Using Stepwise Feature Introduction in Practice: An Experience Report

Ralph-Johan Back, Johannes Eriksson, and Luka Milovanov

Turku Centre for Computer Science,
Åbo Akademi University, Department of Computer Science,
Lemminkäisenkatu 14, FIN-20520 Turku, Finland
{backrj, joheriks, lmilovan}@abo.fi

Abstract. Stepwise Feature Introduction is an incremental method and software architecture for building object-oriented system in thin layers of functionality, and is based on the Refinement Calculus logical framework. We have evaluated this method in a series of real software projects. The method works quite well on small to medium sized software projects, and provides a nice fit with agile software processes like Extreme Programming. The evaluations also allowed us to identify a number of places where the method could be improved, most of these related to the way inheritance is used in Stepwise Feature Introduction. Three of these issues are analyzed in more detail here: diamond inheritance, complexity of layering and unit testing of layered software.

1 Introduction

Stepwise Feature Introduction (SFI) [1] is a bottom-up software development methodology based on incremental extension of the object-oriented system with a single new feature at a time. It proposes a layered software architecture and uses Refinement Calculus [2, 3] as the logical framework.

Software is constructed in SFI in thin layers, where each layer implements a specific feature or a set of closely related features. The bottom layer provides the most basic functionality, with each subsequent layer adding more and more functionality to the system. The layers are implemented as class hierarchies, where a new layer inherits all functionality of previous layers by sub-classing existing classes, and adds new features by overriding methods and implementing new methods. Each layer, together with its ancestors, constitutes a fully executable software system.

Layers are added as new features are needed. However, in practice we cannot build the system in this purely *incremental* way, by just adding layer after layer. Features may interact in unforeseen ways, and a new feature may not fit into the current design of the software. In such cases, one must *refactor* the software so that the new feature fits better into the overall design. Large refactorings may also modify the layer structure, e.g. by changing the order of layers, splitting layers or removing layers altogether.

An important design principle of SFI is that each extension should preserve the functionality of all previous layers. This is known as *superposition refinement* [4]. A superposition refinement can add new operations and attributes to a class, and may override

N. Guelfi and A. Savidis (Eds.): RISE 2005, LNCS 3943, pp. 2–17, 2006.
© Springer-Verlag Berlin Heidelberg 2006

old operations. However, when overriding an old operation, the effect of the old operation on the old attributes has to be preserved (but new attributes can be updated freely). No operations or attributes can be removed or renamed.

Consider as an example a class that provides a simple text widget in a graphical user interface. The widget works only with simple ASCII text. A new feature that could be added as an extension to this widget could be, e.g., formatted text (boldface, italics, underlined, etc). Another possible extension could be a clipboard to support cut and paste. We could carry out both these extensions in parallel and then construct a new class that inherits from both the clipboard text widget and the styles text widget using multiple inheritance (this is called a *feature combination*), possibly overriding some of the operations to avoid undesirable feature interaction. Or, we could first implement the clipboard functionality as an extension of the simple text widget, being careful to preserve all the old features, and then introduce styles as a new layer on top of this. Alternatively, we could first add styles and then implement a clipboard on top of the styles layer. The three approaches are illustrated in Figure 1.

A component is divided into layers in SFI. Layers will often cut across components, so that the same layering structure is imposed on a number of related components. As an example, consider building an editor that displays the text widget. In the first layer we have a simple editor that only displays the simple ASCII text widget. Because of the superposition property of extension this simple editor can in fact also use the `CutAndPaste`, `Styles` or `BetterText` widgets, but it cannot make use of the new features. We need to add some features to the simple editor so that the functionality of the

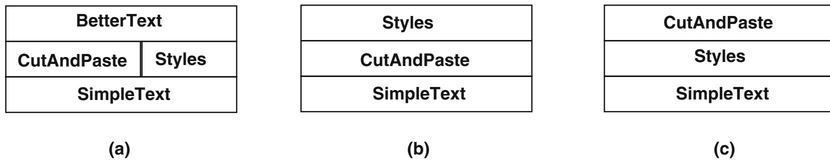

Fig. 1. Alternative extension orders

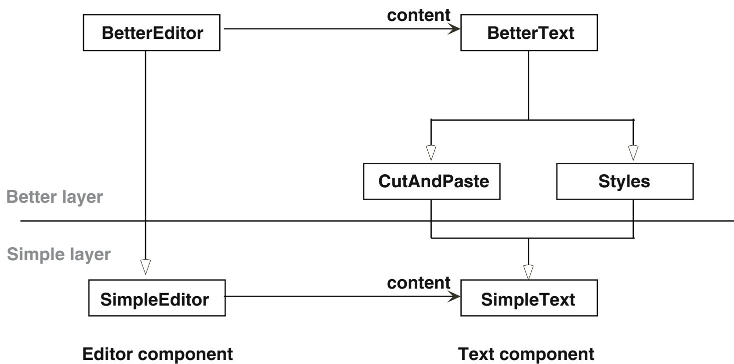

Fig. 2. Interacting components

extended widget can be accessed (menu items for cut and paste, or for formatting, or toolbar buttons for the same purpose). We do this by constructing a new an extension of the simple editor (a better editor), which uses the `BetterText` widget and gives the user access to the new functionality. The situation is illustrated in Figure 2.

The new editor is, however, restricted to only work on the better text widget, because the features it assumes are only available on this level. Hence, there are two layers in the design, the Simple layer and the Better layer.

Stepwise Feature Introduction has been tried out in a number of real software projects. This allows us now to evaluate the merits of this approach and to spot possible drawbacks as well as opportunities for improvement. Our purpose in this paper is to report on these case studies, and to provide a first evaluation of the approach, together with some suggestions on how to improve the method.

The paper is structured as following: Section 2 present the software projects where SFI was applied. We summarize our experience with the methodology in Section 3. In Sections 4–6 we then consider in more detail three interesting issues that arose from our experiments with Stepwise Feature Introduction. The problems with implementing feature combinations using multiple inheritance is discussed in Section 4. The problem of class proliferation is discussed in Section 5, where metaprogramming is considered as one possible way of avoiding unneccessary classes. In Section 6, we show how to adapt unit testing to also test for correct superposition refinement. We end with a short summary and some discussion on on-going and future work.

2 SFI Projects in Gaudi

The software projects where SFI was evauated were all carried out in the Gaudi Software Factory at Åbo Akademi University. The Gaudi Software Factory is an academic environment for building software for the research needs and for carrying out practical experiments in Software Engineering [5]. Our research group defines the setting, goals and methods to be used in the Factory, but actual construction of the software is done in the factory, following a well-defined software process. The work is closely monitored, and provides a lot of data and measures by wich the software process and its results can be evaluated. The software process used in Gaudi is based on agile methods, primarily *Extreme Programming* [6], together with our own extensions.

We will here describe four software projects where Stepwise Feature Introduction was used throughout. The settings for all these projects were similar: the software had to be built with a tight schedule, and the Gaudi software process had to be followed. The programmers employed for these projects (4–6 persons) were third-fifth year students majoring in Computer Science or related areas. Each project had a customer who had final saying on the functionality to be implemented. The projects were also supervised by a coach (a Ph.D. student specializing in Software Engineering), whose main task was to guide the use of the software process and to control that the process was being followed. There has also been one industrial software project [7] with SFI, but this is outside the scope of this paper, as it was not carried out in the Gaudi Factory, and the software process used was not monitored in a sufficiently systematic manner.

All of the projects used SFI, but the ways in which the method was applied differed from project to project. We describe the projects in chronological order below. For each project, we present the goals: both for the software product that was to be built, and for the way in which SFI was to be evaluated in this project. We give a general overview of the software architecture, show how SFI was implemented, what went right and what went wrong, and discuss the lessons learned from the project.

2.1 Extreme Editor

The Extreme Editor project [8] was the first application of SFI in practice. It ran for three months during the summer 2001 and involved six programmers. The programming language of the project was Python [9]. The software product to be built was an outlining editor which became a predecessor for the Derivation Editor described in Section 2.2. The goal for the project was to obtain the first experience from practical application of SFI with a dynamically typed programming language. There were no technical guidelines for the application of SFI except that the extension mechanism for classes (the feature introduction—Section 1) should be inheritance.

Figure 3 shows the layered architecture of the Extreme Editor. There were eight layers in the system. Each layer introduced new functionality into the system, without breaking the old features. The software was structured into these layers in an ad hoc way. A new layer extended its predecessor by inheriting its corresponding classes and possibly introducing one or more new classes. There were no physical division of the

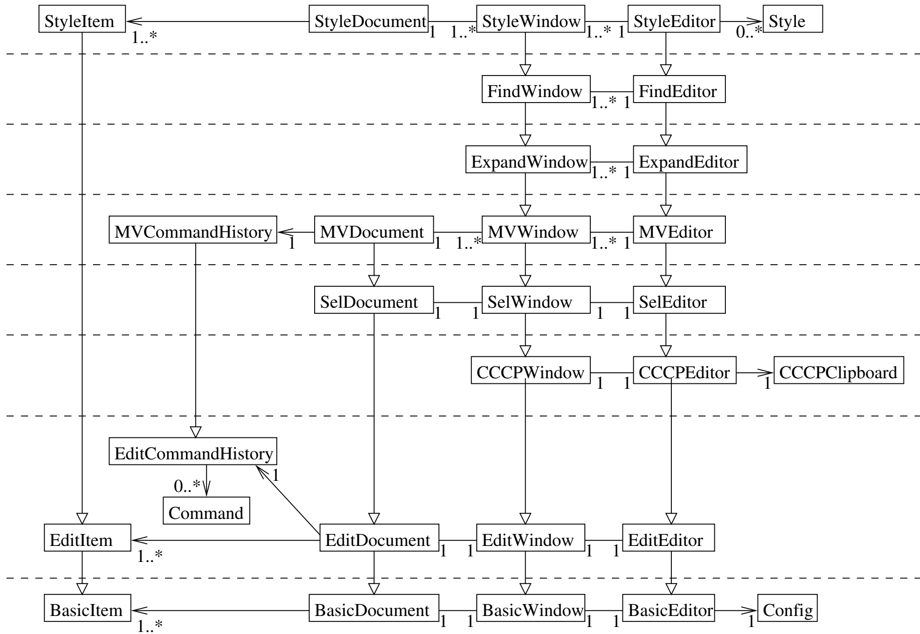

Fig. 3. The layers of the editor

software into the layers on the level of the file system: each class name had a prefix—the name of the layer where the class belongs to. More on the architecture of the software and detailed description of the layers can be found in the technical report [8].

The feedback on SFI from this project was rather positive. The method supported building software with a layered architecture quite well. The developers found it rather easy to add new features as new layers. The fact that functionality of the system was divided into layers made the overall structure of the system clearer to all the programmers. Another advantage was "bug identification": the layered structure made it easy to find the layer in which the bug occured and its location in the source code.

On the other hand, even in the three-month project, as more features were implementer and the more classes were introduced into the system, it was getting harder to navigate among them without any tool support, any automatic documentation describing the layer structure and without a systematic naming convention for classes, layers and methods. We also found that SFI requires a special way of unit testing (the testing classes should be extended in the same way as the ordinary classes of the system). More on unit testing will be discussed in Section 6.

2.2 Editor for Mathematical Derivations

MathEdit [10] was an effort to implement tool support for structured derivations and is currently the largest project developed using the SFI methodology. MathEdit was developed in the Gaudi Software Factory as two successive summer projects, in 2002 and 2003. One of the objectives of the first summer project was to try out feature combination by multiple inheritance. The second objective of the project was to assess how well a new team could embrace an existing SFI codebase and continue its development. The continuation project in 2003 shared only one out of four developers with the 2002 project. Still, it turned out that the new programmers were able to start working productively with the existing code base within the first three weeks.

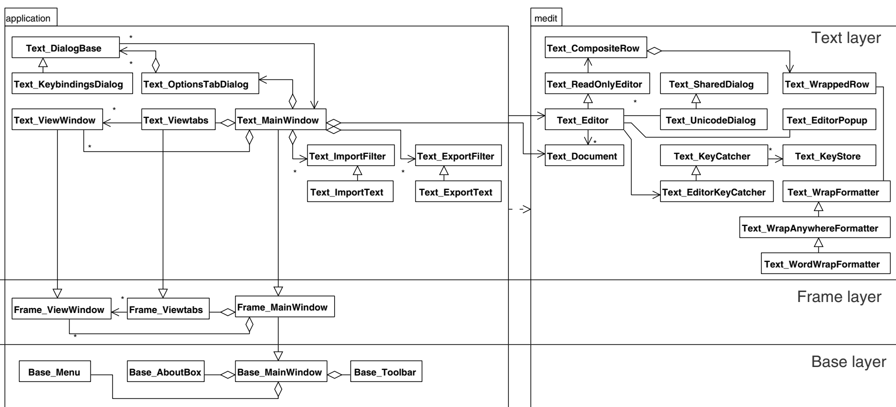

Fig. 4. The first three layers of MathEdit, unit tests excluded

MathEdit consists of totally 16 layers. A description of each layer and its major features can be found in [10]. Classes are associated with layers based on naming conventions. A diagram showing the classes of the first three layers (unit test classes excluded) can be seen in Figure 4. Due to space considerations, we do not display a complete class diagram for MathEdit.

On the highest level, MathEdit is separated into three major components: a document-view component (`medit`), the application-level user interface (`application`) and a mathematical profile plug-in. The `medit` and `application` components are layered as described above. The layering cross-cuts these top-level components, so the layering is *global*. The profile component was designed as a plug-in, so it was not layered—users should be able to write custom profiles without having to care about the internal layer structure of MathEdit.

Combining two layers using multiple inheritance was attempted but ultimately abandoned in the MathEdit project. The main reason were practical problems arising from the use of multiple inheritance; e.g., classes in the graphical toolkit used (Qt) did not support multiple inheritance well. Also, the development team was quite small, so it did not seem fruitful to work on two features in parallel as if they were independent, when it was already known that the features would be combined later on. Instead, a feature was implemented with extensibility in mind, so that it was easy to add the next feature in a new layer. The gains of parallel development would probably be much higher in a larger project, but this still remains to be evaluated.

The MathEdit application is executable with any layer as the top layer, providing the functionality of this layer and all layers below. The gives us the possibilities to fall back to an earlier working version in case of malfunction, and to locate bugs that were introduced in a some unknown layer. We simply implement a test that exposes the bug and run it with different top layers. The lowest layer that exhibits the erroneous behavior is either harboring the bug, or triggering a bug in a previous layer.

2.3 Software Construction Workbench

The Software Construction Workbench (SCW) project (summer – fall 2002) was an effort to build a prototype for IDE supporting Stepwise Feature Introduction and Python. This application was built in Python, as an extension of the System Modeling Workbench [11]. The main feature are modeling software systems in a SFI fashion with UML, automatic Python code generation and execution of the constructed system and support for unit tests in the environment. SCW also included basic support for Design by Contract [12] and reverse engineering.

As far as SFI is concerned, the goals in this project were to get further feedback on the practical application of the methods, to try out the layered approach to unit testing (discussed in Section 6) and to try out new naming conventions (described below in this section). A special feature of this project was that it used its own medicine: the software needed to support software development with SFI was built itself using this method. Since the architecture of the SCW is rather complex, we do not present it here, for the reader's convenience. Instead, we illustrate how the SFI method was applied in this project on a small and simplified fragment of the software.

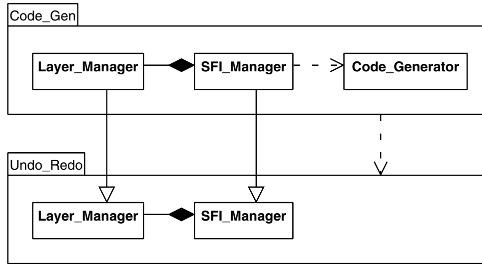

Fig. 5. Layer structure, Python implementation

Figure 5 shows an example of the layer structure implementation. SFI layers were implemented using Python's packaging mechanism: each SFI layer corresponds to a Python package. To show that a layer is a successor of another layer we draw a dependency from the successor to its ancestor; in practice this dependency was the Python `import` statement. The mechanism for extension of classes was inheritance, as shown in Figure 5. Every class, once introduced, keeps its original name through all of its extensions. This was simple to implement: Python packages provide a namespace so we had no conflicts with the names of the classes.

A class can be extended with new methods and/or some inherited methods can be overwritten. When an old method is overwritten in the next class extension, it is a good practice to have a call to the original method inside the body of the extended method. According to the developers, the implemented layered structure of the software together with the name conventions really clarified the software. The Software Construction Workbench project was carried out as two subprojects, such that half of the developers were new in the second subproject, and in the beginning had no understanding of the software at all. Nevertheless, the new programmers were able to take over the code easily because of the division of features into layers. The new programmers also commented that the layering made it much easier to navigate in the code, modify code and search for bugs.

The SCW project showed that in order to use the SFI methods properly and efficiently, tool support is really needed. Because of the way Python packaging was used to implement the SFI layers, it took a lot of time to divide the code into directories corresponding to the layers. Building software according to SFI also promotes refactoring (a practice enforced in our Software Factory). For example, when changing something in the lower layer, it can often affect the successive layers, so they should be slightly changed. According to the developers a simple tool helping with the navigation among the layer structure, i.e. from ancestor to successor and the other way around, would here save a lot of time.

2.4 Personal Financial Planner

The Personal Financial Planner project [13] (FiPla) was the first application of SFI with a statically typed language—Eiffel [14]. The software goal for the project was to build a personal financial planner. The features required of this product type include

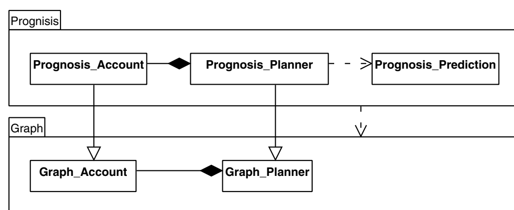

Fig. 6. Layer structure, Eiffel implementation

tracking of actual events (manually or automatically), planning (such as budgeting and creating scenarios), and showing future scenarios.

SFI was evaluated in this project to see how the method would work with a statically typed language. Another goal was to see how well the SFI layers layers correspond to the short release cycles used in the Gaudi Software Factory [5], so that each short iteration starts a new layer. Finally we also wanted to see how well SFI and Design by Contract [12] fit together.

SFI layers in Eiffel are implemented using Eiffel clusters. However, unlike Python packages, a cluster in Eiffel does not provide a namespace for the classes. It means that all names of the classes in the Eiffel software system should have unique name, so it was impossible to have the same naming convention as in the SCW project. For this purpose we used another naming convention, where each class name should have the name of the layer that the class belongs to as a prefix. Figure 6 shows a simplified fragment of the software architecture.

SFI worked well with Eiffel when we applied the methods in the same way as in our Python projects. Structuring the software system into layers according to the small releases defined by the customer turned out to be a good idea. However, a few important technical issues that needed improvement were found. These issues only came up when using SFI with a statically typed language.

The extension of classes using pure inheritance did not work well with Eiffel. The types of the parameters and return value of redefined routines should be at least of conforming types. Eiffel does not support parametrized polymorphism, hence, the number of parameters should be constant in all extensions of a routine. It is possible to overcome these limitations using routine renaming or rewriting a routine completely previously undefining it with Eiffel's *undefine* statement. The last case is, however, not recommended since it will not be a real extension of the routine.

To avoid these problems one must pay more attention to the overall system architecture and plan a bit further ahead then just for the next iteration. Extensive refactoring was needed in some cases, when the planning had not been done carefully enough. To refactor a SFI system efficiently, tool support was again deemed necessary.

3 Experience of Using SFI

In this section we summarize our experience from using SFI in the software projects mentioned above, based on the quantitative and qualitative data that was collected

Table 1. Basic product metrics for SFI projects

	EE	MathEdit	SCW	FiPla
LOC	3300	44372	16334	8572
Test LOC	1360	4198	14565	2548
Total LOC	4660	48570	30899	11120
Classes	52	427	66	59
Test classes	23	53	42	25
Methods	344	3790	610	331
Test methods	85	279	355	177
LOC/class	63	104	247	145
LOC/test class	59	79	347	102
Methods/class	6.6	8.9	9.2	6.0
Test methods/class	3.7	5.3	8.5	7.0

during these projects. Table 1 shows some basic code metrics for the presented SFI projects. It is easy to see that even in small projects like Extreme Editor and FiPla, the number of classes is growing fast. On the one hand this helps with debugging: as the developers were often commenting, it is easy to find the source of a bug in the code because of the separation of functionality into the layers. On the other hand, managing a large number of classes manually eventually becomes quite complicated, suggesting that tool support for navigation among successive layers, classes and methods is needed.

SFI is a bottom-up approach for constructing software systems and is therefore not that well suited for developing graphical user interfaces. Constructing good GUIs is a complicated task in itself and needs to combine different approaches such as bottom-up, top-down, use of state charts and designer tools. Our experience showed that when using SFI, it is better to separate GUI development from the construction of the core of the system.

Stepwise Feature Introduction fits well in our software process and in general in the Extreme Programming philosophy of introducing small changes one at a time in a software project [5, 15]. The division of the system into layers can be driven by the XP iteration planning process. When the development team negotiates with the customer about what new features should be implemented for the next small release, it is then easy to see what should be included in the next layer. Every layer in SFI system together with its ancestor layers represent a functional, working system. Hence, each layer corresponds to a small release, making it easy to package a specific release so that the customer can do the acceptance testing.

We obtained good results regarding the practical usability of SFI from our projects. SFI has a formal basis and provides a sound way of structuring software, and SFI designs often capture the core concerns of the software (the features) more explicitly than many traditional OO designs. We have also identified some shortcomings in the method that we need to work on. The use of inheritance as an extension mechanism can be cumbersome and does introduce some complexity of its own into the system. SFI occasionally makes it difficult to add a feature that does not fit well into the layer hierarchy. In order to make SFI practically usable, it will be necessary to devise another extension

mechanism or introduce SFI-aware development tools (such a tool is currently being worked on, as explained in Section 7).

In the remaining sections, we will discuss in more depth three specific issues that came up during our experiments, and which we think merit a much closer analyzis.

4 Feature Combination and Diamond Inheritance

SFI suggests combining two or more independently developed layers into a feature combination layer (see Section 1). As each layer may contain an extension of the same class, the feature combination layer combines the extensions of a class from each layer into a new subclass using multiple inheritance. It is then the responsibility of the new layer to synchronize the two independent features in a meaningful way.

Multiple inheritance is significantly more complex than single inheritance for both language implementors and programmers. What constitutes correct use of multiple inheritance in object-oriented software is a subject of some controversy. Bir Singh [16] lists the four main uses of multiple inheritance, none of which correspond to the way it is used in SFI (combination of two implementations with a potentially large number of common methods):

- combination of multiple independent protocols;
- mix and match, where two or more classes are designed specially for combination;
- submodularity, to factor out similar subparts for improved system design;
- separation of interface and implementation.

This may suggest that multiple inheritance as implemented in most programming languages might not be ideal for feature combination as originally proposed in SFI. We will here focus on one serious design problem encountered numerous times in our experiments, namely *diamond inheritance*.

Diamond inheritance occurs when two or more ancestors of a class share the same base class. This situation arises fairly often in large systems, especially if the class hierarchy has a common root. In SFI diamond inheritance is likely to occur because of the way inheritance is used as a layer extension mechanism, especially if one uses the suggested feature combination.

An example of a situation in which the diamond pattern typically occurs in an SFI design can be seen in Figure 7. The Basic layer contains two classes, `BasicAccount` and a derived class `BasicCheckingAccount`, the latter supposedly having some extended behavior such as allowing withdrawals greater than the balance. In this case the programmer used inheritance to be able to handle objects of the two account types uniformly, i.e. to achieve polymorphism. Let us now assume that in the next layer (the Better layer), support for multiple currencies is added, and that this feature requires both `BasicAccount` and `BasicCheckingAccount` to be extended into `BetterAccount` and `BetterCheckingAccount` respectively. However, to preserve polymorphism, `BetterAccount` must also be extended to `BetterCheckingAccount`. We notice that mixing the two usage patterns of inheritance results in a diamond structure in the design.

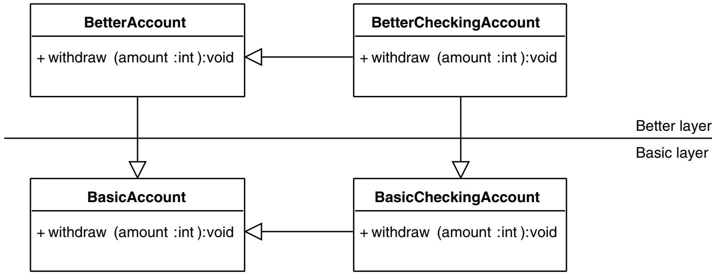

Fig. 7. Diamond inheritance

Diamond inheritance causes difficulty when the same method is implemented in more than one base class. In this example the `withdraw` method of `BetterCheckingAccount` depends on functionality implemented in both `BasicCheckingAccount` and `BetterAccount`, and calls both (in some order). However, each of these calls trigger a call to `BasicAccount.withdraw`, resulting in two calls to this method. The code in `BasicAccount.withdraw` is thus executed twice, which is not the intended behavior (the sum is withdrawn twice from the account). The same situation occurs commonly with constructors—the constructor of the common base class in the diamond is called twice. This results in data structures and resources being initialized twice, potentially leading to resource leaks.

In the example case, when implementing `BetterCheckingAccount.withdraw` we want to call the `withdraw` method of both base classes, but `BasicAccount.withdraw` must be called only once. In Python 2.3, which was actually designed with inheritance diamonds in mind [17], this is possible using the built-in `super` function which creates a linearization of the class hierarchy and returns for a given class the previous class in the linearization. By replacing direct calls to `__init__` with `super` the desired call order can be achieved. The drawback is that since we are no longer explicitly calling the base class method, we might not be sure which method actually gets called without considering the linearization of the whole class hierarchy. Because this would make the class design more complex we have not used `super` in any of our Python projects.

Most SFI projects developed in the Gaudi Software Factory have avoided diamond inheritance by not using multiple inheritance for feature combination or for mixing polymorphic extension with stepwise extension. Consequently we have not been able to add features to a base class in a polymorphic inheritance hierarchy using SFI layers. However, many times features are better implemented using *delegation*, where an object uses another object (the delegatee) to perform an operation. In this case we can simply replace the delegatee with a more advanced version in a higher layer. E.g., MathEdit heavily uses the *Bridge* and *Decorator* design patterns [18] to avoid inheritance diamonds.

5 Avoiding Trivial Classes with Metaprogramming

One problem discovered early on was that some SFI-supporting metaprogramming framework had to be implemented to reduce complexity that was primarily caused by

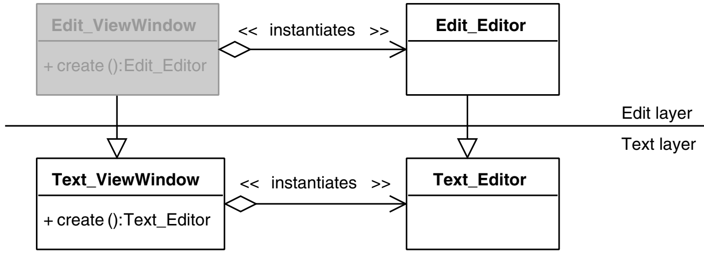

Fig. 8. Subclassing to override factory method

proliferating factory methods. This problem occurs whenever one class (the factory) is responsible for creating instances of another class (the product), and subclassing the product to add a new feature results in having to subclass the factory only to override the factory method so that it creates an instance of the new product class. An example from MathEdit is illustrated in Figure 8.

The class `Text_ViewWindow` creates an object of type `Text_Editor` in the Text layer. In the Edit layer the `Editor` class is extended with a feature that does not affect the behavior of `ViewWindow` in any other way than that it now has to instantiate objects of type `Edit_Editor` instead of `Text_Editor`. A class `Edit_ViewWindow` (grayed out) has to be introduced only to override the factory method that creates the `Editor` object.

Since frequent subclassing and deep inheritance hierarchies are commonplace in SFI designs, this situation will occur whenever a product class is subclassed and results in a large number of trivial factory subclasses, cluttering the design and increasing the code size. To avoid introducing these subclasses the metaprogramming framework of MathEdit implements a routine that given a class name returns the correct Python class for the running layer (Python classes are first-class objects which can easily be passed around; in more static languages one might need to do this in compile time using e.g. macro substitutions). If a certain class is not extended in the current layer, the routine searches backwards in the layer stack until it finds the most recent class definition. For example, calling the routine to get the `ViewWindow` class when running layer Edit would return `Text_ViewWindow`.

A substantial problem we encountered with deep inheritance hierarchies is that the control flow through the call chains of overridden methods becomes difficult to overview. The programmers generally thought it was hard to grasp the order of method calls and how the object state changes in response to the calls, especially with many nested method calls. Also, finding a method declaration in the code could require searching through several classes in the inheritance chain, unless the programmer knew exactly in which layer the method was implemented. One programmer commented that "a drawback with having so many hierarchical levels is that you start to forget methods and variables that you defined on the lowest levels."

Our experience indicates that the refactoring stage of SFI is of very high importance for keeping the design clean. Especially when working with unstable requirements, features can not easily be added on top of each other. The programmers found some of the refactorings to be difficult and error-prone, partly because inheritance creates a

rather tight coupling between classes. However, a good test harness will substantially aid in the detection of such errors.

6 Unit Testing of Superposition

Unit testing is testing of individual hardware or software units or groups of related units [19]. Extensive, automated unit testing has been proposed as an efficient way of detecting errors introduced by changes in the software [6, 20]. A *unit test* exercises some subset of the software's behavior and validates it against its specification. Unit testing frameworks frequently group *test methods* into *test case classes*, which can further be aggregated into *test suites*. The complete *test harness*, consisting of all test suites, can then be executed with a single command.

SFI architectures should maintain the superposition refinement relationship between extensions and their bases—class invariants established in previous layers should not be violated in subsequent layers. A layered unit testing architecture allows us to easily create and maintain a test suite that aids in the detection of such violations, typically caused by programmer error or design mistakes. By writing tests for only new functionality and inheriting existing testing functionality, we introduce the requirement that a test introduced in one layer should also pass in all subsequent layers.

Most of our projects have utilized a unit testing architecture based on inheriting test cases. Our experience has shown it to be useful in practice; especially with many layers programmers easily forget assumptions and requirements introduced in a lower layer—if these are reflected in unit tests for the lower layer, possible violations are detected also when running tests for higher layers.

We assume that for testing we use a unit testing framework that provides us at least with a test case class. When constructing test cases in bottom layers, all test cases inherit the class from the testing framework. Test cases of the extended classes in successive layers should be extensions of the test cases from previous layers using the same extension mechanism as the application classes—inheritance. If an inherited method of an application class is overridden and extended with new functionality, the corresponding test method should be extended accordingly. If the body of the extended method contains a call to its ancestor method, the same technique should be applied in the body of the corresponding test method. This allows us to test both new and old functionality by writing tests just for the new functionality.

An example of a basic testing scenario with two layers can be seen in Figure 9. The Simple layer contains one application class (`SimpleText`) and its associated test class (`SimpleTextTest`), which tests the `insert` method of `SimpleText`. The Better layer extends `SimpleText` into `BetterText` by overriding `insert` and adding the `paste` method; correspondingly the `BetterTextTest` test case extends `SimpleTextTest` to override `testInsert` and adds a new test method for the paste method (`testPaste`). The new `testInsert` method should test directly only the new functionality introduced in `BetterText`, it should call the `testInsert` method from the Simple layer to test that the old functionality of insertion is preserved in `BetterText`. In this way, we test that `BetterText` is in fact a superposition refinement of `SimpleText`.

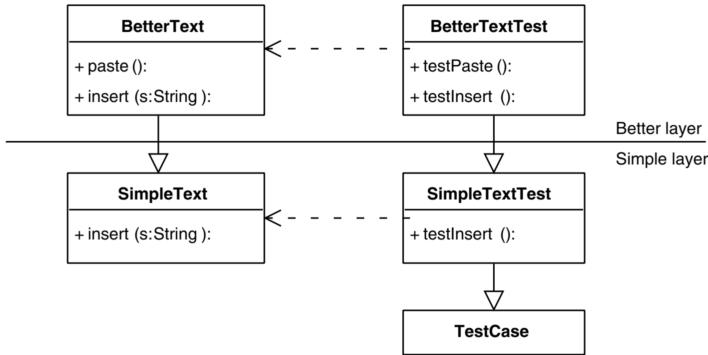

Fig. 9. Layered test cases

We have used the PyUnit [21] unit testing framework, which is essentially a Python version of the JUnit testing framework for Java [20]. The programmers found it easy to write tests in Python using PyUnit. No special compilation cycle for tests is required, and grouping all tests into a single suite makes it easy to run the tests often; programmers were encouraged to run the tests before committing any new code to the main source.

A number of open source testing frameworks for Eiffel are available. The Gobo [22] tool was used in our project for unit testing. Gobo test cases work similarly to PyUnit; the programmer subclasses a predefined test class and adds test methods. However, because the Gobo test framework is not integrated into the EiffelStudio environment, it was necessary to set up two different projects, one for compiling the system and one for compiling the tests and the system. The programmers found this arrangement inconvenient.

7 Conclusion and Future Work

We have above described our results from using Stepwise Feature Introduction, a formally defined method, in practical software engineering projects. The central idea in SFI is that layers are built stepwise as superposition refinements on top of each other; using class inheritance as the extension mechanism. Also, each layer together with lower layers should constitute a fully executable application.

We have carried out several case studies in Stepwise Feature Introduction. Our experience from these studies has shown us that SFI works well for structuring, debugging and testing the software under development. Combining SFI with an agile process like Extreme Programming provides architectural structure and guidance to an otherwise quite ad hoc software process, and has allowed us to deliver good working software in a timely manner. It is easy for developers to learn how to apply SFI, and the layer structure helps developers to understand the software architecture.

The main difficulties in applying the method have been caused by lack of automation and, to some extent, conflicting use of class inheritance. These observations point to a

need for a generic SFI-supporting programming environment. Many of the programming tasks involved in applying SFI can require considerable amount of time. However, most of them can be automated, which would provide a great help for the programmers. The Software Construction Workbench (Section 2.3) was the first attempt to build tool support for SFI, but it was Python-specific. A number of smaller case studies also showed that SCW somewhat too rigidly restricted the software architecture.

SFI-style extensions adds a new dimension to software diagrams, which can become quite large and difficult to overview. We are currently building and experimenting with SOCOS, a prototype tool for constructing and reasoning about software systems, that is intended to support SFI. SOCOS is essentially an editor for *refinement diagrams* [3]. Refinement diagrams have exact semantics and a mathematical base in lattice theory and refinement calculus. A software system is presented to the user as a three-dimensional diagram containing *software parts* and *dependencies* between parts.

The SOCOS system is currently in early stages and the framework is still being worked on. Our current focus is on developing an environment for constructing layered software systems and reasoning about their correctness on both architectural and behavioral levels. Stepwise Feature Introduction, using either inheritance or another layer extension mechanism, is intended to be the main method by which features are added to the system. The goal is to create a tool for correctness-preserving, incremental construction of SFI-layered software systems.

References

1. Back, R.J.: Software construction by stepwise feature introduction. In: ZB '02: Proceedings of the 2nd International Conference of B and Z Users on Formal Specification and Development in Z and B, Springer-Verlag (2002) 162–183
2. Back, R.J.J., Akademi, A., Wright, J.V.: Refinement Calculus: A Systematic Introduction. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1998)
3. Back, R.J.: Incremental software construction with refinement diagrams. Technical Report 660, TUCS – Turku Centre for Computer Science, Turku, Finland (2005)
4. Back, R.J., Sere, K.: Superposition refinement of reactive systems. Formal Aspects of Computing **8** (1996) 324–346
5. Back, R.J., Milovanov, L., Porres, I.: Software development and experimentation in an academic environment: The Gaudi experience. In: Proceedings of the 6th International Conference on Product Focused Software Process Improvement – PROFES 2005, Oulu, Finland. (2005)
6. Beck, K.: Extreme Programming Explained: Embrace Change. The XP Series. Addison-Wesley (1999)
7. Anttila, H., Back, R.J., Ketola, P., Konkka, K., Leskela, J., Rysä, E.: Coping with increasing SW complexity - combining stepwise feature introduction with user-centric design. In: Human Computer Interaction, International Conference (HCII2003), Crete, Greece (2003)
8. Back, R.J., Milovanov, L., Porres, I., Preoteasa, V.: An experiment on extreme programming and stepwise feature introduction. Technical Report 451, TUCS – Turku Centre for Computer Science, Turku, Finland (2002)
9. Lutz, M.: Programming Python. O'Reily (1996)
10. Eriksson, J.: Development of a mathematical derivation editor. Master's thesis, Åbo Akademi University, Department of Computer Science (2004)

11. Back, R.J., Björklund, D., Lilius, J., Milovanov, L., Porres, I.: A workbench to experiment on new model engineering applications. In Stevens, P., Whittle, J., Booch, G., eds.: UML 2003 – The Unified Modeling Language. Model Languages and Applications. 6th International Conference, San Francisco, CA, USA, October 2003, Proceedings Volume 2863 of LNCS, Springer (2003) 96–100
12. Meyer, B.: Object-Oriented Software Construction. second edn. Prentice Hall (1997)
13. Back, R.J., Hirkman, P., Milovanov, L.: Evaluating the XP customer model and design by contract. In: Proceedings of the 30th EUROMICRO Conference, IEEE Computer Society (2004)
14. Meyer, B.: Eiffel: The Language. second edn. Prentice Hall (1992)
15. Back, R.J., Milovanov, L., Porres, I., Preoteasa, V.: XP as a framework for practical software engineering experiments. In: Proceedings of the Third International Conference on eXtreme Programming and Agile Processes in Software Engineering - XP2002. (2002)
16. Singh, G.B.: Single versus multiple inheritance in object oriented programming. SIGPLAN OOPS Mess. **6** (1995) 30–39
17. Simionato, M.: The Python 2.3 method resolution order. http://www.python.org/2.3/mro.html (2003)
18. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley (1995)
19. Institute of Electrical and Electronics Engineers: IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York (1990)
20. Beck, K., Gamma, E.: Test-Infected: Programmers Love Writing Tests. Java Report (1998) 37–50
21. Purcell, S.: PyUnit. http://pyunit.sourceforge.net/ (2004)
22. Bezault, E.: Gobo Eiffel Test. http://www.gobosoft.com/eiffel/gobo/getest/ (2001)