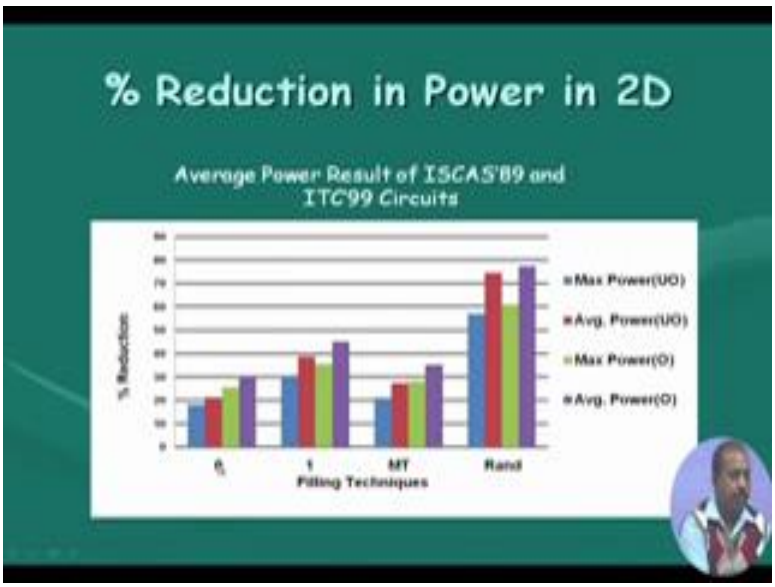
Digital VLSI Testing
Prof. Santanu Chattopadhyay
Department of Electronics and EC Engineering
Indian Institute of Technology, Kharagpur

Lecture – 39
Thermal Aware Testing (Contd.)

(Refer Slide Time: 00:25)

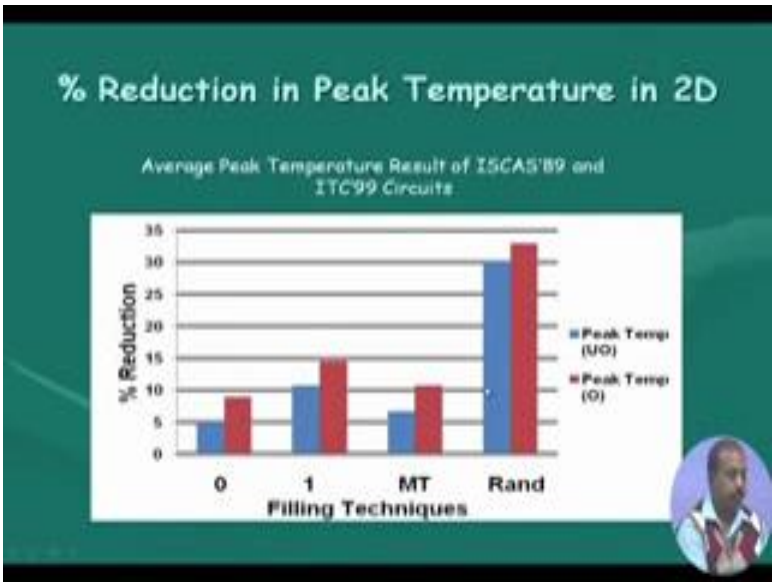


To get the result of this flipping algorithm which apparently seems to be very simple, but it has been observed that it gives relay reasonably good result. So, these are actually the plots of over and average plot of this ISCAS 89 and ITC 99 benchmark circuits. Now, if you use this is 0 filling technique then you see that this max with respect to this 0 filling pattern say, this reduction in maximum power is about 19 percent; average maximum power 19 percent, average power is slightly more than 20 percent etcetera. Over 0 filling, so these are the improvements; over one filling, so these are the improvements of the algorithms. So, you see that they have there are lot of scope of improving this even the power value when we are doing this bit flipping, so checking the bit flips.

So, empty filled improvement are there, but you see for 0 fill and empty fill, so these 2 cases improvements are not that high. Because empty fill actually it reduces those scan transitions and 0 fill in most of the cases the logic gates that we have if we apply a one of the input is 0 the output does not change. So, that is why this transitions are less and the flip-flops they normally also have if it is 0 then it is not going to have lot of transitions,

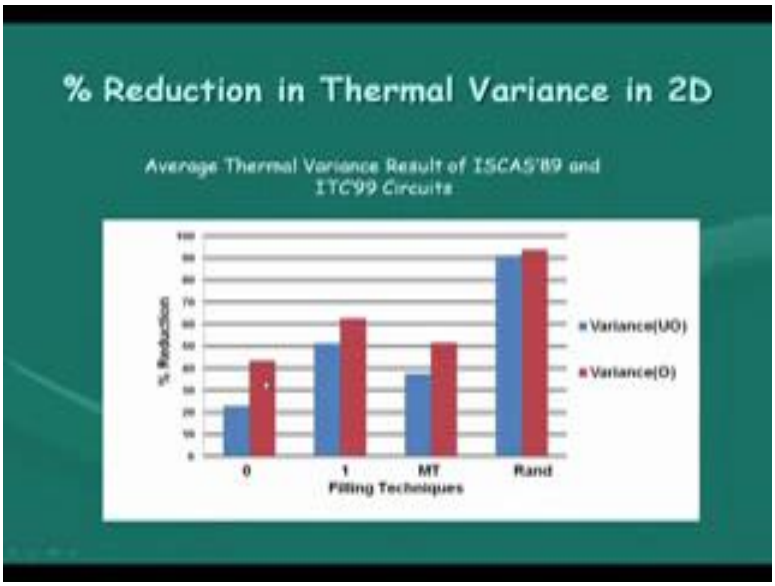
so this way this is 0 fill and empty fill the transitions are less. Of course, this is very much logic family specific may be the work that the technology family that has been used in getting these results, so they are 0 is consuming less power than 1. For random of course, there is lot of improvements that is there.

(Refer Slide Time: 02:02)



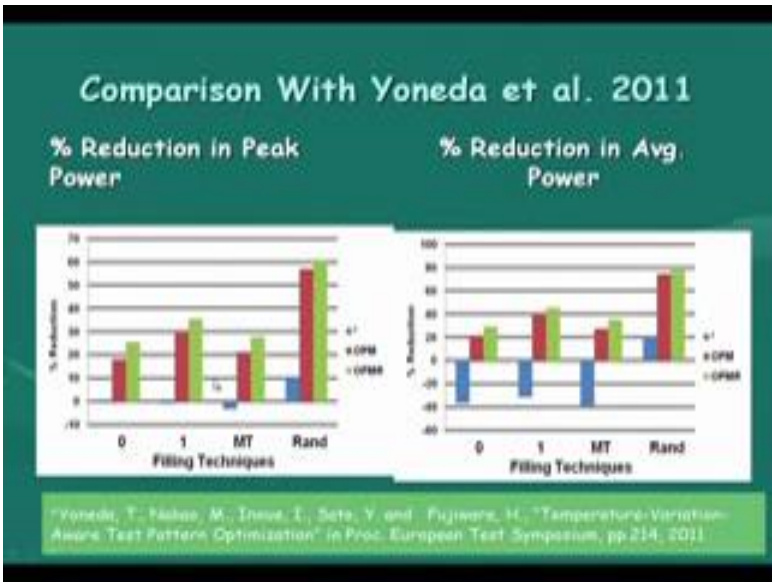
Temperature reduction results you see that though we have not directly try to minimize the temperature, but our power minimization was much more focused power minimization was focused for a particular block it is and it is neighboring block, so that we are looking into and that results in good amount of temperature reduction. So, peak temperature reduction, so this was 5 percent over 0 fill, when the pattern set is unordered, so do not cares are filled, but the ordering of the pattern is not test pattern ordering is not done. Now, this is when the test pattern ordering has been done that is the combined effect of do not care filling and test vector reordering. So, similarly for one filling, so you have got about 10 percent about 15 percent improvement, so this way we can see that filling plus reordering that can give that can result in much better power this temperature saving.

(Refer Slide Time: 03:03)



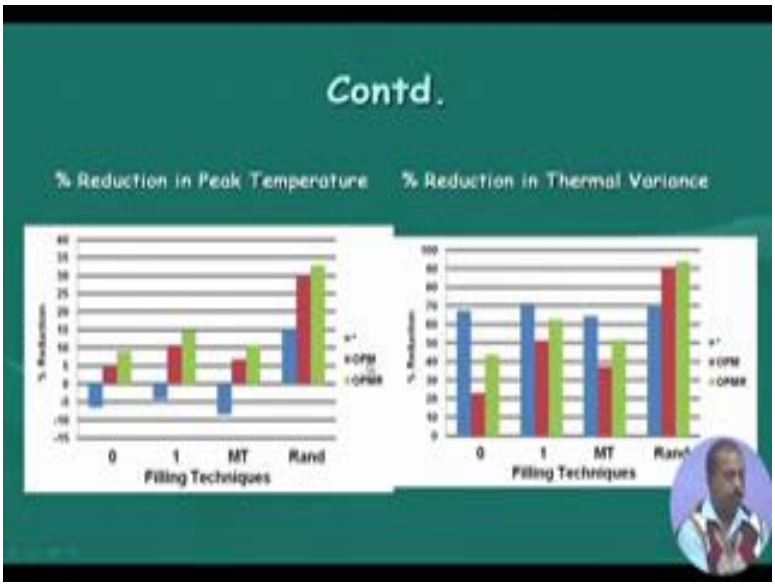
To temperature variance, so that is also another parameter when we are trying to do this thermal minimization, because if the temperature variation across the chip is not uniform then different regions of the chip are differently heated, as a result their delays may be different, so the delay faults may create some problem. So, they may not be detected properly. So, this variance you see that variance reduction is there, so with respect to the unordered set, so this is the reduction in variance. And when you do an ordering of the test resulting test pattern set as well apart from do not care filling then the variance reduces a more than 50 percent, so that is a good saving. So, with random filling you see it, it reaches more than 90 percent saving so; that means, the chip heating pattern will be very uniform, so this delay faults missing the delay faults will be that probability is very less.

(Refer Slide Time: 04:01)



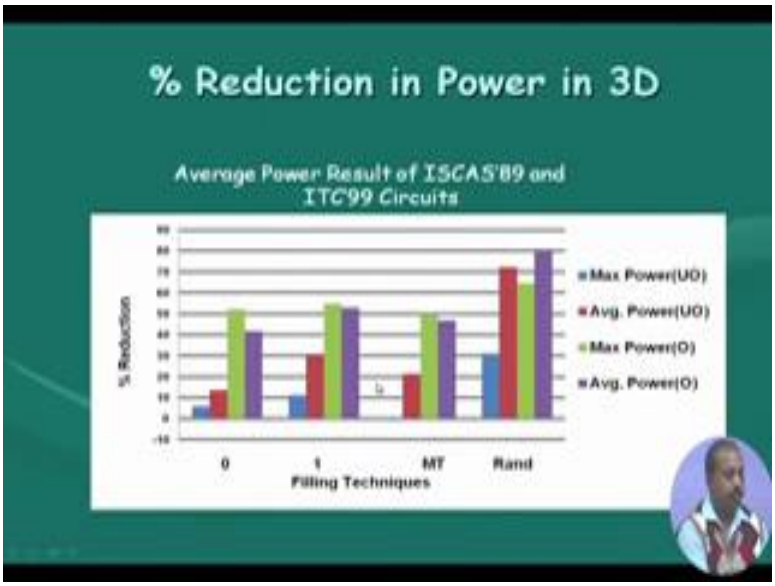
This is the comparison with some previous published work 2011.

(Refer Slide Time: 04:06)



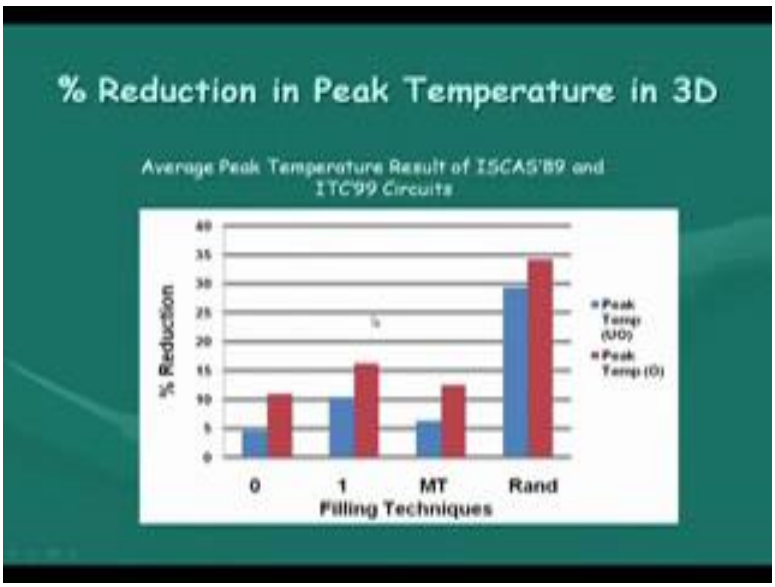
And this is the percent with peak temperature reduction compared to different filling techniques like. So, this is output the peak measurement and this is the reduction part of it.

(Refer Slide Time: 04:23)

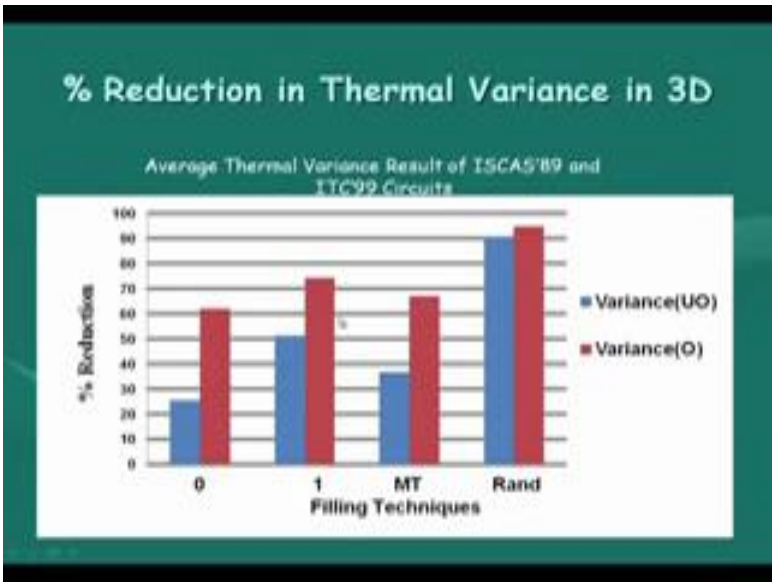


For 3D also, so there are reductions just like 2D, this is the temperature reduction in 3D this is the variance reduction when 3D.

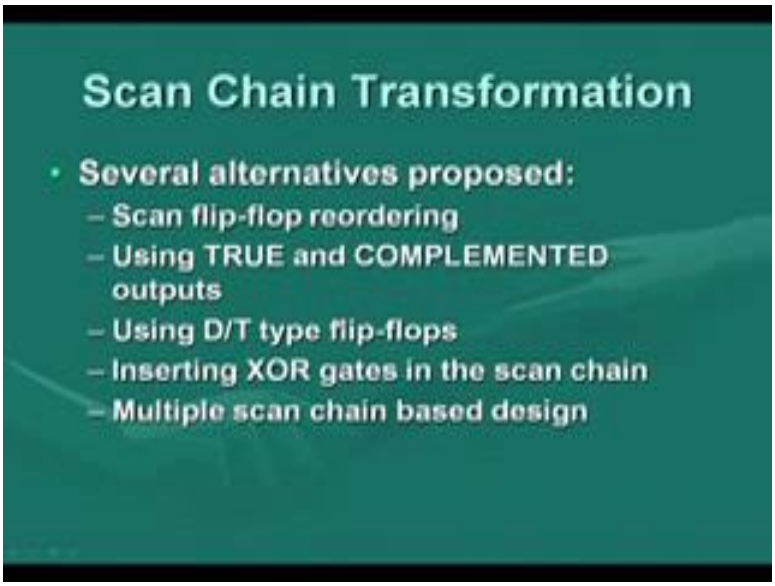
(Refer Slide Time: 04:29)



(Refer Slide Time: 04:31)

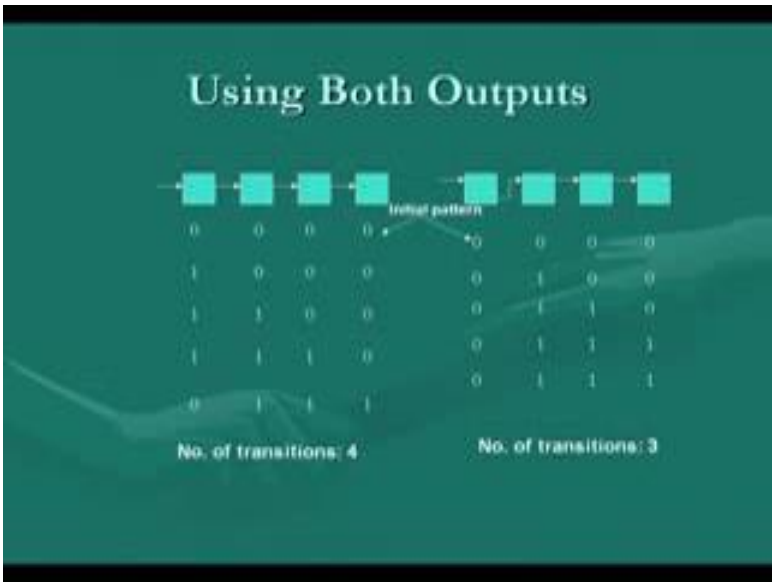


(Refer Slide Time: 04:33)



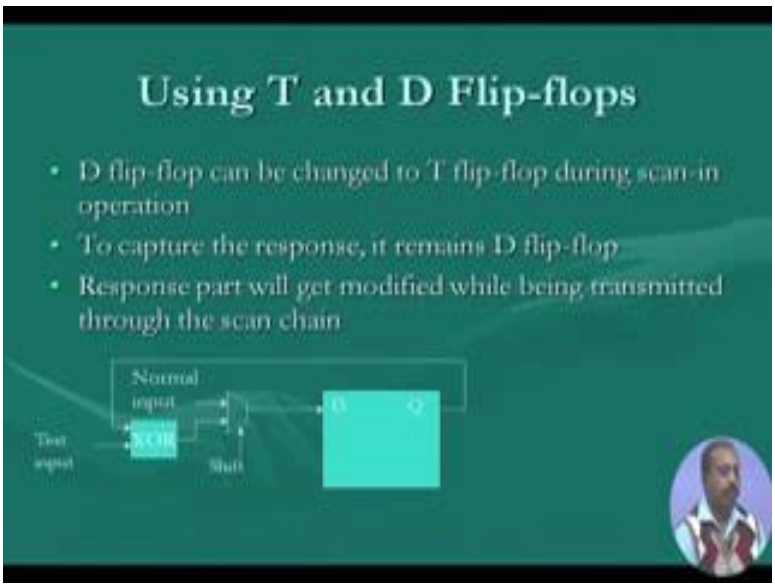
Next, we will look into the scan chain transformation techniques. So, while discussing on this power management techniques we have seen that there are some approaches for minimizing the scan power by flip-flop scan flip-flop reordering, then true complimented outputs, D-T type of flip-flops, inserting XOR gates into scan chain and multiple scan chain based design, so each of them have got effect on the thermal part as well.

(Refer Slide Time: 05:03)



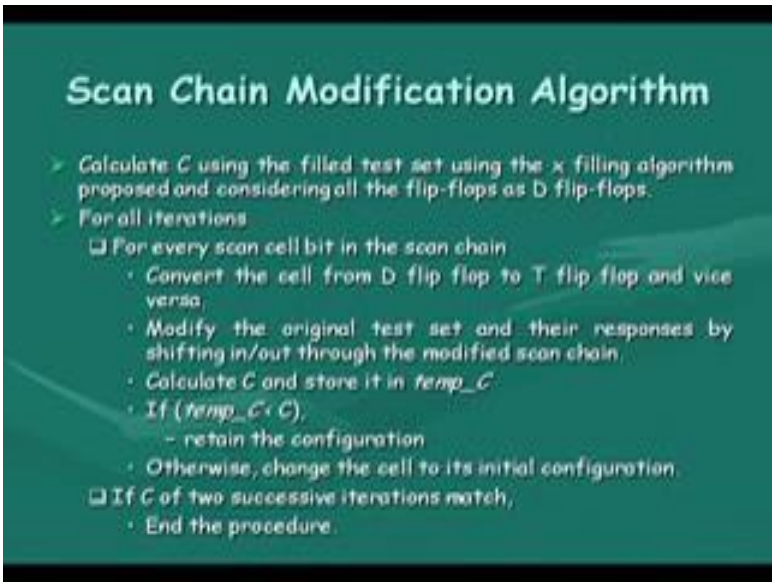
So, using both outputs, so you have seen this example previously that number of transitions gets reduced.

(Refer Slide Time: 05:08)



Similarly, if you use T and D flip-flops, then flip-flop at the time of testing converted to T flip-flop; and this also results in the good number of transition reduction.

(Refer Slide Time: 05:22)



So, scan chain modification algorithm, so what we do, so this is basically that D-T flip-flop modification, so how can we select some of the flip-flops to be converted from D to T, so that we can get temperature improvement. So, first we calculate the criticality of the blocks the max maximum criticality C using the filled test set that has got this do not care filling algorithm that has been proposed here, the flip select algorithm. And assuming that all flip-flops are D type flip-flops, so that is basically the original circuit where all flip-flops is D type flip-flop; and all scan flip-flops are D type flip-flops. And the test pattern ser is filled using the do not care filling technique that we have just discussed previously and we calculate the criticality value the cost functions C which is the maximum criticality value.

Now, so now we do this step. For every scan cell bit in the scan chain we convert the cell from D type to T type or if it was previously T type we convert into D type. So, initially all flip-flops are D type, so we take the first cell convert it into T type. So, as a result, the original test set and their responses will get modified and this modified test pattern and responses they will be shifted in and out this modified scan chain. So, that way the test patterns set will get modified.

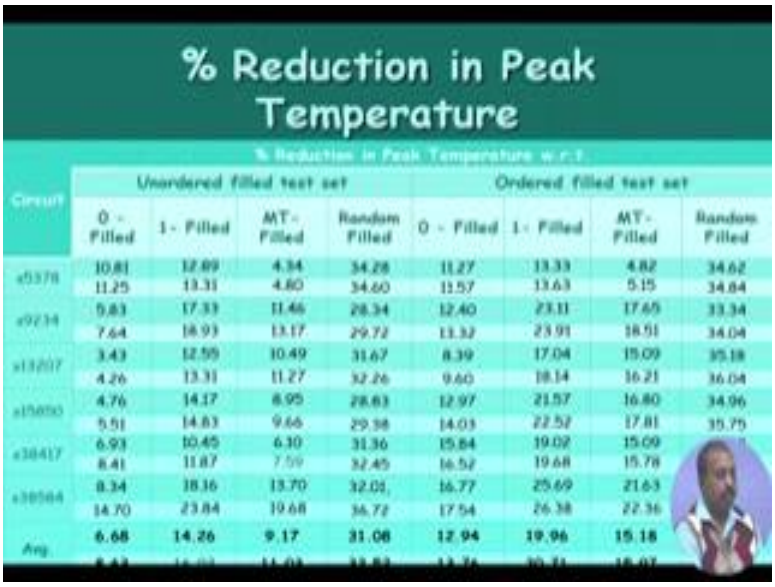
Then you calculate the c value and store it in the temp C variable. So, what was happened is due to this modification the test pattern that needed to be shifted through the scan chain will get modified, and the response bits that will get in the scan chain that will

also get modified. So, the C value that we had computed with respect to the original test pattern set is no more valid, so we need to re-compute it. And those recomputed values are stored in temp C. And if temp C is better than C, so criticality has improved then we retain the configuration, so it is retaining the configuration, so this particular flip from D to T is accepted otherwise we go back to the original configuration. So, the conversion T flip-flop is nullified and it goes back to the D flip-flop.

Now, this process is again repeated. So, this is repeated for every scan cell bit, so if i have got 100 scan cell then thus then this loop is repeated 100 times. After that we start the next iteration of this loop next iteration of this outer-loop. Now, again we try to flip the flip-flop type. So, previously if this flip-flop was converted from D to T now we try taking it back from T to D and why this can improved, this can have an improvement because some other flip-flops have also change, so that may be with respect to them it may be better that this flip-flop was retained to the D value. So, that way if it is done then we can we will get back that configuration, so that is the idea. So, we convert it back from T to D now may be, so this way another iteration is done.

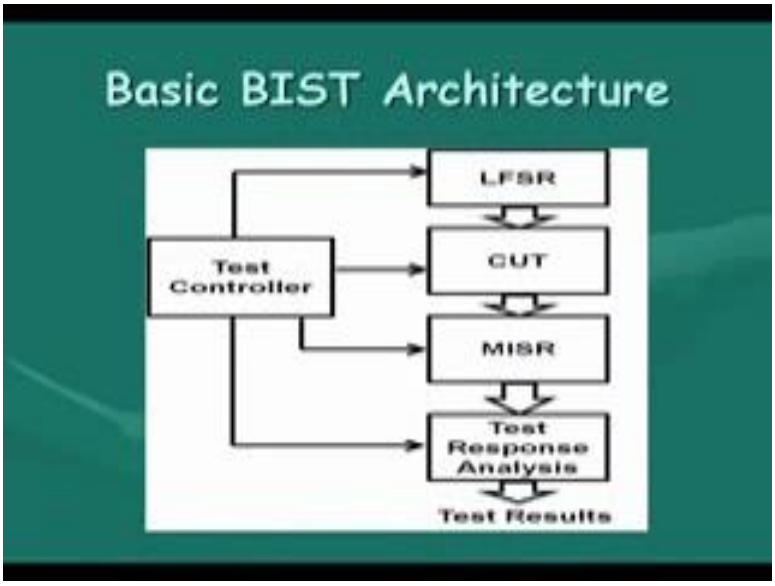
So, if you in 2 successive iterations if you find that there is no improvement that means, there is nothing possibly we cannot do anymore hill climbing to get the better method better result, so it is stop at that point. Again, this is a very greedy algorithm, but it often gives good result.

(Refer Slide Time: 08:56)



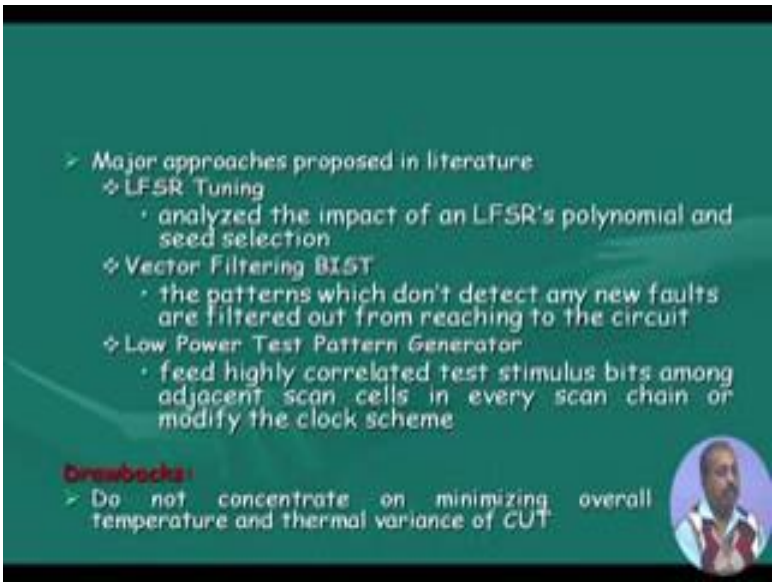
So, this is actually the percentage reduction in peak temperatures. So, these are the various circuits that we have the ISCAS 89 circuits. With respect to this 0 filled test pattern set, so we have got this 10.81 and 10.81 percent reduction in the peak temperature. And similarly this with respect to one filled, we have got 12.89, so this shows various improvements and with respect to ordered filled also there are improvements. So, next we look into the internal testing technique. So, internal testing technique means we have got this test pattern generator and this response analyzer inside the chip, so we have got this basically the BIST type of environment, the built in self test type of environment.

(Refer Slide Time: 09:53)



So, basic BIST architecture is like this as we know that this is the circuit under test, so there is one LFSR, so LFSR is used as the test pattern generator that is fed to the circuit under test. And then this circuits output goes to a multiple input signature register - MISR, so they that is time and space compacted. And then this response is analyzed and when this response is analyzed, so we get the test result whether it is true or not. And this test controller is actually controls all the modules. So, this is the basic BIST architecture that we have for the testing.

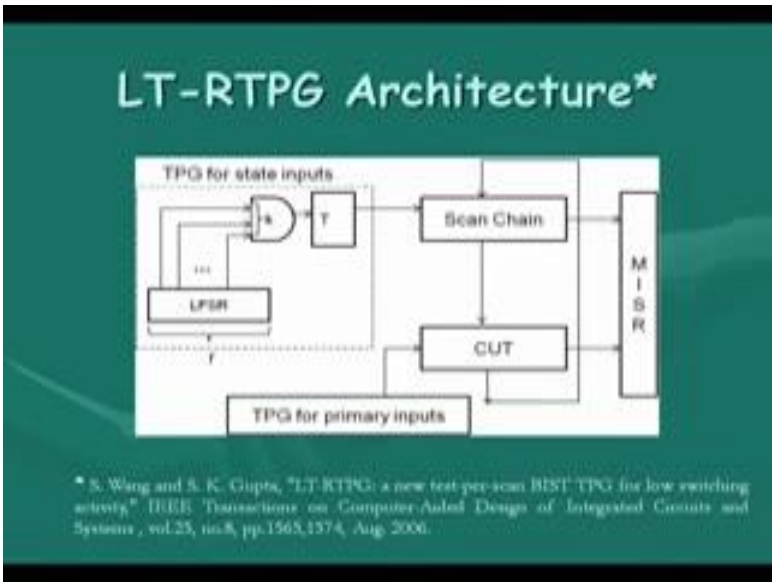
(Refer Slide Time: 10:30)



Now, this architecture maybe modified, and so that we can get some thermal improvement both power and thermal improvement. So, major approaches that have been proposed in the literature are like this LFSR tuning. So, we try tune the LFSR, so we analyze the impact of LFSRs polynomial and seed selection and based on that we know that the LFSR will behave differently as a result it will generate different set of patterns, and possibly it will have different effect on this power consumption as well as the thermal situation. Of course, it will also affect fault coverage that has to be taken.

Vector filtering we have seen previously that is test pattern generators which do not detect hum any new fault, so that are excluded from the test patterns set, so that can also be used. And this low power test pattern generation, so we feed highly correlated test stimulus bits among non adjacent scan cells among adjacent scan cells every scan chain, so this is another possibility that is if 2 successive bits are very much similar to each other then it can generate a low power test sequence, so this is also done. But they do not contribute to the overall peak temperature reduction and thermal variance reduction.

(Refer Slide Time: 11:52)

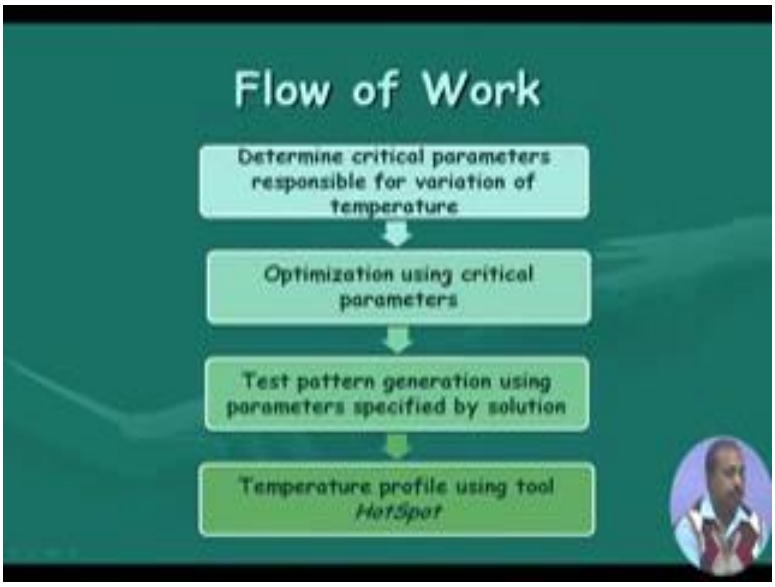


So, another architecture which is known as LT-RTPG, so this is a test patterns scan tape BIST test parts can BIST test pattern for low switching activity. So, this tries to minimize the switching activity. So, what is done this LFSR a number of inputs of them, so they are inputted to a k-input and gate; and this k-input and gate, so this is actually general it is fed to the T flip-flop. So, whenever all these inputs are 1, whenever all this LFSR inputs are 1 then only this T it will be 1. As a result this input that we are feeding to the scan chain, so that is going to change. So, ideally we should take say the LSB of a LFSR and connect it serially to the scan chain, so whenever this LSB switches, so a scan transition is introduced and with that when a number of bits have been shifted into the scan chain, so that is the test pattern.

Now, as soon as you make it a k-input AND gate that and that is controlling this T flip-flop, so possibility of introducing a transition becomes much less. So, that way this scan chain, it will have very less number of changes a very less number of toggling in its input, so that as a result this at the patterns that we are feeding to the circuit, so this is basically going to be less. So, for the primary inputs, so there may be a test pattern are generated which feeds the primary input part, but for the scan chain part, so this transitions are less in the scan chain. So, this test pattern generator, so it may be working in such a fashion that it uses this test patterns, the primary input part of the test pattern very judiciously and, so that it does not have to depend much the scan chain inputs. But scan chain inputs are mostly kept similar, so that is that is the way this circuit operates the LT-RTPG

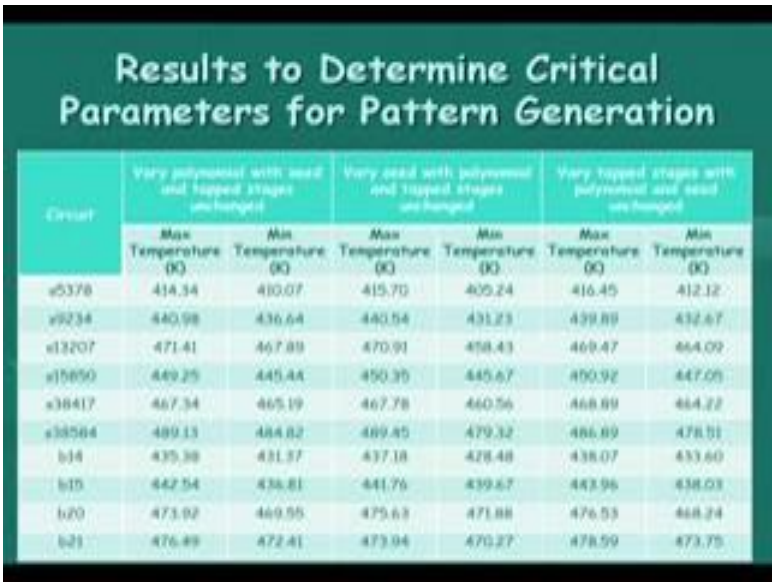
architecture works.

(Refer Slide Time: 13:54)



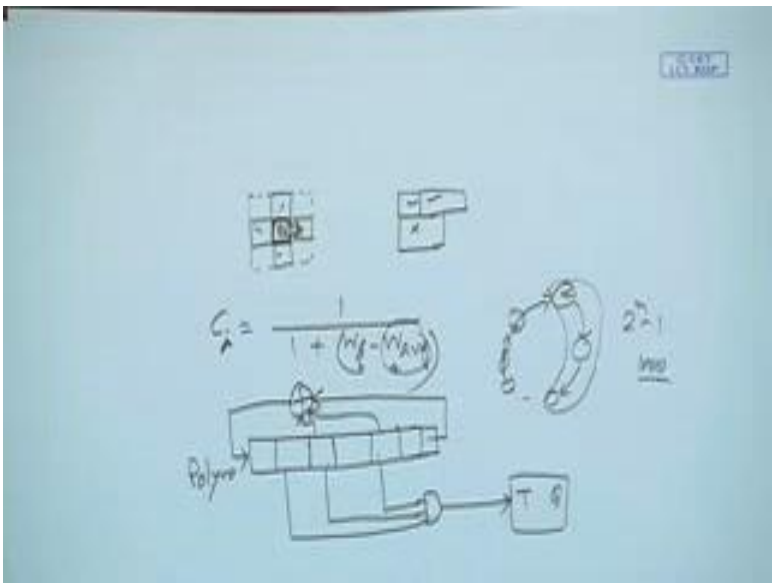
So, if you want to use this for temperature minimization, so first of all we determine critical parameters that are responsible for variation of temperature. So, first we need to know like what are the parameters that are going to affect the temperature variation in the circuit. Then we optimize using critical parameters, so we optimize using the critical parameters like after this study, we will know, what are the parameters to be optimized? And based on that we do perform an optimization based on critical parameter then this test pattern generation has will be done using the parameters specified by the solution. So, this solution will tell us, what are the typical critical settings for the LFSR? And based on that we can do this thing, now, once we have done this test pattern generation, we can do a thermal simulation using the say the tool hotspot or similar such thermal simulation tool to get the temperature profile of the resulting situation.

(Refer Slide Time: 14:57)



So, the initial experimentation that has been done, so that has for we do many things like very polynomial with seed and tapped stages unchanged.

(Refer Slide Time: 15:11)



Like if you have an LFSR, so LFSR there are various issues like so what is if this is the LFSR then the one thing is the polynomial. What is the polynomial that we are using for feeding back to the first cell? So, there is an XOR gate, and this XOR gate actually feeds back the first cell. And what are the cell position that you are tapping in, so this will definitely come and you it may be that it take we tap out these 2 positions and take them

into this LFSR, so that will tell us what is the characteristic polynomial. So, if the characteristic polynomial is changed then it will behave differently, so it will generate different set of patterns.

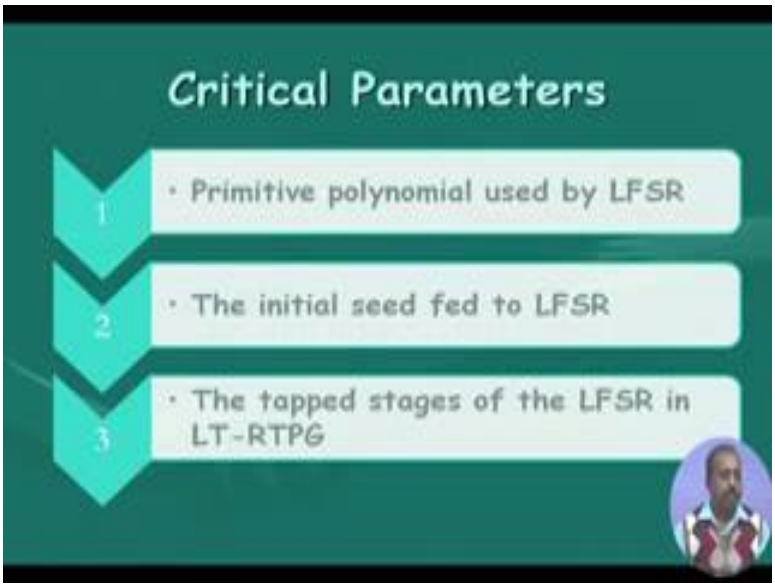
Second important thing is that varying the seed patterns, so particularly if the LFSR is of maximum length then varying the seed pattern, it will generate patterns in different parts of that as of its state space. So, as a result the pattern generated will be different. So, this goes through a number of cycles maximum length LFSR, so it will go through a number of all the 2 power n minus one states in its state transition accepting all 0. So, if you start here and run for 1000 cycles you will get some set of patterns; if you start at this point and run for 1000 cycles you will get this set of patterns, so all this patterns will be generated. So, that way depending upon the seed value also the sequence generated will vary as a result the fault coverage power consumption temperature everything will change, so that is another parameter.

Third parameter that we are looking for is, so the way we did it is that this was feeding one AND gate, and this AND gate we were tap from some places maybe these 3 places we tap and then it was fed to the T flip-flop in the LT-RTPG scheme, so this was fed to the T flip-flop. Now, from which positions we are tapping, so that is also having an effect. So, we just try to see which affect is more critical look like if we do and analysis like this s5378 circuit, we run with the different we run it several times by varying the polynomial with other 2 factors remaining same the tapped seed position and this seed and the tapped stages, so they remain keeping them and unmodified. So, we just vary the polynomial of the LFSR. And see what is the maximum temperature and minimum temperature that is achieved for these particular circuits.

So, you see there is a variation about 4 degree, similarly if you vary the seed with polynomial and tapped stages unchanged, so this gives rise to variation of 15 degrees. And if you change the tapped positions, feeding the AND gate and the polynomial and the seed they are kept unchanged, so this gives a variation of 4 degrees. So, this in general, so this for example, b to 1 circuit, so this gives a variation of 5 degree, this gives a variation of 3 degree, this gives a variation of 4 degree. So, though it apparently seems that 2 degree, 3 degree, 4 degrees saving, so how great it is, but it is also mean useful because you see it is the question of the chip getting damaged and all that. So, it is better that we even is the value is saving is not that high in temperature maybe if we run it for a

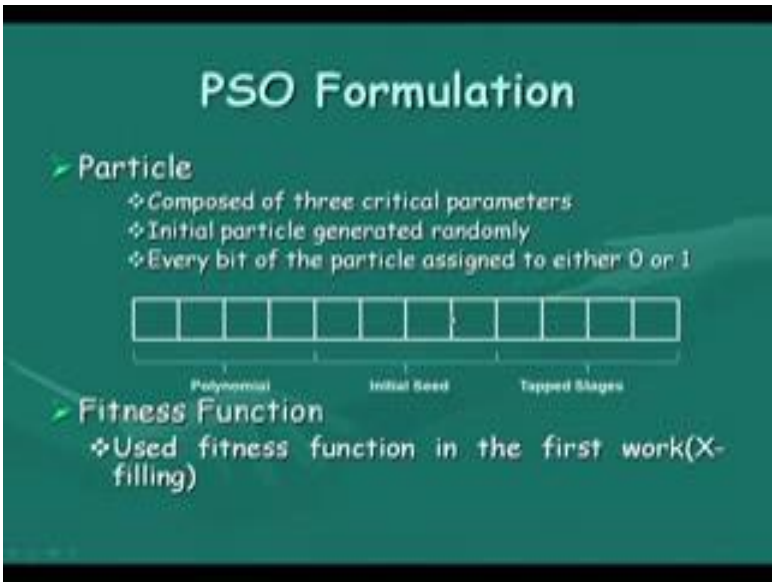
longer time the temperature will build up; and that way it is going to be destructive. So, you see that if there is no clear cut distinction between which one is more important, so possibly we have to optimize all of them to get good temperature aware method.

(Refer Slide Time: 19:02)



So, these all the 3 they turned out to be critical parameter. The primitive polynomial used by the LFSR, the initial seed that we feed to the LFSR, and the tapped stages of LFSR, so all of them become critical.

(Refer Slide Time: 19:20)



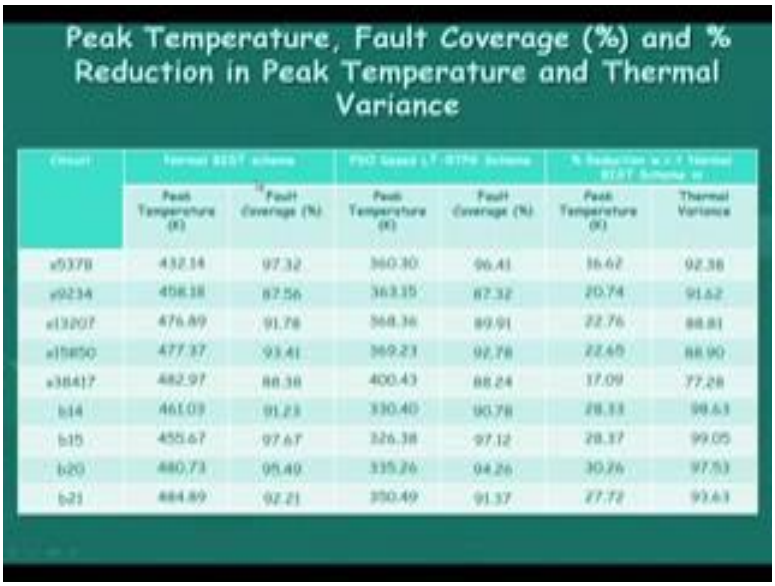
So, when all of them become critical, so how to work. So, again we use a particle swarm

optimization based formulation. So, in this case, if I have got this see the LFSR primary the LFSR is of n bit size then the first n bits of the particle it represents the polynomials. So, if this bit is 0 that means, the position is not used in the feedback function; if the bit is 1 that means, the position used in the feedback function. So, that way since one characteristic polynomial will tell us what will be the connection pattern of the LFSR, so these are the various first n bits this is that will tell us what is the connection pattern. Next part, next I need to tell what is the initial seed and again there are n bits, so this n bit values can be fed here, so this is the initial seed value. So, the initial seed value will be having the proper one. So, this will be having this what with feed value the LFSR will start functioning.

And then what are the tapped stages tap stages feeding the AND gate. Again if it is an n bit LFSR, so I need n positions here, so the bit being 0 will mean that the position is not connected to the AND gate was bit being one means that the position is connected to the AND gate. So, by this structure by this particle we can represent a complete solution. The characteristic polynomial that is chosen for the LFSR, the initial seed value and the tapped stages that are there, so all of them can be utilized, all of them can be specified.

So, initially the particles are generated randomly for the first generation, so they are generated randomly. So, then this every bit of the part, so then this the general fitness has to be evaluated, so fitness is the same as the x filling work that is the criticality value c that we have used, so that criticality value can be may utilized. And then based on that this formulation evolves, and then whatever comes as the final solution so we can after a number of generations that we have done so we can see, what is the best solution that is produced?

(Refer Slide Time: 21:39)



Circuit	Thermal BIST scheme		PSO based LT-RTPG Scheme		% Reduction w.r.t. Thermal BIST Scheme in	
	Peak Temperature (K)	Fault Coverage (%)	Peak Temperature (K)	Fault Coverage (%)	Peak Temperature (K)	Thermal Variance
s5378	432.14	97.32	360.30	96.41	16.62	92.38
s9234	458.18	87.56	363.35	87.32	20.74	91.62
s13207	476.89	91.78	368.36	89.91	22.76	88.81
s15850	477.37	93.41	369.23	92.78	22.65	88.90
s38417	482.97	88.38	400.43	88.24	17.09	77.28
b14	461.09	91.23	330.40	90.78	28.33	98.63
b15	455.67	97.67	326.38	97.12	28.37	99.05
b20	480.73	95.49	335.26	94.26	30.26	97.53
b21	484.89	92.21	350.49	91.37	27.72	93.63

So, if you do that way, so this peak temperature variation, fault coverage and etcetera if you look into those parameter in a normal BIST scheme, where there is no such choice of this polynomial and seed and all that, so you see that this 5378 circuit, so this original temperature of 432.14 degree Kelvin, fault coverage was 97.32. Now, if you use this PSO based LT-RTPG scheme that we have just discussed, so peak temperature reduces to 360.30 degree, and fault coverage also gets reduced. Peak temperature reduction is 16.62 per percent and this thermal variance reduction is 92.38 percent. So, you see that there is a huge improvement in the thermal variation. There is also considerable improvement in the peak temperature part. Of course, fault coverage reduces because of the thing that the emphasis more on this criticality factor which tries to minimize the temperature part rather than the fault coverage part, so fault coverage gets suffered.

So, this can be handled, this can be taken care of by taking help of the this running the PSO further for more number of generation or making the test; making the LFSR to run for more number of cycles. Actually these results are obtained by running the LFSR for 1000 cycles. So, if we run it for more number of cycles and more patterns will be generated and the fault coverage value will improve, so this will be and the peak temperature values will worst, so this will becomes slightly worst than what we have got here. But still we can get saving in the peak temperature may not be say as high as this one, but at least 10 to 12 percent improvements can be obtained. And the fault coverage values can be raise to the original values, so that way this BIST approach. So, we can

choose this polynomial, we can choose this seed and we can choose the tapping positions and use LT-RTPG type of scheme for reducing the thermal variation and peak temperature.