The Feasibility of Carbon Nanotubes for Power Delivery in 3-D Integrated Circuits

Nauman H. Khan  
Department of Computer Science  
Tufts University  
Medford, MA 02155  
Email: nauman@cs.tufts.edu

Soha Hassoun  
Department of Computer Science  
Tufts University  
Medford, MA 02155  
Email: soha@cs.tufts.edu

Abstract—Increased power density and package asymmetry pose challenges in designing power delivery networks for 3-D Integrated Circuits (ICs). The increase in Cu resistivity due to scaling has shifted attention to alternate interconnect technologies. Continued and significant innovations in Carbon Nanotubes (CNTs) manufacturing at CMOS-compatible temperatures with quality low-resistive contacts promise to enable the use of CNTs as a replacement. We investigate in this paper the feasibility of using CNTs for power delivery in 3-D ICs. We evaluate the use of CNTs as Through-Silicon Vias (TSVs) and as wiring for global power delivery grids, fabricated on interposer dies. We assume the CNT interconnect has a mix of single- and multi-walled CNTs with 30% metallic nanotubes. We design a 3-D system-level comparative framework that utilizes select traces from SPEC benchmarks to evaluate improvements of CNTs over Cu. Our results emphasize how CNTs can significantly improve power delivery for 3-D integrated circuits. Using CNTs for on-chip power grid and for TSVs reduces the number of TSVs by 71% when compared to a Cu implementation. For the same substrate area dedicated to power-TSVs, CNTs improve the maximum and average IR drop by 98% and 40%, respectively. Improvements in the Ldi/dt drop are 47% and 18%, respectively.

I. INTRODUCTION

Robust power delivery is one of the ITRS scaling grand challenges due to increasing operating frequencies, increasing power density, and decreasing supply voltages. Increased device density and package asymmetry in a 3-D IC makes power delivery even more challenging. Converting a 2-D IC into a 3-D IC increases the power density by N1/2 where N is the number of dies in the 3-D IC [14]. While the die closest to the package receives power directly from the package, Through-Silicon Vias (TSVs) must be used to deliver power to the dies further away from the package. TSVs however create blockages for on-chip routing and require a keep out zone where active devices cannot be manufactured. Minimizing the number of power TSVs (or their area) thus poses a challenge in designing 3-D power delivery networks (PDNs). While several recent efforts address various 3-D power delivery aspects (via stapling [31], analytical physical model[12], TSV granularity/size trade-offs [14], co-optimization with thermal TSVs [18], early TSV estimation [13]), they all assume Copper (Cu) interconnect.

We investigate in this paper the feasibility of using Carbon Nanotubes (CNTs) for power delivery in a 3-D IC. While CNTs have been intensely investigated for signal interconnect, and in some cases in the context of 3-D ICs [30] [15], little research explores the use of CNT for power grid, or in the context of 3-D power delivery. An initial study derives guidelines on using CNTs for global power grids [21]. To outperform Cu using single-walled CNTs (SWCNT), a minimum metallic nanotube density of 1 per 2.5 nm2 is required. A minimum segment length larger than 20 micron is required for multi-walled metallic CNTs (MWCNT) to outperform an equivalent Cu power grid. While interest in utilizing CNT for interconnect has somewhat waned in the past five years due to fabrication challenges, there continues to be new and promising advances in fabrication technologies to integrate CNT with CMOS processes. One fabrication challenge is utilizing CMOS-compatible temperatures (400-450°C) to grow quality CNTs as earlier work on CNT fabrication required high temperatures to form metal nanoparticles needed for CNT growth and to ensure structural quality which degrades at low temperatures. A short survey of recent exciting advances in low-temperature processes is presented by Nessim et al. [22], and the authors describe their process which combines preheating of incoming gases and using substrate temperatures as low as 400°C. Another fabrication challenge is creating contact with the substrate. Several groups report success (reporting contact resistance as low as 350Ω [19]), with the focus now shifting on developing effective and systematic techniques for measuring contact resistance [17], [16]. Thus, utilizing CNTs to replace Cu continues to be a promising option going forward, as predicted by ITRS, especially in light of scaling requirements and Cu’s electromigration problems and lower current densities. Our study does not investigate the benefits of CNT TSVs in removing heat as such benefits in a 3-D context have been documented [25].

In this paper, we design a CNT vs. Cu comparative study to evaluate using bundles of CNTs (BCNTs) as TSVs and as on-chip power grid. BCNTs provide redundancy and fault tolerance, and reduce resistivities. Unlike the earlier study [21], the goals of our investigation are to compare the minimum number of power-TSVs needed to secure a desired IR drop, and to evaluate the improvements achieved when using CNTs as a power grid in a realistic system-evaluation context. More specifically, we evaluate two distinct scenarios. First,
we evaluate the minimum number of power-TSVs needed to deliver quality power to a representative 3-D IC, shown in Fig. 1, running representative SPEC traces. Second, we evaluate the PDN performance assuming design parameters optimized for a CNT power grid and CNT TSVs. Importantly, instead of focusing on using either SWCNT or MWCNT, we assume a mix of single and multi-walled nanotubes, and that only a portion of the tubes are metallic (conductive).

The rest of the paper is organized as follows. We describe our design setup in Section II. We evaluate the minimum number of TSVs needed for 3-D power delivery assuming various CNT/Cu TSV and power grid configurations in Section III. We compare static IR and Ldi/dt analysis for different configurations in Section IV. We summarize our work in Section V.

II. DESIGN SETUP

We present in this section our design setup to compare the use of CNT vs. Cu in 3-D power delivery. We describe the 3-D stacked architecture in Section II-A. To capture IR and Ldi/dt effects, we use both off-chip and on-chip components of PDN as described in Section II-B. CNT-specific design decisions are presented in Section II-C. Calculated CNT/Cu parameters are summarized in Section II-D.

A. 3-D STACKED ARCHITECTURE AND POWER TRACES

We use a 3-D IC consisting of three dies: a quad-core chipmultiprocessor (PROC), a memory (MEM), and an accelerator engine (ACCL), shown in Fig. 1. Each die is assumed to have an area of 1 cm². We consider the thermal/power profile of each die while considering their placement in the 3-D IC. Since PROC has the highest power consumption, we place it adjacent to the heat sink. We place ACCL farthest from the heat sink due to its lowest power consumption. MEM is placed at the center of the stack to allow shorter access paths from/to both PROC and ACCL. Each core of the PROC utilizes 10W of maximum power, and is composed of five functional blocks: floating point unit (FPU), OOO (the rename, register file, result-bus, and window units), INT (integer arithmetic logic unit), Fetch (combines the instruction cache and branch predictor), and Data (represents the data cache and load-store queue). Memory (MEM) and accelerator engine (ACCL) modules utilize a maximum of 20 and 10W, respectively. MEM is assumed to have the same current trace as the L2 Cache, and ACCL has the same current trace as the FPU block representing an FPU accelerator engine.

We use an architectural-level power model based on Watch [4] to estimate the benchmark specific power dissipation in each functional block. For IR analysis we choose the worst-case representative trace samples based on relative power consumption for each functional block, described in [13]. We extract 1428 out of 10 million traces from four SPEC benchmarks (apsi, bzip, equake, and mcf). For Ldi/dt analysis, we use current traces that represent a variety of current patterns: step, pulse, and resonating. These patterns were derived based on the work of Meeta et al. [10], where four SPEC workloads (apsi, bzip, equake, and mcf) were run for 100 million instructions using Wattch, and 2048 cycle snippets of each benchmark (8192 total traces) representing the current patterns were then extracted.

B. POWER DELIVERY NETWORK (PDN)

Our PDN model consists of off-chip and on-chip networks, as illustrated in Fig. 2 [10]. Values for the off-chip networks are chosen to match the measured off-chip impedance of the Pentium 4 processor [10]. The on-chip network consists of a global level grid-like structure routed in top metal layers. Each grid element is modeled as a resistance and an inductance in series. In addition, current load points and microconnect or TSV points (TSV-P) alternate throughout the grid as shown in Fig. 3. The TSV-Ps are connected to the C4 bumps either directly or through other stacked layers depending on the position of the on-chip PDN in the stack.

The length of each grid element is calculated from the die size (1 cm x 1 cm) and the granularity of on-chip grid. We use a 24 x 24 on-chip power grid. Table I gives the physical dimensions of the grid element for on-chip power delivery network. A fast circuit solver, based on preconditioned Krylov subspace iterative methods [8], is used to solve the SPICE netlist for the modeled configuration. A decoupling capacitance of 33nF/cm² is assumed in our study, corresponding to device capacitance measured using iterative methods [8], is used to solve the SPICE netlist for the modeled configuration. A decoupling capacitance of 33nF/cm² is assumed in our study, corresponding to device capacitance implementation with 1nm gate oxide thickness (from the ITRS roadmap of 90nm, 65nm technology) occupying 20% of die area [12]. The decoupling capacitance is uniformly distributed along the grid elements in our 3-D IC.

![Fig. 1. 3-D stacked architecture.](image1)

![Fig. 2. Modeling power delivery network.](image2)
assume a mixed bundle of CNTs: a mixture of SWCNTs and 100% metallic SWCNTs with identical diameters. We thus followings:

D. Cu and CNT Parameters

We use CNTs for two components of 3-D PDN: as on-chip power grid and as TSVs. Our design configurations are the followings:

- **CNT-Grid-CNT-TSV**: This configuration uses CNTs for both on-chip power grid and TSVs.
- **CNT-Grid-Cu-TSV**: This configuration uses CNTs for on-chip power grid and Cu is used for TSVs.
- **Cu-Grid-CNT-TSV**: This configuration uses Cu-based on-chip power grid and CNTs are used for TSVs.
- **Cu-Grid-Cu-TSV**: This configuration uses Cu for both on-chip power grid and TSVs.

Current CNT fabrication techniques have not yet produced 100% metallic SWCNTs with identical diameters. We thus assume a mixed bundle of CNTs: a mixture of SWCNTs and MWCNTs, having different cross-sectional area. We utilize a tool for computing bundle electrical parameters, Carbon Nanotubes Interconnect Analyzer (CNIA), [29], [11]. This tool utilizes bundle geometry and other properties as input to calculate electrical and thermal parameters. We assume CNT bundle density of 4.5E+11 tubes/cm^2, less than the maximum density of 5E+11 tubes/cm^2 for the selected diameter. Out of three types of CNT chiralities (zigzag, armchair, and chiral) only armchair nanotubes are metallic and the other two are mostly semi conducting. We therefore assume the probability of metallic tubes to be 0.3. We assume an operating temperature of 60°C, typical of modern processors.

The BCNT resistance is dictated by the geometry assumed for the CNT power grid design. Each resulting BCNT segment length is longer than the mean free path when the resistance is dependent on the tube length, as observed experimentally [27], [24]. Imperfect contact resistance is an important parasitic of CNT model. The value of this resistance depends on nanotube diameter and processing technology. For example, Li et al. report an overall resistance of 35Ω for a 25µm long MWCNT with an outer diameter of 100nm [19]. Massoud et al. show that for global level interconnect with large widths, the contact resistance of the bundle is insignificant [20]. For our analysis, we assume negligible contact resistance.

Kinetic and magnetic effects contribute to CNT inductance. Kinetic inductance reflects the net sum of kinetic energy on either side of moving electrons in a nanotube. Earlier experimental observations reported 0.1-4.2nH/µm [28], while later results report no kinetic inductance [5]. Wei Wang et al. give the mathematical details about the relationship of width and kinetic inductance of CNT bundles [29]. For smaller diameters, kinetic inductance is an order of magnitude larger than magnetic inductance. Kinetic inductance depends on the number of shells and number of conduction channels in each shell. Magnetic inductance is produced by time varying currents and the induced magnetic fields. It is dependent on the current loop consisting of the signal line and its return paths. Some work assumes that kinetic inductance dominates magnetic inductance [26], while other work develops a detailed partial inductance model which assumes return paths at infinity [3].

For our analysis, we chose diameter sizes for our bundle that results in kinetic and magnetic inductances of similar magnitudes. We varied diameter mean and standard deviation and computed kinetic inductance. The results are shown in Fig. 4. For all these variations, magnetic inductance remained constant at 1.03E-10H. This shows that kinetic inductance varies exponentially with CNT diameter and as we increase the diameter, kinetic inductance reduces until it becomes comparable to the magnetic one. We choose larger diameters with higher standard deviations for our PDN design.

To compute the R & L values of Cu based on-chip power grid elements, we use Predictive Technology Model (PTM) [1] with resistivity 2.2µΩ-cm and thermal co-efficient of 3.93E-03/°C. These values are shown in the Table II.
III. SUBSTRATE AREA DEDICATED TO TSVS FOR POWER DELIVERY

In this experiment, our goal is to evaluate CNTs to reduce the chip area dedicated to TSVs delivering power in a 3-D IC. We use the four configurations described in Section II to explore the TSV area penalty for each configuration. To find the minimum number of TSVs dedicated for power delivery, we use the Improve Worst Violation (IWV) algorithm described in [13]. We run this algorithm for an IR budget of 12% whereas VDD is assumed to be 1V.

The total number of TSVs for each benchmark are given in Table III (rows 1-4). We use the maximum number of TSVs at each TSV location across all the benchmarks to derive TSV placement across all the benchmarks. Total number of TSVs and the distribution of TSVs over the floorplan for each 3-D IC configuration are given in Table III (row 5) and Fig. 5, respectively. We conclude the followings:

<table>
<thead>
<tr>
<th>Technology</th>
<th>( R )</th>
<th>( L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>1.03E-01Ω</td>
<td>8.30E-11H</td>
</tr>
<tr>
<td>CNT</td>
<td>3.19E-03Ω</td>
<td>1.48E-10H</td>
</tr>
</tbody>
</table>

TABLE II
VALUES OF \( R \) AND \( L \) FOR EACH GRID ELEMENT.

IV. COMPARING POWER DELIVERY OF Cu AND CNT

A. IR Analysis

Once we have calculated the optimal number of TSVs for each configuration, we evaluate CNTs for IR drops across 3-D PDN. To perform IR analysis for the four configurations, we model the power grid as resistive elements. We use the compressed traces from four SPEC benchmarks (apsi, bzip, equake, and mcf) as described in [13]. We study the relative impact of using CNTs over using Cu in a 3-D PDN. For this purpose, we consider the TSV placement for CNT-Grid-CNT-TSV configuration as the baseline and use this TSV placement for rest of the three configurations. We perform IR analysis for each benchmark and report the maximum voltage drop, the average voltage drop, and the standard deviation in Table IV.

We conclude the following:

- For CNT Grid, there is not much difference between the CNT and Cu TSV. This shows that CNT-based on-chip power grid dominates and using Cu TSVs has nominal deterioration in the performance.
- The impact of CNT TSV is more pronounced when using the Cu-based on-chip grid. For PROC die, there is a 32% and 15% improvement in the maximum and average IR drop, respectively. CNT TSVs reduce the standard deviation by 24% which means a smaller variation in voltage across the PROC die.
- The placement of a die in the 3-D stack controls the impact of CNTs on power quality. ACCL, which is closest to package, has the least improvement because it does not have any TSVs. The impact of CNT increases as we move away from the package.
- CNT-based power grid performs better than Cu-based power grid. It has an improvement of 98% when using Cu for both on-chip grid and TSVs. On the other hand, CNT grid has an improvement of 66% over using only Cu-based on-chip grid.

B. \( \text{L}_d/dt \) Analysis

Our strategy to evaluate CNTs for \( \text{L}_d/dt \) performance is similar to the one we used for IR analysis except that we
Fig. 5. Substrate area dedicated to TSVs for power delivery in the 3-D IC. The X- and Y-axis are the grid coordinates; the Z-axis represents the number of TSVs placed at each grid coordinate. The Y-axis have different scales in each figure. The quad symmetry is due to utilizing a quad-core chip multiprocessor.

<table>
<thead>
<tr>
<th>Table IV</th>
<th>IR AND LDi/dt VOLTAGE DROP FOR DIFFERENT 3-D CONFIGURATIONS (NORMALIZED TO CNT-GRID-CNT-TSV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PROC</td>
</tr>
<tr>
<td></td>
<td>CNT Grid</td>
</tr>
<tr>
<td>Max Voltage Drop</td>
<td>1.000</td>
</tr>
<tr>
<td>Avg Voltage Drop</td>
<td>1.000</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>1.000</td>
</tr>
</tbody>
</table>

As observed in the IR analysis, die placement in the 3-D IC controls the relative impact of CNTs on the power quality. The transient behavior of Cu is better than that of CNTs for ACCL. This behavior results because ACCL is connected directly to the package using C4 bumps and there was no change in the off-chip PDN or the number of C4 bumps. Decreasing the number of TSV results in isolating ACCL from the rest of the 3-D IC and reduces the sharing across dies. This reduced sharing results in local improvement for ACCL.

Although the CNT grid performs better, the improvement is not as significant when compared to the results of the

consider inductive and capacitive components along with resistive network. We use the optimal number of TSVs for CNT-Grid-CNT-TSV. The results are presented in and Table IV. We conclude the following:

- The relative impact of CNT, shown in Table IV, shows that CNT TSVs result in better performance for both grid types. For the CNT-grid, we get 5% and 12% improvement in the maximum and average voltage drop in the PROC die, respectively. For the Cu-grid, these improvements are 20% and 13%. These results show that even with fewer TSVs, CNT TSVs outperform the Cu TSVs.

- As observed in the IR analysis, die placement in the 3-D IC controls the relative impact of CNTs on the power quality. The transient behavior of Cu is better than that of CNTs for ACCL. This behavior results because ACCL is connected directly to the package using C4 bumps and there was no change in the off-chip PDN or the number of C4 bumps. Decreasing the number of TSV results in isolating ACCL from the rest of the 3-D IC and reduces the sharing across dies. This reduced sharing results in local improvement for ACCL.

- Although the CNT grid performs better, the improvement is not as significant when compared to the results of the
IR analysis due to the increased inductance of CNTs. For example, CNTs result in an improvement of as high as 98% and 66% for maximum and average IR drop for PROC die. The same numbers for Ldi/dt analysis are limited to 47% and 5%.

V. Conclusion

We performed in this paper a detailed comparative study to evaluate the use of CNTs for power delivery in 3-D ICs. We explored the feasibility of using CNTs as TSVs and as an on-chip power grid implemented on interposer dies. The results herein are specific to the architectural setup and CNT modeling assumptions. However, the results are valuable as they are grounded in up-to-date CNT modeling practices (e.g., assuming 30% metallic CNTs) and measurements (e.g., low contact resistance) and they provide the first system-level trace-based evaluation of improvements possible with CNTs. Our results show that CNTs reduce the number of TSVs, and thus area, for power delivery. Using CNTs for TSVs only and for both TSVs and on-chip power grid reduced the area requirement by 40% and 71%, respectively, over an equivalent Cu implementation. We performed IR and Ldi/dt analysis for different design configurations with CNT-optimal TSV placements. Using CNTs either for the on-chip power grid, or for TSVs, or for both improved power quality. For the same optimal substrate area dedicated to Cu-based power delivery, CNTs can improve the average voltage drop by 40% for IR analysis and by 18% for Ldi/dt analysis.

References