Cyber-Physical Management for Heterogeneously Integrated 3D Thousand-core On-chip Microprocessor

Sai Manoj P.D and Hao Yu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
haoyu@ntu.edu.sg

Abstract—Though 3D TSV/TSI technology provides the promising platform for heterogeneous system integration with design drivers ranged from thousand-core microprocessor to millimeter-cubic sensor, the fundamental challenge is lack of light to deal with significantly increased design complexity. From device level, new state of variables from different physical domains such as MEMS, microfluidic and NVM devices have to be identified and described together with conventional states from CMOS VLSI; and from system level, cyber management of states of voltage-level and temperature has to be maintained under a real-time demand response fashion. Moreover, a cyber-physical link is required to compress and virtualize device level state details during system level state control. This paper shows device-level 3D integration by example of MEMS and CMOS VLSI. In addition, a cyber-physical thermal management for 3D integrated many-core microprocessors is discussed.

I. INTRODUCTION

With the increasing demand of cloud computing for big-data, design of high-throughput data servers has obtained recent interest significantly. The big-data processing at exascale is obviously beyond traditional single-core or multi-core microprocessors. Many-core microprocessors with thousand-core become the emerging need with many recent explorations [1], [2], [3]. The primary challenges come from the low bandwidth and high power density in 2D integration. Moreover, such a complicated computing system requires new means of states identification, reduction and management.

3D integration applies vertical stacking of layers one above other by through-silicon-via (TSV) or through-silicon-interposer (TSI). As such, the communication bandwidth can be improved with small interconnection latency. Moreover, as the loss of I/O is reduced with more data transferred, the communication power can be reduced as well. The other advantage of 3D comes from heterogeneous integration, i.e., devices made from different technologies such as nano-scale non-volatile memory (NVM), MEMS, and even microfluidic [4], [5], [6], [7]. Thereby, one can build a smart cubic microsystem with multiple functionalities. For example, Fig.1 shows one possible 3D integrated thousand-core-on-chip microprocessor with TSV based I/O to connect structured memory and core blocks such as data-bus or clock. The NVM device is considered here to replace the main memory by DRAM. In addition, microfluid works as active cooling channel to dissipate heat [7].

The primary limitations for 3D integrations, especially with applications in thousand-core on-chip, are as follows. Firstly, one needs to identify new physical-domain states. The new nano-scale NVM device such as spin-based STT-RAM may show dynamics not determined by traditional electrical voltages or currents, but by magnetization angles or doping density [6]. Moreover, a reliable utilization of TSV/TSI needs a multiple physical-domain model to characterize cross-coupled electrical-thermal-mechanical delay. Secondly, one needs to reduce the number of states as too many timing violation, power integrity and thermal reliability to check layer by layer. The essential state extraction by macromodeling is required to virtualize the system complexity [4]. Lastly, one needs to perform smart state management for power and thermal. For example, the problem is different now when providing the power supply from many power converters with many voltage levels to thousand cores. Moreover, the long heat dissipation path may require integrated active cooling scheme [7] or new power gating strategy [5].

In this paper, we discuss potential challenges and solutions to build 3D thousand-core system for big-data cloud computing. We show a heterogeneously integrated 3D thousand-core on-chip microprocessor design from perspective of cyber-physical management. Section 2 explains the overall architecture of 3D thousand-core microprocessor and the need for cyber-physical management. Physical modeling in terms of state identification for NVM devices and TSV/TSI is explained in Section 3. Section 4 illustrates how macromodels is formulated for complexity reduction. Section 5 explains cyber system management for 3D thousand-core microprocessor with adaptive flow-rate cooling and power gating. Conclusions are drawn in Section 6.

II. HETEROGENEOUS 3D THOUSAND-CORE SYSTEM

One heterogeneously integrated 3D thousand-core system is shown in Fig.1 for big-data cloud computing. Firstly, many-core microprocessors are organized in a network-on-chip mesh, where core and core communicate by routers. The main memory is designed using NVM devices. Each core visit its local block memory by TSV or TSI with I/O links. Digital power and temperature sensors are realized on-chip to monitor real-time power and thermal profiles, which provide feedback to system to control the power and temperature. For example, one can adjust the flow-rate of microfluid according to the temperature gradient profile.

In this paper, we show a design methodology from cyber-physical perspective to realize such a heterogeneously integrated 3D thousand-core system. Firstly, building a physical model to consider new physical-domain states introduced from non-traditional devices such as nano-scale NVM devices is shown followed by building a physical-model by considering multiple physical-domain states for TSV or TSI delay under coupling from thermal temperature and mechanical stress. Next, with the use of structured and parameterized macro-modeling, we show how to extract the essential states to reduce complicated physical model of 3D thousand-core system. Lastly, by the use of macromodels in a close-feedback-loop with prediction,
managing the system states such as power and thermal in a cyber-physical fashion is shown.

III. PHYSICAL DEVICE MODELING

The heterogeneous integration of different technologies from different physical domains results in the challenges in physical modeling, to identify new physical-domain states or multiple-physical states. In this section, we first discuss physical modeling for the nano-scale NVM devices such as STT-RAM [6] by identifying the new physical-domain state, and also show physical modeling of TSV with cross-coupled delay model from electrical, thermal and mechanical domains.

A. New Physical-domain State

Traditional electric devices are mainly described by modified nodal analysis (MNA) with nodal voltages and branch currents \( (V_m, j_b) \). For nano-scale NVM devices, there are new states to be determined. For example, we need to know magnetization angle to fully describe the dynamics of STT-RAM. Moreover, doping ratio is needed for memristor and crystallization rate for PCM [6]. At the same time, one needs to develop a new MNA state description for both CMOS and NVM devices.

As shown by Fig. 2, we add new branch currents associated with new NVM devices, which are described by introducing new state variables \( s_m \), determine the conductance of all NVM devices. Note that incident matrix for capacitor, resistor, inductor and current source are denoted by \( E_m, E_s, E_l, E_v \), and additional state variables, \( s_m \) for NVM device are linked by incident matrix \( E_m \).

Considering a STT-RAM device, which has two sandwiched ferromagnetic layers and oxide layer in between [6], it needs a new state variable \( \theta \), angle of magnetization between two magnetic layers to describe giant-magnetoresistive (GMR). As such, GMR becomes

\[
R(\theta) = R_L + \frac{R_H - R_L}{2} (1 - \cos(\theta))
\]

\[= R_L + \frac{\Delta R_{GMR}}{2} (1 - \cos(\theta)).\]

The new state vector becomes \( X = [v_n, j_l, j_m, \theta_m]^T \) instead of \( X = [v_n, j_l, j_m]^T \) [6]. One new MNA [6] can be derived correspondingly to fully describe the dynamics of such a hybrid NVM and CMOS system.

B. Multiple Physical-domain State

TSV or TSI is the essential component to integrate NVM memory and microprocessor core in the proposed 3D thousand-core system. As global data-bus or clock with relevant I/Os between memory and core blocks, TSV or TSI can be fabricated in a structured fashion reliably. At the same time, unlike 2D, TSV becomes the path for heat dissipation and also stress passing. All multi-physical domain effects can have significant impact on the electrical delay of TSVs. The electrical model of TSV is thereby not accurate if no thermal and mechanical behavior are considered.

1) TSV Delay with Temperature: TSVs are generally surround-
ed by a liner material (SiO\(_2\) or Si\(_3\)N\(_4\)), of very small radius to avoid diffusion of metal atoms into silicon substrate and to provide isolation. The TSV structure with isolation, for example, in a 3D clock-tree [8], is shown in Fig. 3.

2) TSV Delay with Stress: Due to isolation-layer, there is charging and discharging when signal passes TSV. The work in [8] shows that TSV works as a non-linear capacitor with the following equivalent electrical model

\[
\frac{1}{C_1} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}} \text{Rs} = \frac{\rho h}{2\pi r_{metal}}.
\]

Fig. 4: Typical C-V curve of TSV MOSCAP with non-linear temperature dependence

Fig. 3: (a) Signal and dummy TSV in 3D IC; (b) 3D view of TSV and (c) Equivalent circuit of TSV

Due to isolation-layer, there is charging and discharging when signal passes TSV. The work in [8] shows that TSV works as a non-linear capacitor with the following equivalent electrical model.

\[
\frac{1}{C_1} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}} \text{Rs} = \frac{\rho h}{2\pi r_{metal}}.
\]

Here, \( C_{ox} \) and \( C_{dep} \) are linear capacitance and depletion capacitance of TSV respectively, with \( C_{ox} = \frac{2\pi h \epsilon_{si}}{\ln(\frac{h}{r_{metal}})} \), and \( C_{dep} = \frac{2\pi \epsilon_{si} h}{\ln(\frac{h}{r_{dep}})} \)

Note that \( \rho \) is the resistivity of the TSV metal and \( h \) is height of TSV. \( \epsilon_{si} \) and \( \epsilon_{ox} \) are dielectric constants of silicon and silicon oxide; and \( r_{metal}, r_{ox} \) and \( r_{dep} \) are the outer radius of TSV metal, silicon and depletion regions. As \( r_{dep} \) depends on temperature surrounding TSV, it results in non-linear TSV capacitance with temperature and hence has non-negligible impact on delay.

A typical C-V curve for TSV with liner is shown in Fig. 4, which can be divided into three regions, based on variation of capacitance, separated by flat band (\( V_{FB} \)) and threshold voltage (\( V_t \)). Based on this electrical-thermal coupled model, one can derive the Elmore delay model when using TSV as link between memory and core [8].
TABLE I: Reduction in clock-skew by TSV induced mechanical stress

<table>
<thead>
<tr>
<th>Type</th>
<th>Orig (ps)</th>
<th>Lin (ps)</th>
<th>Redu (ps)</th>
<th>Runtime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4</td>
<td>2.18</td>
<td>1.37</td>
<td>0.87</td>
<td>20.76</td>
</tr>
<tr>
<td>T8</td>
<td>5.77</td>
<td>1.54</td>
<td>4.23</td>
<td>24.27</td>
</tr>
<tr>
<td>T20</td>
<td>10.55</td>
<td>5.2</td>
<td>5.33</td>
<td>19.51</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>5.26</td>
<td>18.20</td>
</tr>
</tbody>
</table>

substrate by changing the lattice structure. The impact of stress ($\sigma$) on mobility ($\mu$) can be described by

$$\frac{\Delta \mu}{\mu} = -\pi \sigma$$

where $\pi$ is piezo-electric constant.

Mechanical stress from TSVs, it can be utilized in a positive manner to reduce the delay on the chip. Insertion of dummy TSVs reduces the on-chip temperature and also the skew. The impact of mechanical stress on clock-skew for a 3D clock-tree design implemented in [8] using IBM benchmark r1, consisting of 45 signal bundles containing 2, 4, 8 and 10 TSVs respectively. Orig is the clock-skew without implementing any optimization technique for insertion of TSVs. Lin and Redu represents the clock-skew after insertion of TSV by linear optimization and reduction in clock-skew after insertion of TSVs by linear optimization respectively. As shown by Table 1, insertion of dummy TSV by linear optimization reduces the stress-induced clock-skew by 45.58% on average.

3) Coupled TSV Model: The coupled electrical-thermal-mechanical TSV delay thereby needs to be addressed during the design for data-bus or clock I/Os. One coupled-dependence Elmore delay model is applied for clock-tree design [8]. For higher temperatures, the effect of temperature on delay becomes significant [8]. To balance the induced clock-skew and mechanical stress from TSVs, dummy TSVs are inserted. Dummy TSVs reduce the on-chip temperature, thus balances the temperature, and the exerted mechanical stress from dummy TSVs reduce the stress gradient. This helps to achieve reduction in clock-skew. Impact of insertion of dummy TSVs on reduction in clock-skew for 4-tier 3D clock-tree design is shown in Fig. 5. It can be clearly observed from Fig. 5, insertion of TSVs reduces the clock-skew and one can adjust dummy TSV density to balance the temperature and also stress gradient.

Fig. 5: Reduction in clock-skew with insertion of dummy TSV

IV. MACROMODEL BASED MAPPING

A heterogeneously integrated 3D thousand-core system can be still described in state equation by

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t)$$

with $x(t)$ and $y(t)$ being input and output matrices. A, B and C represents the state, input and output ports, respectively. One can first reduce the complexity by compressing the input port B and output C to smaller sized b and c, with study of the input and output signal correlation [4].

The state matrix A can be further reduced by identifying the so-called Krylov subspace, constructed from moment matrices by

$$k(A, b) = (A, Ab, A^2b, \ldots, \ldots).$$

By considering first q moments, a small subspace is formed to fit the original system

$$k(A, b; q) = (A, AR_k, \ldots A^{q-1}b, \ldots).$$

For the system level management, one needs inclusion of sensitivity information, which is solved by forming a structured and parameterized subspace [4]. One can form a new state vector $x_{ap}$ in a structured fashion by expanding the original state vector $x(p, s)$ with respect to parameter $p$ in frequency (s) domain

$$x(P, s) = \sum_{i=0}^{K} \sum_{i+p} (x_{1}^{(i+p)}(s)(\delta p)^{i} \times (\delta p)^{p})$$

Reorganize (4) by considering sensitivities

$$s_{x_{ap}}(s) = A_{ap}x(s) + b_{ap}y(s)$$

As such, one can result in a compact state representation with both sensitivity ($x_{1}^{(i)} \times \delta p^{p}$) and nominal response ($x_{0}^{(i)}$) from the original state equation.

Such a structured and parameterized macromodeling is deployed for a 2-layer 3D design is performed in [4]. With compact macromodeling, one can perform simultaneous TSV density optimization to reduce thermal and power hot-spots as shown in Fig. 6. The macromodeling based design shows 127X faster compared to the approach without reduction of states.

V. CYBER SYSTEM MANAGEMENT

The management of system states such as power and temperature requires the use of macromodel. Based on the data from macromodel and sensor, one can perform prediction and correction to generate the real-time response to track the power and temperature profiles. As shown in Fig. 7, one can predict temperature/power demand by macromodel. When corrected by sensor measured data,
The physical device model has been explored to consider new state thousand-core system with state identification, reduction and control. have gating time for runtime control is 8.3% longer than static control, resulting in large data retention overhead. Though power- static control temperature is fluctuating at very high frequency around control compared to uniform flow-rate control.

A number of design examples have been deployed to support the aforementioned design methodology in 3D thousand-core system, including STT-RAM model, thermal-stress delay model of TSV in clock-tree at physical level; but also microfluid cooling and NEMS based power gating at system level.

ACKNOWLEDGEMENT: This project is supported by Singapore MOE TIER-2 (MOE2010-T2-2-037) fund. The authors thank for the fruitful collaboration with Prof. Chuan Seng Tan, Prof. Wei Zhang and Prof. Chip-hong Chang at Nanyang Technological University.

REFERENCES


