Getting to the Gray Box

Some challenges for model reduction

George Verghese (MIT)

ACC 2009

What this talk is about

What this talk is about

1. Large-scale structured dynamic models
    - heterogeneous, mechanistic
    - (not discretizations of a PDE — fluid dynamics, heat flow, elastic structures, electromagnetic transmission, etc.)
    - medical physiology, power systems, systems biology ...
    - settings in which the large model is an evolving summary of accumulated knowledge about a system

    - large uncertainty in parameters
    - no *a priori* fixed set of input and output channels

    - discovering and exploiting structure
      ... challenges, pleasures, strategies, tools

What this talk is about

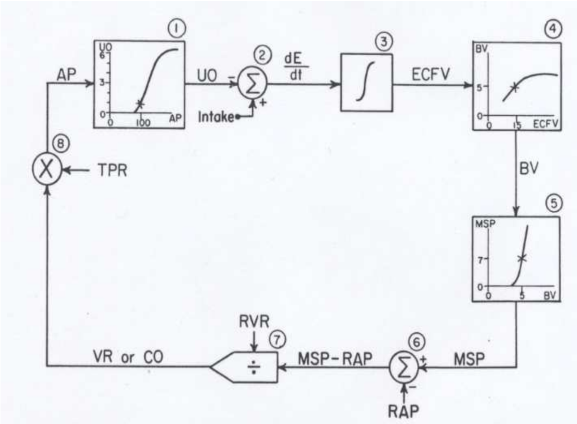
2. Model reduction

    - to structured gray-box models that are interpretable in some meaningful and useful way, in the setting of the original large model
    - not just a one-time interaction to solve a single abstracted computational problem
    - rather, a repeated engagement to refine and consolidate one's understanding of the model and the system
    - a broad approach — using a variety of tools — can yield good insight/results
    - carrying lessons back to the original large model

Medical physiology

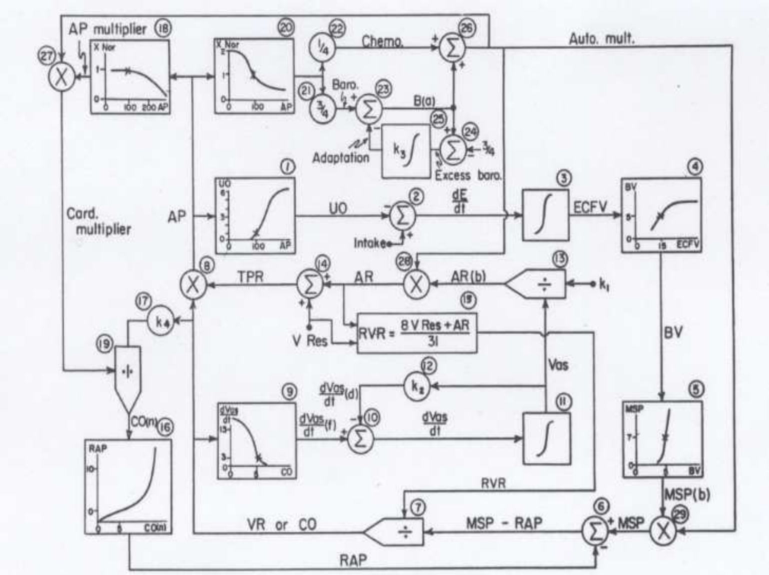
A gray box from physiology

- Structured, mechanistic model
- Physically meaningful parameters, incompletely specified



A gray box from physiology (8 blocks), Guyton *et al.*, 1967

Accretion and refinement (yields 28 blocks)



More familiarly, circulation

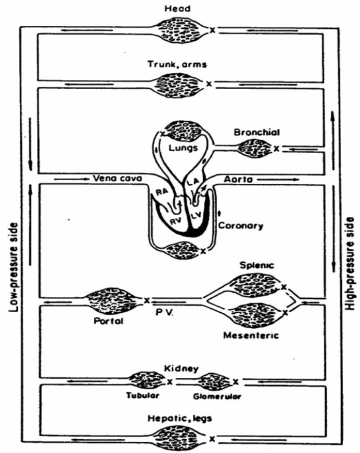
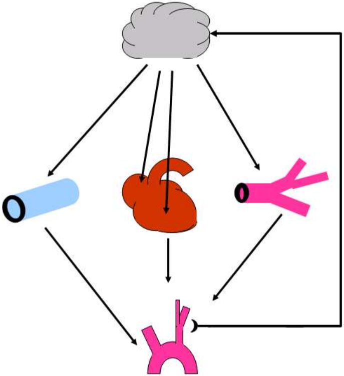


Figure 3: Arrangement of the parallel routes by which the circulation passes from the aorta to the vena cava. Representatives of the different categories of route discussed in the text are indicated. The Xs indicate the location of control points where arterioles may control the flow. *RA*, right atrium; *LA*, left atrium; *RV*, right ventricle; *LV*, left ventricle; *PV*, portal vein. (from Green, H.D.: Circulation: Physical Principles, in Glasser, O. [ed.]: *Medical Physics*, Vol. 1 [Chicago: The Year Book Publishers, Inc., 1949], p. 210. Original illustration kindly furnished by H.D. Green.)

One of many feedback control loops: baroreceptor reflex



Sensor mechanisms: pressure sensors (baroreceptors)

- carotid sinus
- aortic arch

Effector mechanisms:

- heart rate
- cardiac contractility
- arteriolar resistance
- venous tone

A larger-scale gray box (354 blocks), Guyton *et al.*, 1972

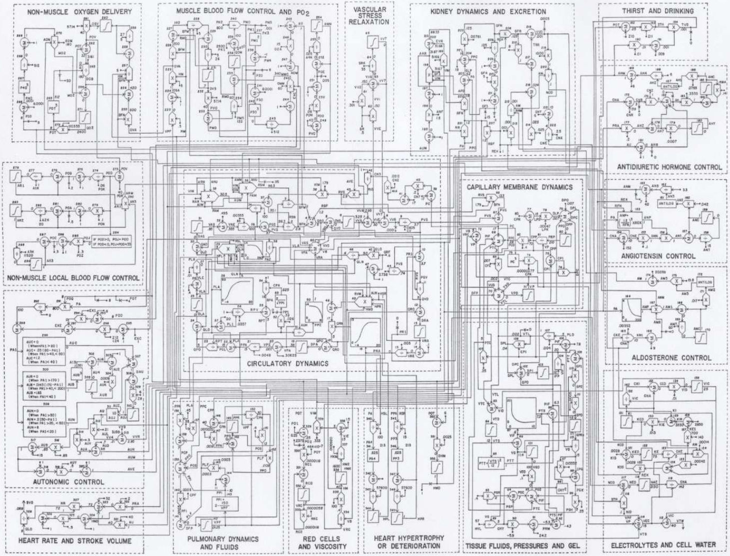
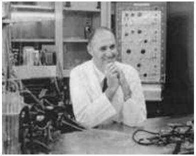


Figure 17–4 *See opposite page for legend.*

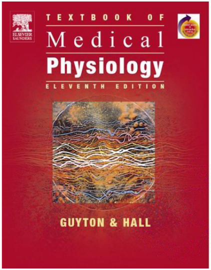
Complex enough for me!

- “large” doesn’t have to be huge to be difficult
- already worthy of structural analysis

Arthur Guyton (1919–2003)



Among the great names of cardiovascular physiology. 42 years as chair of the University of Mississippi Department of Physiology and Biophysics. First edition of his physiology text appeared in 1956.



The Guyton 354-block diagram appears as the Abstract of a 1972 paper ("Circulation: Overall Regulation" in *Ann. Rev. Phys.*). The paper ends with:

"If the general principles of this systems analysis are correct, and we believe they are, then it seems clear that the field of circulatory physiology is on the verge of changing from the realm of a speculative science to that of an engineering science."

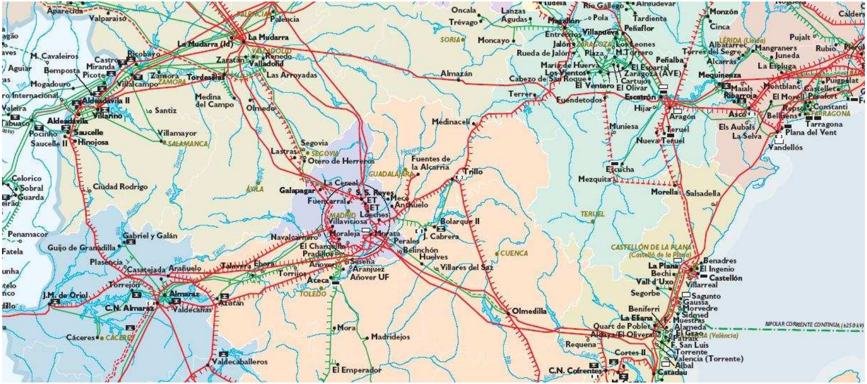
Current efforts in systems physiology

Current efforts in systems physiology

- The Physiome Project (National Simulation Resource at U. of Washington)
  - "The physiome is the quantitative and integrated description of the functional behavior of the physiological state of an individual ... and is built upon information and structure"
  - "... develop and database observations of physiological phenomena and interpret these in terms of mechanism"
- Virtual Physiological Human (EuroPhysiome)
  - "... a way to share observations, to derive predictive hypotheses from them, and to integrate them into a constantly improving understanding of human physiology/pathology, by regarding it as a single system"

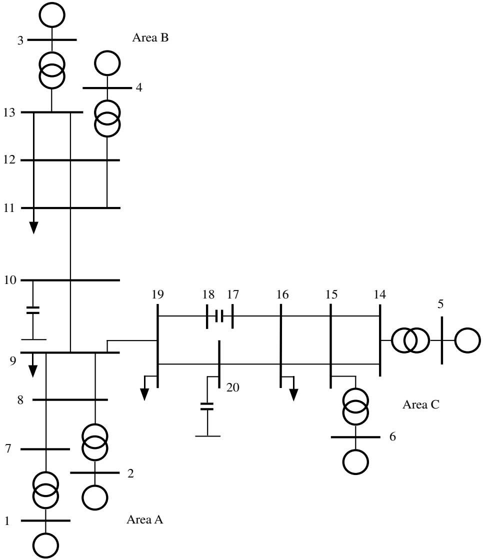
Power systems

Small section of interconnected power system

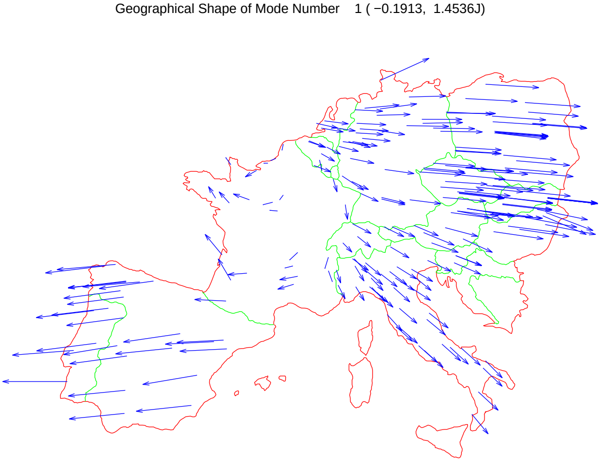


(Red Eléctrica de España)

A small power system model



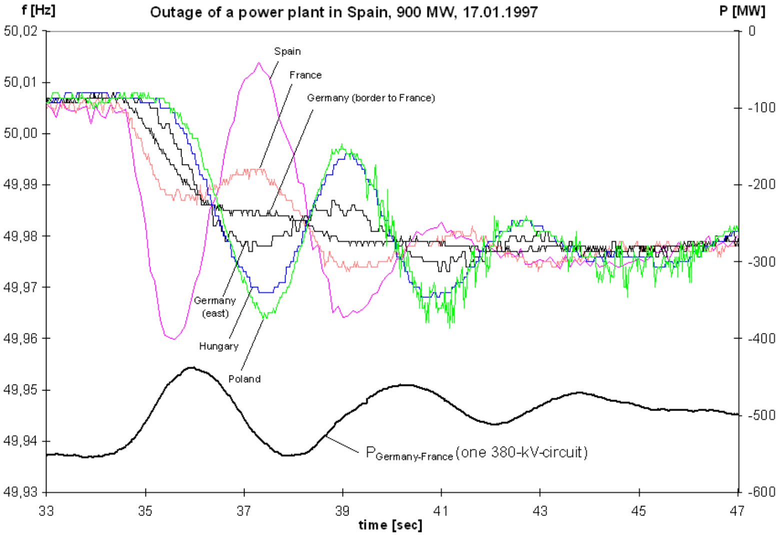
The (swing-mode) dance of a thousand elephants



383 generators; 1,914 buses (nodes); 7,398 state variables; 7,403 algebraic variables

(Luis Rouco, IIT/ICAI, Madrid)

Exciting and recording this swing mode

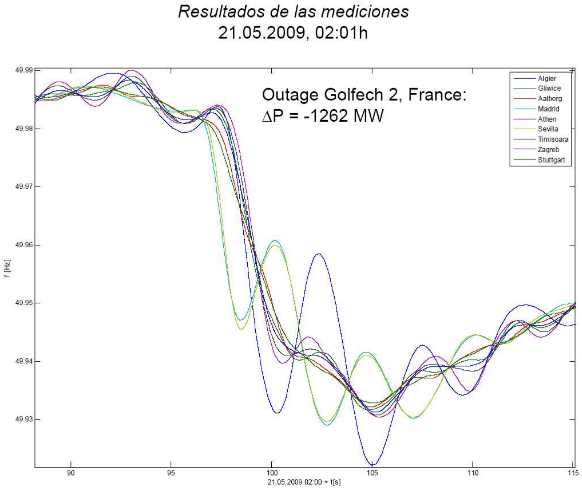


(Juan Manuel Rodriguez, Red Eléctrica de España)

Power system monitoring

(Joachim Lehner, Power Generation and Automatic Control Dept., U. of Stuttgart)

Swing modes, coherence, governor action



(Joachim Lehner, Power Generation and Automatic Control Dept., U. of Stuttgart)

Stepping back ...

Potential value of large-scale gray-box models

Potential value of large-scale gray-box models

- Such models can summarize structural/mechanistic knowledge (and may be the only available starting point)
- allow simulation and exploration with normal and perturbed (pathophysiological) parameters, and for various interventions (valuable in design, training, ...)
- suggest refinements and new experiments

- For hierarchical or multi-scale models
  - constitutive equations and parameters of components at one level can come from analysis of more refined submodels
  - system-level behavior can determine parameters/boundary conditions for submodels

Where large-scale gray-box models fall short

- Difficulty of parameter tuning to match measurements (identification)
  - very many parameters
  - observed signals in any one study are typically not sufficiently rich to reliably estimate all parameters, so parameters are poorly known
  - strategies such as subset selection can help
- Too complicated for the practitioner to use with confidence in real time

Model exploration and reduction

Some questions already for the Guyton model

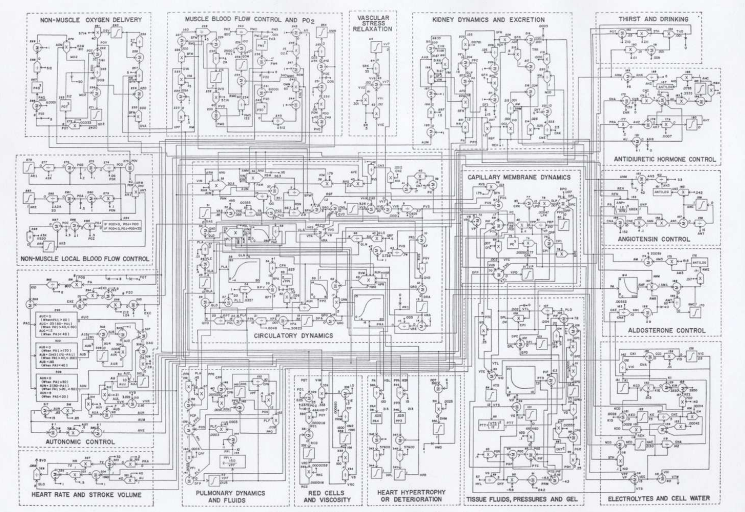


Figure 17–4 *See opposite page for legend.*

Some questions already for the Guyton model

- Is this really modular in some sense (beyond the way it's drawn)? how do we approach this?
- Perhaps there are reasons to expect modularity, from the way the system has evolved?
- Evolution can presumably produce subtle (hard to discover) but beneficial (difficult to ignore) linkages among pre-existing modules

More questions for the Guyton model

Is there a systematic (and at least *semi*-automated) process that reveals the significance of the 8-block structure in the 354-block Guyton diagram?

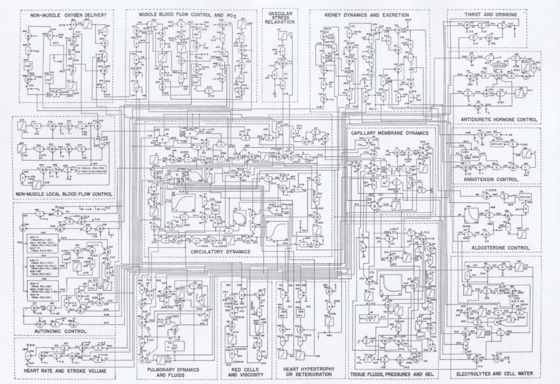
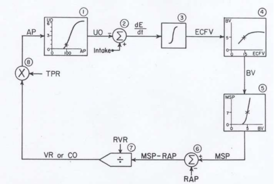


Figure 17–4 *See opposite page for legend.*

$\Rightarrow$



Champion model reducers

Champion model reducers

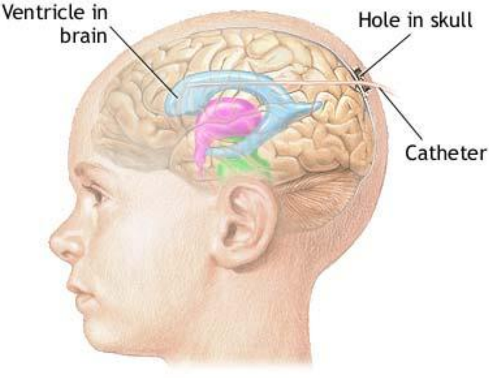
- Doctors!
- having absorbed Gray's *Anatomy of the Human Body*, Guyton's *Medical Physiology*, Harrison's *Principles of Internal Medicine*, pathology, pharmacology, ...
- extracting and applying minimal models in real time, dozens of times a day

- How do they do it?!
- increasingly important to find out
  - aging population and fewer healthcare workers
  - more data being collected, archived
  - standard of care advancing
  - control researchers could do a lot to help

"Integrating data, models and reasoning in critical care"

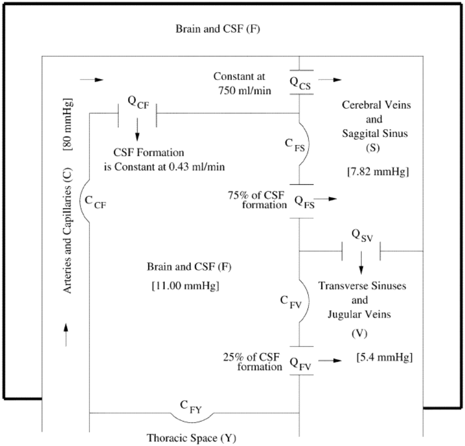
NIH project at MIT (PI: Roger Mark)

Monitoring intra-cranial pressure (ICP)



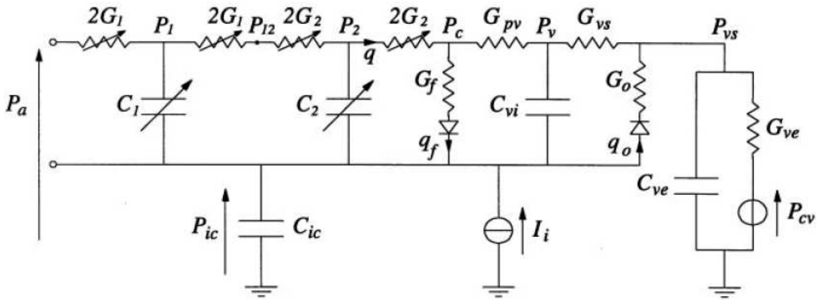
(http://adam.about.com)

Compartmental model of intra-cranial space



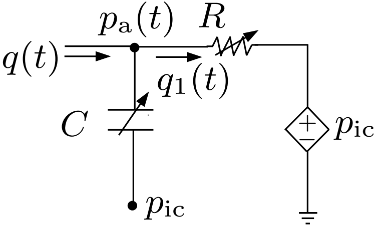
(Stevens *et al.*, *IEEE Trans. Biomed. Eng.*, 2008)

Electrical circuit analog

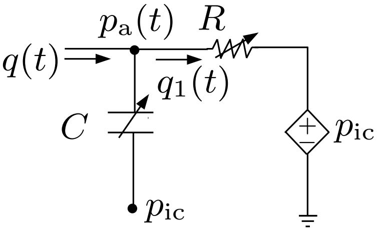


(Ursino and Lodi, *Am. J. Physiol.*, 1998)

Drastically reduced model and noninvasive ICP estimation



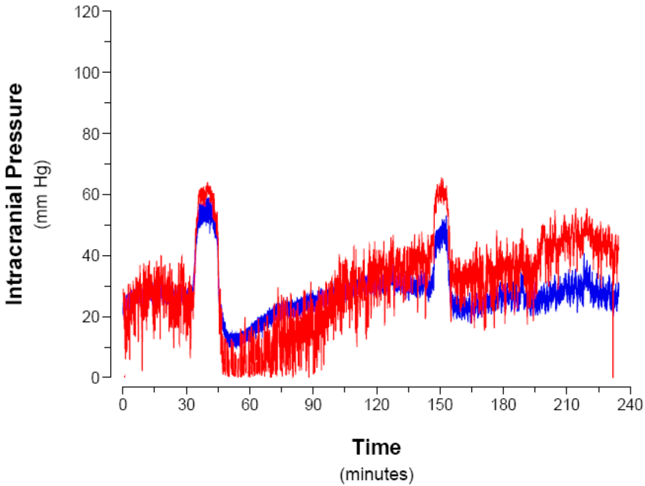
Drastically reduced model and noninvasive ICP estimation



- Key fact: far-end pressure is ICP, not venous pressure (Starling resistor effect)
- Measure flow $q(t)$ with Doppler ultrasound at middle cerebral artery
- Use radial artery pressure as proxy for $p_a(t)$
- Identify parameters, including ICP ($= p_{ic}$)

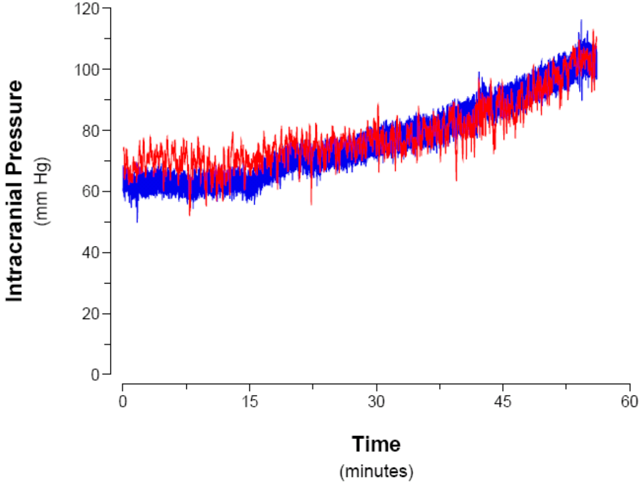
(Kashif *et al.*, *Computers in Cardiology*, 2008)

ICP estimation results



(with F.M. Kashif, T. Heldt, M. Czosnyka, V. Novak)

More ICP estimation results



(with F.M. Kashif, T. Heldt, M. Czosnyka, V. Novak)

Back to more general model reduction

Back to more general model reduction

What do we offer from control theory?

A popular paradigm for model reduction

Model (assume parameter values given):

$$E\,dx(t)/dt = Ax(t) + Bu(t)\ , \qquad y(t) = Cx(t)$$

Projection:

$$x(t) \approx V\widehat{x}(t) \quad \text{and} \quad W^*V = I$$

$$\Downarrow$$

Reduced Model:

$$\begin{aligned}(W^*EV)\,d\widehat{x}(t)/dt &= (W^*AV)\widehat{x}(t) + (W^*B)u(t)\\ \widehat{y}(t) &= (CV)\widehat{x}(t)\end{aligned}$$

Choices for projection matrix V

- **Balanced truncation**: V is determined from dominant eigenspace of $\mathcal{P}\mathcal{O}$, where
  - $\mathcal{P}$ = reachability Gramian
  - $\mathcal{O}$ = observability Gramian

  - stable reduced models from stable parent model
  - explicit bound on $\mathcal{H}_\infty$-norm of input/output error system

- **Proper Orthogonal Decomposition (POD)** focuses on (empirical) reachability Gramian

- **Krylov methods** for (generalized) moment matching of transfer matrix
$$H(s) = C(sE - A)^{-1}B$$
at specified values of s
  - can preserve various desired properties by strategic choice of matching frequencies

Extensions and alternatives

- Direct fitting of low-order model, e.g., in frequency domain
- Optimal Hankel-norm approximation

Extensions and alternatives

- Direct fitting of low-order model, e.g., in frequency domain
- Optimal Hankel-norm approximation

- The preceding methods — including projection — typically don't care about gray-box structure
- nothing to carry back to the large-scale model
- exception: when low-order model in direct fitting is a gray box derived from large model

Recent texts

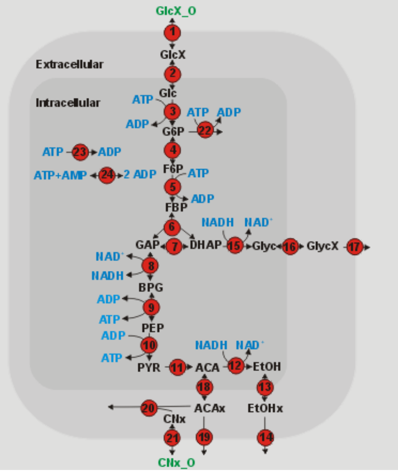
1. *Approximation of Large-Scale Dynamic Systems*
   A.C. Antoulas, SIAM, 2005
   (Detailed and thorough on the topics it focuses on: balanced truncation, Hankel-norm reduction, Krylov subspace methods)

2. *Model Order Reduction: Theory, Research Aspects and Applications*
   W.H.A. Schilders, H.A. van der Vorst, J. Rommes (Eds.), Springer, 2008
   (Predominantly oriented to CAD for circuits, MEMS)

Recent work begins to tackle gray-box issues

- Parameterized model order reduction
  (e.g., work by L. Daniel and collaborators)
- Structure-preserving reduction — preserves block structure
  (several chapters in the preceding text by Schilders *et al.*)

Applying balanced truncation: Glycolysis in yeast



(model from Hynne *et al.*, *Biophys. Chem.* 2001) (from http://jjj.biochem.sun.ac.za/)

Balanced truncation in such a model

Most relevant variable:

$$x_{r,1} = -9.63x_1 - 0.064x_2 - 11.3x_3 - 11.6x_4 - 12.0x_5$$
$$-6.10x_6 + 2.51x_7 - 9.28x_8 - 9.24x_9 - 9.27x_{10} - 9.37x_{11} - 9.36x_{12} - 2.40x_{13}$$

(Bruggeman *et al.* in *Biosimulation in Drug Development*, Bertau *et al.* (Eds.), Wiley 2008)

Balanced truncation in such a model

Most relevant variable:

$$x_{r,1} = -9.63x_1 - 0.064x_2 - 11.3x_3 - 11.6x_4 - 12.0x_5$$
$$-6.10x_6 + 2.51x_7 - 9.28x_8 - 9.24x_9 - 9.27x_{10} - 9.37x_{11} - 9.36x_{12} - 2.40x_{13}$$

- but not clear what to do with this fact!
- authors suggest maybe small coefficient for x_2 is significant
- should something also be made of the fact that several coefficients are close to -9.4? is there any reason the pooled variable $x_1 + x_8 + x_9 + x_{10} + x_{11} + x_{12}$ might be significant?
- though potentially important for model reduction, these are not questions balanced truncation asks

More generally ...

- It's typically not useful to be told some particular linear combination of variables spanning the entire original large model is important (or not important)
- More helpful to know what recognizable (interpretable) variables or components in the large model are critical (or not) to representing the behavior of interest

Don't forget other powerful approximation/reduction approaches ...

1. Averaging or (more generally):
   - tracking (local) means
   - (local) harmonics
   - variances
   - other summarizing functions
2. Singular perturbation (time-scales)
3. Near-decomposability/slow-coherency (spatial scales and time-scales)

All these maintain interpretability in a natural way with respect to the original model. Singular perturbations is widely used (and all these should be taught in model reduction courses!)

Averaging

For quasi-periodic waveforms (as in power systems or power electronics), represent $x(\tau)$ on the interval $(t - T, t]$ by a Fourier series:

$$x(\tau) = \sum X_k(t) e^{j2\pi\tau/T}$$

This is useful when the $X_k(t)$ are slowly-varying Fourier coefficients, or dynamic phasors. They are given by

$$X_k(t) = \frac{1}{T} \int_{t-T}^{t} x(\tau) e^{-jk2\pi\tau/T} \, d\tau$$

Averaging

For quasi-periodic waveforms (as in power systems or power electronics), represent $x(\tau)$ on the interval $(t-T, t]$ by a Fourier series:

$$x(\tau) = \sum X_k(t) e^{j2\pi\tau/T}$$

This is useful when the $X_k(t)$ are slowly-varying Fourier coefficients, or dynamic phasors. They are given by

$$X_k(t) = \frac{1}{T} \int_{t-T}^{t} x(\tau) e^{-jk2\pi\tau/T} \, d\tau$$

- The dynamic equations for $x(t)$ are then approximated by dynamic equations for slowly-varying low-order phasors
- Used quite routinely now in power systems and power electronics

(Sanders *et al.*, *IEEE Trans. Power Electronics*, 1991)

Tracking means and variances in chemical reaction systems

- The (forward Kolmogorov) Chemical Master Equation tracks the evolution of the probability distribution of species molecule numbers in systems of chemical reactions
- Important for small (subcellular) volumes, but is generally of unmanageable size

e.g., complex formation and dissociation: $X_1 + X_2 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X_3.$

- Mass action kinetics (MAK) tracks mean concentrations, and yields

$$\frac{d\mu_3}{dt} = k_1\mu_1\mu_2 - k_2\mu_3$$

Mass Fluctuation Kinetics (MFK)

- Mass fluctuation kinetics (MFK) jointly tracks the means and variances/covariances, yielding

$$\frac{d\mu_3}{dt} = k_1\mu_1\mu_2 - k_2\mu_3 + k_1\sigma^2$$

  with a corresponding equation for $\frac{d\sigma^2}{dt}$
- More generally,

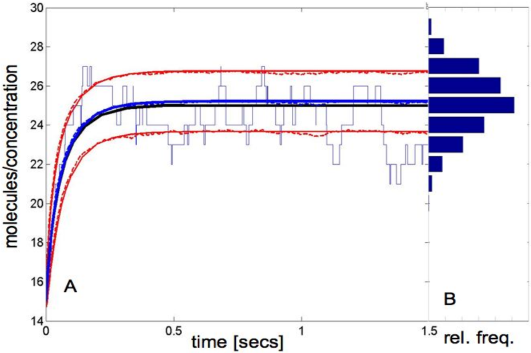
$$\begin{aligned} \frac{d\mu}{dt} &= \mathbf{Sr} = \mathbf{S}\rho + \mathbf{S}\xi \\ \frac{d\mathbf{V}}{dt} &= \mathbf{MV} + \mathbf{VM}' + \frac{1}{\Omega}\mathbf{S\Lambda S}' \end{aligned}$$

(Singh & Hespanha, 2006; Gómez-Uribe & Verghese, 2007; Goutsias, 2007)

Comparisons

MAK and MFK give two different steady-state expressions for the expected concentration:

$$\mu_3 = \frac{k_1}{k_2}\mu_1\mu_2, \text{ and } \mu_3 = \frac{k_1}{k_2}(\mu_1\mu_2 + \sigma^2).$$



Near-decomposability/Slow-coherency

$$\dot{x}(t) = Ax(t) = (A_0 + \epsilon K)x(t)$$

with

$$A_0 = \text{block diag}\{A_i\} \qquad \text{for } \; i = 1, \cdots, N$$

and with A as well as each A_i for $i = 1, \cdots, N$ having a single eigenvalue at 0, with associated eigenvector $[1\ 1\ 1 \cdots 1]^T$.

Near-decomposability/Slow-coherency

$$\dot{x}(t) = Ax(t) = (A_0 + \epsilon K)x(t)$$

with

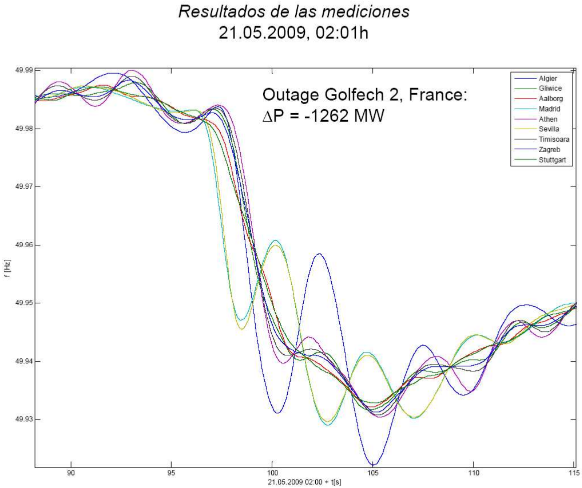
$$A_0 = \text{block diag}\{A_i\} \qquad \text{for } \; i = 1, \cdots, N$$

and with A as well as each A_i for $i = 1, \cdots, N$ having a single eigenvalue at 0, with associated eigenvector $[1\ 1\ 1 \cdots 1]^T$.

- The zero modes of the individual subsystems get globally coupled through the ϵ terms, but the other modes are not much affected
- The system modes accordingly comprise fast $O(1)$ local modes and slow $O(\epsilon)$ global modes
- When only the slow modes are excited, the state variables in each of the N areas move essentially in unison (coherence), i.e., the areas swing against each other.

Simon & Ando, 1961; Ando & Fisher, 1963; Courtois, 1977; Chow & Kokotovic (slow-coherency), 1982, 1985

Slow coherency



(Joachim Lehner, Power Generation and Automatic Control Dept., U. of Stuttgart)

Selective Modal Analysis (SMA)

Notation

- General response of $\dot{x}(t) = Ax(t)$ is of the form

$$x(t) = \alpha_1 v_1 e^{\lambda_1 t} + \cdots + \alpha_N v_N e^{\lambda_N t}$$

where the v_i are the *right* eigenvectors of A, with associated eigenvalues λ_i, so

$$Av_i = v_i \lambda_i$$

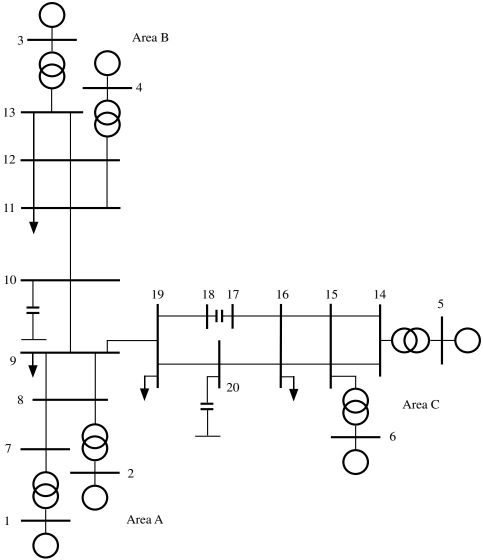
Here $v_i e^{\lambda_i t}$ is termed the ith *mode*

- One can similarly define *left* eigenvectors w_i^T by

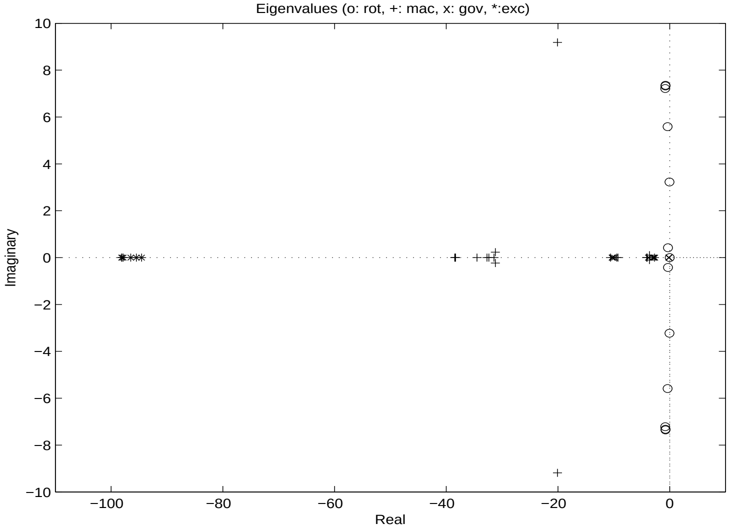
$$w_i^T A = \lambda_i w_i^T$$

normalized such that $w_i^T v_j = 1$ when $i = j$, and 0 otherwise

A small power system model



… and the eigenvalues of the linearized model



Participation factors

- A non-dimensional measure of the participation of the jth state variable in the ith mode is given by

  the jth component of the right eigenvector v_i associated with this mode,

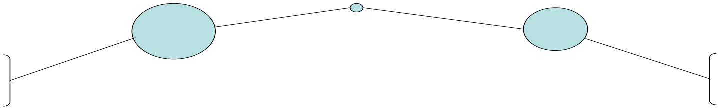
  weighted by the jth component of the left eigenvector w_i^T :

  $p_{ji} \quad = \quad w_{ji}v_{ji}$ or equivalently $P = transpose(inv(V)).*V$

- Participation factors are insensitive to choice of units for the state variables
- Our normalization of the left eigenvectors ensures $\sum_i p_{ji} = 1$, and $\sum_j p_{ji} = 1$.
- Note similarities with Relative Gain Array (but also key differences)

(with Pérez-Arriaga & Schweppe, *IEEE Trans. Power Apparatus & Systems*, 1982; later work by E. Abed *et al.*)

Participations in slow mode of simple mass-spring system



- The peak displacements are approximately the same for all masses (similarly for peak velocities), so the right eigenvector alone will not show any significant differentiation among the three
- The participation factors are proportional to the peak *energies* of the masses (and hence vary approximately as the masses in this case)
- Only the two large masses need to be retained in a reduced model that captures an approximation of this mode

Dynamic patterns in an 'Infinite Bus' power system model

Participation factors		
	-2.7382+j9.2529	-0.2390+j6.7618
δ	+0.0208-j0.0098	**+0.4931+j0.0108**
ω	+0.0208-j0.0098	**+0.4931+j0.0108**
ψ_{fd}	-0.0075-j0.0238	-0.0180-j0.0119
ψ_{kd1}	+0.0036-j0.0023	-0.0118-j0.0191
ψ_{kq1}	+0.0000+j0.0026	-0.0041+j0.0101
ψ_{kq2}	-0.0022-j0.0004	-0.0086+j0.0001
x_{ex1}	**+0.4771+j0.1262**	+0.0283+j0.0062
x_{ex2}	**+0.4652-j0.1362**	+0.0318-j0.0156
x_{ex3}	+0.0222+j0.0535	-0.0037+j0.0085

Conclusions: The lightly damped oscillatory mode is primarily electromechanical, involving perturbations of rotor angle and velocity, δ and ω respectively; the other mode is due to the exciter

Subsystem participation, dynamic patterns

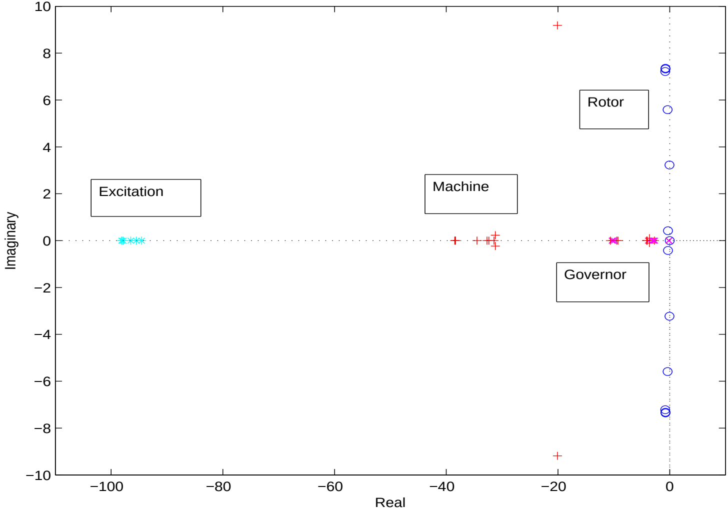
- Participation of the subsystem S in the ith mode:

$$p_{Si} \;=\; \sum_{j \in S} p_{ji} = w_{Si}^T v_{Si}$$

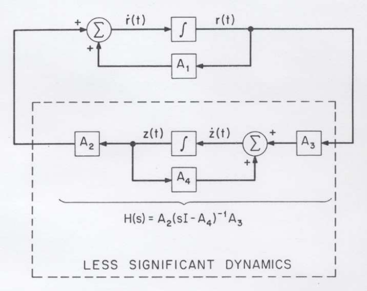
  The subsystem participation is invariant under transformations that only affect the state variables x_S of the subsystem

- Dynamic patterns, i.e., close relationships between subsets of modes and subsets of state variables, can be identified by using the participation factors and the subsystem participations

For small power system model ...



Separating relevant and less relevant variables for model reduction



Less relevant dynamics can be approximated by

- zero-frequency gain $-A_2A_4^{-1}A_3$ to give reduced model $\dot{x}(t) = (A_1 - A_2A_4^{-1}A_3)x(t)$ (familiar from singular perturbation)
- generalizations: $H(\widehat{\lambda}) = A_2(\widehat{\lambda}I - A_4)^{-1}A_3$ for some $\widehat{\lambda}$; multi-mode possibilities

A simple algorithm for reduced-order eigenanalysis

- A first approximation of the mode of interest ${}^{0}\lambda_1$ is captured by

$$\dot{r} = A_1 r, \qquad A_1 \in \Re^{n\times n}$$

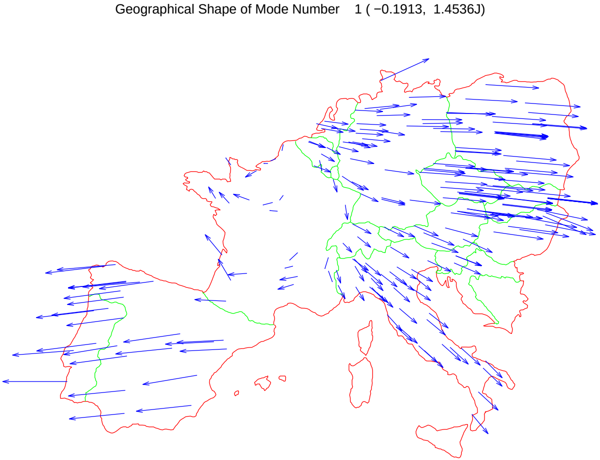
- Selective Modal Analysis successively refines such approximations of the eigenvalue by incorporating the effect of the less relevant dynamics on that mode
- The SMA reduced-order system is:

$$\dot{r} = (A_1 + H({}^{j}\lambda_1))r$$

- Eigenanalysis of $A_1 + H({}^{j}\lambda_1)$ yields, among the eigenvalues, a choice for ${}^{j+1}\lambda_1$, and the iteration continues
- Converges locally precisely when summed participations of r variables exceed summed participations of remaining z variables; ratio determines convergence rate
- Multi-mode extensions available

(with Pérez-Arriaga, Schweppe, Pagola, Rouco, ..., 1981 and later)

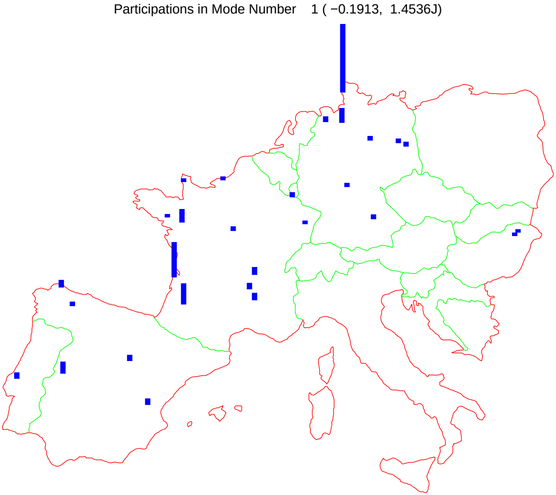
Slowest swing mode



383 generators; 1,914 buses (nodes); 7,398 state variables; 7,403 algebraic variables

(Luis Rouco, IIT/ICAI, Madrid)

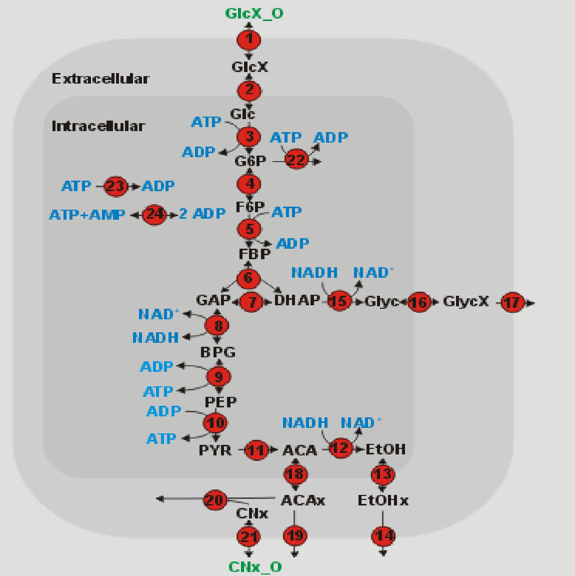
Slow-mode participations



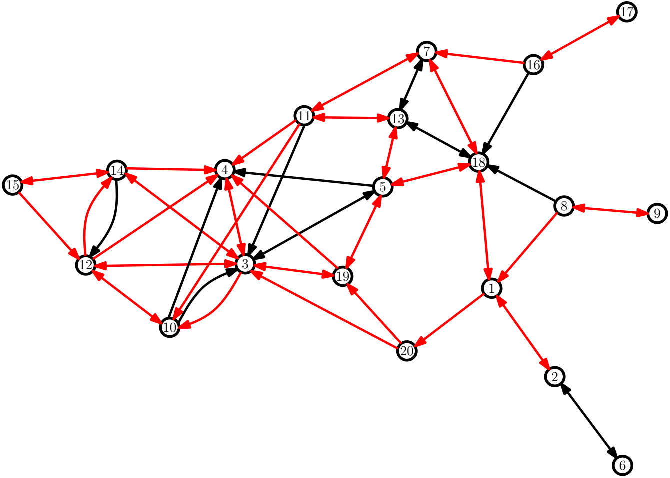
383 generators; 1,914 buses (nodes); 7,398 state variables; 7,403 algebraic variables

(Luis Rouco, IIT/ICAI, Madrid)

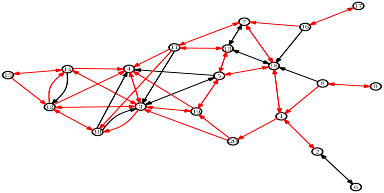
Glycolysis


GlcX_O
GlcX
Extracellular
Glc
Intracellular
ATP
ADP
G6P
ATP ADP
ATP
ADP
ATP+AMP
2 ADP
F6P
ATP
ADP
FBP
NADH NAD+
GAP
DHAP
Glyc
GlycX
NAD+
NADH
BPG
ADP
ATP
PEP
ADP
ATP
PYR
ACA
NADH NAD+
EtOH
ACAx
EtOHx
CNx
CNx_O

Structure of linearized glycolysis model



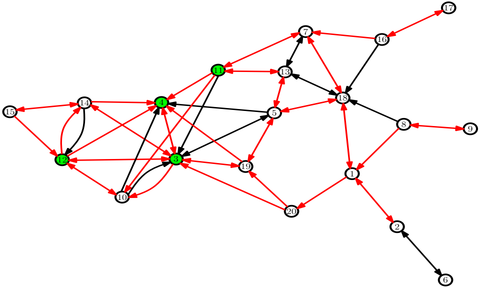
Glycolytic oscillation



- For operating point corresponding to glycolytic oscillation, what small subset of variables suffices to capture the oscillation?
- One approach: exhaustive search, starting from 2×2 models and going up, to find smallest truncated model that is oscillatory
- Almost 60,500 eigenanalyses later, arrive at 6th-order model

(Danø *et al.*, *FEBS Journal*, 2006)

Nodes with high participation in oscillatory mode



- Participation factors computed from a single eigenanalysis of the full (20×20) matrix quickly point to 4 significant variables that capture the oscillation (in a "residualized" model)
- Time-scale decomposition allows further reduction to a 3rd-order model

(with G. Cedersund, U. of Linköping)

Conclusions

- Worthwhile to attend to care and feeding of gray-box models (alongside care and feeding of algorithms or computations)
- Model reduction should pay attention to model too, not just reduction
- Don't settle for having a model handed to you over a wall! — control people can be of great help to the modelers (control folks working in systems biology know this well)
- Consider getting involved in the "other" systems biology, i.e., systems biology of organs and organisms — physiology
- Try out participation factors on linearized model for your favorite physical application — you may find new insights!

Thank you!