

AREA-EFFICIENT AREA PAD DESIGN FOR HIGH PIN-COUNT CHIPS

Louis Luh, John Choma, Jr., and Jeffrey Draper

University of Southern California

Los Angeles, CA 90089

luh@isi.edu, johnc@usc.edu, draper@isi.edu

ABSTRACT

This paper presents an area pad layout method to efficiently reduce the space required for interconnection pads and pad drivers. Unlike peripheral pads, area pads use only the top metal layer and therefore allow active circuitry to be laid out underneath. With identical functional elements grouped together, a group of pad drivers share the same well and can be placed tightly together. The use of silicided diffusion reduces the well contact to diffusion contact spacing requirement. By taking advantage of this spacing requirement and using serpentine gate layout, a driver's size can be effectively reduced without reducing the driving capacity. An embedded multicomputer router interface chip has been implemented using these techniques and has achieved 554 pads in a 9mm $\times$ 6mm chip with a 0.8 μ m single-poly 3-metal N-well CMOS process.

1. INTRODUCTION

Conventional pad design uses only the chip periphery for pad connections and pad drivers. Due to the minimum space requirement between pads and the minimum distance requirement between bonding wires, the maximum pin count is thus limited by the length of a chip's periphery. Besides the disadvantage of the maximum pin count limitation, conventional peripheral pad design exhibits low area efficiency. The I/O pads typically use all layers of metal and N-well (or P-well). The pad driver and ESD circuit of each pad occupies a certain amount of space, disallowing any other circuitry within that space. Furthermore, peripheral pad design often requires that a signal be routed great distances from the core circuitry to the pad frame. This may cause extra delay and/or require a buffer stage.

With the use of area bonding conductive (ABC) adhesives [1, 2] and flipped-chip package or chips-first multi-chip module (MCM) fabrication, area bonding (central I/O circuits [3]) can be used. Area bonding pads require the use of only the top metal layer for I/O connection. This allows other circuits to be built

underneath the pads, and hence the space can be used more efficiently.

To further reduce the chip size, bit-stack layout and grouped I/O circuits can be used [3]. This paper presents an approach, which effectively reduces the space required for I/O circuitry and enables a small sized CMOS VLSI chip to have high pin count.

2. CIRCUIT DESIGN

The proposed pad design was originally designed for a router interface chip for a high-bandwidth (1.3GB/s) embedded multicomputer network. This chip was designed for a single-poly 3-metal N-well process. The top metal layer (metal 3) is reserved for I/O pads, power rails, and the connections between pads and driver circuits. To effectively reduce the space for routing, bit-stack layout and single direction routing for metal 2 layer are used. All the metal 2 wires are routed horizontally. In the datapath blocks, the metal 2 layer is used for signal channels. In the I/O circuit blocks, the metal 2 layer is used for power rails to provide the required currents.

Higher I/O circuit density is achieved through a complete reorganization of the I/O circuits into groups of I/O functions with identical type according to the method described below.

1. Group I/O circuits of identical function into macros. (For example: create a macro of a group of output drivers; create another of a group of tristate buffers; create another of a group of input receivers, etc.).
2. Within each macro, stretch out each I/O circuit into a vertical string of subcomponents, one subcomponent wide. Place the vertical strings side by side horizontally to form a multibit macro. Identical subcomponents of each bit will then fall side by side and be able to share the area of their edge structures.

(a) Place all NMOS transistors close together so they can share ground and substrate contacts. Similarly, place all PMOS transistors close together so they can share N-well, VDD, and well contacts (an example is shown in Figure 1).

(b) Place the sources of the transistors which are connected to VDD or ground close/next to the edges, and place well contacts or substrate contacts right next to the sources (connected without space), so that well/substrate contacts are right on the edges (as shown in Figure 1). Thus, when placing two identical drivers next to each other, their edges are shorted together to share the same ground/VDD and the substrate/well contacts (as shown in Figure 2(b)). This eliminates the gap requirement between two different MOS transistors and also reduces the area required for well/substrate contacts.

3. Use serpentine gate layout to reduce transistors' sizes. Figure 1 shows an example of an output driver which utilizes serpentine gate layout.

4. Construct both the individual circuits and their interconnections using only process layers metal 1 and below as shown in the example in Figure 1. On top of the I/O circuits, metal 2 is reserved for VDD and ground, so that the power rails can be very wide to provide high currents and low voltage drops. The metal 2 power buses at the output end of the macro are locally constricted to allow escape of the I/O wires up to metal 3 for connection to the chip pads. Figure 2 shows an example of a group of eight output drivers.

5. Divide the I/O signals of the core circuitry into groups, each of which requires the same or nearly the same set of I/O circuitry. Pull out the same group of I/O signals from the same edge of the core circuitry block and place the grouped pad drivers and pads right next to the edge to reduce the lengths of interconnections.

This I/O circuit design takes advantage of the low-impedance silicide layer. In item 2(b), by connecting the N-well diffusion/contacts to PMOS source diffusion (or the P substrate diffusion/contact to NMOS source diffusion) (as shown in Figure 2(b)), the sources and the well/substrate contacts are automatically shorted by the silicide layer. This method eliminates the typical space required between MOS transistors and

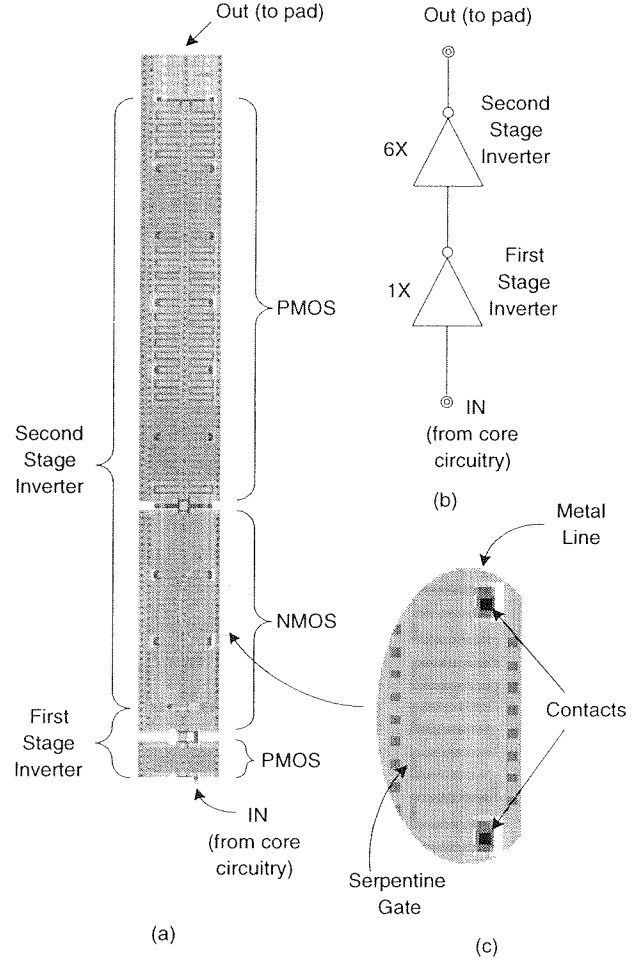


Figure 1: 1-bit strip layout of an output buffer: (a) the layout, (b) the circuit diagram, and (c) an enlarged portion of the driver which shows the long polycide gate is connected to a metal 1 wire at several places to reduce the gate resistance.

well/substrate contacts. The well/substrate contacts also provide additional current for the transistors, which increases the driving capacity without increasing the space. This capability also reduces the space required for guard rings.

The use of serpentine gate layout reduces transistor sizes and enhances the driving capacity of transistors. However, higher gate impedance and higher source/drain impedance can reduce the transistor speed and driving capacity when using serpentine gates. In our design, the gate impedance is reduced by connecting the long polycide (silicide + poly) gate to a metal 1 wire at several points as shown in Figure 1(c). The source/drain impedances are reduced by limiting the length of each finger of a source/drain.

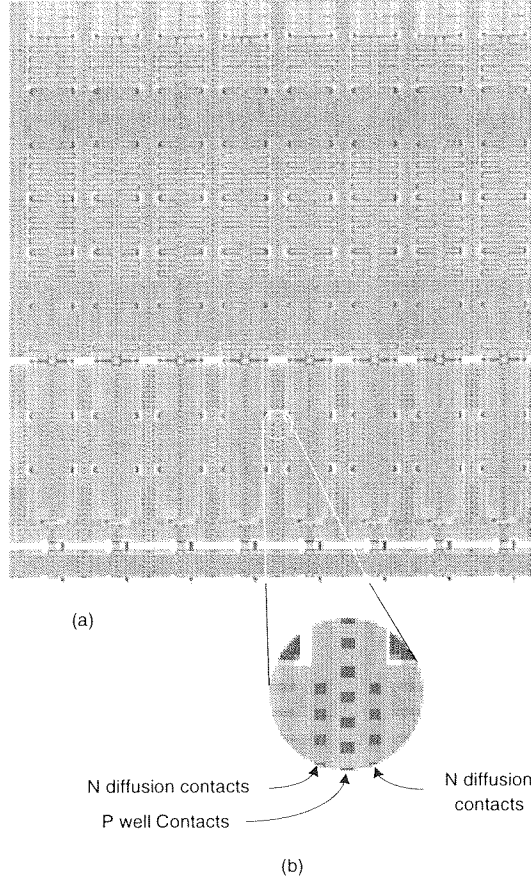


Figure 2: (a) Layout of an 8-output driver bank. (The inputs are on the bottom and the outputs are on the top. Metal 2 wires are laid out horizontally for connecting VDD and ground. For an output driver bank with more drivers, metal 2 wires can be extended vertically to provide more current) (b) Enlarged edge structure of two adjacent pad drivers.

3. EXPERIMENT RESULT

These I/O circuits have been implemented in a $0.8\mu\text{m}$ single-poly 3-metal process for a router interface chip [4]. The pad drivers are designed to drive 100pF load with a rise time and fall time less than 4ns . The ESD protection circuitry is implemented by spark fingers, thick-oxide MOS, diodes, and short diffusion resistors. This circuitry has passed the ESD test of 1000V human body model.

Figure 3 shows the microphotograph of two tristate buffer banks (groups). Each tristate buffer bank contains 22 tristate buffers and two extra drivers for driving the controlling signals of the tristate buffers. To

provide the currents for 22 drivers and to minimize the voltage drop across the VDD and ground wires (metal 2 layer), we use all the metal 2 space on top of the drivers for VDD and ground. The size of each 22 tristate buffer bank is $1028\mu\text{m} \times 266\mu\text{m}$.

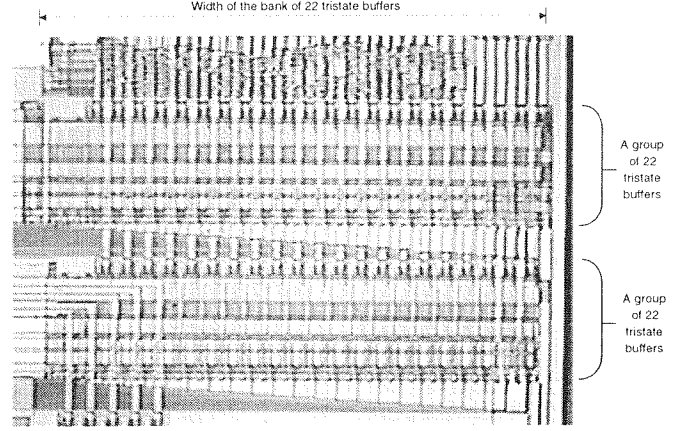


Figure 3: Microphotograph of two banks of tristate buffers. Each bank contains 22 tristate buffers and two drivers for driving controlling signals of the tristate buffers. The size of each bank is $1028\mu\text{m} \times 266\mu\text{m}$.

Figure 4 shows the microphotograph of the entire chip. This chip has 554 I/O pads and approximately 250,000 transistors. Figure 5 shows the locations of grouped pad drivers and pads. All the I/O pads are laid out with a $100\mu\text{m}$ pitch on a $50\mu\text{m}$ grid. The chip size is $9\text{mm} \times 6\text{mm}$. The size of the chip is governed by the sizes of the data path blocks and the controlling blocks. The overall chip layout did not achieve very high area efficiency due to the limited time frame for the chip layout and the use of low-cost silicon compiler tools for the controller blocks.

This chip was packaged by Cray Research's MBGA/MCM "chip-first" interconnection process [5], which builds a multi-layer interconnect on top of die directly attached to the wafer substrate. This process offers superior heat dissipation and can support several dies. However, this process has very high cost and may not be suitable for mass production. This chip has been used in a prototype of a high-bandwidth (1.3GB/s) multi-computer network [4][5] with a system clock up to 50MHz .

4. CONCLUSION

An area-efficient area pad design has been presented. With the use of only the top metal layer (metal 3) for

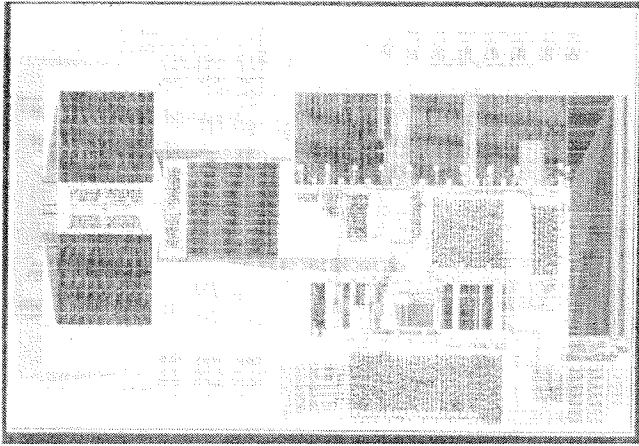


Figure 4: Microphotograph of a router interface chip which used the area pad design. (The size is 9mm $\times$ 6mm.)

chip pads, bit-stack layout, and grouped I/O circuits, the chip space is used more efficiently. The use of silicided diffusions greatly reduces the sizes of pad drivers while enhancing driving capacities. The higher area efficiency is also achieved by the use of serpentine gate layout and the appropriate layout choice of metal layers. The result of the router interface chip has shown that this design approach is ideal for high pin-count complex VLSI circuitry to achieve high area efficiency.

Acknowledgements

This project is sponsored by DARPA (Contract No. DABT63-95-0136).

5. REFERENCES

- [1] J. C. Bolger and K. Gilleo, "Area bonding epoxy adhesive performs for grid array and MCM substrate attach," in *Proc. 44th Electronic Components and Tech. Conference*, May, 1994

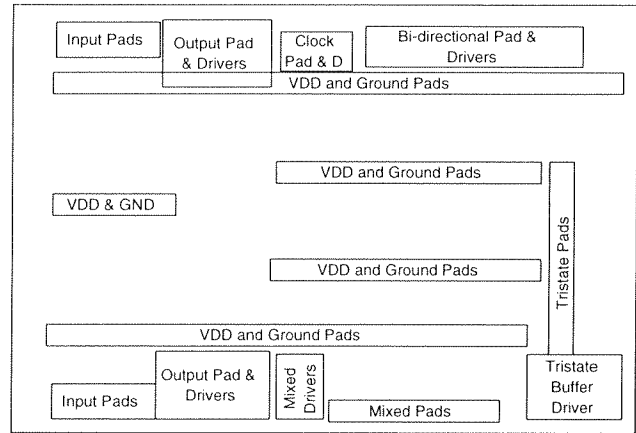


Figure 5: Locations of the grouped pad drivers and pads of the router interface chip.

- [2] J. C. Bolger and J. Popielarczyk, "Area bonding conductive (ABC) adhesives for flex circuit connection to LTCC/MCM substrates," in *Proc. 45th Electronic Components and Tech. Conference*, May, 1995
- [3] R. P. Masleid, "High-density central I/O circuits for CMOS," in *IEEE J. Solid-State Ckts.*, vol. 26, no. 3, pp. March, 1991
- [4] C. S. Steele, J. Draper, J. Koller, and C. La-Cour, "A Bus-Efficient Low-Latency Network Interface for the PDSS Multicomputer," *Proceedings of the International Symposium on High Performance Distributed Computing*, pp. 213-222, August 1997
- [5] J. Koller, "PDSS, The packaging driven scalable system, final project report," *Information Sciences Institute Internal Report*, Nov, 1997