A Formal Hybrid Modeling Scheme for Handling Discontinuities in Physical System Models

Pieter J. Mosterman and **Gautam Biswas**
Center for Intelligent Systems
Box 1679, Sta B
Vanderbilt University
Nashville, Tennessee 37235
pjm, biswas@vuse.vanderbilt.edu

Abstract

Physical systems are by nature continuous, but often exhibit nonlinearities that make behavior generation complex and hard to analyze. Complexity is often reduced by linearizing model constraints and by abstracting the time scale for behavior generation. In either case, the physical components are modeled to operate in *multiple modes*, with abrupt changes between modes. This paper discusses a *hybrid* modeling methodology and analysis algorithms that combine continuous energy flow modeling and localized discrete signal flow modeling to generate complex, multi-mode behavior in a consistent and correct manner. Energy phase space analysis is employed to demonstrate the correctness of the algorithm, and the reachability of a continuous mode.

Introduction

Recent advances in model-based and qualitative reasoning have led to researchers developing large scale models of complex, continuous systems, such as power plants and space station sub-systems. System complexity is handled by replacing nonlinear component behaviors by simpler piecewise linear behaviors, causing the system to exhibit multi-mode behavior[11]. For example, the Airbus A-320 fly-by-wire system includes the *take off*, *cruise*, *approach*, and *go-around* operational modes[13].

Quantitative and qualitative simulation methods (e.g., [6, 12]) typically impose *continuity constraints* to ensure generated behaviors are meaningful. However, system models that accommodate configuration changes and multi-mode components can exhibit discontinuous behavior. Consider the diode-inductor circuit in Fig. 1. When closed switch S_w is opened, and the voltage drop across the diode exceeds $0.6V$ it comes on and *abruptly* enforces this voltage across the inductor. Computational complexity is reduced by modeling the diode as an ideal switch with on and off modes. In reality, parasitic resistive and capacitive effects in the diode would force the *on/off* changes to be continuous with a very fast time constant.

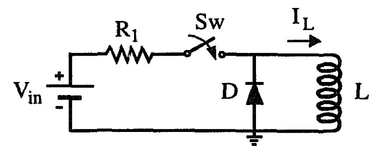

Figure 1: **Physical system with discontinuities.**

Our goal is to derive a uniform approach to analyzing continuous and discontinuous system behavior without violating fundamental physical principles of conservation of energy and momentum. The solution is a hybrid modeling scheme that combines traditional energy-related bond graph elements to model the physical components and finite state automata controlled junctions to model configuration changes.

Characteristics of Physical Systems

A physical system can be regarded as a configuration of connected physical elements. The energy distribution in the system reflects its behavioral history up to that time and defines the traditional notion of system *state*. Future behavior of the system is a function of its current state and input to the system from the present time. State changes are caused by energy exchange between system components, expressed as *power*, the time derivative or flow of energy. Independent of the physical domain (mechanical, electrical, hydraulic, etc.), power is defined as the product of two conjugate *power variables*, *effort*, e, and *flow*, f. Correspondingly, energy comes in two forms: stored effort and flow, called *generalized momentum*, p, and *generalized displacement*, q, respectively. The variables p and q are called *state variables.*

Bond graphs capture continuous energy-related physical system behavior[12]. Its constituent elements are energy storage elements *inductors*, I, and *capaci-*

tors, C, dissipators, R and sources/sinks of effort and flow, S_e and S_f. Sources define interaction with the environment. Idealized, lossless 0- (*common effort*) and 1- (*common flow*) *junctions* connect sets of elements and define the system configuration. Two special junctions called signal transformers complete the set of bond graph primitive elements, the transformer, TF, and the gyrator, GY.

Bond graph models describe system behavior by energy exchange among components. Depending on the type of stored energy, buffer elements impose either effort or flow on their respective junctions. This imposes a causal structure on the system effort and flow variables which is exploited to generate system behavior in the form of quantitative state equations[12] and qualitative relations among variables [1, 8]. In summary, bond graphs provide an elegant formalism to model the continuous behavior of physical systems.

Nature and Effects of Discontinuities

Conservation of energy enforces a time integral relation between energy and power variables which implies *continuity* of power, therefore, effort and flow. Discontinuities in behavior generation can be attributed to simplifying model descriptions[2, 11]. We contend that all discontinuities in the modeling of physical systems can be attributed to *abstracting* component behavior to simplify (i) the time-scale of the interactions, or (ii) the relations among parameters.

Often the time scale of nonlinear behavior in components is significantly less than the time scale at which overall system behavior is studied. Explicit modeling of system behavior at the smaller time scale may greatly increase the time complexity of behavior simulation and introduce numerical stiffness. To avoid this, components like electric switches, valves, and diodes are modeled to have abrupt or discontinuous changes in behavior.

Another cause for discontinuities in models stems from component parameter abstractions. The detailed effects of particular component characteristics, such as fast nonlinear behaviors of transistors and oscillators, are usually not important except for their gross effects on overall system behavior. Behavior generation is simplified by approximating nonlinear behavior as a series of *piecewise* linear behaviors. In other situations certain parameter effects that have negligible effects on gross behavior are omitted from system models.

Since all changes in the state of any physical system are brought about by energy exchange or power, the constraint on power continuity plays an important role in meaningful behavior generation. However, in models with discontinuities, behavior generation schemes have to deal with discontinuities in power variables[9].

The Hybrid Modeling Scheme

In qualitative simulation systems, such as QSIM[6], a higher level global control structure (meta-model) determines when to switch QDE sets during behavior generation. In other approaches[2, 6], transition functions between configurations are specified as rules or state transition tables. In work based on bond graph schemes, researchers have introduced switching bonds[2] controlled by global *finite state automata* to connect and disconnect subsystem models. All these methods fail for systems whose range of behaviors have not been pre-enumerated. Compositional modeling approaches that build system models *dynamically* by composing model fragments[1, 10] overcome this problem. We adopt this methodology and implement a dynamic model switching methodology in the bond graph modeling framework.

We avoid global control structures and pre-enumerated bond graph models. Instead we translate the overall physical model to one bond graph model that covers the energy flow relations within the system. Next, the discontinuous mechanisms and components in the system are modeled locally as *controlled junctions* which can assume one of two states – *on* and *off*. Local *finite state automata* which control each junction constitute the *signal flow model* of the system. It is distinct from the bond graph model that deals with the energy-related behavior of the physical system variables. Signal flow models describe the *transitional*, mode-switching behavior of the system. A *mode* of a system is determined by the combination of the *on/off states* of all the controlled junctions in its bond graph model.

Controlled Junctions

When active (*on*), controlled junctions behave like normal 0- or 1-junctions. Deactivated 0-junctions force the effort and deactivated 1-junctions force the flow at adjoining bonds to become 0, thus inhibiting energy flow. In both cases, the controlled junction exhibits *ideal switch* behavior, and modeling discontinuous behavior in this way is consistent with bond graph theory[12]. Deactivating controlled junctions can affect the behaviors at adjoining junctions, and, therefore, the causal relations among system variables.

Controlled junctions are marked with subscripts (e.g., 1_1, 0_2) in the hybrid bond graph representation (Fig. 2). Each controlled junction has an associated finite state automata that generates the *on/off* signals for the controlled junction. This is called a junction's *control specification* (CSPEC). CSPEC input consists

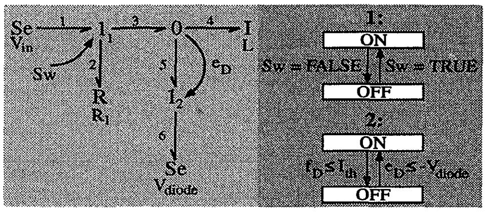

Figure 2: **Hybrid bond graph model.**

of power variables from the bond graph and external control signals. Its output is the *on/off* signal for the controlled junction. In every CSPEC transition sequence, *on/off* signals have to alternate.

Mode Switching in the Hybrid Model

Discontinuous effects establish or break energetic connections in the model when threshold values are crossed. As a consequence, signals associated with bonds at the junction may change discontinuously. Also, when junctions become active, buffers may become dependent, causing an apparent instantaneous change in the energy distribution of the system[9].

The use of controlled junctions is illustrated for the diode-inductor circuit (Fig. 1) in Fig. 2. The manual switch turns *on* or *off* based on an external control signal as shown in CSPEC **1**, and the diode switches *on* or *off* based on CSPEC **2**. Fig. 3 shows a simulation run of this system with parameter values $V_{in} = 10V, R_1 = 330\Omega, L = 5mH, p(0) = 0$. When the switch is closed ($t = 10$), the inductor is connected to the source and builds up a flux, p, by drawing a current. The diode is not active in this mode of operation (10). When the switch is opened ($t = 100$), the current drawn by the inductor drops to 0, causing its flux to discharge instantaneously. Because of the derivative nature of the constituent relation $V_L = L\frac{dp}{dt}$, the result is an infinite negative (the flux changes from a positive value to 0) voltage across the diode (Fig. 4). Because its threshold value, V_{diode}, is exceeded, the diode comes on instantaneously and the mode of operation where the switch was open and the diode inactive (00) is never realized in real time. If it were, the stored energy of the inductor would be released instantaneously in a mythical mode where the model has no real representation, producing an incorrect energy balance in the overall system. Consequently, there would be no flow of current after the diode becomes active. In real time the system switches from mode 10 to 01 directly.

At $t = 350$ the diode turns *off* because its current falls below $I_{th} = 0$. Since there is no stored energy in the system, 00 becomes the final mode. The spike observed is a simulation artifact caused by the simulation time step. The diode was inferred to switch *off* when the current had a small negative value, rather than 0. This small current in the inductor went to 0 instantaneously, which resulted in the spike shown.

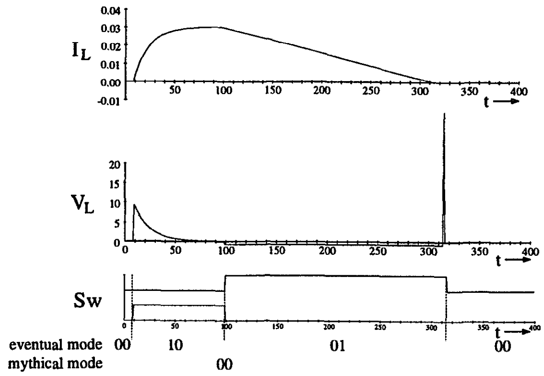

Figure 3: **Diode-inductor simulation.**

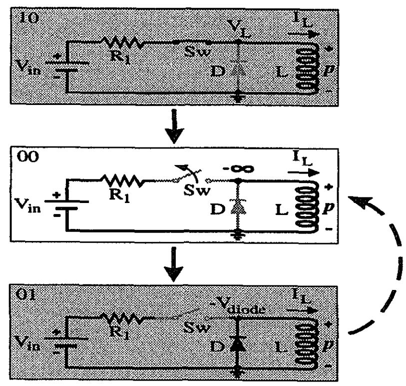

Figure 4: **A series of mode switches may occur.**

A model that undergoes a sequence of instantaneous mode changes has no physical manifestation during these changes, therefore, these modes are termed *mythical*. Thermodynamically, the system is considered to be *isolated*[3] during mythical modes, i.e., there is no energy exchange with the environment. This establishes the energy distribution of the system as a switching invariant. Because the energy distribution in the system defines its state vector, it is referred to as the principle of *invariance of state*. In the diode-inductor example, the flux, p, of the inductor is invariant during switching. The invariance of state principle applies only if the state variables represent the energy distribution (buffer energy values) in the system.

Energy redistribution can occur in the real mode, and the challenge is to compute the initial energy distribution when a real mode is reached after a series of discontinuous mythical changes. At this point, the

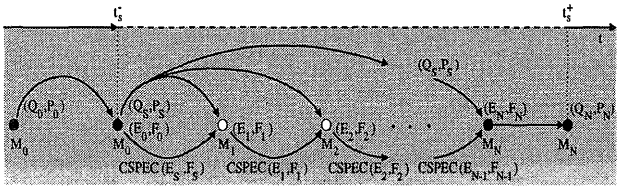

Figure 5: **A sequence of mythical mode switches.**

system is no longer isolated and can exchange energy with the environment.

This is illustrated in Fig. 5. Mythical modes are depicted as open circles and real modes are shown by solid circles. In real mode M_0 a signal value crosses a threshold at time t_s^-, which causes a discontinuous change to mode M_1. Based on the original energy distribution (P_s, Q_s) values for the set of power variables (E_i, F_i) in this new configuration are calculated. The new values cause another instantaneous mode change and the new mode M_2 is reached. Again, the set of new power variables values, (E_i, F_i), is calculated based on the original energy distribution (P_s, Q_s). Further mythical mode changes may occur till a real mode, M_N, is reached. The final step involves mapping the energy distribution, or state variable values, of the departed real mode to the eventual real mode. Real time continuous simulation resumes at t_s^+ so system behavior in real time implies mode M_N follows M_0. The formal *Mythical Mode Algorithm* (MMA) is outlined below.

1. Calculate the energy values (Q_s, P_s) and signal values (E_s, F_s) for bond graph model M_0 using (Q_0, P_0), at the previous simulation step as initial values.
2. Use *CSPEC* to infer a new mode given (E_s, F_s).
3. If one or more controlled junctions switch states:
   (a) Derive the bond graph for this configuration.
   (b) Assign causal directions to bonds[12].
   (c) Calculate the signals (E_i, F_i) for the new mode, M_i, based on the initial values (Q_s, P_s).
   (d) Use *CSPEC* again to infer a possible new mode based on (E_i, F_i) for the new mode, M_i.
   (e) Repeat step 3 till no more mode changes occur.
4. Establish the final mode, M_N, as the new system configuration.
5. Map (Q_0, P_0) to the energy distribution (Q_N, P_N).

Details of the complete simulation algorithm and software for modeling hybrid system behavior are described elsewhere[7].

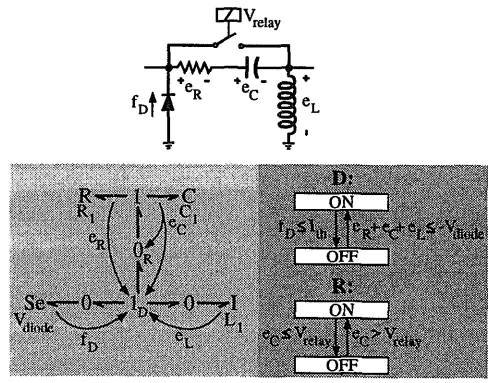

Figure 6: **Diode-relay circuit.**

Divergence of Time

Consider a scenario where the diode requires a threshold current $I_{th} > 0$ to maintain its *on* state. If the inductor has built up a positive flux, the diode comes *on* when the switch opens. However, if the flux in the inductor is too low to maintain the threshold current, the diode goes *off* instantaneously, but in its *off* state the voltage drop exceeds the threshold voltage again. The model goes into a loop of instantaneous changes (see Fig. 4). For instantaneous changes, real time does not progress or diverge, but this violates the physical principle of *divergence of time*[4]. Checking for divergence of time in model behavior is accomplished by a multiple energy phase plot method. Failure to diverge is linked back to the initial values of associated state variables.

Consider the electrical circuit in Fig. 6. The three branches where voltage drops occur in this circuit are represented by 0-junctions. The diode is modeled as an ideal voltage sink and the three branches and elements are connected using 1-junctions. Two switches make up the control flow model

1. The diode switches *on/off* depending on its voltage drop or current. The corresponding controlled junction is 1_D with CSPEC **D**. The input to **D** are the power signals e_R, e_C, e_L, and f_D.
2. The relay is *closed/open* depending on the voltage drop across the capacitor, modeled by controlled junction 0_R with the controlling power variable e_C. A closed relay implies that 0_R is *off* and an open relay implies 0_R is *on*.

To avoid discontinuous changes in power variables during analysis, CSPEC transition conditions are rewritten in terms of the energy variables which are invariant across mythical modes. Since discontinuities

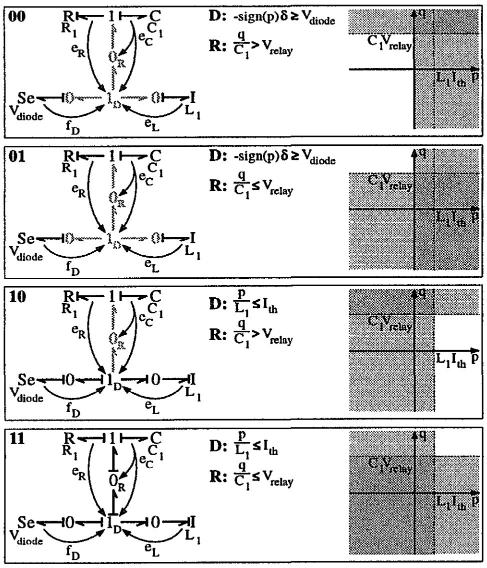

Figure 7: **Multiple energy phase space analysis.**

can cause changes in system configuration, and the relation between power and energy variables, an energy phase space diagram has to be constructed for each switch configuration.

The energy phase space is k-dimensional, where k is the total number of independent buffers in the system. For example, the circuit in Fig. 6 has two energy buffers implying a two dimensional phase space with axes p, the flux in inductor L_1, and q, the charge on capacitor C_1 (Fig. 7). The four modes for the two switches are 00, 01, 10, and 11. The first digit indicates the open/closed state of the diode, and the second digit defines the on/off state of the relay. For each mode, the transition conditions based on the energy variables are grayed out in the phase spaces. For example, in mode 00 the relay turns on if $e_C > V_{relay}$.[1] Substituting $e_C = \frac{q}{C_1}$ generates $q > C_1 V_{relay}$, which is grayed out in the phase space.

The conditions under which the diode turns on are harder to derive because L_1 induces a Dirac pulse, δ.[2] CSPEC **D** switches *on* if $e_R + e_C + e_L \leq -V_{diode}$. When the diode is off, $e_R = e_C = \frac{q}{C_1}$. A deactivated 1-junction has 0 flow so the stored flux in the inductor becomes 0 instantaneously and because $e_L = \frac{dp}{dt}$, this causes e_L to be a Dirac pulse which approaches positive

[1] As part of a larger system (e.g., automobile ignitions), this circuit discharges the inductor through the diode and capacitor. The relay keeps the charge in the capacitor above a small threshold value so that the flux in the inductor does not increase first when it is switched to discharge.

[2] This is a pulse of finite area but infinitesimal width that occurs at a time point.

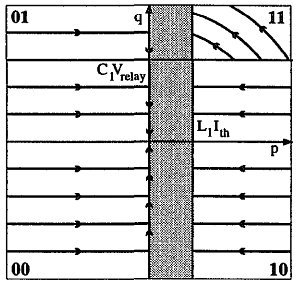

Figure 8: **One energy phase space.**

or negative infinity, depending on whether the stored flux was negative or positive, respectively. If the flux was 0, e_L equals 0. Using the function *sign*

$$sign(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{1}$$

we derive $e_L = -sign(p)\delta \leq -V_{diode}$. The minus signs cancel and e_R and e_C can be neglected, so the condition for switching of the diode becomes $sign(p)\delta \geq V_{diode}$. Assuming the voltage enforced by the diode is 0.6V, this inequality holds for all values of $p > 0$. This area is grayed out in the phase space.

The phase space representation for the four modes (Fig. 7) are superimposed (Fig. 8) to study possible divergence of time violations. If an energy distribution does not have a real energy phase space component, the state vector can never lead to a real mode and time does not diverge for this behavior.

In this example, divergence of time is violated if $I_{th} > 0$ for the diode. This area will be reached for all energy distributions with positive initial flux, p. When $p = 0$ in the 00 and 01 modes, time does diverge. If the flux has a negative initial value, both the flux and capacitor charge converge asymptotically to 0.

Discussion and Conclusions

Hybrid models of physical systems may undergo a series of discontinuous changes. These discontinuities are a result of abstracting the time scale and component parameters in system models. The Mythical Mode Algorithm uses the principle of invariance of state to correctly infer new modes of continuous operation and their state variable values. In pathological cases, system models result in mythical loops, implying the model is physically incorrect. Using the principle of invariance of state, a systematic energy phase space analysis method is developed to verify the correctness of system models. Note that our work verifies the correctness of models, i.e., it ensures that these models do not violate physical principles. This is different from *model validation* which establishes how well a system model behavior matches that of the exact physical situation of interest.

Previous work on model verification by Henzinger *et al*[4] relies on pre-enumeration of global modes of operation, and their method is restricted to variables that have *linear rates of change*. Our method applies more generally to linear and nonlinear models. In other work, Iwasaki *et al.*[5], introduce the concept of *hypertime* to represent the instantaneous switching as an infinitesimal interval. A sequence of switches can be analyzed in hypertime to determine state changes. This approach emulates physical effects of small time constants (e.g., parasitic dissipation) which can greatly increase simulation complexity. Moreover, the modeler often chooses to simplify the model by ignoring parasitic effects. If physical inconsistencies, e.g., non divergence of time arise in behavior generation, the modeler has to add more details in the model increasing its complexity, or adjust landmark values to establish a physically correct but more simple and abstract model. Adding detail may not increase the accuracy of behavior generation because the additional parameters required may be hard to estimate. Also, increasing the computational complexity of models and simulation engines does not guarantee correct models. In the diode-inductor example, an infinitesimal change of time when both the switch and diode are *off* discharges the stored flux and generates incorrect behavior. On the other hand, explicit incorporation of invariance of state ensures that physical consistency of the chosen models can be determined.

Another insight gained is that mythical modes arise from combinations of consistent switching elements, i.e., a single switch cannot cause mythical mode changes. When a number of switches interact via instantaneous relations with no intervening buffers, sequential behavior may ensue. Although these modes are modeling artifacts, they result from justifiable modeling decisions, which have to be dealt with appropriately. In future work we will attempt to demonstrate that reachability analysis can be applied in the multiple energy phase space approach by taking the cross product of all interacting local finite state automata. These sets of interacting automata represent *local modes* of operation. To avoid the computational complexity of the cross product of a number of automata, we will have to develop schemes that efficiently decompose the model into parts that are not instantaneously connected because of intervening energy buffers.

References

[1] G. Biswas and X. Yu. A formal modeling scheme for continuous systems: Focus on diagnosis. *Proc. IJCAI-93*, pp. 1474–1479, Chambery, France, Aug. 1993.

[2] J.F. Broenink and K.C.J. Wijbrans. Describing discontinuities in bond graphs. *Proc. of the Intl. Conf. on Bond Graph Modeling*, pp. 120–125, San Diego, CA, 1993.

[3] G. Falk and W. Ruppel. *Energie und Entropie: Eine Einführung in die Thermodynamik.* Springer-Verlag, Berlin, New York, 1976.

[4] T.A. Henzinger, X. Nicollin, J. Sifakis, and S. Yovine. Symbolic model checking for real-time systems. *Information and Computation*, 111:193–244, 1994.

[5] Y. Iwasaki, A. Farquhar, V. Saraswat, D. Bobrow, and V. Gupta. Modeling time in hybrid systems: How fast is "instantaneous"? *Qualitative Reasoning Workshop*, pp. 94–103, Amsterdam, The Netherlands, 1995.

[6] B. Kuipers. Qualitative simulation. *Artificial Intelligence*, 29:289–338, 1986.

[7] P.J. Mosterman and G. Biswas. Modeling discontinuous behavior with hybrid bond graphs. *Qualitative Reasoning Workshop*, pp. 139–147, Amsterdam, The Netherlands, 1995.

[8] P.J. Mosterman, R. Kapadia, and G. Biswas. Using bond graphs for diagnosis of dynamic physical systems. *Sixth Intl. Workshop on Principles of Diagnosis*, pp. 81–85, Goslar, Germany, 1995.

[9] P.J. Mosterman and G. Biswas. Analyzing discontinuities in physical system models. *Qualitative Reasoning Workshop*, Fallen Leaf Lake, CA, 1996.

[10] P.P. Nayak, L. Joscowicz, and S. Addanki. Automated model selection using context-dependent behaviors. *Proc. AAAI-91*, pp. 710–716, Menlo Park, CA, 1991.

[11] T. Nishida and S. Doshita. Reasoning about discontinuous change. *Proc. AAAI-87*, pp. 643–648, Seattle, WA, 1987.

[12] R.C. Rosenberg and D. Karnopp. *Introduction to Physical System Dynamics.* McGraw-Hill Publishing Company, New York, NY, 1983.

[13] W. Sweet. The glass cockpit. *IEEE Spectrum*, pp. 30–38, Sept. 1995.