# Creating Compact Models of Complex Electronic Systems: An Overview and Suggested Use of Existing Model Reduction and Experimental System Identification Tools

Benjamin Shapiro, *Member, IEEE*

***Abstract*—The electronic system packaging community has a great need to reduce the size of its heat transfer simulations so that it can: simulate and analyze more complex systems, include additional physical phenomena, and improve its ability to search the electronic systems packaging design space. Aside from further improvements in machine speed and numerical algorithm efficiency, this is basically a question of model reduction and experimental identification: one would like to know how to dramatically reduce the size of heat transfer simulations when they are available, and one would further like to identify models directly from experiment when accurate, computationally feasible, numerical simulations are not available (as in the case of turbulent flows through complex geometries).**

**Fortunately, the topic of low-order modeling for design has been widely studied and successfully applied in other fields (mostly in control engineering and fluid dynamics, although also in structural mechanics, chemical deposition, heat transfer and combustion). Specifically, there are thousands of papers on the mathematical techniques of model reduction and experimental system identification. This paper gives a brief overview of these techniques, it suggests how these tools might be effectively used for electronic systems including cases that involve unsteady fluid dynamics, and it summarizes some of the reduced-order modeling lessons learned in other fields. The paper includes some of our initial work in model reducing the unsteady heat conduction equation, a result on component model inter-connections, and an outline of a systems level model for an air cooled personal computer.**

***Index Terms*—Cooling, electronic systems, experimental system identification, model reduction.**

## I. Motivation

THIS paper is aimed at creating compact models of complex electronic systems. The basic motivation is the following: for system design, small not-so-accurate models are more useful than large accurate models. This sweeping statement is, of course, overly simplistic, but it is nevertheless correct in the following ways.

Design (especially preliminary design) is all about choosing system parameters (such as chip geometries and positions, cooling channel shapes) and operating ranges (such as fan speed and power inputs). Systems models are the only way that we can do this rationally. Hence, at the very least, our models should provide a capability to combinatorialy search the parameter space: we should be able to set the parameters, run our model, and observe the results. This is not possible with big models. For example, if we have a simulation that takes one day to run, we have 10 parameters that we are interested in, each parameter can take on five values, and we want to search just 1% of the design space: that would take $(5^{10} \times 1)/100$ days or 267 years. If our model runs in 1 s, this same search would take 1.4 days. This is still long for design purposes, but at least it is reasonable.

Second, as our models get smaller, we can apply far more analysis and design tools (see Fig. 1). These tools serve to guide us through the parameter design space by identifying critical parameters, and by providing an understanding of the parameter inter-dependencies. If our model has 10 internal states and five parameters: we can examine the equilibrium behavior, check stability, analyze the forced response, find limit cycles, compute chaotic attractors if they exist, and perform global optimizations such as branch-and-bound. If, instead, our system has 100 000 internal states, we cannot do any of these things—even if the model can be simulated in seconds on a super-computer. This is because the computation time associated with finding a bifurcation diagram, or with computing a branch-and-bound optimum, scales exponentially with model size and the required computation is not feasible for larger models.

Finally, the process of creating low-order models forces the researcher to isolate and quantify the dominant physical mechanisms. This invariably leads to a better understanding of the system behavior, and it usually reveals effective design decisions that would not have been identified through numerical simulation, experimentation, or “black box” optimization methods such as neural networks and genetic algorithms. This increased understanding is the most valuable benefit derived from low-order modeling efforts, but, it is also the benefit that is hardest to convey to skeptical researchers (primarily because it cannot be quantified in any way). To illustrate this last point, we will just point out two examples: insights from the low-order Moore–Greitzer 3-state model [1] have enabled design avoidance and practical nonlinear control (the system is linearly uncontrollable) of surge and stall in jet-engine experiments [2]; the understanding captured by the low-order combustion model [3] has enabled limit cycle control in practical combustion

Manuscript received January 16, 2002; revised December 1, 2002. This work was presented in part at Thermal Challenges in Next Generation Electronic Systems Conference (THERMES 2002), January 13–16, 2002, Santa Fe, NM. This work was recommended for publication by Guest Editor C. J. Lasance upon evaluation of the reviewers’ comments.

The author is with the Department of Aerospace Engineering, University of Maryland, College Park, MD, 20742 USA.

Digital Object Identifier 10.1109/TCAPT.2002.807991

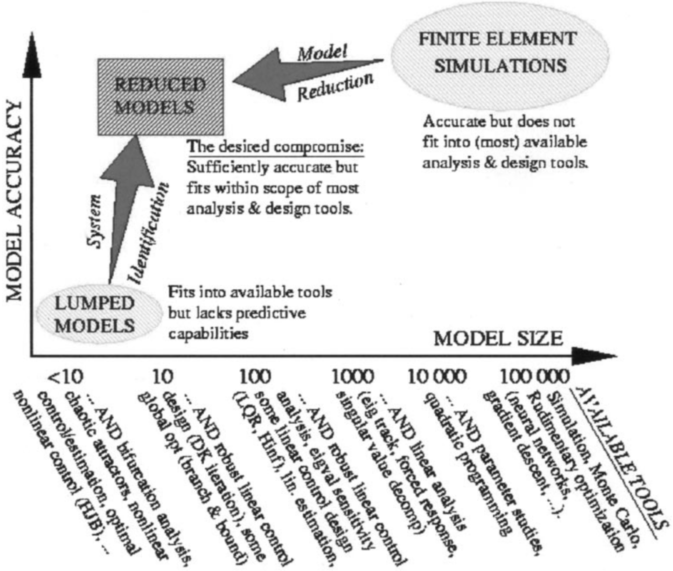



Fig. 1. Model size versus model accuracy tradeoff: we desire small, sufficiently accurate models.

rigs [4]. These practical results would not have been possible without the insights gained from low-order modeling.

Hence our goal in this paper is two-fold. First, we give a broad (and incomplete) overview of the available model reduction and experimental system identification techniques. Second, we provide an indication of how these methods can be used to create compact models of electronic systems. We stress that these tools can aid in the compact modeling effort envisioned in [5], and they should be especially useful for situation that involve complex fluid dynamics.

## II. Component Input/Output Model Reduction Approach

Control engineers spend a great deal of time hooking up components and analyzing/designing the resulting system behavior. It turns out that the most natural way to approach this problem is to view each component as a dynamic input/output map. Here each component model receives time-varying multi-dimensional information from all the surrounding components, this input along with the internal dynamics causes a continual change in the internal states, the resulting multi-dimensional time-varying component output is then an instantaneous function of these dynamic internal states. This same framework is also natural for electronic systems: we would like to have dynamic input/output compact maps for each component, and then have the ability to inter-connect these maps to arrive at systems level compact models. As an example: for a solid chip model, the inputs would correspond to the temperature or heat fluxes from the surrounding environment; the dynamic internal states would correspond to a representation of the internal temperature field; and the outputs would be space and time dependent heat fluxes out of the chip.

The are three key points here.

1) Any systems can be broken up in this fashion.
2) There is no error or approximation associated with this split: it is simply a convenient point of view.

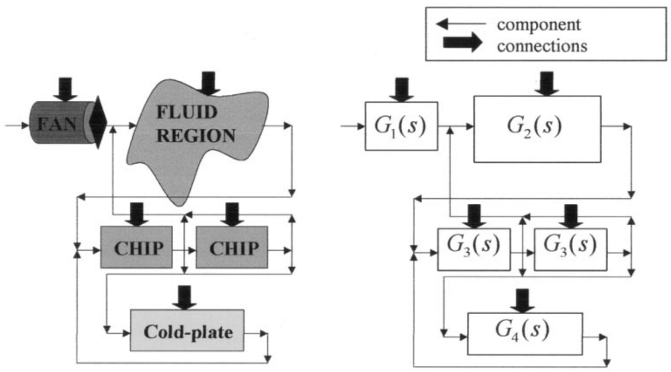



Fig. 2. Interconnection of component models represented by a block diagram. Each block includes dynamics and can be nonlinear. The connection arrows can represent a tremendous amount of information [e.g., the discretized time-varying temperature field $T(x,\, t)$ on a component face]. The behavior of such inter-connections can be analyzed and designed using control engineering tools [6], [7].

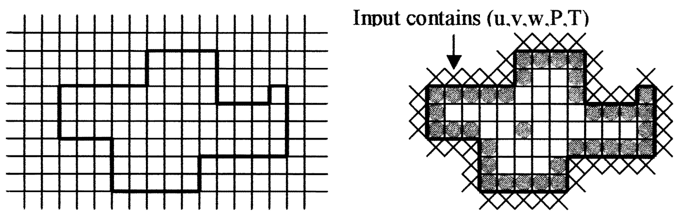



Fig. 3. Partitioning a finite volume or finite element simulation into arbitrary regions. This is just a matter of keeping track of all the region boundary node values: all states stored at the internal boundaries of a region act as outputs to other regions ($\bullet$), all nodes just outside the region act as inputs ($x$). If desired, selected temperatures inside the region can also be listed as outputs (so as to monitor hot-spots). Power input rates to a chip can be listed as inputs. There is no error, approximation, or assumptions associated with this partitioning.

3) There exists a wide array of tools aimed at analyzing and designing the performance of such interconnected input/output maps.

The component inter-connection structure is usually represented by “block diagrams” as shown in Fig. 2, and, given a description of the individual or component models, the resulting system performance can be analyzed completely for linear input/output blocks [6], [7], and partially for nonlinear input/output maps (see for example [8], [9]).

Hence our first recommendation is that the electronic systems low-order modeling problem be couched in this input/output fashion. Accordingly, we discuss how to create compact component input/output models from finite element simulations in Section II-A and from experiment in Section II-B.

### *A. Creating Compact Input/Output Component Models From Finite Element Simulations*

In order to get the full (unreduced) component models $G_j(s)$ of Fig. 2, we can either create a dedicated finite element simulation for that component, or we can take an existing finite element simulation of the type used in [10] and can partition this simulation into component blocks as shown in Fig. 3. Either way, this yields a large, but finite, dynamic input/output map for each solid component or chosen fluid region

$$y(t) \Leftarrow \begin{bmatrix} \dot{x} = f(x,\, u) \\ y = h(x,\, u) \end{bmatrix} \Leftarrow u(t). \tag{1}$$

Here, $(f,\, h)$ corresponds to a discretization of the underlying heat transfer PDE's as written in [10], $x$ is the internal state in the region (the value of all variables at all nodes), $u$ lists all the inputs to the region (variables at immediately adjoining nodes if we follow the convention shown in Fig. 3, or we can also think of this as incoming fluxes, plus any inputs due to power rates into the chip), and $y$ lists all the system outputs (boundary nodes or outgoing fluxes, plus any additional variables that we may wish to observe).

It is the dimension of the internal state variable $x$ that we are trying to reduce in equation (1). So for each component we desire a reduced-order map $(\tilde{f},\, \tilde{h})$ that has the same input $u$ and output $y$

$$y(t) \Leftarrow \begin{bmatrix} \dot{x} = f(x,\, u) \\ y = h(x,\, u) \end{bmatrix} \Leftarrow u(t)$$
$$\rightarrow \quad y(t) \Leftarrow \begin{bmatrix} \dot{\tilde{x}} = \tilde{f}\,(\tilde{x},\, u) \\ y = \tilde{h}\,(\tilde{x},\, u) \end{bmatrix} \Leftarrow u(t) \tag{2}$$

but with a dramatically reduced number of internal states $(\dim(\tilde{x}) \ll \dim(x))$, that provides essentially the same dynamic input-to-output behavior. [The analogy here is this: consider a drum where the input is the hand striking the surface [so a complex input], and the output is the instantaneous displacement of the drum at each point in space [a complex output]. A low order input/output model would represent the internal states of the drum by the amplitude and phase of the first few fundamental drum modes. Even the first few modes will adequately recreate the input/output behavior. This is true because these modes are the natural (i.e., dominant) modes of the drum. The key model reduction step, therefore, is to find a judicious set of modes onto which the dynamics may be projected. This is exactly what model reduction techniques endeavor to do.]

A review of various methods of model reduction for linear systems and their application to the design of a suitable controller for the original system can be found in [11]. Techniques for Pade approximation (another model reduction technique) and an extensive bibliography on the topic can be found in [12]. As an example of a concrete application, the modeling and control of transitional and turbulent shear flows can be found in [13]. A partial overview of model reduction techniques can also be found in the following references: [14]–[23]. Below, we outline two specific techniques that can achieve dramatic reductions in model size $(f,\, h)$ of equation (1).

*Balanced Truncation Model Reduction [6]:* Balanced truncation is aimed at large linear input/output models where $\dot{x} = f(x,\, u) = Ax + Bu$, $y = g(x,\, u) = Cx + Du$. Hence it applies to components whose internal behavior is linear or almost linear (e.g., solid components where conduction is dominant), but it can also be applied successfully to nonlinear maps if the underlying modes do not change appreciably as the states deviate away from their nominal settings.

The goal of balanced truncation is to replicate the same input/output behavior of a linear transfer function with far fewer internal states. Thereby, it seeks to transform

$$y(t) \Leftarrow \begin{bmatrix} \dot{x} = Ax + Bu \\ y = Cx + Du \end{bmatrix} \Leftarrow u(t)$$

into

$$y(t) \Leftarrow \begin{bmatrix} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u \\ y = \tilde{C}\tilde{x} + \tilde{D}u \end{bmatrix} \Leftarrow u(t) \tag{3}$$

with $\dim(\tilde{x}) \ll \dim(x)$. In order to find the reduced-order matrices $\tilde{A},\, \tilde{B},\, \tilde{C},\, \tilde{D}$ one first computes a transformation matrix $z = Tx$ that orders the internal states from most to least participation in the input/output behavior. Some states $z_j$ in the linear input/output map will be affected greatly by the inputs $u(t)$, but will not be clearly visible. Other states will be greatly visible, but will not be affected greatly by the input. What counts is the magnification of the input to each mode, times the amount of visibility, divided by the damping of that mode—this gives the net participation of that mode in the input/output behavior of the linear map. The matrix $T$ orders the modes in decreasing order of input/output importance; it does so by diagonalizing and matching the controllability and observability Grammians [6]. This process is called balancing the system. To arrive at a reduced-order model one truncates the modes of lesser importance in the input/output map, so $\tilde{x} = trunc(z)$, to arrive at a model of any desired dimension.

Balanced truncation model reduction has the following properties.

1) The method is globally optimal [24]. A 5th order balanced truncation model is the best possible 5 dimensional reduced-order model in the sense that it globally minimizes the model reduction error transfer function in the infinity norm.
2) Balanced truncation includes a rigorous *a priori* model-reduction error bound [6], [25]. The error between the output based on the original system $y = Gu$, and the output based on the reduced $k$-dimensional system $G_k$, is bounded as $\|y(t) - y_k(t)\|_2 \leq \varepsilon(k)\|u(t)\|_2$ for any input $u(t)$ and for any initial conditions. Here $\varepsilon = 2(\sigma_{k+1} + \sigma_{k+2} + \cdots + \sigma_n)$ is the sum of the truncated singular values of $G$.
3) We can aim the reduced order model at any desired frequency range. So if we are interested in transients that have a 5–30 Hz content, we can generate a reduced-order model that is focused on this frequency range.
4) Balanced truncation reduced-order models can be computed for models with up to thousands of states [26], [27]. Hence it is possible to reduce some FEM input/output component models directly using balanced truncation. The method is most suited to solid components or fluid regions that have only weak nonlinear terms, or those components that can be accurately linearized about an operating point (so cases with radiation are appropriate). The method is not suited to cases that include highly nonlinear fluid dynamic effects.

*Proper Orthogonal Decomposition (POD) [28]:* The advantage of POD model-reduction is twofold: first, it is aimed specifically at nonlinear situations; second, it has been demonstrated to work superbly in complex fluid dynamic systems [28]–[30]. POD methods have been shown to reduce RANS or LES turbulent simulations from millions of internal states down to 5 or 10 modes with only a 5 or 10% loss in accuracy [29]. The method works directly from simulation data: one runs the desired FEM

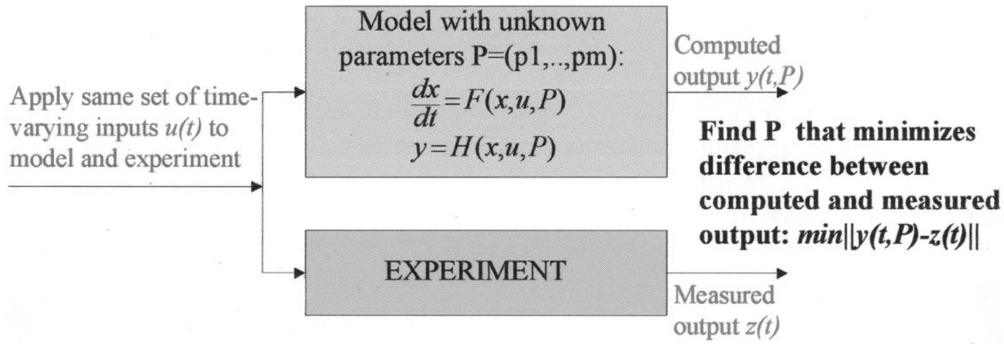



Fig. 4. Experimental system identification aims to find unknown model parameters $P$ by minimizing the error between the input/output behavior of a model and the input/output behavior of a carefully instrumented experiment. Here $P$ can be a set of unknown coefficients, the discretization of an unknown nonlinear curve, or the parameters of a linear or nonlinear transfer function. The optimization is usually solved using some variant of (linear or nonlinear) gradient descent techniques, least squares regression techniques, or black box methods such as genetic algorithms and neural networks.

simulation enough times to generate a statistically representative sample of data, POD then finds the low-order subspace that most closely approximates this data in the 2-norm sense, the full dynamics are then Galerkin projected onto this subspace to create the low-order input/output model. POD retains most of the desirable properties of balanced truncation: it is optimal with reduction-error bounds (although now in a statistical sense), and since it can be applied in the frequency domain [31] it can be frequency weighted to focus on the desired frequency range.

POD is the preferred method of choice for reducing complex fluid simulations, especially those involving turbulence. It can be used in concert with locally averaged fluid equations (one could first apply the method of Reynolds Averaged Navier Stokes [32], and then apply POD to the RANS equations), but it provides more detailed results than methods which average the fluid dynamics over large (nonlocal) domains. For example, if the flow between heat sink fins is being considered [33], then instead of averaging over an analytic "developing channel flow" solution, POD will average over a small set of optimal numerical basis functions. This will give more accurate results in cases where the flow field is not close to an analytically known solution. For recent successful applications of POD techniques see [29], [30], [34].

### *B. Creating Low-Order Input/Output Component Models From Experiment*

Experimental system identification methods will be necessary when accurate FEM simulations are not feasible. This might be because FEM simulations cannot resolve the physics, as is the case in many turbulent flow scenarios, or it can be the case because a particular component is so complex that it makes more sense to identify its behavior experimentally than via simulation, as might be the case for two-phase cooling devices. (There are many instances where first principle modeling is impractical, but the experimental system identification problem is straightforward—for example, the solar-heated house system identification problem in [35], [36].) In both scenarios above, experiments should be employed to identify the missing parts of the system model.

Like model reduction, system identification is an intensely studied field in its own right [3], [24], [35], [37]–[40]. System identification works essentially as shown in Fig. 4. Any system identification problem can be phrased as an optimization. One phrases a cost function $C(P) = \min_P \|y(t, P) - z(t)\|$ that reflects the error between the experiment and the model, and then an optimization problem is solved to find the set of parameters that minimize this cost. The crucial question is whether this optimization is tractable. If the experimental system behavior is linear, the model has a small number of unknowns $P_1, P_2, \ldots, P_m$, and the experiment provides a large set of clean data for many experimental inputs $u(t)$, then the answer is almost certainly yes. If the experimental system exhibits strongly nonlinear behavior, the model has a large number of unknowns $P_1, P_2, P_3 \ldots, P_{100}, \ldots, P_M$, the data is noisy, and the inputs $u(t)$ are limited, then the answer is most likely no. There are continual research efforts to extend the efficacy of system identification methods to larger models, more unknown parameters, and more strongly nonlinear behaviors. For an overview of system identification tools see [24], [35], [37]–[41]..

System identification has enabled low-order modeling of complex systems when it has been directed at the specific, dominant, unknown physical effects. For example, in the case of pre-mixed combustion [3], experimental identification was used to identify the time-delayed nonlinear pressure-to-heat-release input/output map. When this map was combined with a low-order model of the chamber acoustics derived from CFD, the resulting system model allowed the design and implementation of a limit cycle controller in practical combustion rigs [4]. For the case of convective cooling in laptops and desktops, the recent paper by Bachmayer *et al.* [42] is especially relevant. This paper shows how it is possible to experimentally identify the lift and drag coefficients of a fan (here used in water). The resulting 3 state model of the motor dynamics (known), and the fan dynamics (model completed by the experimentally identified lift and drag coefficients), accurately predicts the steady state and transient flow velocities out of the fan. This type of modeling is the next step as compared to the steady-state fan curve fitting models available in ICEPACK or FLUMECAD software.

## III. Example: 2-D Unsteady Conduction Model Reduction by Balanced Truncation

In Figs. 5 and 6 we show how balanced truncation [6] can be used to decrease the size of a time-varying conductive heat

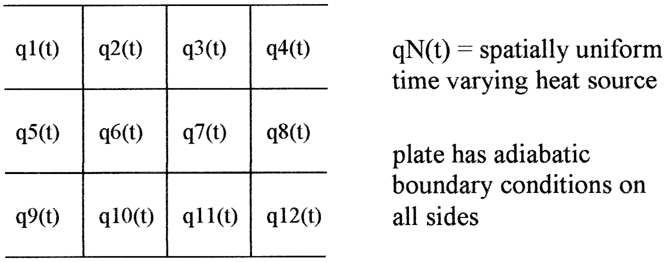



Fig. 5. Example geometry: a rectangular plate described by the heat equation (conduction only) with insulated edges, and with 12 time-varying heat sources. Each source covers one of the 12 squares shown, is constant in space, but can vary in time. The sources are as follows: $q2 = \sin(6\pi t)$, $q3 = 0.3–0.07t$, $q4 = 2\cos(14\pi t)$, $q5 = 0.2\exp(t/10)$, $q9 = -1.5(t-6)$ for $t > 6$ and is zero otherwise, and $q10 = 2$ for $t > 7$ and is zero otherwise. At time $t = 0$, the plate is at a uniform temperature.

transfer simulation by a factor of 5 with only a 5% resulting system error.

Here each "block" in the plate is viewed as an input/output component model where the local (time-varying) external heating plus the temperature values from adjoining blocks is treated as a vector input, and the temperature at the inside boundary of the current block is treated as a vector output. One can compute balanced truncations for linear models with up to thousands of states [26], [27], hence the current 2-D blocks with 576 elements each are within easy reach. For linear systems, model reduction guarantees a rigorous upper bound on the error between the original and reduced model. Here the error in each block is below 2% for any time-varying external heating input. The resulting system error, after all the reduced component models are inter-connected, is about 5% (see Fig. 6).

### *A. Example of the Type of Results Possible Within the Input/Output Model Reduction Framework*

The example above illustrates a salient point, and it also permits a demonstration of why model reduction techniques with known properties and known tuning knobs are far more useful than *ad-hoc* techniques. The central question above is: How do component model reduction errors build up to give system errors? And how can we create a library of low-order component models that, when inter-connected, will minimize systems level model reduction errors?

It is possible to answer this question (for stable linear component models inter-connected in a stable manner) by extending the basic controller reduction result found in [43]. The answer is interesting and physically intuitive, hence we suspect that it will extend to nonlinear systems in some fashion. In order to build accurate system level models from component reduced-order models, one must ensure that the component models are accurate at the resonant frequencies of the full system (see Fig. 7). Basically, compact models for each component need only capture dynamics at those frequencies that will show up in the full system. So if the entire laptop has dynamics in the 5–50 Hz range, it is unnecessary to resolve turbulent flows at the 200–400 Hz range—such high-frequency dynamics will simply not show up in the final system. Hence any reduced fluid-region model need only be accurate in the 5–50 Hz range. Not only is this an easier model reduction question (we can ignore high-frequency details), but it is also fits perfectly into the frequency weighted model reduction tools developed in [25], [44]

## IV. Putting It All Together

Above we have discussed model reduction and experimental system identification tools, and have also illustrated how the input/output framework can be used to unravel model inter-connection issues. In this last section, we discuss how the tools above might be integrated to create compact models of complex electronic systems.

In Fig. 8 we show a schematic of a desktop personal computer. In order to make design decision, we basically need to know how to direct the airflow, and how much heat is generated and removed at critical components. So we desire the smallest possible system model that might be able to answer such questions.

The conceptual flow of the desired compact system model might then have the framework shown in Fig. 9. This figure illustrates how it is possible to split the electronic system behavior into a set of dynamic input/output models. Starting at the top left, the fan speed largely determines the fluid velocity field in the system, but this velocity field can also depend on heat fluxes generated by the chips and circuit boards (especially when the fan is off), hence the velocity map receives these two inputs. The dynamic velocity map is a spatially rough representation of the fluid dynamics, it resolves the fluid dynamics down to cm, or maybe mm, length-scales. This part of the model outputs a spatially averaged velocity (and temperature) field that captures the path of the fluid over the chips but it makes no attempt to resolve the fine scale flow features. The heat carried away by convection is then given by an empirically determined heat transfer relation: at chip 13, averaged velocity, temperature is $v$, $T$, so heat convected away is $q = F(v, T)$. This is done for each chip. Now that the convective fluxes are known, it only remains to solve the conduction/radiation heat transfer in the solid. This is done to low-order by the balanced truncation model. (If the nonlinear radiation effects are significant, one can replace this by a POD model that captures nonlinear radiation effects more accurately.) If necessary, this model can also be coupled to a cold-plate balanced truncation or POD model. To complete the loop, convective heat fluxes go to the reduced-order fluid model. One can pull out any desired system output: here the chip temperatures at any given set of points is being observed.

The solid chip and cold plate input/output models will be created by balanced truncation: each chip will have incoming fluxes as inputs, and outgoing fluxes as outputs; selected chips will also output their temperatures for observation purposes. The turbulent flow field reduced-order model will be created in one of three ways: by reducing a full scale turbulent simulation through POD, by creating a low-resolution (cm length-scale) flow simulation, or by a flow visualization experiment. This model will not attempt to resolve the fine scale turbulence, but will only give a broad indication of the path and velocity of the flow over each chip. The convective heat transfer caused by the turbulent flow field will then be captured by the experimentally identified "local averaged velocity" to "local heat flux" map. It is this experimentally identified map that will capture the details of the turbulent flow heat transfer mechanisms. Since these

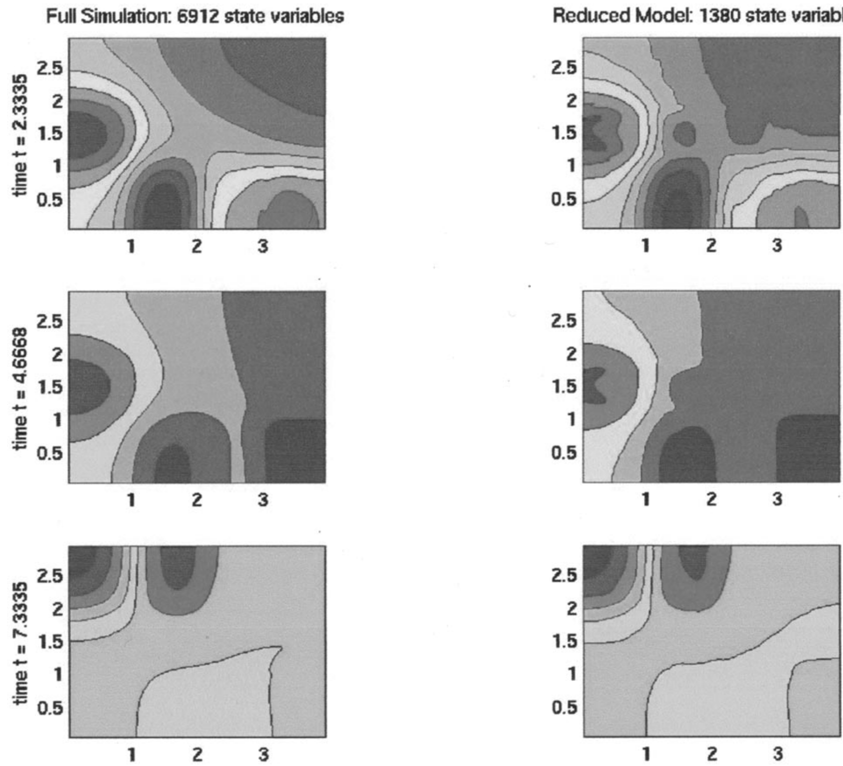



Fig. 6. Temperature field snapshots for a conducting plate divided into "blocks," each block is heated by a time-varying but spatially uniform source. The left snapshots show a full simulation with 6912 free variables. The right snapshots show results generated by connected reduced-order models with 1380 free variables total. Model reduction for each block is optimal, and the error per block is guaranteed to be less than 2% for any time-varying heat input. This yields a 5% systems level error. (The error is defined as the infinity-norm of the full transfer function $G$ from all inputs [the heating sources] to all outputs [the temperature field] minus the reduced transfer function $G_k$ from the same inputs to the same outputs: $\varepsilon = \|G - G_k\|_\infty$.)

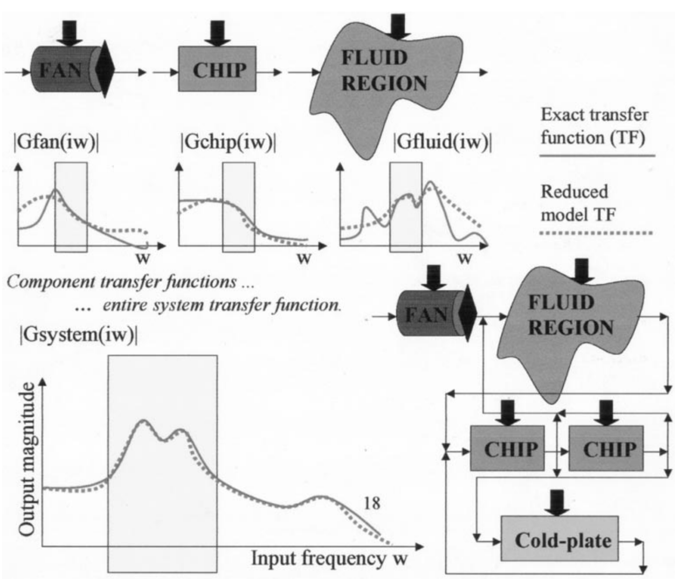



Fig. 7. Model reducing components so as to minimize system level model reduction errors. The top line shows three typical components with their corresponding transfer function (TF) models. (A linear input/output map can always be uniquely represented by a transfer function.) When all of the component TF's are inter-connected (bottom right) then there is a total system frequency response (bottom left). This response will have some resonant modes here highlighted by the shaded rectangle. If the low-order component models are accurate in this middle frequency range (top left, dotted lines) then their inter-connection will accurately capture the full model system response everywhere.

details cannot be captured by current simulations, experimental system identification is the most reasonable approach.

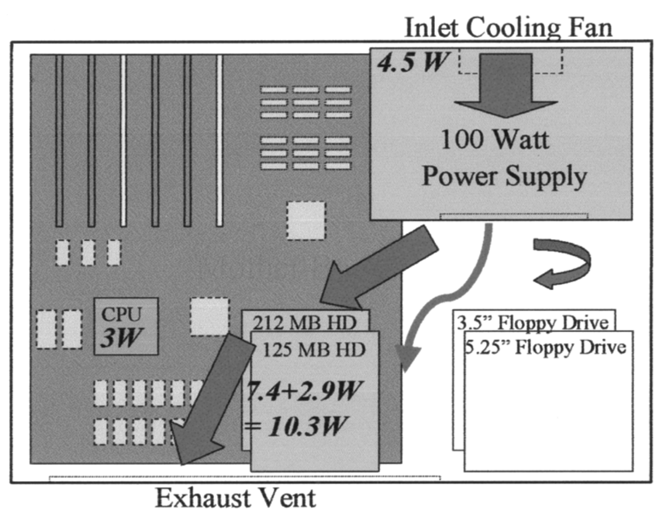



Fig. 8. Typical forced air cooling application (desktop personal computer). Heat is generated in essentially three key places. Arrows represent the direction of fluid flow. At least for preliminary design purposes, and possibly even for more advanced uses, it should only be necessary to capture the thermal characteristics in these three regions. (Figure courtesy of Dr. Y. Joshi.)

Our assessment is that this type of model can capture the dominant system dynamics, and that it will reveal the tradeoffs due to parameters such as fan speed and chip placement. One could certainly point out possible short-comings in Fig. 9. For example, the heat flux on the spatially average flow field might be negligible compared to the flow field generated by the fan and should be neglected, or the empirical identified heat-transfer relation might be insufficient plus it might have large error bars,

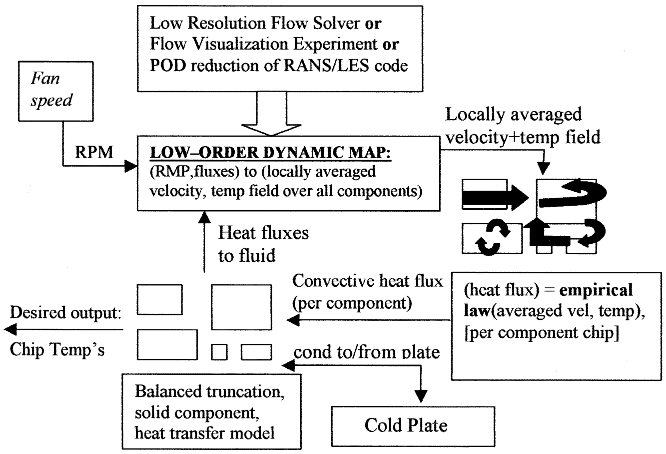



Fig. 9. Putting it all together: A conceptual flow diagram for low-order system model of an electronic system with convective cooling.

or there might be a missing block in the diagram. In all likelihood, these, and other similar issues, will have to be addressed.

Still, the key point is that it is possible to split a complex modeling problem into physical or component blocks as shown, and then one can find reduced-order models for each block, and connect these models to form a small but accurate model of the entire system. This divide and conquer approach has been demonstrated for many complex engineering systems [1], [3], [28]–[31], [34], [42], [45]–[47], and we hope to apply it successfully to electronic systems.

## V. Open Research Question: Geometry as a Model Parameter

Above we have described how model reduction techniques might be applied to electronic systems. However, there are issues raised by electronic systems that are beyond the capabilities of current model reduction techniques. The most critical of these is geometric variations. Basis modes are computed for a specific geometry, hence it is not clear how to effectively parameterize the change in basis functions with geometry changes so as to arrive at a reduced-order model that includes geometric variations. (In terms of the drum analogy we used in Section II-A, if we change the shape of the drum then we have changed the fundamental modes used to create the reduced-order model.)

There are a number of possible work-arounds. The most straight-forward solution is the one we have been implicitly using up till now: we divide the geometric space into a set of blocks, derive reduced order models for each block, and then swap the positions of blocks so as to represent different placements of chips, fans, and cold-plates. This allows us to capture discrete changes in geometry, and it relies on each reduced-order block model having an accurate input/output connection to any of a number of neighbors it might see.

If, instead, the geometry changes continuously, we can interpolate between multiple reduced-order models to create a single model that includes geometric variations. This leads to a (difficult) function fitting problem.

An alternate approach is provided in [48]. The authors considered a 2-D diffuser with a varying ramp angle. They mapped all the different geometries back to a mesh on a straight channel by a continuous change of variables. Hence they were able to compute their POD modes for multiple ramp angles via data that was always transformed back to the simple channel geometry.

## VI. Conclusion

There is a strong need for compact or reduced-order models in the electronic systems thermal management field. The required compact models can be created in one of two ways:

a) one can take an existing large model and reduce its size whilst keeping most of its accuracy—this process is known as model reduction;
b) one can pick a model structure and then use an experiment to identify the missing parts in the model—this is known as experimental system identification.

Fortunately, both techniques have been actively studied for many years, and so the electronic systems community can make use of the existing results. This paper summarizes some of the basic ideas behind these methods, it suggests how they might be applied to electronic systems, and it gives some of our initial results in this direction.


## Acknowledgment

The author would like to thank Dr. Y. Joshi for his valuable input on electronic system issues.


## References


[1] E. M. Greitzer, "Surge and rotating stall in axial flow compressors: Part 1—Theoretical compression system model, Part 2—Experimental results and comparison with theory," *ASME J. Eng. Power*, vol. 98, pp. 190–198, 1976.
[2] G. Gu, S. Banda, and A. Sparks, "An overview of rotating stall and surge control for axial flow compressors," in *Proc. 35th IEEE Conf. Decision Contr.*, 1996, pp. 639–647.
[3] R. M. Murray, C. A. Jacobson, R. Casas, A. I. Khibnik, C. R. J. Johnson, R. Bitmead, A. A. Peracchio, and W. M. Proscia, "System identification for limit cycling systems: A case study for combustion instabilities," in *Proc. Amer. Contr. Conf.*, 1998.
[4] A. Banaszuk, K. Ariyur, M. Krstic, and C. A. Jacobson, "An adaptive algorithm for control of combustion instability," *Automatica*, 2000.
[5] D. S. Boyalakuntla and J. Y. Murthy, "COBRA-based compact models for simulation of electronic chip packages," in *Proc. IPACK Pacific Rim/ASME Int. Electron. Packag. Technol. Conf. Exhibition*, Kauai, HI, 2001.
[6] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Upper Saddle River, NJ: Prentice-Hall, 1996.
[7] G. F. Franklin, J. D. Powell, and A. Emaim-Naeini, *Feedback Control of Dynamic Systems*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[8] I. Horowitz, "Survey of quantitative feedback theory (QFT)," *Int. J. Robust Nonlinear Contr.*, vol. 11, pp. 887–921, 2001.
[9] J.-M. Coron, L. Praly, and A. Teel, "Feedback stabilization of nonlinear systems: Sufficient conditions and Lyapunov and input–output techniques," in *Proc. Trends Contr.. Eur. Perspective*, Berlin, Germany, 1995.
[10] V. H. Adams, Y. Joshi, and D. L. Blackburn, "Three-dimensional study of combined conduction, radiation, and natural convection from discrete heat sources in a horizontal narrow-aspect-ratio enclosure," *Trans. ASME, J. Heat Transfer*, vol. 121, pp. 992–1001, 1999.
[11] N. K. Sinha, "Reduced-order models for linear systems," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, New York, NY, 1992.
[12] A. Bultheel and M. V. Barel, "Pade techniques for model reduction in linear system theory: A survey," *J. Computat. Appl. Math.*, vol. 14, pp. 401–438, 1986.

[13] I. Mezic and S. Narayan, "Overview of some theoretical & experimental results on modeling and control of shear flows," in *Proc. Conf. Decision Contr.*, Sydney, Australia, 2000.

[14] E. Chiprout and T. Nguyen, *Survey of Model Reduction Techniques for Analysis of Package and Interconnect Models of High-Speed Designs*. New York, NY: IEEE, 1997.

[15] T. T. Hartley, H. K. Qammar, and L. D. Murphy, "Model reduction methods for chaotic systems," *Kyberneticak 1993*, vol. 29, 1993.

[16] I. Troch, P. C. Muller, and K. H. Fasol, "Model reduction for simulation and controller design," *Automatisierungstechnik*, vol. 40, 1992.

[17] G. Obinata and B. D. O. Anderson, *Model Reduction for Control System Design*. London, U.K.: Springer, 2001.

[18] D. Hinrichsen and H. W. Philippsen, "Model reduction using balanced realizations," *Automatisierungstechnik*, vol. 38, 1990.

[19] P. N. Paraskevopoulos, *Model Reduction Techniques for 2-D Discrete Systems*. Amsterdam, Netherlands: North Holland, 1985.

[20] A. M. S. Dahshan, A. S. Fawzy, and Y. A. E. Sawy, *Comparisons of Several Model Reduction Techniques*. New York, NY: IEEE, 1985.

[21] S. Shokoohi, *A Survey of Model Reduction Techniques for Large-Scale Systems*. Morgantown, WV: Western Periodicals Co.), 1984.

[22] M. J. Bosley and F. P. Lees, "A survey of simple transfer-function derivations from high-order state-variable models," *Automatica*, vol. 8, 1972.

[23] N. K. Sinha and G. J. Lastman, "Reduced order models for complex systems—A critical survey," in *Proc. IETE Tech. Rev.*, 1990.

[24] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, "A survey of model reduction methods for large-scale systems," in *Applied Math*. Houston, TX: Rice University, Dec. 5, 2000.

[25] G. Wang, V. Sreeram, and W. Q. Liu, "A new frequency-weighted balanced truncation method and an error bound," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1734–1737, Sept. 1999.

[26] D. C. Sorensen and A. C. Antoulas. (2001, Mar.) Projection methods for balanced model reduction. Tech. Rep. TR01-03, Applied Math Dept., Rice Univ., Houston, TX. [Online]. Available: http://www.caam.rice.edu/caam/caam-techrep.html.

[27] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 439–448, Nov. 1995.

[28] L. Sirovich, "Turbulences and the dynamics of coherent structures, Parts I & II," *Quart. Appl. Math*, vol. 45, pp. 561–571, 1987.

[29] C. W. Rowley, T. Colonius, and R. M. Murray, "Dynamical models for control of cavity oscillations," in *Proc. 7th AIAA/CEAS Aeroacoust. Conf.*, 2001.

[30] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," in *Proc. AIAA Computat. Fluid Dyn. Conf.*, Anaheim, CA, 2001.

[31] B. Shapiro and K. Willcox, "Analyzing mistuning of bladed disks by symmetry and reduced-order aerodynamic modeling," *J. Propulsion Power*, 2000.

[32] S. B. Pope, *Turbulent Flows*. London, U.K.: Cambridge Univ. Press, 2000.

[33] C. K. Loh, D. Nelson, and D. J. Chou, "Thermal characterization of fan-heat sink systems in miniature axial fan and micro blower airflow," in *Proc. 17th IEEE Semicond. Thermal Meas. Manag. Symp.*, 2001.

[34] W. R. Graham, J. Peraire, and K. Y. Tang, "Optimal control of vortex shedding using low order models—Part I: Open-loop model development," *J. Numer. Meth. Eng.*, vol. 44, pp. 945–972, 1999.

[35] L. Ljung, *System Identification Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[36] L. Ljung and T. Glad, *Modeling of Dynamic Systems*. Upper Saddle River, NJ: Prentice-Hall, 1994.

[37] K. J. Astrom and P. Eykhoff, "System identification—A survey," *Automatica*, vol. 7, pp. 123–162, 1971.

[38] S. A. Billings, "Identification of nonlinear systems—A survey," *Proc. IEEE D*, vol. 127, pp. 272–285, Nov. 1980.

[39] J. A. M. Fellipe De Souza, "Control, state estimation, and parameter identification of nonlinear distributed parameter systems using fixed point techniques: A survey," in *Proc. 4th IFAC Symp., Contr. Distrib. Parameter Series*, 1987.

[40] R. Isermann, "Parameter estimation of dynamic systems," *Automatisierungstechnik*, vol. 36, pp. 199–207, 1988.

[41] B. Armstrong-Helouvry, P. Dupont, and C. Canudas de Wit, "A survey of models, analysis tools, and compensation methods for the control of machines with friction," *Automatica*, vol. 30, pp. 1083–1138, 1994.

[42] R. Bachmayer, L. L. Whitcomb, and M. A. Grosenbaugh, "An accurate four-quadrant nonlinear dynamical model for marine thrusters: Theory and experimental validation," *IEEE J. Oceanic Eng.*, vol. 25, pp. 146–159, Jan. 2000.

[43] B. D. O. Anderson, "Controller reduction: Concepts and approaches," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 802–812, Aug. 1989.

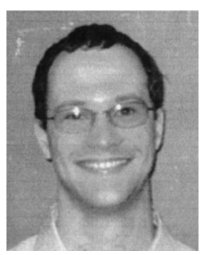
[44] K. Zhou, "Frequency-weighted linfinity norm and optimal hankel norm model reduction," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1687–1699, Oct. 1995.

[45] S. Lall, P. Krysl, and J. E. Marsden, "Structure-preserving model reduction of mechanical systems," Dyn. Stab. Syst., 2000. [Online]. Available: http://element.stanford.edu/~lall/home/publications/publications.htm., submitted for publication.

[46] R. T. M'Closkey, J. S. Gibson, and J. Hui, "Model identification of the JPL microgyroscope," in *Proc. Amer. Soc. Mech. Eng., Dyn. Syst. Contr.*, 1998.

[47] R. A. Sahan, "Low-order dynamical modeling of natural convective air-cooling system in a vertical channel," in *Proc. ITHERM: 7th Intersoc. Conf. Thermal Thermomech. Phenom. Electron. Syst.*, Piscataway, NJ, 2000.

[48] A. I. Khibnik, S. Narayanan, C. A. Jacobson, and K. Lust, "Analysis of low dimensional dynamics of flow separation," *Continuation Methods Fluid Dyn.*, vol. 73, pp. 167–178, 2000.


**Benjamin Shapiro** (M'01) was born in Jerusalem, Israel, in 1973. He received the B.S. degree in aerospace engineering from the Georgia Institute of Technology, Atlanta, in 1995, and the Ph.D. degree from the Control and Dynamical Systems Department, California Institute of Technology (Caltech), Pasadena, in 1999. His Ph.D. thesis was focused on fluid dynamics and symmetry breaking in jet-engines.

He joined the Aerospace Department, University of Maryland, College Park, as an Assistant Professor in 2000. His current research interests include modeling and control of micro-fluidic motion through applied electric fields, modeling and design of multi-channel micro-fluidic networks driven via surface tension, and experimental identification of friction dynamics on the micro-scale. He is part of the interdisciplinary Small Smart Systems Center, University of Maryland. He has a number of collaborations with outside MEMS groups, and is part of the modeling and design MEMS efforts at Nanostream, Inc., University of California at Los Angeles, and the National Institute of Standards and Technology. He also organizes the Minta Martin Seminar Series, Department of Aerospace Engineering.