

# Chapter 1

## Concept-Oriented Modeling of Dynamic Behavior

P.C. Breedveld

**Abstract** This chapter introduces the reader to the concept-oriented approach to modeling that clearly separates ideal concepts from the physical components of a system when modeling its dynamic behavior for a specific problem context. This is done from a port-based point of view for which the domain-independent bond graph notation is used, which has been misinterpreted over and over, due to the paradigm shift that concept-oriented modeling in terms of ports requires. For that reason, the grammar and semantics of the graphical language of bond graphs are first defined without making any connection to the physical modeling concepts it is used for. In order to get a first impression of how bond graphs can represent models, an existing model is transformed into bond graphs as the transformation steps also give a good impression of how this notation provides immediate feedback on modeling decisions during actual modeling. Next, physical systems modeling in terms of bond graphs is discussed as well as the importance of the role of energy and power that is built into the semantics and grammar of bond graphs. It is emphasized that, just like circuit diagrams, bond graphs are a topological representation of the conceptual structure and should not be confused with spatial structure. By means of a discussion of some examples of such confusions it is explained why bond graphs have a slow acceptance rate in some scientific communities.

**Keywords** Labeled di-graph · Node categorization · Energy and co-energy · Legendre transform · Thermodynamic framework of variables · Generalized mechanic framework of variables · Equilibrium-determining variable · Equilibrium-establishing variable

---

P.C. Breedveld (✉)  
University of Twente, EWI/EL/CE, 7500 AE Enschede, The Netherlands  
e-mail: p.c.breedveld@utwente.nl

## 1.1 Introduction

Many models are constructed to obtain better insight into the dynamic behavior of all sorts of real-world systems and processes, however, often without careful consideration of what the meaning of such a model is. This is due to the fact that it is easy to fall into the trap of (unconsciously) extrapolating the meaning of a model to interpretations that violate the original assumptions that led to the abstraction at hand. Apart from the aim to introduce the port-based approach in terms of bond graphs in this book, this contribution is written with the aim to make clear that a model is such an abstraction that focuses on particular aspects of a part of the real world, given a problem with the real world or question about it, even if it is a question about some future realization (design). In all cases, this makes the model highly dependent on the context of that problem, even when the question is quite general: the person(s) making the abstraction will always do so from a particular viewpoint and background. As a result, models are often highly dependent on (scientific) culture and its jargon, even when this culture is mathematics or rather mathematical physics, which has the longest tradition in carefully using unambiguous concepts.

Herein, we will try to carefully step through the process of creating a dynamic model of a system or process that is assumed to obey the principles of classical physics (to be enumerated later). During this journey we will point out why and how things can go wrong the way they often do. In the introduction of the catalogue of an exposition created by Wim Beeren called ‘Traces of Science in Art’ nuclear physicist Walter H. G. Lewin searches for the reason why art and science have so many parallels. “Both try,” he writes, “to expose new ideas; they seek essence and clarity. The goal is a new opinion, a new way of looking at things which we are not used to. Never, therefore, the familiar – and so confirmational – comfort of a ‘warm bed’, but always the uncomfortable but potentially very revealing ‘cold floor’” [1].

During the following journey through the conceptual world of dynamic behavior, some readers, including Walter Lewin when he would read this, may also feel this cold floor, in particular when we discuss one of Lewin’s lectures that are available on YouTube (Part 1: <http://www.youtube.com/watch?v=eqjl-qRy71w>, Part 2: <http://www.youtube.com/watch?v=1bUWcy8HwpM>) as an example of how ignoring implicit assumptions of a modeling technique may lead to misleading paradoxes. The reader is advised to watch the lectures beforehand and is challenged to answer the question of who was cheating, Lewin or his colleagues in the audience . . .

In order to be able to represent the models and modeling issues that will be discussed in an insightful way, the abstractions will not only be expressed in terms of mathematical equations, but primarily in a graphical notation for port-based models called bond graphs [2]. As many negative prejudices exist about this notation and its use, most of them unjustified [3], we will start by simply defining the notation itself that has a graph-theoretic foundation, while keeping it free from any (physical) interpretation for the time being, even when the terminology seems to induce such an interpretation. At a later stage this notation will be used to explain a specific physical interpretation of models known as the port-based approach to modeling.

## 1.2 Graph-Theoretic Foundation of Bond Graphs

### 1.2.1 Introduction

An often heard argument to reject the ideas represented by bond graphs is that they are not well defined, at least mathematically speaking. Therefore, a unique definition of this graphical language will be given first, while being aware of the fact that each language has its dialects, just like any other mathematical paper that needs a proper list of symbols and a set of definitions too. The existence of a dialect does not mean that the meaning becomes ambiguous: for each case unambiguous definitions should always be possible and preferably made explicit along the same lines as the ones to follow.

### 1.2.2 Bond Graphs

A bond graph is a labeled di-graph – directed graph, i.e., ‘a graph in which each link has an assigned orientation’ [4] – of which the two types of edges are called bonds and signals and the labeled nodes are called multiport nodes or, if no confusion is possible, just multiports, where the number of bonds connected to a node corresponds to the number of ports of a multiport, including the concept of a one-port. Each signal connected to a node with inward orientation is said to modulate the multiport, which in turn is called a modulated multiport, unless no bonds but only signals are connected to the node, in which case the signal is just an input of a so-called block diagram-type node. A block diagram is a computational diagram that only contains signal edges and (functional) block nodes that represent an operator on the input(s). A signal is connected to a signal port. It may depend on the context if the concept of the number of ports in a description of a multiport refers to the total number of ports, including or excluding the signal ports, but the description should make this clear. In case of doubt a clear distinction should be made between ports for signals and ports for bonds. Bonds are represented by augmenting the edge with its direction in the graph by means of a small line that turns the representation of the edge into a half arrow and signals are represented by augmenting the edge with its direction by means of an arrow head, usually two small lines that turn it into a full arrow (Fig. 1.1). The labels can consist of contoured descriptions (word bond

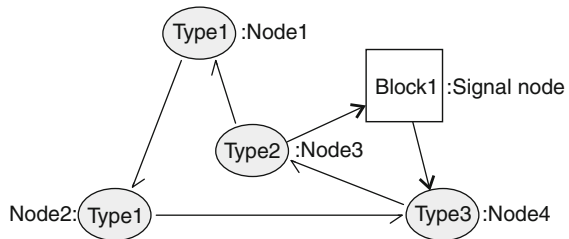


Fig. 1.1 Labeled di-graph with bonds and signals

graph), mnemonic codes consisting of a limited number of alphanumeric symbols, or graphical symbols (in case block diagram symbols represent operations on signals). The node label indicates the *type* of the node. If a particular instantiation of a node type needs to be specified, this is done by means of a colon and a specific name. For linear one-ports this name commonly consists of its constitutive parameter.

Multiple edges of the same type between two nodes may be represented in a combined manner by using condensed edge symbols consisting of two (or more, in case of underlying hierarchy) parallel edges (lines), if necessary with a number (or array of numbers) indicating its dimension(s), i.e., the number(s) of bonds or signals that form this so-called multibond (array) or multisignal (array), respectively. Collections of nodes may be condensed into (a hierarchy of) arrays by underlining. This multibond notation has been defined before [5, 6] and will not be discussed in detail herein as it would distract from the main message and is only meant to allow zooming in order to keep large graphs manageable and insightful.

### 1.2.3 Ports and Bonds

Each port of a multiport, i.e., the interface between a bond and a multiport node, is characterized by two real variables of time that are called effort ( $e$ ) and flow ( $f$ ) and that are considered ‘dynamically conjugated.’ They are scalars in case of single bond edges and vectors in the sense of one-column matrices in case of multibond edges. If their (inner) product is called a power ( $P$ ) these variables are called power conjugated, the bond is called a power bond, and the port to which it is connected is a power port, without linking it to some physical interpretation yet. Bond graphs that only contain power bonds and signals are sometimes called true bond graphs as opposed to the larger class of pseudo-bond graphs in which the efforts and flows are considered to be only dynamically conjugated (terminology to be explained in more detail in Section 1.5.6). The nodes in a true bond graph have true power ports and are connected by true bonds. When it is obvious that no confusion can arise true bond graphs are just called bond graphs and power bonds just bonds. Even though this terminology already suggests some relation with modeling of physical system behavior, this interpretation is deliberately postponed until the discussion of port-based modeling itself. This means that the property of variables being ‘conjugated’ just expresses at this time that they belong to the same port.

The attachment of a signal to a node is called a signal port. There are two types of signal ports: inputs and outputs. A signal interconnects an output to an input. The full arrow of the signal is attached to an input and its open end to an output. This interconnection has a unique meaning: connecting two signal ports by a signal equates the variables on each signal port to those on the signal and thus to each other (Fig. 1.2). This means that the dimensions and units of the variables that are equated by a signal should correspond to each other.

The multiport itself relates all efforts and flows at its ports fundamentally in such a way that each power port has a dependent and an independent (effort or flow)

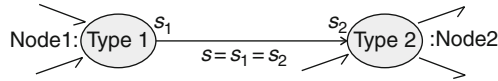


Fig. 1.2 Signal connecting two signal ports in a graph fragment

variable. These relations are called constitutive relations and may also depend on the input signals of the multiport (independent variables of the multiport). Output signals may depend on any of the independent variables of a multiport. In some cases where the desired dynamic behavior of both conjugate variables is taken as a starting point, the constitutive relation itself (constitutive parameter in the linear case) is chosen to be the unknown, such that the fundamental shape of the above relation is violated. However, this does not correspond to a dynamic model anymore, but to the use of a dynamic model structure for changing its dynamic properties in a design context. We will address this anomaly later, as there is a branch of the literature that uses bond graphs for this purpose.

A (multi)bond equates the efforts of the connected ports to each other. Similarly it equates the flows of the ports it connects to each other. As a result, the effort and flow variables can also be considered to live on the (multi)bond (Fig. 1.3). This means that the property of variables being ‘conjugated’ also expresses that they belong to the same bond and that the dimensions of the variables that are equated by a (multi)bond should correspond to each other.

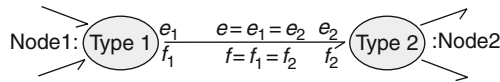


Fig. 1.3 Signal connecting two signal ports in a graph fragment

### 1.2.4 Causal Stroke

From the fact that one of the variables on a port is an independent variable of the constitutive relation and the other one a dependent variable, it follows that the computational direction of the two conjugate variables at a bond is always bi-directional and can in principle be expanded into two opposite signals, in other words, a bond can be considered a bilateral signal flow. The effort is the input of the multiport node at the one side and the output of the node at the other side. Only after a particular choice is made about the two possibilities of these directions, this can be represented by ‘causally augmenting’ the (multi)bond with a so-called causal stroke, a little line drawn orthogonal to the bond at the end of the bond where the effort serves as an output of the bond from a computational point of view. This implicates that it serves as an input for the port that is connected to that side of the bond. Similarly the conjugate flow at that side of the bond is an output of this port and an input to

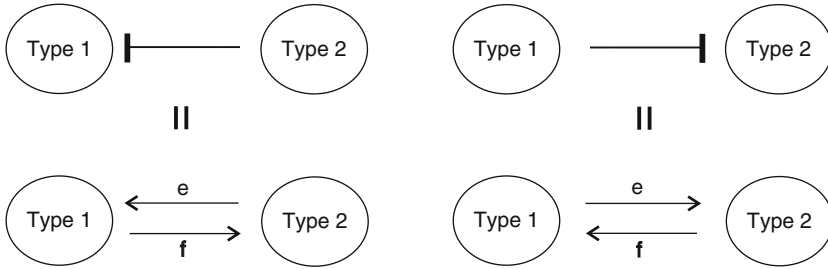


Fig. 1.4 Meaning of the causal stroke: direction of the bilateral signals

the bond. Obviously, the open end of the bond at the other side reverses this story for the port connected to the other side of the bond: effort is output of port and input of bond and flow is input of port and output of bond (Fig. 1.4). The feature of a bond that it can also be drawn without a causal stroke, i.e., without fixing the computational directions of the bilateral signal flow, is one of the main advantages of a port-based approach during modeling and in particular for the re-use of submodels, as it expresses that the causal form of the constitutive relations should not be fixed a priori and left open where possible in order to facilitate interconnection.

A signal represents one arbitrary variable of time that may also be an effort or a flow, but not necessarily. Its computational direction is fixed by definition and represented by a full arrow which clearly distinguishes it from the bilateral signal flow represented by a (power) bond. In the particular case that a signal represents an effort or flow, it can be considered a so-called activated bond, i.e., a bond of which the conjugate variable is negligible and thus considered zero-valued, consequently not playing a role at the side of the edge where there is no arrow and resulting in the inner product of the conjugate variables being zero. As discussed in more detail in Section 1.3 this can be considered as the representation of the concept of an ideal sensor.

Only ports with the same types of efforts and flows can be connected to each other by a bond. A particular type of effort and flow belongs to a domain. Table 1.1 shows a list of domains and the corresponding names of the effort and flow types that belong to that domain. Again we emphasize that, even though the terminology in this table suggests a strong connection with physical system modeling, this connection is postponed until the discussion of physical system modeling itself. In other words, Table 1.1 just defines the terminology and the interconnection constraints, as only ports that belong to the same domain can be connected by a bond. However, as this table is based on a thermodynamic framework of variables (to be discussed in Section 1.5.6) it may be unfamiliar to many readers who have a background in analog system modeling that is commonly using the mechanical framework of variables (Table 1.2) in which some domains are combined to one domain by adding a different kind of state (generalized momentum) [7]. This will be discussed in Section 1.5.

**Table 1.1** Thermodynamic framework of domains and variables

	$f$ flow Equilibrium establishing	$e$ effort Equilibrium determining, intensive state	$q = \int f dt$ Generalized state, extensive state
Electric	$i$ Current	$u$ Voltage	$q = \int i dt$ Charge
Magnetic	$u$ Voltage	$i$ Current	$\lambda = \int u dt$ Magnetic flux linkage
Elastic/potential translation	$v$ Velocity	$F$ Force	$x = \int v dt$ Displacement
Kinetic translation	$F$ Force	$v$ Velocity	$p = \int F dt$ Momentum
Elastic/potential rotation	$\omega$ Angular velocity	$T$ Torque	$\theta = \int \omega dt$ Angular displacement
Kinetic rotation	$T$ Torque	$\omega$ Angular velocity	$b = \int T dt$ Angular momentum
Elastic hydraulic	$\varphi$ Volume flow	$p$ Pressure	$V = \int \varphi dt$ Volume
Kinetic hydraulic	$p$ Pressure	$\varphi$ Volume flow	$\Gamma = \int p dt$ Momentum of a flow tube
Thermal	$T$ Temperature	$fS$ Entropy flow	$S = \int f_S dt$ Entropy
Chemical	$\mu$ Chemical potential	$fN$ Molar flow	$N = \int f_N dt$ Number of moles

An arbitrary node of a bond graph is a multiport, i.e., a node with multiple (power) ports and it may also have signal ports. This multiport can be categorized based on the nature of the relations between the involved (power) port variables and signals. An arbitrary node may be represented in a bond graph by a short description (label) enclosed by an ellipse, resulting in a so-called word bond graph [8]. The only generic properties of this multiport are that it represents the constitutive relations between a number of efforts and a number of flows that are both equal to the number of (power) ports and thus equal itself and that each port has one independent and one dependent variable (either the effort or the flow) in the relation and between a number of arbitrary variables that is equal to the number of input signal ports. Output signal ports result in additional constitutive relations that relate the output to one or more independent variables. For a multiport with  $n$  ports,  $k$  independent flows  $f$ ,  $n - k$  independent efforts  $e$ ,  $r$  outputs  $u$ , and  $s$  inputs  $i$  this can be written as

**Table 1.2** Conventional domains with corresponding flow, effort, generalized displacement, and generalized momentum

	$f$ flow	$e$ effort	$q = \int f dt$ Generalized displacement	$p = \int e dt$ Generalized momentum
Electromagnetic	$i$ Current	$u$ Voltage	$q = \int i dt$ Charge	$\lambda = \int u dt$ Magnetic flux linkage
Mechanical translation	$v$ Velocity	$F$ Force	$x = \int v dt$ Displacement	$p = \int F dt$ Momentum
Mechanical rotation	$\omega$ Angular velocity	$T$ Torque	$\theta = \int \omega dt$ Angular displacement	$b = \int T dt$ Angular momentum
Hydraulic/ pneumatic	$\varphi$ Volume flow	$p$ Pressure	$V = \int \varphi dt$ Volume	$\Gamma = \int p dt$ Momentum of a flow tube
Thermal	$T$ Temperature	$fS$ Entropy flow	$S = \int fS dt$ Entropy	
Chemical	$\mu$ Chemical potential	$fN$ Molar flow	$N = \int fN dt$ Number of moles	

$$\begin{bmatrix} e_1 \\ \vdots \\ e_k \\ f_{k+1} \\ \vdots \\ f_n \\ u_1 \\ \vdots \\ u_r \end{bmatrix} = \begin{bmatrix} e_1(\mathbf{x}) \\ \vdots \\ e_k(\mathbf{x}) \\ f_{k+1}(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \\ u_1(\mathbf{x}) \\ \vdots \\ u_r(\mathbf{x}) \end{bmatrix} \quad \text{with} \quad \mathbf{x} = \begin{bmatrix} f_1 \\ \vdots \\ f_k \\ e_{k+1} \\ \vdots \\ e_n \\ i_1 \\ \vdots \\ i_s \end{bmatrix} \quad (1.1)$$

A categorization of these constitutive relations allows further categorization of the nodes and simpler and more generic labeling. While still making no relation to port-based modeling, these categories will be quite general and when the connection to port-based modeling is made further restrictions can be made, highly depending on the modeling level though, as will be explained when discussing modeling itself. All categories allow the presence of an arbitrary number of outputs (meaning nothing else than making a variable available as independent variable at an input of another multiport). An arbitrary number of inputs is also allowed: this has been introduced already as ‘modulation’ (modulated multiport) and each signal is a modulation signal or modulating signal. Note that one can also consider a multiport as being a node with an arbitrary number of inputs and an arbitrary number of outputs



of which a number ( $n$ ) of inputs are conjugated (paired) to  $n$  outputs while leaving the question open which of the variables in a pair gets the role of input or output as long as the relation is bilateral, i.e., one input and one output. These bilateral pairs are represented by the (power) ports and the other inputs and outputs by the signal ports.

Earlier we noted that in a design context the constitutive relation can be chosen as the unknown, requiring both effort and flow as inputs, which violates the above use of the causal stroke. This refers to the situation that when a given model structure is found, the equations can be rewritten in such a way that the constitutive relation, mostly the constitutive parameter, can be found for a certain desired behavior. The mathematical solution of this kind of question is sometimes mapped on the bond graph by using half causal strokes at both sides of a bond. This graphical approach is addressed by the terminology ‘bi-causality’ and is discussed in more detail in [Chapters 5 and 6](#). However, since it is not clear what the graphical representation adds to the insight of the modeler – on the contrary, it highly confuses many bond graph novices – we will not discuss bi-causality herein any further.

The above defines the principles of a bond graph even though it does not become very useful without a classification of the node types. However, this classification is in part a matter of taste: The more limited a set of nodes is classified, the more powerful conclusions may be drawn from a bond graph, but this also limits the possible models that can be represented at the same time. The most common generic classification is given in the next section. This classification results in particular causal port properties which allow an algorithmic causality assignment. Consequently this assignment can be automated and even be shown during the construction of a graph using a dedicated computer tool like 20-sim ([www.20sim.com](http://www.20sim.com)). In particular causal changes and causal problems during modeling give the modeler immediate feedback on his modeling decisions, which is another powerful aspect of a port-based approach represented by a bond graph. [Section 1.4](#) lists these causal properties and discusses a causality assignment algorithm. Again it is emphasized that in this section no links were made to physical system modeling concepts: only the graphical notation and its interpretation in terms of variables and constitutive relations of various types was discussed. In [Section 1.5](#) physical system modeling is discussed.

## 1.3 Categorization of Nodes

### 1.3.1 Introduction

The following categories of multiport node types can be distinguished in a bond graph:

- *Block diagram nodes*: All nodes that only have signal ports and represent mathematical relations between these signals. These nodes are equal to any type of block diagram definition that exists based on (elementary) mathematical

operations. Their labels commonly consist of rectangles with a (mathematical) symbol that is inspired by the underlying mathematical operation(s). As block diagrams are a well-accepted notation with many dialects, these nodes will not be further categorized explicitly herein. Consequently all other nodes will have at least one (power) port.

- *Power port nodes:* We will discuss the most general form of the constitutive relations of these ports, but in case these relations can be assumed linear, a one-port is characterized by a constitutive parameter and an  $n$ -port by an  $n \times n$ -dimensional matrix. The power will be considered to be the rate of change of a global quantity called ‘energy,’ for which we will assume a conservation principle, anticipating the physical meaning that is commonly given to these concepts. Herein, it is only used as a concept for categorization.

Note that the rest of this categorization assumes that input signals (inputs) are not equal to (a function of) one or more of the outputs (‘port modulation’), even though the bond graph notation allows it in principle. It is easy to see that each of the categories (albeit with an already fixed causality) can be constructed from port modulated sources if port modulation is allowed without further explanation, which would make this categorization more or less useless. However, if a sequence of modeling steps leads to a situation like this, it may create much insight if this kind of identities can be recognized. At the end of Section 1.5.9 an example is given of this situation. We will start with the group of power discontinuous nodes. Due to the assumption of the energy conservation principle these nodes can only reversibly store the energy related to the net power into the node not being equal to zero. This can take place inside the system boundary, resulting in true storage nodes, or in the environment, resulting in sources and sinks, i.e., sources that inject, respectively, absorb power from the perspective of the chosen dichotomy between system and environment. Here it becomes clear that the modeling choice for a particular system boundary is synonymous with the assumption that the storage processes in the chosen environment do not affect the dynamics of the system. In other words, sources can be considered infinitely large storage nodes. The second category we discuss consists of the power continuous nodes.

### 1.3.2 Power Discontinuous Nodes

- *Storage nodes:* All ports of a storage node are storage ports, which means that one of the port variables has to be integrated with respect to time before it plays a role in the constitutive relation of the node or obtained by differentiation with respect to time from a result of the constitutive relation. If the flow variable is integrated with respect to time into a so-called  $q$ -type state variable or if the flow variable is obtained by differentiation with respect to time of the constitutive relation, that is, a function of the effort, the port is called a **C**-type port (or  $q$ -type port). If the effort variable is integrated with respect to time into a so-called  $p$ -type state variable or the effort is obtained by differentiation with respect to time of the con-

stitutive relation that is a function of the flow, the port is called an **I**-type port (or *p*-type port). In case all ports are power ports, there exists a real, scalar function of the independent variables that generates the constitutive relations between the dependent and the independent variables and is equal to the time integral of the net power in the storage node. When the relations are written in a form that only contains integrations with respect to time for each port, the mixed second derivatives of this scalar function (called ‘energy function’) have to be equal in order not to violate the concept of storage. The efforts of *q*-type ports and the flows of *p*-type ports can be found from the energy function by partial differentiation with respect to their conjugate states. Storage nodes have at least one port. In case a storage node has at least one input (signal) port, it is called a modulated storage node, which implies that the concept of storage is violated in principle as the stored energy is changeable while the net power at the port(s) is zero. However, the concept of modulated storage may be useful when the ignored power of the signal port is small with respect to the power at the other ports. In case a storage node only contains **C**-type ports, it is called a (multiport) **C**-element (node label: **C**). In case a storage node only contains **I**-type ports, it is called a (multiport) **I**-element (node label: **I**). When a storage node contains both **C**-type and **I**-type ports, it is called a (multiport) **IC**-element (node label: **IC**). A modulated storage node has node label **MC** or **MI** or **MIC**. Note that port-based modeling may put more restrictions on the modulation of storage nodes, depending on the modeling level, but that the notation allows modulation of storage nodes in principle, as they may be useful when the power related to the modulation is negligible.

Nodes with both storage ports and other ports are not accepted in this categorization in order to keep the labeling (i.e., storage) meaningful later when discussing the concepts used in port-based modeling. Consequently, all other nodes do not contain storage ports.

- *Source nodes*: All dependent port variables of a source node are independent of its independent port variables. This means that the dependent variables are either constant (linear case with one parameter) or the function of an input (modulated source). This means that a multiport source node can always be split into a set of (modulated) one-port sources. When the dependent port variable is an effort the source is called an effort source (node label: **Se**). When the dependent port variable is a flow the source is called a flow source (node label: **Sf**). A modulated source has node label **MSe** or **MSf**.
- *Sensor nodes*: A sensor node allows external observation or availability of one or more independent port variables while the conjugate dependent variables are kept zero. In other words, sensor nodes are zero-valued sources and, as such, need no separate node label, although some dialects use separate labels as this may increase insight given their particular focus.

### 1.3.3 Power Continuous Nodes

- *Resistive nodes*: In this first category of power continuous nodes the power continuity is hidden, as the power entering the resistive ports is converted into thermal power and not explicitly represented by a thermal port, such that energy seems to be ‘dissipated,’ but careful use of concepts shows that only ‘free energy’ can be dissipated and that the use of power as a flow of free energy corresponds to an *implicit assumption*, viz., that the temperature at the thermal port is constant or its fluctuations are slow with respect to the fluctuations of interest, such that the temperature can be considered constant. For a resistive node a semi-positive definite scalar potential function (‘entropy production function’ or ‘dissipation function’) of the independent variables exists that generates its constitutive relations. A resistive node has at least one port. Its node label is **R**. A modulated resistive node has node label **MR**. A resistive node or resistor is sometimes called a dissipative node or dissipator.

It can be proven by means of a linear transformation of the conjugate variables into so-called scattering variables [9, 10] that all power continuous nodes have constitutive relations with a multiplicative form. This means that the vector of dependent port variables can be written as a product of some operator on the vector of independent port variables. When this operator only relates efforts to efforts and flows to flows, a property called ‘non-mixing’ [11], the multiport is called a transformer (node label: **TF**). If the operator is a function of one or more additional node inputs, it is called a modulated transformer (node label: **MTF**). When this operator only relates efforts to flows and flows to efforts, a property called ‘mixing’ [11], the multiport is called a gyrator (node label: **GY**). If the operator is a function of node inputs it is called a modulated gyrator (node label: **MGY**).

- *Reversible transducers*: In case the relations of a **TF** or **GY** are linear, the operator is a constant matrix that is anti- or skew-symmetric due to power continuity. In case the inputs are independent functions of time (externally modulated **MTF** or **MGY**) the anti-symmetric matrix is time variant. In both cases the transduction is reversible in the sense that the sign of the power of each of the ports is always unconstrained, in other words: power can flow in both directions. In case of two-ports the matrix is a  $2 \times 2$ -dimensional anti-symmetric matrix that has only one independent parameter  $n$  for the **TF** or  $r$  for the **GY**:

$$\begin{bmatrix} e_1 \\ f_2' \end{bmatrix} = \begin{bmatrix} 0 & n(\cdot) \\ -n(\cdot) & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ e_2 \end{bmatrix} \quad (1.2)$$

or

$$\begin{bmatrix} e_1 \\ e_2' \end{bmatrix} = \begin{bmatrix} 0 & r(\cdot) \\ -r(\cdot) & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \quad (1.3)$$

By changing the positive orientation of one of the ports, the two constitutive relations of the two ports can be simplified into

$$e_1 = n(\cdot) e_2 \tag{1.4}$$

and

$$f_2 = n(\cdot) f_1 = -f_2' \tag{1.5}$$

for the transformer and

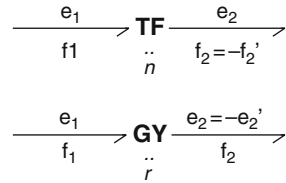
$$e_1 = r(\cdot) f_2 \tag{1.6}$$

and

$$e_2 = r(\cdot) f_1 = -e_2' \tag{1.7}$$

for the gyrator (Fig. 1.5).

**Fig. 1.5** Transformer and gyrator representation



- *Junction nodes*: A junction is a node with power ports that is power continuous and of which the ports are mutually exchangeable without changing its nature: this property is called ‘*port symmetry*.’ Scattering variables can also be used to prove that there exist only two types of power continuous, port symmetric nodes, both with linear constitutive relations (i.e., linearity is not assumed a priori) [9]:

- (1) A **0**-junction or zero-junction (node label: **0**) of which the efforts of all ports are equal

$$e_i = e_j \quad \forall i \neq j \tag{1.8}$$

and the sum of the flows of all ports with a sign in the summation dependent on the direction of the half arrow of the bond connected to the port being zero

$$\sum_{i=1}^n \varepsilon_i f_i \quad \forall \varepsilon_i \in \{-1, +1\} \tag{1.9}$$

where the minus sign holds for bonds oriented away from the junction and the plus sign for bonds oriented toward the junction.

- (2) A 1-junction or one-junction (node label: **1**) of which the flows of all ports are equal

$$f_i = f_j \quad \forall i \neq j \quad (1.10)$$

and the sum of the efforts of all ports with a sign in the summation dependent on the direction of the half arrow of the bond connected to the port being zero

$$\sum_{i=1}^n \varepsilon_i e_i \quad \forall \varepsilon_i \in \{-1, +1\} \quad (1.11)$$

where the minus sign holds for bonds oriented away from the junction and the plus sign for bonds oriented toward the junction.

As a consequence, all other power continuous nodes are port asymmetric. Another result is that modulation of junctions can only take place in the form of a boolean variable that activates or deactivates the node: the junction is then called a switched junction (node labels **X0** and **X1**). This allows a variable interconnection structure.

Note that junctions fall into the category of non-mixing, reversible transducers. They may be seen as a **TF** or **MTF** with a constitutive matrix that contains only 1,  $-1$ , and 0 as matrix elements and modulation consists of changing the absolute values of this matrix. Furthermore, the same holds for multiport substructures that only contain junctions. The transpose of this matrix relates the independent voltages to the dependent ones and thus corresponds to those columns of the reduced incidence matrix of an electrical circuit that relate the link voltages to the branch voltages.

- *Irreversible transduction nodes:* An irreversible transduction node is a resistive node with true power ports with one additional power port of which the flow is equal to the entropy production function of the resistive node. Consequently the conjugate effort equals the power of this port divided by the entropy production function and is called temperature. This illustrates the nonlinear nature of its constitutive relations, such that the earlier conclusion that linear, time-(in)variant transducers are reversible is not violated. Node label: **RS**. In case of modulation (time variance of the entropy production function): **MRS**. Note that an **(M)RS** is an extension of an **(M)R** and when discussing port-based modeling, it will turn out that the **(M)R** rather is a special case of an **(M)RS** which is the result of a modeling assumption (constant temperature or irrelevance of temperature changes) that remains often implicit.

As explained earlier, nonlinearity may be created by port modulation. Therefore, it is possible to use port modulation on an **MTF** or **MGY** to create irreversible behavior, similar to that of the **RS**. While modeling physical systems

this may have the disadvantage that the irreversibility is hidden in the modulation, while designing physical systems with desired behavior it may lead to novel approaches, e.g., damping with reduced generation of heat. Also, this equivalence shows that the constitutive relation of an irreversible transducer has a multiplicative form and does not violate our earlier conclusion as a consequence.

In case the resistive port of the **RS** has a linear relation (the entropy producing port always has a nonlinear relation that relates either the rate of change of the entropy or the temperature to the two independent variables of the **RS**), the two-port **RS** can be characterized by just one parameter  $R$ :

$$\left\{ \begin{array}{l} e = Rf \\ f_{S_{\text{irr}}} = \frac{Rf^2}{T} \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{l} f = \frac{e}{R} \\ f_{S_{\text{irr}}} = \frac{e^2}{RT} \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{l} e = Rf \\ T = \frac{Rf^2}{f_{S_{\text{irr}}}} \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{l} f = \frac{e}{R} \\ T = \frac{Rf^2}{f_{S_{\text{irr}}}} \end{array} \right. \quad (1.12)$$

### 1.3.4 Basic Elements

We now have nine basic bond labels: **(M)C**, **(M)I**, **(M)Se**, **(M)Sf**, **(M)R(S)**, **(M)TF**, **(M)GY**, **(X)0**, and **(X)1**, which, for reasons of clarity can also be introduced bottom-up, in the sense that each is defined in the simplest form possible and where ports are power ports (as mentioned before, in case ports are not power ports, bond graphs are commonly addressed as pseudo-bond graphs with pseudo-bonds). This simplest form is the minimum number of ports and a minimum number of constitutive parameters, which results in the linear form. This results in one-port **C**, **I**, **Se**, **Sf**, and **R**, as well as the two-port **TF**, **GY**, and **RS** that are all characterized by one parameter and the **0**- and **1**-junctions, which are always linear, have no parameter at all. Junctions should have at least two ports to be power continuous, but that would limit the possible structures to chains only. This means that the simplest form of a junction required to be able to build arbitrary structures should have a minimum of three ports. In case the constitutive parameter of an element is replaced by a function of time, the modulated and switched versions of the basic elements, i.e., **MC**, **MI**, **MSe**, **MSf**, **MR(S)**, **MTF**, **MGY**, **X0**, and **X1**, are obtained (cf. Table 1.3).

Note that Paynter [2] originally used only one-letter labels for the node types: **E** instead of **Se**, **F** instead of **Sf**, **T** instead of **TF**, and **G** instead of **GY**. His students Karnopp and Rosenberg [8] noted that interpretation became easier when in some cases two- and three-letter labels were used and they introduced the labeling used herein. In some dialects, e.g., [12], the **0**- and **1**-junctions are replaced by **e**- and **f**-junctions (common effort and common flow junction, respectively) or even using the domain-dependent symbols of effort and flow, like **u** for common voltage junction, **v** for common velocity junction, and **T** for common temperature junction. Thoma [13] uses a dialect that violates our earlier attempt to only focus at the topological structure of relations between concepts: he uses the so-called *s*- and *p*-junctions, where he relates an *s*-junction to a series connection and a *p*-junction

**Table 1.3** Basic node types

Node label	Linear (time-variant) constitutive relation
(M)C	$e - C^{-1}(\cdot) q = 0; q = \int f dt$
(M)I	$f - I^{-1}(\cdot) p = 0; p = \int e dt$
(M)Se	$e = e(\cdot); \frac{de}{dt} = 0$
(M)Sf	$f = f(\cdot); \frac{df}{de} = 0$
(M)R(S)	$e - R(\cdot) f = 0; \left( f_{\text{src}} - \frac{R(\cdot) f^2}{T} = 0 \right)$
(M)TF	$e_1 - n(\cdot) e_2 = 0; f_2 - n(\cdot) f_1 = 0$
(M)GY	$e_1 - r(\cdot) f_2 = 0; e_2 - r(\cdot) f_1 = 0$
0	$e_i = e_j \quad \forall i \neq j; \sum_{i=1}^n \varepsilon_i f_i = 0 \quad \forall \varepsilon_i \in \{-1, +1\}, +1: \text{inward}; -1: \text{outward}$
1	$f_i = f_j \quad \forall i \neq j; \sum_{i=1}^n \varepsilon_i e_i = 0 \quad \forall \varepsilon_i \in \{-1, +1\}, +1: \text{inward}; -1: \text{outward}$
X0	$e_i = \text{if condition then } e_j \text{ else } 0 \text{ endif} \quad \forall j \neq i;$  $f_j = \text{if condition then } \sum_{i=1}^{n-1} \varepsilon_i f_i \quad \forall i \neq j, \varepsilon_i \in \{-1, +1\} \text{ else } 0 \text{ endif}$
X1	$f_i = \text{if condition then } f_j \text{ else } 0 \text{ endif} \quad \forall j \neq i;$  $e_j = \text{if condition then } \sum_{i=1}^{n-1} \varepsilon_i e_i \quad \forall i \neq j, \varepsilon_i \in \{-1, +1\} \text{ else } 0 \text{ endif}$

to a parallel connection. Given that this only has a meaning for physical structures and confuses a topological representation of concepts with a spatial representation, we strongly advise against the use of the terminology in this dialect, as it causes a form of confusion during physical systems modeling that can be quite misleading.

Again it is emphasized that although the terminology (power, energy, entropy, and temperature) seems to imply a link with the physics, the link with the corresponding physical concepts like energy and power will be made during the discussion of port-based modeling. This section lists unambiguous definitions of the most elementary node types. However, any user can add specific nodes to these definitions. The answer whether it makes sense in the context of port-based modeling does not belong in the framework of the description of the notation itself: it is possible to express both sense and nonsense in terms of correct language, and bond graphs are nothing more or less than a graphical language of which we are discussing the semantics and grammar rules. In Section 1.4 we add causal port properties and causality assignment to set the stage for using this graphical language in physical system modeling, but first we shortly discuss how an existing model can be translated or transformed into a bond graph. In Section 1.5 it will become clear that aspects of the causality assignment procedure can also be used during the process of modeling itself as they provide immediate feedback on the modeling decisions being taken.



### 1.3.5 Model Transformation into a Bond Graph

When a model is already available in the form of a domain-dependent iconic representation like an electrical circuit diagram or a mechanical schematic, it can be converted into a domain-independent bond graph by first identifying the domains that are present (step 1) and choosing a reference effort (or velocity) (step 2), next identifying all the port types that are present (step 3) and listing to which effort or effort difference (velocity or velocity difference) they are connected (step 4). All efforts can be represented by a  $\mathbf{0}$ -junction (step 5) and all effort differences can be represented by a  $\mathbf{0}$ -junction that is connected to the two  $\mathbf{0}$ -junctions of which it is a difference by means of a  $\mathbf{1}$ -junction and bond orientations that result in this difference (step 6). Similarly, all velocities can be represented by a  $\mathbf{1}$ -junction (step 5) and all relative velocities ('velocity differences') can be represented by a  $\mathbf{1}$ -junction that is connected to the two  $\mathbf{1}$ -junctions of which it is a difference by means of a  $\mathbf{0}$ -junction and proper bond orientations (step 6). Finally all ports can be connected to this junction structure using the list made during step 4 (step 7). Finally superfluous junctions can be omitted (junction structure simplification; step 8) [14, pp. 297–311]. These eight steps can be used to perform model transformation.

Take, for instance, the model represented in Fig. 1.6 by an iconic diagram which shows an ideal planar pendulum driven by an electric dynamic transducer (motor/generator) connected to a voltage source. The model has four domains in the mechanical framework of variables that we will use, viz., the electromagnetic domain, two orthogonal mechanical translation domains, and the rotational domain. In the thermodynamic framework it would have eight domains, viz., electric, magnetic, three potential, and three kinetic domains for each of the two independent coordinates of the plane as well as the rotation in that plane, but as the iconic diagram model is already based on the mechanical framework of variables, we choose to work herein and show the transition to the thermodynamic framework later. Hence four references (in the mechanical case: reference directions) need to be identified: the electric ground as indicated in the iconic diagram as well as the reference frame with the orthogonal velocities in the  $x$ - and  $y$ -directions that span the plane and the angular velocity in that plane. The already identified elementary behaviors in the electromagnetic domain are (cf. Fig. 1.6) an ideal voltage source, the series resistance of the current loop, i.e., mainly of the motor windings, the series

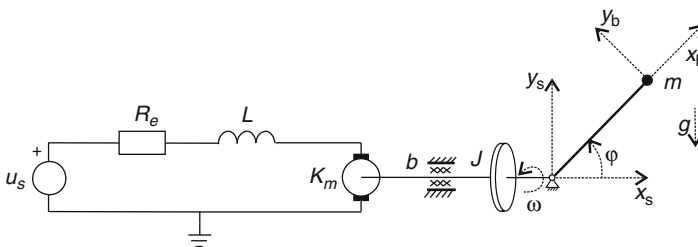


Fig. 1.6 Iconic diagram of example model

inductance of the motor windings, and the electric port of the ideal motor/generator (magnetic field assumed to be constant!). The other port of the motor is in the rotational domain, in which also the motor inertia and the bearing friction can be identified in the iconic diagram. The symbol for the pendulum mass indicates that it is assumed to be a point mass with **I**-type storages in both coordinate directions. If one of these directions is in parallel with the gravitational acceleration, the iconic diagram expresses that there is a constant force of effort (force) on the corresponding **I**-type port, which corresponds to an **Se**-type port.

What remains to be done is to translate the configuration information of the pendulum, i.e., the rigid constraint between the pivot point and the point mass

$$x = l \cos \varphi \tag{1.13}$$

$$y = l \sin \varphi \tag{1.14}$$

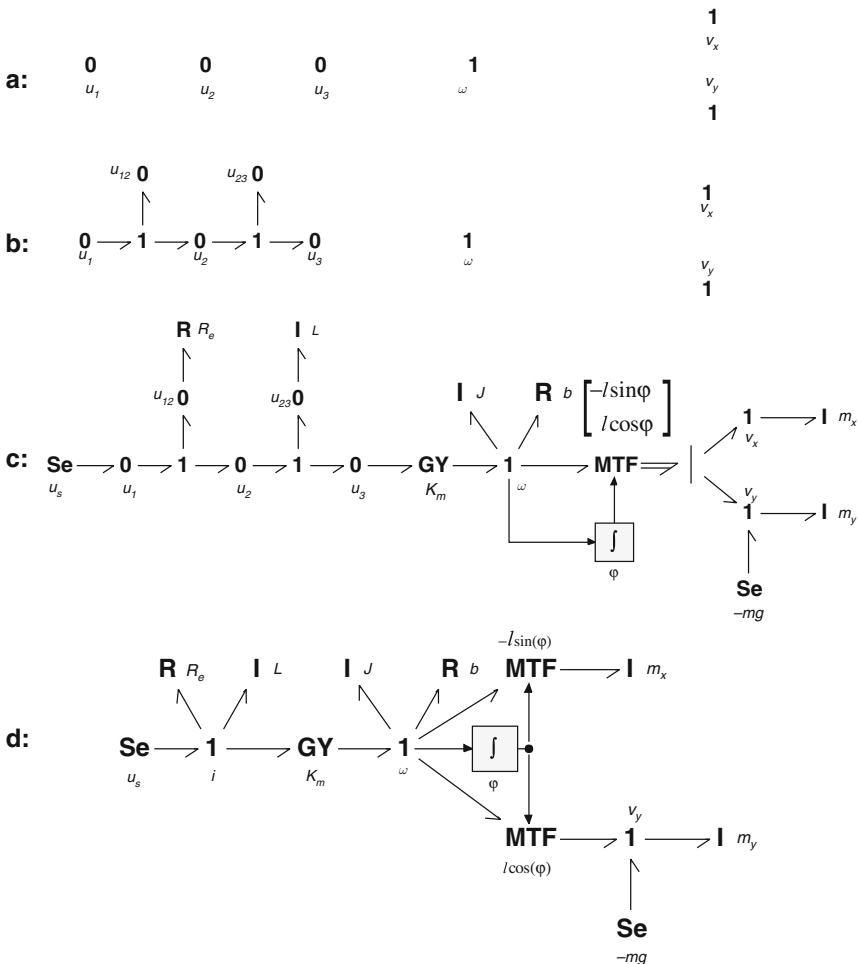


Fig. 1.7 Intermediate steps (a–c) and final bond graph (d) of the example

into velocity relations that generate the remaining part of the junction structure that for the time being can be modeled as a three-port position-modulated transformer (with  $\varphi$  being the modulating state) with flow relations:

$$\dot{x} = (-l \sin \varphi) \dot{\varphi} \quad (1.15)$$

$$\dot{y} = (l \cos \varphi) \dot{\varphi} \quad (1.16)$$

and due to the necessity of power continuity of the constraint:

$$T_\varphi \dot{\varphi} = F_x \dot{x} + F_y \dot{y} = (-l \sin \varphi) F_x \dot{\varphi} + (l \cos \varphi) F_y \dot{\varphi} \quad (1.17)$$

The torque force relation should be

$$T_\varphi = (-l \sin \varphi) F_x + (l \cos \varphi) F_y \quad (1.18)$$

resulting in the  $2 \times 1$  transformation matrix

$$\begin{bmatrix} -l \sin \varphi \\ l \cos \varphi \end{bmatrix} \quad (1.19)$$

that characterizes this  $2 \times 1$ -port **MTF**.

Via the intermediate steps mentioned above (cf. Fig. 1.7a–c) and a straightforward decomposition [15, 16] of the  $2 \times 1$ -port **MTF** into two elementary **MTFs** the final bond graph in Fig. 1.7d is obtained.

## 1.4 Computational Causality and Causality Assignment

### 1.4.1 Introduction

Although there are physical arguments for some of the causal port properties, this section will only address the possible causal port properties and a causality assignment algorithm that is based on these properties. The implications on physical system modeling will be discussed later. The aim of this particular algorithm is to generate a computational structure that is optimally suited for computer simulation, but other algorithms can be defined that have other purposes.

Each port in a true bond graph is characterized by four relevant objects: effort, flow, power, and the constitutive relation between effort and flow that may contain an integration or differentiation with respect to time. In case of pseudo-bonds there are only three relevant objects, effort, flow, and constitutive relation, as the conjugation of effort and flow is not related to power. In case of linearity, the constitutive relation is characterized by just one parameter per port. It depends on the purpose of the model that is being represented by the bond graph which of these objects are independent and which are dependent with respect to a particular port. If the constitutive relation is a known and therefore independent object, either the effort

or the flow has to be an independent variable due to the definition of a constitutive relation, while the (power) conjugate variable is by definition dependent. As a result, the power is a dependent variable too, because it is equal to the product of the power conjugate variables effort and flow. When the constitutive relation is the dependent object – as is the case in the context of design or parameter estimation – both the effort and the flow can be considered independent objects of a port.

Hence, we can distinguish two situations:

1. *Known constitutive relation*: Either the effort or the flow is an independent variable, power can always be computed as the product of effort and flow
2. *Unknown constitutive relation*: Both effort and flow are independent variables

In the first case the effort and flow can be considered bilateral signals at the port (cf. Fig. 1.4). Consequently, the bond connected to that port represents a bilateral signal flow and at the port at the other side of the bond, the roles of effort and flow in the constitutive relation of that port are reversed.

While a bond graph is being constructed a choice has to be made for each port between independent (input) and dependent (output) port variable. If the effort of a bond serves as the input for one of the two ports it connects, it has to be the output of the other port. Due to the bilateral nature of the connection the flow of the first port has to be an output and serves as an input for the second port. When a choice is made, this is indicated in the bond graph by a causal stroke at the end of the bond that is connected to the port where the effort is an independent variable and the conjugate flow a dependent variable. As a result, the open end of the bond (i.e., the end without a causal stroke) is connected to the port where the flow is an independent variable and the conjugate effort a dependent variable (Fig. 1.4). When a bond graph is ready and fully augmented with causal strokes, it can be automatically converted into a set of computable assignment statements. The assignment of this *computational causality* or, if no confusion is possible, just *causality* is an algorithmic process based on the causal properties of each port and/or node type [8, 17]. The graphical representation of this computational structure also gives a modeler immediate feedback on his modeling decisions [18].

In the second case it is possible that at some port both effort and flow are independently given and the constitutive relation (the constitutive parameter of the linear case) is the dependent object. As mentioned before, this situation has been addressed in the literature as ‘bi-causality’ (cf. Chapters 5 and 6). The concept of bi-causality may play a role in design and identification of unknown constitutive relations. In some dialects it is represented by drawing only the upper half of the causal stroke at the port where both effort and flow are given and by putting the lower half of the stroke at the other side of the bond, assuming that the bond is drawn horizontally to be able to explain the position in terms of ‘upper’ and ‘lower.’ Naturally the orientation of a bond in the drawing plane of the graph has no meaning as it is topological, so the concept of ‘lower’ will be defined with respect to a bond in the side of the bond where the little stroke of the half arrow is located. Obviously the concept of ‘upper’ is the other side of the bond. It is clear that in a context of modeling existing dynamic behavior with the purpose of numerical simulation the situation that a constitutive relation is unknown is not relevant and is merely used for

computational purposes related to design and parameter estimation/identification in which one may have doubts about the added value of graphical representation in a bond graph. For this reason, only the first case will be considered in the remainder of this section.

### 1.4.2 Causal Port Properties

Ports may have different causal properties:

- (1) The causality of the port is *fixed*. This occurs in two situations:
  - (a) Either the effort or the flow is a given input of the port, independent of the conjugate variable. This corresponds to the definitions of the **(M)Se** and the **(M)Sf** nodes.
  - (b) The constitutive relation has a form that can only be made explicit in one causal form (non-invertible) any other node except the junctions.
- (2) There is a preference for a particular causality. For example, in case there is a preference for integration over differentiation with respect to time, the ports of the storage nodes, **(M)C** and **(M)I**, may be given a *preferred causality*, viz., effort-out (**C**) and flow-out (**I**), respectively. This will turn out to be the most common preferred causality later, when a causality assignment algorithm is discussed aimed at generating a set of differential equations that is in an optimal form for numerical simulation.
- (3) There is a constraint between the causality of two or more ports of a node. This is the case for the elementary two port nodes **(M)TF** and **(M)GY**: the **(M)TF** always has only one stroke at the node, while the **(M)GY** has either both strokes at the node or both open ends. The junctions also have such a *causal constraint*: an  $n$ -port **0**-junction has one stroke at the node and  $n - 1$  open ends, while an  $n$ -port **1**-junction has one open end at the node and  $n - 1$  causal strokes. The same holds for the **X0** and the **X1** in principle, but in many cases they will be given a fixed causality, given the discontinuous and consequently non-invertible nature of their constitutive relations.
- (4) In all other cases there is *arbitrary causality* or *free causality*, which means that the causality can be chosen freely or is determined by the causal properties of the port at the other side of the bond.

### 1.4.3 Causality Assignment

Causality assignment or causal augmentation is an algorithmic procedure that can be applied to a bond graph based on the causal port properties of its nodes. An important process during this assignment is the *propagation of causality* via causal constraints. This means that when the port of a node with constrained causal port properties gets its causality assigned in such a way that it determines the causality of one or more of its other ports on the basis of the constraint, the causality is

propagated via these other ports, which in turn can cause propagation at the other nodes to which these ports are attached until the process stops. The causal bonds resulting from such a propagation process form a so-called *causal path*. When causality propagation results in a closed causal path or *causal cycle*, in most cases a global constraint can be applied on that cycle that depends on the particular domain represented by the graph.

In short, the following steps are taken:

- (1) Choose an unassigned port with a *fixed causality of type 1a*, assign it, and check that the causality of the port at the other side of the bond, which is assigned as a result
  - (a) either is not in conflict with a fixed causality of that port: If there is a conflict with a fixed causality of type 1a, the bond graph is ill-posed; if there is a conflict with a fixed causality of type 1b, the constitutive relation should be changed into an invertible one or approximated by iteration during numerical solution, or
  - (b) corresponds to the preferred causality of that port: If not, the consequences should be checked based on the origin of the preference, or
  - (c) is propagated via the causal constraint of the port of the other side of the bond to one or more ports for which all these checks should be repeated until propagation ends, or
  - (d) is an arbitrary or free causality.
 

Continue with (1) until all ports with fixed causality of type 1a are assigned.
  
- (2) Choose an unassigned port with a *fixed causality of type 1b*, assign it, and check that the causality of the port at the other side of the bond, which is assigned as a result
  - (a) either is not in conflict with a fixed causality of that port: A conflict with a fixed causality of type 1a cannot occur here, as it should have become apparent during step 1; if there is a conflict with another fixed causality of type 1b, one of the two constitutive relations should be changed into an invertible one, approximated by iteration or the resulting combination of involved constitutive relations should be rewritten analytically, or
  - (b) corresponds to the preferred causality of that port; if not, the consequences should be checked based on the origin of the preference: preferably the constitutive relation should be changed as assigning a non-preferred causality in this manner unnecessarily decreases the order of the set of state equations represented by the bond graph, or
  - (c) is propagated via the causal constraint of the port of the other side of the bond to one or more ports for which all these checks should be repeated until propagation ends, or
  - (d) is an arbitrary or free causality.
 

Continue with (2) until all ports with fixed causality of type 1b are assigned.

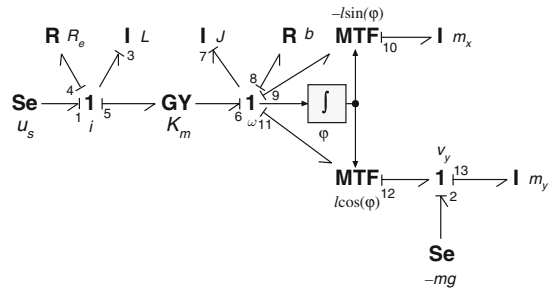
- (3) Choose an unassigned port with a *preferred causality*, assign it, and check that the causality of the port at the other side of the bond, which is assigned as a result
- either corresponds to the preferred causality of that port; if not, this means that the corresponding states are not independent of each other (in case of integral preferred causality) or that the states are not independently controllable (in case of differential preferred causality), or
  - is propagated via the causal constraint of the port of the other side of the bond to one or more ports for which all these checks should be repeated until propagation ends, or
  - is an arbitrary or free causality.
- Continue with (3) until all ports with preferred causality are assigned. If all bonds have obtained their causality after this step a set of ordinary differential equations (ODE) is obtained, if not, the result will be a set of mixed differential and algebraic equations (DAE).
- (4) Choose an unassigned port with an *arbitrary causality*, choose a causality, and check that the causality of the port at the other side of the bond, which is assigned as a result
- either is propagated via the causal constraint of the port of the other side of the bond to one or more ports for which all these checks should be repeated until propagation ends, or
  - is an arbitrary or free causality too; make an inventory of all the algebraic loops that occur as a result of the first choice and compare it with the properties of the algebraic loop(s) that would occur when the first choice would have been different. If solution of the algebraic loops by iteration is no option, the set of constitutive relations that is part of the causal path can be written in an explicit form analytically. Note that the choice of the arbitrary causality may influence the loop gain of the algebraic loop and may be adapted to obtain a set of algebraic relations that is optimal for numerical solution.
- Continue with (4) until all ports with arbitrary causality are assigned.

#### 1.4.4 Causality Assignment Example

Straightforward application of the causality assignment to the bond graph in Fig. 1.7 results in Fig. 1.8, where the numbers refer to the sequential order in which the causal strokes were assigned. Note that the two **I**-type storage elements representing the pendulum point mass do not get their preferred causality. Section 1.5 will explain and solve this issue.

A causal bond graph is commonly synonymous with a set of ordinary first-order differential equations (ODE) that can be automatically generated from the graph when the constitutive relations of each port are well defined (linear relations are

**Fig. 1.8** Causality assignment to bond graph of example model



used by default). In case dependent storage ports or algebraic loops are detected during causality assignment, like in Fig. 1.8, the user knows that the generated set of equations is a mixed set of differential and algebraic equations (DAE) in principle – automated symbolic processes may still generate a set of ODE’s – such that an appropriate choice of a numerical integration method has to be made [19]. Modeling and simulation tools like 20sim can make efficient use of these aspects ([www.20sim.com](http://www.20sim.com)), such that a first simulation result of a modeling idea or a dynamic effect on a model change can be easily obtained, thus providing fast feedback on modeling decisions, which improves the modeling and model debugging processes.

## 1.5 Port-Based Physical System Modeling

### 1.5.1 Introduction

*Modeling* can be defined as the creation of an abstraction of a (part of a) problem and its context in such a way that the abstraction helps the modeler deal with that problem, where a problem should be considered in the most general sense of the word: the need for more insight can be seen as a problem as well. Obviously, there is no such thing as THE model, as each modeler will make his own modeling decisions and it is highly unlikely that these are always the same. Furthermore, even though a man-made device may have been constructed from the perspective of one particular *functional* model, the context of its physical realization will rarely be identical, such that different models are needed that depend on a particular context. This means that it makes no sense trying to find out if one particular model is ‘true’ or ‘false’: the key issue is whether or not the model is *competent* to help the modeler answer his question in the most general sense of the word. In many cases this competence can be quantified by the error between the numerical results of a model and measurements of relevant aspects in the real world that the model tries to capture. However, in some cases a qualitative correspondence in the sense of the nature of the behavior (e.g., oscillation, relaxation, and (in)stability) may already lead to the solution of a problem without considering numerical errors.



### 1.5.1.1 Intended Behavior Versus Parasitic Behavior

In case a model needs to be created for trouble-shooting purposes of a man-made device with an intended functional behavior, the designer of that device is often not the most successful modeler as the problems often result from the difference between the behavior of the actual physical realization and the conceptual picture of reality that the designer used to create a particular functionality and that is a model as well. The trouble-shooter should model the physical effects that are ‘parasitic,’ i.e., not contributing to the intended functionality. Parasitic effects are effects that are neglected initially, yet influence the behavior in such a way that they should be included in a model to be able to capture the behavior that is of interest. They should be made part of the model in order to be able to solve the problem at hand.

### 1.5.2 Basic Physical Principles

In case of the general context of physical systems, i.e., systems of which the behavior is supposed to obey the basic principles of (classical) physics, a model of the dynamic behavior of a physical system is also supposed to correspond to these basic principles like *energy conservation*, and *positive entropy production* and the *conservation principles* of the generic physical quantities like *momentum*, *electric charge*, *magnetic flux* (linkage), and *entropy* (while allowing that entropy is produced as well!). These principles can be considered generic modeling assumptions.

### 1.5.3 Modeling Versus Model Transformation

Many textbooks that claim to discuss modeling ‘pollute’ the picture of modeling sketched above, as they do not describe the process of modeling itself, but the process of transforming one model description (usually in the form of natural language supported by sketches or diagrams) into another one, usually in the form of a set of (differential) equations or a block diagram representing a computational structure or even the input code of a specific simulation package. The modeling decisions themselves, i.e., the decisions about the physical phenomena that should be represented by the model in order to obtain a competent description of reality in a particular problem context, commonly do not play a large role. For instance, when Feynman et al. [20] in his Lectures on Physics part 1 Fig. 4.7 shows a pendulum, his text makes clear that his ‘pendulum’ is already a model consisting of ideal concepts: a point mass, an infinitely stiff, massless constraint, and an ideal pivot point. The transformation into equations then becomes a process with a unique result. Zill [21, p. 209] makes a difference between a ‘physical pendulum’ – any object that swings back and forth – and a ‘simple pendulum’ for which he explicitly states most assumptions. However, his concept ‘mass’ is implicitly assumed to be a point mass and his concept ‘rod’ is implicitly assumed to be infinitely stiff. He then generates the corresponding differential equation (= model transformation)

by writing the constraint of the rod in terms of accelerations and relates this via Newton's second law to the tangential force. By referring to his sketch he equates this force to the tangential component of the gravitational force and thus obtains a nonlinear second-order differential equation in the angular displacement of the pendulum from its equilibrium position. No attention is paid to the background of the modeling decisions when they cease to lead to a competent model and this is typical for many descriptions of a modeling process.

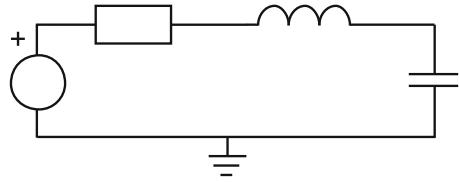
The transformation of the iconic diagram of the simple pendulum model into a second-order differential equation is unique, given a choice of relevant variables (angular, tangential, horizontal, or vertical displacement) and parameters (pendulum length and mass, gravitational acceleration). It is a suitable exercise in model transformation for engineering students, but has nothing in common with the modeling process itself, i.e., the decision under which circumstances this model is a competent abstraction from reality that helps answer a question. As it is easier to create an appearance of objectivity in exams by testing model transformation abilities on the basis of one possible result, while model decisions are much less strict and by definition subjective to some extent, most educational systems that have deviated from the master-pupil model for efficiency reasons have also deviated from teaching the actual process of modeling as a consequence.

Before the time of computer simulation, many could counter this criticism by demonstrating that bringing in more complex aspects into a model leads to differential equations that cannot be solved analytically, giving not much room for competence to help to solve a problem. In other words, the availability of solvable model structures was relatively small and the decision process restricted to the selection of a model structure that would best capture the data. However, using computer simulation, the behavior of much more complex models can be studied, which means that carefully taken modeling decisions and explicit assumptions are a key to competent models for real-world situations.

### ***1.5.4 Model Representations***

Apart from mathematical model representations in terms of differential equations, many graphical model representations exist: block diagrams and signal flow graphs are used to represent the mathematical structure of a model, Venn diagrams and Petri nets represent the logi(sti)c structure of systems in which the events are considered to be discrete in time, electrical circuit diagrams represent both the relations between the electrical properties of physical components and the relations between conceptual elements. Iconic diagrams similar to circuit diagrams exist for the mechanic, hydraulic, pneumatic, and thermal domains, even though the symbols are often not as well defined as in the electrical case, the picture of a simple pendulum model being an already used example. In all cases, however, the (arbitrarily) chosen metric of the topological graph has no meaning: The location of the nodes with respect to each other is commonly chosen such that the picture gives an organized impression.

**Fig. 1.9** Simple electrical circuit diagram



For instance, in an electrical circuit diagram the unwritten, heuristic rule is that the reference node is at the bottom, nodes related to sources at the left, and nodes related to loads at the right, while the nodes are kept as much as possible at a grid and the symbols of the labeled edges are connected to the nodes by straight lines (Fig. 1.9).

By contrast, in a linear graph [22], which can be seen as an abstraction of a circuit diagram, the edge symbols are replaced by edge mnemonic codes that are not part of the edge anymore, while the edges are commonly drawn as curved lines. In his earlier mentioned lecture with the challenging title ‘complete breakdown of intuition,’ Lewin incorrectly implies that the arbitrary area enclosed by the edges of a circuit diagram can be related to an actual area enclosed by a current-carrying wire, a current that can be (among other things) the result of a magnetic field surrounded by this wire. Note that the arbitrary area is only present in the graph in order to represent the nodes separate from each other and that the layout can be changed without changing its meaning. The influence of a magnetic field on an actual (‘spatial’) area enclosed by actual wires thus has to be separately represented in a graph (circuit diagram) by the concept of an ideal inductor (‘conceptual concentration’), which would have prevented the paradoxical conclusions Lewin makes and which even makes him state that ‘Kirchhoff is for the birds.’ Even though this may have been an educational trick to draw the attention of the audience, downplaying famous scientists like Kirchhoff and even downplaying his own colleagues of whom he says in a suggestive way that their correct claim that he is cheating ‘is telling you something about them’ does not seem to fit in this explanation. The point where he is actually cheating is when he claims that the voltmeters are connected to the same points: he has connected the voltmeters to different points in space, viz., immediately across the spatially separate resistors, interpreting the wires as ideal connectors which can then be considered to be ‘the same points’ from the perspective of electric potentials. However, the wires between these points constitute the majority of the winding that picks up the induced current, thus generating an induced voltage between his points of measurement. Like one of the comments on YouTube puts it: ‘There is no problem in interpreting the flux term in Faraday’s equation as an EMF, it is done historically and everyone knows that is not a gradient of a scalar potential.’ (EMF stands here for electromotive force, the induced voltage, and is often called back-EMF when this voltage is generated by a spinning motor and thus a measure for its speed.)

The reason for spending so many words on this issue is that these kinds of confusions are deeply rooted in our scientific culture. A good way to prevent oneself from falling into the traps of confusing modeling with model transformation, of

confusing spatial information with a topology representing conceptual structure and unjustified extrapolation of existing models is to use the following checklist:

**Checklist for modeling physical systems:**

- Make problem context explicit as much as possible and iterate at any time when needed
- Make implicit assumptions explicit
- Keep record of all assumptions being made
- Keep the spatial configuration information separate from the topology that represents conceptual structure, even if the variables happen to coincide

Note that the example of the model with the ideal pendulum in Fig. 1.6 is an example of a situation where the position variable (the angular position  $\varphi$ ) merely describes a configuration state.

Next we study the conceptual foundations of each dynamic model.

### ***1.5.5 The Concepts of System Boundary and Environment***

The system boundary separates the system or subsystem from its environment. The environment of a system is that part of the rest of the universe that has some form of relation with the system. The environment can influence the behavior of the system, but not its dynamic characteristics. The system boundary is the boundary between a system and its environment that can be defined on the basis of a boundary criterion. Common boundary criteria are based on the concept of state that will be discussed first.

### ***1.5.6 The Concepts of State, Equilibrium, and Change of State***

All dynamic models are based on the concept of state. States can be divided into extensive states, i.e., states that are proportional to the extent (either spatial, material, or both) of the physical object of which the behavior is described, and intensive states, i.e., states that are independent of the extent of this physical object. An extensive state can be related to a conservation principle (momentum, charge, matter, magnetic flux, displacement, etc.), i.e., if  $q$  is a state it satisfies the property  $\oint dq = 0$ , and its dialectic complement, the ‘rate of change of state’,  $dq/dt = f$  is generally considered a flow of the corresponding quantity in a general sense, as conservation is also interpreted in the sense that an exchange of a stored quantity has to take place via the intermediate space. In other words it is assumed that conservation is not achieved by simultaneous annihilation and generation at the same rate. The state and its rate of change are considered dialectic complements as a state has no meaning when it cannot change and a flow has no meaning if it does not result in change of state, even when this state is not included in the model. The latter

remark refers to the situation that storage of ‘what is flowing’ is not always relevant to understand behavior. This is the case when constant (= ‘stationary’) flows play a role in a model. In other words, flows can exist that describe an exchange with a non-storing concept, i.e., they are not identical to a rate of change of state, although they may contribute to it. This process of shaping concepts can be seen as ‘conceptual concentration’: even though the flow is in principle always related to a change of state, it may be decoupled from it by concentrating the change of state in a particular concept (‘storage’) and concentrating the description the contribution to the change of state in a separate summation of contributions, i.e., flows that are not a time derivative of a state individually, but contribute to it via a so-called balance equation. This approach allows that the storage related to flows is not part of the model, but of the environment, such that there is no need to describe it as storage.

A system is said to be in its equilibrium state when the system is not exposed to changing environmental influences (there may be a constant influence though) and all the rates of change of state are zero and consequently all states constant. This means that flows can be considered ‘*equilibrium-establishing variables*’: the net result of contributing flows to a rate of change of state is unequal to zero until the equilibrium state is reached. This leaves room for some flows not being zero, but having a constant value in equilibrium. In other words, the concept of a flow can be generalized by realizing that each contribution to a rate of change of state in a balance equation can be considered a flow, such that a flow itself does not have to be equal to a rate of change itself. This means that an equilibrium situation can occur if the sum of the non-zero, constant flows in a balance equation results in a zero rate of change of state. Such a special case of equilibrium is called a stationary state.

When a state of an object is identified, it is considered the same for the whole extent of the object or concept being considered. The extent itself is also a state, even though it is often constant. When we increase the extent by bringing two objects together that are in mutual equilibrium, i.e., do not result in flows when left to interact freely with each other, the intensive state remains the same by definition. In other words, the equilibrium between two objects is determined by the intensive states being equal. On the other hand, a difference in intensive state(s) will generate (a) change(s) of state or flow(s) until an equilibrium is reached. This brings us to the heart of the description of dynamic physical processes: flows that establish equilibrium and intensive states or efforts that determine equilibrium and thus can be considered ‘*equilibrium-determining variables*.’ As a consequence, efforts and flows are said to be dynamically conjugated. Note that this distinction for the concepts of effort and flow is available for a thermodynamic type of variable classification (cf. Table 1.1). We will see later that this distinction is lost in the mechanical framework of variables (cf. Table 1.2) where the potential and kinetic domains are combined into one domain, which is unfortunate for the process of identifying concepts during modeling. Combined with checks of units and dimensions, the check on the nature of physical variables forms a powerful instrument for taking modeling decisions and checking the consistency of models.

### 1.5.7 Boundary Criteria

The states that are commonly used as boundary criteria are ‘amount of matter,’ e.g., not only the fixed amount of matter in a rigid body, but also a fixed amount of fluid in Lagrangian coordinates or the gas in a closed balloon, and ‘available volume,’ e.g., not only a container with a fixed volume, but also a ‘control volume’ for the study of fluid flow in Eulerian coordinates. In case of the ‘matter criterion’ the system consists of a fixed amount of matter or rather the boundary is defined as a border that does not allow exchange of matter. In case of the ‘volume criterion’ the system consists of a fixed volume or rather the boundary is defined by a border that does not allow volume changes. Note that the two latter formulations are more general as they allow mixed boundaries like fluid flowing through a piece of flexible tube: the openings are spatially fixed with respect to the wall boundaries and allow no change of volume, while they do allow exchange of matter (fluid), while the tube wall does not allow an exchange of matter, but it does allow a change of available volume due to the flexibility of the tube. In principle, other conserved quantities like electric charge could serve as a boundary criterion too, but this is seldom used.

In case of a spatial structure of components that are modeled as a network structure, the network structure is only constant if the subsystem boundaries maintain their relative locations with respect to each other, in other words, each subsystem should keep the same neighbors. This means that material subsystem boundaries can only be used when they do not violate the network structure. For instance, rigid bodies in a chain structure should not pass each other. In case the volume criterion is used to define a subsystem boundary a similar constraint holds that is automatically satisfied if all volume boundaries are defined with respect to the same spatial reference.

### 1.5.8 The Port Concept

Due to the dynamic conjugation of *intensive state* (equilibrium-determining variable or effort) and rate of change of *extensive state* (equilibrium-establishing variable or flow), the concept of a ‘port’ can be introduced as the medium of exchange of the concepts effort and flow with other conceptual entities that represent dynamic properties of physical objects. This way of looking at relationships between physical concepts can be seen as a paradigm change with respect to the input–output way of thinking that is induced by modeling in terms of block diagrams (with fixed, explicit causality) as well as with respect to the completely a-causal description of, for instance, balance equations. The possibility to represent the chosen causality for a port in a bond graph as an algorithmic consequence of a particular interconnection, not only generating a format of the model equations well suited for simulation but also providing feedback to the modeler about his modeling decisions, is a feature that makes a bond graph extremely well suited to construct and to represent port-based models.

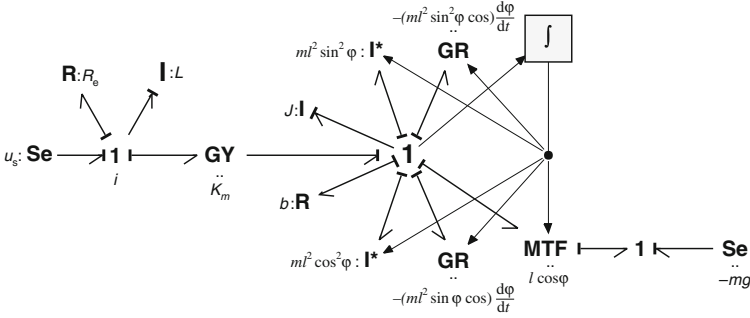
### 1.5.9 Bond Graph Construction

A bond graph can be written during modeling in a manner similar to the model transformation into a bond graph representation:

Identify domains, choose a reference per domain, and identify common equilibrium-determining variables. These are commonly efforts, but if the classical mechanical framework of variables is used the kinetic effort, i.e., the velocity, is *dualized* into a flow. Represent common efforts by **0**-junctions and common velocities by **1**-junctions. Identify the basic concepts that are needed to represent the relevant phenomena per domain in terms of ports. For instance, an ideal transformer that connects two domains has two ports and each of them has to be identified. Next it has to be determined to which effort (velocity) or effort difference (velocity difference) a port is to be connected. An effort difference can be constructed by means of a **1**-junction and a velocity difference can be constructed by means of a **0**-junction, but all effort differences should be explicitly represented by **0**-junctions, while all velocity differences should be explicitly represented by **1**-junctions. Next all ports can be connected to **0**-junctions (or **1**-junctions in case of mechanical ports) and the bond graph can be simplified when possible. Note that the reference effort or velocity is not represented by a junction (as it would result in an unnecessarily complex graph that leads to the same result after simplification), unless one is interested in the value of the conjugate variable acting on the reference (e.g., the reaction force or the ground current).

Returning to the causal bond graph of the example in Fig. 1.8 we can now conclude that the differential causality of both **I**-elements results from the rigid constraint between the rotor inertia and the point mass of the pendulum in the two degrees of freedom of the plane. Even without writing equations the modeler gets this feedback from causality assignment and since this can be automated this provides a means for immediate feedback on modeling decisions while drawing the bond graph. It immediately shows that infinitely rigid, massless mechanical constraints are in principle not physically realizable, but that it still may be a good decision to neglect the elastic behavior of a stiff mechanical connection as it introduces high frequent behavior in the model that may not be relevant for the problem context and is at the same time costly for numerical integration. In case the transformers would have a constant transformation ratio (no modulation), the bond graph allows for a straightforward model transformation of both dependent **I**-elements to the same junction as the rotor inertia, with which they can be combined into one **I**-element with integral causality, an operation analogous to application of the Huygens–Steiner theorem (e.g., [23]). However, in case of modulated transformers an additional gyrator [24] is generated (cf. Fig. 1.10) for both  $x$ -direction

$$T_x = (-l \sin \varphi) \frac{d}{dt} (-ml \sin \varphi) \dot{\varphi} = \frac{d}{dt} (ml^2 \sin^2 \varphi) \dot{\varphi} - (ml^2 \sin \varphi \cos \varphi) \dot{\varphi}^2 \quad (1.20)$$



**Fig. 1.10** Virtual inertia's and gyristors resulting from a transformation of the I-elements to the body-fixed frame in the example of Fig. 1.7

and y-direction

$$T_y = (l \cos \varphi) \frac{d}{dt} (m l \cos \varphi) \dot{\varphi} = \frac{d}{dt} (m l^2 \cos^2 \varphi) \dot{\varphi} + (m l^2 \sin \varphi \cos \varphi) \dot{\varphi}^2 \quad (1.21)$$

However, in this particular case of rotation, the sum of the two time-dependent virtual inertias becomes a true (constant) inertia

$$\frac{d(m l^2 \sin^2 \varphi) \dot{\varphi}}{dt} + \frac{d(m l^2 \cos^2 \varphi) \dot{\varphi}}{dt} = \frac{d m l^2 \dot{\varphi}}{dt} \quad (1.22)$$

such that the total inertia becomes

$$J + m l^2 \quad (1.23)$$

and the gyristor contributions compensate each other as a consequence:

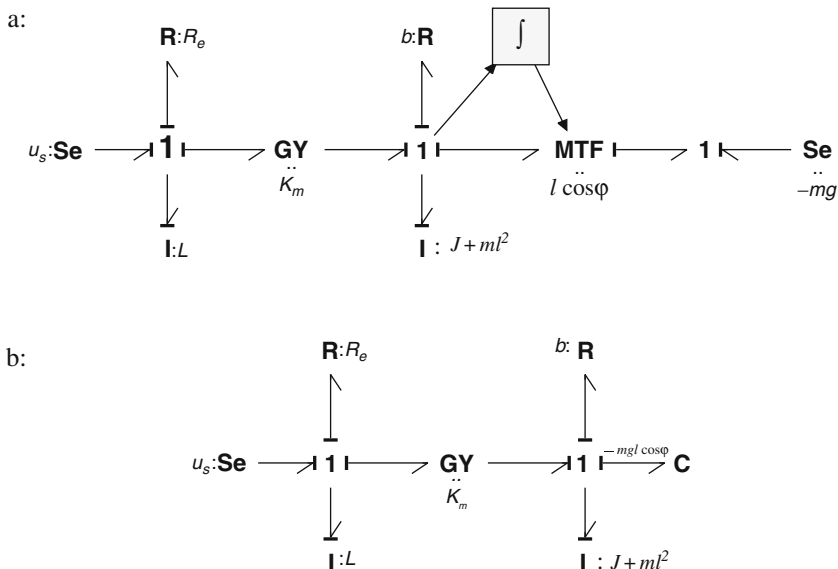
$$-(m l^2 \sin \varphi \cos \varphi) \dot{\varphi}^2 + (m l^2 \sin \varphi \cos \varphi) \dot{\varphi}^2 = 0 \quad (1.24)$$

Note that if the constraint would be elastic, the gyristors (**GR**) would stay in the model and represent the fictitious forces like the centrifugal force (in case of a rigid constraint, the corresponding velocity and thus the corresponding power is zero, such that the contribution becomes irrelevant for the behavior).

The result of transforming both **I**-elements leaves only the **Se**-element representing gravity at the other side of the transformer. Taking a closer look at the combination of configuration state-modulated **MTF** and **Se** shows that internal port-modulation results in a nonlinear **C**-element (cf. Fig. 1.11):

$$T_C = -T_g = -(l \cos \varphi) F_g = -(l \cos \varphi) (-mg) = m g l \cos \varphi \quad (1.25)$$





**Fig. 1.11** Resulting inertia combined with Se via a port-state-modulated transformer (a) resulting in an IC second-order loop and (b) with nonlinear C

The causal path between the **I** and the **C** immediately shows the potential periodic behavior of the pendulum. For small angles around  $-\pi/2$  this **C**-element even becomes a linear **C**-element with parameter  $1/mgl$  and the relation for the oscillation frequency  $f$  can be seen directly from the graph to be the common result:

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{IC}} = \frac{1}{2\pi}\sqrt{\frac{mgl}{J + ml^2}} = \frac{1}{2\pi}\sqrt{\frac{g}{\frac{J}{ml} + l}} \tag{1.26}$$

When compared to immediately writing a damped second-order model in the angle  $\varphi$  for the pendulum, one may observe that the final result is identical, but that the approach allows for physically motivated model extrapolation, e.g., a ‘spring-pendulum’ or a ‘pinned pendulum’ as well as connection to other domains like the electric drive in the example. The example also shows how bond graph construction provides immediate feedback on modeling decisions based on the causal port properties of the concepts used in the model to represent the modeling choices, based on energy conservation and power continuity that is built in the notation. Next the role of energy conservation is further explored for multiport storage.

### 1.5.10 Energy Conservation

For physical systems a function  $E(\mathbf{q}) = E(q_1, \dots, q_n)$  of the (conserved) extensive states  $q_1, \dots, q_n$  can be defined that in turn is conserved. This conservation is related to the evident requirement that the models we make are symmetric with respect to time translation, in other words, a dynamic model should result in the same behavior when starting from a particular initial state, independent of the exact moment in time. Obviously, without such a constraint, modeling would be a senseless activity. Due to the principle of energy conservation, the energy is an extensive state too, albeit a function of all other relevant extensive states. In principle, it is a first-degree homogenous function of the extensive states as a consequence of being extensive. A homogenous function  $F(\mathbf{x})$  satisfies the following property:

$$F(\alpha\mathbf{x}) = \alpha^n F(\mathbf{x}) \quad (1.27)$$

where  $\alpha$  is an arbitrary parameter and  $n$  is the degree, also called order, of the homogenous function. Consequently, if  $\alpha$  systems with energy  $E(\mathbf{q})$  that are in mutual equilibrium are combined, the energy  $E(\alpha\mathbf{q})$  of the resulting combination is equal to  $\alpha$  times the energy  $E(\mathbf{q})$ , i.e.,  $\alpha^1 E(\mathbf{q})$ , so the energy is indeed a first-degree homogenous function in principle. Note that if there would not be mutual equilibrium, the resulting energy depends on the process of combining and thus would change the model.

In many modeling cases energy functions are used that are not first-degree homogenous, due to keeping one of the states constant and not considering it a state anymore, but a constitutive parameter. If one takes the example of an amount of matter moving with respect to some inertial reference, the extensive states are its amount of moles  $N$  and its momentum  $p$ . The (kinetic) energy is

$$E_{\text{kin}}(p, N) = \frac{p^2}{2MN} \quad (1.28)$$

where  $M$  is the molar mass (a material parameter). Since

$$E_{\text{kin}}(\alpha p, \alpha N) = \frac{\alpha^2 p^2}{2M\alpha N} = \alpha \frac{p^2}{2MN} = \alpha^1 E_{\text{kin}}(p, N) \quad (1.29)$$

the kinetic energy is a first-degree homogenous function and thus an extensive quantity. However, when using the concept of a rigid body, the constant amount of moles is commonly not considered a (constant) state anymore, but part of the constitutive parameter 'mass'  $m = MN$ , such that the kinetic energy becomes a second-degree homogenous (quadratic) function of the momentum  $p$ :

$$E_{\text{kin}}(\alpha p) = \frac{\alpha^2 p^2}{2m} = \alpha^2 E_{\text{kin}}(p) \quad (1.30)$$

and in this case equal to the negative value of its Legendre transform  $\mathcal{L}$  with respect to  $p$

$$\mathcal{L}(E_{\text{kin}}(p))_p = \frac{\partial E_{\text{kin}}(p)}{\partial p} p - E_{\text{kin}}(p) = vp - E_{\text{kin}} = \mathcal{L}(v) \quad (1.31)$$

i.e., the complementary energy or co-energy

$$E_{\text{kin}}^*(v) = -\mathcal{L}(v) = vp - E_{\text{kin}} = \frac{1}{2}mv^2 \quad (1.32)$$

where  $v$  is the velocity of the body

$$v = \frac{\partial E_{\text{kin}}(p)}{\partial p} = \frac{p}{m} \quad (1.33)$$

The partial derivatives of the energy with respect to the states  $\mathbf{q}$

$$\frac{\partial E(\mathbf{q})}{\partial q_i} = e_i(\mathbf{q}) \quad (1.34)$$

are called *efforts* and are, again in principle (!), zero-degree homogenous functions of the extensive states and are intensive states as a consequence:

$$e_i(\alpha\mathbf{q}) = \frac{\partial E(\alpha\mathbf{q})}{\partial \alpha q_i} = \frac{\alpha \partial E(\mathbf{q})}{\alpha \partial q_i} = \alpha^0 \frac{\partial E(\mathbf{q})}{\partial q_i} = \frac{\partial E(\mathbf{q})}{\partial q_i} \quad (1.35)$$

Note that the velocity is an intensive state in the kinetic domain, i.e., the domain linked to the extensive momentum state, and thus the equilibrium-determining variable of the kinetic domain, and that there is an intensive state of the material domain  $\mu_{\text{tot}}$  (total material potential) too:

$$\mu_{\text{tot}}(p, N) = \frac{\partial E_{\text{kin}}(p, N)}{\partial M} = -\frac{p^2}{2MN^2} = -\frac{Mv^2}{2} = \mu_{\text{tot}}(v) \quad (1.36)$$

The latter relation between the two intensities can also be seen as the result of the fact that the total Legendre transform of a first-degree homogenous function is equal to zero:

$$L(E(\mathbf{q})) = E(\mathbf{q}) - \sum_i \frac{\partial E(\mathbf{q})}{\partial q_i} q_i = E(\mathbf{q}) - \sum_i e_i q_i = L(\mathbf{e}) = 0 \quad (1.37)$$

This means that one of the intensities is a function of the other ones or, in other words, the number of independent intensive states is one less than the number of extensive states. This makes sense, since the information about the extent itself is omitted from the intensive description. If we would have assumed in our example

that the density of the matter  $\rho$  is constant, the mass can also be described by the volume  $V$  it consumes, i.e.,  $m = \rho V$ , such that

$$\frac{\partial E_{\text{kin}}(p, V)}{\partial V} = \frac{\partial \frac{p^2}{2\rho V}}{\partial V} = -\frac{p^2}{2\rho V^2} = -\frac{\rho v^2}{2} \quad (1.38)$$

The latter term is equal to minus the so-called dynamic pressure that is used to describe fluid flow.

Note that

$$E(\mathbf{q}) = \sum_i \frac{\partial E(\mathbf{q})}{\partial q_i} q_i = \sum_i e_i q_i \quad (1.39)$$

is a property of any first-degree homogenous function (in thermodynamics called Gibbs' equation), while

$$\begin{aligned} d\mathcal{L}(E(\mathbf{q})) &= dE(\mathbf{q}) - d \sum_i e_i q_i = \\ &= \sum_i e_i dq_i - \sum_i e_i dq_i - \sum_i q_i de_i = - \sum_i q_i de_i = 0 \end{aligned} \quad (1.40)$$

The relation

$$\sum_{i=1}^n q_i de_i = 0 \quad (1.41)$$

is called the Gibbs–Duhem relation in thermodynamics [25], where  $n$  is the number of extensive states involved, often written for a constant total amount of moles  $N$  in the form

$$\sum_{i=1}^{n-1} \frac{q_i}{N} de_i = 0 \quad (1.42)$$

where  $q_i/N$  is the molar density of state  $q_i$  (which becomes a molar fraction when this state is another species), i.e., an intensive quantity, which is reflected in the reduction of the amount of involved efforts to  $n - 1$ , while only  $n - 2$  efforts remain independent due to this relation which has the nature of a weighted sum.

When the conserved extensive states change during a dynamic process by exchange due to flows, energy is exchanged too. This means that the ('local') rate of change of the energy, called power  $P$ , is a relevant physical variable too

$$P = \frac{dE(\mathbf{q})}{dt} = \sum_i \frac{\partial E(\mathbf{q})}{\partial q_i} \frac{dq_i}{dt} = \sum_i e_i f_i = \mathbf{e}^T \mathbf{f} = \mathbf{e} \cdot \mathbf{f} \quad (1.43)$$

where we see that these particular intensities (efforts) and rates of change of extensities (flows) are not only dynamically conjugated but also *power conjugated*. Note that, in addition to the energy conservation principle, also Heaviside’s principle is assumed here, viz., that the energy has to transverse the intermediate space and that the conservation is not maintained by annihilation in one ‘place’ and generation at the same rate in another ‘place.’ The concept of ‘place’ is put between quotes in order to emphasize that these ‘places’ may be spatially separated locations, but may also indicate that the separation is merely conceptual, such that an interconnection structure may be related to a spatial structure, but may also refer to a structure of conceptual relations. Abstracting away from physical space often appears to be a conceptual bridge too far for many, leading to all sorts of conflicting interpretations and paradoxes, as will be discussed in more detail in Section 1.5.14.

### 1.5.11 The Power Port Concept and the Bond Graph Notation

In the above treatment a link was made between the physics and the variables effort, flow, and power, for which we defined the bond graph notation earlier. This means that a link between physical concepts and the bond graph notation can be made in a straightforward manner. Given the definition of a C-type port in subsection 1.3.2, Fig. 1.12 shows that the energy can be assumed to be stored in a multiport C-element with  $n$  ports of which the efforts are equal to the partial derivatives of the energy with respect to the extensive state variable stored via the corresponding port and the conjugate flows are equal to the rates of change of this state. This means that the energy function is a generating function of the  $n$  constitutive relations  $e_i(\mathbf{q})$  with  $dq_i = f_i dt$ . In the graph, the generating function is distinguished from the constitutive parameter or relation by the use of a double colon (::) instead of a colon. In order to create a flow when some difference exists between equilibrium-determining variables (efforts), it is necessary that other types of ports than storing ports can be described (e.g., irreversible transduction). The concepts of effort and flow are thus generalized into variables that may contribute to a partial derivative of the energy (generalized effort) and that may contribute to a rate of change of state (generalized flow), but do not have to be identical to them. This corresponds to a multiport-irreversible transducer that relates the effort vector of the multiport to the conjugate flow vector. Due to the irreversible nature of this process, the net power that flows into  $n - 1$  of the ports has to come out at the thermal port of which the flow represents the production of entropy that is related to an irreversible process

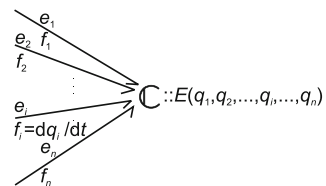


Fig. 1.12 Multiport C-element representing energy storage

as a consequence of the principle of entropy production. The power of this port is thermal power, i.e., a rate of change of heat.

In order to (conceptually) connect these non-storing ports to the storing ports, nodes are needed that neither store nor make a difference between their ports, in other words ports that are power continuous and port symmetric. It can be proven that merely the demands of power continuity and port symmetry lead to two solutions for the constitutive relations that are linear and domain independent and can be considered a combination of the generalized Kirchhoff current (flow) and voltage (effort) laws with the commonness of the conjugate effort and flow, respectively [9]. Again, we emphasize that these nodes do not have to be spatially distinguishable: they may spatially coincide with the fundamental physical behaviors (storage, irreversible transduction, etc.) that are needed to describe the dynamics of variables in one spatial location. This also means that these variables will not be individually measurable in that situation. However, if one wishes to represent such model structures graphically they will need to be given a separate location at a piece of paper or a computer screen. This means that there is no automatic relation between the spatial structure of objects under investigation and the graphical representation of the concepts needed to describe dynamic behavior. This abstraction step in model representation is for many people, both students and experienced researchers, a bridge too far: they tend to confuse the conceptual structure between fundamental dynamic behaviors with a spatial structure, not in the last place, because in some domains physical components are created of which the dominant behavior, at least within a certain range of operation, can be competently modeled by just one elementary behavior and thus represented by just one ideal conceptual element. On top of that, the names of these physical components often coincide with the names of the corresponding conceptual elements. This holds in particular for electrical circuits where the names ‘resistor,’ ‘capacitor,’ ‘transformer,’ etc., are used in both cases, but it also holds for the inertia of a rigid body in the mechanical domain.

### *1.5.12 Causality, Legendre Transforms, and Co-energy*

As discussed in the previous section, the stored energy as a function of the extensive states can be represented by a multiport C-type node in a bond graph (C-element). It can be concluded that when this energy is used to generate the constitutive relations of this multiport, the efforts are dependent variables of the stored extensive states, i.e., an effort-out causality in the bond graph. This causality requires that all flows at the ports are integrated with respect to time in order to obtain the change in state that can be added to the initial state, which is a relevant parameter for each physical storage port. If the interconnection to another port requires that the causality of a port be inverted, this means that the extensive state is replaced in the constitutive relation as an independent variable by the conjugate effort (intensive state) and that the conjugate flow (the rate of change of the state) has to be obtained via differentiation with respect to time which leaves no room for the concept of initial state. This

physical argument, combined with the fact that numerical differentiation amplifies (numerical) noise, results in a preference for effort-out or integral causality for a **C**-type port as opposed to flow-out or differential causality.

However, if a particular model requires non-preferred differential causality the energy cannot be used as a generating function for the constitutive relations and has to be Legendre transformed. As discussed earlier, the negative Legendre transform or co-energy with respect to the dependent extensive state(s) should now be used as generating function and this changes the relation between the efforts at the other ports and the generating function. In case one or more extensive states are considered constant parameters in such a way that the constitutive relations become linear, the energy and co-energy become equal in value, although they still have different independent variables. This results in a common misinterpretation in the sense that energy and co-energy are confused, which may have an impact on the way a constitutive relation can be derived from a generating function:

$$\begin{aligned} E_{q_i}^* &= -\mathcal{L}(E(\mathbf{q}))_{q_i} = \frac{\partial E(\mathbf{q})}{\partial q_i} q_i - E(\mathbf{q}) = e_i q_i - E \\ &= E_{q_i}^*(q_1, \dots, q_{i-1}, e_i, q_{i+1}, \dots, q_n) \end{aligned} \quad (1.44)$$

For example, take a two-port **C** characterized by the following energy function:

$$E(\mathbf{q}) = E(q_1, q_2) = \frac{q_1^2}{2C(q_2)} \quad (1.45)$$

The efforts then are

$$\begin{aligned} e_1 &= \frac{\partial E(q_1, q_2)}{\partial q_1} = \frac{q_1}{C(q_2)} \\ e_2 &= \frac{\partial E(q_1, q_2)}{\partial q_2} = -\frac{q_1^2}{2C^2(q_2)} \frac{dC(q_2)}{dq_2} = -\frac{e_1^2}{2} \frac{dC(q_2)}{dq_2} \end{aligned} \quad (1.46)$$

The co-energy with respect to  $q_1$  is

$$E_{q_1}^*(e_1, q_2) = e_1 q_1 - E(q_1, q_2) = e_1^2 C(q_2) - \frac{e_1^2 C^2(q_2)}{2C(q_2)} = +\frac{e_1^2 C(q_2)}{2} \quad (1.47)$$

and the second effort can be found as the negative value of the partial derivative of this co-energy with respect to  $q_2$

$$\frac{\partial E_{q_1}^*(e_1, q_2)}{\partial q_2} = +\frac{e_1^2}{2} \frac{dC(q_2)}{dq_2} = -e_2 \quad (1.48)$$

which means that

$$e_2 \neq \frac{\partial E_{q_1}^* (e_1, q_2)}{\partial q_2} \quad (1.49)$$

An example will make clear why the latter inequality is listed explicitly.

Take a flat-plate capacitor of which the fringing of the electric field lines can be neglected. For the capacitance  $C$ , such a capacitor can be written as

$$C = \frac{A\varepsilon}{d} \quad (1.50)$$

where  $A$  is the area of the plates,  $d$  the distance between the plates (spatial parameters), and  $\varepsilon$  the dielectric constant (material parameter). The energy can be written (second-degree homogenous function) as

$$E(q) = \frac{q^2}{2C} = \frac{q^2 d}{2A\varepsilon} \quad (1.51)$$

and the co-energy

$$E^*(u) = \frac{Cu^2}{2} = \frac{u^2 A\varepsilon}{2d} \quad (1.52)$$

However, due to the linearity of the constitutive relation both are the same in value and the latter expression is often called ‘energy’ as a consequence. When the distance between the plates is not constant, the distance  $d$  is considered again a state variable, say  $x$ , and the model should be extended with a mechanical port. The mechanical effort (force) of this port can be obtained by partial differentiation of the energy function (that is a third-degree homogenous function)

$$E(q, x) = \frac{q^2}{2C(x)} = \frac{q^2 x}{2A\varepsilon} \quad (1.53)$$

with respect to  $x$ :

$$F(q, x) = \frac{\partial E(q, x)}{\partial x} = -\frac{q^2}{2C^2(x)} \frac{dC(x)}{dx} = \frac{q^2}{2A\varepsilon} \quad (1.54)$$

Note that the energy function is only written as a first-degree homogenous function as the area is also considered as the product of two displacement states, such that all extensive states are identified, even when kept constant.

Using

$$u(q, x) = \frac{\partial E(q, x)}{\partial q} = \frac{q}{C(x)} = \frac{qx}{A\varepsilon} \quad (1.55)$$



$F(u, x)$  can be written (i.e., after a change in causality) as

$$F(u, x) = \frac{u^2 A^2 \varepsilon^2}{2A\varepsilon x^2} = \frac{u^2 A \varepsilon}{2x^2} \quad (1.56)$$

However, if this operation is applied to the co-energy, because it is mistaken for the energy, one obtains

$$F'(u, x) = \frac{\partial E^*(u, x)}{\partial x} = \frac{u^2}{2} \frac{dC(x)}{dx} = -\frac{u^2 A \varepsilon}{2x^2} \quad (1.57)$$

so

$$F'(u, x) = -F(u, x) \quad (1.58)$$

In other words, there is the danger of at least a sign error when deriving the force via partial differentiation of a generating function that is incorrectly considered an energy, even though it has the same numerical value, but not the same independent variables. The reader is challenged to perform a similar analysis in case the voltage is proportional to the third power of the charge, which will show him that the difference may be more than ‘just’ a sign error that is often justified by some action-is-minus-reaction reasoning. This moreover demonstrates the need for making a careful categorization of variables when modeling and the distinction between equilibrium-determining and equilibrium-establishing variables certainly supports this categorization.

### ***1.5.13 The Thermodynamic Versus the Mechanical Framework of Variables***

The above treatment of the concepts of power and energy and its Legendre transforms is a typical thermodynamic approach [25], even though it holds for all domains in which states and energy are useful concepts as demonstrated by the above example of kinetic energy. It results in domains that are identified by one extensive state  $q$  that is conserved, its rate of change or flow  $f$  that establishes the equilibrium, and one intensive state or effort  $e$  that determines the equilibrium (cf. Table 1.1 for particular instantiations). In the mechanical framework in Table 1.2, two of these domains are combined into one, resulting in two types of extensive state per domain, the so-called generalized displacements  $\mathbf{q}$  and the generalized momenta  $\mathbf{p}$ , while the conjugate intensive variables are related to each other’s rates of change by Hamilton’s equations, where the Hamiltonian  $H(\mathbf{q}, \mathbf{p})$  is the energy of the mechanical domain in principle (it can be generalized), i.e., the sum of the kinetic and potential energies:

$$\begin{aligned}\frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial p_i} &= \frac{d}{dt} q_i = v_i \\ \frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial q_i} &= -\frac{d}{dt} p_i = -F_i\end{aligned}\tag{1.59}$$

or

$$\frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial(\mathbf{q}, \mathbf{p})} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{F} \end{bmatrix} = \mathbf{J}^{-1} \begin{bmatrix} \mathbf{v} \\ \mathbf{F} \end{bmatrix}\tag{1.60}$$

or, in non-preferred causality

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{F} \end{bmatrix} = \mathbf{J} \frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial(\mathbf{q}, \mathbf{p})}\tag{1.61}$$

where  $\mathbf{I}$  is a unit matrix and  $\mathbf{J}$  a so-called symplectic matrix or Poisson structure matrix. Note that the so-called external forces can be simply added to this format, but the most important conclusion is that when the generalized forces  $\mathbf{F}$  and generalized velocities  $\mathbf{v}$  are used to describe the physical domains in a generalized mechanical sense, the asymmetry between these variables is lost, as both can be equilibrium establishing and equilibrium determining, which makes the use of these concepts during modeling harder to use and interpret. However, given the large body of literature that is available on all aspects of this approach an extended mathematical framework, inspired by a bond graph approach, was coined as ‘port-Hamiltonian approach’ [26].

Another, even more important drawback of generalization of the mechanical framework is the fact that the displacement variable that can be used to describe storage of potential energy has the same nature as the variables needed to describe the spatial configuration. In many cases the same variable describes configuration as well as energy storage. This dual role of the displacement variable has brought many to believe that force and displacement should be considered to be the conjugate pair of variables for the mechanical domain, which leads to all sorts of confusing analogies. However, it is easy to see that the initial length of a spring between two masses does not influence its dynamic properties for linear motions: only the spring constant and the masses determine the natural frequency of this system. However, in case of planar or spatial motion or even linear motion with contact behavior like a bouncing ball, the changing configuration modulates the energy transfer and in those cases the configuration variable will influence the dynamics, but in a fundamentally different manner than an energy state. This is why it is fruitful to distinguish between (potential) energy states (displacements) and configuration states (positions), even if the variable turns out to be the same mathematical state. Before the time of symbolic processing, reduction of the states to the minimal set of independent states largely supported the analysis. Using modern computer tools, dependent states can either

be eliminated symbolically or solved numerically afterwards without bothering the user, even though the user should carefully inspect the results in both cases.

The close relationship between configuration states and energy states creates a preference for configuration-related variables in a model description, i.e., often generalized momenta are replaced by generalized velocities in the relations, which is rather straightforward in case the linearity of Newton's law is assumed, which always holds in classical mechanics. However, this means that a Legendre transform is performed on the true energy function, the Hamiltonian  $H$

$$H(\mathbf{q}, \mathbf{p}) = T(\mathbf{q}, \mathbf{p}) + V(\mathbf{q}) \quad (1.62)$$

resulting in the Lagrangian  $L$

$$\begin{aligned} L(\mathbf{q}, \dot{\mathbf{q}}) &= \dot{\mathbf{q}}^T \mathbf{p} - H(\mathbf{q}, \mathbf{p}) = \\ &= T(\mathbf{q}, \mathbf{p}) + T^*(\mathbf{q}, \dot{\mathbf{q}}) - (T(\mathbf{q}, \mathbf{p}) + V(\mathbf{q})) = \\ &= T(\mathbf{q}, \mathbf{p}) + T^*(\mathbf{q}, \dot{\mathbf{q}}) - T(\mathbf{q}, \mathbf{p}) - V(\mathbf{q}) = \\ &= T^*(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q}) \end{aligned} \quad (1.63)$$

in other words, the Lagrangian is the difference between kinetic co-energy  $T^*(\mathbf{q}, \dot{\mathbf{q}})$  and potential energy  $V(\mathbf{q})$ , where the position dependence of the kinetic (co-)energy is the result of the coordinate transformation into generalized coordinates.

### 1.5.14 Energy States Versus Configuration States

As the first use of bond graph concepts was highly related to the mechanical domain and heavily influenced by already existing mathematical modeling approaches for mechanical systems, the distinction between energy states and configuration states was usually not made and two types of state variables were distinguished, generalized displacement ( $q$ -type state stored via a **C**-type storage port) and generalized momentum ( $p$ -type state stored via an **I**-type storage port). As a consequence, the asymmetry between effort and flow was lost as this approach combines two domains and treats them as one (refer to thermal inertance paper and thesis). In the two-domain approach the generalized potential domain has a generalized force-type effort and a generalized velocity-type flow (rate of change of generalized displacement, while the generalized kinetic domain has a generalized velocity-type effort and a generalized force-type flow (rate of change of momentum). In bond graph terminology it is easy to see that a unit gyrator expresses that in an inertial frame the net effort of the potential domain is equal to the flow of the kinetic domain, while the effort of the kinetic domain determines the rate of change of displacement (flow) in the potential domain. Similarly, if one assumes quasi-stationary, i.e., non-radiating, electrical circuits, Maxwell's equations that couple the magnetic and electric domains can be reduced to a relation that equates the rate of change of flux (linkage) (i.e., the magnetic flow) with a voltage (i.e., the electric effort) and a

current (i.e., the electric flow) with a magnetomotive force (i.e., the magnetic effort). In these specific cases (inertial frame and quasi-stationary networks) the interdomain relations reduce to identities that are expressed in a bond graph by a unit gyrator called ‘symplectic gyrator’ (**SGY**). As dualizing, i.e., interchanging the roles of effort and flow, the graph at one side of the **SGY** eliminates the **SGY**, the number of variables is reduced, and the roles of effort and flow have become symmetric, i.e., the distinction between equilibrium determining and equilibrium establishing cannot be made anymore. This has caused endless discussions about whether a force is analogous to a voltage or to a current, while the above discussion shows that this is just a matter of which side of the **SGY** is dualized in the sense that the roles of efforts and flows are interchanged. However, given the fact that the energy state of the potential domain (displacement) is similar to and often coincides with a configuration state which is of importance for all domains, it is to be expected that most of the modeling concepts in the model are part of the potential domain, such that it is to be expected that dualization of the kinetic domain is more straightforward. Something similar holds for electrical circuits: The only conceptual connection to the magnetic domain is the ideal coil (inductor), while the rest of the circuit model is part of the electrical domain. In that case it is also more straightforward to dualize the magnetic storage port into an **I**-type port.

Due to the dual role of the displacement variable, it is possible to eliminate the potential energy by choosing another description of the configuration space. For instance, Einstein showed in his general theory of relativity in 1915 that as a mathematical result, it is possible to eliminate potential energy by such a transformation of the space–time that gravity is reduced to a property (curvature) of space–time.

During the modeling process, where one is not yet interested in the mathematically minimal form of a model, but in a form that allows straightforward interpretation in terms of physical concepts, it is wise to represent the two roles of the position variable separated, even if they merge mathematically into one variable when the model is prepared for analysis or simulation.

The confusion that configuration information may cause when not kept separate from other concepts also applies to topological representations. Graphs like circuit diagrams or bond graphs are often interpreted as spatial relations or even with spatial parameters like an area.

Earlier we discussed that assigning more meaning to certain model representations than they actually have leads to all sorts of confusions, similar to Lewin’s incorrect interpretation of his own experiment, which is an example of confusing topological information with spatial information. Although Willems [3] does recognize the basic principle of a bond graph that models of physical systems consist of subsystems that are related by bilateral signal flows, these confusions also make him reject bond graphs when he writes, emphasized by a separate frame at page 66 of his paper:

- (1) ‘The requirement that the product of effort and flow must be power is sometimes not natural, for example, in thermal interconnections.’

- (2) ‘In connecting terminals of mechanical systems, bond graph modeling equates velocities, and sets the sum of the forces equal to zero. In reality one ought to equate positions, not velocities. Equating velocities instead of positions leads to incomplete models.’
- (3) ‘Interconnections are made by means of terminals, while energy is transferred through ports. Ports involve many terminals simultaneously. The interconnection of two electrical wires involves equating two terminal potentials and putting the sum of two terminal currents to zero. The product of effort, namely, the electrical potential, and flow, namely, the electrical current, for an electrical connection has the dimension of power, but it is not power. Power involves potential differences, while the interconnection constraint involves the terminal potentials themselves. It is not possible to interpret these interconnection constraints as equating the power on both sides of the interconnection point.’
- (4) ‘In many interconnections, it is unnecessary to have to worry about conservation of energy.’

In particular points (2) and (3) refer to the earlier mentioned confusion, but we will extensively comment on all of Willems’ arguments, because they represent the beliefs of many and lead to erroneous conclusions, not only about bond graphs but about modeling in general. It should be noted that Willems refers to classical bond graphs based on the mechanical framework of variables with two types of storage and velocity being a flow-type variable as discussed before.

Ad (1) Any self-respecting thermodynamic textbook will explain that the amount of stored heat is not a state as it does not satisfy the criterion that its cyclic integral equals zero, such that a heat flow is not a rate of change of state. If a model is constrained to the thermal domain, heat or thermal energy is only a function of the stored entropy and can serve as a state in that particular case only. In that case also pseudo-bond graphs can be used to represent such models, so the constraint that the two conjugated variables of a bond should always be power conjugated is not as hard as Willems suggests. Furthermore, he writes in his main text that the use of entropy flow ‘seems artificial.’ He probably means that for many students of thermodynamics the concept of an entropy flow is counterintuitive, yet it is a well-accepted concept in physics, which is absolutely necessary to describe the dynamic interaction of the thermal domain with other domains like the mechanical domain as the well-known treatment of an ideal Carnot cycle demonstrates [25].

Ad (2) The statement ‘In reality one ought to equate positions, not velocities’ shows that he does not make a distinction between configuration states (positions) and energy states (displacements). The only difference between a constraint at the displacement level and a constraint at the velocity level is an initial condition, for instance, the initial distance between two masses connected by a spring. However, the dynamic behavior is not influenced by this initial condition. The only argument to use (additional) constraints at the position level can be that numerical integration generates errors that may lead to a change in the contribution of the initial condition, which means that, due to numerical inaccuracies, the masses in the example drift apart or closer together while they are oscillating.

However, adding such a constraint is a numerical issue and not part of the conceptual model. The easiest way to see this is to make a similar statement about the electric charge in an electric circuit: While simulating the behavior of an electric circuit containing capacitors the total charge balance may be distorted due to numerical errors. Apart from the fact that nobody seems to care about this, probably because charge is much less ‘visible’ than configuration and charge cannot modulate the interconnection structure, a constraint at the charge level can still be added in the numerical solution procedure to improve accuracy. However, position variables also describe the configuration state and often one variable plays both roles. For planar and spatial mechanisms the configuration variables can influence the dynamics via modulation when the coordinate transformations required to describe the interconnection structure are position dependent. These coordinate transformations are represented in bond graphs by a modulated multiport transformer [27]. In conclusion one may state that the constraints in the junction structure of the bond graph are energetic constraints at the velocity level, while the configuration constraints are dealt with at the signal level, both for modulating the MTFs in the generalized junction structure and for finding potential relations between initial conditions, e.g., two springs connected between the same masses.

Ad (3) He writes: ‘Interconnections are made by means of terminals, while energy is transferred through ports.’

Here the actual configuration (‘interconnection’) is mixed up with conceptual relationships.

He continues: ‘Ports involve many terminals simultaneously.’

First of all, it should be clear about which domain we talk: in a mechanical system a terminal may correspond to a port, while mostly two electrical terminals are required to create an electrical port. However, Willems fails to give examples of ports consisting of more than two terminals and unless he is referring to multiports, which would be confusing, more than two terminals per port are not possible, unless one uses a rather uncommon definition of a terminal.

He then writes: ‘The interconnection of two electrical wires involves equating two terminal potentials and putting the sum of two terminal currents to zero. The product of effort, namely, the electrical potential, and flow, namely, the electrical current, for an electrical connection has the dimension of power, but it is not power. Power involves potential differences, while the interconnection constraints involves the terminal potentials themselves. It is not possible to interpret these interconnection constraints as equating the power on both sides of the interconnection point.’

The conceptual error in this argument lies in the use of the concept ‘wire.’ Willems uses this as something for which a current is a meaningful concept as soon as one terminal of the wire is connected to the terminal of another wire. However, a current is a ‘through-variable,’ which means that a connection to a second terminal at the other side, commonly with a component in between, is required to make the concept of current meaningful. Not only the current requires two terminals, also the voltage, as it is a relative concept and always measured with respect to some reference (‘ground’). This means that all potentials are in fact potential differences,

while Willems only considers potential differences with respect to potentials that are not the ground potential.

This makes clear that any two electrical terminals, also if one is grounded and not represented as such, form a port. However, in many treatments of electrical circuits the logical order is reversed: First, a distinction is made between potential and potential differences. Next a relation between the potential differences and the potentials is derived via the incidence matrix and only then it is recognized that one of the rows (balance equations) refers to a reference node (ground) which should be omitted from the incidence matrix to obtain the so-called reduced incidence matrix. This culture may have led Willems to drawing this conclusion.

In his main text (page 68) Willems even writes about this:

‘In electrical circuits, energy is not transmitted along terminals but rather through ports. On the other hand, in modeling and interconnection, terminals matter, not ports.’

He then explains the difference in his Fig. 11 where it becomes clear that he assumes that connecting two terminals with each other suffices for electrical interconnection, which is not the case: A current can only flow after a second connection that creates a loop.

This shows that he does not appreciate the difference between a drawing of some physical components soldered together (= ‘mechanical’ connection) which does not lead to a device that has any electrical meaning as no current can flow and an electrical circuit diagram in which connections may coincide with physical connections (soldering points), but may as well be conceptual.

Ad (4) Any textbook on (classical) physics will explain that if one models the dynamics of physical systems, one should always consider the conservation of energy, the most general principle of physics. However, the power involved in some of the relations in a system may be negligible with respect to the power of other relations and this is probably what Willems tries to say; however, he formulates it in such a way that it becomes an implicit modeling assumption that power is negligible. A bond graph represents such a modeling decision by ‘bond activation,’ turning the bond into a signal for one of the conjugate variables while assuming the other to be negligible. This shows that Willems, like many others, including those who use bond graphs, has not taken the effort of studying what a bond graph representation really is about, but that he has considered explanatory text in a particular context to be a unique and generic definition.

Another highly related confusion is the concept of a ‘memristor’ introduced by Chua [28]. When one studies Chua’s paper from a bond graph perspective it becomes clear that he refers to what in a bond graph would be optimally represented by a state-modulated transformer for which Chua coined the name ‘memristor’ due to the seemingly missing link between the  $q$ -type and  $p$ -type variables in the mechanical frame work of variables (Table 1.2). Note that the more general thermodynamical framework would never suggest this relation. Not much attention has been paid to this concept until recently, where some researchers came across an interesting phenomenon that can be best described as a state-modulated resistor, where the state can be set to different values without much ‘leakage,’ in other words:

a resistor with a memory. They immediately saw the latter as a reason to rewrite their equations into a form that more or less resembles Chua's equations and next claimed that the memristor was finally found [29], which again demonstrates the common confusion between physical components and ideal conceptual elements.

### *1.5.15 Conservation and Continuity Within a Domain*

As each conserved state determines a domain, additional connection constraints can be found for various port types. For instance, a bond connected to one side of a 0-junction may be connected to a **C**-type storage port or a source port, as these ports do not violate the balance equation. However, in principle, one should be more careful when connecting an **I**-type, **R**-type, **TF**-type, or **GY**-type port, because these ports cannot 'absorb' the conserved state related to the flow. However, all domains with relative equilibrium-determining variables have a non-displayed balance for the reference node (this balance equation is dependent on the balance equations for the rest of the network and corresponds to the row that is omitted in an incidence matrix to turn it into a reduced incidence matrix of an electrical circuit, for example). This additional balance compensates for this flow, such that it is still possible to connect these ports without violating the balance equation. Note that the **I**-type port in principle is a connection to a **GY**-type port that connects to the storage in another domain. Some domains have absolute equilibrium-determining variables, like temperature and pressure, but since in most cases it is not practical to choose the absolute zero point as a reference, usually another reference state is chosen, such that these variables are treated as differences with respect to an arbitrary reference and an additional balance too.

Further note that labeling energy with a domain is only possible in case of one-port storage as multiport storage means that other domains have access to the stored energy, such that it does not belong to one domain when stored. As we have seen before, one-port storage is the result of neglecting all ports of other states because they are considered constant, but it does not exist in principle, because a first-degree homogenous energy function of a one-port storage element would lead to a zero-order homogenous relation between effort and state, which would mean that the effort would be a constant. So, also in principle, it is not appropriate to assign a domain to energy, as it can be reached via more than one port, unless the flows of all other ports are kept zero.

## **1.6 Conclusion**

The warm bed of dynamic models in the form of a set of differential equations or a block diagram in an a priori fixed causality has been replaced by the cold floor of port-based modeling concepts with causal preferences and constraints. Hopefully the new revelation caused by this disturbing experience is the insight that the



causality assignment procedure gives the modeler immediate feedback on his modeling decisions, because the causal consequences of interconnection constraints can be seen immediately in the context of the physical concepts instead of being the result of an abstract analysis that hides the relation to the physics.

Another key issue of this contribution is to make clear that three kinds of structures, viz., configuration structure (spatial structure), physical structure, and conceptual structure are often mixed up during the modeling of dynamic behavior of physical systems. The earlier mentioned lecture where Lewin gives an area (spatial) interpretation to a conceptual (topological) representation and incorrectly assigns similarity of points in a conceptual representation to points in a spatial representation is an example of such a mix-up. Willems' problems to understand bond graphs are an example of a confusion of physical structure ('soldering' physical components together not resulting in a configuration that has a meaning in an electrical sense) and conceptual structure that shows the relations between ideal elements that represent dominant behaviors of these components in a particular context, viz., a circuit in that allows electrical phenomena, viz., currents. Again it is emphasized that much attention was paid to these cases, not for the sake of having a specialized scientific discussion with opponents, but because these examples clarify how easy it is to become the victim of the conceptual trap of forgetting about implicit assumptions.

The common criticism that bond graphs are not well defined has been countered by showing that their key properties can be defined up to a certain level, but that, like any other language, dialects remain possible and it is up to the user to define the exact meaning of specific symbols.

## References

1. Beeren W., Roessink M. (eds) (1998) *Sporen van wetenschap in kunst/Traces of Science in Art*, 155pp., ISBN 90-6984-224-6.
2. Paynter H.M. (1961) *Analysis and Design of Engineering Systems*. MIT Press, Cambridge, MA.
3. Willems J.C. (2007) The behavioral approach to open and interconnected systems. *IEEE Control Syst. Mag.* Dec:46–99.
4. Bondy J.A., Murty U.S.R. (1976) *Graph Theory with Applications*. North-Holland, Oxford, ISBN 0-444-19451-7.
5. Breedveld P.C. (1982b) Proposition for an unambiguous vector bond graph notation. *Trans. ASME, J. Dyn. Syst. Meas. Control* 104(3):267–270.
6. Breedveld P.C. (1986) A definition of the multibond graph language. In *Complex and Distributed Systems: Analysis, Simulation and Control*, Tzafestas S., Borne P. (eds) Vol. 4 of 'IMACS Transactions on Scientific Computing'. North-Holland, Amsterdam, pp. 69–72.
7. Breedveld P.C. (1982a) Thermodynamic bond graphs and the problem of thermal inertance. *J. Franklin Inst.* 314(1):15–40.
8. Karnopp D.C., Rosenberg R.C. (1968) *Analysis and simulation of multiport systems*. MIT Press, Cambridge, MA.
9. Hogan N.J., Fasse, E.D. (1988) Conservation principles and bond graph junction structures. *Proc. ASME 1988 WAM. DSC* 8:9–14.

10. Paynter H.M., Busch-Vishniac I.J. (1988) Wave-scattering approaches to conservation and causality. *J. Franklin Inst.* 325(3):295–313.
11. Breedveld P.C. (1985) Multibond graph elements in physical systems theory. *J. Franklin Inst.* 319(1/2):1–36.
12. Blundell A.J. (1982) *Bond Graphs for Modelling Engineering Systems*. Ellis Horwood, Chichester and Halsted Press, New York, NY, 151p.
13. Thoma J. (1975) *Introduction to Bond Graphs and Their Applications*. Pergamon Press, Oxford.
14. Breedveld P.C. (2009) Port-based modeling of dynamic systems. Chapter 1 and appendix A and B2. In *Modeling and Control of Complex Physical Systems – The Port-Hamiltonian Approach*, Stramigioli S., Macchelli A., Duindam V., Bruyninckx H. (eds). Springer, Berlin, pp. 1–52, 97–311, 323–328.
15. Breedveld P.C. (1984). Decomposition of multiport elements in a revised multibond graph notation. *J. Franklin Inst.* 318(4):253–273.
16. Breedveld P.C. (1995) Exhaustive decompositions of linear two-ports. *Proceedings of SCS 1995 International Conference on Bond Graph Modeling and Simulation (ICBGM'95)*, SCS Simulation Series 27(1):11–16, Jan 15–18, Las Vegas, Cellier F.E., Granda J.J. (eds). ISBN 1-56555-037-4.
17. Dijk J. van, Breedveld P.C. (1991a) Simulation of system models containing zero-order causal paths – part I: Classification of zero-order causal paths. *J. Franklin Inst.* 328(5/6):959–979.
18. Breedveld P.C. (2007) Port-based modeling of engineering systems in terms of bond graphs. In *Handbook of Dynamic System Modeling*, Fishwick P.A. (ed). Chapman & Hall, London, pp. 26.1–26.29, ISBN 1-58488-565-3.
19. Dijk J. van, Breedveld P.C. (1991b) Simulation of system models containing zero-order causal paths – part II: Numerical implications of class-I zero-order causal paths. *J. Franklin Inst.* 328(5/6):981–1004.
20. Feynman R, Leighton R, Sands M. (1989) *The Feynman Lectures on Physics*. 3 volumes 1964, 1966, Addison-Wesley, Reading, Mass, ISBN 0-201-50064-7.
21. Zill D.G. (2005) *A First Course in Differential Equations*. 9th edition. Brooks/Cole, Belmont, CA, ISBN-13: 978-0-495-10824, Lib. of Congress number: 2008924906.
22. Wellstead P.E. (1979) *Introduction to physical systems modeling*. Academic, London. ISBN: 0-12-744380-0.
23. Timoshenko S. (1976) *Strength of Materials: Elementary Theory and Problems*. Vol. 1 of *Strength of Materials*, 3rd edition, R.E. Krieger, Huntington, NY (First ed. D. Van Nostrand Company, inc., 1940).
24. Allen R.R. (1981) Dynamics of mechanisms and machine systems in accelerating reference frames. *Trans. ASME J. Dyn. Syst. Meas. Control* 103(4):395–403.
25. Callen H.B. (1960) *Thermodynamics*. Wiley, New York, NY.
26. Maschke B.M., van der Schaft A.J., Breedveld P.C. (1995) An intrinsic Hamiltonian formulation of the dynamics of LC-circuits. *Trans. IEEE Circuits Syst. I Fundam. Theory Appl.* 42(2):73–82.
27. Karnopp D.C. (1978) The energetic structure of multibody dynamic systems. *J. Franklin Inst.* 306(2):165–181.
28. Chua L.O. (1971) Memristor-the missing circuit element. *IEEE Trans. Circuit Theory CT-* 18(5):507–519.
29. Strukov D.B., Snider G.S., Stewart D.R., Williams, R.S. (2008) The missing memristor found. *Nature* 453. doi:10.1038/nature06932.