

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 December 2008 (11.12.2008)

PCT

(10) International Publication Number
WO 2008/151013 A2

(51) International Patent Classification:
G06F 12/00 (2006.01)

(21) International Application Number:
PCT/US2008/065312

(22) International Filing Date: 30 May 2008 (30.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/809,514 1 June 2007 (01.06.2007) US

(71) Applicant (for all designated States except US): **Microsoft Corporation** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **Callahan, David**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **Grover, Vinod K.**; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA,

CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

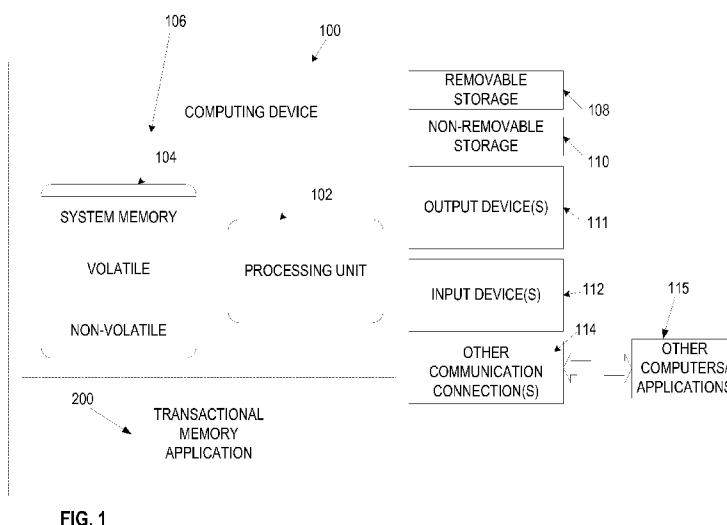
Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- without international search report and to be republished upon receipt of that report

(54) Title: EFFICIENTLY LOCATING TRANSACTIONAL CODE BLOCKS IN A TRANSACTIONAL MEMORY SYSTEM



(57) Abstract: Various technologies and techniques are disclosed for creating and/or locating transactional code blocks in a transactional memory system. A user such as a software developer can decorate a particular function with an identifier to indicate that the particular function is transaction-safe. A normal version and a transactional version are then created for each function of a software application that is marked as transaction-safe. A normal version is created for each function that is not marked as transaction-safe. For the normal version of each function that is marked as transaction-safe, a stub pointer in the normal version is pointed to the transactional version. The proper version of the function is then called depending on the execution context.

EFFICIENTLY LOCATING TRANSACTIONAL CODE BLOCKS IN A TRANSACTIONAL MEMORY SYSTEM

BACKGROUND

[001] Software transactional memory (STM) is a concurrency control mechanism
5 analogous to database transactions for controlling access to shared memory in
concurrent computing. A transaction in the context of transactional memory is a
piece of code that executes a series of reads and writes to shared memory. STM is
used as an alternative to traditional locking mechanisms. Programmers put a
declarative annotation (e.g., atomic) around a code block to indicate safety
10 properties they require and the system automatically guarantees that this block
executes atomically with respect to other protected code regions. The software
transactional memory programming model prevents lock-based priority-inversion
and deadlock problems.

[002] Software transactional memory (STM) systems can take certain sequential
15 programs and allow for portions of those programs to be executed concurrently
(e.g., in parallel) using transactions. The source code for these sequential programs
is typically written by programmers using one of various types of programming
languages. The source code is typically enclosed in one or more functions that
contain the logic that is later executed by a computer. The term “function” is used
20 broadly herein as covering functions, methods, procedures, statement blocks,
and/or other portions of logic that are executed by a computer. With software
transactional memory systems, every function that can be called from a
transactional context must have two versions, one that can be called from
transactions and one that can be called from non-transactions. Determining which
25 version of the function to call is context dependent. This determination must be
done at runtime for calls to virtual functions or through function pointers.

SUMMARY

[003] Various technologies and techniques are disclosed for creating and/or
locating transactional code blocks in a transactional memory system. A user such
30 as a software developer can decorate a particular function with an identifier to
indicate that the particular function is transaction-safe. A normal version and a

transactional version are then created for each function of a software application that is marked as transaction-safe. A normal version is created for each function that is not marked as transaction-safe. For the normal version of each function that is marked as transaction-safe, a stub pointer in the normal version is pointed to the transactional version. The proper version of the function is then called depending on the application context.

[004] In one implementation, a compiler generates the transactional and non-transactional version of the functions for use with the transactional memory system. At compile time, a stub pointer is allocated for each function in a software application. For each respective function that is not marked as a transaction-safe function, a normal version of the respective function is generated, and the stub pointer is filled in with an entry point of a runtime error routine. For each respective function that is marked as a transaction-safe function, code is created for a normal version and a transactional version of the respective function that is transaction-safe. Then, in the stub pointer for the normal version, an entry point is filled to the transactional version. In a stub pointer for the transactional version, an entry point is filled to the runtime error routine.

[005] This Summary was provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[006] Figure 1 is a diagrammatic view of a computer system of one implementation.

[007] Figure 2 is a diagrammatic view of a transactional memory application of one implementation operating on the computer system of Figure 1.

[008] Figure 3 is a high-level process flow diagram for one implementation of the system of Figure 1.

[009] Figure 4 is a process flow diagram for one implementation of the system of Figure 1 illustrating the stages involved in using a code generator to generate code

for transactional and non-transactional function and fill in the stub pointers as part of the compile and/or linking process.

[010] Figure 5 is a diagram for one implementation of the system of Figure 1 illustrating the decision tree for choosing the proper version of the function to call
5 depending on execution context.

[011] Figure 6 is a logical diagram for one implementation that illustrates two hypothetical functions, one with just a normal version and the other with a normal version and a transactional version.

DETAILED DESCRIPTION

10 [012] For the purposes of promoting an understanding of the principles of the invention, reference will now be made to the embodiments illustrated in the drawings and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope is thereby intended. Any alterations and further modifications in the described embodiments, and any further
15 applications of the principles as described herein are contemplated as would normally occur to one skilled in the art.

[013] The system may be described in the general context as a transactional memory system, but the system also serves other purposes in addition to these. In one implementation, one or more of the techniques described herein can be
20 implemented as features within a framework program such as MICROSOFT® .NET Framework, or from any other type of program or service that provides platforms for developers to develop software applications. In another implementation, one or more of the techniques described herein are implemented as features with other applications that deal with developing applications that execute
25 in concurrent environments.

[014] In one implementation, a transactional memory system is provided that programmatically creates two versions of each function that is marked as transaction-safe: a normal version of the function and a transactional version of the function. For functions that are not marked as transaction safe, only a normal
30 version is created and a stub pointer at the entry point of the normal version is pointed to an error routine. Then, for functions that are marked as transaction-safe,

a stub pointer of the entry point of the normal version is pointed to the transactional version of the function. A stub pointer at the entry point of the transactional version is pointed to an error routine. For functions that are not marked as transaction-safe, the stub pointer at the entry point of the normal version points to the error routine. The proper version of the function is then called depending on the execution context.

[015] As shown in Figure 1, an exemplary computer system to use for implementing one or more parts of the system includes a computing device, such as computing device 100. In its most basic configuration, computing device 100 typically includes at least one processing unit 102 and memory 104. Depending on the exact configuration and type of computing device, memory 104 may be volatile (such as RAM), non-volatile (such as ROM, flash memory, etc.) or some combination of the two. This most basic configuration is illustrated in Figure 1 by dashed line 106.

[016] Additionally, device 100 may also have additional features/functionality. For example, device 100 may also include additional storage (removable and/or non-removable) including, but not limited to, magnetic or optical disks or tape. Such additional storage is illustrated in Figure 1 by removable storage 108 and non-removable storage 110. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Memory 104, removable storage 108 and non-removable storage 110 are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by device 100. Any such computer storage media may be part of device 100.

[017] Computing device 100 includes one or more communication connections 114 that allow computing device 100 to communicate with other

computers/applications 115. Device 100 may also have input device(s) 112 such as keyboard, mouse, pen, voice input device, touch input device, etc. Output device(s) 111 such as a display, speakers, printer, etc. may also be included. These devices are well known in the art and need not be discussed at length here. In one
5 implementation, computing device 100 includes transactional memory application 200. Transactional memory application 200 will be described in further detail in Figure 2.

[018] Turning now to Figure 2 with continued reference to Figure 1, transactional memory application 200 operating on computing device 100 is illustrated.

10 Transactional memory application 200 is one of the application programs that reside on computing device 100. However, it will be understood that transactional memory application 200 can alternatively or additionally be embodied as computer-executable instructions on one or more computers and/or in different variations than shown on Figure 1. Alternatively or additionally, one or more parts of
15 transactional memory application 200 can be part of system memory 104, on other computers and/or applications 115, or other such variations as would occur to one in the computer software art.

[019] Transactional memory application 200 includes program logic 204, which is responsible for carrying out some or all of the techniques described herein.

20 Program logic 204 includes logic for providing a transactional memory system 206; logic for allowing a user/developer to decorate functions of a software application with an attribute or other identifier to indicate the function is transaction-safe 208; logic for creating two versions (e.g., a normal version and a transactional version) for each function that is marked as transaction-safe 210; logic for creating one
25 version (e.g., a normal version) for each function that is not marked as transaction-safe 212; logic for allocating a stub pointer for each function (e.g., before the beginning of the code section or other place) 214; logic for making the stubs point to the right places (e.g., the transactional version or the runtime error routine) 216; and other logic for operating the application 220. In one implementation, program
30 logic 204 is operable to be called programmatically from another program, such as using a single call to a procedure in program logic 204.

[020] Turning now to Figures 3-6 with continued reference to Figures 1-2, the stages for implementing one or more implementations of transactional memory application 200 are described in further detail. Figure 3 is a high level process flow diagram for transactional memory application 200. In one form, the process of Figure 3 is at least partially implemented in the operating logic of computing device 100. The process begins at start point 240 with providing a transactional memory system (e.g., a software transactional memory system) (stage 242). The system allows a user/developer to decorate a particular function with an attribute or other identifier to indicate the function is transaction-safe (stage 244). The compiler and/or linker create two versions of the particular function: one for use without transactions (e.g., the normal version) and one for use with transactions (e.g., the transactional version) (stage 246). In the normal version, the system stores a stub pointer points to the transactional version of the particular function (stage 248). In one implementation, the stub pointer(s) described herein are stored before the respective compiled function body text. In other implementations, the stub pointer can be stored in other locations, so long as it is associated with a respective version of the function. In the transactional version, the system stores a stub pointer that points to a runtime error routine (stage 249). The proper version of the function is called depending on execution context (stage 250). The process ends at end point 252.

[021] Figure 4 illustrates one implementation of the stages involved in using a code generator to generate code for transactional and non-transactional functions and to fill in the stub pointers as part of the compile/and or linking process. In one form, the process of Figure 4 is at least partially implemented in the operating logic of computing device 100. The process begins at start point 270 with the code generator allocating an extra pointer (e.g., the stub pointer) for each function at compile time (stage 272). For each function (i.e., while there are more functions) (decision point 274), various tasks are performed. For example, if the function is not marked as being a transaction-safe function (decision point 276), then the system generates the normal version (i.e., regular code) for the function, and fills in

the stub pointer of the normal version with the entry point of the runtime error routine (stage 286).

[022] If the function is marked as a transaction-safe function (decision point 276), the system creates code for the normal version F and its transactional version FT (stage 278). The system then allocates a stub pointer for the transactional version and the normal version of the function F (stage 280). In the stub for the normal version, the entry point of the transactional version is filled (stage 282). In the stub for the transactional version, the entry point of a runtime error routine is filled (stage 284). These stages are repeated appropriately for each function. Then, for each call site, the system makes the proper call depending on the context (stage 288). The call site decision process is described in further detail in Figure 5. The process ends at end point 290.

[023] Turning now to Figure 5, a diagram 300 for one implementation of the system of Figure 1 illustrates a possible decision tree for choosing the proper version of the function to call depending on execution context. In one form, the decision process of Figure 5 is at least partially implemented in the operating logic of computing device 100. For a direct call 306 in a non-transactional context 302, the entry point of the normal version is called directly for the function 312. For a direct call 306 in a transactional context 304, the entry point of the transactional version is called for the function 314. For a pointer call 308 in a non-transactional context 302, the call is the same as before 316. For a pointer call 308 in a transactional context 304, the address of the word pointing to the stub function is computed from the dereference of the function pointer value and then that address is used in the call 318. The stub is populated with a pointer to entry point of the transactional version and therefore the correct version is executed. A virtual or interface call 310 in a non-transactional context 302 is not affected, and is the same as before 320. A virtual or interface call 310 in the transactional context 304 performs the vtable lookup and then computes the stub function entry point of the transactional version of the function 322. The computed entry point is then used to make the call.

[024] Figure 6 is a logical diagram 400 for one implementation that illustrates two hypothetical functions, one with just a normal version and the other with a normal version and a transactional version. The “BAR” function 402 is a function that was not marked as transaction-safe in the source code. Thus, the entry pointer 412 to the function 402 points to the error routine 410. The “FOO” function 414 was marked as transaction-safe in the original source code, so a normally compiled version (the normal version) is provided 406, along with the transactional version 408. The entry point 414 of the normal version of FOO 406 points to the transactional version of FOO 408. The entry point 416 of the transactional version of FOO 408 points to the error routine 410. In one implementation, this error routine is embedded in the stubs (entry points) of these functions to allow errors to be caught at runtime when inappropriate use is encountered.

[025] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims. All equivalents, changes, and modifications that come within the spirit of the implementations as described herein and/or by the following claims are desired to be protected.

[026] For example, a person of ordinary skill in the computer software art will recognize that the client and/or server arrangements, user interface screen content, and/or data layouts as described in the examples discussed herein could be organized differently on one or more computers to include fewer or additional options or features than as portrayed in the examples.

What is claimed is:

1. A computer-readable medium having computer-executable instructions for causing a computer to perform steps comprising:

provide a transactional memory system (206);

5 create a normal version and a transactional version of each function of a software application that is marked as transaction-safe (210);

create a normal version of each function that is not marked as transaction-safe (212); and

10 for the normal version of each function that is marked as transaction-safe, point a stub pointer of the normal version to the transactional version (216).

2. The computer-readable medium of claim 1, wherein the stub pointer is allocated before a beginning of a code section of the normal version of each function (214).

15 3. The computer-readable medium of claim 1, wherein for each function that is not marked as transaction-safe, point the stub pointer of the normal version to an error routine (216).

4. The computer-readable medium of claim 1, wherein the transactional version of each function that is marked as transaction-safe has a stub pointer that points to an error routine (284).

20 5. The computer-readable medium of claim 1, wherein the normal version of each function that is not marked as transaction-safe has a stub pointer that points to an error routine (286).

6. A method for locating transactional code blocks in a transactional memory system comprising the steps of:

25 providing a transactional memory system (242);

allowing a user to decorate a particular function with an identifier to indicate that the particular function is transaction-safe (244);

creating two versions of the particular function, a normal function for use without transactions and a transactional function for use with transactions (246);

30 and

storing a stub pointer in the normal function that points to the transactional function (248).

7. The method of claim 6, further comprising:

calling a proper version of the particular function depending on execution
5 context (250).

8. The method of claim 7, wherein if the execution context is a non-transactional context with a direct call to the particular function, then the normal function is called (250).

9. The method of claim 7, wherein if the execution context is a transactional
10 context with a direct call to the particular function, then the transactional function is called (250).

10. The method of claim 7, wherein if the execution context is a non-transactional context with a pointer call to the particular function, then the pointer call operates as normal (316).

11. The method of claim 7, wherein if the execution context is a transactional
15 context with a pointer call to the particular function, then a stub size is subtracted from a de-reference of a function pointer value to calculate a resulting value, and then the resulting value is called (318).

12. The method of claim 7, wherein if the execution context is a non-
20 transactional context with a virtual call to the particular function, then the virtual call operates as normal (320).

13. The method of claim 7, wherein if the execution context is a transactional
context with a virtual call to the particular function, then a vtable lookup is performed and an entry point computed for the stub pointer of the transactional
25 function, and then the computed entry point is used to make the virtual call (322).

14. The method of claim 6, wherein the identifier is an attribute (244).

15. The method of claim 6, wherein a compiler creates the two versions of the particular function (246).

16. The method of claim 6, wherein the stub pointer that points to the
30 transactional function is stored at the beginning of the normal function (248).

17. A computer-readable medium having computer-executable instructions for causing a computer to perform the steps recited in claim 6 (200).

18. A method for creating transactional and non-transactional versions of functions for use in a transactional memory system comprising the steps of:

5 at compile time, allocating a stub pointer for each function of a plurality of functions in a software application (272);

 for each respective function of the plurality of functions that is not marked as a transaction-safe function, generating a normal version of the respective function that is not transaction-safe, and filling in the stub pointer with an entry
10 point of a runtime error routine (286); and

 for each respective function of the plurality of functions that is marked as a transaction-safe function, perform a process comprising the steps of:

 creating code for a normal version and a transactional version of the respective function that is transaction-safe (278);

15 in the stub pointer for the normal version, filling an entry point of the transactional version (282); and

 in a stub pointer for the transactional version, filling an entry point of a runtime error routine (284).

19. The method of claim 18, further comprising:

20 for each call site in each function in the software application, make proper function calls depending on a context (288).

20. A computer-readable medium having computer-executable instructions for causing a computer to perform the steps recited in claim 18 (200).

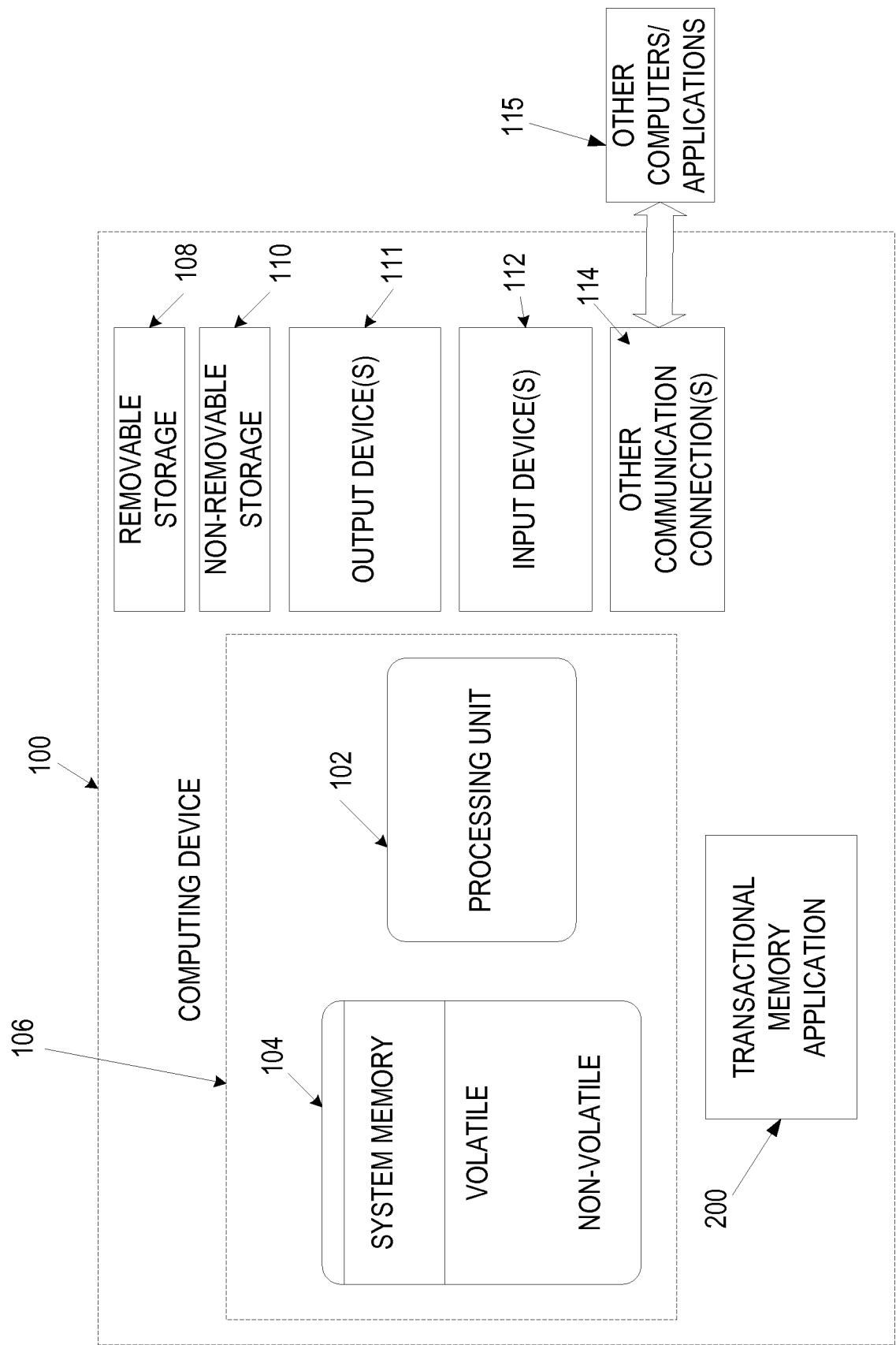


FIG. 1

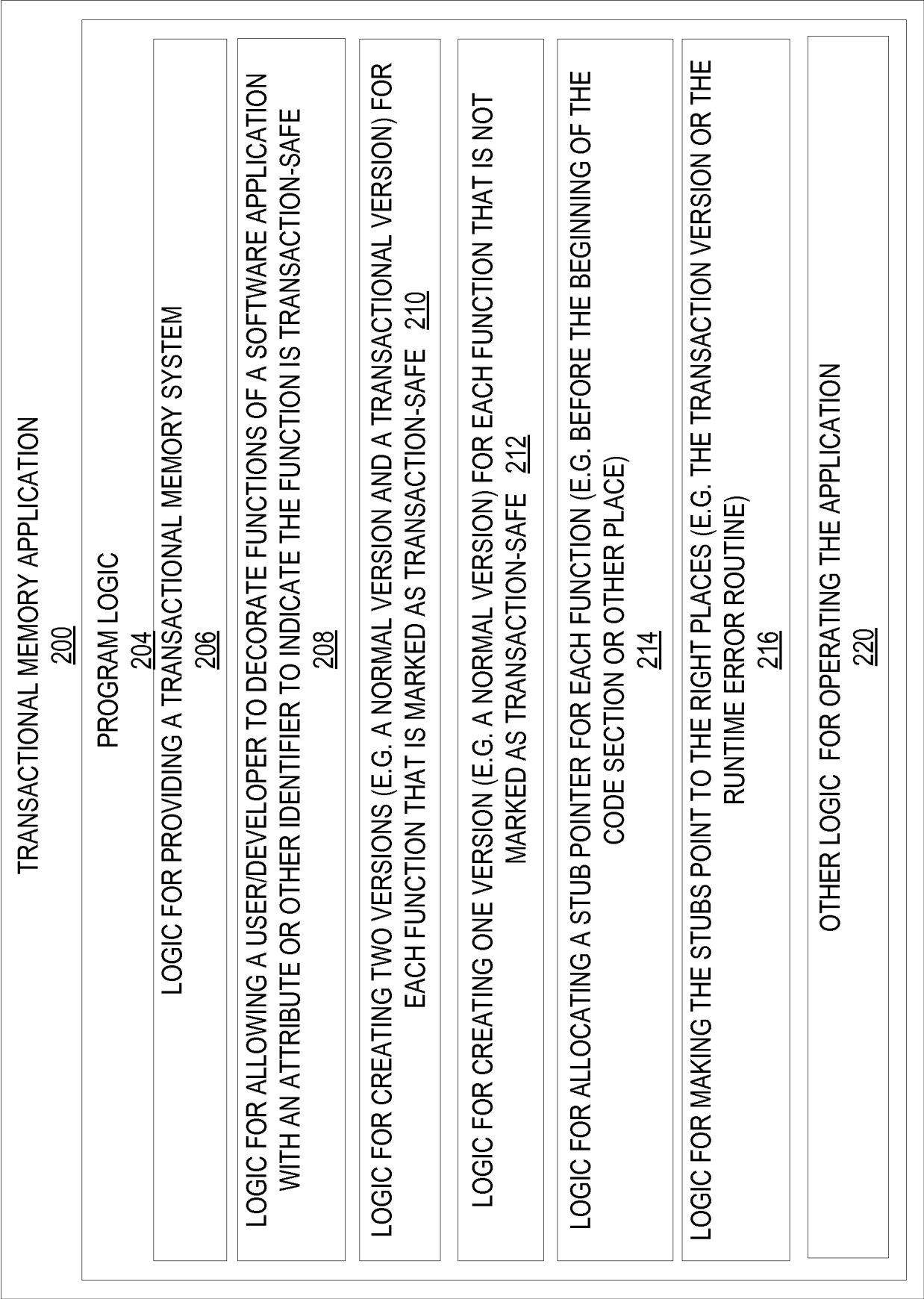


FIG. 2

3 / 6

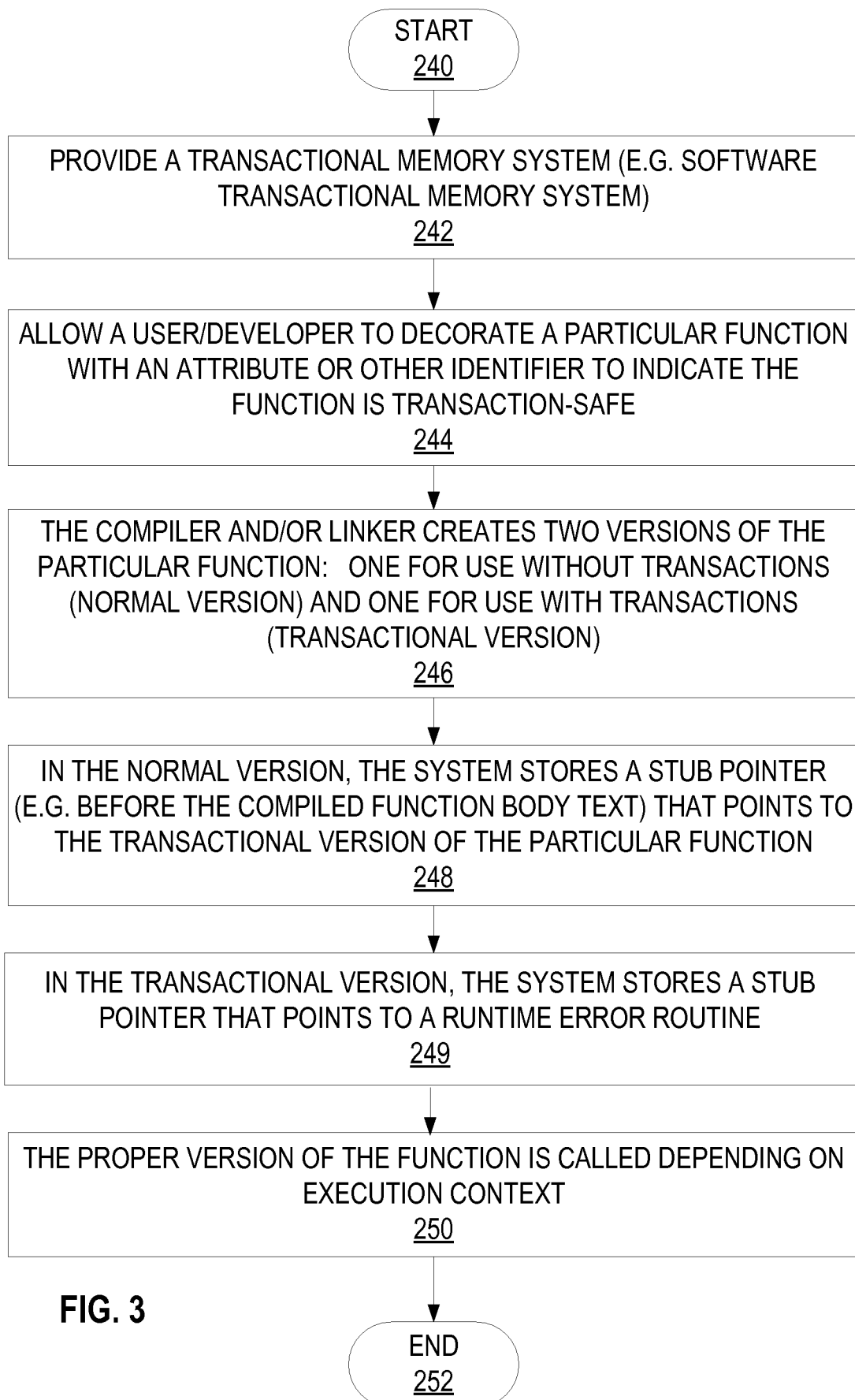
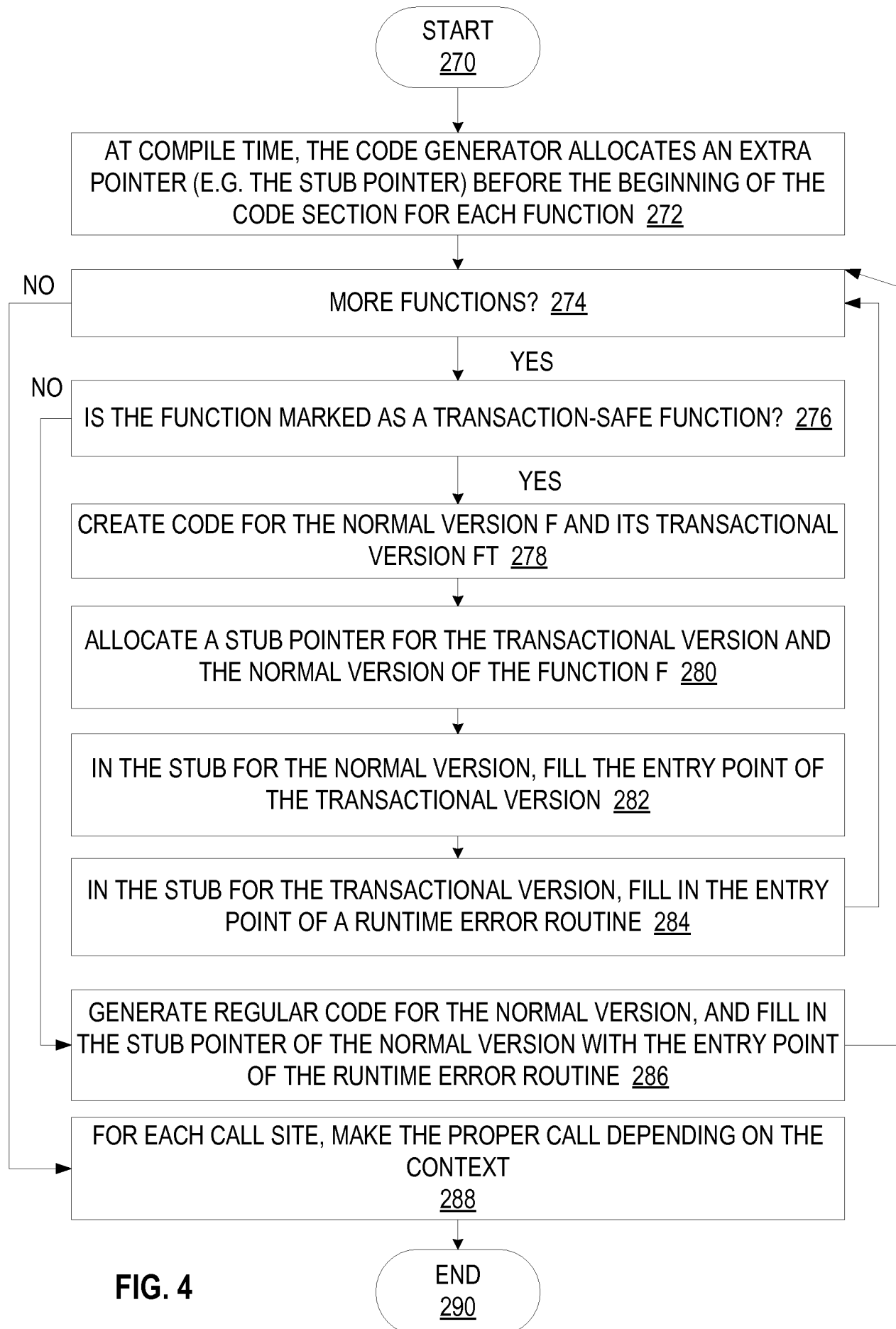


FIG. 3

4 / 6



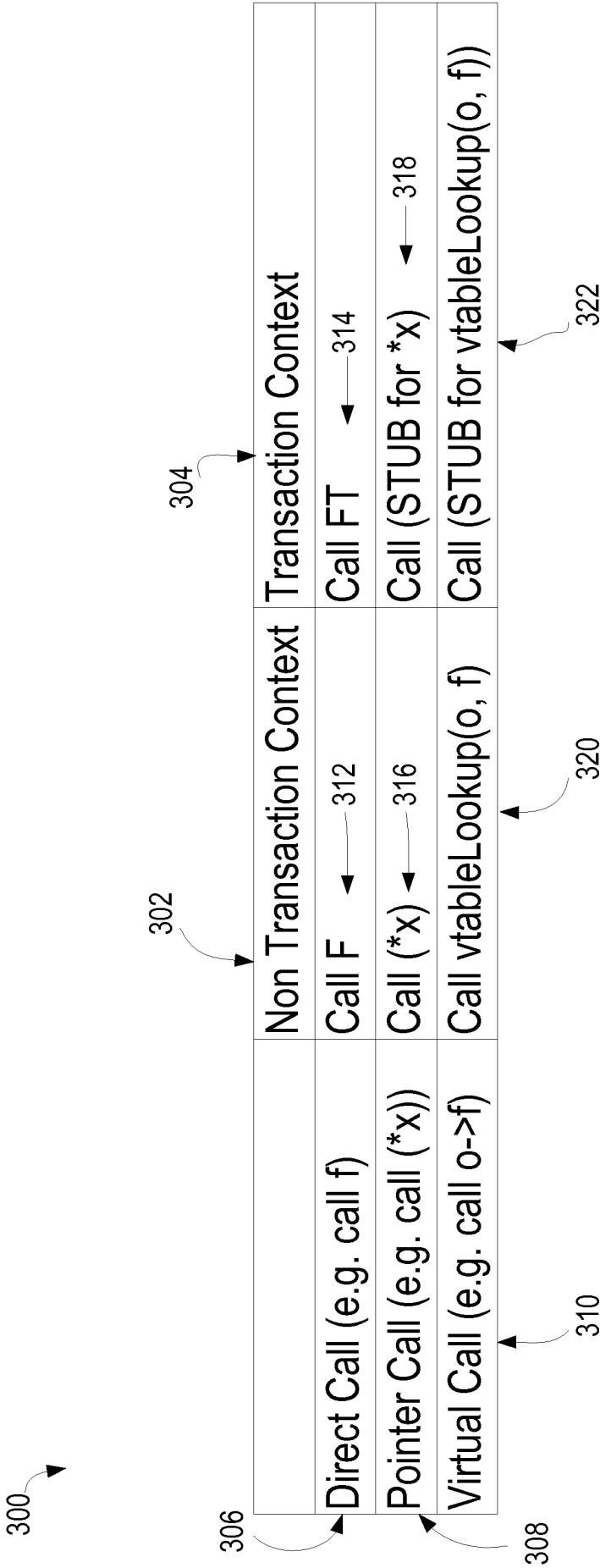


FIG. 5

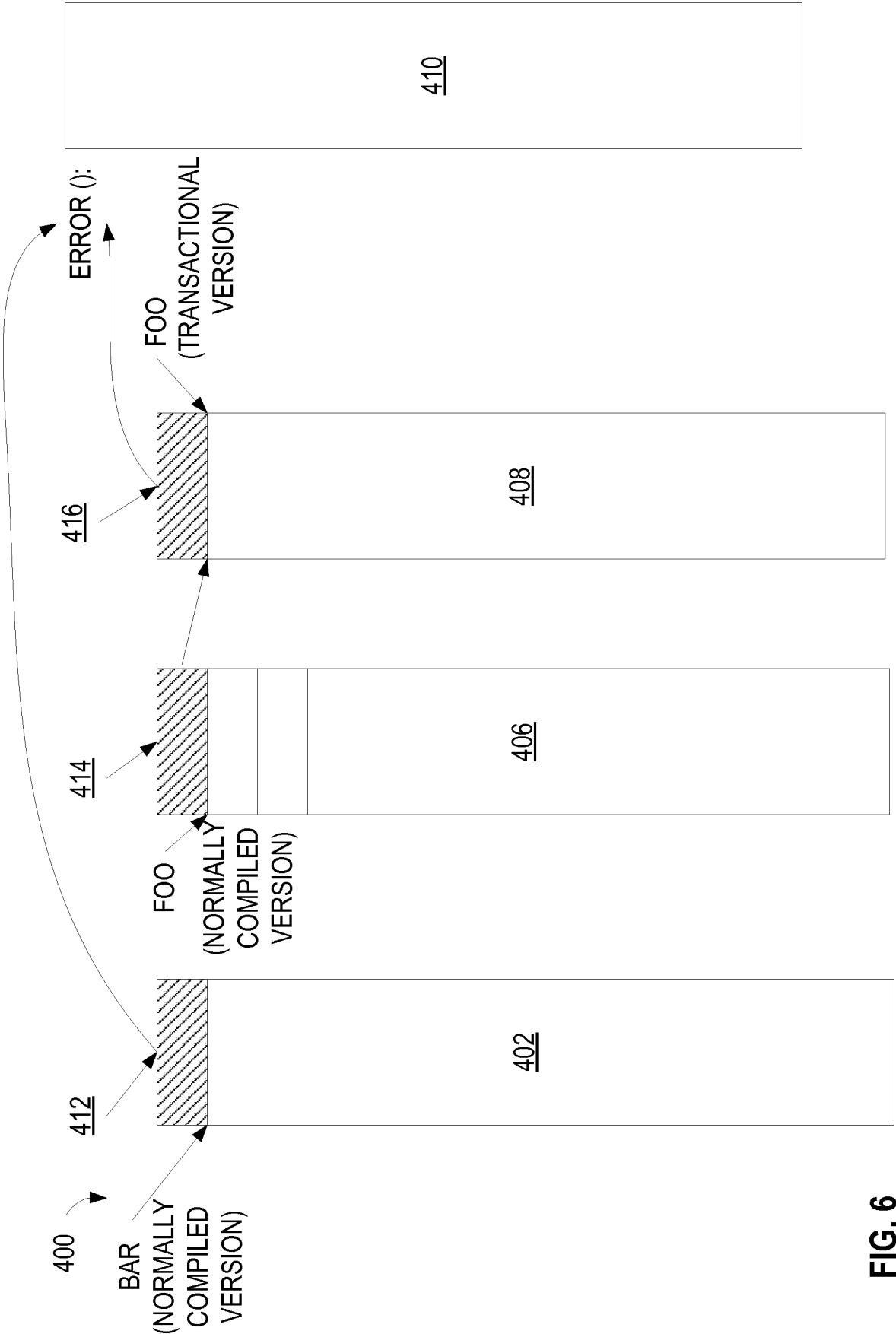


FIG. 6