**
`pre_f(`I_1 i_1`,...,`I_n i_n`)`
To ensure that the precondition is evaluated, we annotate `f` with the following **requires** annotation:
`//@` **requires** `pre_f(`i_1`,...,`i_n`);`

```
/*@ pure @*/
public static Boolean pre_Withdraw(
    final Number id, final Number amount, final atm.ATMtypes.St St) {
  //@ assert (Utils.is_nat(id) && inv_ATM_AccountId(id));
  //@ assert (Utils.is_nat1(amount) && inv_ATM_Amount(amount));
  //@ assert Utils.is_(St,atm.ATMtypes.St.class);
  Boolean andResult_3 = false;
  //@ assert Utils.is_bool(andResult_3);
  if (SetUtil.inSet(St.get_currentCard(), St.get_validCards())) {
    Boolean andResult_4 = false;
    //@ assert Utils.is_bool(andResult_4);
    if (St.get_pinOk()) {
      Boolean andResult_5 = false;
      //@ assert Utils.is_bool(andResult_5);
      if (SetUtil.inSet(St.get_currentCard(),
          ((atm.ATMtypes.Account) Utils.get(St.accounts, id)).get_cards())) {
        if (SetUtil.inSet(id, MapUtil.dom(St.get_accounts()))) {
          andResult_5 = true;
          //@ assert Utils.is_bool(andResult_5);
        }
      }
      if (andResult_5) {
        andResult_4 = true;
        //@ assert Utils.is_bool(andResult_4);
      }
    }
    if (andResult_4) {
      andResult_3 = true;
      //@ assert Utils.is_bool(andResult_3);
    }
  }
  Boolean ret_29 = andResult_3;
  //@ assert Utils.is_bool(ret_29);
  return ret_29;
}
```

Listing 15 Code-generated version of the `pre_Withdraw` operation.

```
//@ requires pre_Withdraw(id,amount,St);
public static Number Withdraw(final Number id, final Number amount) {
 //@ assert (Utils.is_nat(id) && inv_ATM_AccountId(id));
 //@ assert (Utils.is_nat1(amount) && inv_ATM_Amount(amount));
 final Number newBalance =
     ((atm.ATMtypes.Account) Utils.get(St.accounts, id)).get_balance().doubleValue()
         - amount.longValue();
 //@ assert Utils.is_real(newBalance);
 {
   VDMMap stateDes_1 = St.get_accounts();
   atm.ATMtypes.Account stateDes_2 = ((atm.ATMtypes.Account) Utils.get(stateDes_1, id));
   //@ assert stateDes_2 != null;
   stateDes_2.set_balance(newBalance);
   //@ assert (V2J.isMap(stateDes_1) && (\forall int i; 0 <= i && i < V2J.size(stateDes_1);
       (Utils.is_nat(V2J.getDom(stateDes_1,i)) && inv_ATM_AccountId(V2J.getDom(stateDes_1,i)
       )) && Utils.is_(V2J.getRng(stateDes_1,i),atm.ATMtypes.Account.class)));
   //@ assert Utils.is_(St,atm.ATMtypes.St.class);
   //@ assert \invariant_for(St);
   Number ret_7 = newBalance;
   //@ assert Utils.is_real(ret_7);
   return ret_7;
 }
}
```

Listing 16 Code-generated version of the `Withdraw` operation.

5.4 Translating postconditions to JML

Postconditions in VDM-SL and JML are semantically similar, although VDM-SL represents the postcondition function as a derived function definition (as was done for preconditions). Furthermore, in VDM and JML postconditions of operations and methods, respectively, can access both the before and after states. Returning to the `Withdraw` operation, one could specify a postcondition requiring that exactly the value specified by the `amount` parameter is withdrawn from the account—see Listing 17.

```
1 Withdraw : AccountId * Amount ==> real
2 Withdraw (id, amount) == ...
3 post
4 let accountPre = accounts~(id),
5     accountPost = accounts(id)
6 in
7  accountPre.balance =
8  accountPost.balance + amount and
9  accountPost.balance = RESULT;
```

Listing 17 The `Withdraw` operation guarded by a postcondition.

The JML translator produces a **`pure`** Java method to represent the postcondition function. This Java method is invoked from the **`ensures`** clause to check that the postcondition holds. The invocation of the postcondition method of the `Withdraw` operation is shown in Listing 18.

```
1 //@ requires pre_Withdraw(id,amount,St);
2 //@ ensures post_Withdraw(id,amount,\result
      ,\old(St.copy()),St);
3 public static Number Withdraw(final Number
      id, final Number amount) {...}
```

Listing 18 Code-generated version of the `Withdraw` operation.

Note in particular how the before and after states are passed to the `post_Withdraw` method. Reasoning about before state is achieved using JML's **`\old`** expression. For the `Withdraw` operation, the before state is constructed as **`\old`** `St.copy()`. Since `St.copy()` is a deep copy of the state (as explained in Sect. 3), the evaluation inside the **`\old`** expression ensures that the result indeed is a representation of the before state.

The JML translator deep copies the state because Java represents every composite data type using a class. So without deep copying the state, only the address of the before state object reference is copied. In effect, only a single object would exist to represent the pre- and poststates. This would never work, since state changes made by the operation would affect what was intended to be a representation of the before state. Therefore, the state is deep copied to get a separate object to represent the before state.

5. Translating the postcondition of an operation

Let `op` be a method code-generated from a VDM-SL user-defined operation and let the signature of `op` be:
`static` `R op(`I_1 i_1`,...,`I_n i_n`)`
Then `op` has a code-generated p
ostcondition method
`post_op` that is **`pure`** and which in addition to the parameters of `op`, also takes the result and the before and after states of `op` as arguments, i.e.

```
/*@ pure @*/ static boolean
post_op(I1 i1,...,In in,
  R RESULT,S _s,S s)
```

To ensure that the postcondition is evaluated we annotate `op` with the following **`ensures`** annotation:

```
//@ ensures post_op(i1,...,in,\result,
  \old(s.copy()),s);
```

While the primary concern is to preserve the behaviour of the specification across the translation, deep copying values may significantly affect system performance in a negative way. In particular, because it is difficult (in general) to avoid deep copying values unnecessarily when they are passed around in the generated code. To address this issue, the Java code-generator (and hence the JML translator) offers an option that, when selected by the user, omits deep copying of values (other than the old state). While the purpose of this is to generate performance-efficient code, this option is, however, only safe to use if Java objects that originate from VDM-SL values are not modified via aliases.

Yi et al. [20] identify and address a number of problems with the **`\old`** expression. In particular, the authors of that work conduct experiments showing that passing deep copies of the old state may drastically increase a system's memory usage. To address this, Yi et al. propose the **`\past`** expression as a more memory-efficient alternative to the **`\old`** expression. Yi et al. further show that the **`\past`** expression can be implemented as an extension of the OpenJML runtime assertion checker by means of aspect-oriented programming principles. However, OpenJML does not officially support the **`\past`** expression yet, which is why the translation rules do not currently rely on this expression. As we see it, the ideas proposed by Yi et al. could potentially support the development of a more performance-efficient way to handle old states in the JML translator.

Similar to rule 3, rule 5 also assumes that the state of the module enclosing `op` exists. If the state component does not exist, rule 5 changes to not include the state parameters in the definition of `post_op`. Furthermore, if `op` does not return a result (the return type is void), then the definition of `post_op` does not include the `RESULT` parameter.

The example above considers the postcondition of an operation (i.e. `Withdraw`). As described in Sect. 2, the postcondition of a function is not allowed to access state. Therefore, the code-generated version of the postcondition function is not passed the state.

6. Translating the postcondition of a function

Let `f` be a method code-generated from a VDM-SL user-defined function and let the signature of `f` be:
`static` `R f(`I_1 i_1`,...,`I_n i_n`)`
Then `f` has a code-generated postcondition method `post_f` that is **`pure`** and which in addition to the parameters of `f` also takes the result of `f` as an argument, i.e.
`/*@ pure @*/ static boolean post_f(`I_1 i_1`,...,`I_n i_n`,R RESULT)`
To ensure that the postcondition is evaluated we annotate `f` with the following **`ensures`** annotation:
`//@ ensures post_f(`i_1`,...,`i_n`,\result);`

5.5 Translating record invariants to JML

A record can, like any other type definition in VDM-SL, be constrained by an invariant. As an example, Listing 19 shows a record definition modelling a bank account.

```
1 Account ::
2  cards : set of Card
3  balance : real
4  inv a == a.balance >= -1000;
```

Listing 19 A VDM-SL record definition modelling a bank account.

An `Account` comprises the available balance as well as the debit cards associated with the account. We further constrain an `Account` to not have a balance of < -1000, which is expressed using an invariant.

As described in Sect. 3, a record definition is translated to a class that emulates the behaviour of a value type using `copy` and `equals` methods.

Since a record invariant is required to hold for every record value, or object instance in the generated code, we represent it using an **`instance invariant`** in JML as shown in Listing 20. Note that the **`instance invariant`** is formulated as an implication such that invariant violations are not reported when invariant checks are disabled. As we see in Sect. 5.6, this has to do with the way VDM-SL handles atomic execution.

```
1  //@ nullable_by_default
2  final public class Account
       implements Record
3  {
4   public VDMSet cards;
5   public Number balance;
6   //@ public instance invariant atm.
        ATM.invChecksOn ==>
        inv_Account(cards,balance);
7    ...
8   /*@ pure @*/
9   public boolean equals(final Object
        obj)
10  {...}
11  /*@ pure @*/
12  public atm.ATMtypes.Account copy()
        {...}
13  /*@ pure @*/
14  public VDMSet get_cards() {...}
15  public void set_cards(final VDMSet
        _cards)
16  {...}
17  /*@ pure @*/
18  public Number get_balance() {...}
19  public void set_balance(final
        Number _balance) {...}
20  /*@ pure @*/
21  /*@ helper @*/
22  public static Boolean inv_Account(
        final VDMSet _cards, final
        Number _balance){
23    return _balance.doubleValue() >=
          -1000L;
24  }
25 }
```

Listing 20 Code-generated version of the `Account` record.

The code-generated record `Account` defines an *invariant method* `inv_Account` that takes all the record fields of `Account` as input and evaluates the invariant predicate. This method is invoked directly from the JML invariant, as shown in Listing 20. Note that `inv_Account` is a **`static`** method according to rule 2. In addition, this method is annotated as a **`helper`** to avoid the invariant check triggering another invariant check, which eventually would cause a stack-overflow.

7. Translating a record invariant

Let D be a code-generated record definition with fields $f_1, \ldots, f_n$ of types $F_1, \ldots, F_n$, respectively, and let D be constrained by an invariant. Then D has an invariant method inv_D that is annotated as a **helper** to allow it to be invoked from the invariant clause of D. The invariant method can also be annotated as **pure** since it originates from a function definition. The annotated signature of inv_D thus becomes:

```
/*@ pure @*/
/*@ helper @*/
boolean inv_D(F1 f1,...,Fn fn)
```

Let further invChecksOn be a variable that is true if invariant checking is enabled and false otherwise. To represent the record invariant of D we annotate D with the **invariant** annotation:

```
/*@ public instance invariant
invChecksOn ==> inv_D(f1,...,fn); @*/
```

As we see in Sect. 5.6, atomic execution sometimes requires extra assertions to be inserted into the generated code in order to guarantee that the record invariant semantics of VDM-SL are preserved.

All the methods inside a record class—except for the constructor and the "setter" methods—do not modify the state of the record class, and therefore they are marked as **pure**. Updates to a record object in the generated code are made using the "setter" methods of the generated record class, or by using the record modification expression [13]. Use of "setter" methods instead of direct field access to manipulate the state of a record (which is how field access is achieved in VDM-SL) forces the record object into a *visible state* (as described in Sect. 2.2) after it has been updated, thus triggering the invariant check according to the VDM-SL semantics. For example, in VDM-SL we could set the balance of an account as shown in Listing 21.

```
acc.balance := newBalance;
```

Listing 21 Updating the Account balance in VDM-SL.

This assignment produces the Java code shown in Listing 22. Note that for this particular case there is no need to generate any additional JML assertions since the state of acc becomes visible after the call to set_balance. This causes the invariant check of Account to trigger.

```
acc.set_balance(newBalance);
```

Listing 22 Updating the Account balance in the generated code.

5.6 Atomic execution

There are situations where multiple assignment statements in VDM-SL need to be evaluated atomically in order to avoid unintentional violation of a state invariant. In our example, this is the case when the ATM returns the card to the owner, which is done as the last step of a transaction. Returning the debit card also requires us to invalidate the PIN code currently entered. These two things have to be done atomically to avoid violating the state invariant of the ATM module, which is checked using the inv_St function, derived from the state invariant shown in Listing 10 in Sect. 4. Therefore the body of the ReturnCard operation is executed inside an **atomic** statement block as shown in Listing 23. Note that the invariant is evaluated internally by the interpreter, and therefore the example in Listing 23 makes no explicit mention of the invariant.

```
1 ReturnCard : () ==> ()
2 ReturnCard () ==
3 atomic (
4  currentCard := nil;
5  pinOk := false;
6 )
7 pre currentCard <> nil
8 post currentCard = nil and not pinOk;
```

Listing 23 Removal of the debit card from the ATM in VDM-SL.

JML does not include a syntactic construct similar to that of the **atomic** statement. Instead atomic execution must be achieved using different means—for example by manipulating state directly using field access or **helper** methods.

To be consistent with the way record state is updated, and to reflect the way that VDM-SL handles atomic execution, we believe a better approach is to use a flag that indicates if invariant checks are enabled or not. Since this flag should not affect the generated code, we make it a **ghost** field such that it is only visible at the specification level. Since this **ghost** field must be accessible everywhere in the translation, we make it a static field of the class, as shown in Listing 24. The **ghost** field must be added to one of the generated Java classes since Java does not really have global variables. Note that this flag does not affect pre- and postconditions since these checks must always be evaluated.

```
1 /*@ public ghost static boolean
      invChecksOn = true; @*/
```

Listing 24 Ghost field used to control invariant checking.

The declaration of invChecksOn allows us to formulate invariants such that violations are reported only if invariant checking is enabled. An example of this is shown in Listing 25 for the record state class of the ATM module.

```
1  //@ public instance invariant atm.
       ATM.invChecksOn ==> inv_St(
       validCards,currentCard,pinOk,
       accounts);
```

Listing 25 The invariant of the record state class.

The invChecksOn flag provides the means to emulate the behaviour of atomic execution in a Java environment as shown in Listing 26. Specifically, the JML **set** statement is used to disable/enable invariant checking before/after executing the body of the ReturnCard method.

```
1  //@ requires pre_ReturnCard(St);
2  //@ ensures post_ReturnCard(\old(St
       .copy()),St);
3  public static void ReturnCard() {
4   atm.ATMtypes.Card atomicTmp_1 =
        null;
5   //@ assert ((atomicTmp_1 == null)
        || Utils.is_(atomicTmp_1,atm.
        ATMtypes.Card.class));
6   Boolean atomicTmp_2 = false;
7   //@ assert Utils.is_bool(
        atomicTmp_2);
8   { /* Start of atomic statement */
9     //@ set invChecksOn = false;
10    //@ assert St != null;
11    St.set_currentCard(Utils.copy(
          atomicTmp_1));
12    //@ assert St != null;
13    St.set_pinOk(atomicTmp_2);
14    //@ set invChecksOn = true;
15    //@ assert \invariant_for(St);
16  } /* End of atomic statement */
17 }
```

Listing 26 Code-generated version of the ReturnCard operation.

8. Enabling and disabling invariant checking

Declare in a code-generated module M a globally accessible JML **ghost** field invChecksOn to control invariant checking:

```
/*@ public ghost static
boolean invChecksOn = true; @*/
```

Before executing a code-generated atomic statement (in any of the code-generated modules) invariant checking is disabled using the following JML **set** statement:

```
//@ set M.invChecksOn = false;
```

After the code-generated atomic block has finished executing invariant checking is re-enabled using:

```
//@ set M.invChecksOn = true;
```

When all the statements have been executed, it must be ensured that no invariants have been violated. For the example in Listing 26, the only thing that needs to be checked is that the state component of the ATM class, i.e. St does not violate its invariant. This is checked by asserting that **\invariant_for**(St) holds.

9. Resuming invariant checking

Let $d_1, \ldots, d_n$ be state designators of records that have been updated, or affected by an update, during execution of a code-generated atomic statement block. Further assume that $d_1, \ldots, d_n$ have been updated in the given order, i.e. d_i was updated (for the first time) before d_{i+1} and that d_i may be of one of m_i record types $D_{i1}, \ldots, D_{im_i}$. Immediately after executing the code-generated atomic statement block, it is checked that the state designators $d_1, \ldots, d_n$ do not violate any invariants using the following sequence of **assert** statements:

```
//@ assert d1 instance of D11 ==>
  \invariant_for((D11) d1);
...
//@ assert d1 instance of D1m1 ==>
  \invariant_for((D1m1) d1);
...
//@ assert dn instance of Dn1 ==>
  \invariant_for((Dn1) dn);
...
//@ assert dn instance of Dnmn ==>
  \invariant_for((Dnmn) dn);
```

The **\invariant_for** construct is not currently implemented in OpenJML. Instead the invariant check, for an object, can be inlined as a method call (rather than explicitly using **\invariant_for**). However, throughout this paper we use **\invariant_for** to check record invariants as we believe it makes the examples easier to understand.

The JML translator keeps track of state designators of records that potentially have been updated as part of executing the code-generated atomic statement block. This is done by analysing the left-hand sides of the assignment statements. Immediately after invariant checking is re-enabled, i.e. the code-generated atomic statement block has finished execution, it is checked that no record violates its invariant.

There are a few things related to rule 9 that are worth clarifying. First, for assignments to composite state designators such as a.b.c:=42, the invariants of the individual state designators a, b and c have to be checked, if these are defined. For this particular example, we say that c was updated and that a and b were affected by the update. Second, the order in which the invariants are checked follows that used by the Overture VDM interpreter. Third, regardless of how many times a state designator is updated, the corresponding invariant is only checked once (for each state designator) since this is how atomic execution works in VDM, i.e. the update(s) are performed atomically, and afterwards the constraints that the state designators are subjected to are checked. Fourth, rule 9 includes all the state designators that have been updated or affected by an update. No particularly complex situations can arise that makes it diffi-

cult to identify these state designators since all VDM-SL's data types use call-by-value semantics, and therefore no aliasing can occur. Essentially this means that for the assignment statement `a.b.c:=42`, the only invariants (if defined) that have to be checked are those of the state designators `a`, `b` and `c` since aliases do not exist. Therefore, the JML translator can determine (using static analysis) that assertions only have to be generated for these state designators. Naturally this simplifies the translation process, since the JML translator does not have to identify additional state designators (other than those that appear on the left-hand side of the assignment) that are affected by the assignment.

A state designator can be "masked" as a union type, and in such situations, it cannot always be statically determined what the runtime type of a state designator will be. To demonstrate this, consider the record types `R1` and `R2` and a state designator declared as **`dcl`** `r : R1 | R2 := ...`. Further assume that `R1` and `R2` code-generate to classes $\texttt{R1}_\texttt{C}$ and $\texttt{R2}_\texttt{C}$. After updating `r` atomically in the generated code, it is ensured that **`\invariant_for`**$(\texttt{(R1)}_\texttt{C})$ `r` holds if `r` is of type $\texttt{R1}_\texttt{C}$, and similarly that the equivalent condition is true if `r` is of type $\texttt{R2}_\texttt{C}$. Since rule 9 has to take all possible types into account, the invariant checks are formulated as implications.

Although the VDM type system allows state designators to be "masked" as union types, most of the time it is possible to statically determine the runtime type of a state designator. For example, in Listing 26 no **`instanceof`** check is needed since the static type of the state component is `St`. This is an example where the JML translator simplifies the checks proposed by rule 9.

There are more aspects to rule 9 worth discussing—especially when state designators are based on arbitrarily complex data structures such as nested records. These are addressed in Sect. 7.1.

5.7 Translating module state to JML

As described in Sect. 2.1, a module state invariant constrains the record type used to represent the state component of the enclosing module. Therefore, a module state invariant can essentially be seen as a record invariant that can be translated into JML-annotated Java without introducing additional translation rules. This subsection instead explains how a VDM-SL state definition is translated into a form that allows the rules related to record invariants to be applied (see Sect. 5.5).

In our example, each account can be accessed from an ATM using one of the debit cards associated with it. In addition to the bank accounts, the state of the ATM also keeps track of the debit cards that the system considers valid, the debit card that is currently inserted into the ATM, and whether the PIN code entered by the user is valid. The state (including the state invariant) as specified in VDM-SL is shown in Listing 10 and described in Sect. 4. Based on the state definition, a record class is generated that represents the state type as shown in Listing 27. Recall that the fields in this class are nullable according to rule 1.

```
1  final public class St implements
       Record {
2   public VDMSet validCards;
3   public atm.ATMtypes.Card
       currentCard;
4   public Boolean pinOk;
5   public VDMMap accounts;
6
7   //@ public instance invariant atm.
       ATM.invChecksOn ==> inv_St(
       validCards,currentCard,pinOk,
       accounts);
8   /* Record methods omitted */
9  }
```

Listing 27 The record class used to represent the state type.

In addition, an instance of the record class is created to represent the state component as shown in Listing 28. The state component is annotated with the **`spec_public`** modifier so that it can be referred to from the **`requires`** and **`ensures`** clauses of **`public`** methods. Also note that the module is not constrained by an invariant. This is handled entirely by the record invariant shown in Listing 27.

```
1  final public class ATM {
2   /* Fields omitted */
3
4   /*@ spec_public @*/
5   private static atm.ATMtypes.St St
       = new atm.ATMtypes.St(SetUtil.
       set(),
6                 null, false, MapUtil.
                      map());
7   /* Module methods omitted */
8  }
```

Listing 28 The state component in the `ATM` module.

10. Translating the state component

Annotate state components of module classes with the **`spec_public`** modifier to ensure that the state components can be referred to from the **`requires`** and **`ensures`** clauses of **`public`** methods.

6 Checking VDM types using JML

In this section, we describe how the translator uses JML to check the consistency of VDM types when they are code-generated.

Throughout this section, we construct a function called `Is(v,T)` that takes as input a Java value `v` and a VDM type

`T` and produces a JML expression that can be used to check whether `v` represents a value of type `T`. We use `Is(v,T)` to check whether a Java value remains consistent with the VDM type that produces it. The check produced by `Is(v,T)` can be added to the generated Java code to ensure that no type violations occur.

This section covers some of the different classes of VDM types that the JML translator supports, and explains using our case study example, how JML is used to check a Java value against the VDM type that produces it. Finally, we summarise and provide the complete definition of `Is(v,T)` in Fig. 1.

6.1 Where to generate dynamic type checks

Most of the types available in VDM are also present in Java in some form or other. The VDM and Java type systems do, however, have some differences that require us to generate extra checks to ensure that a Java value remains consistent with the VDM type that produces it.

In addition to producing the JML expression needed to check the consistency of a type, i.e. `Is(v,T)`, we also need to consider where to add the check to the generated code. The description below summarises the VDM-SL constructs that must be considered when adding these checks to the generated Java code. We use the term *parameter* to refer to an identifier whose value does not change. A parameter can be defined using a **`let`** construct, which is different from a state designator or variable that can be locally defined using a **`dcl`** statement or globally using a state definition (see Sect. 2). The constructs to be considered are:

- **`return`** statement: If a functional description has a specified result type in its signature, then the returned value must be checked against the specified type.
- Parameters of functions and operations: The arguments passed to a functional description must be checked against the specified types of the corresponding formal parameters upon entry to the functional description.
- State designators: After updating a local or global state designator, the new value assigned must respect the type of the state designator.
- Variable or parameter declaration: After initialising a variable or parameter it must be checked against its declared type.
- Value definition: An explicitly typed value definition must specify a value consistent with its type.

All of the constructs in the list above—with the exception of the value definition—can be checked using a JML **`assert`** statement. The reason for this is that the code-generated versions of these constructs appear inside methods in the generated code. Since a VDM value definition code-generates to a **`public static final`** field (a constant), it is checked using a **`static`** invariant.

6.2 Translating basic types

In our example, we may wish to check that the amount being withdrawn from an account is valid—for example by requiring that it is a natural number larger than zero, as shown in Listing 29.

```
1 let amount : nat1 = expense - profit
2 in
3   Withdraw(accId, amount);
```

Listing 29 Use of explicit type annotation to ensure that a valid amount is being withdrawn.

In the generated Java code, shown in Listing 30, this is checked by analysing the value of the `amount` variable using the `Utils.is_nat1` method available from the Java code-generator's runtime library. This method is invoked from a JML annotation in order to check that `amount` is different from **`null`** and that it represents an integer larger than zero.

```
1 Number amount =
2   expense.longValue() - profit.
      longValue();
3 //@ assert Utils.is_nat1(amount);
4 return Withdraw(accId, amount);
```

Listing 30 Use of JML to check that a valid amount is being withdrawn.

11. Checking of the `nat1` type

Let `v` be a value or object reference in the generated code that originates from a variable or pattern of type **`nat1`** and further define `Is(v,nat1) = Utils.is_nat1(v)`. To ensure that `v` represents a value of type **`nat1`**, generate a JML check to ensure that `Is(v, nat1)` holds.

The approach used to check other basic types follows the principles demonstrated using Listing 29 and Listing 30—the main difference being that each basic type uses a dedicated method from the Java code-generator's runtime library. Therefore, we omit the details of how other basic types of VDM are checked using JML, and instead provide the complete set of rules in Fig. 1.

We note that a record type or a quote type can be checked in a way similar to that of a basic type. The reason for this is that the Java code-generator produces a Java class for each of the record definitions and quote types in the VDM model. Therefore, all there is to checking whether an object reference represents a given record or quote class is to check whether

$$
\texttt{Is(v,T)} = \begin{cases}
\texttt{Utils.is\_bool(v)} & \textbf{if } \texttt{T} = \textbf{bool} \\
\texttt{Utils.is\_nat(v)} & \textbf{if } \texttt{T} = \textbf{nat} \\
\texttt{Utils.is\_nat1(v)} & \textbf{if } \texttt{T} = \textbf{nat1} \\
\texttt{Utils.is\_int(v)} & \textbf{if } \texttt{T} = \textbf{int} \\
\texttt{Utils.is\_rat(v)} & \textbf{if } \texttt{T} = \textbf{rat} \\
\texttt{Utils.is\_real(v)} & \textbf{if } \texttt{T} = \textbf{real} \\
\texttt{Utils.is\_char(v)} & \textbf{if } \texttt{T} = \textbf{char} \\
\texttt{Utils.is\_token(v)} & \textbf{if } \texttt{T} = \textbf{token} \\
\texttt{Utils.is\_(v,String.}\textbf{class}\texttt{)} & \textbf{if } \texttt{T} = \textbf{seq of char} \\
\texttt{Utils.is\_(v,}S_{CG}\texttt{.}\textbf{class}\texttt{)} & \textbf{if } \texttt{T} \text{ is a record or quote type } \texttt{S} \text{ that generates to a Java class with the fully qualified name } S_{CG} \\
\texttt{(v == }\textbf{null}\texttt{ || Is(v,S))} & \textbf{if } \texttt{T} = \texttt{[S]} \\
\texttt{V2J.isTup(v,n) \&\& Is(v,}T_1\texttt{) \&\&...\&\& Is(v,}T_n\texttt{)} & \textbf{if } \texttt{T} = T_1 * \ldots * T_n \\
\texttt{Is(v,}T_1\texttt{) ||...|| Is(v,}T_n\texttt{)} & \textbf{if } \texttt{T} = T_1 | \ldots | T_n \\
\texttt{V2J.isSet(v) \&\& (\}\textbf{forall int}\texttt{ i; 0 <= i \&\& i < V2J.size(v); Is(V2J.get(v,i),S))} & \textbf{if } \texttt{T} = \textbf{set of } \texttt{S} \\
\texttt{V2J.isSeq(v) \&\& (\}\textbf{forall int}\texttt{ i; 0 <= i \&\& i < V2J.size(v); Is(V2J.get(v,i),S))} & \textbf{if } \texttt{T} = \textbf{seq of } \texttt{S} \\
\texttt{V2J.isSeq1(v) \&\& (\}\textbf{forall int}\texttt{ i; 0 <= i \&\& i < V2J.size(v); Is(V2J.get(v,i),S))} & \textbf{if } \texttt{T} = \textbf{seq1 of } \texttt{S} \\
\texttt{V2J.isMap(v) \&\& (\}\textbf{forall int}\texttt{ i; 0 <= i \&\& i < V2J.size(v); Is(V2J.getDom(v,i),D) \&\& Is(V2J.getRng(v,i),R))} & \textbf{if } \texttt{T} = \textbf{map } \texttt{D } \textbf{to } \texttt{R} \\
\texttt{V2J.isInjMap(v) \&\& (\}\textbf{forall int}\texttt{ i; 0 <= i \&\& i < V2J.size(v); Is(V2J.getDom(v,i),D) \&\& Is(V2J.getRng(v,i),R))} & \textbf{if } \texttt{T} = \textbf{inmap } \texttt{D } \textbf{to } \texttt{R} \\
\texttt{Is(v,D) \&\& inv\_T(v)} & \textbf{if } \texttt{T} \text{ is a named invariant type with domain type } \texttt{D} \text{ and invariant method } \texttt{inv\_T}
\end{cases}
$$

Fig. 1 Complete definition of `Is(v,T)`

the object reference is an instance of said class. The rules for checking record and quote types are included in Fig. 1.

6.3 Translating collections

In the generated code, the `VDMSet`, `VDMSeq` and `VDMMap` collection classes are used as raw types. Therefore the code-generator does not take advantage of Java generics to make compile-time guarantees about the types of the objects a collection stores. This approach has the advantage of making it easier to store Java objects and values of different types in the same collection without having to introduce additional types. Although this allows the type system of VDM to be represented in Java, it has the disadvantage that no compile-time guarantees can be made about the types of the objects that a collection stores.

In the ATM example, we use the `TotalBalance` function, shown Listing 31, to calculate the total balance available from a set of accounts.

```
1 TotalBalance : set of Account ->
    real
2 TotalBalance (acs) ==
3  if acs = {} then
4    0
5  else
6   let a in set acs
7   in
8     a.balance + TotalBalance(acs \
        {a});
```

Listing 31 Function that calculates the total balance available from a set of accounts.

When the `TotalBalance` function is code-generated to JML-annotated Java, the code-generator adds JML assertions to ensure that the set of accounts is consistent with the collection type used in VDM. Since an `Account` record is represented using a Java class with the same name, we have to check that every element in the set is an instance of said Java class. As shown in Listing 32, this is checked using a quantified expression. This expression uses a bound variable `i` to iterate over all the accounts and check that each element is an instance of the `Account` record class. Although sets are unordered collections, the quantified expression takes advantage of `VDMset` being implemented as an ordered collection. The formulation of the range expression in the quantified expression further ensures that the assertion can be checked using a tool such as the OpenJML runtime assertion checker, i.e. the assertion is executable.

```
1  /*@ pure @*/
2  public static Number TotalBalance(
       final VDMSet acs) {
3   //@ assert (V2J.isSet(acs) && (\
        forall int i; 0 <= i && i <
        V2J.size(acs); Utils.is_(V2J.
        get(acs,i),atm.ATMtypes.
        Account.class)));
4   if (Utils.empty(acs)) {
5     Number ret_1 = 0L;
6     //@ assert Utils.is_real(ret_1);
7     return ret_1;
8   } else { ... /*Compute sum
        recursively */}
9  }
```

Listing 32 Code-generated version of the `TotalBalance` operation.

The JML translator only uses Java 7 features since OpenJML did not support Java 8 at the time the JML translator was developed. Iterating over collections (as shown in Listing 32) may also be achieved using Java 8 features such as lambda expressions. For example, one could imagine a method used to check collection types that would take as input two arguments (1) the collection itself and (2) a predicate method (e.g. lambda expression) that would be evaluated for each of the elements in the collection. In that way the generated JML annotations would not have to rely on sets implemented as ordered collections. Since lambda expressions in Java are mostly syntactic sugar for anonymous inner classes, lambda expressions could in principle be represented solely using Java 7 features. However, using this approach, the generated JML annotations would not be concise, although this is only a concern if a human will read them.

12. Checking of sets

Let `v` be a value or object reference in the generated code that originates from a variable or pattern of the VDM set type **`set of`** `T` and further define

```
Is(v,set of T) = V2J.isSet(v) &&
(\forall int i; 0 <= i &&
i < V2J.size(v); Is(V2J.get(v,i),T))
```

To ensure that `v` represents a value of type **`set of`** `T`, generate a JML check to ensure that `Is(v,set of T)` holds.

The VDM sequence types **`seq`** and **`seq1`** are checked in a way similar to sets. The difference between checking the **`seq`** and **`seq1`** collection types is that the **`seq1`** type requires at least one element to be present in the sequence. Checking a map, which like a set is an unordered collection, takes advantage of `VDMMap` imposing an order on the domain and range values. The main difference between checking a map and a set is that both the domain and range values of a map have to be checked. Checking the injective map type **`inmap`** is similar to checking a standard map, except that the injectivity property must hold. We refrain from providing examples of how to check each of the collection types in VDM since they are similar to what has already been shown. Instead we summarise the rules for checking all of the collection types in Fig. 1.

6.4 Translating named invariant types to JML

Since the Java code-generator does not generate additional class definitions for named invariant types, the invariant imposed on such a type cannot be expressed as a JML invariant. This is only possible for a record since it translates to a class definition.

Instead, we identify places in the generated code where a named invariant type may be violated, as described in Sect. 6.1, and check that the invariant holds. Also, it is worth noting that a named invariant type, unlike a record type, does not have an explicit type constructor. Therefore, an expression can only violate a named invariant type if the expression is explicitly declared to be of that type.

The ATM in our example is not capable of dispensing cents and also imposes a limit on the amount of money that can be withdrawn. Therefore, the amount of money can be represented as a named invariant type. An attempt to withdraw an amount of money that exceeds 2000 will yield a runtime error. The named invariant type used to represent the amount withdrawn from an account is shown together with the `Withdraw` operation in Listing 33.

```
types
Amount = nat1
inv a == a < 2000;

operations
Withdraw : AccountId * Amount ==> real
Withdraw (id, amount) == ...
```

Listing 33 The amount to withdraw modelled using a named invariant type.

On entering the code-generated version of `Withdraw`, shown in Listing 34, we assert that `amount` meets the named invariant type `Amount`. The assertion does two things: firstly it performs a dynamic type check to ensure that `amount` is a valid domain type of `Amount` and secondly it checks that the invariant predicate holds. For the example in Listing 34, this means checking that `amount` is of type **`nat1`** and smaller than 2000. Note that meeting the invariant condition does not imply compatibility with the domain type of the named invariant type and vice versa. For example, -1 is smaller than 2000 but it is not of type **`nat1`**. Likewise, 2001 is of type **`nat1`**, but it exceeds 2000, so neither -1 nor 2001 is of type `Amount`.

```
public static Number Withdraw(final
     Number id, final Number amount
    ){
...
//@ assert (Utils.is_nat1(amount)
    && inv_ATM_Amount(amount));
 ...
}
```

Listing 34 Checking a named invariant type of an operation parameter in JML.

The code-generated invariant method for type `Amount` is shown in Listing 35. Since the named invariant type check, shown in Listing 34, is evaluated from left to right using short-circuit evaluation semantics [21], the invariant method is only invoked if the value subject to checking is compatible with the domain type of the named invariant type. Therefore, it is safe to narrow (or cast) the type of the argument passed to the invariant method before performing the invariant check.

```
/*@ pure @*/
/*@ helper @*/
public static Boolean
    inv_ATM_Amount(final Object
    check_a) {
 Number a = ((Number) check_a);
 return a.longValue() < 2000L;
}
```

Listing 35 The named invariant type method for `Amount`.

13. Checking of named invariant types

Let `v` be a value or object reference in the generated code that originates from a variable or pattern of the VDM named invariant type `T` based on the domain type `D` and constrained by invariant predicate `e(p)`, i.e. `T` is defined as

```
types
T = D
inv p == e(p)
```

Then `T` has an invariant method, responsible for running the code-generated version of the `e(p)` check, with a signature defined as:

```
public static boolean inv_T(Object o)
```

Further define `Is(v,T) = Is(v,D) && inv_T(v)`

To ensure that `v` represents a value of type `T`, generate a JML check to ensure that `Is(v,T)` holds.

Note that the invariant method `inv_T` in rule 13 defines the input parameter `o` to be of type `Object`, thus allowing `inv_T` to accept inputs of any type. Therefore, `inv_T` must narrow the type of the input parameter `o` before performing the invariant check (see the example in Listing 35). This approach has the advantage that it allows simpler JML checks since the argument type does not need to be narrowed before the invariant method is invoked. Had the input parameter of the invariant method been defined using the smallest possible type, then the argument type would need to be narrowed for situations where the argument is masked as a union type. Although this would complicate the JML checks, it would have the advantage of allowing type narrowing to be removed from the invariant methods.

7 Other aspects of VDM-SL affecting the JML-generation

There are other aspects of VDM-SL that further complicate the generation of VDM-SL models to JML-annotated Java. In this section, we use examples to demonstrate these issues and explain how they may be overcome.

7.1 Complex state designators

State designators may be composite data structures such as records with fields that themselves are records. Such a data type forms *complex state designators* that when modified require careful handling during the translation process. To demonstrate this, consider the three VDM-SL record definitions `R1`, `R2` and `R3` in Listing 36. Note in particular how the invariants of `R1` and `R2` depend on the field of `R3`. This transitive dependency complicates checking of invariants in the generated code. To demonstrate this, the operation in List-

ing 36 instantiates R1 as r1 and modifies it to violate the R1 invariant, which causes a runtime error to be reported.

```
types
R1 :: r2 : R2
inv r1 == r1.r2.r3.x <> -1;
R2 :: r3 : R3
inv r2 == r2.r3.x <> -2;
R3 :: x : int
inv r3 == r3.x <> -3;

operations
op: () ==> nat
op () ==
(
 dcl r1 : R1 := mk_R1(mk_R2(mk_R3
     (5)));
 r1.r2.r3.x := -1;
 return 0;
)
```

Listing 36 Record nesting in VDM-SL.

The operation op in Listing 36 produces the method in Listing 37. For this example, r1 is the same in both listings, r2 is the same as stateDes_1 in Listing 37, and r3 is the same as stateDes_2. Note that in Listing 37 we have removed fully qualified names of record classes and other JML checks that are not relevant.

```
public static Number op() {
 R1 r1 = new R1(new R2(new R3(5L)));
 R2 stateDes_1 = r1.get_r2();
 R3 stateDes_2 = stateDes_1.get_r3();
 stateDes_2.set_x(-1L);
 //@ assert \invariant_for(stateDes_1);
 //@ assert \invariant_for(r1);
 Number ret_1 = 0L;
 return ret_1;
}
```

Listing 37 Code-generated version of the operation from Listing 36.

Immediately after completing the state update, i.e. invoking stateDes_2.set_x(-1L), the following things happen:

1. The state of stateDes_2 becomes *visible* thus triggering the invariant check of stateDes_2.
2. The invariant check of stateDes_1 is run by asserting **\invariant_for**(stateDes_1) and finally,
3. the invariant check of r1 is run by asserting **\invariant_for**(r1), which causes a runtime error to be reported.

Strictly speaking the objects pointed to by stateDes_1 and r are also in visible states after executing the update to stateDes_2 and therefore the invariants of those objects should also hold. In particular, a state is *visible* for an object o "when no constructor, destructor, non-static method invocation with o as receiver, or static method invocation for a method in o's class or some superclass of o's class is in progress [16]". So in theory the invariant checks should not have to be run explicitly (step 2 and step 3). The reason that the JML translator generates these checks anyway has to do with the strategies JML tools use to check invariants.

Tools such as JML runtime checkers may assume no problems with ownership aliasing to avoid having to keep track of what objects and types are in visible states. Although this reduces the overhead of checking invariants, it also means that some invariant violations might go unnoticed. Alternatively, tools can check every applicable invariant for classes and objects in visible states, but this adds a significant overhead to the program execution.

Since aliasing can never occur in VDM-SL, it becomes simpler to keep track of what objects are in a visible state in the generated code and thus generate JML checks that explicitly trigger the invariants checks. This has the advantage that invariant violations do not go unnoticed even though a JML tool adopts a more practical approach to checking invariants.

For the example in Listing 37, the important thing is to ensure that the violation of the invariant of R1 is reported after executing the state update. This is done by asserting the entire chain of state designators. The JML translator is able to generate these checks since it keeps track of state designators of records that may have been affected by updates to other state designators.

14. Checking transitive dependencies

Let d_n be a state designator of a record in the generated code that has been updated non-atomically, and let $d_k, \ldots, d_1$, for $k = n-1$, be state designators that were affected by the update to d_n. Further assume that d_i may be of one of m_i record types $D_{i1}, \ldots, D_{im_i}$. Immediately after executing the update to d_n the state of d_n becomes visible. To ensure that the invariant is evaluated for all affected state designators, execute the following sequence of assertions:

```
//@ assert dk instance of Dk1 ==>
  \invariant_for((Dk1) dk);
...
//@ assert dk instance of Dkmk ==>
  \invariant_for((Dkmk) dk);
....
//@ assert d1 instance of D11 ==>
  \invariant_for((D11) d1);
...
//@ assert d1 instance of D1m1 ==>
  \invariant_for((D1m1) d1);
```

Note that the code in Listing 37 omits the **instance of** checks, proposed by rule 14, since the types of the affected state designators can be determined statically.

Regarding rule 9, similar issues with transitive dependencies may occur in the generated code when dealing with atomic execution. Recall that invariant checking is disabled before a code-generated atomic statement block is executed. Once the atomic execution has completed, invariant checking is re-enabled, and therefore rule 9 must also take into account all the state designators that were affected by the atomic execution.

7.2 Recursive types

It is possible to formulate recursive types for which the generated JML checks can only perform limited type checking. To demonstrate this, consider the recursive VDM type definition in Listing 38. For this example, `S` represents an infinite number of types including **nat1** as well as all possible dimensions of sequences that store elements of type **nat1**, i.e. **seq of nat1**, **seq of seq of nat1** and so on.

```
types
S = nat1 | seq of S;
```

Listing 38 Example of recursive type definition in VDM.

The issue with this kind of type definition is that `Is(v,S)` in theory becomes an expression of infinite length. The JML translator stops generating type checks whenever it encounters type cycles. For the particular example in Listing 38, this means that a Java value or object reference `v` is only considered to respect `S` if `Utils.is_nat1(v)` holds. For the rest of this section, we discuss the current limitations of type checking recursive types and describe how these limitations may be addressed.

The approach used to check types could be changed to also take the depth of the recursion `n` into account, i.e. use `Is(v,T,n)` to generate the type checks. The current approach used by the JML translator thus corresponds to generating checks using `Is(v,T,1)`. `Is(v,S,2)` then generates checks for types **nat1** and **seq of nat**, whereas `Is(v,S,3)` additionally generates a check for the type **seq of seq of nat1**.

Alternatively, checking a recursive type `T` (such as `S` shown in Listing 38) can be done using a code-generated recursive method that is constructed in a way that allows a value `v` to be validated against `T`. Although static provers may not be able to perform checking of such types, it should be possible using runtime assertion checking. However, in order to enable this style of type checking, the JML translator would have to be extended with functionality that enables these methods to be generated such that they can be invoked from the generated JML assertions.

The limitation of the JML translator for the example shown in Listing 38 is a consequence of `S` being defined using the union type constructor "|". However, it is possible to check more practical examples of recursively defined types such as the linked list `LL` shown in Listing 39.

To demonstrate this, consider the construction of a linked list value in VDM that contains the numbers 1, 2 and 3 as shown in Listing 40. In the generated code, this value is represented using the code shown in Listing 41.

```
types
LL ::
  element : nat
  tail : [LL]
```

Listing 39 Example of a linked list defined using a record type.

```
mk_LL(1, mk_LL(2, mk_LL(3, nil)))
```

Listing 40 Example of a linked list value in VDM.

```
new LL(1L,new LL(2L,new LL(3L, null)));
```

Listing 41 Example of a linked list value in Java.

Each time an object of type `LL` is instantiated in Java the constructor checks the types of the current `element` and the `tail`—see Listing 42. For this linked list example, it is therefore possible to type check `LL` since the VDM type is represented using a recursively defined class in the generated code.

```
public LL(final Number _element,
    final LL _tail) {
 //@ assert Utils.is_nat(_element);
 //@ assert (_tail == null || Utils
     .is_(_tail,LL.class));
 ...
}
```

Listing 42 Type checking a linked list using JML.

7.3 Detecting problems with the generated code

As explained in Sect. 5.4, deep copying objects may significantly affect the performance of the generated code. Therefore, the user may not always want to have these copy calls generated. However, from a general perspective this may result in code that does not preserve the semantics across the translation. JML specifications can help detect such problems. To demonstrate this, consider the VDM-SL operation in Listing 43. This operation assumes the existence of a two-dimensional vector `Vector2D`, defined as a record (a value

type). In Listing 43, v2 is created as a deep copy of v1, and therefore the assignment to v1 has no affect on v2, and op therefore returns 1 (see the postcondition).

If this example is translated to Java with deep copying *disabled*, the code shown in Listing 44 is produced. Note that this listing omits the generated JML assertions to focus on the postcondition.

```
1 op : () ==> nat
2 op () == (
3 dcl v1 : Vector2D := mk_Vector2D
    (1,2);
4 dcl v2 : Vector2D := v1; -- Copy
    value
5 v1.x := 2;
6 return v2.x;)
7 post RESULT = 1
```

Listing 43 Use of value types in VDM.

```
1 //@ ensures post_op(\result);
2 public static Number op() {
3   Vector2D v1 = new Vector2D(1L,2L)
      ;
4   Vector2D v2 = v1;
5   v1.set_x(2L);
6   Number ret_1 = v2.get_x();
7   return ret_1;
8 }
```

Listing 44 Generated Java code without copy calls.

If this code is executed using the OpenJML runtime assertion checker, an error is reported because the method returns 2, which is different from the result obtained by executing the corresponding VDM-SL operation. Since deep copying is disabled only the v1 reference is copied, and therefore the update to v1, i.e. v1.set_x(2L), also affects v2.

The detection of the postcondition violation as reported by the OpenJML runtime assertion checker is shown in Listing 45. However, if the code is generated with deep copying enabled (at the cost of performance), then v2 will be constructed as Utils.copy(v1) and the method will change to return 1, as expected.

```
Ex/DEFAULT.java:17: JML
    postcondition is false
     public static Number op() {
Ex/DEFAULT.java:16: Associated
    declaration: Ex/DEFAULT.java:17:
     //@ ensures post_op(\result);
```

Listing 45 Detection of a postcondition violation.

8 Translation assessment

In this section, we provide an assessment of the translation. We first describe how the correctness of the translation was assessed, and afterwards we discuss the scope and treated feature set in relation to existing JML tools.

8.1 Translation correctness

The translation rules have been validated by running examples through the JML translator and analysing the generated Java/JML using the OpenJML runtime assertion checker. Some of the examples used to test the tool constitute *integration tests* that have been developed by the authors. In addition, we have used the tool to analyse an *external specification* (originally used as part of an industrial case study) that the authors have not been involved in the development of. A summary of the different examples used to test the translation is given below. Additional details about the examples can be found via the references provided.

The *integration tests* currently consist of 85 examples that cover testing of all the translation rules. Each test (typically) forms a minimal example that exercises a small part of the entire translation (such as a single rule). The workflow for running these tests is as follows: first, the test model is translated to JML-annotated Java using the JML translator. Next, the generated Java/JML is compiled and executed using the OpenJML runtime assertion checker. Finally, the (actual) output reported by the OpenJML runtime assertion checker is compared to the expected output in order to confirm that the behaviour of the test model is preserved across the translation. For example, if the execution of a test model produces a precondition violation, then the equivalent error is expected to be produced when the generated Java/JML is executed using the OpenJML runtime assertion checker. All the examples used to test the JML translator are available via Overture's GitHub page [19] or can be found in a technical report that presents a more complete definition of the translation [22].

Compared to the *integration tests*, the *external specification* is a large example that is rich in terms of DbC elements. The model was originally developed to study the properties of an algorithm used to obfuscate Financial Accounting District (FAD) codes, which are six digit numbers used to identify branches of a retailer. The customer required that obfuscated FAD codes were still six digit numbers, remained unique (per branch), and that the entire range of FAD codes (0-999999) was still available. In addition, the obfuscation had to be a light-weight calculation (rather than a look-up in a table). The properties of the algorithm were described using VDM contracts to allow the algorithm to be validated using VDM's test automation features [23].

Investigating whether the algorithm met the requirements necessitated the generation and execution of one million tests that initially could not be handled by any of the VDM tools (either due to intractable execution times, or because the VDM interpreter ran out of memory). Motivated by this,

the specification was translated into a JML-annotated Java program [24], and all one million tests were executed using a code-generated version of the VDM specification. In that way, the properties of the obfuscation algorithm could be validated by executing a code-generated version of the VDM specification using the OpenJML runtime assertion checker.

8.2 Translation scope and treated feature set

As explained in Sect. 7.2, it is possible to formulate recursive types that currently are not supported by the JML translator. Aside from that, all VDM-SL's types and contract-based elements are supported. However, the JML translator does not currently support the object-oriented and real-time dialects of VDM, called VDM++ [25] and VDM-RT [15].

The Java code-generator that we extend currently only uses Java 7 features in the generated code. OpenJML is the only JML tool that we are aware of that supports this version of Java. Specifically, as of December 2016, OpenJML version 0.8.5 was released with support for Java 8, i.e. the latest official Java version (at the current time of writing). Other JML tools, on the other hand, lack support for recent Java versions (in particular Java 7 and 8). Therefore, these tools cannot currently be used to analyse the generated Java/JML.

The JML translation is only valuable if the JML features that it relies on are supported by JML tools. Specifically, we have aimed to develop a translation that generates Java/JML that can be analysed using OpenJML. However, the translation would benefit from the **`\invariant_for`** construct, which OpenJML does not currently support. Instead we offer an alternative way to represent this construct in order to achieve compatibility with OpenJML (see Sect. 5.6 for details).

9 Related work

In [26], Vilhena considers the possibilities for automatically converting between VDM++ and JML and the approach is demonstrated using a proof-of-concept implementation. That work considers a bidirectional mapping, whereas we only consider a one-way translation from VDM-SL to Java/JML. The bidirectional mapping proposed by Vilhena only produces the JML specification files (where non-model methods do not define bodies). Therefore, Vilhena's mapping does not generate annotations at the statement level, which is an essential part of our work. The implementation of the bidirectional mapping was originally targeting the Overture tool, but it never reached maturity to be included in the release of the tool.

Rules for translating from a subset of VDM-SL to JML are proposed by Jin and Yang [27]. Their approach also considers implicit functional descriptions, but it provides limited support for translation of record definitions and named invariant types. In the early phases of the software development process, the authors propose to formulate requirements in natural language or using the Unified Modelling Language (UML) [28] and then formalise them in VDM-SL to eliminate ambiguity. Subsequently the authors manually apply their rules to the VDM-SL specification to produce an initial version of the software implementation. Their work does, however, not take generation of the bodies of functions and operations into account. Therefore, the authors only produce the method signatures for the Java methods when translating the functional descriptions of the VDM-SL model.

The translation rules proposed by Jin et al. have been implemented as an Eclipse plugin by Zhou et al. in [29]. The plugin takes a VDM-SL specification as input, which is type-checked using VDMTools [30], and outputs JML-annotated Java classes that must be completed manually by the developer.

Translations from other formal notations or modelling languages to JML-annotated Java have also been developed. As an example, Rivera et al. present the EventB2Java tool [31]—a code-generator, which is capable of translating both abstract and refinement Event-B [32] models into JML-annotated Java. EventB2Java has the advantage over other Event-B code-generators that it does not require user intervention as part of the code-generation process, which is similar to our approach.

In [33], Lensink et al. present a prototype code-generator that translates a subset of the Prototype Verification System (PVS) [34] to an intermediate representation in Why [35] suitable for program verification. Subsequently the Why representation is translated to JML-annotated Java. In their work the authors focus on translating executable PVS constructs, which is similar to what we do for VDM-SL. A key feature of their code-generator is that it, in addition to specification code, also translates proven properties, which is outside the scope of our work.

Hubbers and Oostdijk propose AutoJML [36]—a tool for translating UML state diagrams into JML-annotated Java Card code [37]. A state diagram describes a Java Card applet from which AutoJML produces Java skeleton code annotated with JML. In the generated code, the different states are represented as constant values, and an additional Java field is used to represent the current state of the applet. A JML **`invariant`** is used to specify the valid state values for this field, and a JML **`constraint`** is used to describe the valid state transitions. This is comparable to the way we enable and disable invariant checking, which we do by toggling the `invChecksOn` **`ghost`** field using **`set`** statements.

In [38], Klebanov proposes an approach similar to that of Hubbers et al. Instead of using UML state diagrams, Klebanov uses automata-based programming to describe

the behaviour of a smart card application, which is generated to JML-annotated Java Card code. Klebanov argues that use of automata-based programming over UML state diagrams is a better way to describe application-specific behaviour. A similar argument can be made for VDM-SL, which is suitable for capturing the dynamic aspects of a system.

10 Conclusion and future plans

In this paper, we have demonstrated how VDM-SL models can be translated to JML-annotated Java programs that can be checked for correctness using JML tools. The JML translator uses JML to represent the DbC elements of VDM-SL and generates checks that help ensure the consistency of VDM-SL types across the translation.

The principles for pre- and postconditions in VDM-SL and JML are similar although there are subtle semantic differences between the two notations. These differences are mostly caused by the fact that JML is built on top of Java, where object types use reference semantics. VDM-SL, on the other hand, solely uses value types. Therefore, it is necessary to employ deep cloning principles when representing value types in JML-annotated Java code.

Checking state and record invariants in the generated code is complicated due to two reasons: firstly, atomic execution in VDM requires a way to control when invariant checking must be done. We achieve this by using a **`ghost`** field to indicate when invariant checking is enabled, and update it before entering and leaving the **`atomic`** statement. Secondly, we have demonstrated that transitive dependencies between records sometimes require extra JML checks to be generated to ensure that the invariant checks are evaluated when they should.

The differences between the type systems of VDM-SL and Java further necessitate extra checks to be produced. These checks are needed to ensure that the generated code does not violate any of the constraints imposed by the types in the VDM-SL model. Overture performs these dynamic type checks internally, whereas they must be made explicit in Java.

Although DbC languages often support many of the same DbC concepts, it is the semantic differences between the languages that make developing a translation challenging. In this paper, we have shown several examples of such differences and how they can be addressed. Naturally, translating between other specification language pairs may reveal other differences and design details that are of interest to researchers and practitioners working on comparable tasks. However, based on the experiences gained by developing the VDM-SL-to-JML translation, we list some of the design details that we believe are likely to challenge the development of translations between other specification language pairs:

Invariants: The times when invariants are evaluated vary across specification languages. For example, in VDM they have to hold at all times (except inside **`atomic`** statements), whereas in JML they must hold in visible states. When invariants have different semantics the translation must find a way to either produce or reduce the number of invariant checks at the appropriate places in the code.

Type systems: The differences between type systems require careful attention when developing a translation. Especially, when the destination language (e.g. JML) uses a more "coarse-grained" type system than the source language (e.g. VDM-SL). For such situations, extra checks must be produced to ensure that types are used consistently across the translation. In our work, we use the function `Is(v,T)` to produce these extra checks.

Atomic execution: Languages may use dedicated constructs to represent atomic execution (e.g. VDM) or by allowing invariants not to hold at certain times (e.g. JML). In this paper, an example was given of how a dedicated construct can be emulated in a language that does not support one natively.

Old state: Despite pre- and postconditions being similar concepts in different specification languages, it is likely that the notion of old state may require careful handling when developing a translation between two specification languages. In our work, a deep cloning principle was employed to ensure the correct construction of the old state.

In the future, we plan to use this work in the context of test automation. In VDM, it is possible to specify a trace definition in a way similar to that of a regular expression. This trace can then be expanded into a large collection of tests that can be executed against the model. This is a useful way to detect deficiencies in the model, such as missing preconditions, postconditions and invariants [23].

We plan to code-generate the trace expansion such that the tests can be executed against the code-generated version of the model. The work presented in this paper can then be used to detect contract or type violations and give verdicts to the code-generated trace tests. We believe that this will be particularly advantageous for execution of large collections of tests. We expect this approach to significantly increase execution speed for test cases and also allow more tests to be executed. In addition, we plan to look into JML-generation for other VDM dialects such as VDM++. However, since VDM++ is object-oriented and supports concurrency, we envisage that this will give rise to a completely new set of challenges not addressed by the work in this paper.

So far the analysis of the generated Java/JML has primarily been limited to runtime assertion checking. Another item of future work is to formally verify the generated code against

the JML specification. In particular, by investigating to what extent this is possible, and whether the JML translation can be optimised in a way that better supports formal verification through static analysis. For example, currently the translation produces auxiliary methods for invariants and pre- and postconditions that are used as part of the JML specification. However, use of method calls in specifications complicates static analysis due to, for example, the possibility of exceptions or non-terminating behaviour [39].

We hope that our work will serve as inspiration for other researchers who seek to bridge the gap between other specification notations and implementation technologies that support the DbC approach. We believe that the rules proposed in this paper can be useful for others who want to translate between specification languages such as ASM, B and Z and implementation technologies such as Spec#, Sparc-Ada and Eiffel.

Acknowledgements The authors would like to thank Victor Bandur, Nick Battle and the anonymous reviewers for their valuable feedback on earlier versions of this paper. The work of Leavens was supported in part by the US National Science foundation under Grants CCF 1518789 and CNS 1228695.

References

1. Meyer, B.: Object-Oriented Software Construction. Prentice-Hall International, Upper Saddle River (1988)
2. Bjørner, D., Jones, C. (eds.): The Vienna Development Method: The Meta-Language. Lecture Notes in Computer Science, vol. 61. Springer (1978)
3. Fitzgerald, J., Larsen, P.G.: Modelling Systems—Practical Tools and Techniques in Software Development, 2nd edn. Cambridge University Press, Cambridge (2009). doi:10.1017/CBO9780511626975
4. Fitzgerald, J.S., Larsen, P.G., Verhoef, M.: Vienna Development Method. Wiley Encyclopedia of Computer Science and Engineering. Wiley (2008)
5. Woodcock, J., Davies, J.: Using Z—Specification, Refinement, and Proof. Prentice Hall International Series in Computer Science. Hertfordshire, UK (1996)
6. Jones, C.B.: Software Development A Rigorous Approach. Prentice-Hall International, Englewood Cliffs (1980)
7. Wing, J.M.: Writing Larch interface language specifications. ACM Trans. Program. Lang. Syst. **9**(1), 1–24 (1987). doi:10.1145/9758.10500
8. Burdy, L., Cheon, Y., Cok, D., Ernst, M.D., Kiniry, J.R., Leavens, G.T., Leino, K.R.M., Poll, E.: An overview of JML tools and applications. Int. J. Softw. Tools Technol. Transf. **7**, 212–232 (2005)
9. Larsen, P.G., Battle, N., Ferreira, M., Fitzgerald, J., Lausdahl, K., Verhoef, M.: The overture initiative integrating tools for VDM. SIGSOFT Softw. Eng. Notes **35**(1), 1–6 (2010). doi:10.1145/1668862.1668864
10. The Overture tool website. http://overturetool.org/ (2015)
11. Jørgensen, P.W.V., Couto, L.D., Larsen, M.: A code generation platform for VDM. In: Proceedings of the 12th Overture workshop (2014)
12. Cok, D.: OpenJML: JML for Java 7 by Extending OpenJDK. In: Bobaru, M., Havelund, K., Holzmann, G., Joshi, R. (eds.) NASA Formal Methods. Lecture Notes in Computer Science, vol. 6617, pp. 472–479. Springer, Berlin. doi:10.1007/978-3-642-20398-5_35(2011)
13. Larsen, P.G., Lausdahl, K., Battle, N., Fitzgerald, J., Wolff, S., Sahara, S., Verhoef, M., Tran-Jørgensen, P.W.V., Oda, T., Chisholm, P.: The VDM-10 Language Manual. Tech. Rep. TR-2010-06, The Overture Open Source Initiative (2010)
14. Andrews, D., Bruun, H., Damm, F., Dawes, J., Hansen, B., Larsen, P., Parkin, G., Plat, N., Totenel, H.: A Formal Definition of VDM-SL. Tech. Rep. 1998/9, Leicester University (1998)
15. Lausdahl, K., Larsen, P.G., Battle, N.: A deterministic interpreter simulating a distributed real time system using VDM. In: Qin, S., Qiu, Z. (eds.) Proceedings of the 13th International Conference on Formal methods and Software Engineering. Lecture Notes in Computer Science, vol. 6991, pp. 179–194. Springer, Berlin. doi:10.1007/978-3-642-24559-6_14. ISBN 978-3-642-24558-9 (2011)
16. Leavens, G.T., Poll, E., Clifton, C., Cheon, Y., Ruby, C., Cok, D., Kiniry, J.: JML Reference Manual, revision 2344 edn. (2013)
17. Müller, P., Poetzsch-Heffter, A., Leavens, G.T.: Modular invariants for layered object structures. Sci. Comput. Program. **62**(3), 253–286 (2006). doi:10.1016/j.scico.2006.03.001
18. The Apache Maven Project website. https://maven.apache.org (2016)
19. The Overture tool Github repository. https://github.com/overturetool/overture (2016)
20. Yi, J., Robby, Deng, X., Roychoudhury, A.: Past expression: encapsulating pre-states at post-conditions by means of AOP. In: Proceedings of the 12th Annual International Conference on Aspect-oriented Software Development, AOSD '13, pp. 133–144. ACM. doi:10.1145/2451436.2451453 (2013)
21. McCarthy, J.: A Basis for a Mathematical Theory of Computation. In: Western Joint Computer Conference (1961)
22. Tran-Jørgensen, P.W.V.: Automated translation of VDM-SL to JML-annotated Java. Department of Engineering, Aarhus University, Tech. rep. (2016)
23. Larsen, P.G., Lausdahl, K., Battle, N.: Combinatorial testing for VDM. In: Proceedings of the 2010 8th IEEE International Conference on Software Engineering and Formal Methods, SEFM '10, pp. 278–285. IEEE Computer Society, Washington, DC, USA. doi:10.1109/SEFM.2010.32. ISBN 978-0-7695-4153-2 (2010)
24. Tran-Jørgensen, P.W.V., Larsen, P.G., Battle, N.: Using JML-based code generation to enhance test automation for VDM models. In: Proceedings of the 14th Overture Workshop (2016)
25. Fitzgerald, J., Larsen, P.G., Mukherjee, P., Plat, N., Verhoef, M.: Validated Designs for Object-Oriented Systems. Springer, New York (2005). doi:10.1007/b138800
26. Vilhena, C.: Connecting between VDM++ and JML. Master's thesis, Minho University with exchange to Engineering College of Aarhus (2008)
27. Jin, D., Yang, Z.: Strategies of modeling from VDM-SL to JML. In: International Conference on Advanced Language Processing and Web Information Technology, pp. 320–323 (2008)
28. Rumbaugh, J., Jacobson, I., Booch, G.: Unified Modeling Language Reference Manual, 2nd edn. Pearson Higher Education, Boston (2004)
29. Zhou, J., Jin, D.: Research on modeling from VDM-SL to JML for systematic software development. Control and Decision Conference (CCDC). 2010 Chinese, pp. 2312–2317. IEEE, Xuzhou (2010)
30. Larsen, P.G.: Ten years of historical development: "Bootstrapping" VDMTools. J. Univers. Comput. Sci. **7**(8), 692–709 (2001)
31. Rivera, V., Cataño, N., Wahls, T., Rueda, C.: Code generation for event-B. Int. J. Softw. Tools Technol. Transf. 19:1–22 (2015)
32. Abrial, J.R.: Modeling in Event-B: System and Software Engineering. Cambridge University Press, Cambridge (2010)

33. Lensink, L., Smetsers, S., van Eekelen, M.: Generating verifiable Java code from verified PVS specifications. In: Goodloe, A., Person, S. (eds.) NASA Formal Methods. Lecture Notes in Computer Science, vol. 7226, pp. 310–325. Springer, Berlin. doi:10.1007/978-3-642-28891-3_30 (2012)
34. Owre, S., Rushby, J.M., Shankar, N.: PVS: a prototype verification system. In: Kapur, D. (ed.) 11th International Conference on Automated Deduction (CADE). Lecture Notes in Artificial Intelligence, vol. 607, pp. 748–752. Springer, Saratoga (1992)
35. Filliâtre, J.C.: Why: a multi-language multi-prover verification tool. Research Report 1366, LRI, Université Paris Sud. http://www.lri.fr/~filliatr/ftp/publis/why-tool.ps.gz (2003)
36. Hubbers, E., Oostdijk, M.: Generating JML specifications from UML state diagrams. In: Forum on Specification and Design Languages FDL'03, Frankfurt, Germany, pp 263–273, September 23–26, 2003
37. Zhen, Z.: Java Card Technology for Smart Cards. Prentice-Hall, Boston (2000)
38. Klebanov, A.: Automata-based programming technology extension for generation of JML annotated Java card code. In: Proceedings of the Spring/Summer Young Researchers' Colloquium on Software Engineering, pp. 41–44 (2008)
39. Cok, D.R.: Reasoning with specifications containing method calls and model fields. J. Object Technol. **4**(8), 77–103 (2005)