

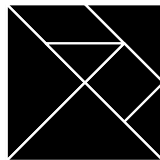
GSDLAB TECHNICAL REPORT

An algebraic semantics for bidirectional model synchronization

Zinovy Diskin

GSDLAB-TR 2014-04-01

August 2014



Generative Software
Development Lab



Generative Software Development Laboratory
University of Waterloo
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

WWW page: <http://gsd.uwaterloo.ca/>

The GSDLAB technical reports are published as a means to ensure timely dissemination of scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the authors or by other copyright holders, notwithstanding that they have offered their works here electronically. It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may not be reposted without the explicit permission of the copyright holder.

A Product Line of Lax Lenses.

Zinovy Diskin^{1,2}

Abstract

The goal of the present document is to support the taxonomy for bidirectional model synchronization developed in [1] with a formal semantics. The taxonomy is 3D so that each synchronization type is characterized by a triple of coordinates (x, y, z) , in which x classifies the *organizational symmetry* of the case, y is for the *informational symmetry*, and z is for *incrementality* of the update propagation operations. Different types of delta lenses (algebraic structures modeling bx) can be classified by points on the YZ -plane. As for the x -coordinate, it says, roughly, whether update propagation is uni- or bi-directional, but as it was shown in [1], there are several important refinements of the two-valued uni, bi-classification so that actually axis X has four rather than two points.

A formal semantics for the enriched X -classification seems to be an entirely novel aspect for the (delta) lens literature. There are also several contributions for the very delta lens framework within the plane YZ . First of all, we build a product line of delta lenses so that each concrete delta lens structure is characterizes by two parameters. To this end, we will build a framework in which an asymmetric delta lens appears as a special case of the symmetric one, which is in a sense dual to the Johnson and Rosebrugh construction, in which a symmetric lens is presented as a span of asymmetric lenses [2]. The second novelty is more essential: we present lax versions of major lens laws of compositionality (the infamous PutPut), and invertibility.

Keywords: Synchronization Taxonomy, Formal semantics, Model Synchronization, Model Transformation, Model Driven Engineering

1. Introduction

The goal of the present document is to support the taxonomy for bidirectional model synchronization developed in [1] with a formal semantics. The taxonomy is 3D so that each synchronization type is characterized by a triple of coordinates (x, y, z) , in which x classifies the *organizational symmetry* of the case, y is for the *informational symmetry*, and z is for *incrementality* of the update propagation operations. It is convenient to present the space as a product $X|timYZ$, where X is the org-symmetry axis and $YZ = Y \times Z$ is the plain of info-symmetry and incrementality, which actually classifies the computational framework

(= a system of update propagation operations) supporting the synchronization. Different types of delta lenses (algebraic structures modeling bx) can be classified by points on this plane. As for the x -coordinate of the cases placed in axis X , roughly, it says whether update propagation is uni- or bi-directional, but as it was shown in [1], there are several important refinements of the two-valued uni, bi-classification so that actually axis X has four rather than two points.

A formal semantics for the enriched X -classification seems to be an entirely novel aspect for the (delta) lens literature. There are also several contributions for the very delta lens framework within the plane YZ . First of all, we build a product line of delta lenses so that each concrete delta lens structure is characterizes by two parameters. To this end, we will build a framework in which an asymmetric delta lens appears as a special case of the

Email address: zdiskin@gsd.uwaterloo.ca (Zinovy Diskin)

¹University of Waterloo, Canada

²McMaster University, Canada

symmetric one, which is in a sense dual to the Johnson and Rosebrugh construction, in which a symmetric lens is presented as a span of asymmetric lenses [2]. The second novelty is more essential: we present lax versions of major lens laws of compositionality (the infamous Put-Put), and invertibility. The semantics is algebraic and essentially based on operations over models, updates, and intermodel correspondence mappings—all considered as abstract nodes and arrows.

We will define a family of algebraic structures using the following common pattern. Let the structure to be defined is S . We first define the carrier of S (as a rule, it will be a previously defined structure with, perhaps, new sets added), on which several new operations that S has to have are defined. To specify these operations, we give their arities, and, perhaps, some elementary technical equations (we say *laws*) they must satisfy (some of these laws will be skipped to save space). Then we introduce substantial laws capturing semantics of the operations, and say that S is *well-behaved* if its operations satisfy these laws. The presentation is necessarily abstract, although informal explanations and intuitive meaning of the constructs are provided. Concrete examples illustrating how these abstract constructs work can be found in papers [3, 4, 5]; some of the constructs were implemented with TGG [6, 7].

2. Formalizing Informational Symmetry: Alignment Frameworks with Consistency

2.1. Model spaces and alignment.

Definition 1 (Model spaces). A *model space* is a category \mathbf{M} , whose objects are called *models*, and morphisms (arrows) are *directed deltas* or *updates*. In detail, a model space is a directed graph $\mathbf{M} = (\mathbf{M}_\bullet, \mathbf{M}_\Delta, \mathbf{s}, \mathbf{t})$ with a set \mathbf{M}_\bullet of nodes called *models*, a set \mathbf{M}_Δ of arrows called *deltas*, and functions $\mathbf{s}: \mathbf{M}_\Delta \rightarrow \mathbf{M}_\bullet$, $\mathbf{t}: \mathbf{M}_\Delta \rightarrow \mathbf{M}_\bullet$, which assign a source and a target model to every delta. For a delta a with $\mathbf{s}(a) = A$ and $\mathbf{t}(a) = A'$, we write $a: A \rightarrow A'$ or else $a \in \mathbf{M}_\Delta(A, A')$, where $\mathbf{M}_\Delta(A, A')$ is the set of all deltas from A to A' .

Being a category means that the graph is endowed with two additional structures. First, arrows can be sequentially composed: for any pair $a_1: A \rightarrow A'$

and $a_2: A' \rightarrow A''$, there is defined their composition $a_1; a_2: A \rightarrow A''$, and the associativity law holds: $(a_1; a_2); a_3 = a_1; (a_2; a_3)$. Second, every model $A \in \mathbf{M}_\bullet$ is assigned with a special delta $\text{id}_A: A \rightarrow A$ called *identity*, and $\text{id}_A; a = a = a; \text{id}_{A'}$ for any $a: A \rightarrow A'$.

Intuitively, one may think about \mathbf{M}_\bullet as the class of all models conforming to a fixed but implicit metamodel. Deltas from \mathbf{M}_Δ can be thought of as either structural (mappings) or behavioral (edit sequences) specifications of updates; details and a thorough discussion can be found in [4]. Identity deltas can be seen as *idle* updates that do nothing. We will use terms delta and update interchangeably.

We will also assume the following *update (delta) division* condition. Given a pair of updates with a common source, $a_1: A \rightarrow A_1$, $a_2: A \rightarrow A_2$, let a_2/a_1 denotes the set $\{a: A_1 \rightarrow A_2: a_1; a = a_2\}$ of updates that continue update a_1 and result in update a_2 . We require this set to be non-empty for any pair (a_1, a_2) :

$$\text{(UDiv).} \quad a_2/a_1 \neq \emptyset$$

Definition 2 (Alignment frameworks). An *alignment framework* is given by the following data.

- (a) A pair of model spaces (\mathbf{M}, \mathbf{N}) called the *source* and the *target* space resp.
- (b) A set \mathbf{R} of *correspondence mappings* or just *corrs* between \mathbf{M} - and \mathbf{N} -models. That is, there are total functions $\mathbf{s}: \mathbf{M}_\bullet \leftarrow \mathbf{R}$ and $\mathbf{t}: \mathbf{R} \rightarrow \mathbf{N}_\bullet$, and we write $r: A \leftrightarrow B$ or $r \in \mathbf{R}(A, B)$ if $A = \mathbf{s}.r$ and $r.\mathbf{t} = B$. Note that we write the function symbol to the right or to the left of the argument to match the direction of function in our diagrams, in which space \mathbf{M} will be on the left, and space \mathbf{N} is on the right.

In our applications, a corr $r: A \leftrightarrow B$ is to be thought of as a set of bidirectional links from elements of model A to elements of model B so that if $e_1 \in A$ and $e_2 \in B$ are linked (then we could write $r(e_1, e_2)$), we consider them as two different representations of the same object. Moreover, one of the elements e_i , or even both, can be derived in its model by posing suitable queries against the model (see [8, 9] for details). However, in the present paper, elements of \mathbf{R} are just abstract entities called *corrs* and denoted by bidirectional arrows between models in different spaces. Similarly, updates are abstract entities denoted by arrows between models in the same space. In our

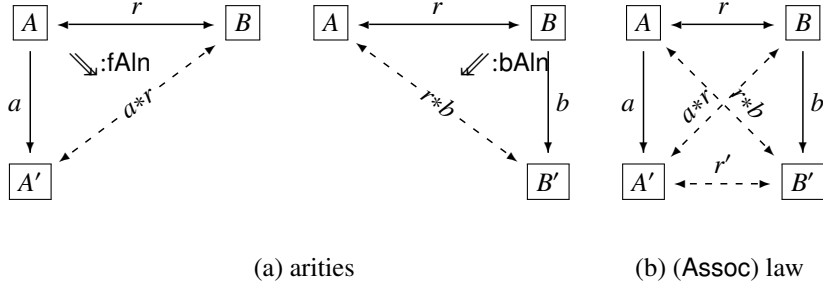


Figure 1: Alignment operations and their laws

diagrams, update arrows will be always vertical whereas corrs are horizontal.

(c) A pair of operations over corrs and updates, $fAln$ and $bAln$, which are called *forward* and *backward (re)alignment*. Their arities are shown in Fig. 1(a): operations' input nodes are framed, input arrows have solid bodies, and output arrows are dashed. We write $a * r$ for $fAln(a, r)$ and $r * b$ for $bAln(b, r)$.

We denote an alignment framework by a stroked arrow $R: M \leftrightarrow N$. Although the arrow is bidirectional to recall the symmetry of the notion, we still distinguish spaces M and N by their order in the line: the left space is the *source* space of the framework, and the right one is the *target* space.

Definition 3 (Well-behavedness). An alignment framework is called *well-behaved (wb)* if the following four laws are respected.

(a) Identity updates do not actually need realignment:

$$(IdAln) \quad idA * r = r = r * idB$$

for any corr $r: A \leftrightarrow B$.

(b) The result of applying a sequence of interleaving forward and backward alignments does not depend on the order of application as shown in Fig. 1(b):

$$(Assoc) \quad (a * r) * b = a * (r * b)$$

for any corr r and any updates a, b .

We will call diagrams like those shown in Fig. 1(a,b) *commutative* if the arrow at the respective operation output is indeed equal to that one computed by the operation. For example, diagram Fig. 1(b) is commutative if $r' = a * r * b$.

(c) Alignment is compositional: for any consecutive updates $a: A \rightarrow A'$, $a': A' \rightarrow A''$, $b: B \rightarrow B'$, $b': B' \rightarrow B''$, the following holds:

$$(AlnAln) \quad (a; a') * r = a' * (a * r) \text{ and } r * (b; b') = (r * b) * b'$$

(d) Corr division condition holds. Given a pair of corrs with a common source, $r_1: A \leftrightarrow B_1$, $r_2: A \leftrightarrow B_2$, let r_2/r_1 denotes the set $\{b: B_1 \rightarrow B_2: r_1 * b = r_2\}$ of updates that map the first corr into the second via realignment. Similarly, $r_2 \setminus r_1$ is $\{a: A_1 \rightarrow A_2: a * r_1 = r_2\}$ for a given pair of corrs $r_i: A_i \leftrightarrow B$. We require these sets to be non-empty:

$$(CDiv) \quad r_2/r_1 \neq \emptyset \neq r_2 \setminus r_1$$

We will assume all our alignment frameworks are well-behaved by default. \square

Remark 1 (Categorical treatment). It is easy to check that conditions (a,b,c) of well-behavedness make an alignment framework a functor $R: M \times N \rightarrow \mathbf{Set}$ or, equivalently, $R: (M^\circ) \times N \rightarrow \mathbf{Set}$ (where $_\circ$ denotes dualization of a category, i.e., inversion of all its arrows). Such functors are called *profunctors* or *distributors* [?][Benabou], and usually denoted by stroked arrows $R: M \rightrightarrows N^\circ$. However, in our context it is more natural and convenient to work in a more symmetric setting of functor $R: M \times N \rightarrow \mathbf{Set}$.

The following three conditions are equivalent:

- (i) $R: M \leftrightarrow N$ is an alignment framework,
- (ii) functor $R: M \times N \rightarrow \mathbf{Set}$ satisfies (CDiv),
- (iii) category $Col(R)$ (with Col the *collage functor* $Col: \mathbf{Dist} \rightarrow \mathbf{Cat}/2$) satisfies the arrow division condition for any pair of its arrows with a common source, i.e., conditions (UDiv) and (CDiv).

Definition 4 (Consistency and private updates). A *consistency framework* is an alignment framework endowed with two extra structures modeling intermodel consistency. Formally, it is a triple of the following components.

(a) An alignment framework $R: M \leftrightarrow N$ as defined above.

(b) A subset $K \subset R$ of corrs called *consistent*. In other words, for any pair of models $A \in M_\bullet$ and $B \in N_\bullet$ and the respective set of corrs $R(A, B)$, there is a defined a subset of *consistent* corrs $K(A, B) \subset R(A, B)$ between A and B (i.e., $K(A, B) = R(A, B) \cap K$). If a corr $r \in K(A, B)$, we say that models A and B are *consistent via r* and write $A \sim_r B$.

We do not exclude the cases when $K(A, B)$ has more than one corr, or is empty. In the latter case we write $A \# B$ and call the models (*entirely*) *inconsistent*. If, on the contrary, $K(A, B) \neq \emptyset$, we say that the models are *potentially consistent* and write $A \sim B$. For $A \in M_\bullet$, we write $A.K_\bullet$ for the set $\{B \in N_\bullet : A \sim B\}$. Similarly, $\bullet K.B$ denotes the set $\{A \in M_\bullet : A \sim B\}$ (where we again write a function symbol to the right or to the left of the argument to suggest the direction of function in our diagrams).

We assume the following *totality* condition:

(Total) $A.K_\bullet \neq \emptyset \neq \bullet K.B$ for all $A \in M_\bullet, B \in N_\bullet$.

That is, any model has at least one potentially consistent counterpart on the other side, and hence synchronization is always possible.

(c) Subspaces of *private updates* are given, $M^{\text{priv}} \subset M$ and $N^{\text{priv}} \subset N$, such that $M^{\text{priv}}_\bullet = M_\bullet$ and $N^{\text{priv}}_\bullet = N_\bullet$. That is, for any two models $A, A' \in M_\bullet$, there is a set (perhaps, empty) of updates $M_\Delta^{\text{priv}}(A, A') \subseteq M_\Delta(A, A')$ called *private*, composition of two private updates is private, idle updates are private, and divisions of private updates are private two: $a_2/a_1 \in M_\Delta^{\text{priv}}$ as soon as $a_1, a_2 \in M_\Delta^{\text{priv}}$. Likewise for the N side.

Non-private updates are called *public*. Thus, we have partitions: $M_\Delta = M_\Delta^{\text{priv}} \uplus M_\Delta^{\text{pub}}$ and $N_\Delta = N_\Delta^{\text{priv}} \uplus N_\Delta^{\text{pub}}$.

Definition 5 (Well-behavedness). Informally, a consistency framework is *well-behaved*, if (the underlying alignment framework is *wb* and) the two facets of consistency, consistent corrs and private updates, fit together so that private updates do not affect consistency. Formally, we require the following two laws to hold: (a) For any corr $r: A \leftrightarrow B$ and any private updates $a: A \rightarrow A', b: B \rightarrow B'$,

we have $a * r \in K$ iff $r \in K$ iff $r * b \in K$. (b) For any two consistent corrs with the same source, $r_1 \in K(A, B_1)$ and $r_2 \in K(A, B_2)$, we have $r_2/r_1 \in N_\Delta^{\text{priv}}$. That is, two consistent N -counterparts of an M -model A only differ privately (but share the same public part determined by A). Similarly, $r_2 \setminus r_1 \in M_\Delta^{\text{priv}}$ for any two corrs with the same target, $r_1 \in K(A_1, B)$ and $r_2 \in K(A_2, B)$. \square

To unify notation, we will sometimes use symbol K polymorphically for denoting both consistent corrs and private updates. That is, we write $K(X, Y)$ with $X, Y \in M_\bullet \cup N_\bullet$ meaning that

$$\begin{aligned} K(X, Y) &\stackrel{\text{def}}{=} M_\Delta^{\text{priv}}(X, Y) && \text{if both } X, Y \in M_\bullet, \\ K(X, Y) &\stackrel{\text{def}}{=} N_\Delta^{\text{priv}}(X, Y) && \text{if both } X, Y \in N_\bullet, \\ K(X, Y) &\stackrel{\text{def}}{=} K(X, Y) && \text{if } X \in M_\bullet \text{ and } Y \in N_\bullet, \\ K(X, Y) &= \perp \text{ (not defined)} && \text{if } X \in N_\bullet, Y \in M_\bullet. \end{aligned}^3$$

Then we can say that a consistency framework is a pair (R, K) with R an alignment framework and K a polymorphic function as above, which satisfy all necessary conditions. We will denote a consistency framework as $K: M \leftrightarrow N$ without explicit mentioning the underlying alignment framework $R: M \leftrightarrow N$.

Remark 2 (Categorical treatment). The polymorphic use of K described above is a known categorical construction of presenting a distributor by its *collage* category, actually a barrel. It can be proven that functor $\text{Col}: \mathbf{Dist} \rightarrow \mathbf{Cat}/2$ is an equivalence of categories (see Joyal's Catlab; <http://nlab.mathforge.org/joyalcatlab/>)

If to follow Leibniz's (formal) vs. Newtonian (physical) style of building mathematical models, then the definitions should ensure that K is a sub-profunctor with division of R . Correspondingly, category $\text{Col}(K)$ is a subcategory of $\text{Col}(R)$, which inherits the division property.

2.2. Informational asymmetry

Definition 6 (Info-asymmetry). Let $K: M \leftrightarrow N$ be a wb consistency framework. We say that space M is *dominated by N via K* , and write $M \leq_K N$ (or $N \geq_K M$), if the only private updates on the M -side are identities.

Proposition 1. Let $M \leq_K N$. Then for any $B \in N_\bullet$, there is a uniquely defined model $A \in M_\bullet$ called the (*abstract*) *view of B* , such that (a) $K(A, B)$ is a singleton whereas (b) $K(A', B) = \emptyset$ for any $A' \neq A$.

Proof. Let $r \in K(A, B)$ and $r' \in K(A', B)$. Then there is a private update $a_{r,r'} : A \rightarrow A'$ such that $a_{r,r'} * r = r'$ by the definition of consistency framework. However, $M \leq_K N$ implies that a is identity, i.e., $A = A'$, and $r = r'$ by the IdAln law. \square

Corollary 1. If $M \leq_K N$, then function $f^K : M_\bullet \leftarrow N_\bullet$ is defined such that (a) set $K(f^K.B, B)$ is a singleton for any model $B \in N_\bullet$, while (b) $K(A', B) = \emptyset$ for any $A' \neq f^K.B$.

Model $f^K.B$ is called the *view* of model B determined by consistency framework K , and f^K is the view computation function; following the terminological tradition of lenses, we call this function *get-the-view* and denote it by get_\bullet^K . Condition (b) formalizes the requirement that any non-idle view update makes the view inconsistent with its source. Condition (a) states that the view is given with its unique traceability mapping — the unique element of set $K(\text{get}_\bullet^K.B, B)$, which we will denote by $\text{get}_{\leftrightarrow}^K.B$. As a rule, we will omit superindex K if it is clear from the context. Note also that function get_\bullet is surjective due to Totality condition in Definition 4.

Proposition 2. If $M \leq_K N$ and $N \leq_K M$, then the respective view computation functions $\text{get}_{\bullet,1} : M_\bullet \leftarrow N_\bullet$ and $\text{get}_{\bullet,2} : M_\bullet \rightarrow N_\bullet$ are mutually inverse: $(\text{get}_{\bullet,1}.B).\text{get}_{\bullet,2} = B$ for any $B \in N_\bullet$, and $A = \text{get}_{\bullet,1}.(A.\text{get}_{\bullet,2})$ for any $A \in M_\bullet$. Hence, both get_\bullet -functions are bijections. Conversely, if $M \leq_K N$ and the respective $\text{get}_{\bullet,1} : M_\bullet \leftarrow N_\bullet$ is bijective, then $N \leq_K M$ and the respective $\text{get}_{\bullet,2} : M_\bullet \rightarrow N_\bullet$ is the inverse of $\text{get}_{\bullet,1}$. \square

2.3. Info-symmetry types

Our goal is to specify a set of properties of consistency frameworks, which would set their (complete and disjoint) taxonomy. We call these properties *info-symmetry types*. We will begin with setting some general terminology about types, then consider dualization, and finish this subsection with a formal definition of the taxonomy.

Let \mathbf{CFwk} be the class of all consistency frameworks. If π is a property of consistency frameworks, then we write $K \models \pi$ to say that a framework K satisfies the property, and we thus have a set $\llbracket \pi \rrbracket = \{K \in \mathbf{CFwk} : K \models \pi\}$ of all frameworks satisfying the property. If $K \in \llbracket \pi \rrbracket$, i.e., $K \models \pi$, we will say that framework K is an *instance of type* π , or K is *of type* π . Thus, we loosely use the term *type* both syntactically — as a synonym for the term *property*,

and semantically — to refer to the corresponding set of instances.

Having a set of types/properties $\{\pi_1 \dots \pi_n\}$, we can form a type $\pi = \pi_1 \vee \dots \vee \pi_n$ so that $\llbracket \pi \rrbracket = \llbracket \pi_1 \rrbracket \cup \dots \cup \llbracket \pi_n \rrbracket$. Then we call type π *abstract* because instantiation of this type means instantiation of one of the subtypes $\llbracket \pi_i \rrbracket$: $K \in \llbracket \pi \rrbracket$ iff $K \in \llbracket \pi_i \rrbracket$ for some i . We call a set of types $\{\pi_1 \dots \pi_n\}$ (*taxonomically*) *complete* if $\mathbf{CFwk} = \llbracket \pi_1 \rrbracket \cup \dots \cup \llbracket \pi_n \rrbracket$. We call types *disjoint* if $K \models \pi_i$ implies $K \not\models \pi_j$ for all $j \neq i$; then sets $\llbracket \pi_i \rrbracket$ are disjoint.

Two types are logically equivalent, $\pi_1 \Leftrightarrow \pi_2$, iff they have the same extension $\llbracket \pi_1 \rrbracket = \llbracket \pi_2 \rrbracket$. \square

For considering informational and organizational symmetries, the following notion will be central.

Definition 7 (Dualization). Any consistency framework $K : M \Leftrightarrow N$ determines its dual framework $K^\circ : N \Leftrightarrow M$ in the following way. We first dualize the underlying alignment framework: (i) the source space is N and the target space is M ; (ii) for any corr $r, r.s^\circ \stackrel{\text{def}}{=} r.t$ and $t^\circ.r \stackrel{\text{def}}{=} s.r$; (iii) operations are $\text{fAln}_{R^\circ}(b, r) \stackrel{\text{def}}{=} \text{bAln}_R(b, r)$ and $\text{bAln}_{R^\circ}(a, r) \stackrel{\text{def}}{=} \text{fAln}_R(a, r)$. Then we dualize the consistency structure: $K^\circ(B, A) \stackrel{\text{def}}{=} K(A, B)$, $K^\circ(B, B') = K(B, B')$, $K^\circ(A, A') \stackrel{\text{def}}{=} K(A, A') = M_\Delta^{\text{prv}}(A, A')$.

Definition 8 (Symmetric and asymmetric types). Any type/property π gives rise to a *dual type* π° such that $K \models \pi^\circ$ iff $K^\circ \models \pi$. We denote type \leq_K° (see Definition 6) by \geq_K , and write $N \geq_K M$ for $(K : M \Leftrightarrow N) \models \leq^\circ$. A type π is called *almost concrete* if $\pi \Leftrightarrow \pi_1 \vee \pi_1^\circ$ (and hence $\llbracket \pi \rrbracket = \llbracket \pi_1 \rrbracket \cup \llbracket \pi_1^\circ \rrbracket$) for some concrete type π_1 . A type is called *symmetric* if $\pi \Leftrightarrow \pi^\circ$. We call types π_1, π_2 *mutually dual* if $\pi_1 \Leftrightarrow \pi_2^\circ$ (and hence $\pi_1^\circ \Leftrightarrow \pi_2$).

Proposition 3. For any $K : M \Leftrightarrow N$, $(K^\circ)^\circ = K$. For any type π , $(\pi^\circ)^\circ \Leftrightarrow \pi$.

Proposition 4 (Dualization and consistency). A consistency framework $K : M \Leftrightarrow N$ is well-behaved(wb) iff its dual $K^\circ : N \Leftrightarrow M$ is wb too. Thus, the notion of consistency is symmetric and not affected by dualization.

Definition 9 (Info-symmetries). Let $K : M \Leftrightarrow N$ be a consistency framework.

(a1) If neither side has non-idle private updates, we call K *poorly info-symmetric* and write $M \approx_K N$ or $K \approx_K : M \Leftrightarrow N$.

(a2) If both sides have non-idle private updates, we call K *richly info-symmetric*, and write $M \times_{\kappa} N$ or $K_{\times} : M \leftrightarrow N$

We say a consistency framework K is *info-symmetric* if it is either poorly or richly symmetric.

(b) If only one side has non-idle private updates, we call the consistency framework *info-asymmetric*. If this side is the target space, we write $M <_{\kappa} N$ or $K_{<} : M \leftrightarrow N$. If it is the source space, we write $M >_{\kappa} N$ or $K_{>} : M \leftrightarrow N$.

Theorem 1 (Info-symmetry taxonomy). Let $K : M \leftrightarrow N$ be a consistency framework. There are four mutually exclusive and jointly complete logical possibilities (concrete info-symmetry types): (a1) $M \approx_{\kappa} N$, (a2) $M \times_{\kappa} N$, (b1) $M <_{\kappa} N$, and (b2) $M >_{\kappa} N$, which can be grouped in two abstract types: info-symmetry (a) = (a1) \vee (a2), and info-asymmetry (b) = (b1) \vee (b2). The latter type is almost concrete as types (b1) and (b2) are mutually dual. \square

Two important remarks are in order.

Remark 3 (Implicit metamodel mappings). Each of the four relationships above is a property of a triple (M, K, N) rather than a pair (M, N) . When we talk about the info-symmetry relation between UML models and Java code, or between class diagrams and relational schemas, we implicitly assume some mapping between the metamodels is given, and this mapping determines a corresponding consistency framework between the model spaces (see [3, 10] for details). In this sense we can use a loose notation and write $M \times_{\text{inf}} N$ meaning that we have a consistency framework K such that $M \times_{\kappa} N$.

Remark 4 (Spaces vs. models). Info-symmetry is a relationship between model spaces rather than individual model states. Suppose, for example, that we have a richly symmetric situation $M \times_{\kappa} N$, and two synchronized models A and B evolving in their respective spaces (we will consider this in detail later in 4). A specific state A_0 of model A can lack private updates because A_0 is a poor model that does not have private data, and hence any update $a : A_0 \rightarrow A_1$ is public. But state A_1 (or consecutive states) can well have private data and hence private updates. When we have been writing $A \times_{\text{inf}} B$ in the paper, we actually meant the respective relationship between model spaces defined by the metamodels of A and B (w.r.t. some implicit mapping between the metamodels as explained in Remark 3 above).

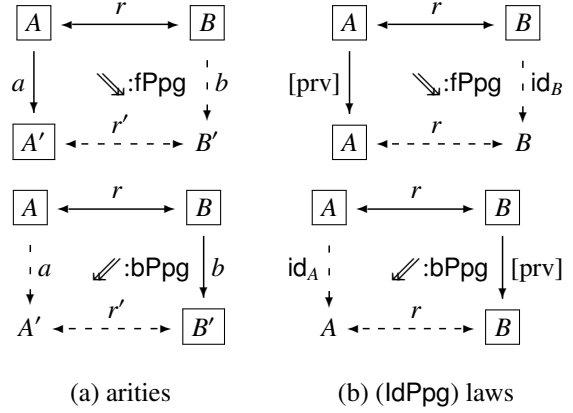


Figure 2: Operations of update propagation

3. Formalizing Incrementality: Delta Lenses

We will describe a family of algebraic structures called (*delta*) *lenses*, which model non-incremental and incremental model synchronization. Each such a structure will be defined over some underlying consistency framework $K : M \leftrightarrow N$, and comprises two operations of forward (from M to N) and backward (from N to M) update propagation.

3.1. Incremental update propagation and lenses

Definition 10 (Lenses). Let $K : M \leftrightarrow N$ be a consistency framework. An a (*delta*) *lens* over K is a pair of operations over corrs and updates, $fPpg$ and $bPpg$, which are called, resp., *forward* and *backward update propagation*. The arities of the operations are specified in Fig. 2(a) with output arrows dashed and output nodes not framed.

We will often use a linear notation and write $a.fPpg(r) = b$ (read “update a is propagated over corr r to b ”) and $a = bPpg(r).b$ (“ a is obtained by backward propagation of b over r ”) for the cases specified in the diagrams. Similarly, we will write $r' = r.fPpg(a)$ (read “ r' is the result of re-alignment caused by update a ”) and $r' = bPpg(b).r$ (“ r' is re-alignment caused by b ”).

We denote a delta lens by a double arrow $\lambda : M \rightleftharpoons N$ to recall two operations.

Definition 11 (Dual lenses). Given a consistency framework $K : M \leftrightarrow N$ and a lens $\lambda : M \rightleftharpoons N$ over it, the dual lens, $\lambda^\circ : N \rightleftharpoons M$ consist of the operations $(bPpg, fPpg)$ over the dual consistency framework $K^\circ : N \leftrightarrow M$.

Definition 12 (Well-behaved lenses). A lens $\lambda: M \rightleftharpoons N$ is called *well-behaved (wb)* if it satisfies several laws specified below.

Correctness. If $(r: A \leftrightarrow B) \in K(A, B)$ and $(b, r') = \text{fPpg}(a, r)$ for some update $a: A \rightarrow A'$, then $r' \in K(A', B')$ where $b: B \rightarrow B'$. Similarly for operation bPpg . That is, update propagation ensures consistency restoration. \square

Privacy no-op. For any update $a: A \rightarrow A'$ the following two conditions are equivalent: (i) a is private, (ii) $a.\text{fPpg}(r) = \text{id}_B$ for any consistent corrs $r: A \leftrightarrow B$. Similar equivalences are required for any update $b: B \rightarrow B'$ (the lower diagram). That is, private updates defined by the underlying consistency framework are exactly updates mapped to identity, i.e., in fact, not propagated to the other side. \square

Recall that private updates determined by the underlying consistency framework are updates that do not affect consistency, and hence, need not be propagated. Stevens [11] called this property *Hypocraticness*. Hence, Privacy no-op implies *Hypocraticness*, but also requires the converse implication.

As idle updates are private, they are propagated to idle updates and we say that the lens is *stable*: if nothing changes on one side, nothing happens on the other side as well [5] (a very special case of *Hypocraticness*).

Proposition 5 (Conjecture). Is condition (ii) in Privacy no-op equivalent to the following weak version: $a.\text{fPpg}(r) = \text{id}_B$ for some consistent corr $r: A \leftrightarrow B$ (see the upper diagram in Fig. 2(b))

Compatibility with alignment. If $a.\text{fPpg}(r) = \text{id}_B$ for $a: A \rightarrow A'$ and $r: A \leftrightarrow B$, then $a * r = r.\text{fPpg}(a)$. Dually for bPpg . That is, a *wb* delta lens can re-align a corr if the change was caused by a *private* update. Re-alignment along a public update is, in general, not possible as the lens must first propagate, and only then re-align the update. \square

Compositionality. Consider two consecutive updates $a_1: A \rightarrow A_1$, $a_2: A_1 \rightarrow A_2$ and their composition $a_{12} = a_1; a_2: A \rightarrow A_2$. In reasonable synchronization scenarios, if both updates a_i are insertions, or if both are deletions, then $a_{12}.\text{fPpg}(r) = a_1.\text{fPpg}(r); a_2.\text{fPpg}(r_1)$ for any consistent corr $r: A \leftrightarrow B$ and $r_1 = r.\text{fPpg}(a_1)$ (Fig. 3 shows how arrows fit together, but the marker $:\text{fPpg}$ to be labeling the outer rectangle ABB_2A_2 is not shown).

In general, for a corr $r: A \leftrightarrow B$, the equality $a_{12}.\text{fPpg}(r) = a_1.\text{fPpg}(r); a_2.\text{fPpg}(r_1)$ with $r_1 =$

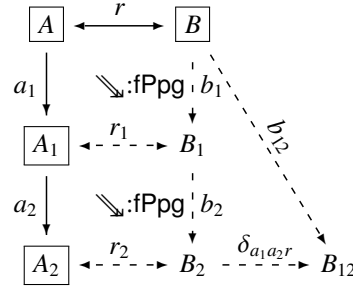


Figure 3: Compositionality for fPpg

$r.\text{fPpg}(a_1)$ does not hold if update a_1 involves deletion of some public data (say, A^-) of model A whereas update a_2 restores these data so that the composed update a does not actually change A . However, the story on the B -side is more complex.

Update $b_1 = a_1.\text{fPpg}(r)$ must delete the r -image of data piece A^- in model B , say, B^- , and the respective part of private data based on B^- , say, $B^{-\text{prv}}$, as well. Thus, we have $b_1: B \rightarrow B_1$ with B^- and $B^{-\text{prv}}$ deleted from B_1 . Then update $b_2 = a_2.\text{fPpg}(r_1): B_1 \rightarrow B_2$ must restore data B^- in model B_2 , but as the respective private component $B^{-\text{prv}}$ is lost, b_2 replaces it with some standard minimal set of N -private data, say, B_0 . Thus, model B_2 is built from B^- and B_0 . In contrast, update $b_{12} = a_{12}.\text{fPpg}(r): B \rightarrow B_{12}$ keeps data B^- unchanged in B_{12} , and hence the latter is built from B^- and $B^{-\text{prv}}$. Minimality of B_0 is specified by a uniquely determined delta $\delta_0^-: B_0 \rightarrow B^-$, and then we should have a delta $\delta: B_2 \rightarrow B_{12}$ obtained by pairing id_{B^-} and δ_0^- such that $b_1; b_2; \delta = a_{12}.\text{fPpg}(r)$. This delta is private (as its public component is idle). These considerations motivate the following formal law: For any two consecutive updates, $a_1: A \rightarrow A_1$, $a_2: A_1 \rightarrow A_2$ and a corr $r: A \leftrightarrow B$, there is a *private* delta $\delta_{a_1a_2r}: B_2 \rightarrow B_{12}$ such that $b_1; b_2; \delta_{a_1a_2r} = (a_1; a_2).\text{fPpg}(r)$ as shown in Fig. 3

If both updates are insertions, or both are deletions, then $\delta_{a_1a_2r}$ is required to be an identity, and, hence, $a_1.\text{fPpg}(r); a_2.\text{fPpg}(r_1) = (a_1; a_2).\text{fPpg}(r)$. Moreover, if at least one of the updates is private, then $\delta_{a_1a_2r}$ is required to be an identity too.

A similar law is stated for a corr and two consecutive updates on the N -side. \square

Invertibility. We begin with an informal discussion. Given an update $a: A \rightarrow A'$ and a consis-

tent corr $r: A \leftrightarrow B$, let $a_{rr}: A \rightarrow A'_{rr}$ denotes the update $a.fPpg(r).bPpg(r)$ resulting from forward and then backward propagation (see the left and central squares in diagram Fig. 4a). In general, $a_{rr} \neq a$ because when a is propagated to the N-side, the private part of a is lost and cannot be restored. In more detail, we can consider update $a: A \rightarrow A'$ as a pair (a^{prv}, a^{pub}) , whose components update, resp. the private, say, A^{prv} , and the public, A^{pub} , parts of A . Then update a_{rr} is also a pair $(id_{A^{prv}}, id_{A^{pub}})$, and there should be a delta $\delta_{ar} = (a^{prv}, id_{A^{pub}}): A'_{rr} \rightarrow A'$ so that $a_{rr};\delta_{ar} = a$. These considerations motivate the following formal law: for any $a: A \rightarrow A'$ and consistent corr $r: A \leftrightarrow B$, there is defined a *private* delta $\delta_{ar}: A'_{rr} \rightarrow A'$ such that $a_{rr};\delta_{ar} = a$ with symbols A'_{rr} and a_{rr} explained in Fig. 4(a). Dually, for any $b: B \rightarrow B'$ and consistent corr $r: A \leftrightarrow B$, there is defined a private update δ_{br} such that $b;\delta_{br} = b_{rr}$ as shown in diagram Fig. 4(b). We call δ_{ar} and δ_{br} *invertibility deltas*.

Corollary 2. An immediate consequence of the existence of invertibility deltas are the following laws: $a.fPpg(r).bPpg(r).fPpg(r) = a.fPpg(r)$ and $b.bPpg(r).fPpg(r).bPpg(r) = b.bPpg(r)$ described by Fig. 4(a')(b') and called *weak invertibility* in [5]. Thus, invertibility deltas imply weak invertibility. \square

We present two results about lenses.

Theorem 2. For a wb-lens, if update $a: A \rightarrow A'$ is public, $a \in M_{\Delta}^{pub}$, then for any consistent corr $r: A \leftrightarrow B$, update $a.fPpg(r)$ is also public and belongs to N_{Δ}^{pub} . Dually, if $b: B \rightarrow B'$ is public, then $bPpg(r).b$ is also public.

Proof. Follows from the definition of privacy/publicity and the privacy of invertibility deltas. \square

Theorem 3. A lens $\lambda: M \rightleftharpoons N$ is wb iff its dual $\lambda^{\circ}: M \rightleftharpoons N$ is wb too. That is, well-behavedness is a symmetric notion. \square

3.2. Info-Asymmetry and Lenses

The consistency framework underlying a lens influences its properties.

Definition 13 (Asymmetric lenses). [ZD: insert the def from the jot11 paper [4].] We denote an asymmetric lens by $\lambda^{\leq}: M \rightleftharpoons N$ with the superscript pointing to the dominated side (the view).

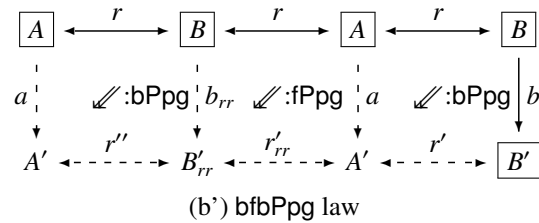
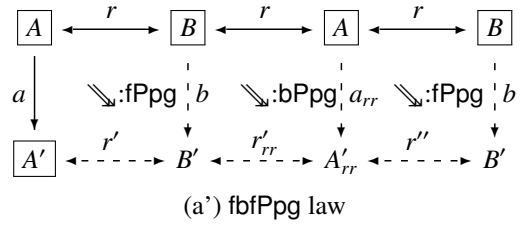
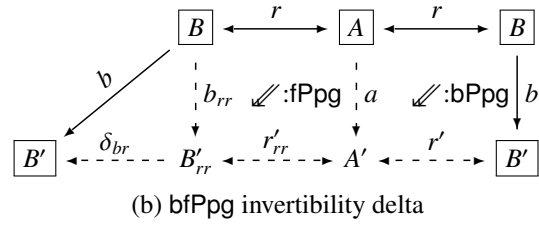
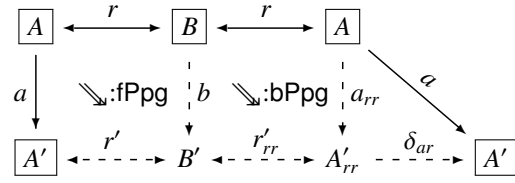


Figure 4: Round-tripping laws. (a) and (b) are Invertibility deltas. (a') and (b') are weak invertibility. Scenario in diagrams (b,b') “run” from the right to the left.

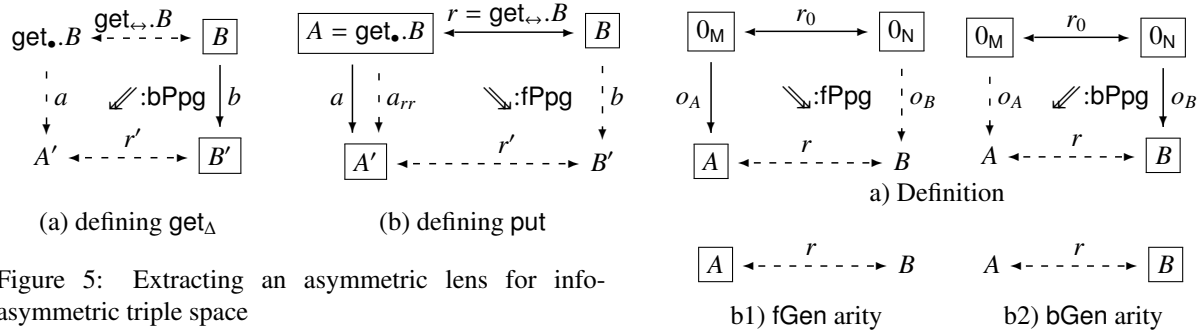


Figure 5: Extracting an asymmetric lens for info-asymmetric triple space

Theorem 4 (Info-asymmetry and lenses). Any delta lens $\lambda: M \rightleftharpoons N$ over an asymmetric alignment framework $K^\leq: M \leftrightarrow N$ (i.e., $M \leq_{\text{inf}} N$ holds), gives rise to an asymmetric delta lens as defined in [4].

Proof. We will employ functions $\text{get}_\bullet^K: M \leftarrow N$ and $\text{get}_\bullet^K: R \leftarrow N_\bullet$, provided by the asymmetry of $K: M \leftrightarrow N$, and show how they can be used to define functions get_Δ and put such that the PutGet law holds.

Diagram Fig. 5(a) shows how to define get_Δ for a given update $b: B \rightarrow B'$. We first compute $A = \text{get}_\bullet^K.B$ and $r = \text{get}_\bullet^K.B$, then apply bPpg and obtain arrows a and r' as shown in the diagram. Because $\bullet K.B$ is a singleton by the info-asymmetry condition, we necessarily have $A' = \text{get}_\bullet^K.B'$ and $r' = \text{get}_\bullet^K.B'$. We now define $\text{get}_\Delta(b) = a$. It preserves identities as bPpg does.

Definition of put is shown by diagram Fig. 5(b). Recall that get_\bullet^K is surjective; having an update a and model B such that $A = \text{get}_\bullet^K.B$ as shown in the diagram (b), we compute $r = \text{get}_\bullet^K(B)$ and apply fPpg , which produces arrows b and r' as shown. We define $\text{put}(a, B) \stackrel{\text{def}}{=} b$. Because $\bullet K.B'$ is a singleton, we conclude that $A' = \text{get}_\bullet^K.B'$, and it remains to prove the PutGet law. Let $a_{rr} \stackrel{\text{def}}{=} \text{get}_\Delta.b = \text{bPpg}(r).(a.\text{fPpg}(r))$ (see diagram Fig. 4(a)), and δ_{ar} is the respective invertibility delta. The latter is always private, and by asymmetry of the alignment framework, δ_{ar} must be an identity, hence, $a = a_{rr}$. \square

Theorem 5 (Info-bijectivity and lenses). Let $\lambda: M \rightleftharpoons N$ be a lens over a bijective consistency framework $K^\cong: M \leftrightarrow N$. Then categories M and N are isomorphic via mutually inverse functors $\text{get}_1: M \leftarrow N$ and $\text{get}_2: M \rightarrow N$.

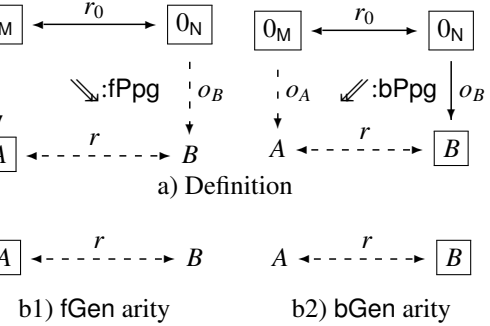


Figure 6: Non-incremental propagation

3.3. Non-incremental update propagation.

We will begin with an important addition to the notion of model space.

Definition 14 (Initial models). Let M be a model space. A model 0_M is called *initial*, if for any model $A \in M_\bullet$, there is a unique delta $o_A: 0_M \rightarrow A$. It can be proven by standard categorical means that all initial models are isomorphic.

Intuitively, the initial model comprises all data that must be in any M -model and so is the minimal model possible; delta o_A embeds model 0_M into A . For many structural model spaces, e.g., class diagrams, the initial model is empty. In contrast, behavior models often must have certain states and transitions initializing the behavior, e.g., initial states.

Definition 15 (Well-behaved lenses and initial models). A consistency framework $K: M \leftrightarrow N$ is called *wb*, if in addition to laws stated in Definition 5, $K(0_M, 0_N)$ is a singleton $\{r_0\}$.

A lens $\lambda: M \rightleftharpoons N$ is *wb*, if, in addition to laws stated in Definition 12, the following holds: for any $A \in M_\bullet$, $o_A.\text{fPpg}(r_0) = o_B$ for some $B \in N_\bullet$, and similarly, for any $B \in N_\bullet$, $o_A = \text{bPpg}(r_0).o_b$ for some $A \in M_\bullet$. \square

Now we note that a *wb* lens $\lambda: M \rightleftharpoons N$ provides also non-incremental propagation operations defined as specified in Fig. 6(a). What we have defined are two operations, *forward*, fGen , and *backward*, bGen , *generation*, whose arities are specified in Fig. 6(b): $A.\text{fGen} = (r, B)$ and $(A, r) = \text{bGen}.B$

Each of these operations actually comprises two ordinary operations: fGen consists of $\text{fGen}_\bullet: M_\bullet \rightarrow N_\bullet$.

and $fGen_{\leftrightarrow}: M_{\bullet} \rightarrow R$; and $bGen$ consists of $\bullet bGen: M_{\bullet} \leftarrow N_{\bullet}$ and $\leftarrow bGen: R \leftarrow N_{\bullet}$. For example, in diagram Fig. 6(b1), we have $A.fGen_{\leftrightarrow} = r$ and $A.fGen_{\bullet} = B$.

The following result shows how non- and incremental update propagation are related (see [12] for details).

Theorem 6. Let $\lambda: M \rightleftharpoons N$ be a wb lens, $a: A \rightarrow A'$ an update, and $r: A \leftrightarrow B$ a consistent corr. Let $(a, r).fPpg = (r', b)$ and $A.fGen = (r_{||}, B_{||})$. Then there is a unique private delta $\delta_{ar}: B_{||} \rightarrow B$ such that $r_{||}; \delta_{ar} = r'$. \square

Thus, although a lens provides incremental update propagation in both directions, we can choose to make change propagation in one or both directions non-incremental by employing operations $fGen$ and $bGen$ defined above. In this way a lens determines several *computational frameworks* depending on which (if any) directions of update propagation are chosen to be non-incremental.

Definition 16 (Incremental specializaion). Let $\lambda: M \rightleftharpoons N$ be a wb lens. Its *incremental specializaion* is an algebraic structure $\lambda^{inc}: M \rightleftharpoons N$ comprising two operations (Op_1, Op_2) defined as follows.

(i) If $Op_1 = fGen$ and $Op_2 = bGen$, i.e., neither direction is incremental, we call the specialization *non-incremental* and write $\lambda^{\cup}: M \rightleftharpoons N$.

(ii) If $Op_1 = fPpg$ and $Op_2 = bPpg$, i.e., both directions are incremental, we call the specialization *fully incremental*, and write $\lambda^{\times}: M \rightleftharpoons N$.

(iii) If only one direction is incremental, specialization is *semi-incremental*. If the incremental direction is from the source to the target, i.e., $Op_1 = fPpg$ while $Op_2 = bGen$, we write $\lambda^{>}: M \rightleftharpoons N$. If the incremental direction is from the target to the source, i.e., $Op_1 = fGen$ while $Op_2 = bPpg$, we write $\lambda^{<}: M \rightleftharpoons N$.

We will often call an incremental specialization $\lambda^{inc}: M \rightleftharpoons N$ of a lens just a lens.

Theorem 7 (Incrementality taxonomy). Let $\lambda: M \rightleftharpoons N$ be a wb lens. There are four mutually exclusive and jointly complete logical possibilities (concrete incrementality types) for the incremental specialization $\lambda^{inc}: M \rightleftharpoons N$ with $inc \in \{\cup, \times, >, <\}$. The last two types are mutually dual, and can be grouped into almost concrete type of semi-incrementality. \square

3.4. Lenses, incrementality and info-symmetry

Our work above demonstrates that the type of a lens $\lambda: M \rightleftharpoons N$ is determined by two indexes: one is its info-symmetry type $inf \in \{\approx, \times, >, <\}$ (determined by the info-symmetry type of the underlying consistency framework), and the other is the type of its incremental specialization $inc \in \{\cup, \times, >, <\}$. We will write a lens with double-indexing λ_{inf}^{inc} and call each of so double-indexed specialization a *computational frameworks*. The total is 16 types of computational frameworks. However, amongst these 16 types, several are basically the same up to permutation of the source and the target, e.g., lenses $\lambda_{<}^<$ and $\lambda_{>}^>$ possess the same properties and dualization mutually converts them each into the other. The same is true for lenses of types $\lambda_{<}^>$ and $\lambda_{>}^<$. However, lenses $\lambda_{>}^>$ and $\lambda_{<}^<$ are essentially different. The lower inf -index shows that both lenses support view maintenance. The upper inc -index of lens $\lambda_{<}^<$ describes it as a pair $(fPpg, bGen)$ that supports non-incrementally computed (by $bGen$) and updatable (by $fPpg$) view. In contrast, the inc -index of lens $\lambda_{>}^>$ describes it as a pair $(fGen, bPpg)$ that supports incremental view computation but non-incremental view update propagation; the latter makes this computational framework non-applicable for databases (but it can still be used for model compilation and incremental reverse engineering). As Fig. ?? shows, there are 10 really different lens types.

4. Formalizing Organizational symmetry: Synchronization Cases and Types

Organizational symmetry is about change propagation, and in this section we first introduce a key notion of a changing model as a *trajectory* in the respective model space. Then we define a synchronization case as a pair of trajectories satisfying certain mutual synchronization conditions. Finally, we define organizational polices as special constraints on synchronization cases, which can be classified by their symmetry w.r.t. the source and the target models.

4.1. Models as trajectories

We consider a changing model as a *trajectory* in the respective model space. Let M be a metamodel, and $M = (M_{\bullet}, M_{\Delta})$ be the space of its instances and deltas as described above. A *model-as-trajectory* is a mapping

$A: I \rightarrow \mathbf{M}_\bullet$, whose domain I is a finite linearly ordered set $\{i_0 < i_1 < \dots < i_n\}$ of *version numbers* or *indexes*. Thus, A appears as the model's immutable identity whereas its state $A(i)$ changes as index i runs over I . To simplify notation, below we will write A_i for $A(i)$, and by the abuse of terminology often call model's states just models.

To traverse set I , we will use operations $i-1, i-2, \dots$ and $i+1, i+2, \dots$ with the evident meaning ($i_0 - 1$ and $i_n + 1$ are not defined).

If $i \in I$ is a version number and $i-1$ is its predecessor, we have a (directed) delta $a_i: A_{i-1} \rightarrow A_i$ specifying the change. We may consider the pair $(i-1, i)$ as an arrow from $i-1$ to i , and delta a_i as the A -image of this arrow in \mathbf{M}_Δ . This makes I a directed graph, and A a graph mapping. Moreover, we can make I a category \mathbf{l} with nodes $\bullet = I$ and arrows $\mathbf{l}_\Delta = \{(i_1 i_2) \in I \times I: i_1 < i_2\}$, arrow composition $(i_1 i_2); (i_2 i_3) = (i_1 i_3)$ and identities ii . Then a model trajectory is a functor (a mapping of categories) $A: \mathbf{l} \rightarrow \mathbf{M}$, i.e., a graph mapping such that $A(i_1 i_3) = A(i_1 i_2); A(i_2 i_3)$ and $A(ii) = \text{id}_{A_i}$. For any non-initial index i , we write a_i for delta $A(i-1, i): A_{i-1} \rightarrow A_i$ to be read "the update that created model A_i ".

4.2. Synchronization cases

Synchronization of two models is about maintaining certain correspondences between two trajectories, say, $A: \mathbf{l} \rightarrow \mathbf{M}$ and $B: \mathbf{j} \rightarrow \mathbf{N}$, in two computationally related model spaces. We will assume the spaces are related by a wb delta lens $\lambda: \mathbf{M} \rightleftharpoons \mathbf{N}$ comprising operations of forward and backward delta propagation as described in Def. 10.

Partitioning of model deltas into private and public determines a similar partitioning of indexes $I = I^{\text{prv}} \uplus I^{\text{pub}}$ with

$$I^{\text{prv}} = \{i \in I: a_i \in \mathbf{M}_\Delta^{\text{prv}}\}, \text{ and } I^{\text{pub}} = \{i \in I: a_i \in \mathbf{M}_\Delta^{\text{pub}}\},$$

and similarly $J = J^{\text{prv}} \uplus J^{\text{pub}}$. Thus, $i \in I^{\text{prv}}$ (or $i \in I^{\text{pub}}$) means that model A_i is the result of a private (resp. public) update $a_i: A_{i-1} \rightarrow A_i$.

Since public updates destroy consistency, as soon as a public update is committed on one side, it must be propagated to the other side to restore consistency. According to Theorem 2, the propagated update is also public, but we call it *passive*, whereas the original update is *active*.

For example, suppose that we have registered a public update $a_i: A_{i-1} \rightarrow A_i \in \mathbf{M}_\Delta^{\text{pub}}$ on the A -side, and hence $i \in I^{\text{pub}}$. There are two possibilities for the case.

(a) Update a_i was initiated on the A side (we say a_i is *active*) and then was propagated to the B side. This means that there is a version number $i \triangleright \in J$ and update $b_{i \triangleright}: B_{(i \triangleright)-1} \rightarrow B_{i \triangleright}$ produced by this propagation (we then say that $b_{i \triangleright}$ is *passive*), so that the corr resulted from this propagation, $r_{i, i \triangleright}: A_i \leftrightarrow B_{i \triangleright}$, restores consistency. Considering all such updates gives us an order preserving bijection $\triangleright: I^{\text{act}} \rightarrow J$ for some ordered subset $I^{\text{act}} \subseteq I^{\text{pub}}$.

(b) Update a_i is the result of propagation from the other side (now a is *passive*) of some (active) update $b_{i \blacktriangleright}: B_{(i \blacktriangleright)-1} \rightarrow B_{i \blacktriangleright}$ for some version number $i \blacktriangleright \in J$, and we again have a consistent corr $r_{i, i \blacktriangleright}: A_i \leftrightarrow B_{i \blacktriangleright}$ resulted from this backward propagation. Considering all such updates gives us an order preserving bijection $\blacktriangleright: I^{\text{pas}} \rightarrow J$ for some ordered subset $I^{\text{pas}} \subseteq I^{\text{pub}}$. Clearly, sets I^{act} and I^{pas} are disjoint and their union is I^{pub} , $I^{\text{pub}} = I^{\text{act}} \uplus I^{\text{pas}}$.

Viewing the same pair of trajectories from the B -side, gives us partitioning $J^{\text{pub}} = J^{\text{act}} \uplus J^{\text{pas}}$ and order-preserving bijections $\triangleleft: I \leftarrow J^{\text{act}}$ and $\blacktriangleleft: I \leftarrow J^{\text{pas}}$. Moreover, for a correct synchronization case, we should have bijections \triangleright and \blacktriangleleft to be mutually inverse and set an isomorphism $I^{\text{act}} \cong J^{\text{pas}}$. Similarly, bijections \blacktriangleright and \triangleleft are to be mutually inverse as well, and $I^{\text{pas}} \cong J^{\text{act}}$.

Thus, for a pair of correctly synchronized trajectories as above, we have a partitioning $I^{\text{pub}} = I^{\text{act}} \uplus I^{\text{pas}}$, a partitioning $J^{\text{pub}} = J^{\text{act}} \uplus J^{\text{pas}}$, and two pairs of order-preserving bijections,

$$\triangleright: I^{\text{act}} \rightarrow J^{\text{pas}} \text{ and } \blacktriangleright: I^{\text{pas}} \rightarrow J^{\text{act}},$$

and

$$\blacktriangleleft: I^{\text{pas}} \leftarrow J^{\text{act}} \text{ and } \triangleleft: I^{\text{act}} \leftarrow J^{\text{pas}},$$

such that $\triangleright = \blacktriangleleft^{-1}$ and $\blacktriangleright^{-1} = \triangleleft$. Hence, we have an isomorphism $\triangleright \triangleleft \subset I^{\text{pub}} \times J^{\text{pub}}$ (we could denote it by $\blacktriangleright \blacktriangleleft$ as well) such that for any pair $i \triangleright \triangleleft j \in I^{\text{pub}} \times J^{\text{pub}}$, models A_i and B_j are consistent via a corr $r_{i \triangleright \triangleleft j}$ computed during update propagation.

Figure 7 provides more details for the mechanism. Suppose that models A_i and B_j have been synchronized, and after that each of the models evolved with its own sequence of private updates (labeled with [prv] in the figure). Suppose that a public update $a_{i \oplus 1}$, where $i \oplus 1$ denotes

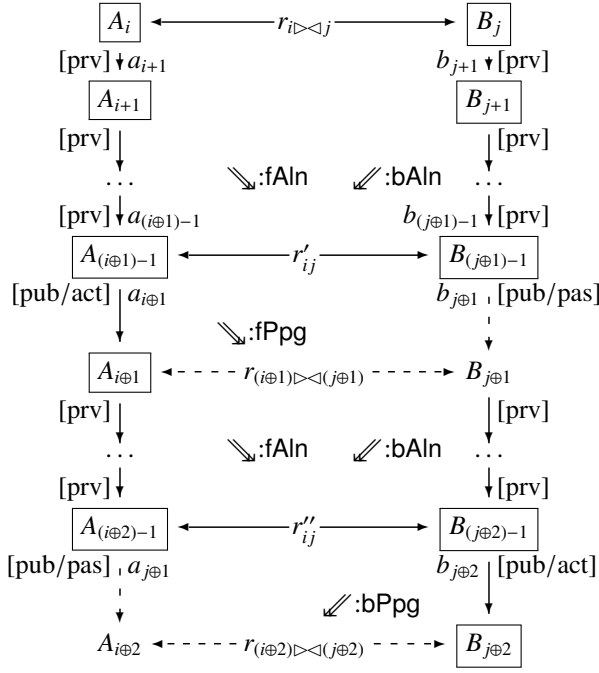


Figure 7: A general sync case

the next index in the set I^{pub} (which is different from the next index, $i+1$, in the entire set I), was committed on the A-side and destroyed consistency. Hence, this update was propagated to side B to restore consistency. In order to preserve the results of private changes on the B side, this propagation has been done for the last corr r'_{ij} before the public update $a_{i\oplus 1}$ as shown in the figure. This corr r'_{ij} is computed by formula (1)

$$[a_{i+1}; a_{i+2}; \dots; a_{(i\oplus 1)-1}] * r_{i \triangleright \triangleleft j} * [b_{j+1}; b_{j+2}; \dots; b_{(j\oplus 1)-1}] \quad (1)$$

where $j\oplus 1$ denotes the immediate successor of j in J^{pub} , and $*$ denotes the realignment operation specified in Def. 2 and 3. Note that the set of I -indexes from i to $i\oplus 1$ and the set of J -indexes from j to $j\oplus 1$ can be of different length. It is also possible that, e.g., $(i\oplus 1) - 1 = i$, which means there were no private updates on the left side preceding the public one, in which case the left term in square brackets above is an identity update. The same holds for the right term in square brackets.

Now update $a_{i\oplus 1}$ can be propagated to the right side by

applying operation fPpg to the pair $(a_{i\oplus 1}, r'_{ij})$, which results in update $b_{j\oplus 1}$ and the new corr $r_{(i\oplus 1) \triangleright \triangleleft (j\oplus 1)}$. The latter is consistent because r'_{ij} is consistent (since previously only private update were made on both sides), and a correct propagation preserves consistency.

In the next synchronization step of the case, public update $b_{j\oplus 2}$ was first committed on the B -side, and hence propagated to the left side with operation bPpg . The input corr for the operation was again provided by the alignment framework, again computed by formula (1) with i and j replaced by $i\oplus 1$ and $j\oplus 1$ (and, respectively, $i\oplus 1$ and $j\oplus 1$ in formula (1) are replaced by $i\oplus 2$ and $j\oplus 2$). The result is a new pair of synchronized models $r_{(i\oplus 2) \triangleright \triangleleft (j\oplus 2)}$, and the system is ready to yet another synchronization step as above.

The next definition gives an accurate formalization of the discussion above. As mappings denoted by black triangles are derived from those denoted by blank triangles, we can specify the required correspondences by only using blank triangle mappings.

Definition 17 (Synchronization case). Given a fully incremental delta lens $\lambda^\times : M \rightleftharpoons N$, a (consistent) synchronization case is a pairs of trajectories, $A : I \rightarrow M$ and $B : J \rightarrow N$, with the following additional structure.

(a) Sets I^{pub} and J^{pub} are further partitioned, $I^{\text{pub}} = I^{\text{act}} \uplus I^{\text{pas}}$ and $J^{\text{pub}} = J^{\text{act}} \uplus J^{\text{pas}}$, and two isomorphisms are given: $\triangleright : I^{\text{act}} \rightarrow J^{\text{pas}}$ and $\triangleleft : J^{\text{pas}} \leftarrow I^{\text{act}}$. This establishes an isomorphism $\triangleright \triangleleft \subset I^{\text{pub}} \times J^{\text{pub}}$, and for each pair of corresponding indexes $i \triangleright \triangleleft j$, there is a consistent corr $r_{i \triangleright \triangleleft j} : A_i \leftrightarrow B_j$. Particularly, for the initial indexes we have $i_0 \triangleright \triangleleft j_0$ and $r_0 = r_{i_0 \triangleright \triangleleft j_0}$.

(b) Let $i\oplus 1$ denotes the next index in the linearly ordered set I^{pub} , which, in general, may be greater than the next index $i+1$ in set $I \supset I^{\text{pub}}$ due to several private update indexes following i before $i\oplus 1$. Similarly, $j\oplus 1$ is the next index in J^{pub} .

If $i \triangleright \triangleleft j$ and $i\oplus 1 \in I^{\text{act}}$, then $j\oplus 1 \in J^{\text{pas}}$ and

$$a_{i\oplus 1} \cdot \text{fPpg}(r'_{ij}) = b_{j\oplus 1} \text{ and } r_{(i\oplus 1) \triangleright \triangleleft (j\oplus 1)} = r'_{ij} \cdot \text{fPpg}(a_{i\oplus 1}),$$

where r'_{ij} is computed by (1).

If $i \triangleright \triangleleft j$ and $i\oplus 1 \in I^{\text{pas}}$, then $j\oplus 1 \in J^{\text{act}}$ and

$$a_{i\oplus 1} = \text{bPpg}(r'_{ij}) \cdot b_{j\oplus 1} \text{ and } r_{(i\oplus 1) \triangleright \triangleleft (j\oplus 1)} = \text{bPpg}(b_{j\oplus 1}) \cdot r'_{ij}.$$

where r'_{ij} is again computed by the same formula (1).

Modifications of this definition for the cases of a non-incremental lens $\lambda^u: M \rightleftharpoons N$ and a semi-incremental lens $\lambda^>: M \rightleftharpoons N$ or $\lambda^<: M \rightleftharpoons N$ are straightforward. \square

We will denote a synchronization case as defined above by a double arrow $\sigma: A \rightleftharpoons B$, and the set of all such cases by $\mathbf{Sync}_\lambda(A, B)$ with the subscript referring to the underlying lens providing the computational framework. Given a synchronization case σ , we will sometime subscript its constituting elements and write $A_\sigma: I_\sigma \rightarrow M$ etc.

4.3. Organizational Policies and Types

Class $\mathbf{Sync}_\lambda(A, B)$ encompasses all possible synchronizations of models A and B . If a special *organizational* policy between the models is assumed, some synchronization cases can be a priori prohibited. For example, we can prohibit update propagation from A to B and thus make A an entirely passive receiver of changes from side B , if we require $I_\sigma^{\text{act}} = \emptyset$ for any case σ allowed by the policy. Dually, we make B a passive receiver of changes A by requiring $J_\sigma^{\text{act}} = \emptyset$ for any legal case σ . To establish a more refined organizational policy, we can prohibit propagating some, but not all, updates from either side so that both sets I^{act} and J^{act} are not empty. Thus, an org-policy specifies a special subclass of class $\mathbf{Sync}_\lambda(A, B)$.

Below we make these ideas formal.

Definition 18 (Org-policy). Let $\lambda: M \rightleftharpoons N$ be a wb lens.

(i) An *org-policy* is a pair of sets $\Pi = (P, Q)$ with $P \subseteq M_\Delta^{\text{pub}}$ and $Q \subseteq N_\Delta^{\text{pub}}$, whose elements are called *propagatable* deltas. We require Π to be *complete* in the following sense. Let $P.\lambda$ denotes all target updates propagated from P with lens λ , $P.\lambda \stackrel{\text{def}}{=} \{a.f\text{Ppg}(r): a \in P, r \in K\}$, and, dually, $\lambda.Q$ denotes the set of all source updates propagated from Q , i.e., the set $\{b.\text{bPpg}(r).b: b \in Q, r \in K\}$. Then we require $M_\Delta^{\text{pub}} = P \cup \lambda.Q$ and $P.\lambda \cup Q = N_\Delta^{\text{pub}}$.

(ii) We say a synchronization case $\sigma: A \rightleftharpoons B$ *conforms* to policy $\Pi = (P, Q)$ and write $\sigma \models \Pi$, if $a_i \in P$ for all $i \in I_\sigma^{\text{act}}$, and $b_j \in Q$ for all $j \in J_\sigma^{\text{act}}$. In other words, in order to claim conformance $\sigma \models \Pi$, we require $\{a_i \in M_\Delta: i \in I_\sigma^{\text{act}}\} \subset P$ and $\{b_j \in N_\Delta: j \in J_\sigma^{\text{act}}\} \subset Q$.

In this way, an org-policy Π determines a corresponding synchronization *type*, i.e., a class of synchronization cases $\llbracket \Pi \rrbracket = \{\sigma \in \mathbf{Sync}_\lambda(A, B): \sigma \models \Pi\}$ conforming to the policy.

Remark 5 (Organization vs. technology). Importantly, sets P and Q constituting a policy are determined organizationally rather than technologically in the sense that propagation operations can be defined for non-propagatable deltas. For example, when code is generated from a UML model by a forward operation $f\text{Ppg}$ or $f\text{Gen}$, the backward operation $b\text{Ppg}$ is defined for all code deltas, and is important for checking correctness of code generated from the model wrt. its conformance to the model as prescribed by the invertibility law. However, only some (or none) of code deltas are allowed to be propagated back to the model.

On the other hand, if some updates are not propagatable by technical reasons, e.g., in the database context, there is no a reasonable update propagation policy ensuring the uniqueness of the source database, these updates obviously cannot be organizationally allowed. That is, an organizational policy is built within the boundaries of the technical restrictions.

Definition 19 (Org-symmetries). Let $\Pi = (P, Q)$ is an organizational policy as defined above.

(a1) If both directions propagate all possible updates: $P = M_\Delta$, $Q = N_\Delta$, we call the policy *richly org-symmetric* and write $\Pi_{\times}: A \rightleftharpoons B$ or $A \times_{\Pi} B$.

(a2) If both directions only propagate some of the possible updates: $\emptyset \neq P \subsetneq M_\Delta$, $\emptyset \neq Q \subsetneq N_\Delta$, we call the policy *poorly org-symmetric* and write $\Pi_{\leq}: A \rightleftharpoons B$ or $A \leq_{\Pi} B$.

We call the policy *org-symmetric* if it is poorly or richly symmetric.

(b) If one direction propagates all, and the other some, of the updates, we call the policy *org-semi-symmetric*. That is, either (b1) $P = M_\Delta$ and $\emptyset \neq Q \subsetneq N_\Delta$, in which case we write $\Pi_{\geq}: A \rightleftharpoons B$ or $A \geq_{\Pi} B$, or (b2) $\emptyset \neq P \subsetneq M_\Delta$ and $Q = N_\Delta$, and we write $\Pi_{\leq}: A \rightleftharpoons B$ or $A \leq_{\Pi} B$.

(c) If one direction propagates all, and the other none, of the updates, we call the policy *(strictly) org-asymmetric*. That is, either (c1) $P = M_\Delta$ and $Q = \emptyset$, in which case we write $\Pi_{>}: A \rightleftharpoons B$ or $A >_{\Pi} B$, or (c2) $P = \emptyset$ and $Q = N_\Delta$, and we write $\Pi_{<}: A \rightleftharpoons B$ or $A <_{\Pi} B$.

Theorem 8 (Org-symmetry taxonomy). There are six mutually exclusive and jointly complete logical possibilities (concrete org-symmetry types): (a1) $A \leq_{\Pi} B$, (a2)

$A \times_{\Pi} B$; (b1) $A \times_{\leq \Pi} B$, (b2) $A \times_{\geq \Pi} B$; (c1) $A >_{\Pi} B$ and (c2) $A <_{\Pi} B$. They can be grouped in three abstract types: org-symmetry (a)=(a1) \vee (a2), org-semi-symmetry (b)=(b1) \vee (b2), and org-asymmetry (c)=(c1) \vee (c2). Types (b) and (c) are almost concrete as types (b1) and (b2), as well as (c1) and (c2) are mutually dual. \square

Remark 6. Each of the six types is a property of a triple (A, Π, B) rather than a pair (A, B) . However, we may use a loose notation and write, say, $A \times_{\text{org}} B$ meaning that we have a synchronization $\Pi_{\times} : A \rightleftharpoons B$ providing $A \times_{\Pi} B$. This loose notation is used in [1].

Remark 7. A very special subtype of this type is *interleaving*, for which, in addition, the following holds: $J^{\text{priv}} = \emptyset = J^{\text{priv}}$. In other words, the two models actually share the same set of version indexes, and all changes on either sides are at once propagated to the other side in the interleaving mode.

5. Related Work

A majority of work on semantic foundations for model transformations and bx assumes an operational rule-based semantics, e.g., Maude’s term rewriting rules for ATL [13], transformations of symbolic graphs for QVT-R (the check-only mode) [14] and for general inter-modeling patterns in [15], or TGG as a general bx engine [16]. A much more declarative approach to bx semantics, in which update propagation procedures are considered as abstract algebraic operations (along the lines of the *lens* approach to the view update problem [17]), was proposed by Stevens in her seminal papers [18, 11] and developed in [19]. The original *state-based* lenses were later modified and subsumed by asymmetric [4] and symmetric [5] *delta* lenses; implementation of these delta lenses via TGG is described in, resp., [7] and [6].

The formal semantics presented above is a major development of the delta lens framework. We enrich the notion of consistency framework with the constructs of private and public updates, and show that asymmetry defined in terms of private updates coincides with asymmetry defined in terms of a view computation function. Also, we refine compositionality and invertibility laws for delta lenses by making them lax, and present some simple results about them. Accurate definitions of symmetry

and duality of delta lenses are novel, and organization of the variety of delta lenses into a two-dimensional space is novel too.

The org-symmetry dimension has been discussed in the literature as unidirectional vs. bidirectional transformations [20, 16, 21]. We present a more fine-grained taxonomy by introducing organizational semi-symmetry, and give the dimension a formal semantics via the notion of a synchronization case.

References

- [1] Z. Diskin, H. Gholizadeh, A. Wider, K. Czarnecki, A three-dimensional taxonomy for bidirectional model synchronization, *Journal of System and Software*.
- [2] M. Johnson, R. D. Rosebrugh, Spans of lenses, in: K. S. Candan, S. Amer-Yahia, N. Schweikardt, V. Christophides, V. Leroy (Eds.), *Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014)*, Athens, Greece, March 28, 2014., Vol. 1133 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014, pp. 112–118. URL <http://ceur-ws.org/Vol-1133/paper-18.pdf>
- [3] Z. Diskin, Model Synchronization: Mappings, Tiles, and Categories, in: J. M. Fernandes, R. Lämmel, J. Visser, J. Saraiva (Eds.), *GTTSE*, Vol. 6491 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 92–165.
- [4] Z. Diskin, Y. Xiong, K. Czarnecki, From State- to Delta-Based Bidirectional Model Transformations: the Asymmetric Case, *Journal of Object Technology* 10 (2011) 6: 1–25.
- [5] Z. Diskin, Y. Xiong, K. Czarnecki, H. Ehrig, F. Hermann, F. Orejas, From state-to delta-based bidirectional model transformations: the symmetric case, in: *MODELS*, Springer, 2011, pp. 304–318.
- [6] F. Hermann, H. Ehrig, F. Orejas, K. Czarnecki, Z. Diskin, Y. Xiong, Correctness of model synchronization based on triple graph grammars, in: *MODELS*, Springer, 2011, pp. 668–682.

- [7] A. Anjorin, S. Rose, F. Deckwerth, A. Schürr, Efficient model synchronization with view triple graph grammars, in: *Modelling Foundations and Applications*, Springer, 2014, pp. 1–17.
- [8] Z. Diskin, Model synchronization: mappings, tile algebra, and categories, in: R. Lämmel et al. (Ed.), *Postproceedings GTTSE 2009*, LNCS#6491, Springer, 2011.
- [9] Z. Diskin, T. Maibaum, K. Czarnecki, Intermodelling, queries, and kleisli categories, in: *Fundamental Approaches to Software Engineering*, Springer, 2012, pp. 163–177.
- [10] Z. Diskin, S. Kokaly, T. Maibaum, Mapping-aware megamodeling: Design patterns and laws, in: *Software Language Engineering*, Springer, 2013, pp. 322–343.
- [11] P. Stevens, Bidirectional model transformations in QVT: semantic issues and open questions, *Software and System Modeling* 9 (1) (2010) 7–20.
- [12] Z. Diskin, An algebraic semantics for bidirectional model synchronization, Tech. Rep. GSDLab-TR 2014-04-01, University of Waterloo (2014).
URL <http://gsd.uwaterloo.ca/node/588>
- [13] J. Troya, A. Vallecillo, A rewriting logic semantics for ATL, *Journal of Object Technology* 10 (2011) 5: 1–29.
- [14] E. Guerra, J. de Lara, An algebraic semantics for qvt-relations check-only transformations, *Fundam. Inform.* 114 (1) (2012) 73–101.
- [15] E. Guerra, J. de Lara, F. Orejas, Inter-modelling with patterns, *Software and System Modeling* 12 (1) (2013) 145–174.
- [16] A. Schürr, F. Klar, 15 Years of Triple Graph Grammars, in: *ICGT*, 2008, pp. 411–425.
- [17] J. N. Foster, M. B. Greenwald, J. T. Moore, B. C. Pierce, A. Schmitt, Combinators for bidirectional tree transformations: A linguistic approach to the view-update problem, *ACM Transactions on Programming Languages and Systems (TOPLAS)* 29 (3) (2007) 17.
- [18] P. Stevens, Bidirectional model transformations in qvt: Semantic issues and open questions, in: *Model Driven Engineering Languages and Systems*, Springer, 2007, pp. 1–15.
- [19] Z. Diskin, Algebraic Models for Bidirectional Model Synchronization, in: *MoDELS*, Vol. 5301 of LNCS, Springer, 2008, pp. 21–36.
- [20] M. Antkiewicz, K. Czarnecki, Design Space of Heterogeneous Synchronization, in: R. Lämmel, J. Visser, J. Saraiva (Eds.), *GTTSE*, Vol. 5235 of *Lecture Notes in Computer Science*, Springer, 2007, pp. 3–46.
- [21] K. Czarnecki, J. N. Foster, Z. Hu, R. Lämmel, A. Schürr, J. Terwilliger, Bidirectional Transformations: A Cross-Discipline Perspective, in: *ICMT*, 2009.