

2. Energy-Based Bond Graph Model Reduction

L. S. Louca¹, D. G. Rideout², T. Ersal³, and J. L. Stein³

¹Department of Mechanical and Manufacturing Engineering, University of Cyprus, Nicosia 1678, Cyprus

²Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada

³Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Abstract Model reduction refers to reducing the complexity of a given model to achieve a balance between model simplicity and accuracy. This chapter presents a set of model reduction techniques that are particularly amenable to bond graph models due to the common energy-based nature of these techniques and the bond graph language. Three techniques are presented that are developed with model order reduction, model partitioning, and simultaneous order and structure reduction in mind. Each technique utilizes a different energy-based metric that can be easily calculated from a bond graph model. These underlying metrics are presented first, followed by the algorithms, each with a simple illustrative example, as well as summaries of larger case studies performed with those algorithms to highlight their benefits. All three techniques are applicable to nonlinear models in differential-algebraic form, are realization preserving in the sense that the original meanings of the states and parameters are preserved, are trajectory dependent and thus explicitly take the specific inputs and parameter values into account, and can reduce models directly at the bond graph level.

2.1. Introduction

The viability of a model for engineering system analysis, design, and control development rests on its *accuracy* and *simplicity*. Model accuracy is critical for understanding, optimizing, and controlling the dynamics of a given system effectively. Model simplicity, on the other hand, is essential for gaining insights and for tractability in system identification and optimization. Simpler models also lead to lower-order controllers that are easier to implement.

Seeking model accuracy and simplicity simultaneously, however, typically engenders a tradeoff: increasing the accuracy of a system model often necessitates increasing the complexity of the model to a level more commensurate with the complexity of the real system. In other words, the requirements of model accuracy and simplicity often compete, and must hence be traded off. This competition typ-

ically grows as engineering systems become larger, more complex, and more integrated; a trend in many modern engineering disciplines. Thus, there is a growing need for system models that mitigate this competition and balance accuracy and simplicity by only capturing the dynamics necessary for their respective applications. The literature, in recognition of this need, deems a dynamic system model *proper for an application* [1] if it provides the accuracy required for that application with minimal complexity. Note that the definition of a proper model in this context is application dependent and different from the control theoretic definition of a proper transfer function, which refers to a transfer function in which the degree of the numerator does not exceed the degree of the denominator.

Obtaining a proper model is not a trivial task. It is not always obvious which phenomena are important for a specific application, and, hence, what to include in the model and what to neglect. Therefore, dynamic system models are seldom proper at the outset. To remedy this problem, the literature proposes many techniques for obtaining proper models; a broad review of these techniques can be found in [2]. Some proper modeling techniques begin with simple models and increment their complexity until they meet their respective accuracy requirements; a process known as *model deduction*. Most techniques, however, begin with excessively complex models and then *reduce* them until they become proper.

This chapter considers the reduction approach to proper modeling, and describes a set of model reduction techniques that are particularly amenable to bond graph models in the sense that the techniques take advantage of the explicit energetic nature of bonds in a bond graph model and yield reductions not only at the equation level, but also directly at the graph level.

Three techniques are covered in detail in this chapter. The first one, the *Model Order Reduction Algorithm (MORA)*, uses the L^1 norm of power, referred to as *activity*, to rank the energy storage and dissipation elements in a bond graph and reduce the model by eliminating the least active elements. The second technique is a decoupling identification and partitioning algorithm that applies a variation of the activity metric, namely, *relative activity*, to the junction structure in a bond graph to find local sites of weak coupling and ultimately partition models into “driving” and “driven” subsystems, which can lead to reduction of model order and junction structure. The third technique is based on another metric, namely, the *Energetic Contribution Index (ECI)*, which considers not only the magnitudes of, but also the correlations between energy flow trajectories in the bonds for an improved assessment. The ECI-based reduction technique ranks the bonds in a bond graph model according to their ECI and identifies various possibilities for eliminating bonds from the model, thereby obtaining simultaneous model order and structure reduction.

The chapter starts with the description of the metrics that the techniques are based on. Then, the reduction techniques are presented together with simple examples illustrating their mechanics. Next, brief summaries of larger scale case studies are given to demonstrate the techniques’ performances on larger scale system models. Finally, a discussion of the advantages and limitations of the techniques is given along with possible directions for future research.

2.2. Model Reduction Metrics

In the heart of any reduction technique lies a metric to evaluate which phenomena can be neglected in a given model and which are important to keep. This section presents the three metrics that underpin the subsequent reduction techniques. These metrics are the activity, the relative activity, and the Energetic Contribution Index (ECI).

All three metrics are based on the concepts of power and energy, which are common underlying phenomena in all physical systems. A formalism such as bond graphs, which represent the power/energy topography of a dynamic system, is therefore particularly amenable to implement these metrics and the algorithms based on them. The power of each bond in a bond graph model is readily available as the product of the corresponding generalized effort (e) and generalized flow (f). The metrics are applicable to other formalisms, as well, as long as the energetic interactions are made explicit.

The metrics assume a general system that can be described by differential-algebraic equations as follows:

$$\begin{aligned} \dot{\mathbf{x}}_i &= \mathbf{F}(\mathbf{x}_i, \mathbf{x}_d, \mathbf{u}), & \mathbf{x}_i(0) &= \mathbf{x}_{i0} \\ \mathbf{0} &= \mathbf{G}(\mathbf{x}_i, \mathbf{x}_d) \end{aligned} \quad (2.1)$$

where $\mathbf{x}_i \in \mathfrak{R}^{n_i}$ is the independent state vector, and n_i is the number of independent states, $\mathbf{x}_d \in \mathfrak{R}^{n_d}$ is the dependant state vector, and n_d is the number of dependent states, $\mathbf{u} \in \mathfrak{R}^m$ is the input vector, and m is the number of inputs, $\mathbf{F} : \mathfrak{R}^{n_i} \times \mathfrak{R}^{n_d} \times \mathfrak{R}^m \rightarrow \mathfrak{R}^{n_i}$ is a nonlinear vector function, $\mathbf{G} : \mathfrak{R}^{n_i} \times \mathfrak{R}^{n_d} \rightarrow \mathfrak{R}^{n_d}$ is a nonlinear constraint vector function, $\mathbf{x}_{i0} \in \mathfrak{R}^{n_i}$ is the initial condition vector of the independent states. In bond graph terms, this corresponds to bond graph models that can include derivative causalities and Lagrange multipliers [3, 4]. Given any input excitation \mathbf{u} and initial conditions \mathbf{x}_{i0} , the state equations in Eq. (2.1) can be solved analytically or numerically to obtain the time response of the state variables $\mathbf{x}_i(t)$ and $\mathbf{x}_d(t)$. The solutions of the states can then be used to determine the required efforts and flows for calculating the power of any bond in the model.

2.2.1. Activity: A Metric for Assessing Aggregate Energy Flow

The power associated with each element in a bond graph provides an indication of the element's contribution to the total energetic behavior of the system. Thus, if an element has high power associated with it, then this suggests that it is storing or absorbing a "significant" portion of the power that it is supplied into the system through the inputs and, therefore, this high power element can be expected to be important to the system's behavior.

However, using power as a modeling metric would lead to instantaneous and time varying decisions about the importance of elements, thus producing models

that are valid for a single time instance. Instead of instantaneous power, an aggregate measure of power flow in or out of an element during a given time window is desired. Previous work by Rosenberg and Zhou [5] proposed and implemented RMS power, chosen from several alternative metrics that produce a single quantity from power. While the physical meaning of RMS power is clear in steady-state sinusoidal signals, its physical interpretation in assessing element importance in a dynamic system is not. Therefore, a new measure of the power response that has a simpler definition and clearer physical interpretation is used here. This new power response measure is called activity, A , and is defined as the L^1 norm of power; i.e.,

$$A = \int_{\tau_1}^{\tau_1+T} |P(t)| dt = \int_{\tau_1}^{\tau_1+T} |e(t)f(t)| dt \quad (2.2)$$

where $P(t)$ is the element power, $e(t)$ and $f(t)$ are the effort and flow of an energy element, respectively, τ_1 is the beginning of the time period T over which the model has to accurately predict the system behavior. Element power can also be calculated using the constitutive law of each energy element as follows:

$$\begin{aligned} \mathbf{I}: P_I(t) &= e_I f_I = e_I \Phi_I \left(\int e_I dt \right) = e_I \Phi_I(p) \\ \mathbf{C}: P_C(t) &= e_C f_C = \Phi_C \left(\int f_C dt \right) f_C = \Phi_C(q) f_C \\ \mathbf{R}: P_R(t) &= e_R f_R = \Phi_R(f_R) f_R \end{aligned} \quad (2.3)$$

where $\Phi_I(p)$, $\Phi_C(q)$, and $\Phi_R(f)$ are known scalar, in general nonlinear, constitutive functions, p is the generalized momentum, and q is the generalized displacement.

Activity has the units of energy, representing the total amount of energy that flows in and out of an energetic element over the specified time window T . However, activity is a different quantity from energy because of the absolute value in its definition. For example, the change in stored energy of an ideal energy storage element (\mathbf{I} or \mathbf{C}) over a power oscillation cycle is zero, however, the activity over the same cycle is nonzero, strictly positive and monotonically increasing as shown in Fig. 2.1.

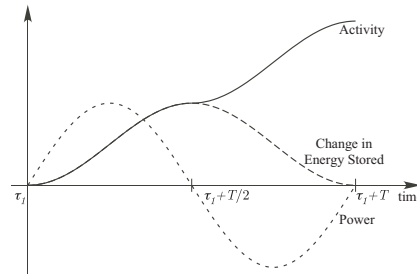


Fig. 2.1. Energy stored in an energy storage element versus the Activity for an oscillation of Power

In summary, to calculate the activity of each element, first the time response of the state variables is determined by integrating the state equations in Eq. (2.1). Then, element power is calculated from Eq. (2.3), and finally, the activity is calculated as defined in Eq. (2.2) for each energy element in the model.

2.2.1.1. Activity Index

To get a relative measure of element importance, the element activity is compared to a quantity that represents the “overall activity” of the system. This quantity is called *total activity* and is defined as the sum of all the energetic element activities of the system; i.e.,

$$A^{total} = \sum_{i=1}^k A_i = \sum_{i=1}^k \left\{ \int_{\tau_1}^{\tau_1+T} |P_i(t)| dt \right\} \quad (2.4)$$

where A_i is the activity of the i^{th} element and k is the total number of energy elements in the system.

The total activity, A^{total} , represents the total amount of energy that flows through the system’s energy elements over the given time T , and, therefore, it indicates the activity level of the system. This quantity is used to calculate a normalized measure of element importance called *element activity index* or just *activity index*, and is given by:

$$AI_i = 100 \frac{A_i}{A^{total}} = 100 \frac{\int_{\tau_1}^{\tau_1+T} |P_i(t)| dt}{\sum_{i=1}^k \left\{ \int_{\tau_1}^{\tau_1+T} |P_i(t)| dt \right\}} \quad i = 1, \dots, k \quad (2.5)$$

The activity index is calculated for each element in the model and it represents the portion of the total system energy that flows through a specific element. Thus, Eq. (2.5) provides a ranking of all k elements relative to the total energy flowing through all the elements in the system. It is proposed that an element with a low activity index has a small contribution to the system dynamic response, thus is unnecessary under the given scenario and, therefore, can be eliminated from the model to generate a reduced model. This elimination procedure is described in the section on *Model Reduction Algorithms*.

2.2.2. Relative Activity

As described above, eliminating energy storage and dissipative elements with low activity index will be shown to be an effective means of reducing model order.

Extending the application of activity to the bond graph junction structure can give further model reduction opportunities by unearthing weakly coupled sets of elements in the model. Bonding of graph elements represents a connection of physical devices at power ports, at which the power variables (effort and flow) of the connected elements must be constrained to be equal. If only one of the power variables is required to be equated between the elements, then one-way coupling exists which can lead to model partitioning and reduction.

“Junction structure” refers to the 1- and 0-junctions capturing the constraint equations that link the constitutive law variables of elements. Decoupling or one-way coupling among elements creates negligible constraint equation terms, which manifest themselves as bonds with relatively low aggregate power flow at a 1- or 0-junction [6]. The activity of a bond attached to a 0- or 1-junction, compared to the activities of the other bonds at that junction, can then be used as a measure of the relative importance of the associated constraint equation term compared to the other terms in the equation. Such local comparison entails dividing each bond activity A_i at the junction by the maximum bond activity A_{\max} at the junction to get “Relative Activity” RA ; i.e., for bond i at a junction,

$$RA_i = \frac{A_i}{A_{\max, \text{ junction}}} \quad (2.6)$$

Suppose any relative activity that falls below a user-defined threshold ε is deemed negligible. Negligible relative activity of bond i implies that for a 0-junction, the flow f_i can be neglected in the flow constraint equation, and for a 1-junction, the effort e_i can be neglected in the effort equation.

A power bond with low relative activity can be “conditioned”, or converted to a modulated source and signal. This leads to the partitioning and model reduction algorithm described in the *Model Reduction Algorithms* section.

2.2.3. Energetic Contribution Index (ECI)

Energetic Contribution Index (ECI), previously referred to as *Relative Importance* in the literature [7, 8], is a metric that is based on the same energy-based intuition behind metrics such as RMS power or activity. However, unlike these other metrics, ECI works directly with energy instead of power, and is geared towards capturing not only the *magnitudes* of, but also the *correlations* between the energy flow patterns in a model for an improved assessment of the relative importance of both the energetic *components* and their *interactions*, thereby enabling simultaneous *structure* and *order* reduction.

The ECI is developed as follows. Let $e_i(t)$ denote the energy in a given bond i , which is given by the time integral of the product of the generalized effort and flow variables associated with the bond, i.e.,

$$e_i(t) = \int_{t_0}^t \text{effort}(\tau) \text{flow}(\tau) d\tau \quad (2.7)$$

Construct the energy vector

$$\mathbf{e}(t) \triangleq [e_1(t) \ e_2(t) \ \dots \ e_n(t)]^T \quad (2.8)$$

where n is the number of bonds in the given bond graph. The energy trajectory matrix of the system for a time window of $[t_0, t_f]$ is then given by

$$\mathbf{E} \triangleq \begin{bmatrix} \mathbf{e}^T(t_0) \\ \mathbf{e}^T(t_1) \\ \vdots \\ \mathbf{e}^T(t_f) \end{bmatrix} \quad (2.9)$$

As an intermediate step towards developing the ECI, let our first goal be to optimally compress the information contained in \mathbf{E} . In mathematical terms, this corresponds to looking for an optimal lower-rank approximation to \mathbf{E} , or, equivalently, to minimizing the following error residual:

$$J = \int_{t_0}^{t_f} \|\mathbf{e}(t) - \mathbf{e}_r(t)\|^2 dt \quad (2.10)$$

where $\mathbf{e}_r(t)$ denotes an approximation to $\mathbf{e}(t)$ that resides in an r -dimensional subspace. The solution to this minimization problem is given by the Kahunen-Loève Expansion (KLE) [9, 10], also known as principal component analysis [11], the method of empirical orthogonal functions [12], proper orthogonal decomposition [13], singular value decomposition [14], empirical eigenfunction decomposition [15-17], or the method of quasi-harmonic modes [18]. Specifically, let us define the energy Gramian

$$\mathbf{W} \triangleq \int_{t_0}^{t_f} \mathbf{e}(t) \mathbf{e}^T(t) dt \geq 0 \quad (2.11)$$

with real eigenvalues $\sigma_1^2 \geq \dots \geq \sigma_n^2 \geq 0$ and orthonormal eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Then, the following are known from the KLE theory [19]:

1. The energy vector $\mathbf{e}(t)$ can be decomposed into the orthogonal components $\mathbf{v}_i c_i(t)$, i.e., $\mathbf{e}(t) = \sum_{i=1}^n \mathbf{v}_i c_i(t)$, where $c_i(t) = \mathbf{v}_i^T \mathbf{e}(t)$ is the component function.
2. The magnitude, as measured by the squared norm, of each component $\mathbf{v}_i c_i(t)$ is given by its corresponding eigenvalue, i.e., $\int_{t_0}^{t_f} \|\mathbf{v}_i c_i(t)\|^2 dt = \sigma_i^2$.
3. The error residual $J = \int_{t_0}^{t_f} \|\mathbf{e}(t) - \mathbf{e}_r(t)\|^2 dt$ is minimized by $\mathbf{e}_r(t) = \sum_{i=1}^r \mathbf{v}_i c_i(t)$ and is equal to $\sum_{i=r+1}^n \sigma_i^2$.

In other words, the KLE provides an optimal set of orthogonal basis vectors that minimizes the error residual, i.e., for a given r there is no better choice of an orthogonal set of vectors than the first r vectors given by KLE. Traditionally, model

reduction using KLE involves projecting the equations on the subspace spanned by these basis vectors. However, to avoid a change in realization, no projection will be performed at this point.

Instead, let us interpret the abovementioned facts as follows: the eigenvectors provide an optimal orthogonal basis for the full space of $\mathbf{e}(t)$, and the corresponding eigenvalues quantify how much signal energy is captured by each eigenvector. Since the observed quantity is energy, the eigenvectors give a new basis to express the energy flow in the system, so they can be interpreted as energy exchange modes in the system. Within each mode, the (absolute value of the) j^{th} component tells how much the j^{th} bond contributes to that mode. Thus, an intuitive measure for the energetic contributions of the bonds is proposed as a weighted combination of the absolute values of the eigenvectors $\bar{\mathbf{v}}_i$, where the weights are the eigenvalues σ_i^2 , i.e.,

$$\tilde{\mathbf{c}} \triangleq \sum_{i=1}^n \sigma_i^2 |\mathbf{v}_i| \quad (2.12)$$

where the j^{th} component of $\tilde{\mathbf{c}}$ provides an indication for the energetic contribution of bond j . The absolute value $|\mathbf{v}_i|$ indicates the absolute value of each component of the vector, and not a norm of the vector. The vector $\tilde{\mathbf{c}}$ can be normalized with respect to its maximum element to give a relative measure of energetic contribution, i.e.,

$$\mathbf{c} \triangleq \tilde{\mathbf{c}} / \max(\tilde{\mathbf{c}}) \quad (2.13)$$

where \mathbf{c} is referred to as the Energetic Contribution Index (ECI) vector of the bonds.

Naturally, since the energy vector will depend on the particular inputs, initial conditions, and parameter values used for simulation, it will be trajectory-dependent, and so will be the ECI vector. This will allow for tailoring the reduction to specific scenarios of interest as will be illustrated later.

2.2.3.1. A Note on Computing the ECI

Note that the energy Gramian in Eq. (2.11) provides an analytical way of computing the ECI. However, obtaining the energy Gramian analytically may not always be easy or even possible, especially in nonlinear systems. In that case, the Gramian can be obtained through numerical integration. That would involve a discretization of the integral in Eq. (2.11), which, in its simplest form, can be written as follows:

$$\mathbf{W} \approx \mathbf{e}(t_1)\mathbf{e}^T(t_1)\Delta t_1 + \mathbf{e}(t_2)\mathbf{e}^T(t_2)\Delta t_2 + \dots + \mathbf{e}(t_f)\mathbf{e}^T(t_f)\Delta t_f \quad (2.14)$$

with $\Delta t_k = t_k - t_{k-1}$. For brevity, let $e_i^k \triangleq e_i(t_k)\sqrt{\Delta t_k}$, and $e_{ij}^k \triangleq e_i(t_k)e_j(t_k)\Delta t_k$. Then, Eq. (2.14) can be manipulated as follows:

$$\begin{aligned}
\mathbf{W} &\approx \begin{bmatrix} e_1^1 \\ e_2^1 \\ \vdots \\ e_n^1 \end{bmatrix} \begin{bmatrix} e_1^1 & e_2^1 & \dots & e_n^1 \end{bmatrix} + \begin{bmatrix} e_1^2 \\ e_2^2 \\ \vdots \\ e_n^2 \end{bmatrix} \begin{bmatrix} e_1^2 & e_2^2 & \dots & e_n^2 \end{bmatrix} + \dots + \begin{bmatrix} e_1^f \\ e_2^f \\ \vdots \\ e_n^f \end{bmatrix} \begin{bmatrix} e_1^f & e_2^f & \dots & e_n^f \end{bmatrix} \\
&= \begin{bmatrix} e_{11}^1 + e_{11}^2 + \dots + e_{11}^f & e_{12}^1 + e_{12}^2 + \dots + e_{12}^f & \dots & e_{1n}^1 + e_{1n}^2 + \dots + e_{1n}^f \\ e_{21}^1 + e_{21}^2 + \dots + e_{21}^f & e_{22}^1 + e_{22}^2 + \dots + e_{22}^f & \dots & e_{2n}^1 + e_{2n}^2 + \dots + e_{2n}^f \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1}^1 + e_{n1}^2 + \dots + e_{n1}^f & e_{n2}^1 + e_{n2}^2 + \dots + e_{n2}^f & \dots & e_{nn}^1 + e_{nn}^2 + \dots + e_{nn}^f \end{bmatrix}_{n \times n} \\
&= \begin{bmatrix} e_1^1 & e_1^2 & \dots & e_1^f \\ e_2^1 & e_2^2 & \dots & e_2^f \\ \vdots & \vdots & \ddots & \vdots \\ e_n^1 & e_n^2 & \dots & e_n^f \end{bmatrix}_{n \times f} \begin{bmatrix} e_1^1 & e_2^1 & \dots & e_n^1 \\ e_1^2 & e_2^2 & \dots & e_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_1^f & e_2^f & \dots & e_n^f \end{bmatrix}_{f \times n} = \mathbf{S}^T \mathbf{S}
\end{aligned} \tag{2.15}$$

Now recall the relationship between the singular value decomposition of an arbitrary matrix \mathbf{S} , and the eigenvalue decomposition of $\mathbf{S}^T \mathbf{S}$:

$$\mathbf{S}^T \mathbf{S} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T)^T (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) = \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^T \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma}) \mathbf{V}^T \tag{2.16}$$

In other words, the eigenvalues of $\mathbf{S}^T \mathbf{S}$ are the squares of the singular values of \mathbf{S} , and the eigenvectors of $\mathbf{S}^T \mathbf{S}$ are the right singular vectors of \mathbf{S} . Thus, the eigenvalues, σ_i^2 , and eigenvectors, \mathbf{v}_i of $\mathbf{S}^T \mathbf{S}$, and hence of \mathbf{W} , can be computed from the singular value decomposition of \mathbf{S} . This, actually, gives a numerically better way of computing σ_i^2 and \mathbf{v}_i , because calculation of singular values is a numerically better-posed problem than calculating eigenvalues. The algorithm that is proposed later in this chapter is based on this way of calculating the ECI.

2.3. Model Reduction Algorithms

The previous section defined activity A , relative activity RA , and energetic contribution index ECI . These metrics underpin the following algorithms for reduction of model order and/or structure and system partitioning. These algorithms assume that a model satisfying the accuracy goal is given, but is overly complex, and hence a balance between accuracy and simplicity is sought through a reduction approach. This initial model is hereafter referred to as the full model.

2.3.1. Model Order Reduction Algorithm (MORA)

The first model reduction algorithm based on the activity metric is shown in Fig. 2.2 and is called *Model Order Reduction Algorithm (MORA)*. Given the full model, the goal of MORA is to order the importance of the energy elements in that model as given by their activity, and reduce the size of the model based on a user supplied threshold of the percent of the total activity to be retained in the reduced model.

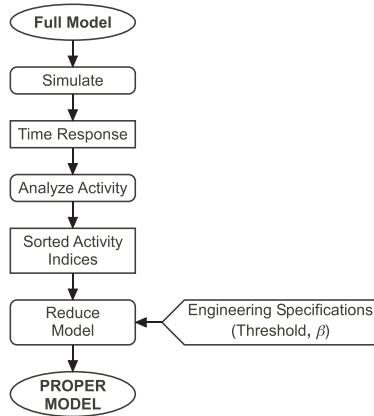


Fig. 2.2. Model Order Reduction Algorithm (MORA)

At the end of the simulation of the full model, the time response of the output variables in Eq. (2.3) is available for the activity analysis. Given these output variables, the element power is calculated. Then, the activity metric is calculated using Eq. (2.2). Finally, the activity index in Eq. (2.5) of each element is calculated. The activity indices are then sorted in descending order to identify the elements with high activity, considered more important, and low activity, considered less important. The result of the sorting process is the ranking of the element importance and it is represented with the vector \mathbf{r} as shown in Eq. (2.17). The first element of this vector points to the element with the highest activity and the last points to the one with the lowest activity.

$$\mathbf{r} = \begin{Bmatrix} r_1 \\ \vdots \\ r_k \end{Bmatrix} \equiv \begin{Bmatrix} \text{highest activity element} \\ \vdots \\ \text{lowest activity element} \end{Bmatrix} \quad (2.17)$$

Sorting of the activity indices provides the critical information needed to reduce the full model. Based on a user supplied threshold (β) of how much of the total activity is required to retain sufficiently accurate predictions for a particular engineering design task, MORA trims appropriate energy elements from the bond

graph. This threshold defines the borderline between the elements that are going to be retained in the model and those that are going to be eliminated. This process is shown graphically in Fig. 2.3.

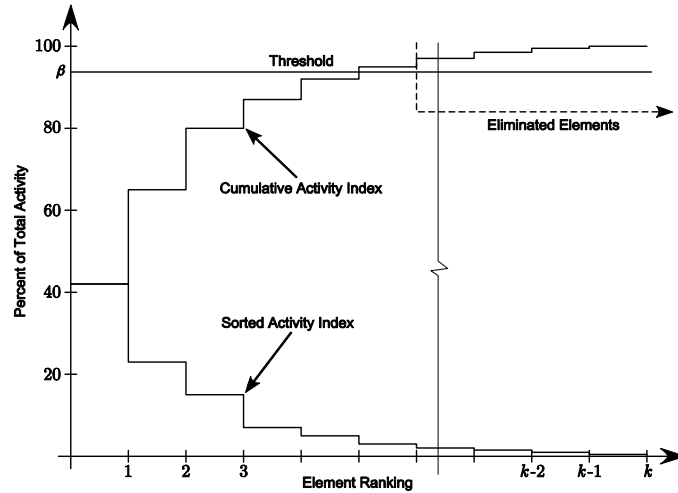


Fig. 2.3. Activity Index sorting

The reduction algorithm sums the sorted activity indices until the Cumulative Activity Index (CAI) exceeds the user-specified threshold. The summation process starts with element that is considered the most important (r_1) and the cumulative activity index is set to be equal to the activity index of element r_1 . Then, the next important element (r_2) is added and the cumulative activity index is increased by the addition of the activity index of r_2 to the previous cumulative activity index. For a given number of included elements the cumulative index is given by:

$$CAI_i = CAI_{i-1} + AI_{r_i}, \quad i = 1, \dots, k, \quad CAI_0 = 0 \quad (2.18)$$

This procedure is repeated until the cumulative activity index exceeds the specified threshold (β) as shown in Fig. 2.4. The elements that have been accounted for in the cumulative activity index are the significant elements, and therefore, are the only elements included in the reduced model. The remaining elements of the full model are eliminated.

MORA generates reduced models by eliminating ideal elements from the full model. In the case where the full model is represented as a bond graph, the elimination of low activity elements is achieved simply by removing these low activity energy elements and their connecting bonds. The junction structure of the bond graph is retained in the reduced model, and therefore, the reduced model realization is the same as the full model, so its physical meaning and relation to the physical system is retained.

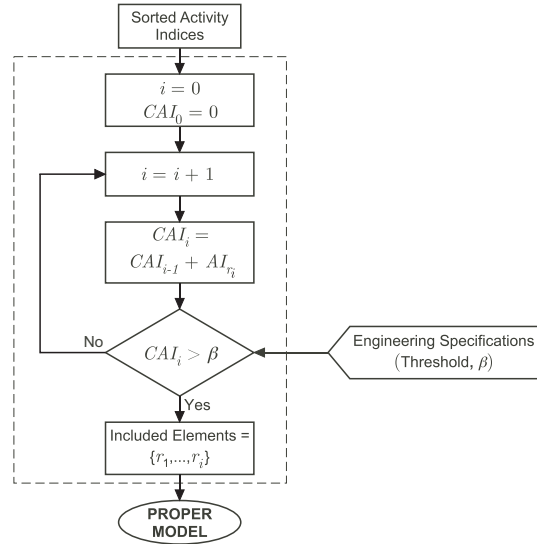


Fig. 2.4. Detail of Reduce Model procedure in Fig. 2.2

2.3.1.1. Illustrative Example: Nonlinear Quarter Car Model

A quarter car ride quality model is used to illustrate MORA. This model is a standard vehicle dynamics model used in automotive engineering for ride quality evaluation [20] and it is depicted in Fig. 2.5. Though it is already a relatively small model (two DOF, requiring four state variables), it still serves the purpose of illustrating the proposed ideas, as well as showing that even a relatively small model may not be proper.

The assumed full model of the vehicle is shown in Fig. 2.5 and consists of the sprung mass (car body, engine, etc.) and the unsprung mass that accounts for the wheel and axle masses supported by the tire. The suspension is modeled as a spring and a damper in parallel, which connects the unsprung to the sprung mass. The tire is also modeled as a spring and a damper in parallel, and represents the transfer of the road force to the unsprung mass through the tire's elastic and viscoelastic properties. It is assumed that the model is excited by large amplitude road inputs, and therefore, the nonlinear force/deflection characteristics (stiffening) of the suspension and tire springs are included. In addition, the tire is modeled such that it can only carry compressive loads to capture wheel lift off that can occur on tough roads under high forward vehicle speeds. The tire force for contact conditions is given by $e(q) = a_1 \cdot q + a_3 \cdot q^3$, whereas for no contact (lift off) the tire force is zero, i.e., $e(q) = 0$. This model feature introduces a model discontinuity that demonstrates the applicability of the activity metric to highly nonlinear systems.

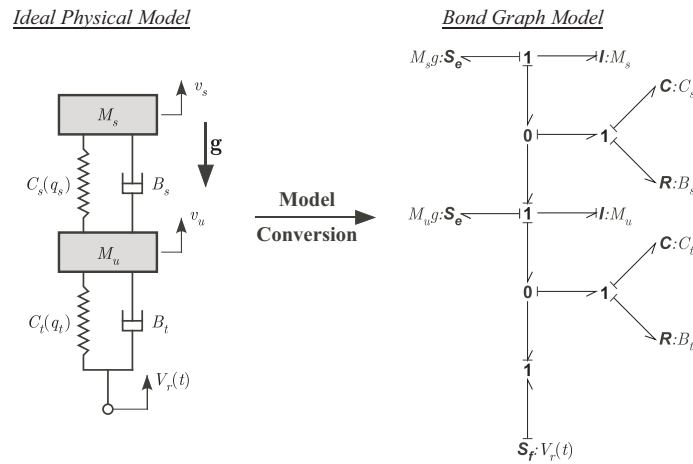


Fig. 2.5. Nonlinear quarter car model representations – Full model

Table 2.1. Vehicle Parameters

Ideal Energy Element	Parameter
Sprung Mass	$M_s = 267.0 \text{ kg}$
Suspension Damping	$B_s = 700.0 \text{ N} \cdot \text{s} / \text{m}$
Suspension Stiffness	$a_1 = 18,872 \text{ N} / \text{m}, \quad a_3 = 2 \times 10^5 \text{ N} / \text{m}^3$
Unsprung Mass	$M_u = 36.6 \text{ kg}$
Tire Damping	$B_t = 200.0 \text{ N} \cdot \text{s} / \text{m}$
Tire Stiffness	$a_1 = 193,915 \text{ N} / \text{m}, \quad a_3 = 2 \times 10^8 \text{ N} / \text{m}^3$

The bond graph of the full model is also given in Fig. 2.5 and has six ideal energy elements, two of each type (**I**, **C**, and **R**), and their parameters are given in Table 2.1. Note that the model includes only the system dynamics in the vertical direction; the constant forward speed is used only to convert the spatial road description $Z_r(x)$ into a temporal vertical velocity input, $V_r(t)$, at the road/tire interface as shown in Fig. 2.6. The vehicle is assumed to be traveling at a constant forward speed V_F on a flat road and at some point $x = x_0$ it reaches a “smooth” curb. The transition of the road profile from low to high level ($Z = 0.2 \text{ m}$) is described by a cycloid function, whereas the length of the curb is $L = 0.5 \text{ m}$.

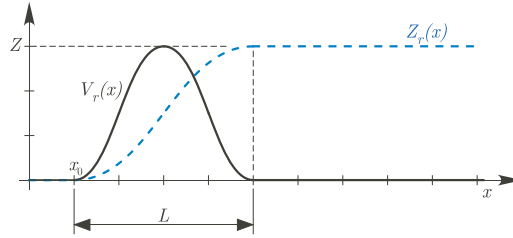


Fig. 2.6. Road profile - Road elevation, $V_r(x)$ and velocity, $Z_r(x)$

Due to the simplicity of this model the state and output equations are derived by hand and transferred into MATLAB [21]. This step could have been done using a bond graph modeling environment (e.g., 20Sim [22], CAMP-G [23]) to generate the time responses needed for the “Analyze Activity” step of MORA (see Fig. 2.2). The dynamic equations are numerically integrated to first produce the time response of the state variables and then the required set of outputs as defined in Eq. (2.3).

This illustrative example is divided into three studies. The first study explores the proposed model reduction procedure by showing the outputs of many of the intermediate steps leading to the proper model for a single “high forward speed” input. Because the proper model is scenario dependent, a second study explores the effect of changing the input to a “smoother” input (lower vehicle forward speed than that used in the first study) on the proper model produced. Finally, a third study explores the effect of the user defined “model reduction threshold”, (β), on the reduced model accuracy, by generating a series of reduced models under the same input conditions.

High Forward Vehicle Speed Scenario

In this first study, the vehicle model is exercised as it travels over the curb at a constant forward speed, $V_F = 5 \text{ m/s}$. This high forward speed generates a severe velocity input that approximates an impulse function; the duration of the input is only 0.1 s . The activity is calculated as a function of time by setting the lower bound, τ_1 , to zero and varying the time window, T , of the integration in Eq. (2.2). As shown in Fig. 2.7, the activities remain at zero until the vehicle hits the curb, at which point power starts to flow into the system. The activities increase due to the nonzero power flow until they approach a steady state value as the system transients die out. Note the discontinuity in the slope of the activities (especially for tire stiffness and damping) at around 1.5 s . The high forward speed causes the wheel to lift off as the vehicle drives over the curb and contact is restored at about 1.5 s . This causes an impact force that results in the rapid increase in the activities.

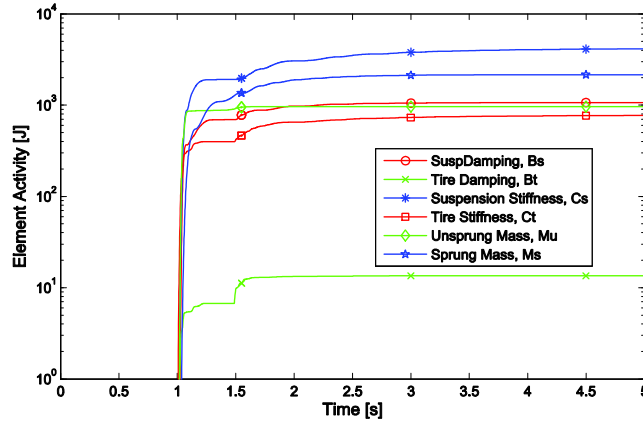


Fig. 2.7. Element Activity, $V_f = 5 \text{ m/s}$

The steady state value of the activities ($\tau_1 = 0 \text{ s}$ and $T = 5 \text{ s}$) is used to generate a reduced model. This value accounts for all the power flow during the entire dynamic event (both transient and steady state effects), and thus, a reduction based on these activities hypothetically should produce a model that accurately predicts the system behavior over the same time interval.

The sorted steady state values of the element activities are shown in Table 2.2. There is only one element, namely the tire damping, that is clearly separated from the others. The tire damping has a 0.15% activity index where the next most important element, the tire stiffness, has an 8.46% activity index. The most important element is the suspension stiffness, which accounts for more than 45% of the total activity of the system. The sprung mass, suspension damping, and unsprung mass are the next most important elements.

Table 2.2. Element Activity and Ranking, $V_f = 5 \text{ m/s}$

Rank	Element Name	Activity [J]	Activity Index [%]	Cumulative Activity Index [%]
r_1	Suspension Stiffness	4139	45.44	45.44
r_2	Sprung Mass	2155	23.65	69.09
r_3	Suspension Damping	1066	11.70	80.79
r_4	Unsprung Mass	966.2	10.61	91.39
r_5	Tire Stiffness	770.7	8.457	99.85
r_6	Tire Damping	13.49	0.1481	100.00

Based on a model reduction threshold, $\beta = 95\%$, the element elimination is completed as outlined in Fig. 2.4. MORA identifies the tire damping as the only element that can be eliminated from the full model for this threshold. Note that the

algorithm keeps adding energy elements until the CAI becomes bigger than the predefined threshold. As Table 2.2 shows, the reduced model includes five out of the six elements and it retains 99.85% of the total activity. The tire damping is eliminated from the model and the reduced model bond graph is given in Fig. 2.8. The reduced model maintains the same number of states (4), but has one energy element eliminated. The equivalent ideal physical model is also shown in Fig. 2.8, which is developed using a systematic procedure for the physical interpretation of ideal element elimination as suggested by the activity metric [24].

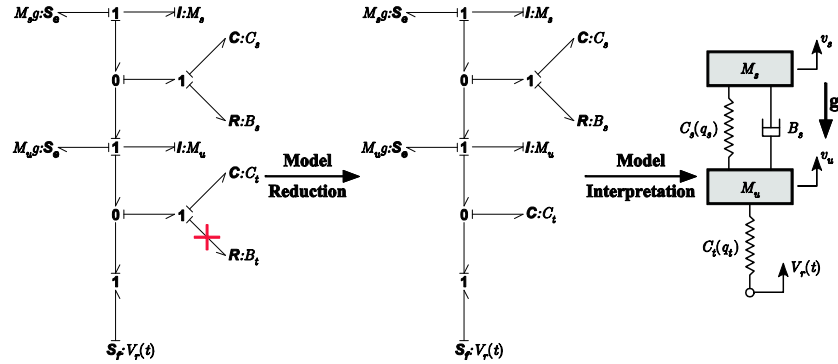


Fig. 2.8. Model reduction, $V_r = 5 \text{ m/s}$, $\beta = 95\%$

For assessing the accuracy of the reduced model, the dynamic response of the reduced model is compared to the one of the full model. For all four states, the reduced model accurately predicts the system behavior, but for a more rigorous comparison, a quantitative comparison is used. The average relative error (ε) is calculated using the following expression:

$$\varepsilon = 100 \int_{\tau_1}^{\tau_1+T} |w(t) - w_r(t)| dt \bigg/ \int_{\tau_1}^{\tau_1+T} |w(t)| dt \quad (2.19)$$

where $w(t)$ and $w_r(t)$ are the responses of the full and reduced model, respectively. This accuracy measure is not unique and other metrics can be used leading to similar conclusions.

The average error, ε , of each state is calculated using the metric in Eq. (2.19), and the accuracy for the sprung mass velocity, unsprung mass velocity, suspension deflection, and tire deflection is 0.38%, 6.1%, 0.10% and 0.57%, respectively. The average errors for the sprung mass and suspension spring states are expected to be lower, since these components have a small high frequency component and the eliminated tire damping has a higher affect on the high frequency mode.

For visual comparison, also the error of the tire contact force is plotted as shown in Fig. 2.9. Notice that the error has a large high frequency component and

a much smaller low frequency one, which indicates the reduced model retains more the lower than the higher frequency content of the full model.

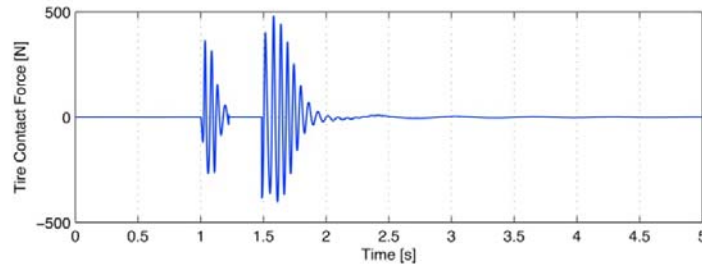


Fig. 2.9. Tire contact force error, $V_F = 5 \text{ m/s}$

Low Forward Vehicle Speed Scenario

A lower forward speed is used to demonstrate how the element activities and the reduced model are affected by the system input. All the other simulation conditions remain the same as in the high forward speed study.

Table 2.3 shows the steady state activity indices as sorted by MORA. The table shows that the most important element for this scenario is the suspension stiffness, which utilizes most of the energy that flows through the system (60.86%). In decreasing importance are the sprung mass, tire stiffness, and suspension damping. A significant decrease in activity is then observed. The least important elements are the unsprung mass and tire damping, with the tire damping absorbing almost no energy (0.02% activity index).

Table 2.3. Element Activity and Ranking, $V_F = 1 \text{ m/s}$

Rank	Element Name	Activity [J]	Activity Index [%]	Cumulative Activity Index [%]
r_1	Suspension Stiffness	1775	60.86	60.86
r_2	Sprung Mass	719.2	24.66	85.53
r_3	Suspension Damping	200.8	6.89	92.41
r_4	Tire Stiffness	196.6	6.74	99.15
r_5	Unsprung Mass	24.21	0.82	99.98
r_6	Tire Damping	0.5521	0.02	100.00

Using the same threshold as in first study, MORA eliminates the elements that contain less than 5% of the total activity (95% accumulated activity). MORA suggests that two elements, namely the unsprung mass and tire damping (shaded ele-

ments in Table 2.3), should be eliminated from the model. The reduced model retains 99.15% of the total activity and it is generated by eliminating the unsprung mass (M_u) and tire damping (B_t) from the bond graph (see Fig. 2.10). The reduced model has only three states instead of four, since the unsprung mass, an energy storage element, has been eliminated. The equivalent ideal physical model is also shown in Fig. 2.10. The elimination of the unsprung mass removes the inertial effects without removing the gravity force.

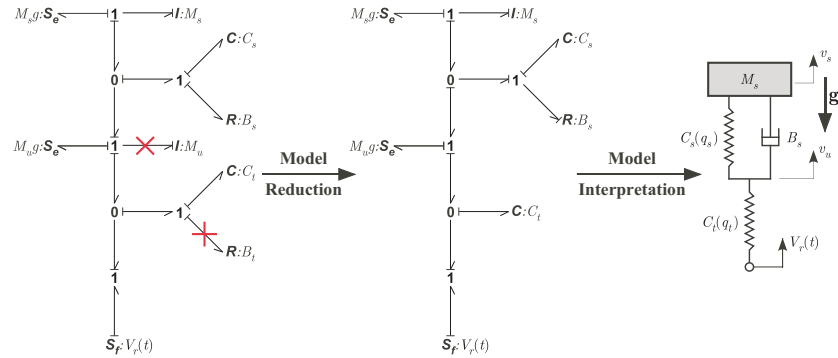


Fig. 2.10. Model reduction, $V_r = 1 \text{ m/s}$, $\beta = 95\%$

For assessing the accuracy of the reduced model, the dynamic response of the reduced model is compared to the one of the full model. The reduced model has only three states, however, the eliminated unsprung mass velocity state can still be computed from the remaining three states. The predictions of all states of the reduced model are "similar" to those of the full model. The error, ε , for each of the four state variables, sprung mass velocity, unsprung mass velocity, suspension deflection, and tire deflection is 1.17%, 1.41%, 0.08%, and 0.25%, respectively. The accuracy of the unsprung mass velocity is the worst, since the state associated with this variable is eliminated from the reduced model.

Element Rank and Model Accuracy

The underlying assumption of MORA is that the lower the activity of an element is, the lower the contribution of that element to the overall model behavior is and, therefore, the element can be eliminated without significant degradation of the model predictions. If this conjecture is true, it is expected that as elements with higher activity are eliminated from the model, the accuracy of the reduced model will be more strongly affected. To test these assumptions, a series of reduced models are generated by varying the reduction threshold, β , in such manner that one additional element, namely, the one with the lowest activity, is eliminated from the next reduced model.

The conditions for this study are the same as the ones used in the second scenario, i.e., vehicle drives over the curb at a low forward speed. Therefore, the same activity analysis as shown in Table 2.3 is used to generate the reduced models in this study. The first reduced model is generated by setting the reduction threshold, $\beta = 99.5\%$, which, according to MORA, maintains five elements and only the tire damping is eliminated from the full model. The accuracy of the states is also calculated using the average error metric defined in Eq. (2.19), and the errors are given in Table 2.4. To have a single accuracy measure for this reduced model, the mean value of all the state variable accuracies is calculated. In addition to the state variables, the tire contact force is used for comparison purposes, and it is plotted versus the full model response in Fig. 2.11. The accuracy of the tire contact force is also given in Table 2.4.

Table 2.4. Accuracy level of reduced models

Reduced Model	Reduction Threshold	Included Elements	Average error as given by metric in Eq. (2.19)					
			State Variables					Tire Contact Force [%]
			v_s [%]	v_u [%]	q_s [%]	q_t [%]	Mean [%]	
1	99.5	5	0.25	0.27	0.02	0.05	0.15	0.02
2	95	4	1.17	1.41	0.08	0.25	0.73	0.31
3	90	3	31.7	26.9	2.15	5.64	16.5	2.41
4	85	2	781	26.9	56.7	5.64	218	62.2

Setting the reduction threshold to 95%, 90%, and 85% generates three additional reduced models. Each of these reduced models has one less energy element as compared to the previous reduced model. For generating the four reduced models, 1, 2, 3, and 4 elements are eliminated respectively. No additional reduced models are generated since further reduction produces models with no physical meaning. The results are given in Table 2.4, as well as Fig. 2.11.

The first two reduced models (Reduced Model 1 and Reduced Model 2) have almost identical predictions for the tire contact force as it is shown in Fig. 2.11 and have an error of 0.02% and 0.31%, respectively. The third reduced model (Reduced Model 3) has some visible differences with an accuracy of 2.41%, while the last reduced model (Reduced Model 4) has an average error of 62.2% and fails to capture the system behavior. As can be seen in Fig. 2.11, the tire contact force prediction of the fourth reduced model has undamped oscillations and it is very different from the one of the full model. This prediction is expected, since this model has no dissipation elements included and consists of only a spring and mass connected in series. Note that in the top plot of Fig. 2.11 there are five curves plotted, one for the full model and one for each one of the four reduced models. However, only three curves (Reduced 2, Reduced 3, Reduced 4) are visible, since the Full and Reduced 1 curves are almost identical with Reduced 2. Similarly, in the bottom plot of Fig. 2.11 there are five curves plotted but only three are visible.

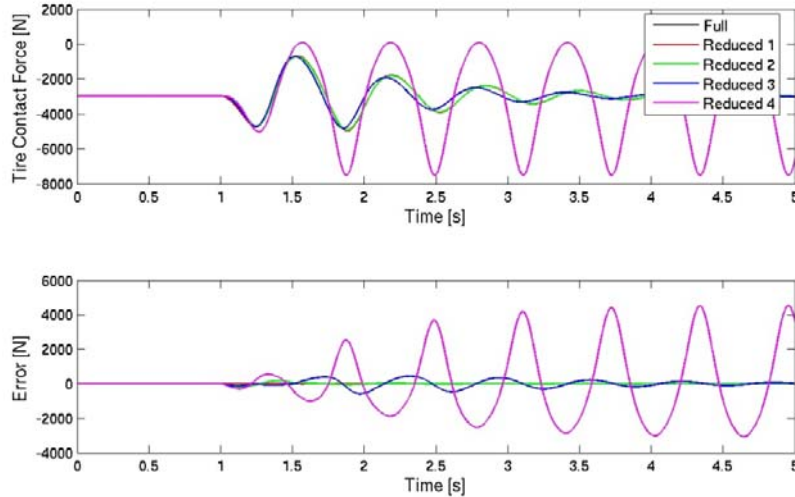


Fig. 2.11. Time histories of full vs. reduced model predictions of tire contact force and the errors in the reduced model predictions

The results in Table 2.4, for both the average error and the tire contact force error, show that the error increases as more elements are eliminated. As elements with a larger percentage of the total activity are eliminated, correspondingly larger errors are produced. For the contact force, the error from the full to the first reduced model is 0.02%, from first to second reduced model is 0.29%, from second to third reduced model is 2.1%, and from third to fourth reduced model is 59.8%. The average errors of the four states exhibit a similar trend in the accuracy drop. This shows a relation between accuracy and element activity such that the higher the activity of an eliminated element, the higher the loss of accuracy will be.

2.3.2. Decoupling Identification and Partitioning Algorithm

The algorithm in the previous section establishes the connection between activity and an energetic element's contribution to the overall system dynamics. MORA reduces model order by eliminating inactive elements, while retaining the junction structure. Another proper modeling challenge is to find groups of elements within a bond graph which form partitions, i.e., sets of elements between which there is weak coupling, and to eliminate partitions which are not required to predict outputs of interest. The partitioning problem is tackled in this section by applying the activity metric to the junction structure.

A power bond with low relative activity RA at a junction, as defined in the *Model Reduction Metrics* section, can be “conditioned”, or converted to a mod-

ulated source due to the fact that one of the two power variables that are shared by the elements connected by the bond does not contribute significantly to one of the associated constitutive laws. The conditioning is illustrated in Table 2.5. Scenario (i), Case A, is a case where a bond has low RA at a 0-junction. In other words, activity $A_1 \ll A_{m+1}, \dots, A_n$. Assuming that the low activity is due to relatively low flow in the 0-junction flow summation [24], the low- RA bond is removed from the flow summation by replacing it with a modulated effort source. The modulating signal is the effort out of the junction. This effort is applied by the source to the element at the other end of the original bond, i.e., the end that is not adjacent to the 0-junction at which the bond is relatively inactive. The half-arrow direction (direction of algebraically positive power flow) of the modulated source is the same as that of the original low-activity bond. Case A in Table 2.5 shows a possible partitioning site between the elements to the left and to the right of bond 1, for now a power bond with causal information flow in both directions has been replaced by a modulating signal carrying information only from left to right. At this site, the elements to the right of bond 1 are “driven” by the elements to the left, and do not significantly back-excite the elements to the left.

For a bond with negligible RA at a 1-junction, the bond becomes a modulated flow source upon conditioning, thereby removing a term from the junction effort summation as illustrated in Case B of Scenario (i).

If the RA of an external junction structure bond (I , C , or R element) is negligible, then the element can be eliminated from the model, as its causal output makes an insignificant contribution to the remainder of the system. In Scenarios (ii) and (iii) of Table 2.5, external elements are represented by the symbol Z for a generalized impedance.

Table 2.6 shows an internal bond connecting a junction to an $(M)TF$ or $(M)GY$ element. If both bonds are locally inactive compared to their respective junctions, then the transformer or gyrator can be eliminated. If one bond is locally inactive, a modulated source in sequence with an $(M)TF$ or $(M)GY$ results. The transformer or gyrator can be incorporated into the source, resulting in the equivalent conditioned junction structure shown in Table 2.6. Conversion of $(M)TF$ or $(M)GY$ elements connected to 1- or 0-junctions respectively, results in MSf elements.

Note that, as shown in the Tables 2.5 and 2.6, the locally inactive bonds must be causally weak, i.e., they must not provide the flow input to a 1-junction, nor the effort input to a 0-junction. Causality reassignment may be required.

Table 2.5. Conversion (conditioning) of low relative activity power bonds

<p>Scenario (i)</p>	<p>Case A, $A_1 \ll A_{m+1, \dots, n}$</p>	<p>Case B, $A_1 \ll A_{2, \dots, m}$</p>
<p>Scenario (ii)</p>	<p>Case A, $A_1 \ll A_{2, \dots, m}$</p>	
<p>Scenario (iii)</p>	<p>Case A, $A_1 \ll A_{2, \dots, m}$</p>	

Table 2.6. Transformer and gyrator conversion examples

<p>Transformer</p>	<p>$A_{1a} \ll A_{m+1, \dots, n}$</p>	<p>Equivalent</p>
<p>Gyrator</p>	<p>$A_{1a} \ll A_{m+1, \dots, n}$</p>	<p>Equivalent</p>

Conditioning of all appropriate bonds allows for visual identification of one-way coupled groups of bond graph elements, or “driving and driven partitions”, and subsequent model reduction.

Given a *conditioned* model, i.e., a bond graph with all low- RA power bonds converted to modulated sources, we make the following definitions:

A *subgraph* S is a set of elements from that bond graph that has no power bonds connected to any bond graph element outside the set. The subgraph may be connected to the rest of the bond graph by modulating signals, or may be unconnected.

When all *new* modulating signals (due to bond conditioning) between two subgraphs are directed from one subgraph to the other, then the subgraph from which the signals originate is the *driving subgraph* S^+ , and the other is the *driven subgraph* S^- .

A *subgraph loop* S_L is a set of subgraphs in which each $S \in S_L$ is connected to the previous by modulating signals directed into S , and to the next by modulating signals directed out of S . A subgraph loop can itself be a subgraph. See Fig. 2.12.

A *driving partition* is a subgraph P^{\rightarrow} that is connected to at least one other subgraph *strictly* by modulating signals directed outwards from P^{\rightarrow} to that subgraph. A subgraph loop can be a driving partition.

A *driven partition* is a subgraph P^{\leftarrow} that is not an element of a subgraph loop, and is connected to at least one other subgraph *strictly* by modulating signals directed inwards towards P^{\leftarrow} from that subgraph.

A flow chart of the algorithm to search for locally inactive bonds is shown in Fig. 2.13, and Fig. 2.14 depicts the model conditioning algorithm.

Model conditioning is followed by a search for partitions. The distinction between subgraphs and partitions is drawn because modulating signals may exist, prior to conditioning, which create subgraph loops and prevent partitioning as shown in Fig. 2.12b – even though all new modulating signals from bond conditioning carry information from S^{\rightarrow} to S^{\leftarrow} . Such modulated element signals must be assessed to quantify whether or not the subgraph loop can be broken and partitions created as in Fig. 2.12c.

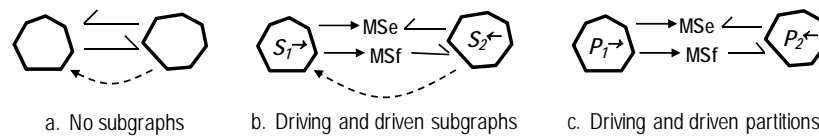


Fig. 2.12. Subgraphs, partitions and loops

The contribution of a modulating signal can be evaluated if the modulated energetic element can be divided into modulated and non-modulated elements, the total effect of which is the same as the original element. The signal can be eliminated if the modulated elements have low Relative Activity, in which case the primary contribution of the element was from the non-modulated portion. Three possible approaches to element division are [25]:

1. *Series expansion of element constitutive laws*: For a system that operates close to an equilibrium point, the user may attempt to expand the constitutive law containing the modulating signal as an infinite series such as a Taylor Series. A separate element will result for each term of the series. The first (equilibrium) term of the expansion will be non-modulated. If local comparison of the activities of the individual elements suggests elimination of all but the first, then modulation is not necessary. See Fig. 2.15.
2. *Analytical separation of constitutive laws*: In Fig. 2.16, a C and modulated R element are bonded to a 1-junction. The modulated resistor constitutive law consists of clearly separable modulated and non-modulated components. The element can be broken into separate elements with the same causality as the original, affixed to a common junction – in this case the original 1-junction whose flow serves as the input to both. If as in Fig. 2.16 the element is attached to a 0-junction, the separate R_o and $\phi(x)$ elements still require a common 1-junction, at which the two resistance components can be added.

3. *Separation of the signal into nominal and varying components*: If the constitutive law is not separable, but a nominal value of the modulating signal can be calculated, then the element can be duplicated and the signal split into constant and varying components. The first copy of the element is modulated by the constant nominal value. This element is essentially non-modulated, as the constant signal can be incorporated into the constitutive law as a parameter. The second copy is attached to a common junction (according to standard bond graph equivalence relations) as shown in Fig. 2.17 and is modulated by the difference between the total and nominal signal components. A local activity comparison then determines whether the modulated copy of the element can be neglected.

The final partitioning algorithm is depicted in Fig. 2.18. Finding and separating partitions allows immediate model reduction, even if the partitions are not subjected to techniques to eliminate non-contributing states. As shown in Fig. 2.19, an output associated with a driving partition element can be predicted by simulating the driving partition by itself. The energetic elements and junction structure of the driven partition can be eliminated. To predict a driven partition output, the driving model structure can be eliminated after the necessary modulating signals are generated and stored in an input file.

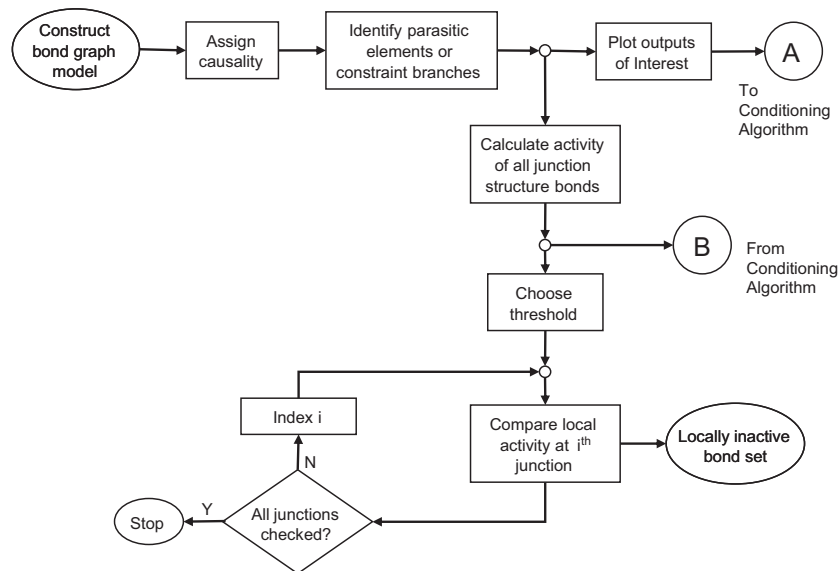


Fig. 2.13. Inactive bond search algorithm

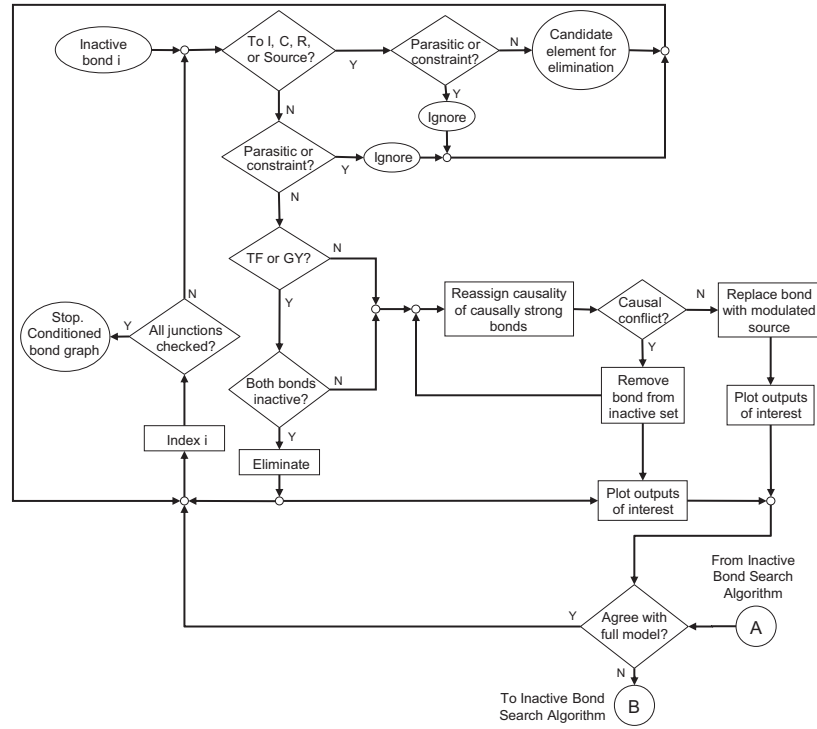


Fig. 2.14. Conditioning algorithm

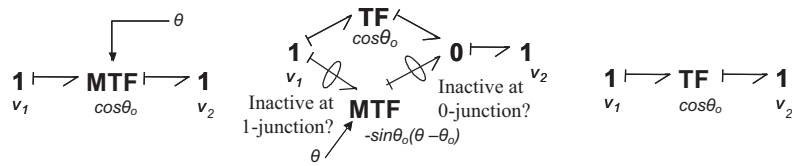


Fig. 2.15. Taylor expansion of state-modulated transformer constitutive law

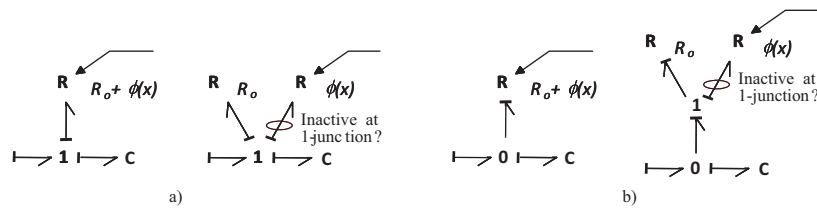


Fig. 2.16. Modulated element with separable constitutive law

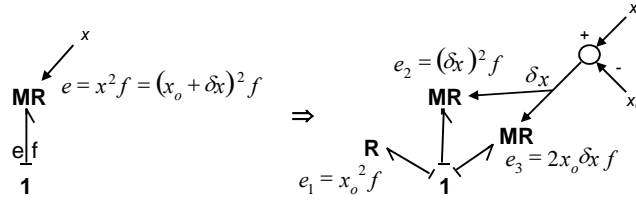


Fig. 2.17. Constitutive law and modulating signal decomposition

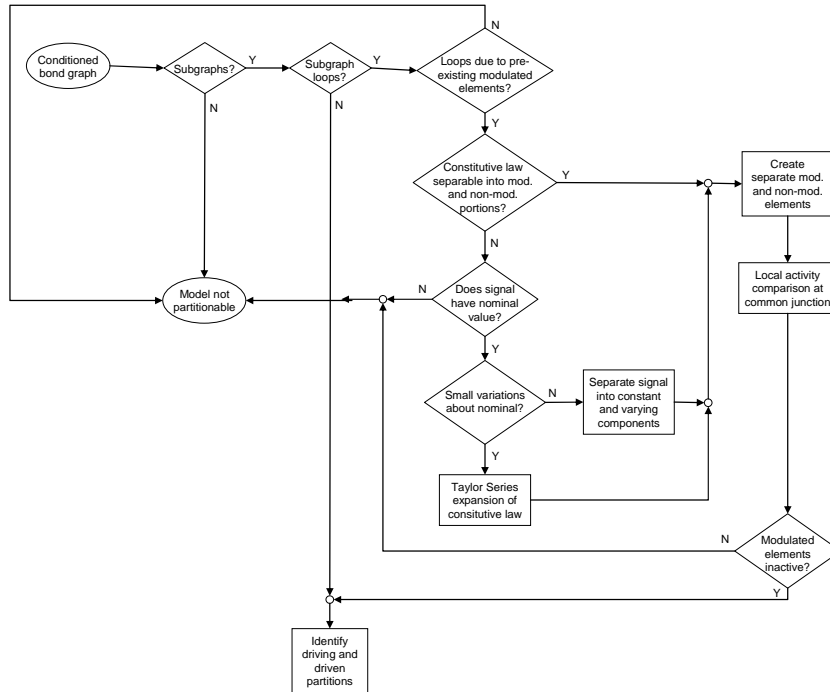


Fig. 2.18. Model partitioning algorithm

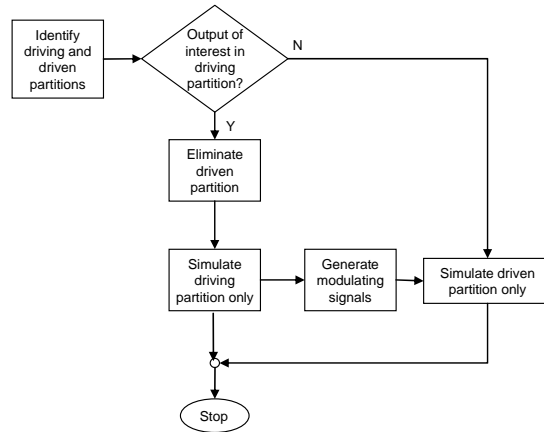


Fig. 2.19. Implications for model reduction

2.3.2.1. Illustrative Example: Two-Mass System

In Fig. 2.20, two masses and three parallel spring-damper arrangements are connected through a lever. Inertia of the lever is neglected. The lever is assumed to be long enough, so that the endpoints approximately translate in the v_1 and v_2 directions. A step input force F_1 is applied to the mass m_1 . The goal is to predict displacements x_1 and x_2 .

For this system, increasing the ratio a/b will increasingly attenuate the force F_b transmitted through the lever to k_2-c_2 . Thus, depending on the energetic element parameters, the response of m_1 may not be affected significantly by the dynamics of m_2 . The velocity v_b , which would be significantly lower than v_a for large a/b , would still be necessary to define the spring velocity v_{k3} and set m_2 into motion.

Note that neglecting the lever inertia creates an algebraic loop. With state variables defined as the mass velocities and spring displacements, the differential-algebraic equations of motion are given below.

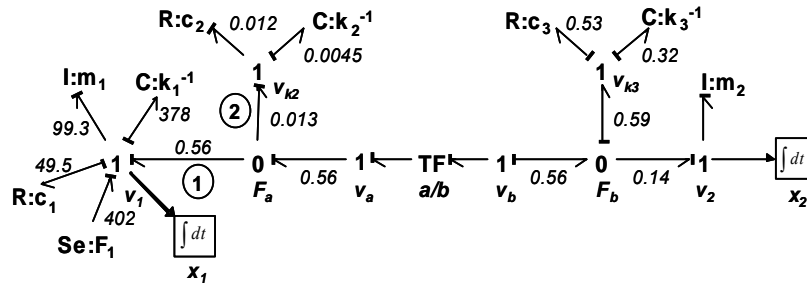


Fig. 2.21. Example system bond graph relative activities

Conditioning bonds 1 and 2 in Fig. 2.21 results in the conditioned model and partitions of Fig. 2.22, and the output time series shown in Fig.2.23. The trivial driven subgraph containing k_2-c_2 is discarded in Fig. 2.22, as it does not affect the outputs of interest. For the partitioned system, the equations of motion below illustrate the partitioned state variable vector and the presence of the input $v_1(t)$ to the driven partition.

$$\begin{aligned}
 \text{Driving: } \begin{cases} \dot{v}_1 \\ \dot{x}_{k2} \end{cases} &= \begin{bmatrix} -c_1 & -k_1 \\ 1 & 0 \end{bmatrix} \begin{cases} v_1 \\ x_{k2} \end{cases} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \{F_1(t)\} \\
 \text{Driven: } \begin{cases} \dot{v}_2 \\ \dot{x}_{k3} \end{cases} &= \begin{bmatrix} -c_3 & -k_3 \\ m_2 & 0 \end{bmatrix} \begin{cases} v_2 \\ x_{k3} \end{cases} + \begin{bmatrix} \frac{b}{a} \frac{c_3}{m_2} \\ \frac{b}{a} \end{bmatrix} \{v_1(t)\}
 \end{aligned} \tag{2.21}$$

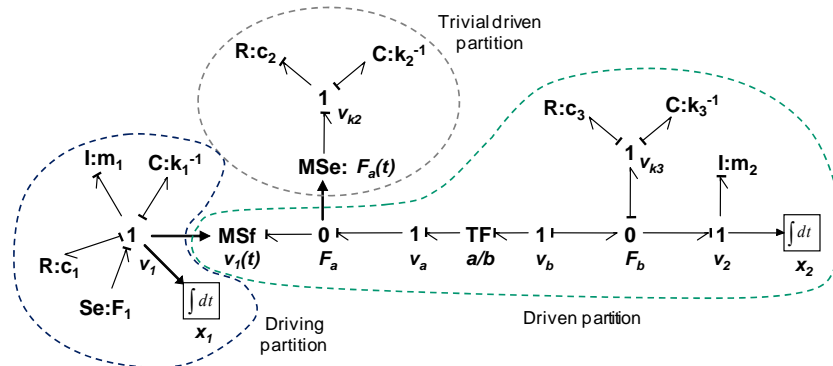


Fig. 2.22. Conditioned bond graph showing partitions

Eliminating c_2 breaks the algebraic loop, thus facilitating equation formulation and computation. Simply conditioning the model removes the output from c_2 to the rest of the system, eliminating the algebraic loop and reducing computation

time by 35% even though no states are eliminated. Simulation of the Driving or Driven partition, or both in parallel or in sequence, requires even less time. The computation steps and time reductions are more significant for larger models with more conditioned bonds and more balanced partitions. In [26], for example, a 279-state three-dimensional engine model was partitioned, with sequential driving-driven partition simulation reducing the computation time by 54%. Fig. 2.23 shows accurate predictions of the mass displacements using the partitioned model.

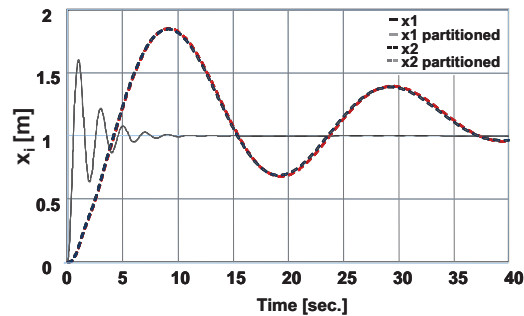


Fig. 2.23. Partitioned and original model output

A larger nonlinear vehicle pitch plane model is subjected to the inactive bond search, conditioning and partitioning algorithms in the *Case Studies* section. The case study demonstrates not only the benefits of conditioning and partitioning for model reduction, but also the breaking of subgraph loops due to pre-existing modulating signals.

In summary, the conditioning and partitioning algorithms allow the simulation-based design engineer to

- systematically assess coupling within an arbitrary lumped-parameter model of a nonlinear system, rather than using *a priori* decoupling assumptions
- monitor the coupling strength as the system parameters or environment change
- validate *a priori* decoupling or one-way coupling assumptions and subsequent model reductions
- identify “partitions” – collections of dynamic elements between which one-way coupling exists
- determine the significance of modulating signals from driven to driving dynamics, thus maximizing partitioning opportunities
- simulate “driving” and “driven” partitions separately or in parallel.

When decoupling can no longer be assumed, relative activity directs the analyst towards the specific locations of increased or decreased two-way power flow within a system. Retention of physical parameters maximizes insight into the coupling among dynamic system elements. As local bond activities exceed the threshold, the required increase in model complexity is automatically suggested – modulating signals and sources can simply be reverted back to power bonds.

Finally, it is worth noting that the presence of outputs of interest only in driving partitions, and therefore the ability to eliminate large numbers of energetic and junction structure elements, is not assured.

2.3.3. ECI-Based Model Reduction Algorithm

An algorithm for model reduction using the ECI can be outlined as follows:

1. Simulate the full model for a scenario of interest (i.e., a choice of inputs, initial conditions, parameters, and time window), and record the energies of the bonds of interest. Let n be the number of bonds of interest, and m the number of samples along the energy trajectories of each bond.
2. Arrange the data in a $m \times n$ matrix \mathbf{E} as defined in Eq. (2.9) such that the columns are the energy trajectories of the corresponding bonds.
3. Calculate $\mathbf{S}_{k,j} = \sqrt{\Delta t_k} \mathbf{E}_{k,j}$; $k = 1, \dots, m$; $j = 1, \dots, n$, where Δt_k is the time step between the $k-1^{\text{st}}$ and the k^{th} samples as determined by the numerical integration scheme.
4. Perform singular value decomposition on \mathbf{S} , i.e., $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)_{m \times n}$ with $\sigma_1 \geq \dots \geq \sigma_n \geq 0$, and the columns of \mathbf{V} are \mathbf{v}_i .
5. Calculate the relative ECI of the bonds using Eq. (2.12) and (2.13).
6. Arrange bonds in decreasing order of ECI. Let p be an index for the rows of this ordered list.
7. If $ECI_p / ECI_{p+1} > r$ for some row $p < n-1$ and user-defined ratio $r > 1$, then bonds in rows $p+1, \dots, n$ are subject to reduction. There may be more than one such threshold, i.e., more than one level of reduction. It is up to the modeler to decide on the ratio r and which threshold to use for reduction.
8. Remove the elements that are disconnected from the rest of the model as a result of step 7.

Note that if all bonds are subject to the analysis, this algorithm gives a unified approach to the reduction problem in the sense that not only the order, but also the structure of the model can be reduced. This will be hereafter referred to as a global application of the ECI. It is also possible to perform the analysis locally, e.g., only for the bonds connected to the components representing the states for the purposes of model order reduction, or only for the bonds connected to a junction element for the purposes of model partitioning.

2.3.3.1. Illustrative Example: Slider-Crank System

This section provides an example to illustrate the mechanics of the ECI-based model reduction algorithm and emphasize its advantages, namely, its applicability to nonlinear systems, ability to achieve graph-level reduction, and ability to reduce

the order and structure of the model, while taking into account the scenario of interest and preserving the realization of the model.

Consider the system shown in Fig. 2.24, where a mass-spring-damper system is connected to the slider of a crank mechanism. A rotational spring and damper are connected to the crank arm, and the rotational spring is undeflected when $\alpha = \pi/2$. There is viscous friction between the slider and ground. The parameter values are given in Table 2.7 and the bond graph model of the system is given in Fig. 2.25, which is the full model in this example. The bonds are numbered such that each bond with a unique energy receives a unique index. Bonds connected to power-through junctions therefore have the same index. The full model includes the dynamics of the links and masses, as well as the kinematics

$$\begin{aligned}
 \mathbf{v}_G &= \mathbf{v}_B + \boldsymbol{\omega}_{BC} \times \mathbf{r}_{G/B} = \boldsymbol{\omega}_{AB} \times \mathbf{r}_{B/A} + \boldsymbol{\omega}_{BC} \times \mathbf{r}_{G/B} \\
 &= \left(-L_1 \dot{\alpha} \sin \alpha - \frac{L_2}{2} \dot{\beta} \sin \beta \right) \mathbf{i} + \left(L_1 \dot{\alpha} \cos \alpha + \frac{L_2}{2} \dot{\beta} \cos \beta \right) \mathbf{j} \\
 &= v_{Gx} \mathbf{i} + v_{Gy} \mathbf{j} \\
 \mathbf{v}_C &= \mathbf{v}_B + \boldsymbol{\omega}_{BC} \times \mathbf{r}_{C/B} = \boldsymbol{\omega}_{AB} \times \mathbf{r}_{B/A} + \boldsymbol{\omega}_{BC} \times \mathbf{r}_{C/B} \\
 &= \left(-L_1 \dot{\alpha} \sin \alpha - L_2 \dot{\beta} \sin \beta \right) \mathbf{i} + \left(L_1 \dot{\alpha} \cos \alpha + L_2 \dot{\beta} \cos \beta \right) \mathbf{j} \\
 &= v_{Cx} \mathbf{i} + v_{Cy} \mathbf{j}
 \end{aligned} \tag{2.22}$$

along with the constraint

$$v_{Cy} = L_1 \dot{\alpha} \cos \alpha + L_2 \dot{\beta} \cos \beta = 0 \tag{2.23}$$

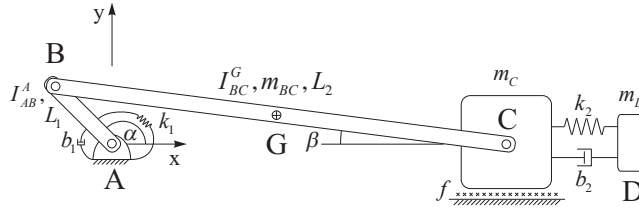


Fig. 2.24. Schematic representation of the example system

Table 2.7. The parameters of the example system in Fig. 2.24

Parameter	Value	Parameter	Value
I_{AB}^A	$3.53 \cdot 10^{-5} \text{ kg} \cdot \text{m}^2$	k_2	100 N/m
I_{BC}^G	$8.84 \cdot 10^{-3} \text{ kg} \cdot \text{m}^2$	f	1 N·s/m
m_{BC}	0.42 kg	b_1	0.01 N·m·s/rad
m_C	10 kg	b_2	0.1 N·s/m
m_D	0.1 kg	L_1	0.05 m
k_1	1 N·m/rad	L_2	0.5 m

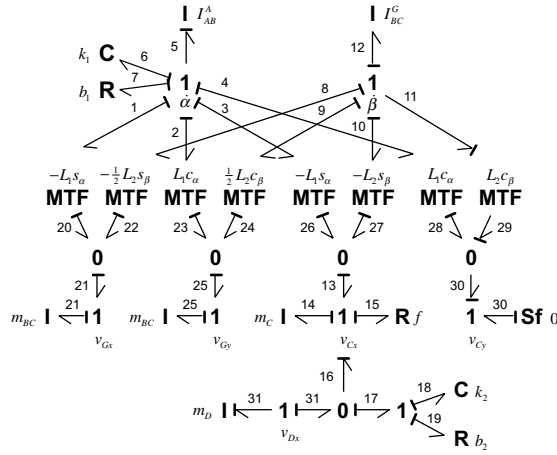


Fig. 2.25. Bond graph of the example system in Fig. 2.24

Scenario 1

Consider the scenario, in which the springs k_1 and k_2 are given initial displacements of 1 rad and 0.01 m, respectively, where a positive sign indicates extension, and the free response of the system is observed. Let the output of interest be the position of the mass m_D . When the ECI analysis is applied to this scenario, the results summarized in Table 2.8 are obtained for a global analysis with $r = 2$ and a simulation time-window of 5 seconds.

The dashed lines in Table 2.8 indicate the 5 thresholds for $r = 2$, and hence 5 different levels of reduction, which can be explained physically as follows:

Level 1: This threshold points to a well-known structural simplification that can be made in the bond graph; namely, the null flow-source can be removed along with the 1-junction v_{C_y} without affecting the accuracy of the model.

Level 2: The moment of inertia of the second link is removed. Even though it is larger than the moment of inertia of the first link, the second link rotates less due to the kinematics, and therefore the energy associated with its rotational dynamics is very low.

Level 3: The rotational and translational dynamics of the first link, and the translational dynamics and kinematics of the second link in y-direction are removed. Furthermore, the translational kinematics of the point G in x-direction is reduced by neglecting the terms involving β and its derivatives, i.e. the expression for \mathbf{v}_G in Eq. (2.22) reduces to

$$\mathbf{v}_G = -L_1 \dot{\alpha} \sin \alpha \mathbf{i} \quad (2.24)$$

Table 2.8. ECI of Bonds for Scenario 1 and Reduction Thresholds for $r = 2$

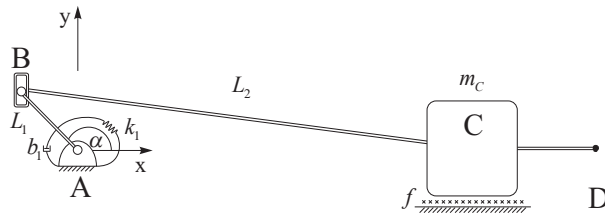
Bond	ECI	Bond (cont'd)	ECI (cont'd)
7	100%	1	1.86%
26	64.93%	21	1.69%
3	64.93%	17	1.37%
13	56.60%	31	0.91%
14	39.84%	18	0.48%
6	27.11%	22	0.21%
15	21.25%	8	0.21%
11	10.25%	23	0.13%
29	10.25%	2	0.13%
28	10.25%	5	0.08%
4	10.25%	25	0.07%
27	10.02%	24	0.07%
10	10.02%	9	0.07%
19	2.79%	12	0.02%
16	1.90%	30	0
20	1.86%		

Level 4: The dynamics and kinematics of the mass-spring-damper system connected to the slider are removed, as well as the translational dynamics and kinematics of the second link in x-direction.

Level 5: The kinematics associated with β are removed. As a result, the expression for \mathbf{v}_C in Eq. (2.22) reduces to

$$\mathbf{v}_C = -L_1 \dot{\alpha} \sin \alpha \mathbf{i} \quad (2.25)$$

and the constraint Eq. (2.23) is not needed. Fig. 2.26 shows the schematic representation of this reduced system, and Fig. 2.27 shows the corresponding bond graph. Fig. 2.28 compares the output of this reduced model to the output of the full model.

Fig. 2.26. Schematic representation of the 5th-level reduced model for Scenario 1

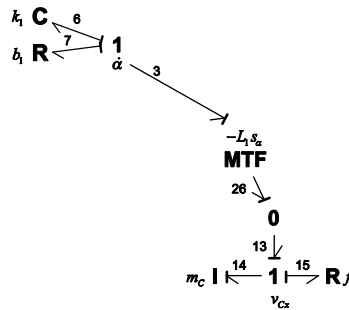


Fig. 2.27. Bond graph of the 5th-level reduced model for Scenario 1

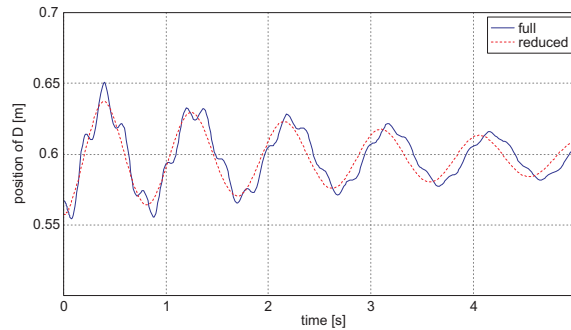


Fig. 2.28. Output of the full model vs. output of the 5th-level reduced model for Scenario 1

So far this example illustrated the mechanics of the ECI-based model reduction algorithm and highlighted the following benefits of it: applicability to nonlinear systems, ability to achieve graph-level reduction, preservation of the original realization of the model, and ability to reduce the structure of the model, i.e. reducing not only the dynamics, but also the kinematics. The next part highlights the method's ability to furnish different reduced models for different scenarios.

Scenario 2

Consider now the scenario, in which an initial momentum of $0.05 \text{ kg} \cdot \text{m/s}$ is given to the mass m_D , and the free response of the system is observed. As in Scenario 1, let the output of interest be the position of the mass m_D . When the ECI analysis is applied to this scenario, the results summarized in Table 2.9 are obtained for a global analysis with $r = 2$ and a simulation time-window of 3 seconds.

Table 2.9 indicates 4 thresholds and hence 4 different levels of reduction, however, since the level 4 corresponds to practically discarding the model completely, it will be ignored. The remaining reduction levels can be explained physically as follows.

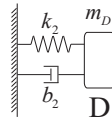
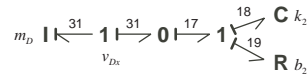
Table 2.9. ECI of Bonds for Scenario 2 and Reduction Thresholds for $r = 2$

Bond	ECI	Bond (cont'd)	ECI (cont'd)
17	100%	21	0.02% 1
19	85.70% 4	4	0
18	21.93%	29	0
31	19.22% 3	28	0
16	1.81%	11	0
13	1.14%	10	0
3	1.14%	27	0
26	1.14%	22	0
7	0.83%	8	0
14	0.48%	2	0
6	0.35%	23	0
15	0.21% 2	25	0
5	0.03%	24	0
20	0.02%	9	0
1	0.02%	12	0
		30	0

Level 1: All kinematics involving β , the translational kinematics and dynamics of the second link in y-direction as well as its rotational dynamics, and the constraint Eq. (2.23) are removed. Without any kinematics involving β , the expression for \mathbf{v}_G in Eq. (2.22) reduces to Eq. (2.24).

Level 2: The rotational and translational dynamics of the first link, and the kinematics and dynamics of the second link in x-direction are removed.

Level 3: The kinematics involving α , the rotational stiffness k_1 and damping b_1 , the kinematics and dynamics of mass m_D along with the friction between m_D and the surface are removed. Schematically, the system reduces down to Fig. 2.29, and the bond graph reduces to Fig. 2.30. Fig. 2.31 compares the output of this reduced model to the output of the full model.

Fig. 2.29. Schematic representation of the 3rd-level reduced model for Scenario 2Fig. 2.30. Bond graph of the 3rd-level reduced model for Scenario 2

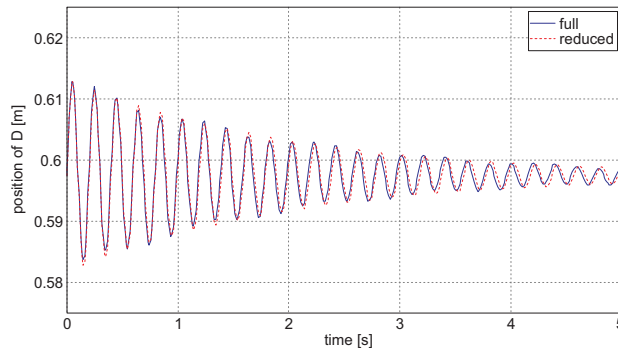


Fig. 2.31. Output of the full model vs. output of the 3rd-level reduced model for Scenario 2

Notice the difference between the reduced models for the two scenarios (Fig. 2.27 and 2.30), and how the ECI tailors the reduction according to the scenario of interest.

2.4. Case Studies

The three algorithms described in this chapter were demonstrated through relatively simple illustrative examples. However, their applicability is not limited by the size or complexity of the system. In fact, the motivation for developing these modeling metrics and algorithms was to overcome the bottlenecks of the modeling procedure of real life engineering systems. The following case studies represent this class of systems and reveal the strengths of the algorithms when it comes to real systems. However, due to the lack of space, they are described only briefly and references to more detailed analyses are given.

2.4.1. *Reduction of a Heavy Tractor Semi-Trailer and a Hybrid Hydraulic Truck using MORA*

Vehicle handling and stability, which are critical for heavy trucks, are typically predicted by means of modeling and simulation. Such models are based on assumptions of the critical components that need to be included for obtaining accurate predictions. Depending on the modeler's knowledge and understanding of the system, these assumptions may lead to an oversimplified or overcomplicated model. MORA can be used in this case to systematically verify the modeling assumptions and evaluate the complexity of such vehicle models.

The first system to be considered here is the M916A1/870A2 military heavy-duty tractor semi-trailer. The full model is generated by synthesizing 20 rigid bo-

dies with 33 rigid body DOF and 122 forces/moments. The full model has 91 states and approximately 120 parameters. The equations of motion are formulated using Kane's method and the resulting model is highly nonlinear due to the nonlinear constraint forces and the three-dimensional rigid body kinematics.

A specific maneuver is selected to calculate activity and reduce the model. It is assumed that the vehicle is traveling with a constant speed of 60 mph and at time $t = 1\text{ s}$ the driver performs a lane change maneuver to avoid an obstacle. This maneuver is assumed to be executed by turning the steering wheel first left to avoid the obstacle, and then right to resume the original heading in the adjacent left lane. The full model is used to calculate the system response as it is performing this lane change maneuver. This is a relatively severe maneuver producing a maximum of about 0.2 g of lateral acceleration.

The simulation also produces the required outputs needed for calculating the power, the activity, and finally the activity index of each energy element in the model. There are 169 elements (121 forces and 8 rigid bodies with 6 directions each), for which the activities are calculated to determine their relative importance for this maneuver. The inertial forces of rigid bodies are projected onto each degree of freedom to produce six activities for each body. The sorted activity indices along with the cumulative activity indices for this steering maneuver are plotted in Fig. 2.32. Notice in that the cumulative activity is at approximately 80% after including the first 13 most important (active) elements. The most important elements are the trailer and tractor translational inertia in the longitudinal and lateral direction and the force-generating elements in the tire model in the longitudinal direction.

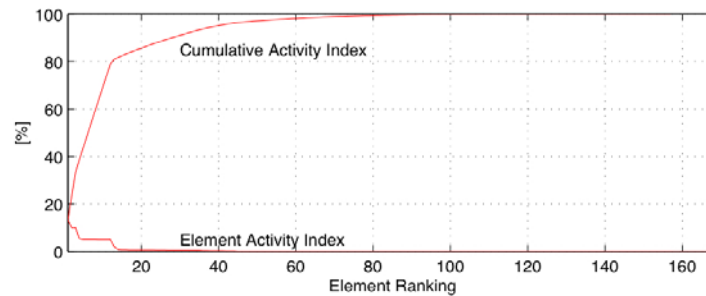


Fig. 2.32. Sorted activity indices and cumulative index

A series of reduced models are produced based on the activity index and MORA. More specifically 19%, 23% and 37% of the elements are eliminated to generate three reduced models. The accuracy of the reduced models degrades as more elements are eliminated, however, even after removing 37% of the elements the reduced model retains only the most important elements to predict the system response with sufficient accuracy. The model with 37% of the elements eliminated has a maximum error of 8% in predicting the lateral acceleration and yaw rate. It

is also shown that the frequency content of the model reduces as the model size is reduced according to the activity metric, i.e., low activity elements are associated with high frequencies. More details on this real life modeling case study can be found in [27].

Another real life case study is the development and reduction of an integrated hybrid vehicle model composed of an engine, drivetrain, hydraulics, and vehicle dynamics. The model is configured for a medium size truck using the bond graph formulation and implemented in 20Sim. After developing the model, MORA is applied to generate a reduced vehicle model that provides more design insight, while having improved computationally efficiency. Compared to the full model, the reduced model for the hybrid truck, as generated by MORA, produces almost identical results, has half the size, and computes the system response 2.5 times faster. More specifically, these benefits come with only a 0.11% loss in accuracy in the predictions of fuel economy over a complete driving cycle on an uneven terrain. More details on this case study are given in [28].

2.4.2. Partitioning of a Nonlinear Pitch Plane Truck Model

A pitch plane model of a Class VI delivery truck based on an International 4700-series vehicle was constructed to predict forward speed V and pitch angle θ . The vehicle accelerates at full-throttle from a standstill, on a road that is flat and smooth for the first 1200 feet of travel, after which a 1:10 slope is encountered. The maneuver lasts 120 seconds. The model is conditioned with a 4% threshold to determine if the longitudinal dynamics form a partition that drives the pitch response, thus allowing prediction of V and θ with smaller individual submodels. Details of the model formulation and parameters can be found in [25]. A schematic and conditioned bond graph are shown in Fig. 2.33 and 2.34, respectively. The model permits large angular motions of the sprung mass and uses nonlinear constitutive laws for aerodynamic drag and tire slip and rolling resistance. In Fig. 2.34, the heavy dashed lines indicate new modulating signals to sources arising from bond conditioning. The dotted lines are pre-existing modulating signals from the driven subgraph to the driving, thus precluding partitioning unless the subgraph loop can be broken.

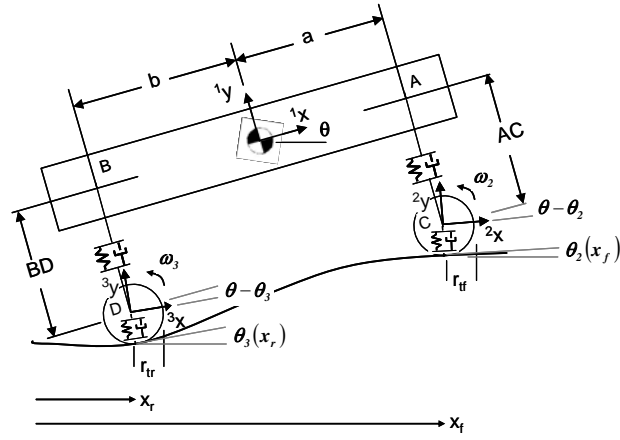


Fig. 2.33. Pitch plane truck schematic

The conditioned bonds are physically interpreted in Table 2.10.

Table 2.10 Truck model conditioned bonds

Bond	Description	RA
1	longitudinal force on sprung mass from front tires/susp.	0.75
2	gyrational longitudinal force on sprung mass	0.75
3	long. vel. component of rear hub due to rotation	~0
4	moment about c.g. from long. rear tire/susp. force	2.8
5	long. vel. component of front hub due to rotation	~0
6	moment about c.g. from long. front tire/susp. force	0.58
7	body-fixed y rear susp. force component tangent to road	1.5
8	rear tire vel. normal to road , body-fixed long. component	~0
9	body-fixed long. rear susp. force component normal to road	0.23
10	front tire vel. normal to road , body-fixed long. component	~0
11	body-fixed long. front susp. force normal to road	~0

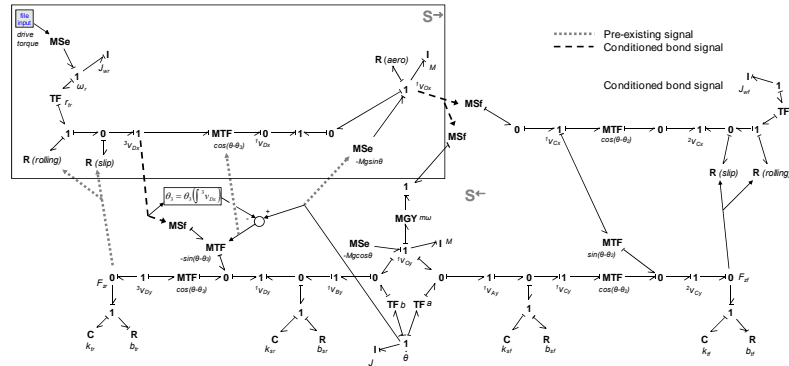


Fig. 2.34. Pitch plane truck model conditioned bond graph

Partitioning is contingent upon breaking the bond subgraph loop created by the following two modulating signals: the rear tire normal force modulating signal (output of the F_{zr} 0-junction), and the pitch angle signal (output of the θ 1-junction). The driving subgraph rear tire rolling and slip resistances are functions of the normal force F_{zr} from the driven subgraph. A nominal signal value of -39950N was identified as the static tire load, and the difference between the total and nominal values gave the varying component. The pitch angle nominal value was set at the road inclination (adjusted by the truck's static pitch angle), and departure of the pitch angle from the road angle served as the varying component of the signal. The modulated elements had low RA and could be eliminated, breaking the subgraph loop.

To estimate the outputs of interest (forward speed and pitch angle), only the driving partition is required, leading to significant model reduction. As reported in [25], conditioning the model breaks four algebraic loops, reducing the number of computation steps by 38% and time by 43%. Simulating the driving partition only gives accurate predictions, with reductions in computation steps and time of 61% and 52%, respectively.

Subjecting the truck model to a rough road increases the relative activity of bonds 1, 2, and 7 in Fig. 2.34 above the threshold. Partitioning is no longer recommended, and significant discrepancies arise between the predictions of a fully-coupled and partitioned model.

The reader is also referred to [26] for details of an even larger case study involving a 14 body, 279-state model of an inline six-cylinder diesel engine. The partitioning algorithm was used to decouple engine dynamics into a driving partition comprised of reciprocating elements, and a driven partition containing the block moving in three dimensions on its mounts. This partitioning is consistent with the assumption of decoupling between reciprocating dynamics and engine block motion reported in prior literature [29] for balanced engines running at low speed. Introduction of a misfire into the engine increased vibration, and the algo-

rithms predicted that partitioning the model would bring unacceptable prediction errors. Simulation times were reduced by 20% and 53% for the conditioned and partitioned models respectively, in predicting the responses of both partitions.

2.4.3. ECI-Based Reduction of a HMMWV Model

The ECI-based model reduction algorithm has been used for the reduction of a multibody model of a High Mobility Multipurpose Wheeled Vehicle (HMMWV) in [8]. The three-dimensional 22-body model of the HMMWV comprised the chassis; the four independent double-wishbone suspensions comprising lower and upper A-arms, wheel hubs, suspension springs and dampers; the front and rear anti-roll bars; the four tires with vertical tire stiffness and damping, and longitudinal and lateral slip models; and the steering mechanism consisting of the steering link, idler arm, Pitman arm, and tie rods. The anti-roll bars were not modeled as 3D bodies, but their effect was taken into account through spring elements applying restoring forces on the velocity differences between the left and right suspension struts.

There are many families of maneuvers that could be employed in vehicle dynamics studies, and different families of maneuvers would require different reduced models. To illustrate this, this case study considered three scenarios representative of three different families of maneuvers.

The first scenario was a two-double-lane-change maneuver on a flat road. Such a maneuver could be employed in, e.g., vehicle rollover and handling studies. The vehicle was accelerated from rest to a constant velocity, the lane-change maneuvers were performed, and the vehicle was brought to a stop. The output of interest was the roll acceleration of the vehicle.

The second scenario was a shaker table scenario. This scenario might be of interest when studying, e.g., the suspension characteristics and ride-quality of a vehicle. In this scenario the tires were removed from the model, and a sinusoidal sweeping displacement was applied to all four wheel hubs. The output of interest was the vertical position of the chassis.

The third scenario was driving straight on a flat road. This scenario could be useful, e.g., when studying the acceleration characteristics of the vehicle, sizing the engine, or designing a cruise controller. In this particular scenario the input dictated the wheel speed. The output of interest was the longitudinal acceleration of the vehicle.

Application of the ECI revealed what can be physically removed from the model in each scenario. For example, in Scenario 1, ECI analysis showed that most of the translational and rotational dynamics of the smaller suspension and steering linkages (e.g., A-arms, Pitman arm, etc.) had little influence on the system behavior. In Scenario 2, the analysis showed that the dynamics of the steering mechanism were not critical to model for analyzing the vehicle's vertical vibrations, but part of the kinematics of the mechanism may be important. In Scenario 3, the

model essentially reduced down to a disk rolling without slip, or, equivalently, to a point mass. Such physical interpretations provide a deeper insight into the vehicle dynamics, and the ECI analysis provided these insights automatically.

Table 2.11 highlights the computational benefits obtained with ECI-based model reduction in each scenario, and Figures 2.35-2.37 compare the outputs of the full and reduced models to show the fidelity compromised for the purposes of reduction.

Table 2.11. 20-sim processing and simulation results for the HMMWV case study

	% Decrease		
	Scenario 1	Scenario 2	Scenario 3
Equations	38.9	71.2	98.1
Independent states	22.1	81.2	94.7
Dependent states	38.9	85.2	100
Constraints	32.0	16.7	100
Simulation time (s)	88.2	94.7	99.8

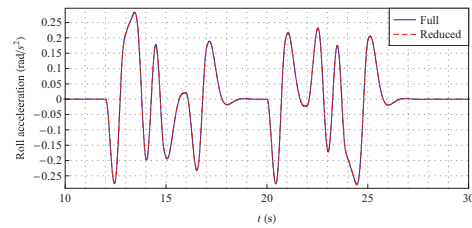


Fig. 2.35. Output comparison for Scenario 1

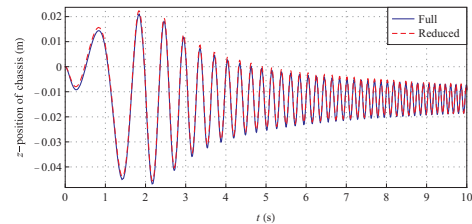


Fig. 2.36. Output comparison for Scenario 2

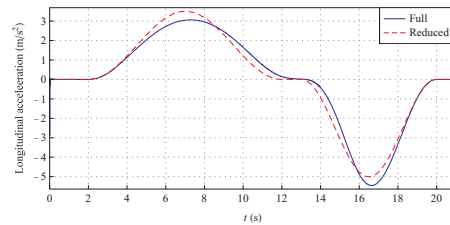


Fig. 2.37. Output comparison for Scenario 3

This case study shows that ECI analysis can be successfully applied to complicated systems, as well, to seek for a balance between model simplicity and fidelity for the specific scenarios of interest. The details of this case study can be found in [8, 30].

2.5. Discussion

This chapter has defined a suite of metrics and algorithms for energy-based model reduction. The use of power and energy in the metrics recommends the bond graph formalism, in which power flow paths among elements are explicitly indicated, and the generalized effort and flow variables are at hand for all elements. Indeed, bond graph based software programs have started implementing some of the ideas presented in this chapter. For example, 20sim has appended its element library with 1- and 0-junctions that can calculate relative activity of all connected bonds. AMESim [31] has implemented activity analysis and can automatically provide the analysis results to the user at the end of a simulation, based on which the user can then reduce the model manually according to MORA. Calculation of these metrics can also be accomplished in non-bond graph based packages such as ADAMS or MATLAB, as well, but bond graph based packages certainly facilitate the implementation of these techniques.

The three techniques introduced in this chapter share some common properties that give them certain advantages over some of the existing techniques. The first advantage is that they can be applied to the reduction of nonlinear models while preserving the model's realization. This is a distinct advantage over, for example, frequency-based methods that are typically applicable to linear models only, or projection based methods that are by their very nature non-realization preserving. As the case studies suggest, the techniques can be applied even in the presence of the nonlinear elements such as suspension springs and rolling resistance, or even model discontinuities such as wheel lift off, and the reduced models still preserve their original physical interpretations.

A second advantage is that all energy elements in the model can be treated equally. This is, for example, unlike the Model Order Deduction Algorithm - MODA [32] that adds only compliance elements to the model to increase its complexity and requires the user to decide by other non specified means whether to include other energy storage or dissipations elements. The analyst may, however, choose not to treat all elements equally when applying the partitioning and ECI algorithms. These algorithms can be applied either globally to all model elements, thereby serving as a means of simultaneous model reduction and partitioning, or locally to subsets of bond graph elements or junction structure only, thus achieving a selective reduction or partitioning.

A third advantage of the techniques is the insight they provide to the user. Even if the metrics are calculated and the algorithms are applied without actually reducing the model, the analyst can gain considerable insight into the interactions

among the elements and the criticality of accurate estimation of their parameters. MORA, for instance, shows the relative ranking of all elements in the model. This parameter sensitivity is different from typical sensitivity measures in that the parameter ranking is with respect to the entire system dynamics and not to some outputs [33, 34]. For example, the ride quality example shows, not surprisingly, that for the harsh input the suspension deflects nearly twice as much as with the smooth input. Thus, from a conventional sensitivity point of view, the suspension spring displacement is sensitive to the input magnitude. However, activity shows that the relative importance of the suspension spring to the system total activity slightly reduces as the vehicle speed increases. In other words, the relative value of the spring is less important to the overall system dynamics as the vehicle speed is increased. The elements that are important relative to the input are the unsprung mass, whose activity index is about 12 times more important to the model when the input becomes harsher, and the suspension damper, which is nearly twice as important. Note the tire damping also becomes more important by a factor close to 12, but still remains the least important element in the model. This, of course, is likely to change for even more severe inputs, i.e., higher forward vehicle speed. These parameter ranks as a function of the input provides the designer with the insight to see where the most benefit may be generated from modifying the design. Some of this type of insight could be obtained with other techniques; however, the approach of this chapter allows this insight to be gained as part of the modeling process, and not as part of a downstream analysis process, which is less likely to have an effect on modeling decisions. Furthermore, other techniques such as design of experiments [35, 36] require multiple model evaluations that can be costly and time consuming to calculate system sensitivities especially for large size models. With the activity metric, however, the system's sensitivity to its parameters can be obtained with only a single or a few simulations.

Indeed, one of the most costly and perhaps difficult tasks associated with modeling is obtaining numerical values for the parameters. As a first observation, reducing the parameter estimation burden may not appear to be an advantage of the aforementioned techniques, or any other reduction technique in general, because a full model is required to assess what may or may not be important. However, often during the modeling process, parameter values are approximated from previous test results, borrowed from a similar model of a similar product, or simply estimated. The hierarchy created by MORA or ECI, or the relegation of certain elements to driven partitions, can be viewed as a priority task list. The most important elements are the elements whose parameters must be known most accurately. If, as in the MORA illustrative example, tire damping is not important, then spending a lot of time and effort to determine its characteristics is not justified. On the other hand, more resources should be allocated for getting better estimates of the suspension parameters since they are the most important elements according to activity.

It is also reasonable to expect significant savings in numerical simulation time from a reduced model. Although the saving may not be compelling in the simple illustrative examples used above, the simulation time of much larger systems may

be dramatically reduced if a large number of low activity, low-ECI, or driven partition elements are removed from the full model, leading to a significant reduction in the size of the state vector [8, 26-28, 37]. Even modest improvements in simulation time for an individual run can give significant aggregate time savings for applications that require many iterations such as optimization or Monte Carlo simulation. However, it is important to note that the numerical efficiency of a reduced model is not assured, as it may not be always easy or even possible to obtain the reduced model equations in explicit form. As elements are trimmed from the model, dependencies can be created between energy storage elements as well as between storage elements and inputs. This, of course, can be easily seen if the model is in bond graph form. Thus, while the full model might have been a type 1 causal system for which explicit equations can be easily found, this might not be true for the reduced model. This can affect not only the effort required to formulate the state equations but also the efficiency of numerically integrating them, as well. The user can use causality to help direct the model reduction process to try to avoid these potential problems. These problems could exist in the full model, as well, so reducing the model does not necessarily make the model harder to formulate and solve numerically. The simple illustrative example of the conditioning algorithm demonstrates that bond conditioning can break algebraic loops and improve numerical efficiency even if no partitions are identified or eliminated. The reader is referred to [26] for an example of significant savings due to conditioning alone in a large engine dynamics model. The algorithms of Fig. 2.13 and 2.14 specify that “parasitic” or “constraint” branches of a bond graph ought not to be subject to conditioning, and that causality assignment be attempted to ensure that conditioned bonds are causally weak. “Parasitic” elements refer to small masses or stiff springs that are added to, for example, multibody mechanical systems [38] normally comprised of rigidly constrained bodies. A compliance at a pin joint can, if sufficiently stiff and damped, break dependencies between the momenta of the connected bodies without compromising the kinematic constraints at a macroscopic level. Such elements are tuned to have low relative activity [39], yet are purposefully retained to facilitate formation of explicit ordinary differential equations.

Reduction of initially complex models allows the designer the option of deciding explicitly and quantitatively the tradeoff between model accuracy and simplicity. The user can easily generate an array of models to satisfy different needs by specifying the desired level of accuracies (cumulative activity, ECI threshold, or low RA threshold) needed for the different tasks. Noting the MORA and ECI illustrative examples, model accuracy degrades as more elements are eliminated from the model. The change in error between two consecutive MORA-reduced models was proportional to the activity of the eliminated element. This monotonic relation between the error and the element activity provides evidence to support the assumption that activity and the other energy-based metrics can be used as measures of elements’ contribution to the accuracy of the model predictions. While there does not currently exist proof that this relationship will always be true, results suggest that such a relationship may exist. The correlation of activity, relative activity, or ECI to a measure of the state trajectory deviation of the full model versus

a reduced model is unknown. Fathy and Stein [40] have shown that for linear system models with certain initial conditions and inputs, the reduced model attained through balanced truncation will be identical to a bond graph model in which the lowest activity elements are eliminated. While general equivalence between balanced truncation models and models reduced using activity has not yet been proven, the theoretical justification of activity as a reduction metric for linear systems expressed in a particular form increases confidence in the activity approach. The work of Fathy and Stein is relevant to partitioning inasmuch as the partitioning algorithm calculates activity for elements of a set of constraint terms, and elimination of negligible-activity terms is held to have no significant effect on the dynamic response of the remaining elements of the set. Many further extensions in this area are required before the three algorithms can be subject to formal proof. Some contributions have also been made toward quantifying model accuracy based on the states of the full model versus the reduced (e.g., Sendur et. al., [41, 42]), but this also remains an area for further research.

One of the advantages of the model reduction techniques presented in this chapter is also one of their limitations, namely, scenario dependence. While this can be viewed as useful in uncovering which parameters are important under which conditions, it also means that, in general, one reduced model cannot be found that is suitable for all conditions and, for that matter, all design variations. Rerunning of the model to check one's assumption under different conditions is needed. Quantification of the "range of validity" of a reduced model remains an open research question – in other words, how far from the inputs, initial conditions, and parameters can one deviate during the model-driven design process without having to recalculate the metrics and reapply the algorithms?

Also common to the three methods, and again a potential advantage and disadvantage, is the setting of a time window over which to calculate activity, relative activity or energy trajectories. Design insight can be gained from the metrics by utilizing the freedom provided in choosing this window. As can be seen in Fig. 2.7 for the activity metric, for example, the metric values for elements can change over time with respect to one another. For example, in Fig. 2.7, the unsprung mass at the time just after the input is turned on (t just after 1 s) has the second higher activity. At about 1.2 s, it becomes smaller than the activity the sprung mass, then, at around 2 s, it becomes smaller than the activity the suspension damping, as well. Obviously, the relative importance of these elements to the system model can be manipulated by the choice of the integration time window. This can be of value to the designer who chooses the windowing parameters wisely. For example, placing the window tightly around the main transient event versus including all of the steady state information could be used to differentiate between those elements important to maximum values (loads, displacements, etc.) versus those responsible for efficiency. The formal exploitation of this issue is an ongoing research topic. Prior research by Kypuros and Longoria [43] applied different road input frequencies to a pitch-plane vehicle model at different times of a simulation maneuver, and generated four MORA-based reduced models. The four models resulted from changing the activity integration time window to correspond with the

intervals of a given input frequency. They then calculated a moving average of activity and plotted the results to show the time dependency of proper model complexity. The moving average was proposed as a means of generating variable complexity models, if the model complexity changes could be automated. Recent research by Rideout and Haq [44] has used moving average of absolute power (a metric very closely related to activity) to detect when changes to model complexity are required, and to automate the switching off and on of elements as their energetic contribution falls below or rises above a threshold. The method of [44] brings with it a computational penalty, but automatically predicts the time window over which a given degree of model reduction is appropriate, and can predict the accuracy of a collection of such proper models of the same system if they were run sequentially.

It is also important to note that although the techniques presented in this chapter take explicitly into account the particular inputs, initial conditions, parameter values and time windows of interest, they do not take the outputs of interest into account. Therefore, a model arrived at through these techniques is not necessarily the lowest-complexity model that sufficiently predicts a specific output variable. Making the techniques output-dependent is an important area for future research.

Finally, despite this chapter's emphasis on energy-based model *reduction*, the proposed metrics can also be used for model *simplification*, which refers to finding a more succinct realization without compromising accuracy [45]. Some well-established rules for bond graph-level simplification include eliminating loose, power-through, or constraining junctions, merging adjacent junctions of same type, eliminating a null effort (flow) source connected to a 1-junction (0-junction), lumping dependent elements, or some structural equivalencies [46, 47]. It is also possible to leverage, for example, the activity metric presented in this chapter and introduce the *junction inactivity* concept to achieve further structural simplification in a bond graph model. A junction element, 1- or 0-junction, is called *inactive*, if all the bonds that are connected to the junction element have a negligible activity [45]. This concept can be considered as the generalization of the idea that 1-junctions with zero flow and 0-junctions with zero effort can be eliminated from a bond graph without sacrificing the accuracy of the model, because a 1-junction (0-junction) will be inactive not only if its flow (effort) is zero, but also if the efforts (flows) are zero. For the details of this concept and a simplification algorithm based on it, the reader is referred to [45, 48].

2.6. Conclusion

To maximize their utility, mathematical models need to achieve a balance between accuracy and simplicity. One way of achieving this balance systematically is to start with an accurate, but overly complex model and reduce it. Towards this end, three model reduction algorithms based on three different metrics are given in this chapter, that are particularly suitable for reducing bond graph models due to their

common energy-based nature. However, neither the metrics, nor the algorithms are restricted to bond graphs.

Typical benefits that accrue to the user of these algorithms are increased insight into physical system dynamics, fewer model parameters to estimate, and improvements in computation time without significant reduction of model accuracy that render the model more suitable for iterative applications such as optimization and Monte Carlo simulation. Furthermore, these algorithms also have the advantages of being applicable to nonlinear models and realization preserving, where the latter is important to preserve the original meanings of the states and parameters. Finally, the algorithms are trajectory dependant, thereby taking the specific scenarios of interest (i.e., inputs, initial conditions, parameters, time windows) into account for reduction. Thus, the algorithms are capable of yielding different reduced models for different scenarios of interest.

2.7. References

- [1] B.H. Wilson and J.L. Stein: An algorithm for obtaining minimum-order models of distributed and discrete systems. Winter Annual Meeting of the American Society of Mechanical Engineering **41**, 47-58 (1992)
- [2] T. Ersal, H.K. Fathy, L.S. Louca, D.G. Rideout, and J.L. Stein: A review of proper modeling techniques. Journal of Dynamic Systems Measurement and Control **130**, 061008 (2008)
- [3] T. Ersal, J.L. Stein, and L.S. Louca: A bond graph based modular modeling approach towards an automated modeling environment for reconfigurable machine tools. International Conference on Integrated Modeling and Analysis in Applied Control and Automation (2004)
- [4] T. Ersal, J.L. Stein, and L.S. Louca: A modular modeling approach for the design of reconfigurable machine tools. ASME 2004 International Mechanical Engineering Congress and Exposition (2004)
- [5] R.C. Rosenberg and T. Zhou: Power-based model insight. Automated Modeling for Design **8**, 61-67 (1988)
- [6] D.G. Rideout, J.L. Stein, and L.S. Louca: Systematic identification of decoupling in dynamic system models. Journal of Dynamic Systems, Measurement and Control **129**, 503-513 (2007)
- [7] T. Ersal, H.K. Fathy, and J.L. Stein: Realization-preserving structure and order reduction of nonlinear energetic system models using energy trajectory correlations. Journal of Dynamic Systems Measurement and Control **131**, 031004 (8p) (2009)
- [8] T. Ersal, B. Kittirungsi, H.K. Fathy, and J.L. Stein: Model reduction in vehicle dynamics using importance analysis. Vehicle System Dynamics **47**, 851-865 (2009)
- [9] M.M. Loève: Probability theory. Princeton, N.J., Van Nostrand (1955)

- [10] K. Karhunen: Zur spektraltheorie stochastischer prozesse. *Annales Academiae Scientiarum Fennicae* **37** (1946)
- [11] H. Hotelling: Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**, 417-441, 498-520 (1933)
- [12] E.N. Lorenz: Empirical orthogonal functions and statistical weather prediction. *49* (1956)
- [13] J.L. Lumley: The structure of inhomogeneous turbulent flow. *Atmospheric Turbulence and Radio Wave Propagation*, 166-178 (1967)
- [14] G.H. Golub and C.F. Van Loan: *Matrix computations*. North Oxford Academic, Oxford (1983)
- [15] L. Sirovich: Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of Applied Mathematics* **45**, 561-570 (1987)
- [16] L. Sirovich: Turbulence and the dynamics of coherent structures. II. Symmetries and transformations. *Quarterly of Applied Mathematics* **45**, 573-582 (1987)
- [17] L. Sirovich: Turbulence and the dynamics of coherent structures. III. Dynamics and scaling. *Quarterly of Applied Mathematics* **45**, 583-590 (1987)
- [18] C.L. Brooks, M. Karplus, and B.M. Pettitt: *Proteins: A theoretical perspective of dynamics, structure and thermodynamics*. Wiley, New York (1988)
- [19] B.C. Moore: Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **26**, 17-32 (1981)
- [20] L. Segel: *Analysis and prediction of the dynamic behavior of motor vehicles*. (1994)
- [21] The MathWorks Inc., *MATLAB/SIMULINK*, v7, Natick, MA, 2005
- [22] Controllab Products BV, *20-sim*, v3.6, Enschede, The Netherlands, 2005
- [23] Cadsim Engineering, *CAMP-G*, Davis, CA, 2001
- [24] L.S. Louca and J.L. Stein: Ideal physical element representation from reduced bond graphs. *Journal of Systems and Control Engineering* **216**, 73-83 (2002)
- [25] D.G. Rideout, J.L. Stein, and L.S. Louca: Extension and application of an algorithm for systematic identification of weak coupling and partitions in dynamic system models. *Simulation Modelling Practice and Theory* **17**, 271-292 (2009)
- [26] D.G. Rideout, J.L. Stein, and L.S. Louca: Systematic assessment of rigid internal combustion engine dynamic coupling. *Journal of Engineering for Gas Turbines and Power* **130**, 022804 (2008)
- [27] L.S. Louca, D.G. Rideout, J.L. Stein, and G.M. Hulbert: Generating proper dynamic models for truck mobility and handling. *Heavy Vehicle Systems* **11**, 209-236 (2004)

- [28] L.S. Louca and U.B. Yildir: Modeling and reduction techniques for studies of integrated hybrid vehicle systems. *Journal of Mathematical and Computer Modeling of Dynamic Systems* **12**, 203-218 (2005)
- [29] D.M.W. Hoffman and D.R. Dowling: Fully coupled rigid internal combustion engine dynamics and vibration-part i: Model development. *Journal of Engineering for Gas Turbines and Power* **123**, 677-684 (2001)
- [30] T. Ersal: Realization-preserving simplification and reduction of dynamic system models at the graph level. Ph.D. Dissertation, University of Michigan (2007)
- [31] IMAGINE S.A., AMESim, v4.1, Roanne, France, 2003
- [32] B.H. Wilson and J.L. Stein: An algorithm for obtaining proper models of distributed and discrete systems. *Journal of Dynamic Systems, Measurement and Control* **117**, 534-540 (1995)
- [33] B.G. Christensen, J.B. Ferris, and J.L. Stein: An energy-enhanced design of experiments method applied to multi-body models. *ASME International Mechanical Engineering Congress and Exposition DSC 69-1*, 527-534 (2000)
- [34] B.G. Christensen, J.B. Ferris, and J.L. Stein: An energy-enhanced design of experiments method. *ASME International Mechanical Engineering Congress and Exposition DSC 69-2*, 795-799 (2000)
- [35] R.M. Heiberger: *Computation for the analysis of designed experiments*. Wiley, New York, NY. (1989)
- [36] D.C. Montgomery: *Design and analysis of experiments*. Wiley, New York, NY (2005)
- [37] M. Kokkolaras, L.S. Louca, G.J. Delagrammatikas, N.F. Michelena, Z.S. Filipi, P.Y. Papalambros, J.L. Stein, and D.N. Assanis: Simulation-based optimal design of heavy trucks by model-based decomposition: An extensive analytical target cascading case study. *Heavy Vehicle Systems* **11**, 403-433 (2004)
- [38] D. Karnopp and D. Margolis: Analysis and simulation of planar mechanism systems using bond graphs. *Journal of Mechanical Design* **101**, 187-191 (1979)
- [39] D.G. Rideout and J.L. Stein: An energy-based approach to parameterizing parasitic elements for eliminating derivative causality. *International Conference on Bond Graph Modeling* **35-2**, 121-127 (2003)
- [40] H.K. Fathy and J.L. Stein: Fundamental concordances between balanced truncation and activity-based model reduction. *2nd International Conference on Integrated Modeling and Analysis in Applied Control and Automation* (2005)
- [41] P. Sendur, J.L. Stein, H. Peng, and L.S. Louca: An algorithm for the selection of physical system model order based on desired state accuracy and computational efficiency. *ASME International Mechanical Engineering Congress and Exposition* **72**, 891-902 (2003)
- [42] P. Sendur, G.S. Stein, L.S. Louca, and H. Peng: An algorithm for the assessment of reduced dynamic system models for design. *International*

- Conference on Simulation and Multimedia in Engineering Education, 92-101 (2003)
- [43] J.A. Kypuros and R.G. Longoria: Variable fidelity modeling of vehicle ride dynamics using an element activity metric. *ASME 2002 International Mechanical Engineering Congress and Exposition* **71**, 525-534 (2002)
- [44] D.G. Rideout and K.T. Haq: Active modelling: A method for creating and simulating variable-complexity models. *Journal of Dynamic Systems Measurement and Control* (accepted)
- [45] T. Ersal, H.K. Fathy, and J.L. Stein: Structural simplification of modular bond-graph models based on junction inactivity. *Simulation Modelling Practice and Theory* **17**, 175-196 (2009)
- [46] D. Karnopp and R.C. Rosenberg: *System dynamics: A unified approach*. John Wiley & Sons (1975)
- [47] J.R. Rinderle and B.L. Subramaniam: Automated bond graph modeling and simplification to support design. Winter Annual Meeting of the American Society of Mechanical Engineers **34**, 45-68 (1991)
- [48] T. Ersal, H.K. Fathy, and J.L. Stein: Orienting body coordinate frames using Karhunen-Loève expansion for more effective structural simplification. *Simulation Modelling Practice and Theory* **17**, 197-210 (2009)

A
activity, 3
activity index, 5

D
driven partition, 23
driving partition, 23

E
Energetic Contribution Index (ECI), 6

M
model conditioning, 20
model decoupling, 20

Model Order Reduction Algorithm (MORA), 10
model partitioning, 20
model reduction, 2

P
proper model, 2

R
relative activity, 5

S
subgraph, 22
subgraph loop, 22