Compact Nonlinear Thermal Modeling of Packaged Microprocessors

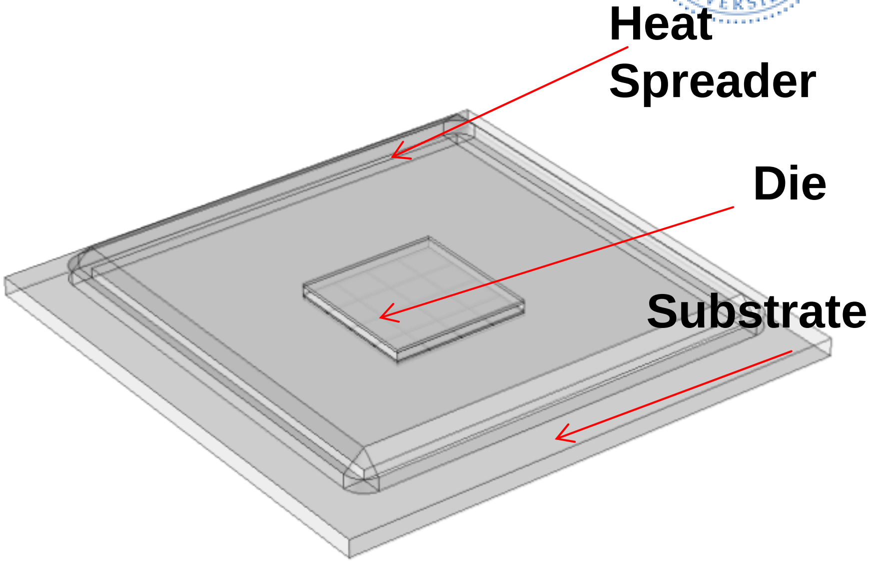
Zao Liu, Sheldon Tan,
Hai Wang, Yingbo Hua
Department of Electrical Engineering
University of California, Riverside, CA
Ashish Gupta
Intel Corporation
Chandler, AZ

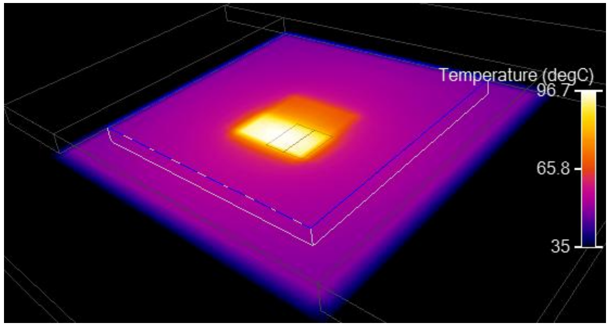
Content

- Introduction to thermal modeling
- Problem of subspace-based thermal modeling
- Proposed method
- Experimental result
- Conclusion

Thermal modeling of packaged microprocessor

- Temperature has become a major concern for high performance microprocessors
- Even for severe for multi/many core anc emerging 3D stacked systems
  - Longer thermal paths
  - Loaded dependent hoptspots
  - Large thermal gradients and dynamic thermal effect related reliability issues.
- Compact thermal model at package levels is vital for efficient thermal aware design and management.
  - Enable thermal-aware design flow
  - Enable accurate online thermal management and regulation





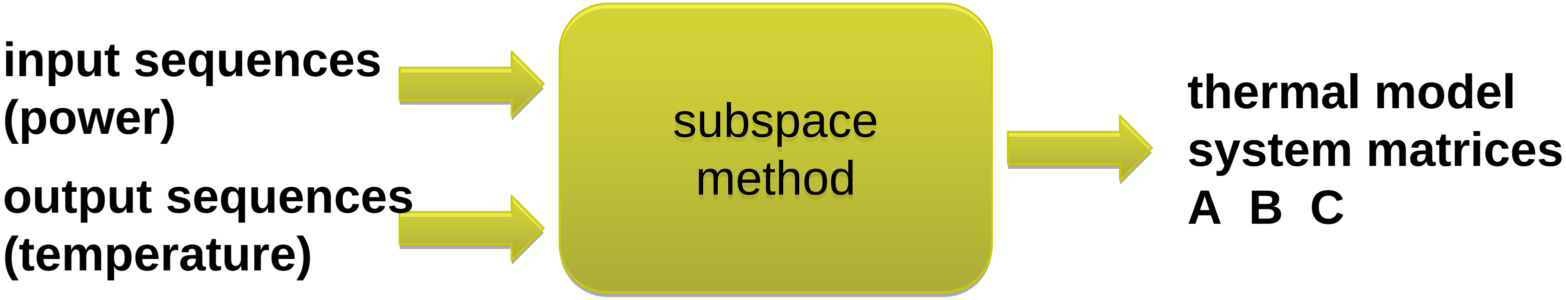
Simulated Temp distribution using Cu sink (390 W/m K)

Bottom-up thermal modeling methods

- FDM (finite difference)
- FEM (finite element) [Lasance:SEMITHERM'95, Christiaens:TCPMT'98]
  - Limitation:
    - Knowledge of detailed thermal structures is not easy to obtain
    - Impractical for large scale circuits
- HotSpot [Huang:DAC'04, Skadron:ISCA'03]
  - Mainly for architectural level design exploration
  - Limitation:
    - Accuracy losses in lumped model

Behavioral thermal modeling method

- Matrix pencil or subspace methods [Overschee:book'06] [Eguia and Tan: TVLSI'10]
  - Obtains thermal model through input/output information
  - No need for detailed thermal structures



input sequences (power)

output sequences (temperature)

subspace method

thermal model system matrices
A B C

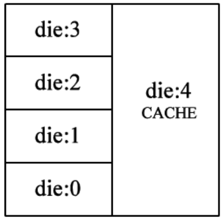
- Compact model suffers from accuracy losses due to non-linearity of the practical thermal system

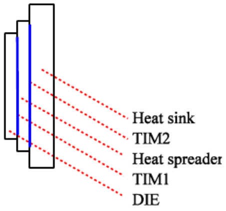
Content

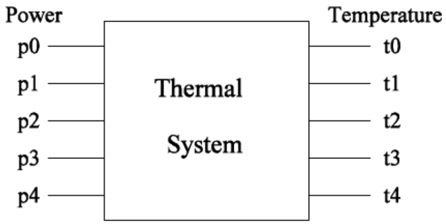
- Introduction to thermal modeling
- Problem of subspace-based thermal modeling
- Proposed method
- Experimental result
- Conclusion

Thermal modeling for given power density under correlated power inputs

- Current:

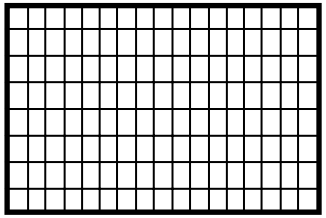


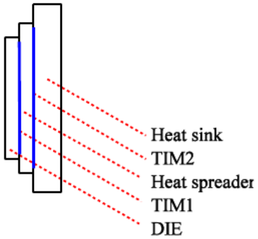


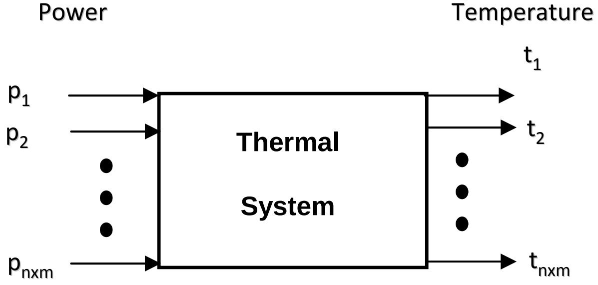


- Proposed:

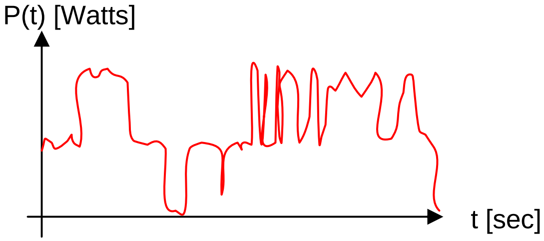
nxm sections



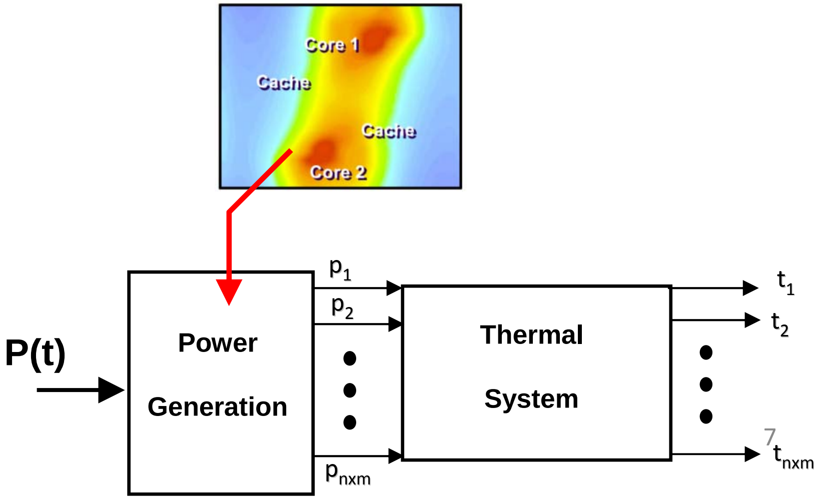




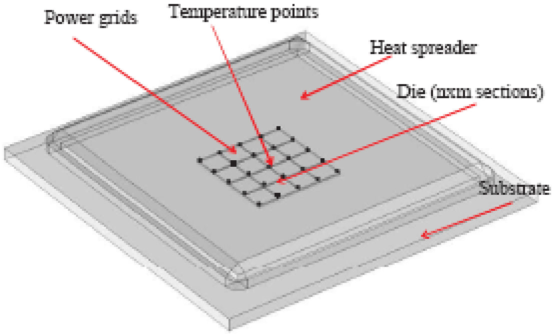
p_1, p_2, …, p_{nxm} generated from a power map and are highly correlated & overall processor power P(t)



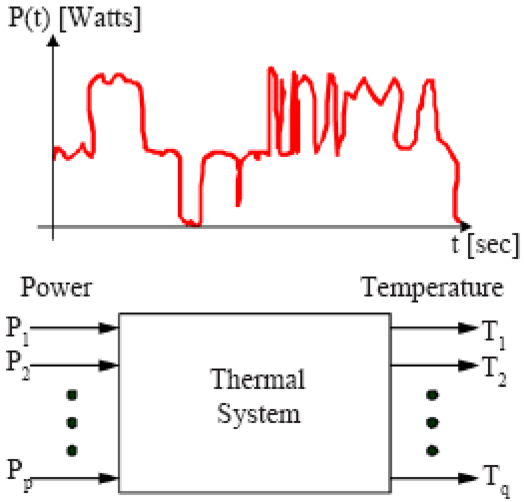
  - Desired : Stand Alone Application
    - Inputs: P(t), Power Map, Identification method (POF or Subspace ID), t_1, t_2, …, t_{nxm}
    - Output: H(s) → nxm transfer function matrix



State space model of thermal system



chip partition



The abstracted model with power as input

The linear model of the thermal system can be described by state space equation:

$$x(t+1) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t)$$

Given input $u(t)$ and output $y(t)$, the state matrices A, B, C can be identified by subspace method.

Subspace based modeling flow

Input/Ouput Data
$u(t), y(t)$

QR

Weighted
Hankel Matrix
HR^T

SVD

$HR^T=U\,\Sigma V^T$

Reduced SVD
$HR^T=U_r\,\Sigma_r V_r^T$

Extended Observability
/Controlibility Matrix
$O(M)=U_r\,\Sigma_r^{1/2}$
$C(N)=\ \Sigma_r^{1/2}V_r^T(R^T)^{-1}$

C = *first top block row of* $O(M)$
$[F(t)\ B^u]$=*first left block column of* $C(N)$
$A=O_{\downarrow}(M)^{\dagger}O^{\uparrow}(M)$

$$W=\begin{bmatrix}U_{future} & W_{past} & Y_{future}\end{bmatrix}$$
$$=\begin{bmatrix}Q_1 & Q_2 & Q_3\end{bmatrix}\begin{bmatrix}R_{11} & R_{12} & R_{13}\\ 0 & R_{22} & R_{23}\\ 0 & 0 & R_{33}\end{bmatrix}$$

R_{23}^T is the weighted Hankel matrix HR^T

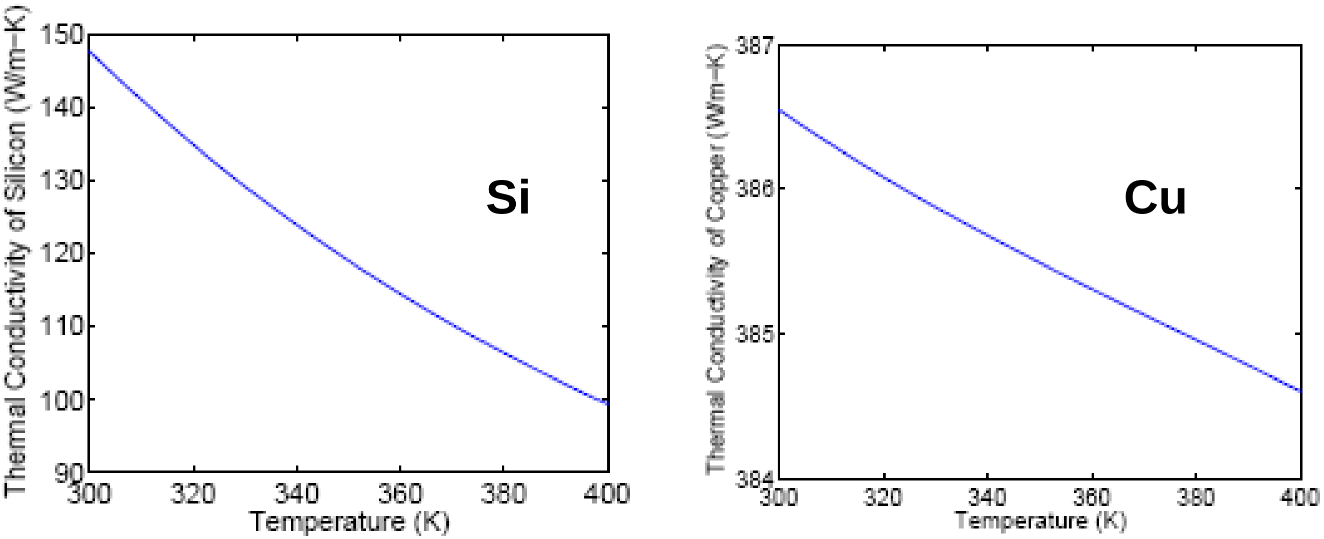
R_{22} is the weight matrix R

$$H(N,M)=\begin{bmatrix}h_1(N,1) & h_1(N,2) & \dots & h_1(N,N)\\ h_2(N,1) & h_2(N,2) & \dots & h_2(N,N)\\ \vdots & \vdots & \ddots & \vdots\\ h_M(N,1) & h_M(N,2) & \dots & h_M(N,N)\end{bmatrix}$$
$$=\begin{bmatrix}C\\ CA\\ \vdots\\ CA^{M-1}\end{bmatrix}\begin{bmatrix}B & SB & \dots & S^{N-1}B\end{bmatrix}=O(M)C(N)$$

$$O^{\uparrow}(M)=\begin{bmatrix}CA\\ CA^2\\ \vdots\\ CA^{M-1}\end{bmatrix},\ O_{\downarrow}(M)=\begin{bmatrix}C\\ CA\\ \vdots\\ CA^{M-2}\end{bmatrix}$$

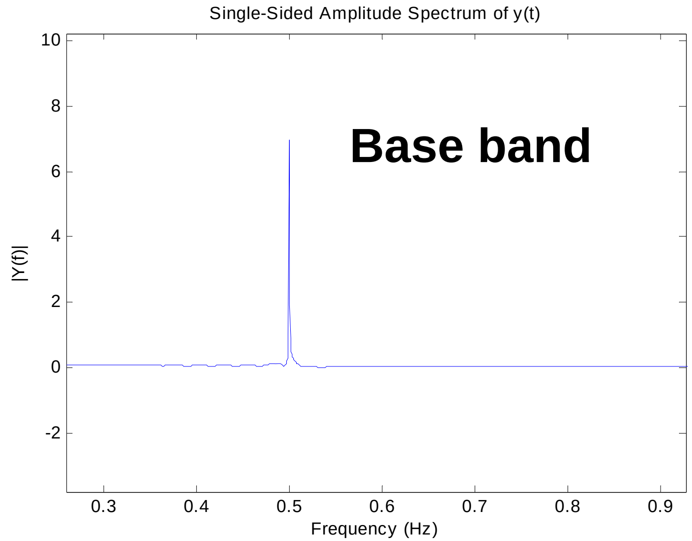
Thermal systems are actual nonlinear!

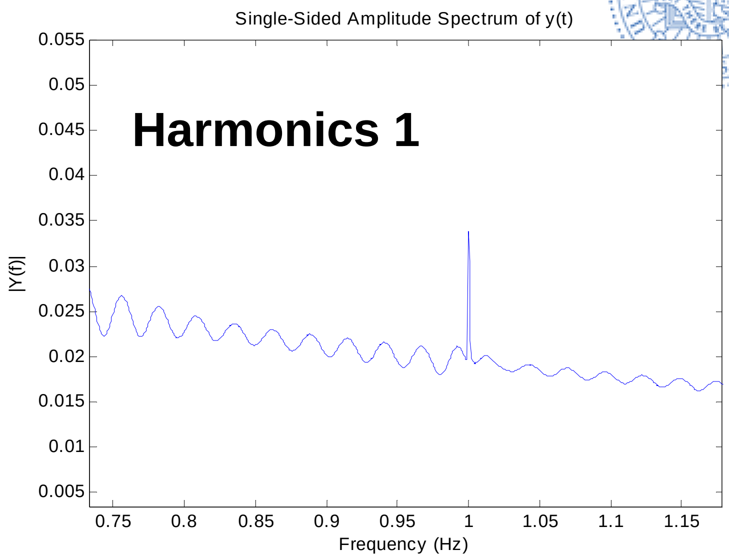
- Nonlinearity is caused by the temperature dependent properties of the package materials.

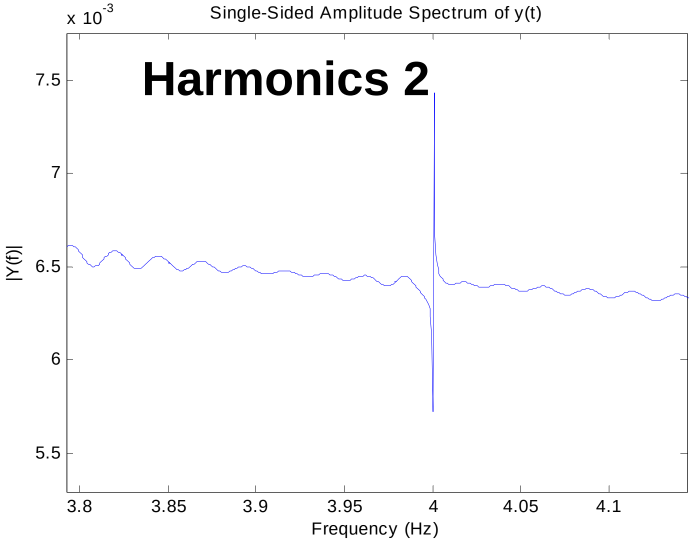


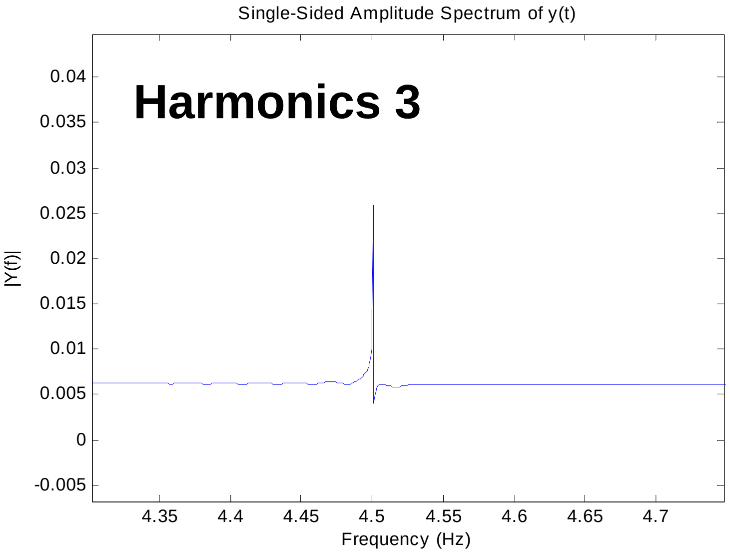
Example: Temperature dependence of thermal conductivity

Harmonics of the thermal response





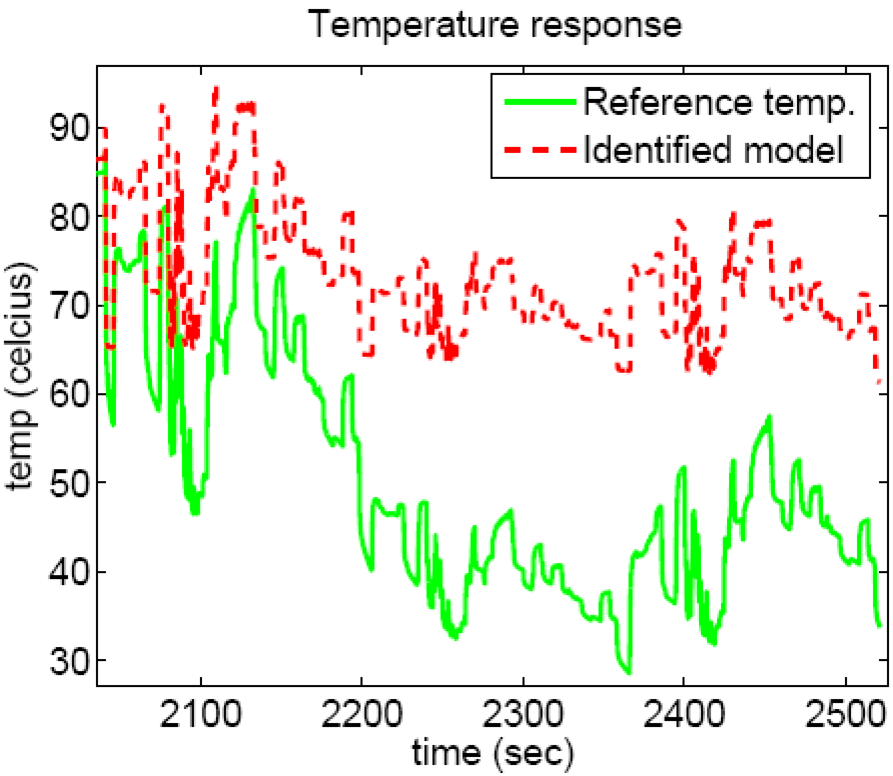




- The output frequencies (The input signal is Pin=Psin(πt), t=0:0.1:1599.9)

Accuracy loss of linear model

- Nonlinearity results in accuracy losses of the compact linear state space model (order=4).



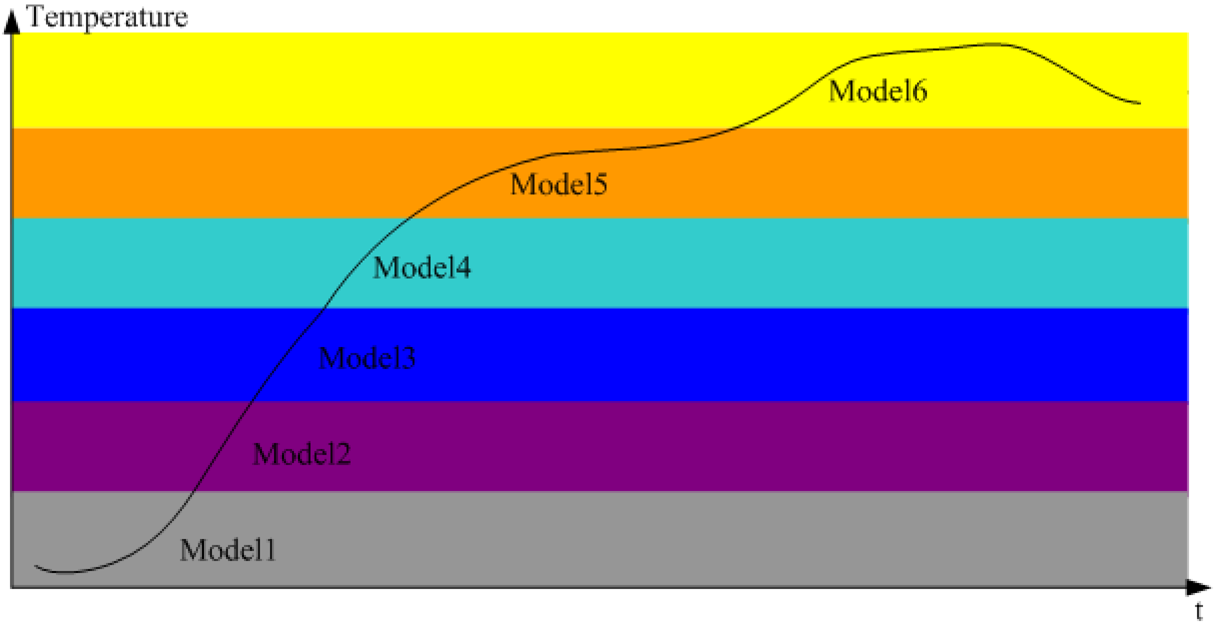
Temperature output of the identified model (linear)

Content

- Introduction to thermal modeling
- Problem of subspace-based thermal modeling
- Proposed method
- Experimental result
- Conclusion

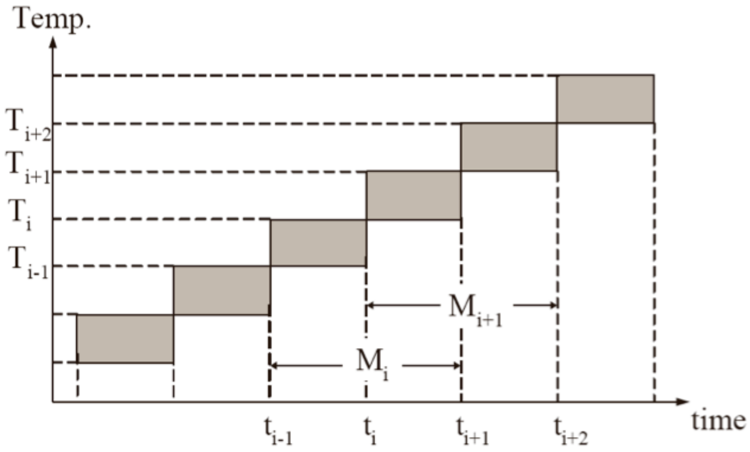
Proposed PWL modeling method

- Use piecewise linear model to approximate the thermal response of the chip for different temperature ranges

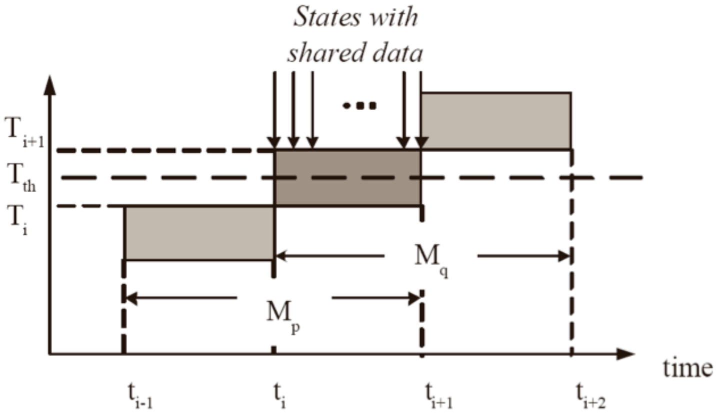


Outline of the proposed method

- Preparing training data sets for model identification in different temperature ranges
- Improved subspace method is used to identify the sub-models for each temperature range
- Linear transformation is used to build the piecewise linear model

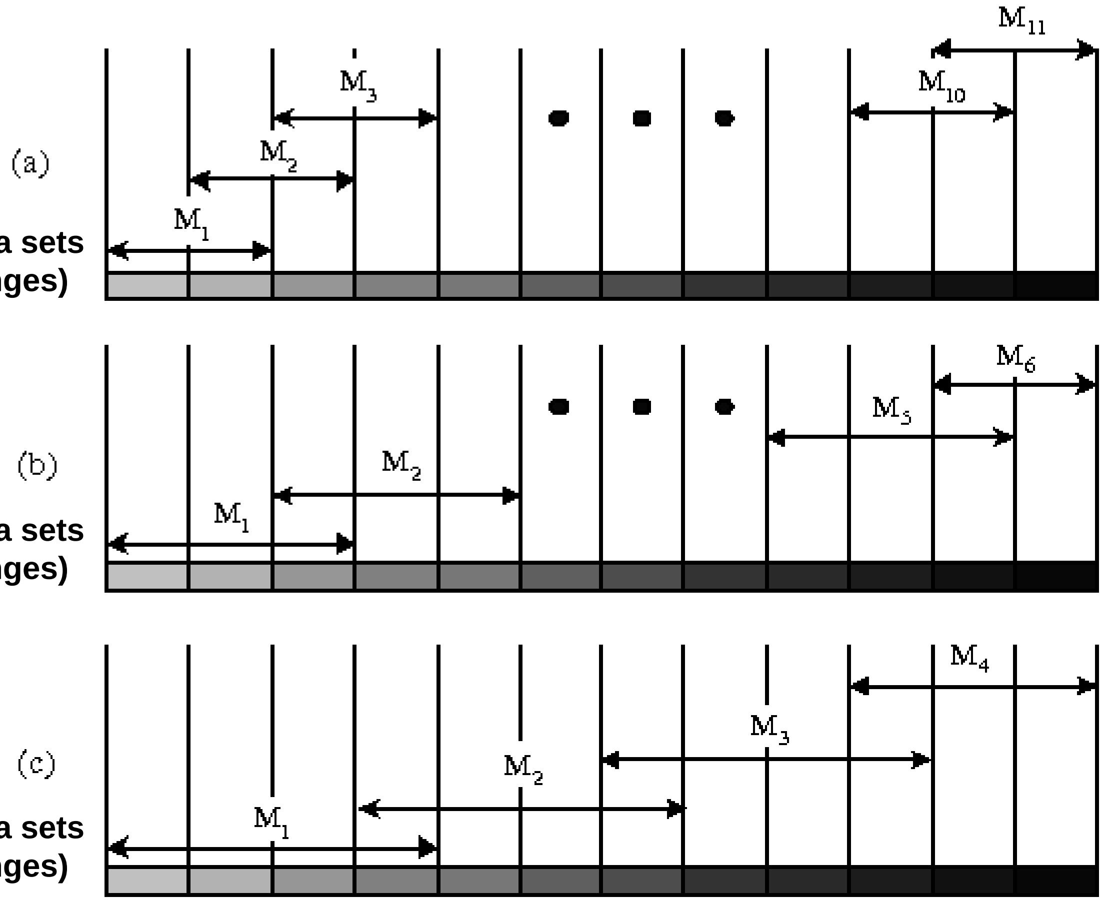


Identification of different models at different temp ranges



Modeling transition from M_p to M_q

Data partition for model identification



12 Data sets
(12 ranges)

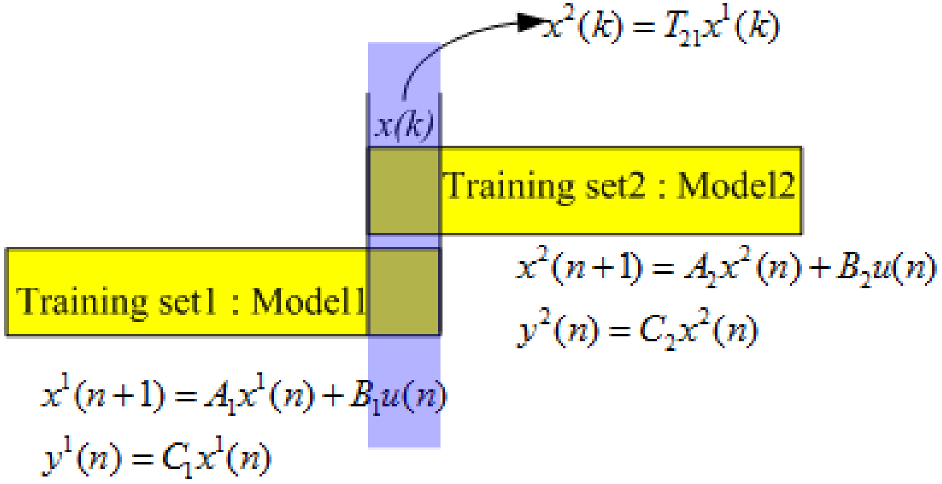
12 Data sets
(12 ranges)

12 Data sets
(12 ranges)

Data partition scheme to build piecewise linear model (a) use 11 sub-models (b) use 6 sub-models (c) use 4 sub-models

Build PWL model from sub-models

- To build PWL model, it is necessary to convert the state vectors of different sub-models to the same basis
- The identified sub-models are not on the same basis.
- At the transition region, the states in Model1 and Model2 differ by a linear transformation $\boldsymbol{T_{21}}$: $x^2(k_1:k_N)=T_{21}x^1(k_1:k_N)$
- $\boldsymbol{T_{21}}$ could be determined using least square method.
- $x^2(k)=T_{21}x^1(k)$ transforms Model2 to the basis of Model1

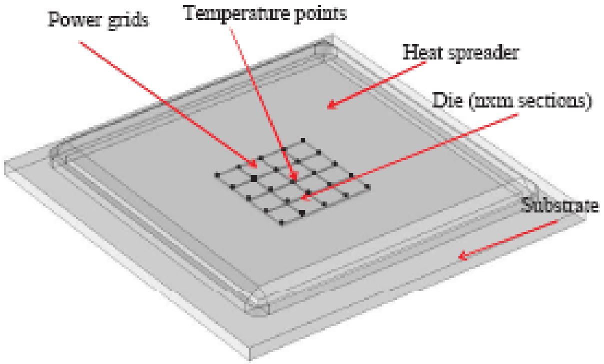


Content

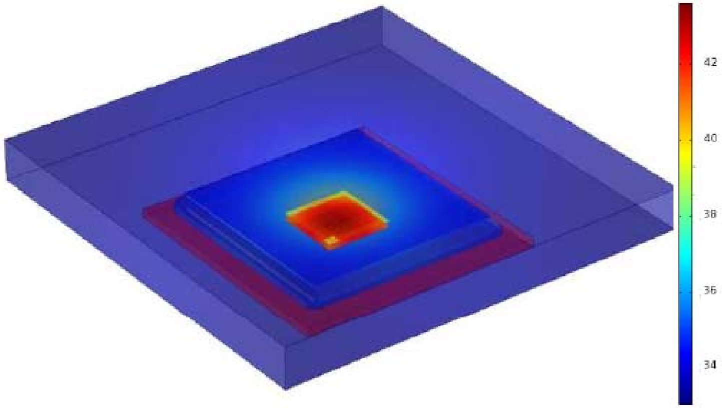
- Introduction to thermal modeling
- Problem of subspace-based thermal modeling
- Proposed method
- Experimental result
- Conclusion

Experimental setup

- Treat the meshed thermal chip package as a 16-input (power) and 25-output (temperature) system.
- Use COMSOL to simulate its transient temperature response to obtain the temperature data for piecewise linear system identification
- Full simulation time steps: 20412

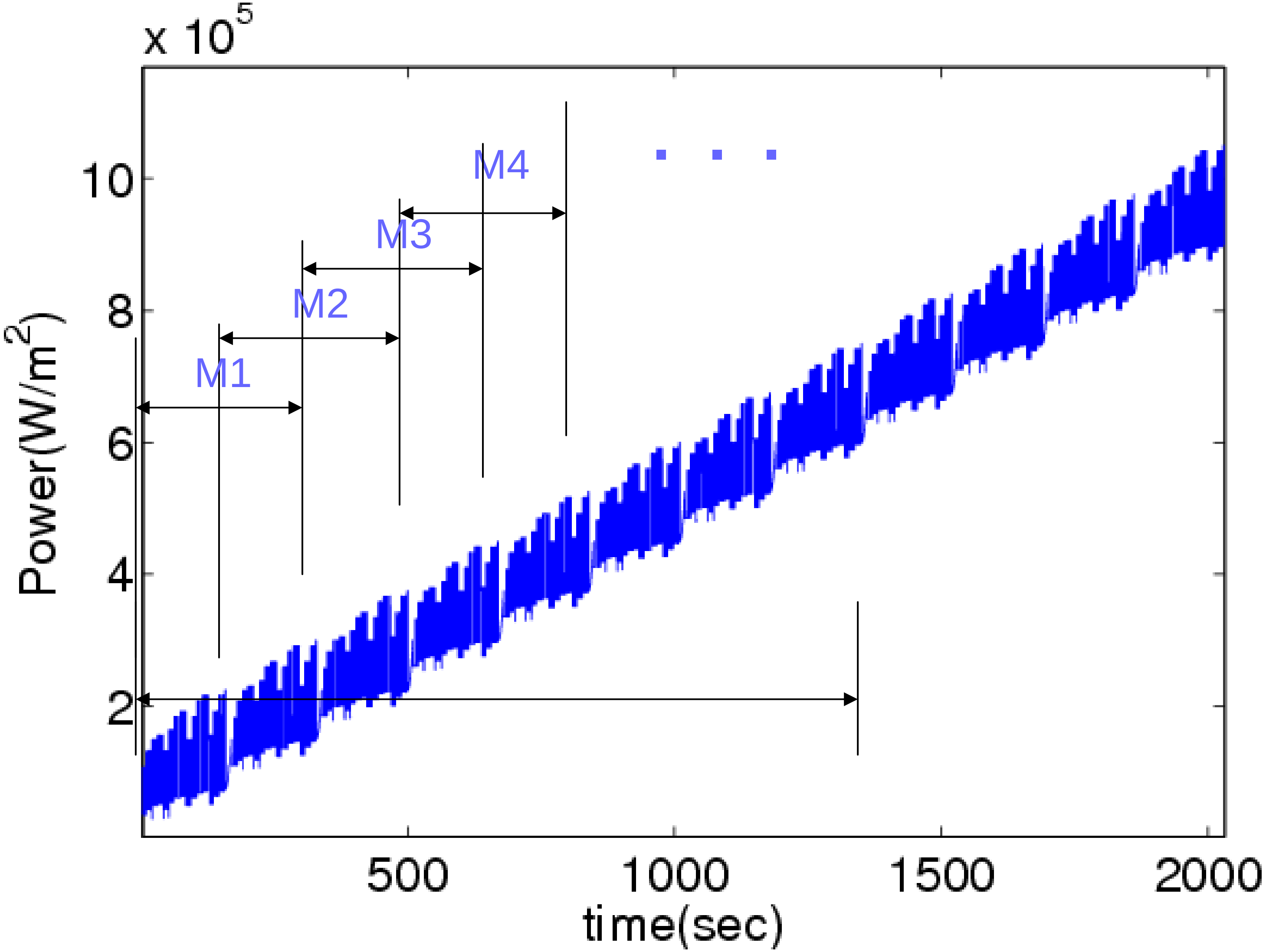


Chip partition



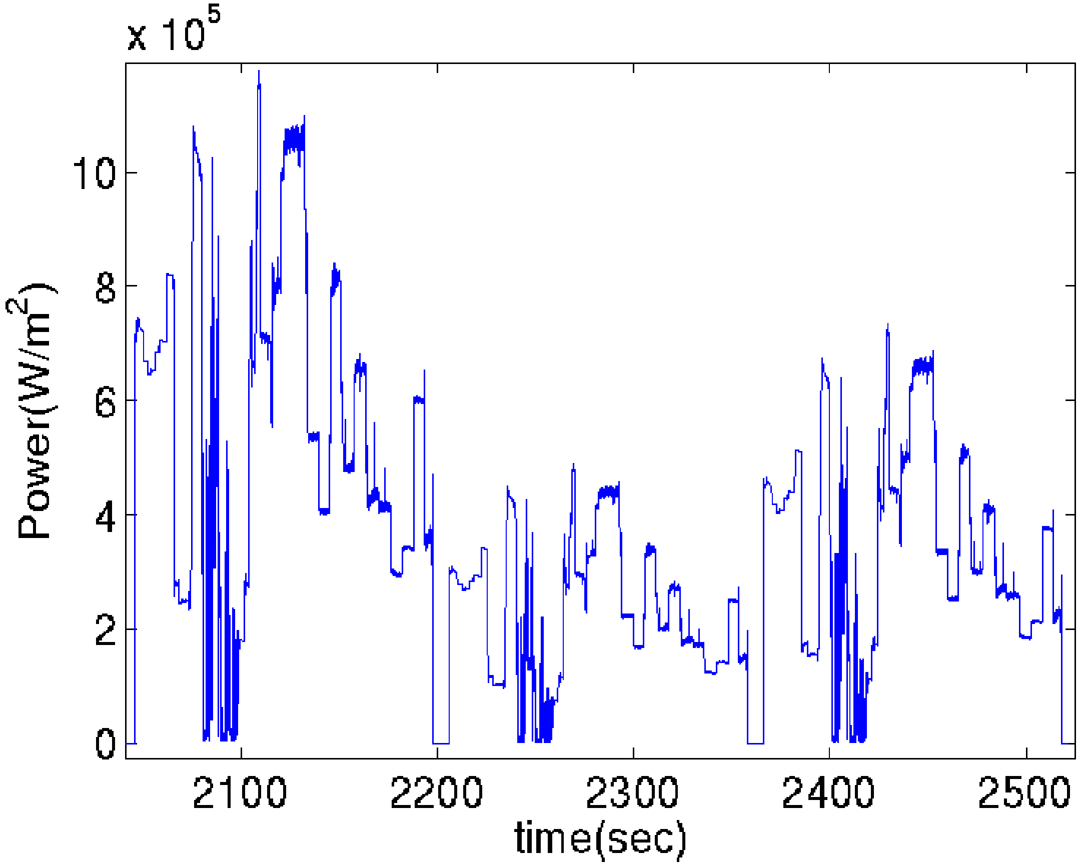
Steady state temperature distribution

Input power waveform (I)



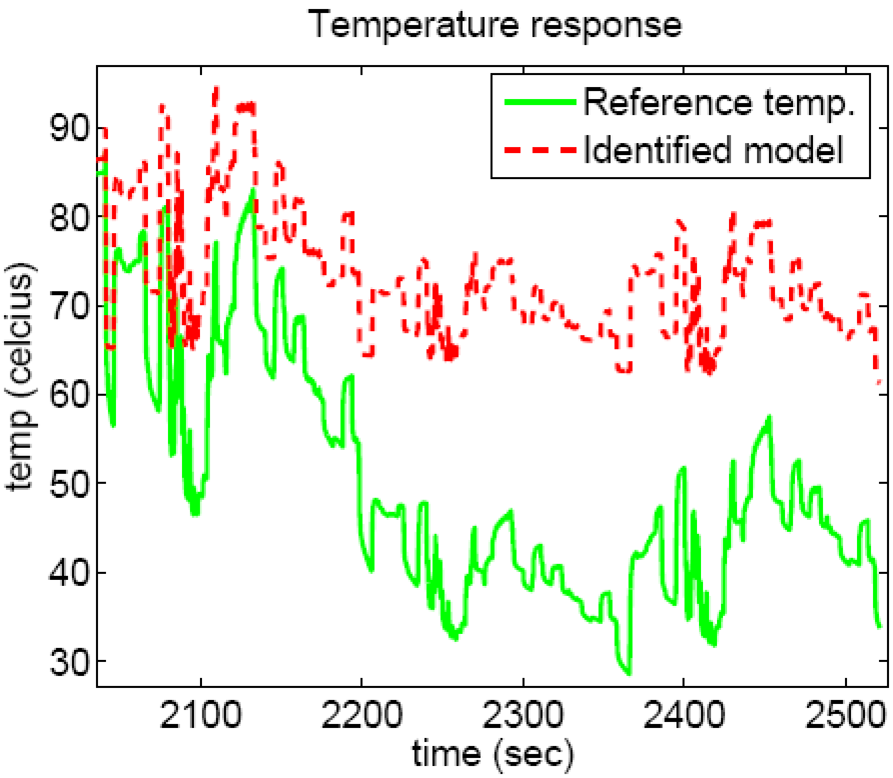
Input waveform used for model identification (PRBS signals with stairs-like envelop)

Input power waveform (II)

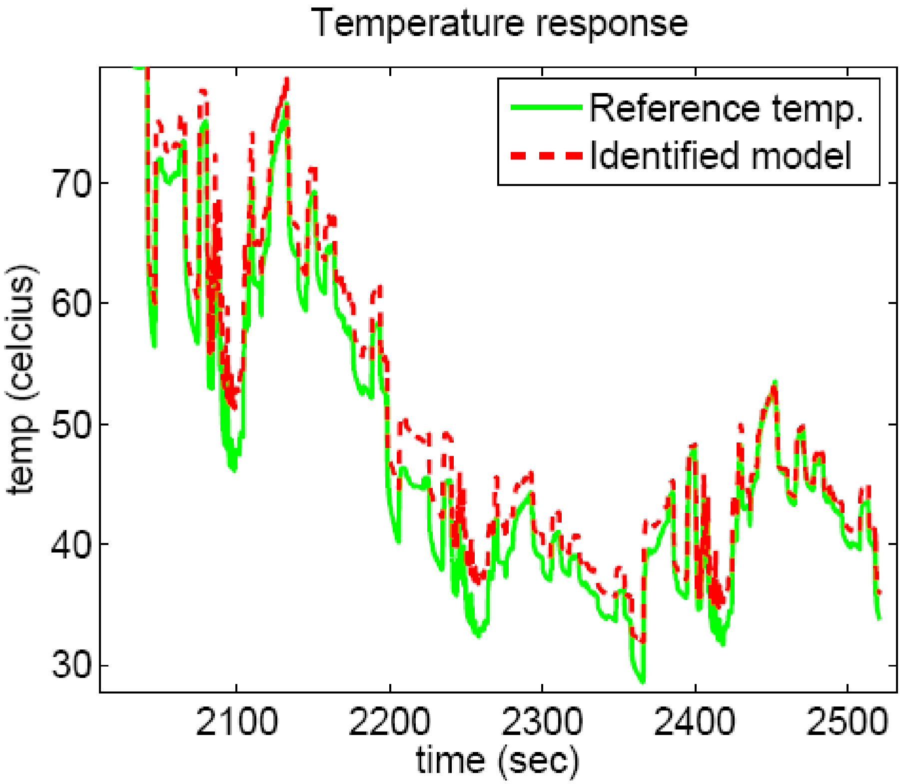


Input waveform used for model validation (Intel’s signal)

Transient response of 16-input and 25 output system at section(1,1)

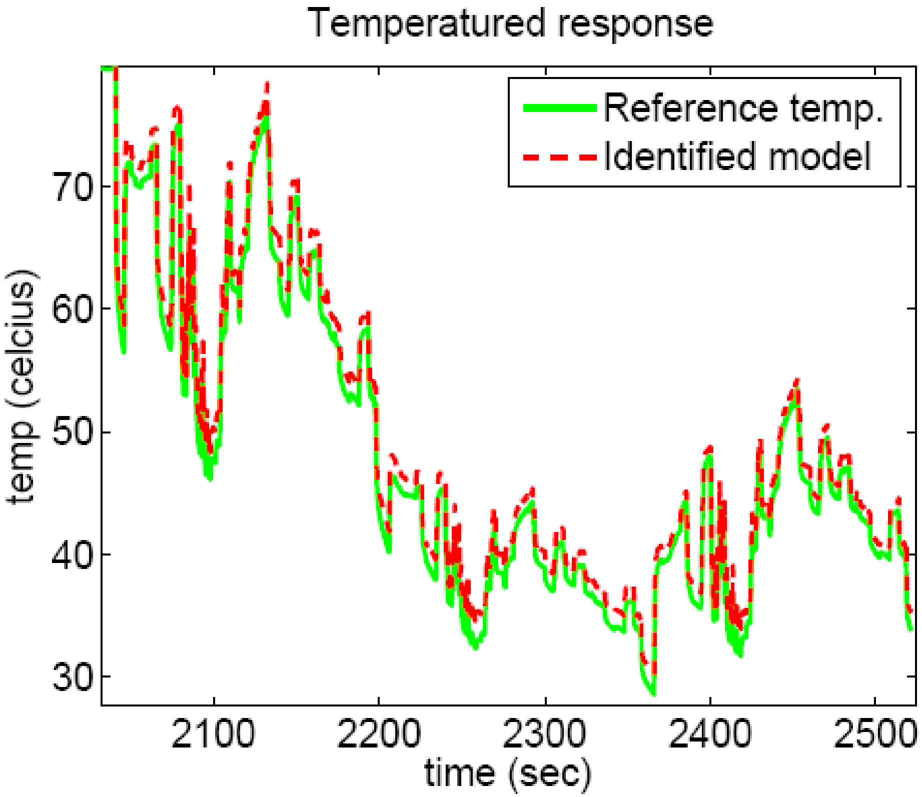


Use 1 linear model:
Order = 4

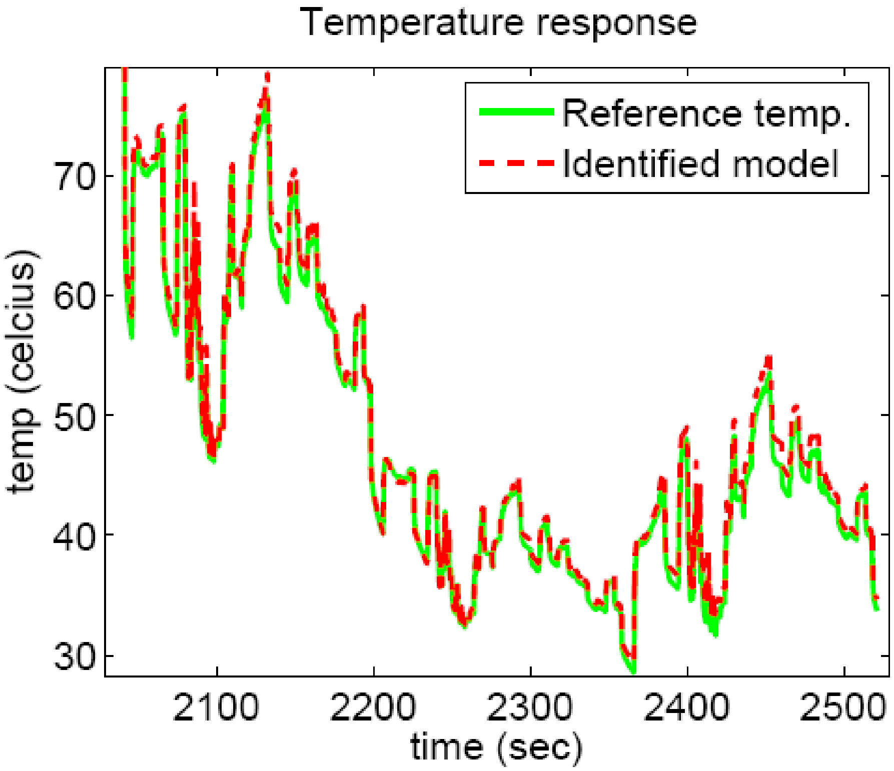


Use piecewise linear model:
Number of sub-model used: 4
Order =4

Transient response of 16-input and 25 output system at section(1,1)



Number of sub-model used: 6
Order = 4



Use piecewise linear model:
Number of sub-model used: 11
Order =4

Summary of errors of 16-temperature output of the piecewise linear models

Error with PWL models (order:4)

Num. of linear models in use	11	6	4
Maximum of mean errors	2.1%	3.9%	5.9%

Mean errors are calculated during the entire transient simulation

Performance comparison with linear model

Comparison items	**Error**	**Identification time**	**Simulation time**
PLM(order:4)	2.1%	63.8 sec	7.88 sec
LM (order:15)	2.3%	627.1 sec	22.2 sec

PLM – Piecewise linear model
LM – Linear model

Conclusion

- Piecewise linear model scheme has been proposed to consider nonlinear effects in thermal systems.
- Linear sub-models are identified for different temperature ranges using subspace identification method.
- A linear transformation method has been proposed to build piecewise linear model.
- Our experiment results show that Piecewise linear model is more efficient for fast thermal modeling and simulation of packaged microprocessor.