**SLIDE 1: TITLE**

Thank you for the introduction. This is a joint work of UCSD and Qualcomm Research.

**SLIDE 2: OUTLINE**

My talk is structured as follows. I begin with motivation.

**SLIDE 3: 3DIC VALUE PROPOSITION**

As the semiconductor industry nears the end of the CMOS roadmap, 3DICs have emerged as a promising solution to continue Moore’s Law trajectory of value scaling. 3DICs are fundamental to the whole idea of More than Moore. In the context of this work, power reduction benefit is the key value proposition for 3DICs. 3D power estimation tools are required to enable fast and accurate implementation-space exploration to evaluate power benefits of 3DICs.

**SLIDE 4: 3D POWER ESTIMATION CHALLENGES**

3D power estimation is challenging because the benefit varies with netlist topologies, constraints, etc.. Moreover, no golden 3D implementation flow exists. This introduces a chicken-and-egg loop because we are trying to embed netlists not created for 3D into 3D. Therefore, we can only rely on 2D implementations for 3D guidance.

But, no tool exists today that predicts 3D power benefits based on 2D implementations.

**SLIDE 5: OUTLINE**

In the context of this work,

**SLIDE 6: 3D PRELIMINARIES**

This figure shows a classic 2DIC. The height and width of the die are H and W, respectively. When this implementation is changed to 3DIC, we can create two vertically stacked dies. The height and width of each die is divided by the square root of two with respect to those of the 2DIC. The dies are interconnected by vertical interconnects.

Shrunk2D is another way to emulate a 3DIC by doing P&R in a die with the same height and width of the 3DIC. This flow proposed by Panth et al. is the best “3D” flow today.

**SLIDE 7: 3D POWER ESTIMATION TOOL**

The implementation-space of a 3DIC is high-dimensional. Multiple choices are available in setting constraints such as timing and design rules, layout contexts such as aspect ratio, utilization, EDA tool flows, and technology choices such as Vt flavors, libraries, etc.

Therefore, we need an accurate power estimation tool to quickly determine the best set of parameters from this high-dimensional space that delivers 3D power benefits.

**SLIDE 8: KEY CONTRIBUTIONS**

In this work, we provide a tight upper bound on 3D wirelength reduction.

We are the first to develop a 3D power benefits estimation tool, 3DPE based on 2D implementations. We predict the percentage delta power benefits of 3DIC relative to 2DIC implementations. The error range of our predictions is within 10%.

We propose a novel parameter selection methodology based on sensitivity of SP&R outcomes to wireload model and RC scaling. This is an unexplored approach to assess how gate-level netlists will react to 3D vs. 2D implementation contexts.

We propose a “stress testing” validation approach and application of 3DPE in model-guided implementation.

**SLIDE 9: FLOW IMPROVEMENTS**

We use the latest Shrunk2D or S2D flow from Georgia Tech as our baseline. We have transplanted the flow and replicated published results at UCSD.

We have enhanced the S2D flow in various ways.

**SLIDE 10: OUTLINE**

What is an upper bound on the wirelength reduction in 3D?

**SLIDE 11: TIGHT UPPER BOUND ON 3DIC WL BENEFITS (1)?**

The figure on the left shows a 3D grid graph with net n1 in red color. The right-figure shows the cross-section view of the net. Our cost model is as follows. The Z-direction cost is zero because the heights of vertical interconnects are assumed to be very small in 3DIC. X, Y cost is 1 unit per hop.

Key 1 of this derivation is we start with an optimal 3D routing; hence the reduction is an upper bound. Key 2 is without loss of generality, we can stretch the graph arbitrarily in one direction, and hence the bound is tight. Therefore, this example shows a tight upper bound.

The length of net n1 is therefore, 1 due to the segment BC.

**SLIDE 12: TIGHT UPPER BOUND ON 3DIC WL BENEFITS (2)?**

Net n1 becomes net n1’ in 2D and its length is 3 due to 3 hops between segments AB, BC and CD. The wirelength reduction is therefore 66.7%.

In our experiments, we do not explore RC scaling factors below 33.3% as guidance from this bound.

**SLIDE 13: OUTLINE**

Next, I describe our modeling methodology.

**SLIDE 14: TESTCASES AND IMPLEMENTATION**

We use a wide range of IPs that resembles building blocks of modern SoCs. The table shows the list of our testcases. We use five types of testcases -- CPU, GPU, modem, multimedia and peripheral engine.

**SLIDE 15: IMPLEMENTATION-SPACE PARAMETERS**

Here is list of various implementation-space parameters we use in our experiments. The parameters span across various constraints, layout contexts and technology choices.

**SLIDE 16: FLOW AND TOP-10 MODELING PARAMETERS**

However, to restrict the dimensionality and runtime of our modeling problem, we seek to explore the 10 most influential parameters.

In our flow, we use engineered WLMs to perform synthesis. Then, we perform both 2D and Shrunk2D P&R. **We use S2D as a proxy for 3D**.

For both P&R flows, we use scaled RC cap tables. We then extract parameters for modeling.

The top-10 parameters include six constraints such as clock period, max transition time, etc. We also use two implementation and two technology parameters such as utilization, multi-Vt libraries, respectively.

**SLIDE 17: MACHINE LEARNING METHODOLOGY**

With parameters extracted from 2DIC implementation, we perform modeling. We use artificial neural networks to capture the complex interactions between parameters.

We define the ANN architecture with one input and one output layer, plus two hidden layers. We search for the best number of the epochs of back propagation and the number of neurons per layer using the loop here to achieve bounded errors.

We obtain our ground truth from S2D runs.

**SLIDE 18: OUTLINE**

Now, I present our results.

**SLIDE 19: BOUNDED-ERROR MODELS**

This plot shows the actual percentage delta power benefit in the X-axis and the predicted values in the Y-axis for the five types of testcases. We derive separate models for each of the power components – internal, switching and leakage. Then we compose these models to create a model for total power.

We challenge ourselves to predict delta power. Consider this example in which 2DIC power is 90mW and the corresponding 3DIC power is 80mW. The delta is 10mW. To achieve 10% error on actual 3DIC power, the predictions can range between 72mW to 88mW. However, to achieve 10% error on the delta, the predictions must range between 79mW to 81mW. This is a difficult task.

Across all our test data points, the range of errors is 9%.

**SLIDE 20: NOVEL “STRESS-TESTING”**

We do not have ground truth from true 3DIC implementations, so we must test if our 3DPE models are capable of returning unlikely predictions.

We perform “stress testing” of the models. We perform Monte Carlo-like simulations by varying the mean and variance of each parameter in the models.

The figure shows a histogram of percentage predicted delta power. The maximum value is 39% for data points that are practically realizable.

We reject data points that are not practically realizable. For example, data points in which the number of cells, utilization and the cell area are mismatched. Or, the wirelength and the number of cells are mismatched.

**SLIDE 21: MODEL-GUIDED IMPLEMENTATION**

The hypothesis here is 3DPE should guide implementation if the predictions are reliable. We refer to this as model-guided implementation. We test this hypothesis with an implementation here. This figure shows WLM cap in the X-axis and 3D power in the Y-axis. Minimum 3D power is achieved at 0.45pF. Our models predict the cap to be 0.75pF, using which the delta power is 0.34mW. Therefore, 3DPE model guidance is better than S2D by 5%.

**SLIDE 20: OUTLINE**

**SLIDE 22: SUMMARY**

In summary, power reduction is a key value proposition for 3DICs. Lack of a golden 3D flow makes prediction of 3D power benefits a difficult problem.

We develop the 3DPE tool that predicts the percentage delta power benefits of 3DIC relative to 2DIC implementations. 3DPE is accurate with an error range of 10%.

We also propose stress testing and model-guided implementation approaches with 3DPE. Our ongoing works include extending 3DPE from block-level to SoC-level predictions and developing a true 3D flow.

**SLIDE 23: ACKNOWLEDGMENTS**

We thank Prof. Zelikovsky, Prof. Lim, Dr. Panth and Dr. Jung for various discussions and generosity in flow development at UCSD.

**SLIDE 24: THANK YOU**

Thank you for your attention.