

Input Reordering for Power and Delay Optimization

Masanori Hashimoto, Hidetoshi Onodera and Keikichi Tamaru
 Department of Electronics and Communication, Kyoto University
 Sakyo-ku, Kyoto-shi, 606-01, Japan
 E-mail: hasimoto@tamaru.kuee.kyoto-u.ac.jp

Abstract---It is known that input reordering of a gate affects the power dissipated by the internal capacitance of the reordered gate, which has been utilized for power reduction so far. We show that the reordering also has a significant effect on the power dissipation of the gate which drives the reordered gate. It is because the input capacitance depends on signal values of other inputs. We propose a reordering algorithm considering the power dissipation in the driving gate, the reordered gate and the gates driven by the reordered gate. Experimental results using 21 benchmark circuits show that our method reduces the power dissipation in all the circuits by 3.6 % on average. There is a possibility that power dissipation is reduced by 17.2% maximum. In the case of delay and power optimization, our method reduces delay by 7.0 % and power dissipation by 3.1 % on average.

I. INTRODUCTION

Power dissipation has been a strong concern in VLSI design, not only for meeting strong requirements for the use in portable environment, but also for reducing package costs, maintaining high reliability, etc. In the various stages of the VLSI design, many techniques for power reduction have been proposed, such as supply-voltage scaling[1], technology mapping for low power[2], gate sizing[3], input reordering[4-7], and so on. The technique of input reordering has two advantages. The first advantage is that input reordering has little effect on the layout area. The second is that other techniques can be combined easily with input reordering. In [4-7], the authors discussed input reordering for power reduction such that the reordering reduces the power dissipation inside the reordered gate. The input reordering, however, affects not only the power dissipated inside the reordered gate but also the power dissipated by the fan-in gates and the fan-out gates, which has not been utilized previously (See Fig.1). Where the fan-in gates are the gates which drive the reordered gate and the fan-out gates are the gates driven by the reordered gate. In this paper, we discuss the effect of input reordering on power dissipation in the fan-in gates and the fan-out gates as well as in the reordered gate, and propose an improved method for power optimization which exploits the effect.

This paper is organized as follows. Section II discusses the effect of input reordering on power dissipation and delay. Section III discusses a strategy of input reordering for power and delay optimization at each gate. Section IV introduces an algorithm of input reordering for power and delay optimization for the whole circuit. Section V shows the experimental result of our method. Finally Section VI concludes the discussion.

II. THE EFFECT ON POWER DISSIPATION AND DELAY

In this section, we discuss the major effect of input reordering in the fan-in gate, the reordered gate and the fan-out gate. So far, only the effect for the reordered gate has been considered for performance optimization. We show that there are notable

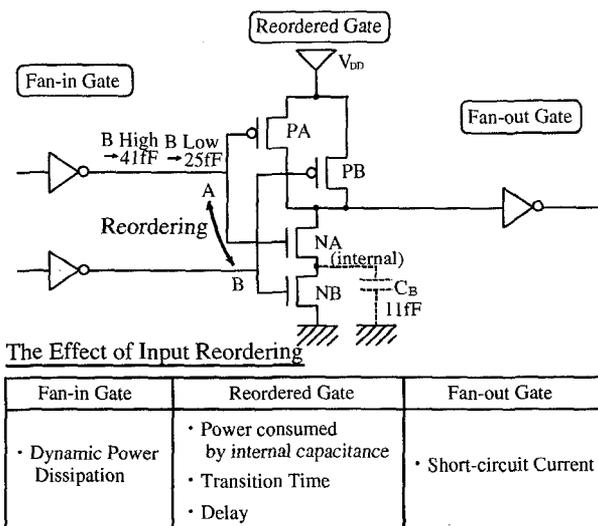


Fig. 1. The Effect of Input Reordering in 2-input NAND

effects of input reordering on power dissipation in the fan-in gates and the fan-out gates, which could be utilized for performance optimization as well.

A. Fan-in Gate

The dynamic power dissipated by a fan-in gate varies by the input reordering of the reordered gate (the gate which the fan-in gate drives). This is because the input capacitance of the reordered gate, i.e. the load capacitance of the fan-in gate, depends on the signal values of other inputs of the reordered gate. We demonstrate the difference numerically using an example from a real 0.7 μm standard cell library. Figure 1 shows a 2-input NAND gate with inputs A and B, two nMOSFETs NA and NB in series, being NA closer to the output. When the input B keeps low, the input capacitance of A is 25 fF. When the input B keeps high, the input capacitance of A becomes 41 fF which is 64 % larger than the previous case. This big difference comes from the difference in the source voltage of NA. With the input B low, the source of NA is floating from ground and the source voltage of NA is $V_{DD} - V_{TH}$ when the input A becomes high, where V_{DD} is the supply voltage and V_{TH} is the threshold voltage of MOSFETs. The voltage difference across the gate of NA is V_{TH} . On the other hand, with the input B high, the source of NA becomes tied down to ground, and hence the voltage difference becomes V_{DD} . The difference of the input capacitance (16 fF) is larger than the internal capacitance ($C_B = 11 \text{ fF}$) which is the sum of the diffusion capacitances of the source(NA) and the drain(NB).

TABLE I
TYPICAL CHARACTERISTICS OF 4-INPUT NAND

	Pin A	Pin D	Pin D/Pin A
Power(μ J)	3.8	6.8	179%
Rise Delay(ns)	0.26	0.44	169%
Fall Delay(ns)	0.18	0.23	128%
Rise Transition Time(ns)	0.41	0.73	178%
Fall Transition Time(ns)	0.34	0.33	97%

B. Reordered Gate

Internal capacitances in a reordered gate have an influence on the power dissipation, delay time, and transition time of the reordered gate. References [4-7] discusses methods for power reduction by input reordering such that the number of charging and discharging the internal capacitances could be reduced. Let us take a 4-input NAND gate as an example to investigate how power dissipation and delay vary input by input. TABLE I lists the power dissipation (dissipated energy, rigorously), rise/fall delay times and transition times when the output load capacitance is 60 fF. The gate is driven by input A or D, where input A is closest to the output and input D is closest to ground. The dissipated power(energy), rise delay time and rise transition time of input D are larger than those of input A by 79 %, 69 %, 78 %, respectively.

C. Fan-out Gate

Input reordering of the reordered gate affects the power dissipation of a fan-out gate. Because the reordering changes the transition time of the input signal of the fan-out gate which leads to the change in the short-circuit current of the fan-out gate. If the transition time is short, the short-circuit power dissipation in the fan-out gate becomes small.

III. REORDERING STRATEGY

In this section, we discuss the reordering strategies for each effect discussed in the previous section. Some of the strategies do not conflict, but some may be inconsistent. For example, the strategy of power optimization for the reordered gate is consistent with that for the fan-out gate, whereas the strategy may conflict with that of delay optimization and that of power optimization for the fan-in gate, as shown in this section. The overall algorithm which resolves the conflict will be shown in the next section.

The signal probability and the transition density which are used for the discussion from this section are defined as follows. The signal probability $P(x)$ is defined as the average fraction of clock cycles in which steady state value of node x is a logic high. The transition density $D(x)$ is the average number of transitions per second at node x [8]. Hereafter, we use $P(x)$ and $D(x)$ without further explanation.

A. Power Dissipation in Fan-in Gate

We use a 3-input NAND gate (Fig.2) as an example to explain the strategy of power dissipation in the fan-in gate. As discussed in Section II.A, input capacitance becomes small when the source of the input transistor is floating from ground. Therefore, if inputs are reordered such that the number of input transitions

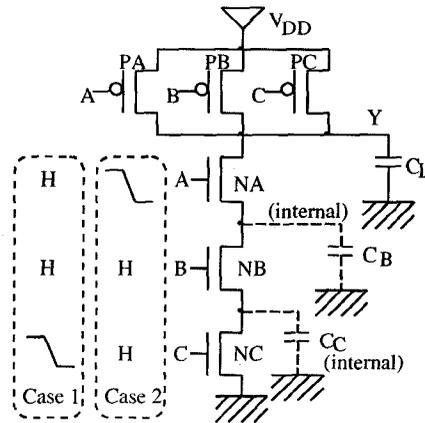


Fig. 2. 3-input NAND

with the source floating becomes large, the power dissipation in its fan-in gates becomes small. The probability with the source of NA floating can be represented as $1 - P(BC)$. So the transition density of A with the source of NA floating can be represented as $D(A)\{1 - P(BC)\}$. Similarly the transition density of B with the source of NB floating can be represented as $D(B)\{1 - P(C)\}$. Thus we should reorder the inputs so as to increase S (the sum of the transition density with the source floating).

$$S = D(A)\{1 - P(BC)\} + D(B)\{1 - P(C)\} \quad (1)$$

Similarly we can represent the sum S in 3-input NOR case and so on.

B. Power Dissipation in the Reordered Gate and Fan-out Gate

We explain the strategy of power dissipation in the reordered gate and the fan-out gate using a 3-input NAND gate (Fig.2) as an example. First, we examine the situation when the output changes from low to high. Let us consider two cases when the inputs (A, B, C) change from (1, 1, 1) to (1, 1, 0)[Case 1], and from (1, 1, 1) to (0, 1, 1)[Case 2]. In Case 1, the internal capacitances C_B and C_C as well as the output load capacitance C_L are charged, whereas only the output load capacitance C_L is charged in Case 2. Therefore the power dissipation is larger in Case 1 than in Case 2. Also, the output transition time of Case 1 becomes longer than that of Case 2, because the total amount of capacitances to be charged is larger, which leads to the increase in short-circuit power dissipation in fan-out gates. As a result, we can draw a strategy such that the input which changes the output value frequently should be placed near the output.

Secondly, we examine the situation when the output keeps high while the inputs are changing. In such a case, we should avoid charging the internal capacitances as much as possible. This can be achieved by keeping the nMOSFET off which is close to the output. In other words, the input with high signal probability should be placed far from the output. Similar discussion can be done for NOR gates.

Based on the discussions above, the reordering policies are summarized as follows:

Policy 1: The input which changes the output value frequently should be placed near the output.

Policy 2(NAND): The input with high signal probability should be placed far from the output.

Policy 2(NOR): The input with low signal probability should be placed far from the output.

Then, we translate the above policies into a more tractable form. The transition density of the output Y can be calculated as the sum of transition densities which are passed from each input to the output, as follows:

$$D(Y) = P(BC)D(A) + P(AC)D(B) + P(AB)D(C) \quad (2)$$

Using the transition densities, we can rewrite Policy 1 as follows. The inputs should be placed from the output side to ground side in the descending order of the transition densities from each input to the output.

Now let us examine whether this strategy conforms Policy 2 or not. In the case of NAND gates, if the signal probability of a certain input is high, the transition density which is passed from other inputs to the output tends to be large, and hence the input tends to be placed far from the output. If the signal probability of a certain input is low, the transition density which is passed from other inputs to the output tends to be small, and hence the input tends to be placed close to the output. We can conclude that Policies 1 and 2 do not contradict in the case of NAND gates. Similarly in the case of NOR gates, we can see that Policies 1 and 2 do not contradict.

C. Delay

The delay of a gate differs not only input by input but also by the direction of output transition(rise/fall). Even in transitions driven by a parallel connected transistor(eg. output rise/fall for NAND/NOR gates), there exists input-pin dependency as seen in TABLE I. Also, the fall/rise delay of the pin with the smallest rise/fall delay is not necessarily the smallest. We therefore need to consider both rise and fall pin-to-pin delays for each input instead of reducing them to a single pin-to-pin delay as is done in conventional timing optimization approaches [9, 10]. This implies that we should associate two delays(fall and rise delays) with each output. In order to evaluate the contribution of each delay to the overall circuit delay, we calculate the slack at each output and use it as the measure of delay.

For the delay optimization, we use the input order which makes $\min(\text{rise_slack}, \text{fall_slack})$ the largest. This strategy does not increase the delay of critical path.

IV. OPTIMIZATION ALGORITHM

In the previous section, we present three strategies for the power reduction in the fan-in gates, the power reduction in the reordered and the fan-out gates and the delay optimization for each gate, respectively. In this section, we discuss an algorithm which combines the three strategies for the performance optimization of the whole circuit.

A. Optimization in Each Gate

We use the slack as the measure of the delay constraint, so that the strategy of delay optimization can be easily combined with the other power reduction strategies. The important point here is how to combine the power reduction strategy for the fan-in gates with that for the reordered and the fan-out gates.

In order to judge the superiority of the two strategies, we estimate the power dissipated in the fan-in gates and the reordered gate. The contribution of the fan-out gates is omitted from the consideration since it is small compared to those of the fan-in gates and the reordered gate. We use a 3-input NAND gate (Fig.2) as an example to explain the method for the power estimation. Referring to each term in the right-hand side of Eq.(2), we assume that the power dissipated by charging and discharging the internal capacitances is represented below.

$$PW_{internal} = \frac{1}{2}(V_{DD} - V_{TH})V_{DD}\{C_B P(AC)D(B) + (C_B + C_C)P(AB)D(C)\} \quad (3)$$

Referring to Eq.(1), we assume that the power dissipated by charging and discharging of the input capacitances is represented below.

$$PW_{input} = \frac{1}{2}V_{DD}^2\{C_{AF}D(A)\{1 - P(BC)\} + C_{AN}D(A)P(BC) + C_{BF}D(B)\{1 - P(C)\} + C_{BN}D(B)P(C) + C_{CN}D(C)\} \quad (4)$$

where C_{AF} , C_{BF} are the input capacitances with each source floating, and C_{AN} , C_{BN} , C_{CN} are those with each source connecting to ground.

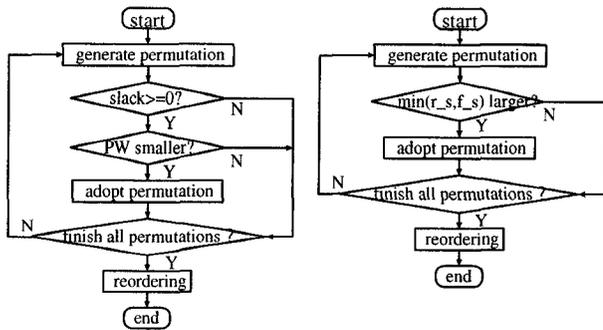
Using these two estimated power dissipations, we consider that the input ordering which minimizes the total power $PW(= PW_{input} + PW_{internal})$ is the best for low power. This ordering can not be found easily, so we try all the permutations and choose the one with the smallest PW . When the delay constraint is imposed, we calculate the slack for each permutation and select the ordering with the smallest PW and positive slack. This flow is shown in Fig.3(a).

In the case of delay optimization, the ordering that gives the largest $\min(\text{rise_slack}, \text{fall_slack})$ can not be found easily, so we calculate $\min(\text{rise_slack}, \text{fall_slack})$ for all the permutations and select the best order. This flow is shown in Fig.3(b)

B. Optimization of the Whole Circuit

For delay optimization, each gate is reordered with the strategy in Section IV-A, in a breadth-first search order starting from a gate with all the inputs driven by primary inputs. Since a reordering of a certain gate may change the slack of a gate not only in the fan-out direction but also in the fan-in direction, delay optimization is not a single path process. Even if all the gates in the circuit have been reordered once, there is a possibility that further delay reduction can be achieved. Thus iterative optimization is required. The delay optimization loop finishes when the delay of critical path can not be decreased. In the case of power optimization, we apply the algorithm in Section IV-A to each gate once, assuming input reordering does not change the transition density.

In the case of delay and power optimization, delay optimization is executed first for minimizing the critical path delay. After that, power optimization is processed under the delay constraint as shown in Fig. 3(a).



(a) Power with delay constraint (b) Delay

Fig. 3. Optimization Algorithm in Each Gate

V. EXPERIMENTAL RESULT

In this section, we show some experimental results. The circuits shown in TABLE II are used for the experiments. The maximum number of fan-in is four for the gates in the benchmark circuits. The transition density D at each gate is computed by logic simulation (Verilog-XLTM), the signal probability P is calculated using SBDD (shared binary decision diagram)¹.

TABLE II lists the result of power optimization without delay optimization. Power dissipation is computed by circuit simulation with HSPICETM using a 0.7 μ m process parameter. Input patterns are randomly generated with a signal probability of 0.5. The number of applied patterns is 100, which is the recommended number for the power estimation at circuit level by Ref.[11]. The same pattern is used for different configurations in input ordering of the same circuit.

The column "Initial" represents the power dissipation of the initial circuit. The columns under "Best" show the power dissipation of the circuits which are reordered for low power with the following two strategies.

A: The strategy which considers the dissipated power in the fan-in gates, the reordered gate and the fan-out gates when reordering (proposed).

B: The strategy which considers the dissipated power only in the reordered gate and the fan-out gates (similar to Ref.[4]). The column "Worst" represents the power dissipation of the circuit when the worst reordering is applied for high power considering the fan-in gates, the reordered gate and the fan-out gates.

The column "Improvement" shows the percentage of power reduction in the best reordered case from the initial circuit ($\frac{Initial - Best}{Initial} \times 100(\%)$). The column "Diff." explains the percentage of the difference between the largest and the smallest power dissipations ($= \frac{Worst - Best}{Best} \times 100(\%)$). The column "TIME" lists a CPU time for reordering on a SUN SPARC station 20. It does not include the time to calculate transition density by logic simulation.

From TABLE II, we can see that power dissipation of all circuits is reduced. The "Diff." column indicates that there is a possibility of reducing power dissipation by 17.2% maximum. The proposed method (Column "A") reduces power dissipation

by 3.6 % on average, and by 11.7 % maximum, whereas the conventional method, which considers the dissipated power only in the internal capacitances of the reordered gate reduces power dissipation by 1.8 % on average, and by 4.2 % maximum.

In TABLE III, the result of power optimization with delay optimization is shown. Our method reduces power dissipation by 3.1 % and delay by 7.0 % on average.

VI. CONCLUSION

We propose an improved method for power optimization of CMOS gates by input reordering. The dependence of input capacitance on the signal values of other inputs, as well as the possibility of charging/discharging internal capacitances, is utilized for the power reduction. The effect of the method is demonstrated experimentally using 21 benchmark circuits in a 0.7 μ m CMOS technology. The average reduction of power dissipation is 3.6 %. By input reordering there is a possibility that power dissipation is reduced by 17.2% maximum. In the case of delay and power optimization, our method improves delay by 7.0 % and power dissipation by 3.1% on average. Although the amount of improvement in power and delay is not drastic, input reordering can provide a steady improvement with almost zero penalty.

REFERENCES

- [1] A. Chandrakasan and R. W. Brodersen, "Low-power CMOS digital design," IEEE Journal of Solid-State, pp.473-483, April 1992.
- [2] Chi-Ying Tsui, Massoud Pedram and Alvin M. Despain, "Power Efficient Technology Decomposition and Mapping Under an Extended Power Consumption Model," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 13 No. 9 pp.1110-1122, September 1994.
- [3] Guangqiu Chen, Hidetoshi Onodera, and Keikichi Tamaru, "An Iterative Gate Sizing Approach with Accurate Delay Evaluation," Proc. of the 1995 IEEE/ACM International Conference on Computer-Aided Design, pp.422-427, November 1995.
- [4] Wen-Zen Shen, Jiing-Yuan Lin, Fong-Wen Wang, "Transistor Reordering Rules for Power Reduction in CMOS Gates," Proc. ASP-DAC'95/CHDL'95/VLSI'95, pp.1-6, 1995.
- [5] Razak Hossain, Menghui Zheng, and Alexander Albicki, "Reducing Power Dissipation in CMOS Circuits by Signal Probability Based Transistor Reordering," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 15 No. 3 pp.361-368, March 1996.
- [6] E. Musoll and J. Cortadella, "Optimizing CMOS Circuits for Low Power using Transistor Reordering," Proc. European Design and Test Conference ED&TC 96, pp.219-223, 1996.
- [7] S.C. Prasad, K. Roy, "Circuit Optimization for Minimization of Power Consumption under Delay Constraint," Proc. of the 8th International Conference on VLSI Design, pp.305-309, 1995.
- [8] F.N. Najm, "Transition Density, A Stochastic Measure of Activity in Digital Circuits," Proc. of the 28th IEEE Design Automation Conference, pp.644-649, 1991.
- [9] Bradley S. Carlson and Suh-Juch Lee, "Delay Optimization of Digital CMOS VLSI Circuits by Transistor Reordering," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 14 No. 10 pp.1183-1192, October 1995.
- [10] Malgorzata Marek-Sadowska and Shen P. Lin, "Pin Assignment for Improved Performance in Standard Cell Design," Proc. ICCD, pp.339-342, 1990.
- [11] Daniel Brand and Chandu Visweswariah, "Inaccuracies in Power Estimation During Logic Synthesis," Proc. of the 1996 IEEE/ACM International Conference on Computer-Aided Design, pp.388-394, 1996.

¹BDD Manipulator ver 6.03 : Copyright 1992 Kyoto University (by Shin-ichi MINATO)

TABLE II
POWER OPTIMIZATION WITHOUT DELAY OPTIMIZATION

Circuit	Initial (mW)	Best(mW)		Worst (mW)	Improvement(%)		Diff. (%)	TIME (s)	NO. of Gate	NO. of Tr.	NO. of Inputs
		A [†]	B [‡]		A [†]	B [‡]					
misex1	12.1	11.7	11.9	12.6	3.3	1.7	7.7	0.6	51	308	8
pcl	6.04	5.87	5.94	6.61	3.3	1.7	12.6	0.6	55	312	19
cm163a	12.1	11.9	11.7	12.3	1.7	3.3	3.4	0.5	63	300	16
pcler8	6.30	6.05	6.10	7.09	4.0	3.2	17.2	0.6	63	376	27
cm162a	12.4	12.2	12.0	12.4	1.6	3.2	1.6	0.6	68	320	14
pm1	13.0	12.7	12.8	13.5	2.3	1.5	6.3	0.6	75	372	16
cm85a	11.7	11.5	11.6	12.0	1.7	0.9	4.3	0.6	76	328	11
rd53	13.0	12.7	12.9	13.7	2.3	0.8	7.9	0.5	76	496	5
sqrt8	16.6	16.4	16.5	17.5	1.2	0.6	6.7	0.5	81	530	8
cc	17.6	17.0	17.3	18.8	3.4	1.7	10.6	0.6	81	424	21
misex2	14.9	14.3	14.7	15.3	4.0	1.3	7.0	0.6	98	592	25
5xp1	32.4	31.1	32.3	34.1	4.0	0.3	9.6	0.6	158	1070	7
C432	21.8	21.0	21.1	22.8	3.7	3.2	8.6	1.3	171	802	36
count	11.0	10.5	10.7	11.1	4.5	2.7	5.7	0.6	176	828	35
sao2	21.3	18.8	20.4	21.9	11.7	4.2	16.5	0.5	188	1332	10
lal	21.1	20.5	20.9	21.9	2.8	0.9	6.8	0.6	189	970	26
C499	64.3	63.2	63.7	65.9	1.7	0.9	4.3	1.8	218	1572	41
b9	28.0	27.6	27.6	28.3	1.4	1.4	2.5	0.5	234	936	41
9sym	31.8	29.9	31.4	33.7	6.0	1.3	12.7	0.5	251	1776	9
rd73	24.2	22.5	23.9	25.4	7.0	1.2	12.9	0.8	277	2008	7
ALU	31.8	30.4	30.9	33.0	4.4	2.8	8.6	0.8	279	1302	21
Avg.	-	-	-	-	3.6	1.8	8.3	-	-	-	-

A[†] : Proposed Method B[‡] : Conventional Method

TABLE III
DELAY AND POWER OPTIMIZATION

Circuit	Delay(ns)			Power(mW)			TIME (s)
	Initial	Reordered	Improvement(%)	Initial	Reordered	Improvement(%)	
misex1	2.10	2.01	4.3	12.1	11.9	1.7	1.2
pcl	3.24	3.17	2.2	6.04	5.94	1.7	1.2
cm163a	3.28	2.95	10.1	12.1	11.7	3.3	1.2
pcler8	3.99	3.65	8.5	6.30	6.10	3.2	1.4
cm162a	3.39	3.07	9.4	12.4	12.0	3.2	1.3
pm1	2.11	1.91	9.5	13.0	12.8	1.5	1.3
cm85a	3.45	3.23	6.4	11.7	11.6	0.9	1.3
rd53	2.44	2.30	5.7	13.0	12.9	0.8	1.4
sqrt8	2.61	2.39	8.4	16.6	16.5	0.6	1.3
cc	2.46	2.25	8.5	17.6	17.3	1.7	1.3
misex2	2.03	1.87	7.9	14.9	14.3	4.0	1.3
5xp1	2.79	2.55	8.6	32.4	31.1	4.0	1.8
C432	8.30	8.04	3.1	21.8	21.6	0.9	2.8
count	9.47	8.75	7.6	11.0	10.5	4.5	1.8
sao2	3.76	3.53	6.1	21.3	18.9	11.3	2.0
lal	4.01	3.73	7.0	21.1	20.6	2.4	2.0
C499	5.24	4.98	5.0	64.3	64.5	-0.3	4.4
b9	4.75	4.36	8.2	28.0	27.6	1.4	2.6
9sym	3.81	3.63	4.7	31.8	29.9	6.0	2.7
rd73	4.02	3.74	7.0	24.2	22.5	7.0	3.0
ALU	6.14	5.63	8.3	31.8	30.4	4.4	3.1
Avg.	-	-	7.0	-	-	3.1	-