November 2007

OSCI TLM2 USER MANUAL

Software version: TLM 2.0 Draft 2

Document version: DRAFT_JA4

DISTRIBUTION RESTRICTED TO OSCI MEMBERS

DISTRIBUTION RESTRICTED TO OSCI MEMBERS

Contributors

The TLM2.0-draft-2 standard was created under the leadership of the following individuals:

Trevor Wieman, Intel, TLM-WG Chair
Stuart Swan, Cadence, TLM-WG Vice-Chair

This document was authored by:

John Aynsley, Doulos

The following is a list of active technical participants in the OSCI TLM Working Group at the time of the release of TLM2.0-draft-2:

Tom Aernoudt, CoWare
James Aldis, OCP-IP
John Aynsley, Doulos
Guillaume Audeon, ARM
Bill Bunton, ESLX
Mark Burton, GreenSocs
Jack Donovan, ESLX
Othman Fathy, Mentor Graphics
Karthick Gururaj, NXP
Atsushi Kasuya, Jeda
Tim Kogel, CoWare
Laurent Maillet-Contoz, ST Microelectronics
Marcelo Montoreano, Synopsys
Rishiyur Nikhil, Bluespec
Victor Reyes, NXP
Adam Rose, Mentor Graphics
Olaf Scheufen, Synopsys
Alan Su, Springsoft
Stuart Swan, Cadence
Bart Vanthournout, CoWare
Yossi Veller, Mentor Graphics
Trevor Wieman, Intel
Charles Wilson, ESLX

The following people have also contributed to the OSCI TLM Working Group:

Mike Andrews, Mentor Graphics
Matt Ballance, Mentor Graphics
Geoff Barrett, Broadcom
Ryan Bedwell, Freescale
Bishnupriya Bhattacharya, Cadence
Bobby Bhattacharya, ARM
Axel Braun, University of Tuebingen
Herve Broquin, ST Microelectronics
Adam Erickson, Cadence
Alan Fitch, Doulos
Frank Ghenassia, ST Microelectronics
Mark Glasser, Mentor Graphics
Andrew Goodrich, Forte Design
Serge Goosens, CoWare
Thorsten Groetker, Synopsys
Kamal Hashmi, SpiraTech
Holger Keding, Synopsys
Devon Kehoe, Mentor Graphics
David Long, Doulos
Kiyoshi Makino, Mentor
Mike Meredith, Forte Design
David Pena, Cadence
Nizar Romdhane, ARM
Stefan Schmermbeck, Chipvision
Shiri Shem-Tov, Freescale
Jean-Philippe Strassen, ST Microelectronics
Tsutomu Takei, STARC
Jos Verhaegh, NXP
Maurizio Vitale, Philips Semiconductors
Vincent Viteau, Summit Design
Thomas Wilde, Infineon
Hiroyuki Yagi, STARC

Eugene Zhang, Jeda

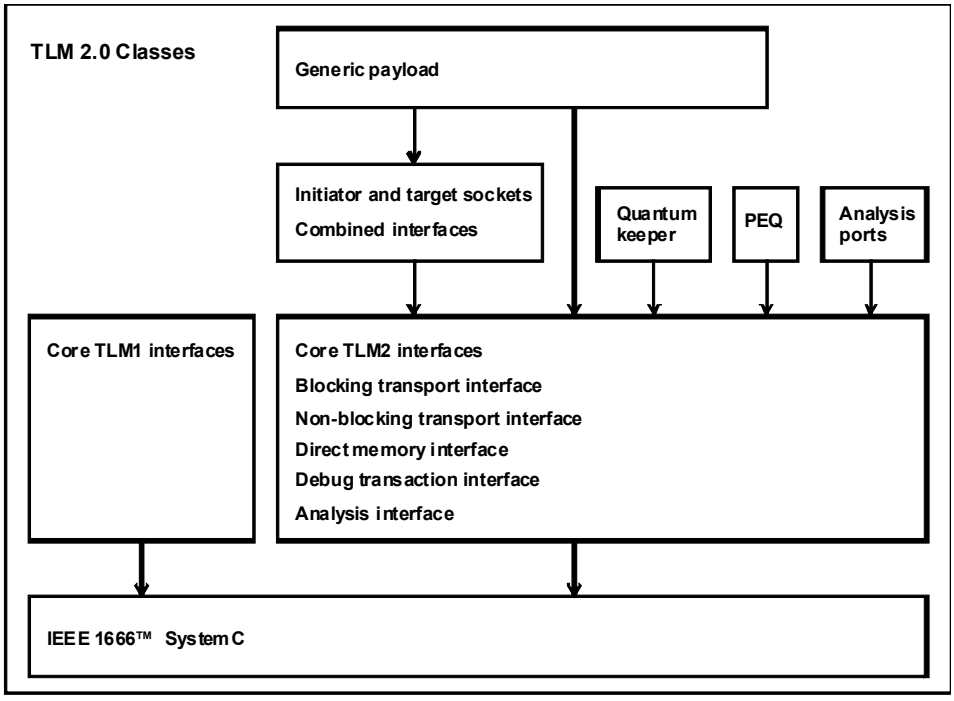
Contents

1. Overview

This document is the User Manual for the OSCI Transaction Level Modeling standard, version 2.0-draft-2. This version of the standard supersedes version 2.0-draft-1, and is not generally compatible with 2.0-draft-1. This version of the standard includes the core interfaces from TLM 1.0.

TLM2 consists of a set of core interfaces, analysis ports, initiator and target sockets, and the generic payload. The core interfaces include the core interfaces from TLM1 together with the blocking and non-blocking transport interfaces, the direct memory interface (DMI), the debug transaction interface, and the analysis interface. The core interfaces support untimed, loosely-timed and approximately-timed coding styles. The generic payload supports the abstract modeling of memory-mapped buses, together with an extension mechanism to support the modeling of specific bus protocols whilst maximizing interoperability. The quantum keeper and payload event queue (PEQ) are utility classes for use with the non-blocking transport interface.

The TLM2 classes are layered on top of the SystemC class library as shown in the diagram below. The core interfaces can be used directly without the other classes, or the generic payload can be used directly with the core interfaces without using the initiator and target sockets. For maximum interoperability, however, all of the classes should be used together.



1.1. Scope

This document describes the contents of the TLM2.0 standard. The main focus of this document is the key concepts and semantics of the TLM2 core interfaces and classes. It does not describe all the supporting code, examples, and

unit test. It lists the core TLM1 interfaces, but does not define their semantics. This document is not a definitive language reference manual. It is the intention that this document will be extended over time to add more practical guidelines on how to use TLM2.0[1]

[1] As it stands, this document does not include a description of the structure of the kit, unit tests, examples, or Doxygen documentation.

2. References

This standard shall be used in conjunction with the following publications:

ISO/IEC 14882:2003, Programming Languages—C++

IEEE Std 1666-2005, SystemC Language Reference Manual

Requirements Specification for TLM 2.0, Version 1.1, September 16, 2007

2.1. Bibliography

The following books provide useful background information:

Transaction-Level Modeling with SystemC, TLM Concepts and Applications for Embedded Systems, edited by Frank Ghenassia, published by Springer 2005, ISBN 10 0 387-26232-6(HB), ISBN 13 978-0-387-26232-1(HB)

Integrated System-Level Modeling of Network-on-Chip enabled Multi-Processor Platforms, by Tim Kogel, Rainer Leupers, and Heinrich Meyr, published by Springer 2006, ISBN 10 1-4020-4825-4(HB), ISBN 13 978-1-4020-4825-4(HB)

ESL Design and Verification, by Brian Bailey, Grant Martin and Andrew Piziali, published by Morgan Kaufmann/Elsevier 2007, ISBN 10 0 12 373551-3, ISBN 13 978 0 12 373551-5

3. Introduction

3.1. Background

The TLM1 standard defined a set of core interfaces for transporting transactions by value or const reference. This set of interfaces is being used successfully in some applications, but has three shortcomings with respect to the modeling of memory-mapped busses and other on-chip communication networks:

a) TLM1 has no standard transaction class, so each application has to create its own non-standard classes, resulting in very poor interoperability between models from different sources. TLM2 addresses this shortcoming with the generic payload.

b) TLM1 has no support for timing annotation, so no standard way of communicating timing information between models. TLM1 models would typically implement delays by calling **wait**, which slows down simulation. TLM2 addresses this shortcoming with the non-blocking transport interface.

c) The TLM1 interfaces require all transaction objects and data to be passed by value or const reference, which slows down simulation. Some applications work around this restriction by embedded pointers in transaction objects, but this is non-standard and non-interoperable. TLM2 addresses this shortcoming with transaction objects whose lifetime extends across several transport calls.

3.2. Transaction-level modeling, use cases and abstraction

There has been a longstanding discussion in the ESL community concerning what is the most appropriate taxonomy of abstraction levels for transaction level modeling. Models have been categorized according to a range of criteria, including granularity of time, frequency of model evaluation, functional abstraction, communication abstraction, and use cases. The TLM2 activity explicitly recognizes the existence of a variety of use cases for transaction-level modeling (see the Requirements Specification for TLM2.0), but rather than defining an abstraction level around each use case, TLM2 takes the approach of distinguishing between interfaces (APIs) on the one hand, and coding styles on the other. The TLM2 standard defines a set of interfaces which should be thought of as low-level programming mechanisms for implementing transaction-level models, then describes a number of coding styles that are appropriate for, but not locked to, the various use cases. Each coding style can support a range of abstraction across functionality, timing and communication.

An untimed functional model consisting of a single software thread can be written as a C function or as a single SystemC process, and is sometimes termed an *algorithmic* model. Such a model is not *transaction-level* per se, because by definition a transaction is an abstraction of communication, and a single-threaded model has no inter-process communication. A transaction-level model requires multiple SystemC processes to simulate concurrent execution and communication.

An abstract transaction-level model containing multiple processes (multiple software threads) requires some mechanism by which those threads can yield control to one another. This is because SystemC uses a co-operative multitasking model where an executing process cannot be preempted by any other process. Process scheduling can be modeled in one of two ways, either by having the user introduce explicit communication and synchronization points into their application, or by having a coding style or underlying modeling framework that forces each process to yield control at certain points. The former style is typified by the CSP (Communicating Sequential Processes) and

KPN (Kahn Process Network) formalisms, and is easily implemented in SystemC using FIFOs, semaphores, or other synchronization primitives. This allows a completely untimed modeling style where in principle simulation can run without advancing simulation time, and all synchronization points are determined in advance and explicitly coded. The latter style is suitable when modeling multiple software threads running in an environment where the threads are not running in lockstep, but communicate from time-to-time, possibly at points in time that cannot be determined completely in advance. In this standard, such a coding style is called *loosely-timed*.

3.3. Coding styles

3.3.1. Untimed coding style

In TLM2, the untimed coding style makes use of the blocking transport interface. This interface does not make any provision for annotating timing information, but since the **transport** method is blocking, it could in principle call **wait** to model a latency or to introduce a random delay into the process scheduling. The application is expected to implement synchronization between processes by calling explicit synchronization statements. For example, a blocking transport call could **wait** on an event notified by a separate process, or could return a flag to instruct the caller to **wait**, giving other processes a chance to execute.

3.3.2. Loosely-timed coding style and temporal decoupling

The loosely-timed coding style makes use of the non-blocking transport interface. This interface provides for timing annotation and for multiple phases and timing points during the lifetime of a transaction. A loosely-timed transactions has exactly two timing points marking the beginning of the request and the beginning of the response (or, equivalently, the end of the transaction).

The loosely-timed coding style is appropriate for the use case of software development in an MPSoC environment, where multiple software threads are running on a virtual model of a hardware platform, possibly under the control of an operating system. The loosely-timed coding styles supports the modeling of timers and course-grained process scheduling, sufficient to model the booting and running of an operating system. It also supports temporal decoupling, where each software thread is permitted to run ahead in a local “time warp” until it reaches the point when it needs to synchronize with the rest of the system. Temporal decoupling can result in very fast simulation for certain systems.

The SystemC scheduler keeps a tight hold on simulation time. The scheduler advances simulation time to the time of the next event, then runs any processes due to run at that time or sensitive to that event. SystemC processes only run at the current simulation time (as obtained by calling the method **sc_time_stamp**), and whenever a SystemC process reads or writes a variable, it accesses the state of the variable as it would be at the current simulation time.

When a process finishes running it must pass control back to the simulation kernel. If the simulation model is written at a fine-grained level, then the overhead of event scheduling and process context switching becomes the dominant factor in simulation speed. One way to speed up simulation would be to allow processes to run ahead of the current simulation time, or to *warp* time.

In general, a process can be allowed to run ahead of simulation time until it encounters a dependency on a variable updated by another process, or needs to interact with another process. At that point, there is a decision to be made. In terms of general simulation techniques, outside the context of SystemC, there are several options. The process can be allowed to continue running, but the scheduler may need to backtrack later if it is discovered that the process assumed the wrong value. The scheduler may launch several alternative runs of the same process in parallel,

assuming a different value for the external variable in each run, then pick the correct run after-the-fact when the value becomes known. This may make sense with dedicated hardware support. A third option is to return control to the simulation kernel when an external dependency is encountered, only resuming the process later in simulation time when the external value becomes known.

In TLM2, running ahead of simulation time is called *temporal decoupling*. Temporal decoupling does *not* permit backtracking, and each individual process is responsible for determining whether it can run ahead of simulation time without breaking the functionality of the model. When a process encounters an external dependency it has two choices: either force synchronization, which means yielding to allow all other processes to run as normal until simulation time catches up, or accept the current value and continue. The synchronization option guarantees functional congruency with the standard SystemC simulation semantics. Continuing with the current value relies on making a very significant assumption concerning communication and timing in the modeled system. It assumes that no damage will be done by sampling the value too early, and that any subsequent change to the value will be picked up in a subsequent process execution. This assumption is valid either if there is some explicit handshaking associated with the value, or if the value only changes relatively infrequently.

Temporal decoupling, and in particular allowing the time of process interactions to slip in time, is characteristic of the loosely-timed coding style.

If a process were permitted to run ahead of simulation time with no limit, the SystemC scheduler would be unable to operate and other processes would never get the chance to execute. This is avoided by the TLM2 quantum keeper, which imposes an upper limit on the time a process is allowed to run ahead, known as the *quantum*. The quantum is set by the application, and the quantum value represents a tradeoff between simulation speed and accuracy. Too small a quantum forces processes to yield and synchronize very frequently, slowing down simulation. Too large a quantum introduces large timing inaccuracies, possibly to the point where important events are missed and the model ceases to function.

For example, consider the simulation of a system consisting of a processor, a memory, a timer, and some slow external peripherals. The software running on the processor spends most of its time fetching and executing instructions from system memory, and only interacts with the rest of the system when it is interrupted by the timer, say every 1ms. The ISS that models the processor could be permitted to run ahead with a quantum of up to 1ms, making direct accesses to the memory model, but only synchronizing with the peripheral models at the rate of timer interrupts. Depending on the detail of the models, this could give a simulation speed improvement of up to 1000X.

In TLM2, temporal decoupling is supported by the **tlm_quantum_keeper** class and the timing annotation of the non-blocking transport interface.

3.3.3. Untimed versus loosely-timed modeling

Explicit temporal decoupling is not required in an untimed model, because an untimed model does not rely on the notion of simulation time for process synchronization and scheduling. An untimed initiator can implicitly run ahead of simulation time until it reaches the next explicit synchronization point. An untimed model relies on the presence of explicit synchronization points in the system model in order to pass control between initiators at predetermined points during the execution.

A loosely-timed model can also benefit from explicit synchronization in order to guarantee deterministic execution, but a loosely-timed model is able to make progress even in the complete absence of explicit synchronization points, because each initiator will only run ahead as far as the end of the time quantum before yielding control. A loosely-timed model can increase its timing accuracy by using synchronization-on-demand, that is, yielding control to the

scheduler before reaching the end of the time quantum. Synchronization-on-demand in a loosely-timed model is equivalent to explicit synchronization in an untimed model.

The non-blocking transport interface can support both the untimed and the loosely-timed coding styles, as well as the approximately-timed coding style. However, the blocking transport interface is simpler to use if neither timing annotation nor temporal decoupling are required.

3.3.4. Approximately-timed coding style

The non-blocking transport interface also supports the approximately-timed coding style, which is appropriate for the use cases of architectural analysis and performance analysis. For approximately-timed modeling, a transaction is broken down into multiple phases, with an explicit timing point marking the transition between phases. Because the loosely-timed and approximately-timed styles both use the same non-blocking transport interface, a degree of interoperability between them is possible.

In the case of the generic payload there are exactly four timing points marking the beginning and the end of the request and the beginning and the end of the response. Specific protocols may need to add further timing points, which may possibly cause the loss of direct compatibiliy with the generic payload.

The approximatedly-timed coding style does not use temporal decoupling, but rather each process executes in lock step with the SystemC scheduler. Process interactions are annotated with specific delays. To create an approximately-timed model, it is generally sufficient to annotate two kinds of delay: the latency of the target, and the initiation interval or accept delay of the target. The annotated delays are implemented by making calls to the SystemC scheduler, that is, **wait**(delay) or **notify**(delay).

3.3.5. Characterization of untimed, loosely-timed and approximately-timed coding styles

The three coding styles can be characterized in terms of how running processes use simulation time and when they yield to the SystemC scheduler.

Untimed. The notion of simulation time is not required, and each process runs up to the next explicit synchronization point before yielding.

Loosely-timed. Simulation time is used, but processes are temporally decoupled from simulation time. Each process keeps a tally of the time it consumes, and may yield because it reaches an explicit synchronization point or because it has consumed its time quantum.

Approximately-timed. Processes run in lock-step with SystemC simulation time. Delays annotated onto process interactions are implemented using timeouts (wait) or timed event notifications.

3.3.6. Switching between loosely-timed and approximately-timed modeling[2]

(This section is to be completed)

3.3.7. Cycle-accurate modeling

Cycle-accurate modeling is beyond the scope of TLM2 at present, but the TLM1 interfaces provide sufficient machinery for cycle-accurate modeling.

3.3.8. Blocking versus non-blocking transport interfaces

The blocking and non-blocking transport interfaces are separate interfaces that exist in TLM2 to support distinct coding styles. Applications should choose an interface depending on the coding style used (untimed, loosely-timed or approximately-timed), which in turn will depend on the use case. It is possible to adapt between the two interfaces or even to combine both interfaces in a single socket; the TLM2 Kit includes examples of doing so. However, mixing interface types and mixing coding styles does incur some cost, both in terms of simulation speed and coding complexity, so should be avoided where possible.

The blocking transport interface is particularly appropriate for models that include explicit, strong synchronization between processes such that the sequence of synchronization points is deterministic and independent of implementation detail. It is also appropriate where an untimed transaction-level model is to be transformed directly to an RTL implementation without any intermediate timed SystemC model.

The non-blocking transport interface is particulary appropriate for models that include timing information, or that are to be refined from a loosely-timed to an approximately-timed model, or where the exact sequence of synchronization points between concurrent processes is indeterminate or dependent upon hardware implementation detail or an operating system scheduler.

The strength of the blocking transport interface is that it allows a simplified coding style for models that do not require specific timing detail. The strength of the non-blocking transport interface is that it supports both temporal decoupling in a loosely-timed coding style and timing annotation for approximately-timed modeling.

3.3.9. Use cases and coding styles

The table below summarizes the mapping between use cases for transaction-level modeling and coding styles:

Use Case	**Coding style**[3]
Software application development	Untimed or loosely-timed
Software performance analysis	Loosely-timed
Hardware architectural analysis	Loosely-timed or approximately-timed
Hardware performance verification	Cycle-accurate
Hardware functional verification	Untimed (verification environment), loosely-timed or approximately-timed

[2] I need some help with this

[3] To be refined. Suggestions please!

3.4. Initiators, targets, and sockets

The TLM2 transport interfaces pass transactions between initiators and targets. An initiator is a module that can initiate transactions, that is, create new transaction objects and pass them on by calling a method of one of the core interfaces. A target is a module that acts as the final destination for a transaction. In the case of a write transaction, an initiator (such as a processor) writes data to a target (such as a memory). In the case of a read transaction, an initiator reads data from a target. An interconnect component is a module that accesses a transaction but does act as an initiator or a target for that transaction, typical examples being arbiters and routers.

In order to illustrate the idea, this paragraph will describe the lifetime of a typical transaction object. The transaction object is created by an initiator and passed as an argument of a method of the transport interface (blocking or non-blocking). That method is implemented by an interconnect component such as an arbiter, which may read attributes of the transaction object before passing it on to a further transport call. That second transport method is implemented by a second interconnect component, such as a router, which in turn passes on the transaction through a third transport call to a target such as a memory, the final destination for the transaction object. (The actual number of interconnect components will vary from transaction to transaction. There may be none.) This sequence of method calls is known as the *forward path*. The transaction is executed in the target, and the transaction object may be returned to the initiator in one of two ways, either carried with the return from the transport method calls as they unwind, or passed by making explicit transport method calls in the opposite direction from target back to initiator, known as the *backward path*. This choice is determined by the return value from the **nb_transport** method.

The forward path is the calling path by which an initiator or interconnect component makes interface method calls forward in the direction of another interconnect component or the target. The backward path is the calling path by which a target or interconnect component makes interface method calls back in the direction of another interconnect component or the initiator.

In order to support both forward and backward paths, each connection between components requires a port and an export, both of which have to be bound. This is facilitated by the *initiator socket* and the *target socket*. An initiator socket contains a port for interface method calls on the forward path and an export for interface method calls on the backward path. A target socket provides the opposite. The initiator and target socket classes overload the SystemC port binding operator to implicitly bind both forward and backward paths.

As well as the transport interfaces, the sockets also encapsulate the DMI and debug transaction interfaces (see below).

When using sockets, an initiator component will have at least one initiator socket, a target component at least one target socket, and an interconnect component at least one of each. A component may have several sockets transporting different transaction types, in which case a single component may act as initiator or target for multiple independent transactions.

It is also possible for components to use SystemC ports and exports directly with the core TLM2 interfaces, without using sockets. The TLM2 kit provides some examples of doing this[4]. Sockets are provided for convenience, and are recommended for maximal interoperability and a consistent coding style.

[4] This refers to the tlm_initiator_port and tlm_target_port convenience ports.

3.5. DMI and debug transaction interfaces

The direct memory interface (DMI) and debug transaction interface are specialized interfaces distinct from the transport interface, providing direct access and debug access to an area of memory owned by a target. The DMI and debug transaction interfaces each bypass the usual path through the interconnect components used by the transport interface. DMI is intended to accelerate regular memory transactions in an untimed or loosely-timed simulation, whereas the debug transaction interface is for debug access free of the delays or side-effects associated with regular transactions.

The transport interfaces are combined with the DMI and debug transaction interfaces in the standard initiator and target sockets. All three interfaces, transport, DMI, and debug, can be used in parallel to access a target. The DMI has both forward (initiator-to-target) and backward (target-to-initiator) interfaces, whereas debug only has a forward interface.

4. Core TLM2 interfaces

In addition to the core interfaces from TLM1, TLM2 adds blocking and non-blocking transport interfaces, a direct memory interface (DMI), and a debug transaction interface.

4.1. Blocking transport interface

4.1.1. Introduction

The new TLM2 blocking transport interface supports an untimed modeling style. This interface has deliberate similarities with the transport interface from TLM1, which is still part of the TLM2 standard, but the TLM1 transport interface and the TLM2 blocking transport interface are not identical. In particular, the new **b_transport** method takes a single argument passed by non-const reference, whereas the TLM1 **transport** method takes a request as a single const reference argument and returns a response by value. TLM1 assumes separate request and response objects passed by value (or const reference), whereas TLM2 assumes a single transaction object passed by reference, whether using the blocking or the non-blocking TLM2 interfaces.

4.1.2. Migration path from TLM1

The old TLM1 and the new TLM2 interfaces are both part of the TLM2 standard. The TLM1 blocking and non-blocking interfaces are still useful in their own right. For example, a number of vendors have used these interfaces in building functional verification environments for HDL designs.

The intent is that the similarity between the old and new blocking transport interfaces should ease the task of building adapters between legacy models using the TLM1 interfaces and the new TLM2 interfaces.

4.1.3. Class definition

```
namespace tlm {

template <typename TRANS = tlm_generic_payload>
class tlm_blocking_transport_if : public virtual sc_core::sc_interface {
public:
  virtual void b_transport(TRANS& trans) = 0;
};

} // namespace tlm
```

4.1.4. The TRANS template argument

The intent is that this core interface may be used to transport transactions of any type. A specific transaction type, **tlm_generic_payload**, is provided to ease interoperability between models where the precise details of the transaction attributes are less important.

For maximum interoperability, applications should use the default transaction type **tlm_generic_payload**. In order to model specific protocols, applications may substitute their own transaction type. Sockets that use interfaces specialized with different transaction types cannot be bound together, providing compile-time checking but restricting interoperability.

4.1.5. Rules

a) The caller is responsible for allocating storage and for constructing the transaction object. The **b_transport** method shall be called with a fully initialized transaction object.

b) The call to **b_transport** shall mark the first timing point of the transaction.

c) The callee may modify or update the transaction object, subject to any constraints imposed by the transaction class **TRANS**.

d) The return from **b_transport** shall mark the final timing point of the transaction.

e) The caller is responsible for deleting or pooling the transaction object after the call.

f) The callee shall assume that the caller will invalidate the transaction object upon return from **b_transport**.

g) The **b_transport** method does not support timing annotation or temporal decoupling.

h) It is recommended that the transaction object should not contain timing information. Timing should be annotated using the non-blocking transport interface.

4.2. Non-blocking transport interfaces

4.2.1. Introduction

The new non-blocking transport interface exists to support the loosely-timed and approximately-timed coding styles. This interface is the central innovation of the TLM2 standard. It uses a similar argument-passing mechanism to the new blocking transport interface in that the **nb_transport** method passes a non-const reference to the transaction object, but there the similarity ends. The **nb_transport** method also passes a phase to indicate the state of the transaction, and a time to annotate the delay of the next phase transition. Together, these arguments support multiple phases within the lifetime of a single transaction object, timing annotation, and temporal decoupling for fast simulation.

The non-blocking transport interface also supports the untimed coding style, but the blocking transport interface may be preferred for its simplicity.

The non-blocking transport interface and the generic payload were designed to be used together for the fast, abstract modeling of memory-mapped buses. However, the non-blocking transport interface can be used separately from the generic payload to model specific protocols. Both the transaction type and the phase type are template parameters.

4.2.2. Class definition

```
namespace tlm {

enum tlm_phase { BEGIN_REQ, END_REQ, BEGIN_RESP, END_RESP };
```



```
enum tlm_sync_enum { TLM_REJECTED = 0, TLM_SYNC = 1, TLM_SYNC_CONTINUE = 2,
TLM_COMPLETED = 3 };

template <typename TRANS = tlm_generic_payload, typename PHASE = tlm_phase>
class tlm_nonblocking_transport_if : public virtual sc_core::sc_interface {
public:
  virtual tlm_sync_enum nb_transport(TRANS& trans, PHASE& phase, sc_core::sc_time& t) = 0;
};

} // namespace tlm
```

4.2.3. The TRANS template argument

The intent is that this core interface may be used to transport transactions of any type. A specific transaction type, **tlm_generic_payload**, is provided to ease interoperability between models where the precise details of the transaction attributes are less important.

For maximum interoperability, applications should use the default transaction type **tlm_generic_payload**. In order to model specific protocols, applications may substitute their own transaction type. Sockets that use interfaces specialized with different transaction types cannot be bound together, providing compile-time checking but restricting interoperability.

4.2.4. The PHASE template argument

The intent is that this core interface may be used for transactions with any number of phases and timing points. A specific type **tlm_phase** is provided for use with the generic payload.

For maximum interoperability, applications should use the default arguments **tlm_phase** and **tlm_generic_payload** with the loosely-timed or approximately-timed coding styles. Applications are free to use the type **tlm_phase** with other transaction types, or to use the generic payload with other phase types, but doing either will restrict interoperability.

4.2.5. The nb_transport call

a) The **nb_transport** method shall not call **wait**, directly or indirectly.

b) The **nb_transport** method may be called from a thread process or from a method process.

c) Exceptionally, if both the blocking and non-blocking transport interfaces are being used together, **nb_transport** may need to call **b_transport**, typically in an adaptor. This is only possible if the initiator is a thread process. Because a blocking method may call **wait**, it is critical that the initiator should synchronize before the call to **b_transport**.[5]

[5] This statements requires clarification. Is nb_transport non-blocking or not? Olaf says non-blocking – nb_transport can only call b_transport if it is guaranteed that b_transport is non-blocking. Otherwise the adaptor would need to call the b_transport from an independent thread. We'll still need to defined the exact return value semantics for these cases.

4.2.6. The trans argument

a) The initiator is responsible for allocating storage and constructing the transaction object passed as the first argument. The **nb_transport** method shall be called with a fully initialized transaction object.

b) The lifetime of a given transaction object may extend beyond the return from **nb_transport** such that a series of calls to **nb_transport** may pass a single transaction object forward and backward between initiators and targets.

c) The caller of **nb_transport** need not be the initiator. In particular, the target may call **nb_transport** to signal a timing point back to the initiator.

d) Since the lifetime of the transaction object may extend over several calls to **nb_transport**, either the caller or the callee may modify or update the transaction object, subject to any constraints imposed by the transaction class **TRANS**. For example, for the generic payload, the target may update the data array of the transaction object in the case of a read command, but shall not update the command field. See clause 9.7 Default values and modifiability of attributes

e) The initiator is responsible for deleting or pooling the transaction object after the final timing point. The initiator, and only the initiator, may delete the transaction object.

4.2.7. The phase argument

a) Each call to **nb_transport** shall mark a timing point of the given transaction.

b) The attributes of a transaction are notionally stable during each phase, only changing at the timing points that mark phase transitions. Any change to the transaction object occurring in the middle of a phase should only become visible to other components at the next timing point.

c) The phase argument is passed by reference. Either caller or callee may modify the phase.

d) Any change to the state of the transaction should be accompanied by a change to the phase argument such that either caller or callee can detect the change by comparing the value of the phase argument from one call to the next.

e) The value of the phase argument represents the current state of the protocol state machine for the communication between caller and callee. Where a single transaction object is passed between more than two components (initiator, interconnect, target), each caller/callee connection requires (notionally, at least) a separate protocol state machine.

f) Whereas the transaction object has a lifetime and a scope that may extend beyond any single call to **nb_transport**, the phase is normally local to the caller. Each **nb_transport** call for a given transaction may have a separate phase object.

g) The enum **tlm_phase** is specific to the generic payload. Other protocols may use this same phase type or may substitute their own phase type (with a corresponding loss of interoperability).

4.2.8. The sc_time argument

a) It is recommended that the transaction object should not contain timing information. Timing should be annotated using the **sc_time** argument to **nb_transport**.

b) The time argument shall be non-negative.

c) When using the loosely-timed coding style, the caller may pass a positive value for the time argument to indicate to the callee that it should behave as if the transaction were received at time **sc_time_stamp() + t**. This is the local time with respect to temporal decoupling, in other words, an offset from the start time of the current time quantum.

d) For a general description of temporal decoupling, see clause 3.3.2 Loosely-timed coding style and temporal decoupling

e) For a description of the quantum, see clause 7.3 Rules for processes using temporal decoupling

f) If the callee is not able to determine how to process the transaction on the basis of predicting any necessary state information as it will be at a future point in time (**sc_time_stamp() + t**), the callee should return a value of TLM_SYNC. It is still the responsibility of the callee to behave as if the transaction were received at the given future time, and in order to do this, the callee would typically create a timed notification that would cause a process to be resumed at the given future time. As an example, the TLM2 kit includes a payload event queue class **tlm_peq** which may be used for this purpose. See clause 8 Payload event queue

g) With a positive value for the time argument, the callee is being invited to become temporally decoupled from the simulation time, or to “live in a time warp”. If the target is a simple slave that only serves one master, that may be acceptable. On the other hand, if the target has dependencies on other asynchronous events, the target may have to wait for simulation time to advance (synchronization-on-demand) before it can predict the future state of the transaction with certainty.

h) Passing a positive value for the time argument is normally associated with the loosely-timed coding style, but is still technically possible in an approximately-timed model. When using the approximately-timed coding style, the only reasonable behavior for the callee would be to create a timed notification with the given delay (perhaps using the payload event queue), and return a value of TLM_SYNC.

i) If **nb_transport** is called consecutively at simulation times and with values for the time argument such that the second call is effectively to be processed before the first call, in other words the timing points would occur in reverse order when taking into account the time of the call and the time argument, the callee has two options. Either put the transaction into a payload event queue (or similar) to delay the arrival of the transaction until the correct time (synchronization-on-demand), or process the transactions in the order of the **nb_transport** calls and assume that the system design can tolerate out-of-order execution (because of the existence of some explicit mechanism in the system to enforce the correct causal chain of events). The former option is characteristic of the approximately-timed coding style, and the latter option of the loosely-timed coding style.

j) On return from **nb_transport**, it is the responsibility of the caller to behave as if it had received notification that the transaction will change state at time **sc_time_stamp() + t**, where **t** is the **sc_time** argument to **nb_transport**. In other words, the time argument is used to annotate a latency to the **nb_transport** call, and it is the caller’s responsibility to realize that latency.

k) On return from **nb_transport**, the caller has three options for implementing an annotated latency. It can run in temporally decoupled mode, it can put the transaction into a payload event queue (or similar), or it can call **wait(t)** (assuming the caller is a thread process).

l) This non-blocking transport interface is explicitly intended to support pipelined transactions. For example, several successive calls from the same caller at the same time to **nb_transport** could each create notifications (these could be, but are not necessarily, SystemC event notifications) of transaction timing points at distinct future times. It is the responsibility of the caller and callee to keep track of these actual or effective notifications.

m) The callee may increase the value of the time argument, but shall not decrease the value. This rule is consistent with time not running backward in a SystemC simulation.

4.2.9. The tlm_sync_enum return value

The meaning of the return value is fixed, and does not vary according to the transaction type or phase type.

TLM_REJECTED. The callee has rejected the transaction. The callee shall not have modified the state of the transaction object, the phase, or the time argument during the call. The caller should yield (suspend in the case of a thread process or return to the kernel in the case of a method process) before attempting to call **nb_transport** again for the same transaction object. This is similar to a return value of false from the methods such as **nb_put** of the TLM1 non-blocking interface, and can be used to code a polling-style interface.

TLM_SYNC. The callee has accepted the transaction. The callee shall not have modified the state of the transaction object, the phase, or the time argument during the call. The caller should ignore the values of the arguments following the call, since they shall not have changed. The callee would typically register the fact that the transaction has been accepted by changing some internal state or by making a further call to **nb_transport**. The component containing the caller should expect a response from the component containing the callee through the backward path between them, and should take no further action with respect to this transaction until it receives that response. Following the call to **nb_transport**, the caller should yield control to the SystemC scheduler in order that other processes may run and initiate a response. One way to do this is to call the **sync** method of class **tlm_quantum_keeper**. This is known as synchronization-on-demand.

TLM_SYNC_CONTINUE. The callee has accepted the transaction. The callee shall have modified the state of the transaction object and the phase argument during the call. In other words, the callee shall have advanced the state of the protocol state machine associated with the transaction. The callee may have increased the value of the time argument. Following the call to **nb_transport**, the caller should inspect the phase argument and transaction object and take the appropriate action. The caller should behave as if the phase transition occurred at time **sc_time_stamp()+t**, where **t** is the time argument. Depending on the protocol, there may or may not be a subsequent response on the backward path. Following the call to **nb_transport**, the caller should sync in order that other processes may run (synchronization-on-demand).

TLM_COMPLETED. The callee has accepted the transaction, and the transaction has been completed. Following the call to **nb_transport**, the caller should inspect the transaction object and take the appropriate action. The caller should behave as if the transaction completed at time **sc_time_stamp()+t**, where **t** is the time argument. The caller should ignore the phase argument, since a transition to the final phase is implicit in this return value. The caller can assume that there will be no further response through the backward path. Depending on the specific details of the protocol and the modeling style, the caller may or may not need to yield control to the SystemC scheduler following the call.

4.2.10. Coding styles and tlm_sync_enum

The recommended usage of the tlm_sync_enum return value across the coding styles is as follows:

tlm_sync_enum	**Coding style**
TLM_REJECTED	-[6]
TLM_SYNC	Loosely-timed and approximately-timed
TLM_SYNC_CONTINUE	Approximately-timed only
TLM_COMPLETED	Loosely-timed and approximately-timed

4.2.11. Coding styles and tlm_phase[7]

The following table summarizes the usage of the **tlm_phase** values across the coding styles:

tlm_phase	**Coding style**	**Path**
BEGIN_REQ	Loosely-timed and approximately-timed	Forward, initiator-to-target
END_REQ	Approximately-timed only	Forward, initiator-to-target
BEGIN_RESP	Loosely-timed and approximately-timed	Backward, target-to-initiator
END_RESP	Approximately-timed only	Backward, target-to-initiator

4.2.12. Transaction lifetime example

A typical transaction lifetime might run as follows. A transaction is created by an initiator and is passed as an **nb_transport** argument to an interconnect component representing a bus bridge, which returns with a value of TLM_SYNC. That bridge component in turn passes the transaction as an **nb_transport** argument to a target, which is able to process the transaction immediately, and returns with a value of TLM_COMPLETED. The bridge component then passes the transaction as an **nb_transport** argument back to the initiator through the backward path between initiator and bridge. Finally, the initiator maintains the transaction object in a memory pool for subsequent reuse.[8]

---

[6] Need to discuss which coding styles TLM_REJECTED falls into

[7] We are lacking a description of the precise rules for using these phases in each of the coding styles

[8] Trevor has commented: “This sounds like a bad example. If my slave can respond immediately, I really don't want the interconnect to split up the transaction requiring the backwards path and imposing additional overhead.”

4.3. Direct memory interface

4.3.1. Introduction

The Direct Memory Interface, or DMI, provides a means by which an initiator can get direct access to an area of memory owned by a target, thereafter accessing that memory using a direct pointer rather than through the transport interface. The DMI offers a large potential increase in simulation speed for memory access between initiator and target, because once established it is able to bypass the normal path of multiple **transport** or **nb_transport** calls from initiator through interconnect components to target.

There are two direct memory interfaces, one for calls on the forward path from initiator to target, and a second for calls on the backward path from target to initiator. The forward path is used to request DMI access to a given address, and returns a reference to a DMI descriptor of type **tlm_dmi**, which contains the bounds of the DMI region and the kind of access (read or write) permitted. The backward path is used by the target to invalidate DMI pointers previously established using the forward path. The forward and backward paths may pass through zero, one or many interconnect components.

The DMI descriptor returns latency values for use by the initiator, and so provides sufficient timing accuracy for loosely-timed modeling.

DMI may be used for debug, but the debug transaction interface itself is usually sufficient because debug traffic is usually light, and usually dominated by I/O rather than memory access. Debug transactions are not usually on the critical path for simulation speed. If DMI were used for debug, the latency values should be ignored.

4.3.2. Class definition

```
namespace tlm {

// From tlm_helpers.h
// enum tlm_endianness { TLM_LITTLE_ENDIAN, TLM_BIG_ENDIAN };[9]

class tlm_dmi
{
public:
  tlm_dmi() { init(); }

  void init();

  unsigned char* dmi_ptr;

  sc_dt::uint64 dmi_start_address;
  sc_dt::uint64 dmi_end_address;

  enum Type { READ = 0x1, WRITE = 0x2, READ_WRITE = READ|WRITE };
  Type type;
```

[9] Descriptions of the generic payload helper functions are still to be inserted. They depend on endianness.


```
  sc_core::sc_time  read_latency;
  sc_core::sc_time  write_latency;

  tlm_endianness endianness;
};

class tlm_fw_direct_mem_if : public virtual sc_core::sc_interface
{
public:
  virtual bool get_direct_mem_ptr(const sc_dt::uint64& address, bool for_reads, tlm_dmi& dmi_data) = 0;
};

class tlm_bw_direct_mem_if : public virtual sc_core::sc_interface
{
public:
  virtual void invalidate_direct_mem_ptr(sc_dt::uint64 start_range, sc_dt::uint64 end_range) = 0;
};

} // namespace tlm
```

4.3.3. get_direct_mem_ptr method

a) The **get_direct_mem_ptr** method shall only be called by an initiator or by an interconnect component.

b) The **address** argument shall be the address for which direct memory access is requested.

c) The **for_reads** argument shall indicate whether the initiator is requesting read access (**for_reads** is **true**) or write access (**for_reads** is **false**) to the given address.

d) The **dmi_data** argument shall be a reference to a DMI descriptor constructed by the initiator.

e) Any interconnect components should pass on the **get_direct_mem_ptr** call along the forward path from initiator to target. In other words, the implementation of **get_direct_mem_ptr** for the target socket of the interconnect component may call the **get_direct_mem_ptr** method of the initiator socket, decoding and where necessary modifying the address exactly as they would for the corresponding transport interface. For example, an interconnect component may need to mask the address (reducing the number of significant bits) according to the address width of the target and its location in the system memory map.

f) An interconnect component need not pass on the **get_direct_mem_ptr** call in the event that it detects an addressing error.

g) Any interconnect components shall pass on the **for_reads** and **dmi_data** arguments without modification in the forward direction, although they may modify the DMI descriptor upon return from the **get_direct_mem_ptr** method, that is, when unwinding the function calls from target back to initiator.

h) If the target is able to support DMI access to the given address, it shall set the members of the DMI descriptor as described below and set the return value of the function to **true**.

i) If the target is not able to support DMI access to the given address, it shall set only the address range and type members of the DMI descriptor as described below and set the return value of the function to **false**.

4.3.4. tlm_dmi class

a) A DMI descriptor is an object of class **tlm_dmi**. DMI descriptors shall be constructed by initiators, but their members may be set by interconnect components or targets. An initiator may re-use a DMI descriptor from one call to the next, in which case the initiator shall call the **init** method to re-initialize the object between calls to **get_direct_mem_ptr**.

b) Since an interconnect component is not permitted to modify the DMI descriptor as it is passed on towards the target, the DMI descriptor shall be in its initial state when it is received by the target.

c) Method **init** shall initialize the members as described below.

d) Member **dmi_ptr** shall be set to point to the storage location corresponding to the value of member **dmi_start_address**. This shall be less than or equal to the address requested in the call to **get_direct_mem_ptr**. The initial value shall be 0.

e) The storage in the DMI region is represented with type **unsigned char***. For a full description of how memory organization and endianness are handled in TLM2, see clause 9.10 Data pointer attribute

f) Members **dmi_start_address** and **dmi_end_address** shall be set to the addresses of the first and the last bytes in the DMI region, where the meaning of the DMI region is determined by the value returned from the **get_direct_mem_ptr** method (**true** or **false**).

g) If **get_direct_mem_ptr** return the value **true**, the DMI region indicated by **dmi_start_address** and **dmi_end_address** is a region for which DMI access is allowed. On the other hand, if **get_direct_mem_ptr** return the value **false,** it is a region for which DMI access is disallowed.

h) A target or interconnect component receiving two or more calls to **get_direct_mem_ptr** may return two or more overlapping allowed DMI regions or two or more overlapping disallowed DMI regions.

i) A target or interconnect component shall not return overlapping DMI regions where one region is allowed and the other is disallowed for the same access type, that is both READ or READ_WRITE or both WRITE or READ_WRITE, without making an intervening call to **invalidate_direct_mem_ptr** to invalidate the first region.

j) In other words, the definition of the DMI regions shall not be dependent upon the order in which they were created unless the first region is invalidated by an intervening call to **invalidate_direct_mem_ptr**. Specifically, the creation of a disallowed DMI region shall not be permitted to punch a hole in an existing allowed DMI region, or vice versa. Moreover READ and WRITE DMI regions are defined independently of one another.

k) Any interconnect components that pass on the **get_direct_mem_ptr** call are obliged to transform these addresses as they do the address **argument**. Any transformations on the addresses in the DMI descriptor shall occur as the descriptor is passed along the backward path. For example, the target may set **dmi_start_address** to a relative address within the memory map known to that target, in which case the interconnect component is obliged to transform the relative address back to an absolute address in the system memory map. The initial values shall be 0 and the maximum value of type sc_dt::uint64, respectively.

l) A target may disallow DMI access to the entire address space (**dmi_start_address** = 0, **dmi_end_address** = maximum value), perhaps because the target does not support DMI access at all, in which case an interconnect

component should clip this disallowed region down to the part of the memory map occupied by the target. Otherwise, if an interconnect component fails to clip the address range, then an initiator would be mislead into thinking that DMI was disallowed across the entire system address space.

m) Member **type** shall be set to indicate the kind of DMI access permitted, READ for readonly access, WRITE for writeonly access, or READ_WRITE for read-write access. If **for_reads** is **true**, type shall be set to READ or READ_WRITE. If **for_reads** is **false**, type shall be set to WRITE or READ_WRITE. The initial value shall be READ_WRITE. An interconnect component is permitted to restrict the kind of access by overwriting a value of READ_WRITE with READ or WRITE.

n) Members **read_latency** and **write_latency** shall be set to the latencies of read and write memory transactions, respectively. The initial values shall be SC_ZERO_TIME. Both interconnect components and the target may increase the value of either latency such that the latency accumulates as the DMI descriptor is passed back from target to initiator on return from the **get_direct_mem_ptr** method. One or both latencies will be valid, depending on the value of the member **type**.

o) The initiator is responsible for respecting the latencies whenever it accesses memory using the direct memory pointer. If the initiator chooses to ignore the latencies, this may result in timing inaccuracies.

p) Since DMI gives direct access to a memory region in the target, the memory organization will have the endianness of the target[10].

q) Member **endianness** shall be set by the target to indicate the endianness of the target.

4.3.5. invalidate_direct_mem_ptr method

a) The **invalidate_direct_mem_ptr** method shall only be called by a target or an interconnect component.

b) A target is obliged to call **invalidate_direct_mem_ptr** before any change that would modify the validity or the access type of any existing DMI region. For example, before restricting the address range of an existing DMI region, before changing the access type from READ_WRITE to READ, or before re-mapping the address space.

c) The **start_range** and **end_range** arguments shall be the first and last addresses of the address range for which DMI access is to be invalidated.

d) An initiator receiving an incoming call to **invalidate_direct_mem_ptr** shall immediately invalidate and discard any DMI region (previously received from a call to **get_direct_mem_ptr**) that overlaps with the given address range.

e) In the case of a partial overlap, that is, only part of an existing DMI region is invalidated, an initiator may adjust the boundaries of the existing region or may invalidate the entire region.

f) Any interconnect components are obliged to pass on the **invalidate_direct_mem_ptr** call along the backward path from target to initiator, decoding and where necessary modifying the address arguments as they would for the corresponding transport interface. Because the transport interface transforms the address on the forward path and DMI on the backward path, the transport and DMI transformations should be the inverse of one another.

g) Since there may be multiple initiators each getting direct memory pointers to the same target, a safe implementation is for an interconnect component to call **invalidate_direct_mem_ptr** for every initiator.

---

[10] "The endianness of the target" is ambiguous. This needs revisiting when we have concluded the discussion on endianness in general.

h) An interconnect component can invalidate all direct memory pointers in an initiator by setting **start_range** to 0 and **end_range** to the maximum value.

i) An implementation of **invalidate_direct_mem_ptr** shall not call **get_direct_mem_ptr**, directly or indirectly.

4.3.6. Optimization using a DMI Hint

a) The DMI hint is a mechanism to optimize simulation speed by avoiding the need to repeatedly poll for DMI access. Instead of calling **get_direct_mem_ptr** to check for the availability of a DMI pointer, an initiator can check the DMI hint of a normal transaction passed through the transport interface.

b) The generic payload provides a DMI hint. User-defined transactions could implement a similar mechanism, in which case the target should set the value of the DMI hint appropriately.

c) Use of the DMI hint is optional. An initiator is free to ignore the DMI hint of the generic payload.

d) For an initiator wishing to use DMI, the recommended sequence of actions is as follows:

i. The initiator should check the address against its cache of valid DMI regions

ii. If there is no existing DMI pointer, the initiator should perform a normal transaction through the transport interface

iii. Following that, the initiator should check the DMI hint of the transaction

iv. If the hint indicates DMI is allowed, the initiator should call **get_direct_mem_ptr**

4.4. Debug transaction interface

4.4.1. Introduction

The debug transaction interface provides a means to read and write to storage in a target, over the same forward path from initiator to target as is used by the transport interface, but without any of the delays, waits, event notifications or side effects associated with a regular transaction. Because the debug transaction interface follows the same path as the transport interface, the implementation of the debug transaction interface can perform the same address translation as for regular transactions.

For example, the debug transaction interface could permit a software debugger attached to an ISS to peek or poke an address in the memory of the simulated system from the point of view of the simulated CPU. The debug transaction interface could also allow an initiator to take a snapshot of system memory contents during simulation for diagnostic purposes, or to initialize some area of system memory at the end of elaboration.

4.4.2. Class definition

```
namespace tlm {

class tlm_debug_payload
{
public:
  sc_dt::uint64 address;
```



```
    bool do_read;
    unsigned int num_bytes;
    unsigned char* data;
};

class tlm_transport_dbg_if : public virtual sc_core::sc_interface
{
public:
  virtual unsigned int transport_dbg(tlm_debug_payload& r) = 0;
};

} // namespace tlm
```

4.4.3. Rules

a) Calls to **transport_dbg** shall follow the same forward path as the transport interface used for normal transactions.

b) The initiator shall construct the debug payload object, and shall set each of the members of the object before passing it as an argument to **transport_dbg**.

c) Member **address** shall be set to the first address in the region to be read or written. Member **address** may be modified by any interconnect component that is performing address decoding or translation. For example, an interconnect component may need to mask the address (reducing the number of significant bits) according to the address width of the target and its location in the system memory map.

d) Member **address** may be modified several times if a debug payload is forwarded through several interconnect components. When the debug payload is returned to the initiator, the original value of member **address** may have been overwritten.

e) Member **do_read** shall be set to **true** to read from (or peek) the target, or **false** to write to (or poke) the target.

f) Member **num_bytes** shall be set to the number of bytes to be read or written. This may be 0[11].

g) Member **data** shall be set to the address of an array from which values are to be copied to the target (for a write), or to which values are to be copied from the target (for a read). This array shall be allocated by the initiator, and shall not be deleted before the return from **transport_dbg**. The size of the array shall be at least **num_bytes**.

h) The implementation of **transport_dbg** in the target shall read or write the given number of bytes using the given address (after address translation through the interconnect), if it is able.

i) The **data** array shall always have the endianness of the host machine. The implementation of **transport_dbg** shall be responsible for converting between the endianness used by the target and the endianness of the host machine such that the debug transaction interface as it appears to the initiator is independent of target endianness.

j) **transport_dbg** shall return a count of the number of bytes actually read or written, which may be less than **num_bytes**. If the target is not able to perform the operation, it shall return a value of 0.

---

[11] I was thinking that a debug transaction may just want to “touch” a target at an address. Olaf: As a debug transaction is supposed to be side-effect free, "touching" an address should be a NOP

k) **transport_dbg** shall not call wait, shall not create any event notifications, and shall not have any side effects on the target or any interconnect component.

5. Sockets and combined interfaces

5.1.1. Introduction

A socket combines a port with an export. An initiator socket has a port for the forward path and an export for the backward path, whilst a target socket has an export for the forward path and a port for the backward path. The sockets also overload the SystemC port binding operators to bind both the port and export to the export and port in the opposing socket.

The combined interfaces create groupings of the core TLM2 interfaces, primarily for use by the initiator and target sockets. Note that these groupings include the transport, DMI and debug transaction interfaces, but do not include any TLM1 core interfaces.

The groups are categorized along two dimensions: forward and backward interfaces, and blocking and non-blocking transport interfaces. The forward interfaces provide method calls on the forward path from initiator socket to target socket, and the backward interfaces on the backward path from target socket to initiator socket. Neither the blocking transport interface nor the debug transaction interface require a backward calling path.

5.1.2. Class definition

```
namespace tlm {

// The forward non-blocking interface:
template <typename TRANS = tlm_generic_payload, typename PHASE = tlm_phase>
class tlm_fw_nb_transport_if
  : public virtual tlm_nonblocking_transport_if<TRANS, PHASE>
  , public virtual tlm_fw_direct_mem_if
  , public virtual tlm_transport_dbg_if
{};

// The backward non-blocking interface:
template <typename TRANS = tlm_generic_payload, typename PHASE = tlm_phase>
class tlm_bw_nb_transport_if
  : public virtual tlm_nonblocking_transport_if<TRANS, PHASE>
  , public virtual tlm_bw_direct_mem_if
{};

// The forward blocking interface:
template <typename TRANS = tlm_generic_payload>
class tlm_fw_transport_if
  : public virtual tlm_blocking_transport_if<TRANS>
  , public virtual tlm_fw_direct_mem_if
  , public virtual tlm_transport_dbg_if
{};

// The backward blocking interface:
```



```
class tlm_bw_transport_if
  : public virtual tlm_bw_direct_mem_if
{};

// Initiator socket
template <
   unsigned int BUSWIDTH = 32,
   typename FW_IF = tlm_fw_nb_transport_if<>,
   typename BW_IF = tlm_bw_nb_transport_if<>
>
class tlm_initiator_socket : public sc_core::sc_port<FW_IF>
{
public:
   typedef FW_IF fw_interface_type;
   typedef BW_IF bw_interface_type;
   typedef sc_core::sc_port<fw_interface_type> port_type;
   typedef sc_core::sc_export<bw_interface_type> export_type;
   typedef tlm_ target_socket<BUSWIDTH, fw_interface_type, bw_interface_type> target_socket_type;

   tlm_initiator_socket( const char* );

   unsigned int get_bus_width() const;

   void bind( target_socket_type& );
   void operator() ( target_socket_type& );
   void bind( tlm_initiator_socket& );
   void operator() ( tlm_initiator_socket& );
   void bind( bw_interface_type& );
   void operator() ( bw_interface_type& );

protected:
  export_type mExport;
};

// Target socket
template <
   unsigned int BUSWIDTH = 32,
   typename FW_IF = tlm_fw_nb_transport_if<>,
   typename BW_IF = tlm_bw_nb_transport_if<>
>
class tlm_target_socket : public sc_core::sc_export<FW_IF>
{
public:
   typedef FW_IF fw_interface_type;
   typedef BW_IF bw_interface_type;
   typedef sc_core::sc_port<bw_interface_type> port_type;
```



```
    typedef sc_core::sc_export<fw_interface_type> export_type;
    typedef tlm_initiator_socket<BUSWIDTH, fw_interface_type, bw_interface_type> initiator_socket_type;

    tlm_target_socket( const char* );

    unsigned int get_bus_width() const;

    void bind( initiator_socket_type& );
    void operator() ( initiator_socket_type& );
    void bind( tlm_target_socket& );
    void operator() ( tlm_target_socket& );
    void bind( fw_interface_type& );
    void operator() ( fw_interface_type& );

    bw_interface_type* operator-> ();

  protected:
    port_type mPort;
  };

  } // namespace tlm
```

5.1.3. Rules

a) The constructor shall pass the character string argument to the constructor belonging to the base class (**sc_port** or **sc_export**) to set the string name of the instance in the module hierarchy.

b) The method **get_bus_width** shall return the value of the BUSWIDTH template argument.

c) Template argument BUSWIDTH shall determine the word length for each individual data word transferred through the socket, expressed as the number of bits in each word. For a burst transfer, BUSWIDTH shall determine the number of bits in each beat of the burst. The precise interpretation of this attribute shall depend on the transaction type. For the meaning of BUSWIDTH with the generic payload, see clause 9.11 Data length attribute.

d) When binding socket-to-socket, the two sockets may have different values for the BUSWIDTH template argument. The BUSWIDTH decorates each individual socket, but has no impact on binding or type compatibility. Executable code in the initiator or target may get and act on the BUSWIDTH.

e) Each of the methods **bind** and **operator()** that take a socket as an argument shall bind the socket instance to which the method belongs to the socket instance passed as an argument to the method.

f) Each of the methods **bind** and **operator()** that take an interface as an argument shall bind the export of the socket instance to which the method belongs to the channel instance passed as an argument to the method. (A channel is the SystemC term for a class that implements an interface.)

g) When binding initiator socket to target socket, the methods **bind** and **operator()** shall bind the port of the initiator socket to the export of the target socket, and the port of the target socket to the export of the initiator socket. This is for use when binding sockets-to-socket at the same level in the hierarchy.

h) When binding initiator socket to initiator socket or target socket to target socket, the methods **bind** and **operator()** shall bind the port of one socket to the port of the other socket, and the export of one socket to the export of the other socket. This is for use when binding child socket to parent socket, or parent socket to child socket, passing transactions up or down the module hierarchy.

i) An initiator socket may be bound to a target socket by calling the bind method of either socket.

j) (Hierarchical binding of sockets needs to be resolved[12])

k) The method **operator->** of the target socket shall call method **operator->** of the port in the target socket (on the backward path), and shall return the value returned by **operator->** of the port.

[12] There are still some issues with hierarchical binding in the current kit

6. Analysis ports[13]

(This section is to be completed)

[13] We need to decide which parts are in (FIFOs?, triple?), and whether to rename for non-analysis usage

7. Quantum keeper

7.1. Introduction

Temporal decoupling permits SystemC processes to run ahead of simulation time for an amount of time known as the quantum, and maintained by the class **tlm_quantumkeeper**. When using the loosely-timed coding style, the delays annotated to the **nb_transport** method may be interpreted as local times defined relative to the current simulation time **sc_time_stamp()**.

Temporal decoupling permits a significant simulation speed improvement by reducing the number of context switches and events.

For a general description of temporal decoupling, see clause 3.3.2 Loosely-timed coding style and temporal decoupling

For a description of timing annotation and the non-blocking transport interfaces, see clause 4.2.8 The sc_time argument

7.2. Class definition

```
namespace tlm {

class tlm_quantumkeeper
{
public:
  static void setGlobalQuantum(const sc_core::sc_time& t) { mGlobalQuantum = t; }
  static const sc_core::sc_time& getGlobalQuantum() { return mGlobalQuantum; }

  tlm_quantumkeeper() : mLocalTime(sc_core::SC_ZERO_TIME) { computeLocalQuantum(); }

  void inc(const sc_core::sc_time& t) { mLocalTime += t; }
  sc_core::sc_time getCurrentTime() const { return sc_core::sc_time_stamp() + mLocalTime; }
  const sc_core::sc_time& getLocalTime() const { return mLocalTime; }
  sc_core::sc_time& getLocalTime() { return mLocalTime; }

  bool needSync() const { return mLocalTime >= mLocalQuantum; }
  void reset() { mLocalTime = sc_core::SC_ZERO_TIME; computeLocalQuantum(); }
  void sync() { sc_core::wait(mLocalTime); reset(); }

protected:
  void computeLocalQuantum() { mLocalQuantum = mGlobalQuantum; }
```



```
private:
  static sc_core::sc_time mGlobalQuantum;
  sc_core::sc_time mLocalQuantum;
  sc_core::sc_time mLocalTime;
};

} // namespace tlm
```

7.3. Rules for processes using temporal decoupling

a) There is a single global quantum maintained by the class **tlm_quantumkeeper**.

b) Temporal decoupling runs in the context of the standard SystemC simulation kernel, so events can be scheduled, processes suspended and resumed, and loosely-timed models can be mixed with other coding styles.

c) For maximum simulation speed, all initiators should use temporal decoupling, and the number of other runnable SystemC processes should be zero or minimized.

d) In an ideal scenario, the only runnable SystemC processes will belong to temporally decoupled initiators, and each process will run ahead to the next quantum boundary before yielding to the SystemC kernel.

e) *Yield* means call **wait** in the case of a thread process, or return from the function in the case of a method process.

f) Calls to the **sc_time_stamp** method will return the simulation time as it was at the start of the current quantum.

g) The process should accumulate any local processing delays and communication delays in a local variable, referred to in this clause as the local time offset.

h) The effective local time is **sc_time_stamp() + local_time_offset.**

i) The local time offset is unknown to the SystemC scheduler.

j) The **tlm_quantumkeeper** class provides some machinery to maintain the quantum and the local time offset. It includes a static global quantum, a local time, and methods to increase the local time, compare it to the quantum, and to synchronize. The implementation of the class is listed above to reveal the details.

k) The given quantum keeper should be considered a default implementation. Applications may derive their own quantum keeper from class **tlm_quantumkeeper** and override the methods.

l) When the local time offset is greater than the quantum, the process shall yield to the kernel.

m) Any access to a non-local object will give the value as at the start of the quantum.

n) All processes using temporal decoupling shall be synchronized to the SystemC simulation time at the start of each quantum, and possibly at additional times using synchronization-on-demand.

o) The quantum should be chosen to be less than the typical communication interval between initiators, otherwise important process interactions may be lost, and the model broken.

p) If an initiator needs to synchronize before the end of the quantum, that is, if an initiator needs to suspend execution such that simulation time can catch up with the local time and any other processes have the chance to execute, the initiator should call the **sync** method. This is synchronization-on-demand.

q) The **sync** method shall suspend the calling process until simulation time equals the effective local time, and shall then recalculate the end time for the next quantum.

r) If both the blocking and non-blocking transport interfaces are being used together, an initiator using temporal decoupling may be in the situation where it directly or indirectly makes a call to **b_transport**. typically in an adaptor. This is only possible if the initiator is a thread process. Because a blocking method may call **wait**, it is critical that the initiator should synchronize before the call to **b_transport**.

8. Payload event queue[14]

8.1. Introduction

The payload event queue is a class that maintains a queue of SystemC event notifications, where each notification carries an associated transaction object. Transactions are written to the queue annotated with a delay, and each transaction pops out of the back of queue at a time calculated from the simulation time when it was written to the queue plus the annotated delay.

The payload event queue is particularly useful when combining the non-blocking interface with the approximately-timed coding style. But there is no obligation to use a payload event queue for this purpose. The **tlm_peq** class not a definitive part of the TLM2 standard, but is included as an example. The full source code is listed below.

8.2. Class definition

```
namespace tlm {

template< typename T>
class tlm_peq : public sc_core::sc_module , public virtual delayed_analysis_if[15]< T >
{
public:
   sc_core::sc_export< delayed_analysis_if< T > > delayed_analysis_export;
   analysis_port< T > ap;

   SC_HAS_PROCESS( tlm_peq );

   tlm_peq( sc_core::sc_module_name nm ) : sc_core::sc_module( nm ) , ap("ap")
   {
      delayed_analysis_export( *this );
      SC_METHOD( wake_up_method );
      dont_initialize();
      sensitive << m_wake_up;
   }

   int size() const { return m_map.size(); }

   int posted_before( const sc_core::sc_time &time ) const {
      int i = 0;
      for( typename std::multimap< sc_core::sc_time , T>::const_iterator iter = m_map.begin();
         iter != m_map.end(); ++iter ) {
```

[14] This description of the PEQ is only partial as yet

[15] We need to address the issue of renaming the analysis_if, because the PEQ is not doing analysis


```
        if( (*iter).first < time ) { i++; }
      }
      return i;
    }

    int posted_at( const sc_core::sc_time &time ) const { return m_map.count( time ); }

    void write( const T &transaction , const sc_core::sc_time &time ) {
      m_map.insert( pair_type( time + sc_core::sc_time_stamp() , transaction ) );
      m_wake_up.notify( time );
    }

  private:
    typedef std::pair<sc_core::sc_time,T> pair_type;

    void wake_up_method() {
      pair_type p;
      sc_core::sc_time now = sc_core::sc_time_stamp();
      assert( m_map.size() > 0 );
      assert( (*(m_map.begin())).first == now );

      for( p = *(m_map.begin()); p.first == now; p = *(m_map.begin()) )
      {
        ap.write( p.second );
        m_map.erase( m_map.begin() );
        if( m_map.size() == 0 ) { return; }
        p = *(m_map.begin());
      }

      m_wake_up.notify( p.first - now );
    }

    sc_core::sc_event m_wake_up;
    std::multimap< sc_core::sc_time , T> m_map;
  };

  } // namespace tlm
```


9. Generic payload

9.1. Introduction

The generic payload is intended to improve the interoperability of memory-mapped bus models, which it does at two levels. Firstly, the generic payload can be used as a standard transaction type (passed as a template argument to the core interfaces) when creating abstract models of memory-mapped busses where the precise details of the bus protocol are unimportant. Secondly, the generic payload can be used as the basis for creating detailed models of specific bus protocols, with the advantage of reducing the implementation cost and increasing simulation speed when there is a need to bridge or adapt between different protocols.

The generic payload is specifically aimed at modeling memory-mapped busses. It includes some of the attributes found in typical memory-mapped bus protocols such as command, address, data, byte enables, single word transfers, burst transfers, streaming, and response status. The generic payload may be used as the basis for modeling protocols other than memory-mapped busses, but memory-mapped busses are its primary purpose.

The generic payload does not include every attribute found in typical memory-mapped bus protocols, but it does include an extension mechanism so that applications can add their own specialised attributes.

For specific protocols, whether standard or proprietary, modeling and interoperability are the responsibility of the protocol owner and are outside the scope of OSCI. It is up to the protocol owner to proliferate models or coding guidelines for their own particular protocol. However, the generic payload is still applicable here, because it provides a common starting point for model creation, and in many cases will reduce the cost of bridging between different protocols in a transaction-level model.

9.2. Extensions and interoperability

The goal of the generic payload is to enable interoperability between memory-mapped bus models, but all busses are not created equal. Given two transaction-level models that use different protocols and that model those protocols at a detailed level, then just as in a physical system, a bridge must be inserted between those models to perform protocol conversion and allow them to communicate. On the other hand, many transaction level models produced early in the design flow do not care about the specific details of any particular protocol. For such models it is sufficient to copy a block of data starting at a given address, and for those models the generic payload can be used directly to give excellent interoperability.

The generic payload extension mechanism permits any number of extensions of any type to be defined and added to a transaction object. Each extension represents a new set of attributes, transported along with the transaction object. Extensions can be created, added, written and read by initiators, interconnect components, and targets alike. The extension mechanism itself does not impose any restrictions. Of course, undisciplined use of this extension mechanism would compromise interoperability, so disciplined use is strongly encouraged. But the flexibility is there where you need it!

The generic payload assists with interoperability in two ways. Firstly, for abstract modeling, it provides an off-the-shelf general-purpose payload that guarantees immediate interoperability and supports extensions (with the restriction that other component must be allowed to ignore those extensions). Secondly, for modeling specific

protocols, it reduces the cost of creating bridges between protocols, sometimes to the point where the bridge becomes trivial to write.

The use of the extension mechanism represents a trade-off between increased coding convenience when binding ports, and decreased compile-time type checking. If the undisciplined use of generic payload extensions were allowed, each application would be obliged to detect any incompatibility between extensions by including explicit run-time checks in each interconnect component and target, and there would be no mechanism to enforce the existence of a given extension. The TLM2 standard prescribes specific coding guidelines to avoid these pitfalls.

There are three, and only three, recommended alternatives for the transaction template argument TRANS of the blocking and non-blocking transport interfaces.

a) Use the generic payload directly, with ignorable extensions

b) Use an empty class derived from the generic payload

c) Use a transaction type unrelated to the generic payload

These three alternatives are defined below in order of decreasing interoperability.

9.2.1. Use the generic payload directly, with ignorable extensions

a) In this case, the transaction type is **tlm_generic_payload**. This is the default value for the TRANS argument. Any model that uses a core transport interface specialized with the generic payload will be interoperable with any other model that uses the same transport interface, provided that those models respect the semantics of the generic payload.

b) In this case, it is strongly recommended that any generic payload extensions should be ignorable. *Ignorable* means that a target or interconnect component shall not fail and shall not generate an error response because of the absence of a given extension, and that the component shall perform its primary function in the same way regardless of whether the given extension is present or absent.

c) If an extension is deemed ignorable, then it is not possible to use compile-time checking to enforce support for that extension in a target.

d) In general, an ignorable extension can be thought of as one for which there exists an obvious and safe default value, such that any interconnect component or target can behave normally in the absence of the given extension by assuming the default value. An example might be the privilege level associated with a transaction, where the default is the lowest level. In the end, the definition of *ignorable* comes down to a matter of judgement.

e) Ignorable extensions may be used to transport auxiliary, side-band, or simulation-related information or meta-data. For example, a unique transaction identifier, the wall-time when the transaction was created, or a diagnostic filename.

f) The generic payload intrinsically supports minor variations in protocol. For example, a particular component may or may not support byte enables. A target that is unable to support a particular feature of the generic payload is obliged to generate the standard error response. This should be thought of as being part of the specification of the generic payload.

g) Note that there are two separate transport interfaces, blocking and non-blocking, and that interoperability between those interfaces depends on the coding style chosen and may require adapters.

9.2.2. Use an empty class derived from the generic payload

a) In this case, the transaction type consists of an empty class derived (directly or indirectly) from **tlm_generic_payload**

b) It is strongly recommended that the class contain no data members or member functions. This ensures a consistent use of the extension mechanism and avoids variations in coding style that may hamper interoperability.

c) The generic payload extension mechanism may be used for ignorable or for mandatory extensions with no restrictions. The semantics of any extensions should be thoroughly documented with the new transaction type.

d) Because the derived class is empty, the transaction type may be upcast to type **tlm_generic_payload** with no loss of information. This provides a good starting point for building interoperable components, because the transaction can them be transported through interconnect components and targets that use the generic payload type, and can be downcast to any other type derived from the generic payload.

e) Generic payload transactions that were created by upcasting from the specific transaction type can be transported through components that support the generic payload, and can be cloned in their entirety, including all extension, and regardless of the original transaction type. Hence the generic payload can act as an “untyped” transaction. But to ensure interoperability, the user should consider the semantics of the new transaction type very carefully when upcasting.

f) There are two recommended patterns of use, outlined below.

g) The first pattern is to use the new transaction type throughout the initiator, interconnect and target, and never upcast to the generic payload. This pattern supports strong compile-time type checking when binding ports or sockets.

h) The second pattern is to upcast the transaction type to the generic payload, pass the generic payload through the interconnect, and downcast the generic payload back to the original transaction type at the target. Upcasting and downcasting in this way should be used in pairs, not in isolation.

i) When upcasting from the transaction type to type **tlm_generic_payload**, the user is obliged to consider the semantics of each extension very carefully to ensure that the transaction can be transported through components that are aware of the generic payload but not the extensions. There is no general rule. Some extensions can be transported through generic payload components without mishap, for example an attribute specifying the security level of the data. Other extensions will require explicit adaption or might not be supportable at all, for example an attribute specifying that the interconnect is to be locked.

9.2.3. Use a transaction type unrelated to the generic payload

a) In this case, the transaction type is a new class unrelated to the generic payload

b) The new transaction type should not be derived from the generic payload, and should not include a member of type **tlm_generic_payload** or a pointer or reference to that type.

c) This choice is only justified when the new transaction type bears little or no resemblance to the generic payload, because all the interoperability benefits of using the generic payload are lost.

9.3. Generic payload attributes and methods

The generic payload class contains a set of private attributes, and a set of public access functions to get and set the values of those attributes. The exact implementation of those access functions is implementation-defined.

The majority of the attributes are set by the initiator and shall not be modified by any interconnect component or target. Only the address, return status and extension attributes may be modified by an interconnect component or by the target. In the case of a read command, the target may also modify the data array.

9.4. Class definition

```
namespace tlm {

class tlm_extension_base
{
public:
   virtual tlm_extension_base* clone() const = 0;
   virtual ~tlm_extension_base() {}
};

template <typename T>
class tlm_extension : public tlm_extension_base
{
public:
   virtual tlm_extension_base* clone() const = 0;
   virtual ~tlm_extension() {}
   const static unsigned int ID;
};

enum tlm_command {
   TLM_READ_COMMAND,
   TLM_WRITE_COMMAND,
   TLM_IGNORE_COMMAND
};

enum tlm_response_status {
   TLM_OK_RESPONSE = 1,
   TLM_INCOMPLETE_RESPONSE = 0,
   TLM_GENERIC_ERROR_RESPONSE = -1,
   TLM_ADDRESS_ERROR_RESPONSE = -2,
   TLM_COMMAND_ERROR_RESPONSE = -3,
   TLM_BURST_ERROR_RESPONSE = -4,
   TLM_BYTE_ENABLE_ERROR_RESPONSE = -5
 };

class tlm_generic_payload {
```



```
public:
  // Constructors, clone, destructor
  tlm_generic_payload();
  tlm_generic_payload(const tlm_generic_payload&);
  virtual tlm_generic_payload* clone() const;
  virtual ~tlm_generic_payload();

  // Access methods
  inline tlm_command get_command() const;
  inline void set_command( const tlm_command );
  inline bool is_read();
  inline void set_read();
  inline bool is_write();
  inline void set_write();

  inline sc_dt::uint64 get_address() const;
  inline void set_address( const sc_dt::uint64 );

  inline unsigned char* get_data_ptr() const;
  inline void set_data_ptr( unsigned char* );

  inline unsigned int get_data_length() const;
  inline void set_data_length( const unsigned int );

  inline unsigned int get_streaming_width[16]() const;
  inline void set_streaming_width( const unsigned int );

  inline bool* get_byte_enable_ptr() const;
  inline void set_byte_enable_ptr( bool* );
  inline unsigned int get_byte_enable_length() const;
  inline void set_byte_enable_length( const unsigned int );

  // DMI hint
  void set_dmi_allowed( bool );
  bool get_dmi_allowed() const;

  inline tlm_response_status get_response_status() const;
  inline void set_response_status( const tlm_response_status );
  inline std::string[17] get_response_string();
  inline bool is_response_ok();
  inline bool is_response_error();
```

[16] The name streaming_width is only provisional.

[17] Olaf: std::string is inefficient. char* is faster. John: we've tried to move SystemC from char* to std::string, and get_response_string should't be on the critical path for speed anyway.


```
    // Extension mechanism
    template <typename T> void set_extension( T* );
    void set_extension( unsigned int , tlm_extension_base* );

    template <typename T> void get_extension( T*& ) const;
    tlm_extension_base* get_extension( unsigned int ) const;

    template <typename T> void clear_extension( const T* );
    void clear_extension( unsigned int );
    void resize_extensions();
};

} // namespace tlm
```

9.5. Memory management[18]

The initiator shall be responsible for memory allocation and memory management for the transaction object. This could be static, automatic (stack) or dynamically allocated (new) storage. The initiator may create a new object for each transaction instance, or may implement a memory pooling strategy to reuse transaction objects.

The initiator shall not delete the transaction object until the lifetime of the transaction is complete. The initiator may determine when the transaction is complete using the transaction phase, which depends on the coding style used.

The initiator shall be responsible for setting the data pointer and byte enable pointer attributes to existing storage, which could be static, automatic (stack) or dynamically allocated (new) storage. The initiator shall not delete this storage before the lifetime of the transaction is complete.

These obligations apply to the generic payload and to transaction types derived from the generic payload. Similar obligations would apply to transaction types unrelated to the generic payload.

9.6. Constructors, clone, destructor

a) The default constructor shall set the generic payload attributes to their default values, as defined in the following clauses.

b) The copy constructor shall make a complete and deep copy of the object, including the data array, byte enable array, and extensions, such that the copy can survive the destruction of the original transaction object with no visible side-effects.

c) The **clone** method shall have the same behavior as the copy constructor. Note that the clone method is virtual (polymorphic), but the copy constructor is not. For example, the clone method may be used to make a new copy of a transaction for analysis beyond the lifetime of the transaction. The clone method may also be required when creating a bridge between two different transactions each derived from the generic payload.

[18] Memory management for the generic payload is underdiscussion. Analysis needs a deep copy, a transaction bridge needs a shallow copy.

d) The virtual destructor is present because of the existence of the virtual **clone** method. The destructor and clone methods are not expected to be overridden, but they may be overridden to implement some user-defined memory management strategy for transaction data and extensions.

9.7. Default values and modifiability of attributes

The default values and modifiability of the generic payload attributes are summarized in the following table:

Attribute	**Default value**	**Modifiable by interconnect?**	**Modifiable by target?**
Command	TLM_IGNORE_COMMAND	No	No
Address	0	**Yes**	No
Data pointer	0	No	No
Data array	-	No	**Yes** (read)
Data length	1	No	No
Byte enable pointer	0	No	No
Byte enable array	-	No	No
Byte enable length	1	No	No
Streaming width	0	No	No
DMI allowed	false	**Yes**	**Yes**
Response status	TLM_INCOMPLETE_RESPONSE	No	**Yes**
Extensions	Number of tlm_extension template instances	**Yes**	**Yes**
Lock	*(Under consideration for future inclusion)*		

9.8. Command attribute

a) The method **set_command** shall set the command attribute to the value passed as an argument. The method **get_command** shall return the current value of the command attribute.

b) The methods **set_read** and **set_write** shall set the command attribute to TLM_READ_COMMAND and TLM_WRITE_COMMAND respectively. The methods **is_read** and **is_write** shall return **true** if and only if the current value of the command attribute is TLM_READ_COMMAND and TLM_WRITE_COMMAND respectively.

c) A read command is a generic payload transaction with the command attribute equal to TLM_READ_COMMAND. A write command is a generic payload transaction with the command attribute equal to TLM_WRITE_COMMAND.

d) On receipt of a read command, the target shall copy the contents of a local array in the target to the array pointed to be the data pointer attribute, honoring all the semantics of the generic payload as defined by this standard.

e) On receipt of a write command, the target shall copy the array pointed to by the data pointer attribute to a local array in the target, honoring all the semantics of the generic payload as defined by this standard.

f) If the target is unable to execute a read or write command, it shall generate a standard error response. The recommended response status is TLM_COMMAND_ERROR_RESPONSE.

g) On receipt of a generic payload transaction with the command attribute equal to TLM_IGNORE_COMMAND, the target shall not execute a write command or a read command. In particular, it shall not modify the value of the local array that would be modified by a write command, or modify the value of the array pointed to by the data pointer attribute. The target may, however, use the value of any attribute in the generic payload, including any extensions.

h) The command attribute shall be set by the initiator, and shall not be overwritten by any interconnect component or target.

i) The default value of the command attribute shall be TLM_IGNORE_COMMAND.

9.9. Address attribute

a) The method **set_address** shall set the address attribute to the value passed as an argument. The method **get_address** shall return the current value of the address attribute.

b) For a read command or a write command, the target shall interpret the current value of the address attribute as the address of the first byte to be read from or written to the local array in the target. This always correspond to the first byte in the array pointed to by the data pointer attribute.

a) A byte at a given index in the data array shall be transferred to or from the address given by the formula **address_attribute + (array_index % streaming_width)**

j) If the target is unable to execute the transaction with the given address attribute (because the address is out-of-range, for example) it shall generate a standard error response. The recommended response status is TLM_ADDRESS_ERROR_RESPONSE.

c) The address attribute shall be set by the initiator, but may be overwritten by one or more interconnect components. This may be necessary if an interconnect component performs address translation, for example to translate an absolute address in the system memory map to a relative address in the memory map known to the target. Once the address attribute has been overwritten in this way, the old value is lost (unless it was explicitly saved somewhere).

d) The default value of the address attribute shall be 0.

9.10. Data pointer attribute

a) The method **set_data_ptr** shall set the data pointer attribute to the value passed as an argument. The method **get_data_ptr** shall return the current value of the data pointer attribute. Note that the data pointer attribute is a pointer to the data array, and these methods set or get the value of the pointer, not the contents of the array.

b) It is an error to call the transport interface with a transaction object having a null data pointer attribute.

c) For a read command or a write command, the target shall copy data to or from the data array, respectively, honoring the semantics of the remaining attributes of the generic payload.

d) Endianness[19]

e) The length of the data array shall be greater than or equal to the value of the data length attribute, in bytes.

f) The data pointer attribute shall be set by the initiator, and shall not be overwritten by any interconnect component or target.

g) The storage for the data array shall be allocated by the initiator.

h) For a write command or TLM_IGNORE_COMMAND, the contents of the data array shall be set by the initiator, and shall not be overwritten by any interconnect component or target

i) For a read command, the contents of the data array shall be overwritten by the target (honoring the semantics of the byte enable) but by no other component.

j) The default value of the data pointer attribute shall be 0, the null pointer.

9.11. Data length attribute

a) The method **set_data_length** shall set the data length attribute to the value passed as an argument. The method **get_data_length** shall return the current value of the data length attribute.

b) For a read command or a write command, the target shall interpret the data length attribute as the number of bytes to be copied to or from the data array, inclusive of any bytes disabled by the byte enable attribute.

c) The data length attribute shall be set by the initiator, and shall not be overwritten by any interconnect component or target.

d) The data length attribute shall not be set to 0. In order to transfer zero bytes, the command attribute should be set to TLM_IGNORE_COMMAND.

e) For burst transfers, the word length for each transfer shall be determined by the BUSWIDTH template parameter of the initiator socket and target socket. BUSWIDTH is independent of the data length attribute. BUSWIDTH shall be expressed in bits. If the data length is less than or equal to the BUSWIDTH / 8, the transaction is effectively modeling a single-word transfer, and if greater, the transaction is effectively modeling a burst. A single transaction can be passed through sockets of different bus widths. The BUSWIDTH may be used to calculate the latency of the transfer.

f) The target may or may not support transactions with data length greater than the word length of the target, whether the word length is given by the BUSWIDTH template parameter or by some other value.

g) If the target is unable to execute the transaction with the given data length, it shall generate a standard error response. The recommended response status is TLM_BURST_ERROR_RESPONSE.

h) The default value of the data length attribute shall be 1.

[19] To be defined, including issues of host endianness, modeled processor endianness, and modeled bus endianness

9.12. Byte enable pointer attribute

a) The method **set_byte_enable_ptr** shall set the pointer to the byte enable array to the value passed as an argument. The method **get_byte_enable_ptr** shall return the current value of the byte enable pointer attribute.

b) Byte enables may be used to create burst transfers where the address increment between each beat is greater than the number of significant bytes transferred on each beat, or to place words in selected byte lanes of a bus. At a more abstract level, byte enables may be used to create “lacy bursts” where the data array of the generic payload has an arbitrary pattern of holes punched in it.

c) The byte enable mask may be defined by a small pattern applied repeatedly or by a large pattern covering the whole data array. See clause 9.13 Byte enable length attribute

k) The number of elements in the byte enable array shall be given by the byte enable length attribute.

d) The byte enable pointer may be set to 0, the null pointer, in which case byte enables shall not be used for the current transaction.

e) If byte enables are used, the byte enable pointer attribute shall be set by the initiator, the storage for the byte enable array shall be allocated by the initiator, the contents of the byte enable array shall be set by the initiator, and the contents of the byte enable array shall not be overwritten by any interconnect component or target.

f) If the byte enable pointer is non-null, the target shall either implement the semantics of the byte enable as defined below or shall generate a standard error response. The recommended response status is TLM_BYTE_ENABLE_ERROR_RESPONSE.

g) In the case of a write command, any interconnect component or target should ignore the values of any disabled bytes in the data array. It is recommended that disabled bytes have no effect on the behavior of any interconnect component or target. The initiator may set those bytes to any values, since they are going to be ignored.

h) In the case of a write command, when a target is doing a byte-by-byte copy from the transaction data array to a local array, the target should not modify the values of bytes in the local array corresponding to disabled bytes in the generic payload.

i) In the case of a read command, any interconnect component or target should not modify the values of disabled bytes in the data array. The initiator can assume that disabled bytes will not be modified by any interconnect component or target.

j) In the case of a read command, when a target is doing a byte-by-byte copy from a local array to the transaction data array, the target should ignore the values of bytes in the local array corresponding to disabled bytes in the generic payload.

k) If the application needs to violate these semantics for byte enables, or to violate any other semantics of the generic payload as defined in this document, the recommended approach would be to create a new transaction type derived from the generic payload. See clause 9.2.2 Use an empty class derived from the generic payload

l) In the case of the read command, instead of using byte enables, the target may be able to copy a contiguous block of bytes from the local array to the transaction data array, then mask the data later. This would place the responsibility for implementing the byte enables with the initiator instead of the target, and may or may not be safe because it causes an area of storage owned by the initiator to be entirely overwritten.

l) The default value of the byte enable pointer attribute shall be 0, the null pointer.

9.13. Byte enable length attribute

a) The method **set_byte_enable_length** shall set the byte enable length attribute to the value passed as an argument. The method **get_byte_enable_length** shall return the current value of the byte enable length attribute.

b) For a read command or a write command, the target shall interpret the byte enable length attribute as the number of elements in the bytes enable array.

c) The byte enable length attribute shall be set by the initiator, and shall not be overwritten by any interconnect component or target.

d) The byte enable to be applied to a given element of the data array shall be calculated using the formula **byte_enable_array_index = data_array_index % byte_enable_length**. In other words, the byte enable array is applied repeatedly to the data array.

e) The byte enable length attribute may be greater than the data length attribute, in which case any superfluous byte enables should not affect the behavior of a read or write command, but could be used by extensions.

f) If the byte enable pointer is 0, the null pointer, then the value of the byte enable length attribute shall be ignored by any interconnect component or target.

i) If the target is unable to execute the transaction with the given byte enable length, it shall generate a standard error response. The recommended response status is TLM_BYTE_ENABLE_ERROR_RESPONSE.

g) The default value of the byte enable length attribute shall be 1.

9.14. Streaming width[20] attribute

a) The method **set_streaming_width** shall set the streaming width attribute to the value passed as an argument. The method **get_streaming_width** shall return the current value of the streaming width attribute.

b) For a read command or a write command, the target shall interpret and act upon the current value of the streaming width attribute

c) Streaming affects the way a component should interpret the data array. A stream consists of a sequence of data transfers occurring on successive notional beats. The streaming width attribute shall determine the width of the stream, that is, the number of bytes transferred on each beat.

d) The address to or from which each byte is being copied in the target shall be reset to the value of the address attribute at the start of each beat.

e) A byte at a given index in the data array shall be transferred to or from the address given by the formula **address_attribute + (array_index % streaming_width)**

f) A streaming width of 0 shall be interpreted in the same way as a streaming width greater than or equal to the value of the data length attribute, that is, the data transfer is a normal transfer, not a streaming transfer.

g) The value of the streaming width attribute shall have no affect on the length of the data array or the number of bytes stored in the data array.

[20] The name streaming width is only provisional

j) If the target is unable to execute the transaction with the given streaming width, it shall generate a standard error response. The recommended response status is TLM_BURST_ERROR_RESPONSE.

h) Streaming may be used in conjunction with byte enables, in which case the streaming width would typically be equal to the byte enable length. It would also make sense to have the streaming width a multiple of the byte enable length. Having the byte enable length a multiple of the streaming width would imply that different bytes were enabled on each beat.

i) The default value of the streaming width attribute shall be 0.

9.15. Lock attribute

(The lock attribute is under consideration for future inclusion. The TLM Working Group considers that the inclusion of a standard locking mechanism for the generic payload is very important for interoperability.)

9.16. DMI allowed attribute

a) The method **set_dmi_allowed** shall set the DMI allowed attribute to the value passed as an argument. The method **get_dmi_allowed** shall return the current value of the DMI allowed attribute.

b) The DMI allowed attribute provides a hint to an initiator that it may try to obtain a direct memory pointer. The target should set this attribute to **true** if the transaction at hand could have been done through DMI. See clause 4.3.6 Optimization using a DMI Hint

h) The default value of the DMI allowed attribute shall be **false**.

9.17. Response status attribute

a) The method **set_response_status** shall set the response status attribute to the value passed as an argument. The method **get_response_status** shall return the current value of the response status attribute.

b) The method **is_response_ok** shall return **true** if and only if the current value of the response status attribute is TLM_OK_RESPONSE. The method **is_response_error** shall return **true** if and only if the current value of the response status attribute is not equal to TLM_OK_RESPONSE.

c) The method **get_response_string** shall return the current value of the response status attribute as a text string.

d) The response status attribute shall be set to TLM_INCOMPLETE_RESPONSE by the initiator, and may be overwritten by the target. The response status attribute should not be overwritten by any interconnect component, because the default value TLM_INCOMPLETE_RESPONSE indicates that the transaction was not delivered to the target.

e) The target may set the response status attribute to TLM_OK_RESPONSE to indicate that it was able to execute the command successfully, or to one of the five error responses listed in the table below to indicate an error. The target should choose the appropriate error response depending on the cause of the error.

Error response	Interpretation
TLM_ADDRESS_ERROR_RESPONSE	Unable to act upon the address attribute, or address out-of-range
TLM_COMMAND_ERROR_RESPONSE	Unable to execute the command
TLM_BURST_ERROR_RESPONSE	Unable to act upon the data length or streaming width
TLM_BYTE_ENABLE_ERROR_RESPONSE	Unable to act upon the byte enable
TLM_GENERIC_ERROR_RESPONSE	Any other error

f) If a target detects an error but is unable to select a specific error response, it may set the response status to TLM_GENERIC_ERROR_RESPONSE.

g) The default value of the response status attribute shall be TLM_INCOMPLETE_RESPONSE.

h) It is recommended that the initiator should always check the response status attribute after calling the method of a transport interface. An initiator *may* choose to ignore the response status if it is known in advance that the value will be TLM_OK_RESPONSE, perhaps because it is known in advance that the initiator is only connected to targets that always return TLM_OK_RESPONSE, but in general this will not be the case. In other words, the initiator ignores the response status at its own risk.

9.17.1. The standard error response

When a target receives a generic payload transaction, the target should perform one and only one of the following actions:

a) Execute the command represented by the transaction, honoring the semantics of the generic payload attributes, and honoring the publicly documented semantics of the component being modeled, and set the response status to TLM_OK_RESPONSE.

b) Set the response status attribute of the generic payload to one of the five error responses as described above.

c) Generate a report using the standard SystemC report handler with any of the four standard SystemC severity levels indicating that the command has failed or been ignored, and set the response status to TLM_OK_RESPONSE.

It is recommended that the target should perform exactly one of these actions, but an implementation is not obliged or permitted to enforce this recommendation.

It is recommended that a target for a transaction type other than the generic payload should follow this same principle, that is, execute the command as expected, or generate an error response using an attribute of the transaction, or generate a SystemC report. However, the details of the semantics and the error response mechanism for such a transaction are outside the scope of this standard.

The conditions for satisfying point a) above are determined by the expected behavior of the target component as would be visible to a user of that component. The attributes of the generic payload have defined semantics which correspond to conventional usage in the context of memory-mapped busses, but which do not necessarily assume that the target behaves as a random-access memory. There are many subtle corner cases. For example:

i. A target may have a memory-mapped register that supports both read and write commands, but the write command is non-sticky, that is, write modifies the state of the target, but a write followed by read will not return the data just written but some other value determined by the state of the target. If this is the normal expected behavior of the component, it is covered by point a).

ii. A target may implement the write command to set a bit whilst totally ignore the value of the data attribute. If this is the normal expected behavior of the target, it is covered by point a)

iii. A read-only memory may ignore the write command without signalling an error to the initiator using the response status attribute. Since the write command is not changing the state of the target but is being ignored altogether, the target should at least generate a SystemC report with severity SC_INFO or SC_WARNING.

iv. A target should not under any circumstances implement the write command by performing a read, or vice versa. That would be a fundamental violation of the semantics of the generic payload.

v. A target may implement the read command according to the intent of the generic payload, but with additional side-effects. This is covered by point a).

vi. A target with a set of memory-mapped registers forming an addressable register file receives a write command with an out-of-range address. The target should either set the response status attribute of the transaction to TLM_ADDRESS_ERROR_RESPONSE or generate a SystemC report.

vii. A passive simulation bus monitor target receives a transaction with an address that is outside the physical range of the bus being modeled. The target may log the erroneous transaction for post-processing under point a) and not generate an error response under points b) or c). Alternatively, the target may generate a report under point c).

In other words, the distinction between points a), b) and c) is ultimately a pragmatic judgement to be made on a case-by-case basis, but the definitive rule for the generic payload is that a target should always perform exactly one of these actions.

9.18. Extension mechanism

9.18.1. Introduction

The extension mechanism is an integral part of the generic payload, and cannot be used separately from the generic payload. Its purpose is to permit attributes to be added to the generic payload.

Extensions can be ignorable or mandatory. An ignorable extension is an extension that may be ignored by any or all interconnect components or targets that receive the generic payload transaction. The main intent of ignorable extensions is to model auxiliary information, simulation artefacts, side-band information, or meta-data that do not have a direct effect on the functionality of the downstream components. A mandatory extension is an extension that any interconnect component or target receiving the transaction is obliged to inspect and to act upon. The main intent of mandatory extension is for use when specializing the generic payload to model the specific details of a protocol.

9.18.2. Rationale

The rationale behind the extension mechanism is to permit TLM ports or sockets that carry variations on the core attribute set of the generic payload to be specialized with the same transaction type, thus allowing them to be bound together directly with no need for adaption or bridging. Without the extension mechanism, the addition of any new attribute to the generic payload would require the definition of a new protocol class, leading to a new template specialization of the core interface class, which would be type-incompatible with the generic payload and with any other such specialization. The extension mechanism allows minor variations to be introduced into the generic payload without breaking the type compatibility of TLM ports, thus reducing the amount of coding work that needs to be done to connect ports that carry slightly different information.

9.18.3. Rules

a) An extension can be added by an initiator, interconnect or target component. In particular, the creation of extensions is not restricted to initiators.

b) Any number of extensions may be added to each instance of the generic payload.

c) In the case of an ignorable extension, it is recommended that any interconnect or target component should be free to ignore the given extension, but this cannot and should not be enforced by the implementation. Having an interconnect or target component generate a standard error response because of the absence of an extension is possible, but is not recommended practice.

d) In the case of an ignorable extension, it is recommended that the presence or absence of a given extension should have no effect on the primary functionality of any component, but may, for example, have an effect on diagnostic reporting, debug, or optimization.

e) There is no built-in mechanism to enforce the presence of a given extension.

f) The semantics of each extension are application-defined. There are no pre-defined extensions.

g) An extension shall be created by deriving a user-defined class from the class **tlm_extension**, passing the name of the user-defined class itself as a template argument to **tlm_extension**, then creating an object of that class. The user-defined extension class may include members which represent extended attributes of the generic payload.

h) The pure virtual function **clone** of class **tlm_extension** shall be defined in the user-defined extension class to clone the extension object, including any extended attributes. This **clone** method is intended for use by the **clone** method of the generic payload. It shall create a complete and deep copy of any extension object such that the copy can survive the destruction of the original object with no visible side-effects.

i) The act of instantiating the class template **tlm_extension** shall cause the public data member **ID** to be initialized, and this shall have the effect of registering the given extension with the generic payload object and assigning a unique ID to the extension. The ID shall be unique across the whole executing program.

j) The generic payload shall behave as if it stored the extensions in a re-sizable array, where the ID of the extension gives the index of the extension in the array. Registering the extension with the generic payload shall reserve an array index for that extension. Each generic payload object shall contain an array capable of storing every extension registered in the currently executing program.

k) The methods **set_extension**, **get_extension**, and **clear_extension** each have two definitions, one a function template, and the other with an ID argument. The function template determines which extension to set, get or

clear using the template argument. The non-templated function determines which extension to set, get, or clear using the ID argument.

l) The function template forms of these three methods should generally be used. The functions with an ID argument are intended for low-level programming such as when cloning a generic payload object.

m) The member function **template<typename T> void set_extension(T*)** shall add the extension object given by the argument to the generic payload object. The argument shall be a pointer to a registered extension.

n) The member function **void set_extension(unsigned int, tlm_extension_base*)** shall add the extension object given by the second argument to the generic payload object at the array index given by the first argument. The given index shall have been registered as an extension ID, otherwise the behavior of the function is undefined.

o) The member function **template<typename T> void get_extension(T*&)** shall return a pointer to the extension object of the given type, if it exists, or a null pointer if it does not exist. The argument shall be a pointer to an object of a type derived from **tlm_extension**. It is not an error to attempt to retrieve a non-existent extension using this function template.

p) The member function **tlm_extension_base* get_extension(unsigned int)** shall return a pointer to the extension object with the ID given by the argument. The given index shall have been registered as an extension ID, otherwise the behavior of the function is undefined.

q) The member function **template<typename T> void clear_extension(const T*)** shall remove the extension given by the argument from the generic payload object. The argument shall be a pointer to an object of a type derived from **tlm_extension**.

r) The member function **void clear_extension(unsigned int)** shall remove from the generic payload object the extension at the array index given by the argument. The given index shall have been registered as an extension ID, otherwise the behavior of the function is undefined.

s) Each generic payload transaction should allocate sufficient space to store every registered extension. This can be achieved in one of two ways, either by constructing the transaction object *after* C++ static initialization, or by calling the method **resize_extensions** *after* static initialization but *before* using the transaction object for the first time. In the former case, it is the responsibility of the generic payload constructor to set the size of the extension array. In the latter case, it is the responsibility of the application to call **resize_extensions** before accessing the extensions for the first time.

t) The method **resize_extensions** shall increase the size of the extensions array in the generic payload to accommodate every registered extension.

10. Core TLM1 interfaces

The following core interfaces from TLM1 are still part of the TLM2.0-draft-2 standard, but are not documented in detail here. The transport method with the signature **transport(const REQ& , RSP&)** was not part of TLM1.0, but has been added in TLM2.0.

```
namespace tlm {

// Bidirectional blocking interfaces
template < typename REQ , typename RSP >
class tlm_transport_if : public virtual sc_core::sc_interface
{
public:
  virtual RSP transport( const REQ& ) = 0;
  virtual void transport( const REQ& req , RSP& rsp ) { rsp = transport( req ); }
};

// Uni-directional blocking interfaces
template < typename T >
class tlm_blocking_get_if : public virtual sc_core::sc_interface
{
public:
  virtual T get( tlm_tag<T> *t = 0 ) = 0;
  virtual void get( T &t ) { t = get(); }
};

template < typename T >
class tlm_blocking_put_if : public virtual sc_core::sc_interface
{
public:
  virtual void put( const T &t ) = 0;
};

// Uni-directional non blocking interfaces
template < typename T >
class tlm_nonblocking_get_if : public virtual sc_core::sc_interface
{
public:
  virtual bool nb_get( T &t ) = 0;
  virtual bool nb_can_get( tlm_tag<T> *t = 0 ) const = 0;
  virtual const sc_core::sc_event &ok_to_get( tlm_tag<T> *t = 0 ) const = 0;
};

template < typename T >
class tlm_nonblocking_put_if : public virtual sc_core::sc_interface
```



```
{
public:
  virtual bool nb_put( const T &t ) = 0;
  virtual bool nb_can_put( tlm_tag<T> *t = 0 ) const = 0;
  virtual const sc_core::sc_event &ok_to_put( tlm_tag<T> *t = 0 ) const = 0;
};

// Combined uni-directional blocking and non blocking
template < typename T >
class tlm_get_if :
  public virtual tlm_blocking_get_if< T > ,
  public virtual tlm_nonblocking_get_if< T > {};

template < typename T >
class tlm_put_if :
  public virtual tlm_blocking_put_if< T > ,
  public virtual tlm_nonblocking_put_if< T > {};

// Peek interfaces
template < typename T >
class tlm_blocking_peek_if : public virtual sc_core::sc_interface
{
public:
  virtual T peek( tlm_tag<T> *t = 0 ) const = 0;
  virtual void peek( T &t ) const { t = peek(); }
};

template < typename T >
class tlm_nonblocking_peek_if : public virtual sc_core::sc_interface
{
public:
  virtual bool nb_peek( T &t ) const = 0;
  virtual bool nb_can_peek( tlm_tag<T> *t = 0 ) const = 0;
  virtual const sc_core::sc_event &ok_to_peek( tlm_tag<T> *t = 0 ) const = 0;
};

template < typename T >
class tlm_peek_if :
  public virtual tlm_blocking_peek_if< T > ,
  public virtual tlm_nonblocking_peek_if< T > {};

// Get_peek interfaces
template < typename T >
class tlm_blocking_get_peek_if :
  public virtual tlm_blocking_get_if<T> ,
  public virtual tlm_blocking_peek_if<T> {};
```



```
template < typename T >
class tlm_nonblocking_get_peek_if :
  public virtual tlm_nonblocking_get_if<T> ,
  public virtual tlm_nonblocking_peek_if<T> {};

template < typename T >
class tlm_get_peek_if :
  public virtual tlm_get_if<T> ,
  public virtual tlm_peek_if<T> ,
  public virtual tlm_blocking_get_peek_if<T> ,
  public virtual tlm_nonblocking_get_peek_if<T>
  {};

} // namespace tlm
```


11. Tutorial[21]

(This section is to be completed.)

11.1.1. Extensions

Let us suppose you wish to add an extension contain the time when the transaction was created to a generic payload transaction. The first step is to create a user-defined extension class:

```
class my_extension:  public tlm::tlm_extension<my_extension>
{
public:
   my_extension(const sc_time& t) : start_time (t) {}
   virtual tlm_extension_base* clone() const {
      tlm_extension_base* t = new my_extension( this.start_time );
      return t;
   }
   sc_time start_time;
};
```

Since **tlm_extension** is a class template, the above code will have the effect of instantiating that template, and thus registering my_extension as an extension.

Next, create a generic payload transaction. As noted in the rules, this transaction object should be created as a stack (automatic) or heap (new) object. Otherwise, you would need to call **resize_extensions**.

```
tlm::tlm_generic_payload* m_gp;
m_gp = new tlm::tlm_generic_payload;
```

Now you can create an object to represent the extension, initializing the object using the current simulation time, and add it to the transaction.

```
my_extension* ext_object;
ext_object = new my_extension( sc_time_stamp() );
m_gp->set_extension( ext_object );
```

[21] To include a detailed tutorial walking through the use of the main classes, and lots of example code fragments.

12. Glossary

Blue = taken from the SystemC LRM

This glossary contains brief, informal descriptions for a number of terms and phrases used in this standard. Where appropriate, the complete, formal definition of each term or phrase is given in the main body of the standard. Each glossary entry contains either the clause number of the definition in the main body of the standard or an indication that the term is defined in ISO/IEC 14882:2003 or IEEE Std 1666™-2005.

adapter: A module that connects a transaction level interface to a pin level interface (in the general sense of the word interface) or that connects together two transaction level interfaces, often at different abstraction levels. Typically, an adapter is used to convert between two transaction-level interfaces of different types. See *transactor*.

approximately timed: A modeling style for which there exists a one-to-one mapping between the externally observable states of the model and the states of some corresponding detailed reference model such that the mapping preserves the sequence of state transitions but not their precise timing. The degree of timing accuracy is undefined. See *cycle approximate*.

attribute (of a transaction): Data that is part of and carried with the transaction and is implemented as a member of the transaction object. These may include attributes inherent in the bus or protocol being modeled, and attributes that are artefacts of the simulation model (a timestamp, for example).

backward path: The calling path by which a target or interconnect component makes interface method calls back in the direction of another interconnect component or the initiator.

bidirectional interface: A TLM 1.0 transaction level interface in which a pair of transaction objects, the request and the response, are passed in opposite directions, each being passed according to the rules of the unidirectional interface. For each transaction object, the transaction attributes are strictly readonly in the period between the first timing point and the end of the transaction lifetime.

blocking: Permitted to call the **wait** method. A blocking function may consume simulation time or perform a context switch, and therefore shall not be called from a method process. A blocking interface defines only blocking functions.

blocking transport interface: A blocking interface of the TLM2 standard which contains a single method **b_transport**. Confusingly, there is also still a blocking transport interface left as a legacy from TLM1 which contains a method named **transport**.

bridge: A module that connects together two similar or dissimilar transaction-level interfaces, each representing a memory-mapped bus or other protocol, usually at the same abstraction level. A bus bridge is a device that connects two similar or dissimilar buses together. A communication bridge is a device that connects network segments on the data link layer of a network. See *transactor*.

caller: In a function call, the sequence of statements from which the given function is called. The referent of the term may be a function, a process, or a module. This term is used in preference to *initiator* to refer to the caller of a function as opposed to the initiator of a transaction.

callee: In a function call, the function that is called by the caller. This term is used in preference to *target* to refer to the function body as opposed to the target of a transaction.

channel: A class that implements one or more interfaces or an instance of such a class. A channel may be a hierarchical channel or a primitive channel or, if neither of these, it is strongly recommended that a channel at least

be derived from class **sc_object**. Channels serve to encapsulate the definition of a communication mechanism or protocol. (SystemC term)

core interface: One of the [N] specific transaction level interfaces defined in this standard, not derived from any other interface, and with a template parameter representing a transaction type,. Each core interface is an *interface proper*. The core interfaces are distinct from the generic payload API.

cycle accurate: A modeling style in which it is possible to predict the state of the model in any given cycle at the external boundary of the model and thus to establish a one-to-one correspondence between the states of the model and the externally observable states of a corresponding RTL model in each cycle, but which is not required to explicitly re-evaluate the state of the entire model in every cycle or to explicitly represent the state of every boundary pin or internal register. This term is only applicable to models that have a notion of cycles.

cycle approximate: A model for which there exists a one-to-one mapping between the externally observable states of the model and the states of some corresponding cycle accurate model such that the mapping preserves the sequence of state transitions but not their precise timing. The degree of timing accuracy is undefined. This term is only applicable to models that have a notion of cycles.

cycle count accurate, cycle count accurate at transaction boundaries: A modeling style in which it is possible to establish a one-to-one correspondence between the states of the model and the externally observable states of a corresponding RTL model as sampled at the timing points marking the boundaries of a transaction. A cycle count accurate model is not required to be cycle accurate in every cycle, but is required to accurately predict both the functional state and the number of cycles at certain key timing points as defined by the boundaries of the transactions through which the model communicates with other models.

declaration: A C++ language construct that introduces a name into a C++ program and specifies how the C++ compiler is to interpret that name. Not all declarations are definitions. For example, a class declaration specifies the name of the class but not the class members, while a function declaration specifies the function parameters but not the function body. (See definition.) (C++ term)

definition: The complete specification of a variable, function, type, or template. For example, a class definition specifies the class name and the class members, and a function definition specifies the function parameters and the function body. (See declaration.) (C++ term)

forward path: The calling path by which an initiator or interconnect component makes interface method calls forward in the direction of another interconnect component or the target.

generic payload: A specific set of transaction attributes and their semantics together defining a transaction level protocol which may be used to achieve a degree of interoperability between untimed, loosely timed and approximately timed models for components communicating over a memory-mapped bus. The same transaction class is used for all modeling styles.

generic payload API: The class interface through which the generic payload is accessed. The generic payload API is distinct from the core interfaces.

initiator: A module that can initiate transactions. The initiator is responsible for initializing the state of the transaction object, and for deleting or reusing the transaction object at the end of the transaction's lifetime. An initiator is usually a master and a master an initiator, but the term *initiator* means that a component can initiate transactions, whereas the term *master* means that a component can take control of a bus. In the case of the TLM 1.0 interfaces, the term *initiator* as defined here may not be strictly applicable, so the terms *caller* and *callee* may be used instead for clarity.

initiator socket: A class containing a port for interface method calls on the forward path and an export for interface method calls on the backward path. A socket also overloads the SystemC binding operators to bind both port and export.

interconnect component: A module that accesses a transaction object, but does act as an initiator or a target with respect to that transaction. An interconnect component may or may not be permitted to modify the attributes of the transaction object, depending on the rules of the payload. An arbiter or a router would typically be modeled as an interconnect component, the alternative being to model it as a target for one transaction and an initiator for a separate transaction.

interface: A class derived from class **sc_interface**. An interface proper is an interface, and in the object-oriented sense a channel is also an interface. However, a channel is not an interface proper. (SystemC term)

Interface Method Call (IMC): A call to an interface method. An interface method is a member function declared within an interface. The IMC paradigm provides a level of indirection between a method call and the implementation of the method within a channel such that one channel can be substituted with another without affecting the caller. (SystemC term)

interface proper: An abstract class derived from class **sc_interface** but not derived from class **sc_object**. An interface proper declares the set of methods to be implemented within a channel and to be called through a port. An interface proper contains pure virtual function declarations, but typically contains no function definitions and no data members. (SystemC term)

interoperability: The ability of two or more transaction level models from diverse sources to exchange information using the interfaces defined in this standard. The intent is that models that implement common memory-mapped bus protocols in the programmers view use case should be interoperable without the need for explicit adapters. Furthermore, the intent is to reduce the amount of engineering effort needed to achieve interoperability for models of divergent protocols or use cases, although it is expected that adapters will be required in general.

lifetime (of an object): The lifetime of an object starts when storage is allocated and the constructor call has completed, if any. The lifetime of an object ends when storage is released or immediately before the destructor is called, if any. (C++ term)

lifetime (of a transaction): The period of time that starts when the transaction becomes valid and ends when the transaction becomes invalid. Because it is possible to pool or re-use transaction objects, the lifetime of a transaction object may be longer than the lifetime of the corresponding transaction. For example, a transaction object could be a stack variable passed as an argument to multiple *put* calls of the TLM1 interface.

loosely timed: A modeling style that represents minimal timing information sufficient only to support features necessary to boot an operating system and to manage multiple threads in the absence of explicit synchronization between those threads. A loosely timed model may include timer models and a notional arbitration interval or execution slot length. Some users adopt the practice of inserting random delays into loosely timed descriptions in order to test the robustness of their protocols, but this practice does not change the basic characteristics of the modeling style.

master: This term has no precise technical definition in this standard, but is used to mean a module or port that can take control of a memory-mapped bus in order to initiate bus traffic, or a component that can execute an autonomous software thread and thus initiate other system activity. Generally, a bus master would be an initiator.

method: A function that implements the behavior of a class. This term is synonymous with the C++ term *member function*. In SystemC, the term *method* is used in the context of an *interface method call*. Throughout this standard,

the term *member function* is used when defining C++ classes (for conformance to the C++ standard), and the term *method* is used in more informal contexts and when discussing interface method calls. (SystemC term)

non-blocking: Not permitted to call the **wait** method. A non-blocking function is guaranteed to return without consuming simulation time or performing a context switch, and therefore may be called from a thread process or from a method process. A non-blocking interface defines only non-blocking functions.

non-blocking transport interface: A non-blocking interface of the TLM2 standard which contains a single method **nb_transport**.

object: A region of storage. Every object has a type and a lifetime. An object created by a definition has a name, whereas an object created by a new expression is anonymous. (C++ term)

payload event queue: A class that maintains a queue of SystemC event notifications, where each notification carries an associated transaction object. Transactions are put into the queue annotated with a delay, and each transaction pops out of the back of queue at the time it was put in plus the given delay. Useful when combining the non-blocking interface with the approximately-timed coding style.

phase: The period in the lifetime of a transaction occurring between successive timing points. The phase is passed as an argument to the non-blocking transport method.

programmers view (PV): The use case of the software programmer who requires a functionally accurate, loosely timed model of the hardware platform for booting an operating system and running application software.

quantum: In temporal decoupling, the amount a process is permitted to run ahead of the current simulation time.

slave: This term has no precise technical definition in this standard, but is used to mean a reactive module or port on a memory-mapped bus that is able to respond to commands from bus masters, but is not able itself to initiate bus traffic. Generally, a slave would be modeled as a target.

socket: See initiator socket and target socket

standard error response: The behavior prescribed by this standard for a generic payload target that is unable to execute a transaction successfully. A target should either a) execute the transaction successfully or b) set the response status attribute to an error response or c) call the SystemC report handler.

synchronization-on-demand: An indication from the **nb_transport** method back to its caller that it was unwilling or unable to fulfil a request to effectively execute a transaction at a future time (temporal decoupling), and therefore that the caller must yield control back to the SystemC scheduler so that simulation time may advance and other processes run.

target: A module that represents the final destination of a transaction, able to respond to transactions generated by an initiator, but not itself able to initiate new transactions. For a write operation, data is copied from the initiator to one or more targets. For a read operation, data is copied from one target to the initiator. A target may read or modify the state of the transaction object. In the case of the TLM 1.0 interfaces, the term *target* as defined here may not be strictly applicable, so the terms *caller* and *callee* may be used instead for clarity.

target socket: A class containing a port for interface method calls on the backward path and an export for interface method calls on the forward path. A socket also overloads the SystemC binding operators to bind both port and export.

temporal decoupling: The ability to allow one or more initiators to run ahead of the current simulation time in order to reduce context switching and thus increase simulation speed.

timing point: A point in time at which the processes that are interacting through a transaction either transfer control or are synchronized. Certain timing points are implemented as function calls or returns, others as event notifications. Timing points mark the boundaries between the phases of a transaction.

TLM1: The first major version of the OSCI Transaction Level Modeling standard. TLM1.0 was released in 2005.

TLM2: The second major version of the OSCI Transaction Level Modeling standard. This document describes TLM2.0-draft-2.

transaction: An abstraction for an interaction or communication between two or more concurrent processes. A transaction carries a set of attributes and is bounded in time, meaning that the attributes are only valid within a specific time window. The timing associated with the transaction is limited to a specific set of timing points, depending on the type of the transaction. Processes may be permitted to read or modify attributes of the transaction, depending on the protocol.

transaction object: The object that stores the attributes associated with a transaction. The type of the transaction object is passed as a template argument to the core interfaces.

transaction level (TL): The abstraction level at which communication between concurrent processes is abstracted away from pin wiggling to transactions. This term does not imply any particular level of granularity with respect to the abstraction of time, structure, or behavior.

transaction level model, transaction level modeling (TLM): A model at the transaction level and the act of creating such a model, respectively. Transaction level models typically communicate using function calls, as opposed to the style of setting events on individual pins or nets as used by RTL models.

transactor: A module that connects a transaction level interface to a pin level interface (in the general sense of the word interface) or that connects together two or more transaction level interfaces, often at different abstraction levels. In the typical case, the first transaction level interface represents a memory mapped bus or other protocol, the second interface represents the implementation of that protocol at a lower abstraction level. However, a single transactor may have multiple transaction level or pin level interfaces. See *adapter, bridge*.

transport interface: The one and only bidirectional core interface in TLM1. The transport interface passes a request transaction object from caller to callee, and returns a response transaction object from callee to caller. TLM2 adds separate blocking and non-blocking transport interfaces.

unidirectional interface: A TLM 1 transaction level interface in which the attributes of the transaction object are strictly readonly in the period between the first timing point and the end of the transaction lifetime. Effectively, the information represented by the transaction object is strictly passed in one direction either from caller to callee or from callee to caller. In the case of **void put(const T& t)**, the first timing point is marked by the function call. In the case of **void get(T& t)**, the first timing point is marked by the return from the function. In the case of **T get()**, strictly speaking there are two separate transaction objects, and the return from the function marks the degenerate end-of-life of the first object and the first timing point of the second.

untimed: A modeling style in which there is no explicit mention of time or cycles, but which includes concurrency and sequencing of operations. In the absence of any explicit notion of time as such, the sequencing of operations across multiple concurrent threads must be accomplished using synchronization primitives such as events, mutexes and blocking FIFOs. Some users adopt the practice of inserting random delays into untimed descriptions in order to test the robustness of their protocols, but this practice does not change the basic characteristics of the modeling style.

valid: The state of an object returned from a function by pointer or by reference, during any period in which the object is not deleted and its value or behavior remains accessible to the application. (SystemC term)

yield: Return control to the SystemC scheduler. For a thread process, to yield is to call **wait**. For a method process, to yield is to return from the function.

