ANALYSIS OF THE BUMP PROBLEM IN BSIM3 USING NOR GATE CIRCUIT AND IMPLEMENTATION OF TECHNIQUES IN ORDER TO OVERCOME THEM

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

SUBRAMANIAM SANKARALINGAM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

PROFESSOR TED HIGMAN

December 2014

Acknowledgements

First I would like to thank the Almighty for helping and guiding me in each step of this work and for making me believe and his blessings helped me in finishing the thesis work perfectly in time.

I would like to express my gratitude to my advisor Prof. Ted Higman for his constant support and encouragement starting from selecting the topic and through the entire period of the work. He helped me setting specific target for each week and the following week we ensured that the target was achieved and we also used to review the idea with the results. This meticulous planning ensured us to finish off the work in time. He also took great efforts in reviewing the report to ensure that we have covered all the areas and always had some suggestions to make it look perfect.

I would also like to thank Prof. Chris Kim and Prof. Adam Rothman for being part of the thesis committee and also for their views and insightful comments during the review.

Finally I would like to thank my mother and dad who always supported and stood by me no matter when I called them and had words of encouragement each time I spoke with them.

Dedication

This thesis is dedicated to my mother and dad for without whom I wouldn't be here.

Abstract

In this paper we will be analyze the bump problem in BSIM3 using “Killer” NOR gate circuit. We refer to the NOR gate circuit as “Killer” gate because it kills our simulation results. This problem is witnessed in all the models employing quasi-static approximation. Quasi-static approximation and Non quasi-static approximation will be explained in detail to give better insight into the bump problem. Finally some techniques will be proposed and analyzed with the help of waveforms in order to overcome the problem.

Table of Contents

List of Figures

INTRODUCTION:

BSIM refers to the Berkeley short channel IGFET models which are compact (fast converging) models used for circuit simulation. These models are used to capture the current versus voltage behavior of the transistor. Currently there are different versions of BSIM models like: BSIM3, BSIM4, BSIM6 and so on. No model is fool proof and every model has some scenarios where they fail to model as desired.

For our study we have taken the case of BSIM3 model which is the most widely used model in both educational institutions and also commercially. As we mentioned earlier, there are many areas where the functionality of BSIM3 model can be improved which will help the users like us to better analyze the results. Now we will take one such problematic case in BSIM3 and will look into that in detail.

One of the major issues in BSIM3 is the bump problem. This bump problem is common in all models employing what is known as the quasi-static approximation and is not just a BSIM3 problem. To an extent it also depends on the way the channel charges are assumed to be distributed between the source and the drain commonly referred to as the charge partition scheme. Let's first try to get a better and in depth understanding of the bump problem.

This will require us to get a handle on the types of modelling in BSIM3,which can can broadly classified into two categories: non quasi-static modelling and quasi-static modelling. The quasi-static approximation assumes that as the terminal voltages vary with time, the charges in the channel arrange themselves instantaneously. For example if

we apply a voltage to some circuit (say an inverter) which switches from low to high, the quasi-static model assumes that a steady state charge carrier profile is achieved at the same instant that input voltage switches from low to high. This is where the transit time of the charge carriers comes into play. In case of MOSFET, transit time is the time taken for the charges to move from source to drain or vice versa. When the time scale of switching is more than the transit time we don't have a problem in the quasi-static approximation. But in the reverse situation when the transit time is more than the switching time, the model assumes that the steady state is attained in such a short span of time but we very well know that it is not possible. Thus we can conclude that the quasi-static approximation is incorrect in this scenario when the transit time is more than the time scale of operation.

The NOR gate is one such circuit which can highlight the bump problem that was explained in the previous section. The only reason that this circuit was chosen is to replicate the problem in Hspice and in turn suggest some ways to overcome them.

A schematic of the "Killer" NOR gate circuit is shown below:

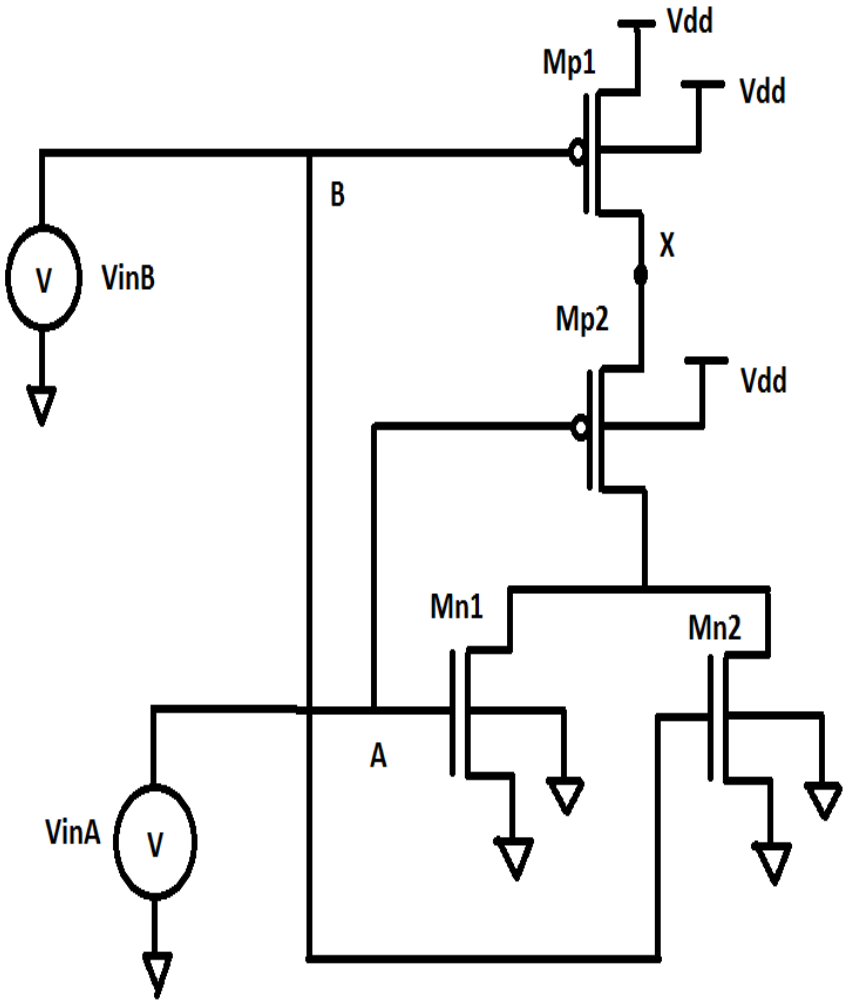

Fig. 1: Schematic of NOR gate circuit

ANALYSIS OF THE BUMP PROBLEM:

The circuit consists of 2 NMOS transistors in parallel and 2 PMOS transistors in series. Let us shed some light on the dimensions of the transistors. We have chosen TSMC

0.18micron technology for simulation purposes. In order to have equal rise and fall time the width of PMOS transistors must be 4 times that of the NMOS transistors. We have chosen minimum dimensions for both NMOS. Hence each of the NMOS transistors has a length of 0.18u and width of 0.36u and PMOS transistors have a width of 1.44u and length of .18u. The only exception in this case is that the top transistor has a length of 5u. This is essential because only beyond this length we see the voltage dropping to unphysical negative value resulting in a bump in the initial part of the transient. We have considered an ideal voltage source of 5v.

We apply 2 pulses, VinA and VinB as inputs and both of them have a rise time and fall time of 5ns and there is a delay of 25 ns between both the pulses. Irrespective of the charge partition scheme that we choose, it is important to note that output voltage V(Q) is comparable in all of them. However the main problem is with the voltage at node X , which reaches a highly unphysical negative value of 7V and 9V with 40/60 and 50/50 charge partition schemes.

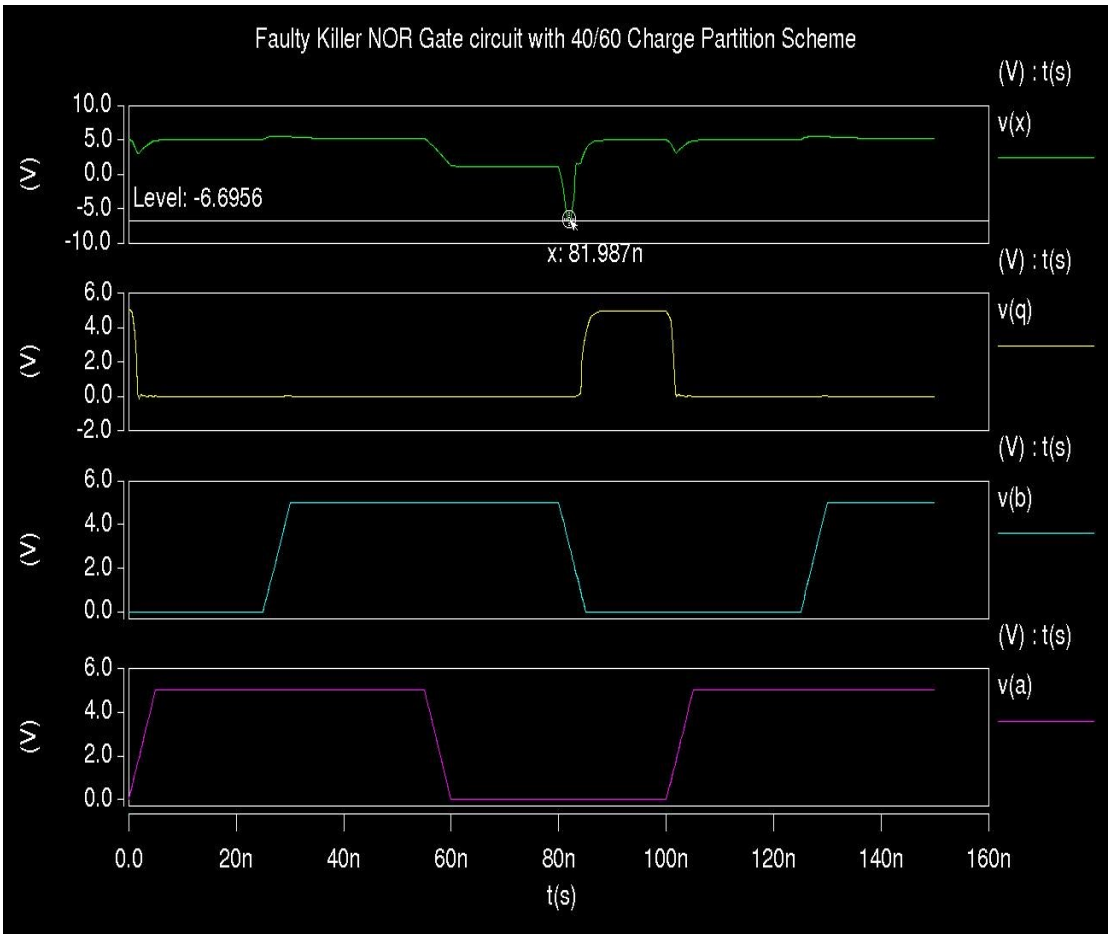

Fig. 2: Faulty NOR gate circuit with 40/60 charge partition scheme

The negative voltage results from negative capacitive current. It can be explained clearly based on the drain current used in BSIM3.

$$\mathbf{I_D(t)=I_D - C_{dg}\, dV_{GS}/dt + C_{dd}\, dV_{DS}/dt - C_{db}\, dV_{BS}/dt}$$

Let us analyze each of the factors from the above equation individually. dV_{BS}/dt becomes zero as body to source voltage of the problematic PMOS transistor,Mp1 in this case, doesn't change with time. Similarly C_{dd} is also zero during saturation region mode of operation. Even I_D is small value. Thus the entire drain current depends only on – C_{dg}

dV_{GS}/dt. This explains the bump problem that we see in BSIM3. Negative current only implies that it will be charging the load capacitor. However when gate to source voltage increases significantly beyond the threshold voltage, Id comes into picture and all the charge in the load capacitor is discharged.

Now let us try to correlate this with the NOR Gate circuit that we designed.

Thus during saturation mode of operation due to above scenario, there is negative drain current flowing through the Mp1 transistor. This current must in turn flow through Mp2 and come from either Mn1 or Mn2. We know that at the particular time instant Mn1 is 0 and hence the NMOS transistor is turned off and hence the current must be coming from Mn2 transistor. We can see that in order to generate highly negative current, the voltage at that point needs to be highly negative so that it can have the ability to supply such negative current.

MODIFICATION OF THE NOR GATE CIRCUIT:

Now we slightly modify the NOR gate circuit. First, we add a couple of inverters before both the NMOS and PMOS transistors. By doing so we ensure that the input pulses pass through the inverters before reaching the MOSFET. The main idea behind this is to achieve realistic rise time and fall time with respect to the applied input pulses. We have to understand that though this will make the voltage further negative, this is the most

ideal scenario and is far more realistic than the rise and fall time specified in the Hspice file.

The following diagram represents the schematic of such a NOR gate circuit and waveforms generated by simulating this circuit is shown further below:

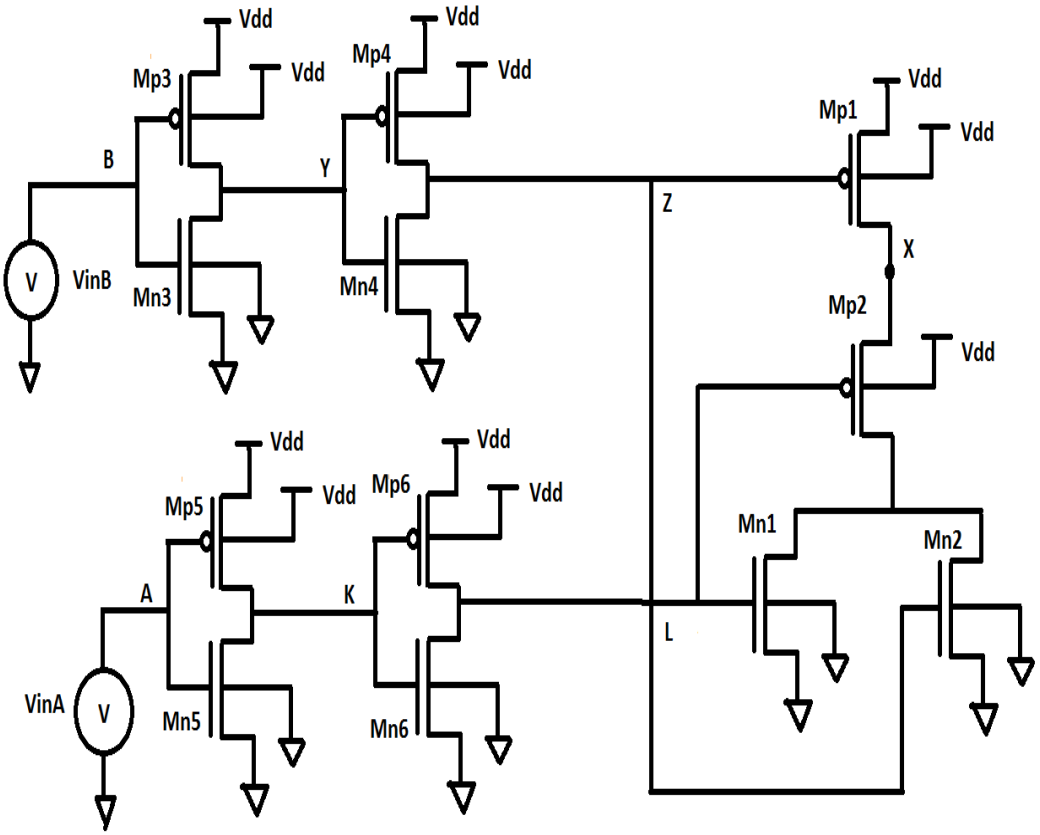

Fig. 3: Modified NOR gate circuit by passing the input pulses through couple of inverters

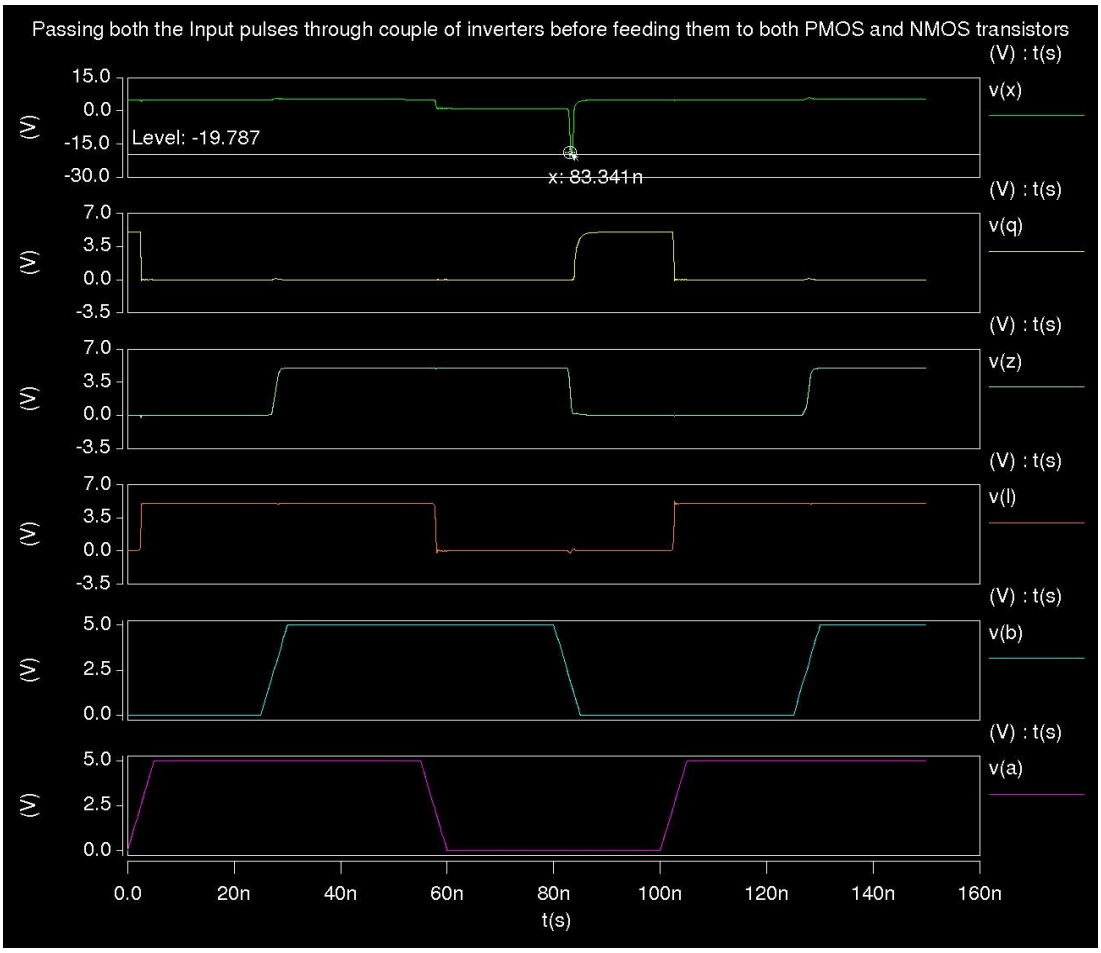

Fig. 4: Waveform related to the modified NOR gate circuit

TECHNIQUES TO OVERCOME THE BUMP PROBLEM IN BSIM3:

XPART parameter:

This is very important parameter in BSIM3. Based on the value being assigned to this parameter we can choose different charge partition schemes for the model.

It is known that MOS charges are somewhat arbitrarily positioned. They can be broadly classified as gate charge, body/bulk charge and the source and drain charges. The sum of the source and drain charges can be effectively called as the channel charge.

Though channel charge can be readily calculated by adding the individual source and drain charge components but it is not that easy to determine the source and drain charges separately. This is where the XPART parameter comes into play. Based on the value being assigned to XPART parameter we specify the charge partition scheme, which determines the way the effective channel charges are distributed between the source and the drain.

Thus XPART parameter is also often referred to as charge partition flag.

Having explained the idea of XPART parameter in brief, now let us see how they are broadly classified as. There are basically three types of charge partition schemes based on value of the XPART parameter.

They are:

40/60 charge partition scheme

50/50 charge partition scheme

0/100 charge partition scheme

Now it is important to understand each of the above charge partition schemes individually and understand their significance which will play an important role in overcoming the bump problem. The number in each of the above charge partition scheme tells us the way or proportion in which the effective charges in the channel are partitioned between the source and the drain.

40/60 CHARGE PARTITION SCHEME:

Ideally the value of XPART parameter can be varied between 0 and 1. It should never be set a negative value. In case of it is set a negative value, transient analysis will be incorrect and there will be no charge calculation. For any value set between 0 and 0.4 (less than 0.5), BSIM3 understands to choose the so called 40/60 charge partition scheme. By 40/60 charge partition scheme, BSIM3 understands that in saturation mode about 40% of the charges in the channel are assigned to the drain and the remaining 60% of them constitute the source charges. It is important to understand the conditions under which this charge partition exists or can be applied. The 40/60 charge partition scheme exists only when the transistor is in the saturation mode of the operation. In the linear region this partition scheme cannot be applied and partition entirely depends only on the bias condition.

50/50 CHARGE PARTITION SCHEME:

From the above explanation it can be straight forward to conclude the subsequent charge partition schemes.

In this charge partition scheme, the entire channel charges are divided between the source and the drain in the 50/50 proportions. In order to choose this charge partition scheme, the XPART parameter has to be set a value of 0.5.

But the important aspect that needs to be considered here is the fact that this scheme applies irrespective of the mode of operation of the transistor i.e it can be either in linear or saturation mode of operation.

0/100 CHARGE PARTITION SCHEME:

This is the final charge partition scheme and BSIM3 chooses this scheme when we set the value of XPART parameter to be any value above 0.5 and 1. In this case all the charges are assumed to be assigned to source and hence it can be concluded that the effective drain charge in this scenario is zero. Similar to the 40/60 charge partition scheme this method also applies only when the transistor is in the saturation mode of operation and entirely depends on the bias condition in the linear mode of operation.

As discussed earlier by choosing the 40/60 charge partition scheme the voltage reaches -7V at 83ns which is way lower than zero voltage. It may be interesting to note that we can overcome this problem by choosing 0/100 charge partition scheme which forces us to ponder on the most viable charge partition scheme and also the reason for existence of three different charge partition schemes.

As stated above the 40/60 charge partition scheme suffers from bump problem due to unphysical rise in the voltage at the initial part of the transient. This is due to the fact that the gate to drain capacitance has finite value at the saturation region and increases gradually in the linear region. This usually causes numerical problems in HSpice simulations.

Now we are left with the analysis of 50/50 charge partition scheme and 0/100 charge partition scheme. It is interesting to note that both of these charge partition schemes are unphysical. By unphysical we mean that they are not derived from device physics and

they are entirely based on hand waving arguments. Now we will take each of the two charge partition schemes individually and analyze them.

50/50 charge partition scheme has the unphysical voltage rise in the initial part of the transient similar to the 40/60 charge partition scheme and also as mentioned above it lacks physical basis due to which it is hardly used in real case.

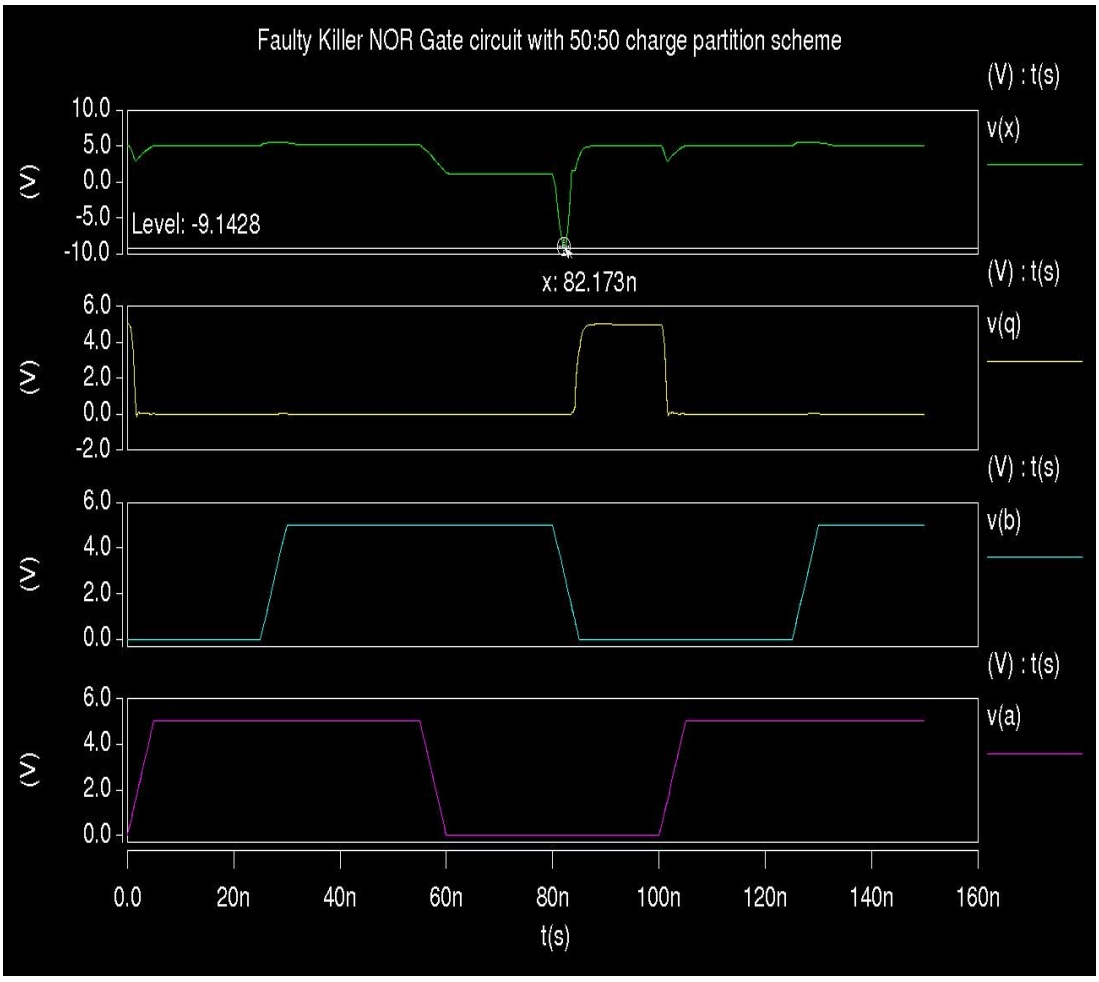

Fig. 5: Faulty NOR gate circuit with 50/50 charge partition scheme

Finally let us have a look at the 0/100 charge partition charge scheme. In this case the gate-drain capacitance has a value of zero at the initial part of the transient. Hence the bump problem that we have in 2 charge partition schemes is not seen here. However since it is quite unphysical charge partition scheme, it results in quite uneven results in the later stages of the transient.

The simulation results of the same Killer NOR gate with 0/100 charge partition schemes is shown below:

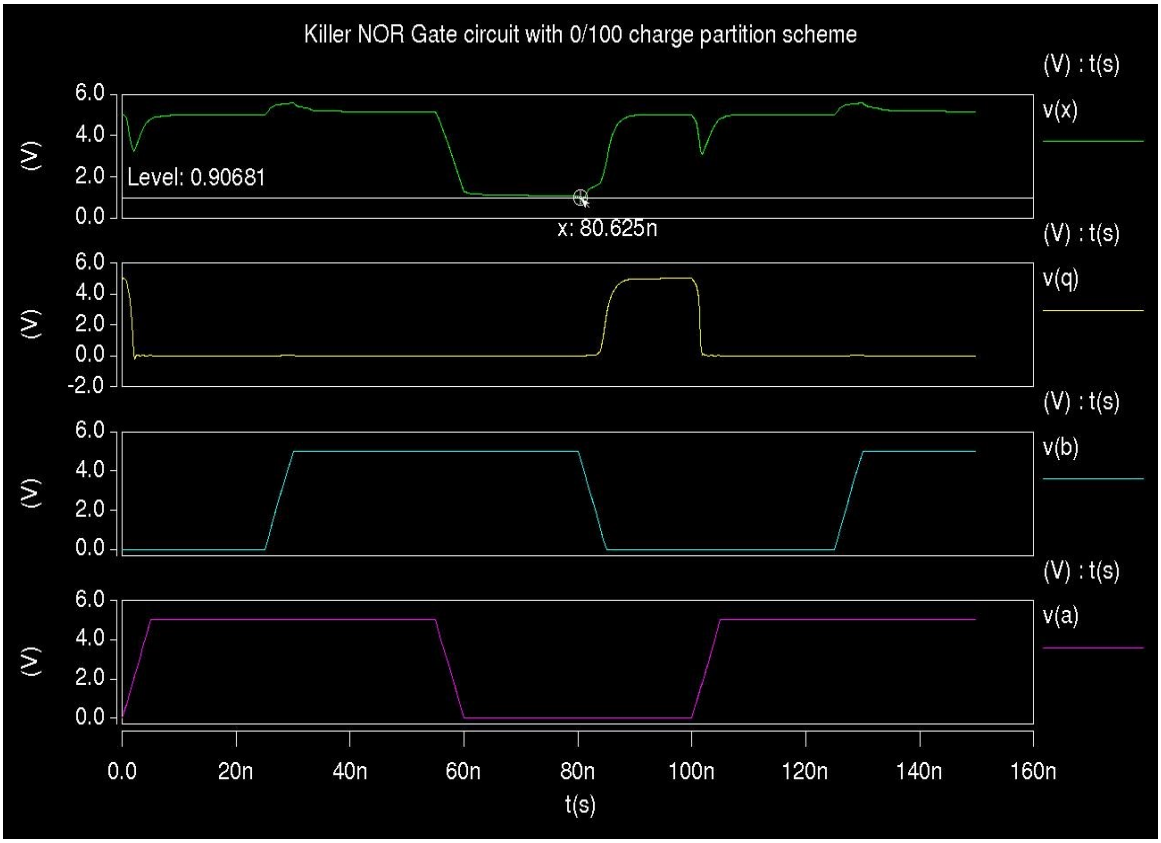

Fig. 6: Faulty NOR gate circuit with 0/100 charge partition scheme

Hence it can be seen that the only difference between 40/60 charge partition schemes and 0/100 charge partition scheme is that in 0/100 charge partition scheme, gate to drain

capacitance is zero during saturation mode of operation and hence we don't witness the bump problem similar to 40/60 charge partition scheme.

So from the above discussions it can be concluded that though 0/100 charge partition schemes can be preferably chosen in order to overcome the bump problem, the fact that it is highly unphysical, forces us to use 40/60 charge partition scheme.

STACKING OF THE PROBLEMATIC TRANSISTOR:

Stacking is a very efficient way of overcoming or reducing the impact of the bump problem in BSIM3. The efficiency depends on the number of transistors that we decide to stack in series. Let us first look at the general approach of stacking and then we will have better understanding of implementing that in our case.

Stacking refers to the technique of placing or splitting the transistors in series by dividing their length by number of transistors into which they are split. This is more useful technique in designing integrated circuits as such stacked transistor will have smaller sub-threshold leakage than a single transistor. This is mainly due to the fact that one or more transistors in series usually have lower drain to source voltage which in turn reduces the DIBL. As DIBL reduces, threshold voltage increases and hence the sub-threshold leakage also reduces.

Though stacking has such an important advantage, it is not that straightforward. It comes with many catches. When we split the transistors into number of equal parts and place them in series, we are effectively increasing the delay through them and hence overall

circuit speed decreases. This is the major drawback of stacking and forces us to limit the number of the transistors that we place in series. Hence it is mostly advisable to use stacking in non-critical paths so that circuit timing is not compromised at any instant.

The next major issue is that by stacking, we are effectively reducing the length of the each transistor based on the number of the transistors that we decide to place in series. We should also keep in mind that reducing channel length beyond certain point will result in short channel effects which will cause some other unusual circuit behavior during simulation.

Now let us take our problematic circuit and try splitting them two transistors each of length 2.5 u. By splitting them into two, voltage at point X V(X) reduces from -7V in the faulty scenario to -0.39V. Then we try to split them into 5 transistors in series such that each transistor has length of 1u. By doing so V(X) further decreases to 0.016V which is physical value . It is important to note that the only reason that we chose to split into two and five transistors in series is owing to the fact the effective length of 5u can be easily divided by 2 and 5 rather than 3 or 4.

Waveforms in each of the above 2 cases discussed are shown below.

REPLACING PROBLEMATIC PMOS TRANSISTOR(TOP TRANSISTOR) BY TWO TRANSISTORS IN SERIES:

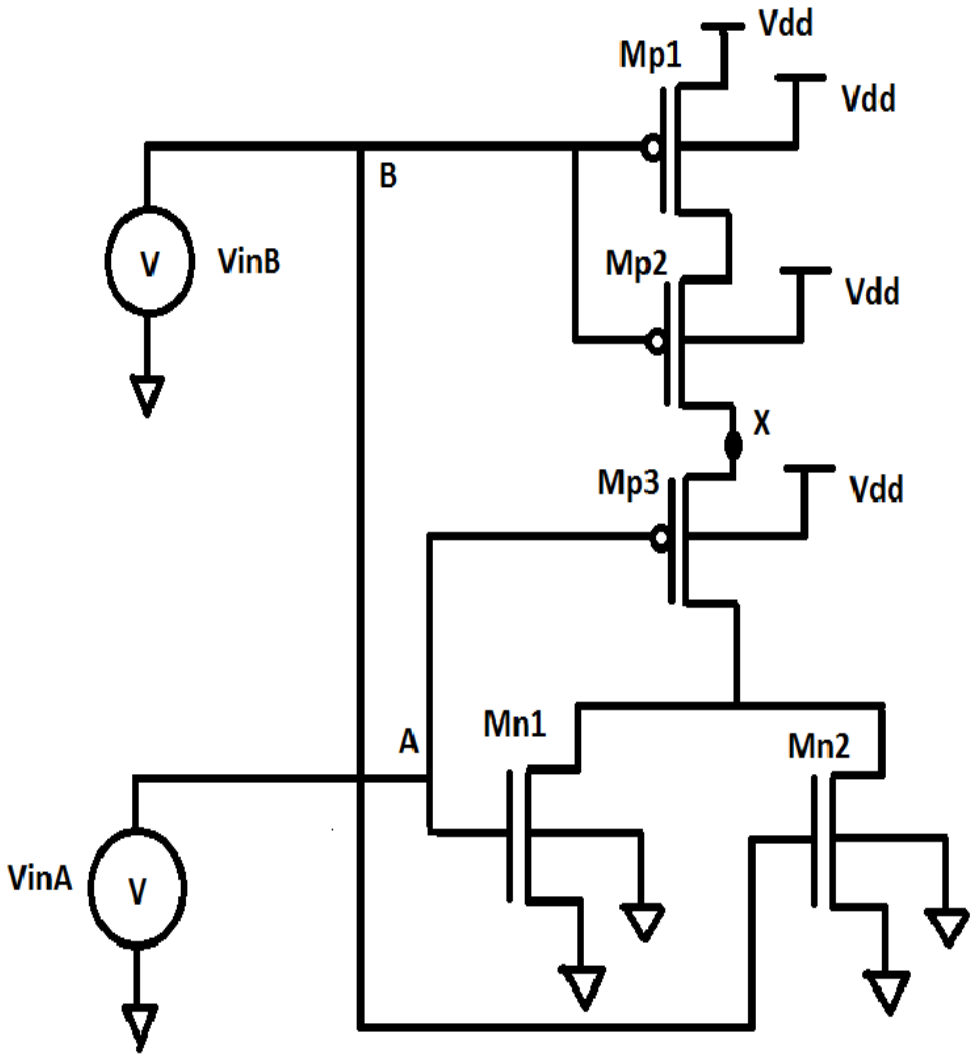

Fig. 7: Schematic of the circuit by replacing the problematic transistor by two transistors in series

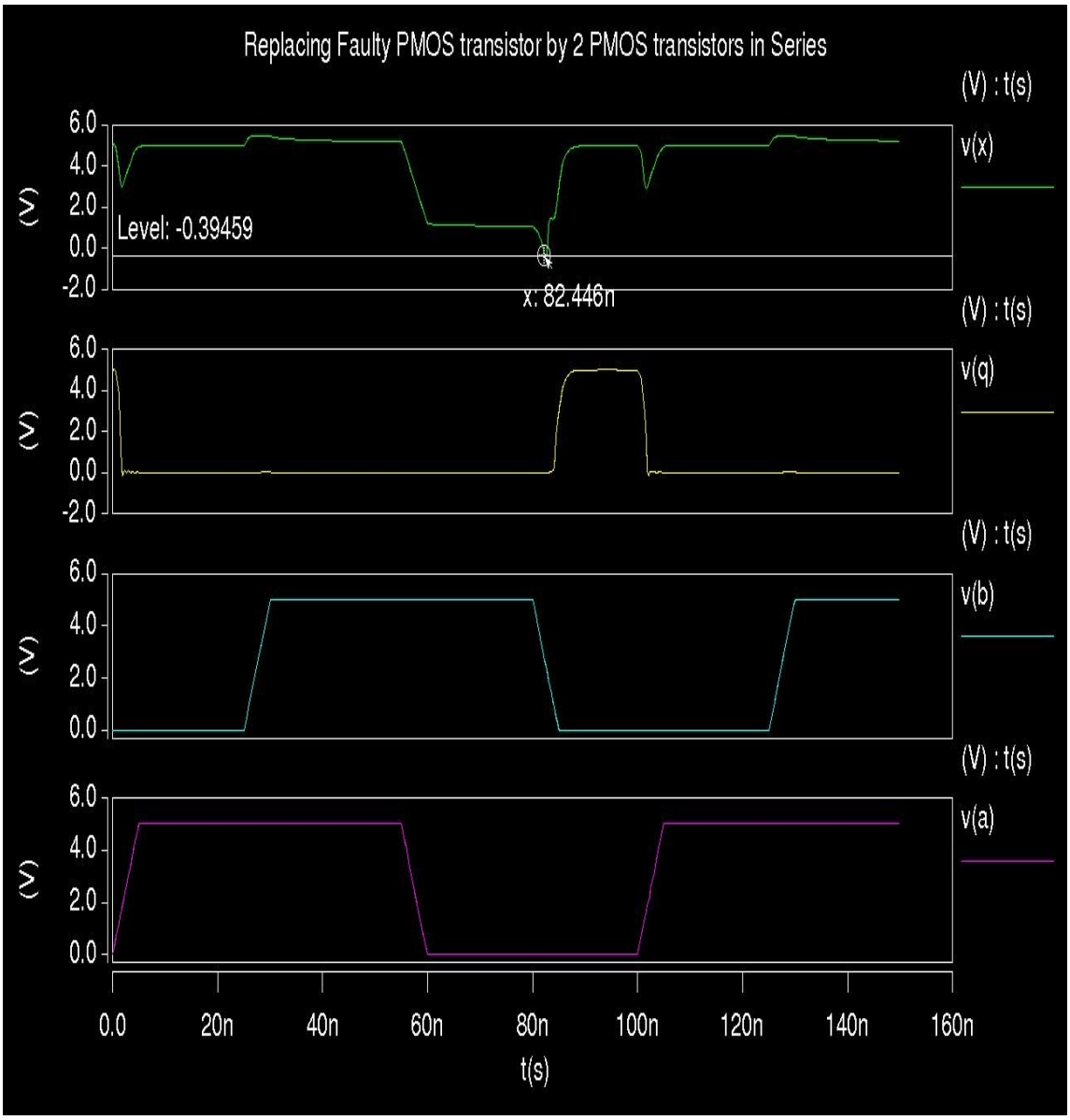

Fig. 8: Waveform related to the above circuit by placing 2 PMOS transistors in series

REPLACING PROBLEMATIC PMOS TRANSISTOR (TOP TRANSISTOR) BY FIVE TRANSISTORS IN SERIES:

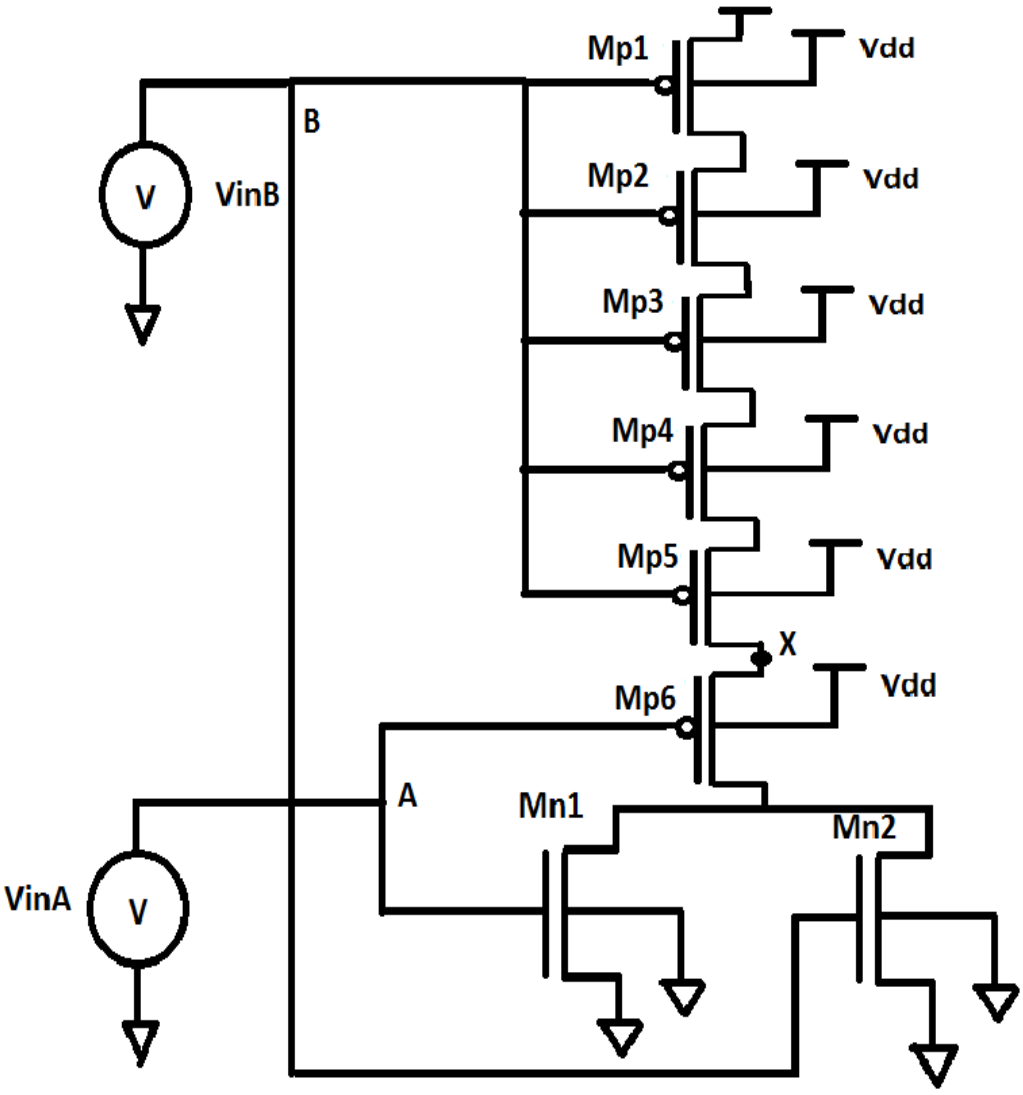

Fig. 9: Schematic of the circuit by replacing the problematic transistor by five transistors in series

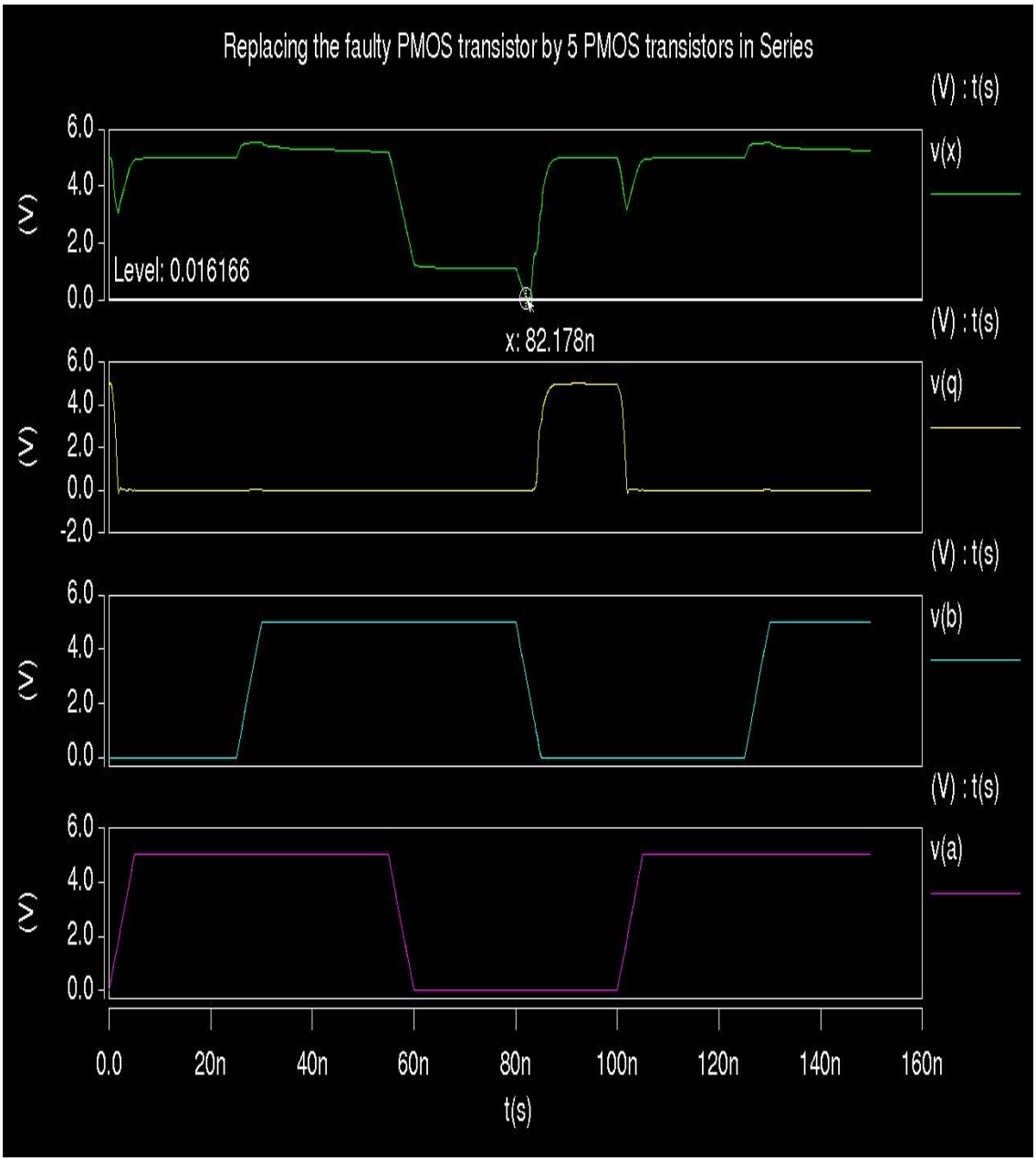

Fig. 10: Waveform related to the above circuit by placing five PMOS transistors in series

EFFECT OF INCREASING THE RISE TIME AND FALL TIME:

The rise time and fall time of the input pulses are very important factors as far as Bump problem is concerned. As discussed earlier in case if the rise time and fall time are faster

than the transit time of the charge carriers, then QS assumptions doesn't hold. Hence the value that we choose for rise time and fall time of the input pulses determines the nature of operation of the circuit. We also described the way to make it an ideal circuit by passing the input pulses through the inverters. Now since we are focusing on the methods to overcome or reduce the bump problem, we can try making the transition of rise time and fall time slower, thus giving more time for the charge carriers to make a transition from the source to drain. In other words we try increasing the rise time and fall time values of both the input pulses.

We know the drain current depends on the component – C_{dg} dV_{GS}/dt and basically by slowing both the rise and fall times we reduce the magnitude of the drain current and hence the voltage at X,V(X) need to be so negative to supply the required drain current.

The following waveforms will show us the merits of slowing both the rise time and fall time of input pulses.

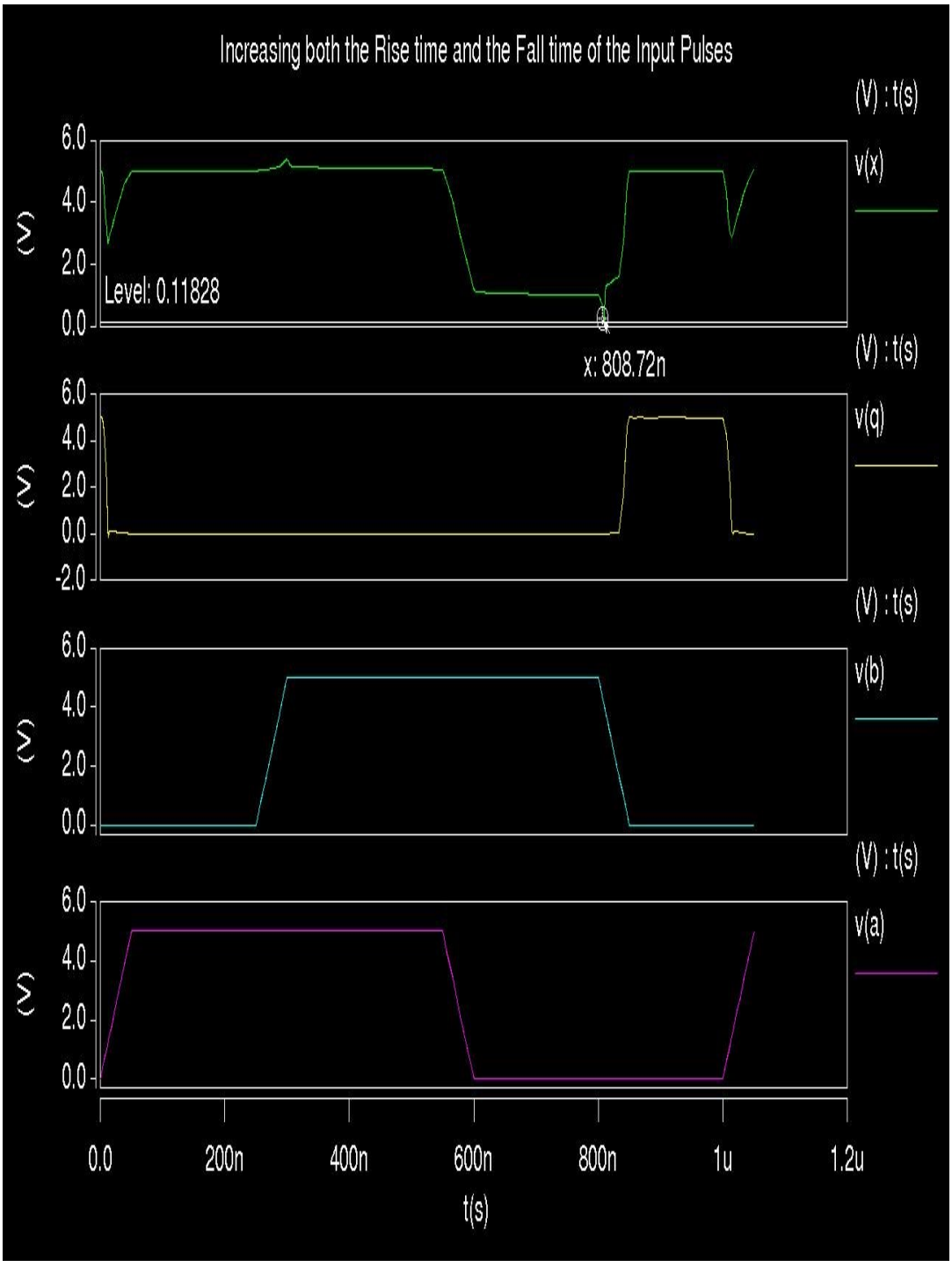

Fig. 11: Waveform showing the effect of increasing the rise and fall time of i/p pulses

EFFECT OF ADDING SOURCE TO BODY JUNCTION CAPACITANCE:

Adding source to body junction capacitance has significant impact in reducing the voltage at node X and hence the spike in initial voltage transients.

There is a connection from Node X to Vdd through source to body junction capacitance of Mp2. Since drain current depends on the factor CjSB dV(x)/dt and when we take CjSB into consideration, the voltage at node X V(X) need not be that negative. Thus the bump problem can be overcome by choosing a suitable value for source to body junction capacitance.

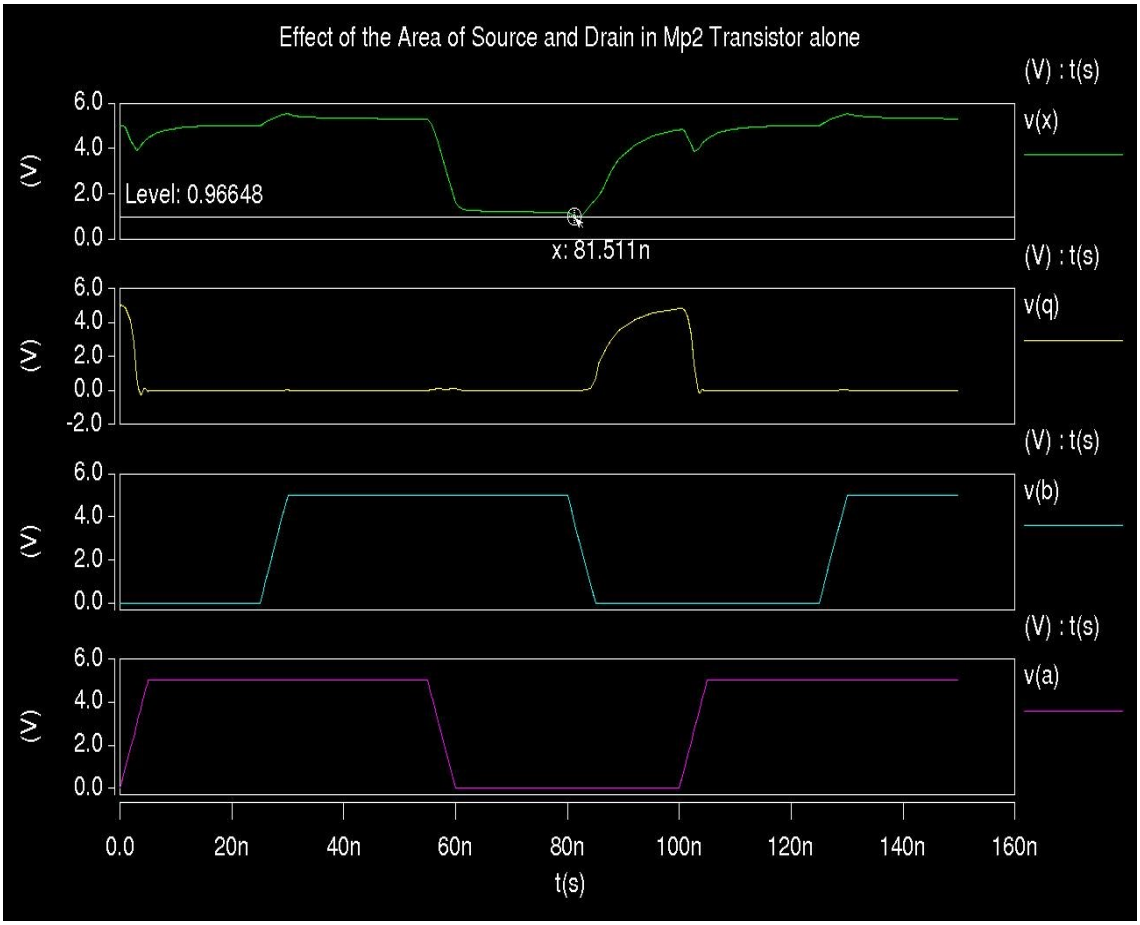

Fig. 12: Waveform displaying the effect of adding source to body junction capacitance

EFFECT OF ADDING DRAIN TO BODY JUNCTION CAPACITANCE:

Similar to the effect of adding source to body junction capacitance, we can also consider the effect of drain to body junction capacitance. Once again by considering this effect, we consider a connection from Node X to supply voltage Vdd which will supply some part of the negative drain current. Hence even in this case voltage V(X) is not that negative.

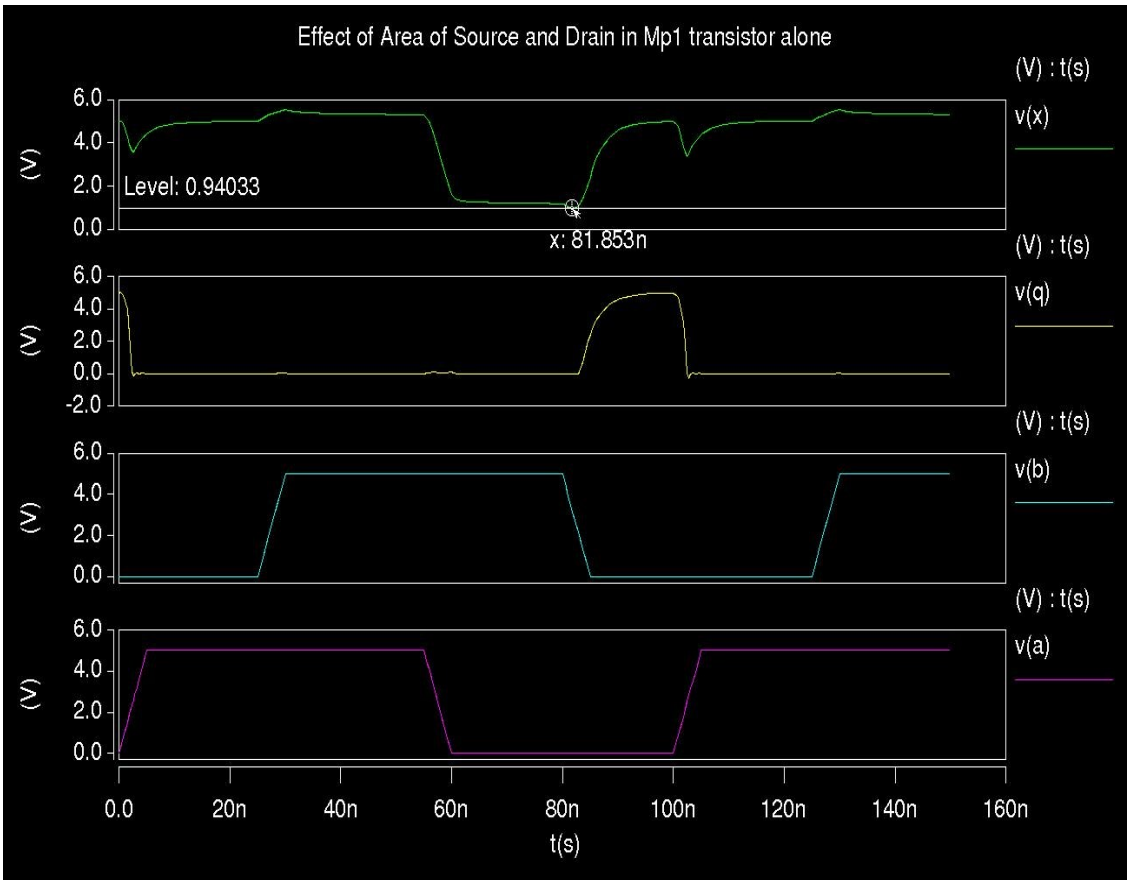

Fig. 13: Waveform displaying the effect of adding drain to body junction capacitance

We have considered the effect of source to body junction capacitance and drain to body junction capacitance separately. The following waveform will give us an idea by taking into consideration the effects of both the capacitances when applied simultaneously. We

can consider the effects of both the capacitances by adding area of source and drain and also their perimeters in device statement.

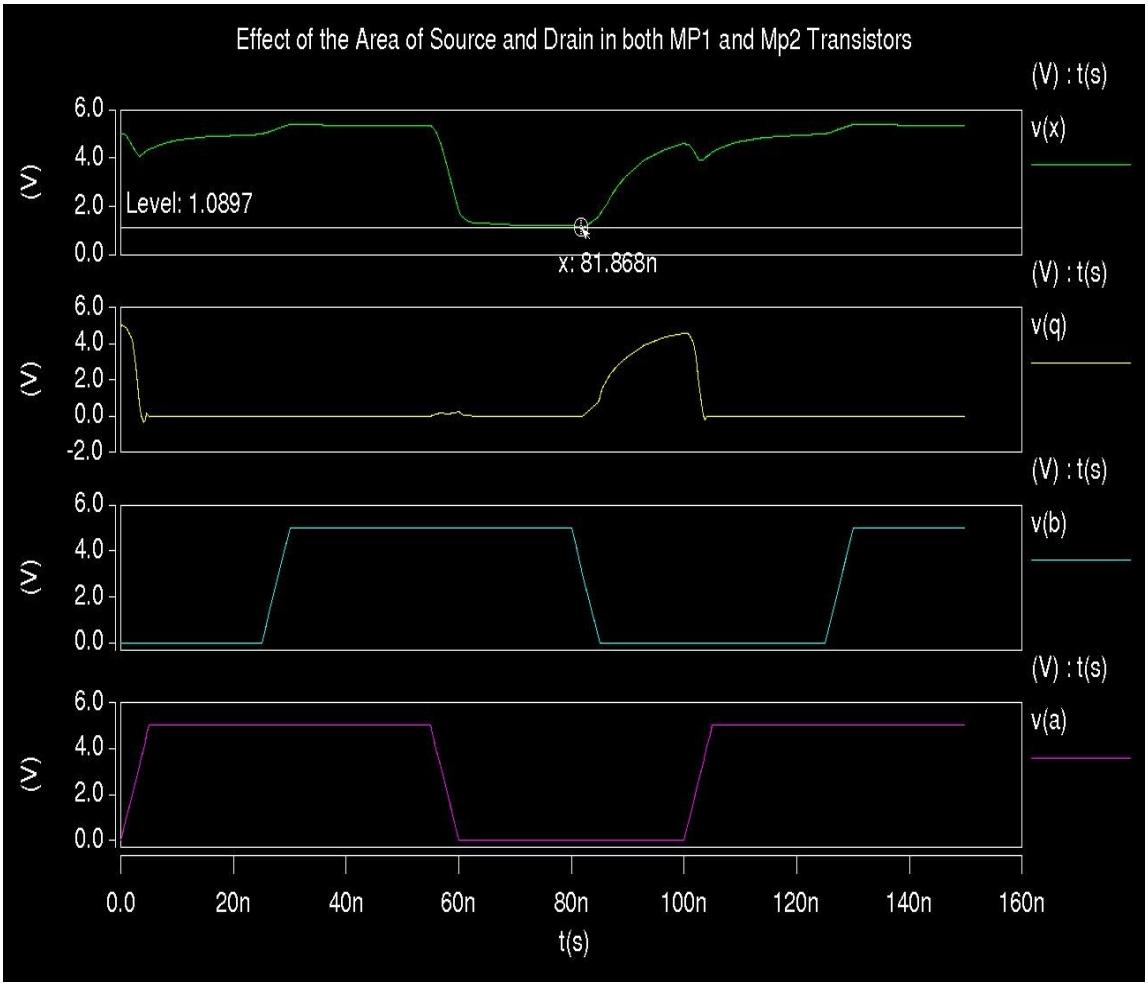

Fig.14: Waveform showing the effect of area of the source and drain in both the transistors

NON QUASI-STATIC MODELLING (NQS MODEL):

A brief introduction was given at the beginning about Non Quasi-static model, called as NQS model in short, when we were talking about different types of model. NQS model is useful in many applications particularly when change in the signal time is lower than the transit time in the device at which point the QS approximation is not applicable.

We saw that QS models have assumptions that the finite time taken by the charges to make a transition from let's say source to drain is ignored. Thus it is assumed that the charges respond to the changes in the voltage with infinite speed.

But we know that this is not the case in reality and the charges do take some time to transit the channel due to which the QS assumption doesn't hold well. Hence in cases when the input signals have a faster transition in comparable to the transit time of the devices the NQS model is preferred over the QS assumption due to greater accuracy in the simulation results and in this manner we can also oversee the spikes in the voltage in the initial part of the transients that was a major issue in the QS modelling.

Numerous papers have discussed different types of NQS modelling based on the current continuity equation. But the problem with all of them are the long simulation time and also complexity, which forces us to overlook them in search for better model with more reasonable circuit simulation time. One such NQS model is proposed by BSIM3 which is based on the Elmore RC circuit.

Having developed good understanding on NQS modelling let us now look at NQS modelling can be controlled or invoked in HSpice. BSIM3 supports a parameter called

nqsMod which enables the users to turn on the NQS model. It defaults to zero value which means it is disabled. In order to invoke them in our simulation we need to set the nqsMod to 1. It is implemented only in BSIM version 3.2.

However since we use BSIM3 version 3.1 in order to support NQS, we have to add the parameter as a model or an instance parameter.

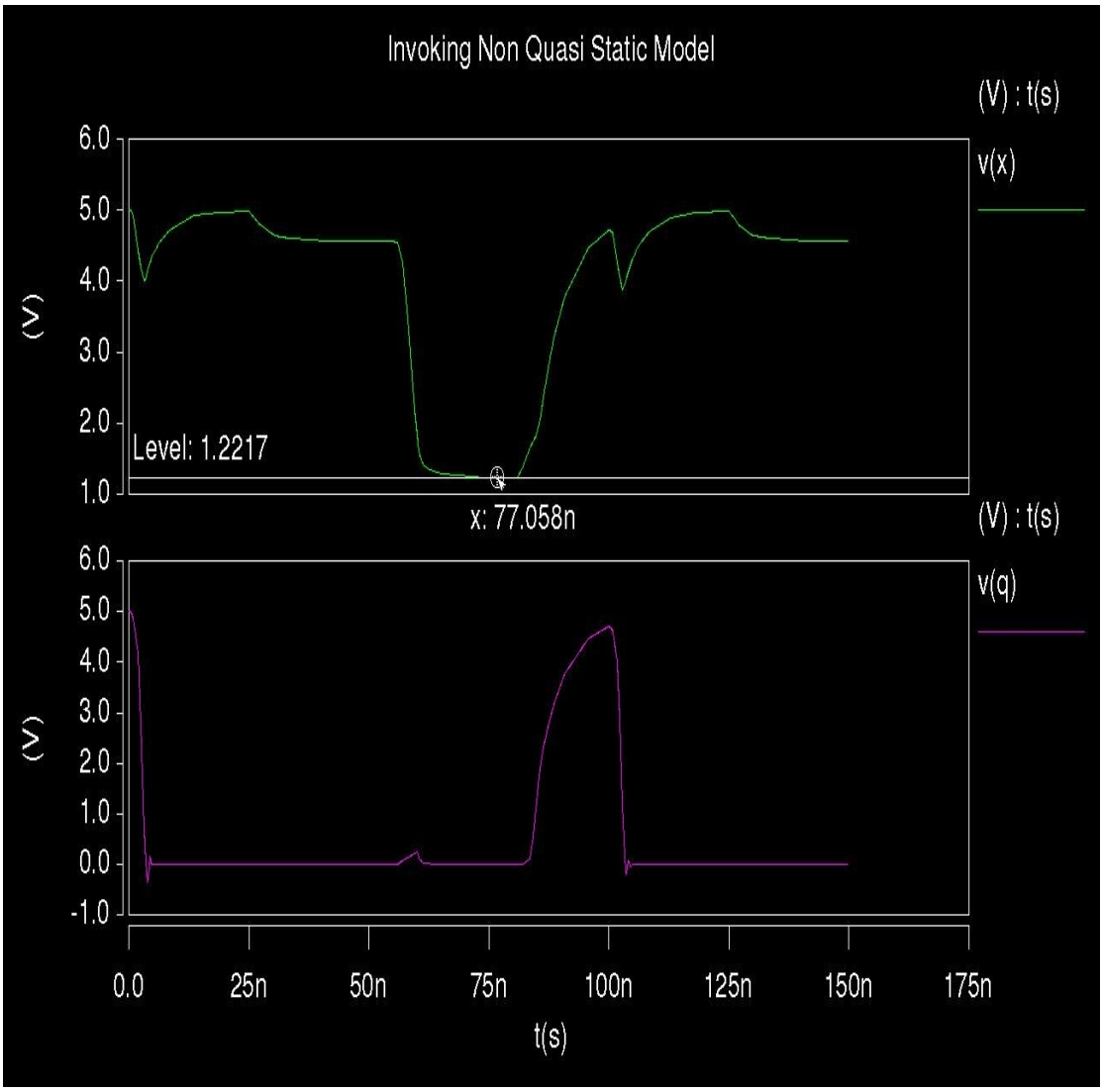

Fig. 15: Waveform showing the effect of invoking NQS modelling

Along with nqsMod parameter there is another parameter called Elm associated with it. ELM parameter refers to Elmore constant of the channel and by default has value of 5. By increasing this Elmore constant it is possible to improve the simulation of the Killer NOR gate circuit making it more positive. The below waveforms represents the case when nqsMod is set to 1 and Elm is set a value of 8.

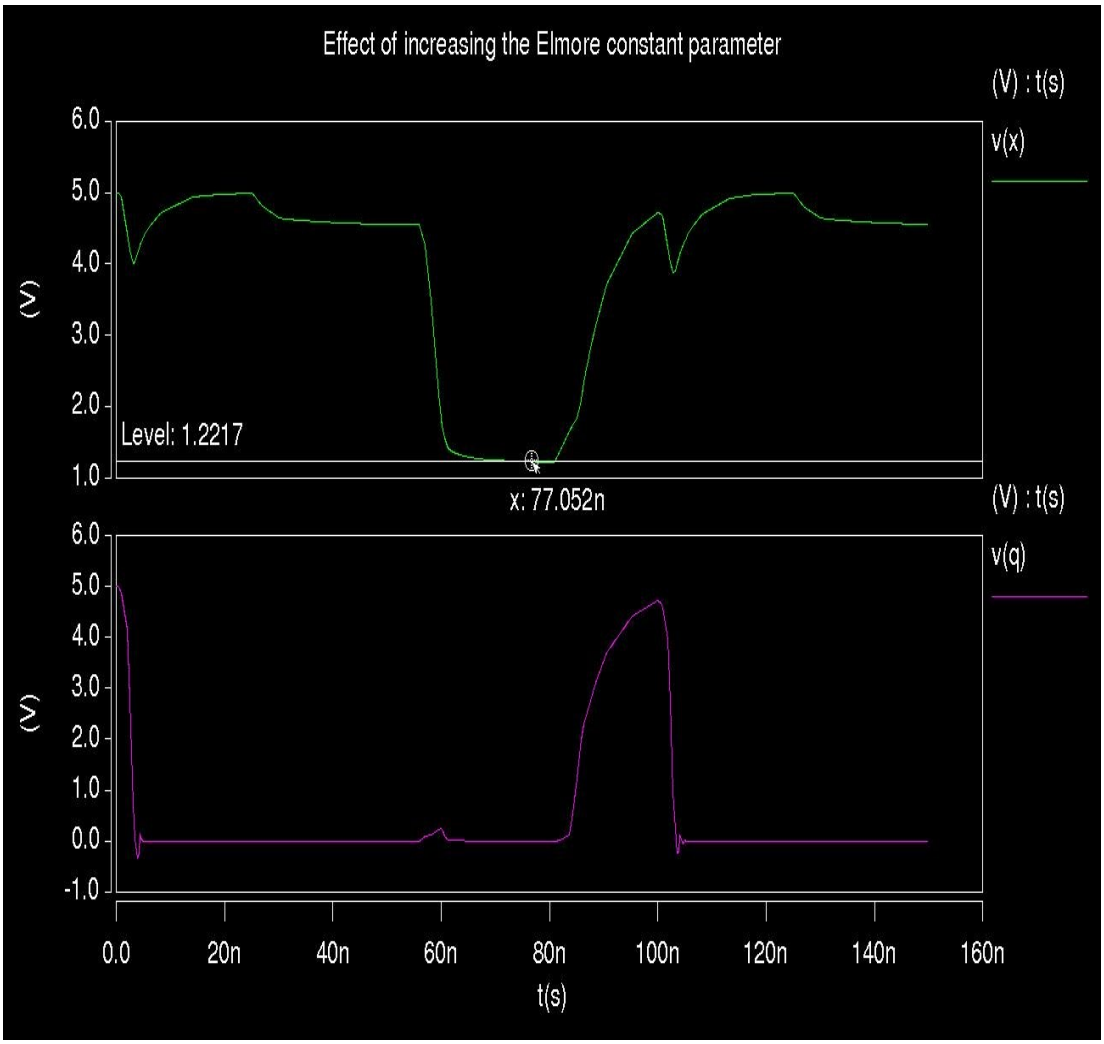

Fig. 16: Waveform showing the effect of increasing the Elmore constant parameter

EFFECT OF MODIFYING THE CAPMOD PARAMETER:

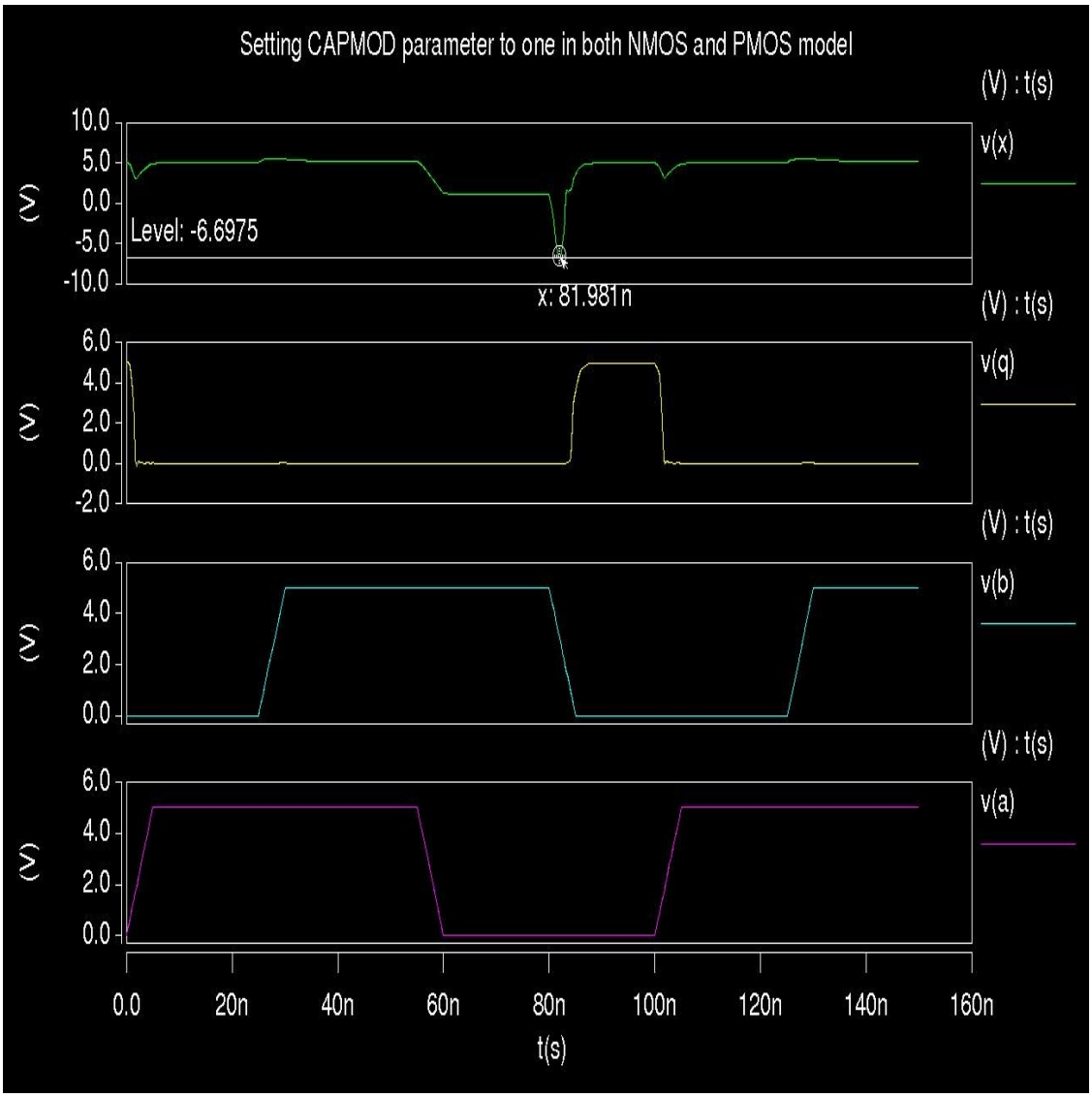

Fig. 17: Setting the CAPMOD parameter to one in both NMOS and PMOS model

CONCLUSION:

In this paper we discuss in detail one of the problematic areas in BSIM3 namely the bump problem, which is not just a BSIM3 problem. A simple NOR gate circuit is taken into consideration to give a better understanding of the problem. Once the bump problem was understood, a few techniques to overcome the problem were suggested and discussed in detail.

The primary suggestion was to stack the problematic transistor by a number of transistors in series. This was explained with the help of schematic and also behavior was analyzed in depth using the generated waveforms. This has its own drawbacks. Factors like speed of the circuit and area restricts us to limit the number of transistors connected in series.

Next solution was the use 0/100 charge partition scheme instead of 40/60 charge partition that is currently being used. Though the problem is completely overcome by using this method, it is highly unphysical method and is seldom used. We may be able to overcome the spikes in voltage in initial part of the transient but we will witness unusual behavior in later part of transient analysis.

Considering the effects of the body to source junction capacitance along with body to junction capacitance of PMOS transistor was also proposed and it was seen that this will supply part of the required current due to connection between drain/source and body which is tied to VDD. Hence the voltage doesn't have to be that negative resulting in significant improvement in the voltage analysis and suppressing the bump problem.

Quasi-static assumptions are no longer valid when the time scale is shorter than the transit time of the charge carriers. Slowing down of the input pulses could help us in overcoming the bump problem. We demonstrated with the help of waveforms, increasing the rise time and the fall time of input pulses could prove to another effective solution to the problem.

Finally we considered invoking non quasi-static modelling in our analysis by setting the parameter nqsMod to one.

In this paper our intention of discussing the bump problem in BSIM3 is not to single out BSIM3, but addressing this issue and following the suggestions discussed in the paper will significantly improve the simulation results and also help in better understanding of the results.

BIBILIOGRAPHY:

1. MOSFET modelling and BSIM3 User's Guide, Yuhua Cheng, Chenming Hu
2. MOSFET Models for SPICE Simulation: Including BSIM3v3 and BSIM3, William Liu
3. CMOS VLSI Design: A Circuits and Systems Perspective, Neil Weste and David Harris