TAMING THE RED LLAMA—MODELING A CMOS-BASED OVERDRIVE CIRCUIT

Lasse Köper and Martin Holters
Department of Signal Processing and Communication
Helmut Schmidt University
Hamburg, Germany
lasse.koeper@hsu-hh.de, martin.holters@hsu-hh.de

ABSTRACT

The Red Llama guitar overdrive effect pedal differs from most other overdrive effects because it utilizes CMOS inverters, formed by two metal-oxide-semiconductor field-effect transistors (MOSFETs), instead of a combination of operational amplifiers and diodes to obtain nonlinear distortion. This makes it an interesting subject for virtual analog modeling, obviously requiring a suitable model for the CMOS inverters. Therefore, in this paper, we extend a well-known model for MOSFETs by a straightforward heuristic approach to achieve a good match between the model and measurement data obtained for the individual MOSFETs. This allows a faithful digital simulation of the Red Llama.

1. INTRODUCTION

The Red Llama overdrive is one of several guitar effect pedals designed by Jeorge Tripps and sold through his company Way Huge in the 1990s. Commercial success was limited and so was the company’s life span. Once production of the devices had stopped, interest in them among guitarists and collectors rose, leading to Jim Dunlop reviving the brand some 15 years later and releasing reissues of some of the original effects, including the Red Llama. While only a limited number of Red Llamas were sold, the effect is still available as part of the Camel Toe, however slightly modified to include a controllable hi-cut filter. One of the most well-known guitar players reported to regularly use the Red Llama/Camel Toe is Mike Campbell.

From a technical perspective, the Red Llama is remarkable because while most overdrive pedals rely on operational amplifiers and diodes for distortion, the Llama employs CMOS inverters to achieve a unique sound with a relatively simple circuit. Consequently, while the simulation of diode clipping and operational amplifiers has received broad attention in virtual analog modeling (e.g. [1, 2, 3] to name just a few), to the best of our knowledge, no work has considered the simulation of CMOS-based circuits for audio applications so far. Therefore, the focus of this paper is on modeling CMOS inverters in the audio effect context with the Red Llama as an example of practical usage to prove the effectiveness of the derived model.

2. THE LLAMA CIRCUIT

Before going into the details of the CMOS inverters, we shall first give an overview of the Red Llama circuit shown in figure 1. We have to note that the reference circuit used for this work is not an original Way Huge device, but rather the Llama do-it-yourself kit sold by mustking.de. Apart from the lower cost, this has the advantage that all components could be measured prior to assembly. The obtained information is listed in table 1.

![Figure 1: Signal path schematic of the Llama](image)

Table 1: Element information

<table>
<thead>
<tr>
<th>part name</th>
<th>nominal</th>
<th>measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>100 kΩ</td>
<td>99.565 kΩ</td>
</tr>
<tr>
<td>R2</td>
<td>100 kΩ</td>
<td>99.625 kΩ</td>
</tr>
<tr>
<td>R3</td>
<td>1000 kΩ</td>
<td>994.23 kΩ</td>
</tr>
<tr>
<td>R4</td>
<td>1000 kΩ</td>
<td>992.89 kΩ</td>
</tr>
<tr>
<td>R5</td>
<td>1000 Ω</td>
<td>997.49 Ω</td>
</tr>
<tr>
<td>R6</td>
<td>2.2 kΩ</td>
<td>2.19 kΩ</td>
</tr>
<tr>
<td>C1</td>
<td>66 nF</td>
<td>66 nF</td>
</tr>
<tr>
<td>C2</td>
<td>51 pF</td>
<td>48 pF</td>
</tr>
<tr>
<td>C3</td>
<td>33 nF</td>
<td>33 nF</td>
</tr>
<tr>
<td>C4</td>
<td>100 pF</td>
<td>95 pF</td>
</tr>
<tr>
<td>C5</td>
<td>10 µF</td>
<td>9.3 µF</td>
</tr>
<tr>
<td>C6</td>
<td>330 µF</td>
<td>316 µF</td>
</tr>
</tbody>
</table>

have to note that the reference circuit used for this work is not an original Way Huge device, but rather the Llama do-it-yourself kit sold by mustking.de. Apart from the lower cost, this has the advantage that all components could be measured prior to assembly. The obtained information is listed in table 1.

![Table 1: Element information](image)

Apparently, the CMOS inverter is used like an operational amplifier in inverting configuration. In fact, to obtain a first intuition about the circuit, we shall consider its small signal behavior, for which we assume the CMOS inverters to act like an ideal operational amplifier with the non-inverting input tied to a fixed reference voltage. Note that this assumption is valid only for inverters with very high input impedance, low output impedance and a sufficiently high gain-bandwidth product. Observing that input and output are AC-coupled and as we are not interested in any bias voltages for the small signal model, we choose this reference voltage to be ground...
for simplicity.

The first inverter together with C1, C2, R1, R2, the gain control and any source impedance $R_\text{in}$ of the input signal (assumed ohmic) forms a second-order bandpass filter

$$H_1(s) = \frac{(R_2 + R_{\text{gain}})C_1s}{(1 + (R_1 + R_{\text{in}})C_1s)(1 + (R_2 + R_{\text{gain}})C_2s)}$$

(1)

with the gain in the pass band approaching

$$G_1 = \frac{R_2 + R_{\text{gain}}}{R_1 + R_{\text{in}}}$$

(2)

It can be decomposed into a first-order highpass filter with cut-off frequency

$$f_c = \frac{1}{2\pi(R_1 + R_{\text{in}})C_1}$$

(3)

and a first-order lowpass filter with cut-off frequency

$$f_c = \frac{1}{2\pi(R_2 + R_{\text{gain}})C_2}.$$  

(4)

The highpass cut-off frequency $f_c$ reaches 23.4 Hz for $R_{\text{in}} = 0\,\Omega$ and decreases even further for higher source impedance. Therefore, it has only very limited impact on typical input signals. The lowpass cutoff frequency $f_c$ ranges between 31.2 kHz for $R_{\text{gain}} = 0\,\Omega$ and 2837 Hz for $R_{\text{gain}} = 1\,\mathrm{M\Omega}$. For higher gain settings, this may represent a significant reduction in the signal’s higher frequency content. Frequency responses for equidistant gain settings at two different assumed source impedances are shown in figure 2.

The first inverter stage amplifies the signal dependent on the resistance of the gain control. However, the inverter does not necessarily go into saturation. That is because the maximum gain of the first stage would be $G_{\text{max}} = -11$. Assuming a medium level guitar signal of 0.5 $\mathrm{V_{pp}}$ and a medium gain of $G = 5$ the output voltage level would be at 2.5 V. With a supply voltage of 9 V the inverter would not go into saturation.

The second inverter stage together with C3, C4, and R3 forms a first-order highpass filter

$$H_2(s) = -\frac{R_3C_3s}{1 + R_3C_4s}$$

(5)

with cutoff frequency

$$f_c = \frac{1}{2\pi R_3 C_4} = 1592\,\text{Hz}$$

(6)

and passband gain

$$G_2 = \frac{C_3}{C_4} = 330 \pm 50\,\text{dB}.$$  

(7)

Its frequency response is shown in figure 3. As can be seen, the second inverter stage has a high gain even for low and medium frequencies. It will therefore quickly reach saturation, making it responsible for most of the circuit’s distortion.

The output of the second inverter stage is AC-coupled to the volume control potentiometer. For an open output, the cut-off frequency is approximately 1.6 Hz. Assuming a load impedance significantly larger than the potentiometer’s 10 kΩ, no acoustically relevant signal components are affected by this filtering.

Figure 4 shows the supply path of the circuit. The LED and its current limiting resistor R6 are just for signaling the state of the
The only nonlinear element in the circuit from figure 1 is the CMOS inverter CD4049. Therefore, a model needs to be found which behaves like the corresponding analog integrated circuit. Internally, the CMOS inverter mainly comprises two transistors connected as shown in figure 5, an n-channel and a p-channel metal-oxide-semiconductor field-effect transistor (MOSFET) [4]. If the input voltage $v_{\text{in}}$ is near to the supply voltage $v_{\text{DD}}$, the n-channel MOSFET will be active and the p-channel MOSFET will be cut off. As a result, the output voltage will be near ground. Analogously, if the input voltage level is near ground the output will be near the supply voltage [5].

For a modeling approach based on the individual transistors, a suitable model for the n- and p-channel MOSFETs needs to be derived. Therefore, the individual transistors inside the inverter have been measured. The IC has been connected to the measurement equipment as shown in figure 6. This configuration prevents an influence of one transistor while measuring the other. All other inputs of the IC have been connected to ground while the outputs have been left open. The measurement results for the n-channel MOSFET are illustrated in figure 7(a). The drain-source voltage $v_{\text{DS}}$ is plotted against the drain current $i_D$. Each curve corresponds to a different gate-source voltage $v_{\text{GS}}$.

The shown mosfet characteristics can be divided into three different regions. Starting from a drain-source voltage of zero, a region can be observed in which the drain-current rises almost linearly with the drain-source voltage. This is referred to as the linear region. With increasing drain-source voltage, the MOSFET slowly goes into saturation. In the saturation region, the drain current is almost constant. The last region is the cut-off region, in which the gate-source voltage stays below the threshold voltage of the transistor. In this case, almost no drain current is flowing through the transistor. Taking a closer look at the measurements of the p-channel MOSFET in figure 7(b) it can be observed that in the saturation region the drain current is no longer constant but linearly increasing. This is due to the channel-length modulation effect, which can be observed in almost any mosfet [6].

We first try to capture the transistor behavior with the simple model [6]

$$i_D = \begin{cases} 0 & \text{if } v_{\text{GS}} \leq v_T \\ \alpha \cdot (v_{\text{GS}} - v_T - \frac{v_{\text{DS}}}{2}) \cdot v_{\text{DS}} & \text{if } v_{\text{DS}} \leq v_{\text{GS}} - v_T \land v_{\text{GS}} > v_T \\ \frac{\alpha}{2} \cdot (v_{\text{GS}} - v_T)^2 & \text{otherwise} \end{cases}$$

where $v_T$ is the threshold voltage and $\alpha$ is a constant depending on the dimensions and physics of the transistor. The equation is divided into three cases. The first case describes the MOSFET in the cut-off region. The second case models the linear region and the last one the saturation region. The model from equation (8) has two parameters namely $\alpha$ and $v_T$, which have to be adapted to the measurement data. Here, we choose the mean squared error (MSE) between model and measurement output as optimization criterion. The MSE is taken over the whole measurement range using equal weights for each curve. Employing the Levenberg-

---

**Figure 5:** Simple CMOS inverter

**Figure 6:** MOSFET measurement configuration

**Figure 7:** Measurements of the CMOS inverter’s MOSFETs
Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20), Vienna, Austria, September 2020-21

Table 2: Optimized parameters for the simple inverter model

<table>
<thead>
<tr>
<th>MOSFET type</th>
<th>parameter</th>
<th>optimum value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-channel</td>
<td>$\alpha$</td>
<td>$5.1021 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>$v_T$</td>
<td>$1.5702$ V</td>
</tr>
<tr>
<td>p-channel</td>
<td>$\alpha$</td>
<td>$8.2246 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$v_T$</td>
<td>$-0.4848$ 76 V</td>
</tr>
</tbody>
</table>

Marquardt algorithm to minimize the MSE yields the optimum parameters listed in table 2.

The output of the n-channel model with the optimum parameters together with the measurement data is depicted in figure 8(a). It can clearly be seen that the model shows significant differences to the measurement data. First of all, the level of the drain current in the saturation region corresponding to a certain gate-source voltage does not match the measurements. The problem is not only the difference to the measurements, but that the drain current’s magnitude is sometimes lower and sometimes higher than the corresponding measurement. Consequently, this issue cannot be solved by using a simple scaling factor. Additionally, the steepness in the linear region of the model is decreased in comparison to the measurement data. The model output for the p-channel MOSFET, which is obtained from equation (8) by reversing the sign of the voltages and the drain current, is shown in figure 8(b). This model shows the same issues as the n-channel MOSFET model. However, it has to be mentioned that the difference in steepness for the linear region is not as severe as for the n-channel MOSFET. Another difference can be observed in the saturation region. The model has no increase in drain current in saturation mode. Looking at equation (8) this becomes obvious as the third case does not depend on the drain-source voltage. Taking everything into account, it can be stated that the model of equation (8) is not well-suited to model the given CMOS inverter.

In the following, the model will be extended to increase its suitability for the emulation of the CMOS inverter. The first approach is to make the parameters $\alpha$ and $v_T$ a function of the gate-source voltage. By doing this, each curve in figure 7 can be modeled with a nearly linear dependency on the gate-source voltage. The dependence of the gate-source voltage can then be found. To find the optimum parameters for each curve, the MSE for each curve is calculated individually. However, this can only be done if a simple function can be found. Additionally, the steepness in the linear region of the model is decreased in comparison to the measurement data. Consequently, this issue cannot be solved by using a simple scaling factor. Furthermore, the conductivity of the channel is proportional to its width-to-length ratio $L_s$. The channel-length modulation now causes a reduction of the actual channel length. This is because near the drain, the electric field pattern is not only dependent on the gate but also on the drain-source voltage in the saturation region. To understand this effect, the parameter $\alpha$ needs to be decomposed into its basic elements as

$$\alpha = \mu_n C_{ox} \frac{W}{L}$$  \hspace{1cm} (11)

with the load carrier mobility $\mu_n$, capacitance $C_{ox}$ of the gate-oxide, and width $W$ and length $L$ of the gate.

Regarding the p-channel MOSFET, the channel-length modulation needs to be modeled as well. For this purpose it is useful to get a detailed insight of the effect itself. The channel-length modulation causes a dependency of the drain current on the drain-source voltage in the saturation region. To understand this effect, the parameter $\alpha$ needs to be decomposed into its basic elements as

$$\alpha = \mu_n C_{ox} \frac{W}{L}$$  \hspace{1cm} (11)

with the load carrier mobility $\mu_n$, capacitance $C_{ox}$ of the gate-oxide, and width $W$ and length $L$ of the gate. Consequently, the conductivity of the channel is proportional to its width-to-length ratio $L_s$. The channel-length modulation now causes a reduction of the actual channel length. This is because near the drain, the electric field pattern is not only dependent on the gate but also on the drain-source voltage in the saturation region.
voltage. Increasing the drain voltage in saturation produces a gap of uninvited silicon between channel and the drain. The resulting channel length reduction can be expressed by [7]

\[
\alpha' = \mu_n C_{\alpha L} \frac{W}{L - \Delta L}.
\]

Assuming that \(\Delta L\) is much smaller than \(L\), we can rearrange equation (12) to [7]

\[
\alpha' = \frac{\alpha n C_{\alpha L}}{W} \left(1 + \frac{\Delta L}{L}\right) = \alpha \cdot \left(1 + \frac{\Delta L}{L}\right).
\]

By introducing a new parameter \(\lambda\), the channel length reduction can be taken into account with

\[
\frac{\Delta L}{L} = \lambda v_{DS}.
\]

Finally, the equation for the saturation region of a p-channel MOSFET including channel-length modulation can be expressed as

\[
i_{D,\text{sat}} = \frac{\alpha}{2} \cdot (v_{GS} - v_T)^2 \cdot (1 - \lambda v_{DS}).
\]

The parameter \(\lambda\) can be deduced by linear extrapolation of the saturation region from figure 7(b). Comparing the resulting line equations

\[
i_{D,\text{sat}} = m \cdot v_{DS} + c
\]

with equation (15), we immediately find \(-\frac{\alpha}{2} \cdot (v_{GS} - v_T)^2 = c\) and \(\lambda = -m/c\). Using the linear extrapolation shown in figure 10, we find values varying around 0.06 V\(^{-1}\), so we choose \(\lambda = 0.06\) V\(^{-1}\)

in the following.

Now the model from equation (9) can be extended by taking equation (15) as the third case. The parameters \(\alpha\) and \(v_T\) are still functions of the gate-source voltage. The second case needs to be adapted as well to make the whole function continuously differentiable. The resulting extended p-channel MOSFET model is thus given by

\[
i_D = \begin{cases} 
0 & \text{if } v_{GS} \geq v_T(v_{GS}) \\
if D,lin & \text{if } v_{DS} \geq v_{GS} - v_T(v_{GS}) \land v_{GS} < v_T(v_{GS}) \\
i_{D,\text{sat}} & \text{otherwise} 
\end{cases}
\]

with

\[
i_{D,\text{lin}} = -\alpha(v_{GS}) \cdot (v_{GS} - v_T(v_{GS})) \cdot v_{DS} \cdot (1 - \lambda v_{DS})
\]

\[
i_{D,\text{sat}} = -\frac{\alpha(v_{GS})}{2} \cdot (v_{GS} - v_T(v_{GS}))^2 \cdot (1 - \lambda v_{DS})
\]

where

\[
\alpha(v_{GS}) = c_{\alpha,0} + c_{\alpha,1} v_{GS} + c_{\alpha,2} v_{GS}^2 + c_{\alpha,3} v_{GS}^3
\]

\[
v_T(v_{GS}) = c_{v_T,0} + c_{v_T,1} v_{GS} + c_{v_T,2} v_{GS}^2 + c_{v_T,3} v_{GS}^3
\]
Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20), Vienna, Austria, September 2020-21

The passivity of the models was analyzed following the approach from [8].

Figure 11: P-Channel MOSFET model parameters in dependence on $v_{GS}$

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure11a}
\includegraphics[width=0.5\textwidth]{figure11b}
\caption{P-Channel MOSFET model parameters in dependence on $v_{GS}$}
\end{figure}

It can be seen that the model from equation (17) actually reduces to the one from equation (9) if $\lambda$ is set to zero. The coefficients for the parameters $\alpha$ and $\lambda$ can be found in table 3. In contrast to the n-channel model, the regression function for the $\alpha$ parameter of the p-channel model is a third-order polynomial, which provides a good fit as seen in figure 11.

Figure 12 shows the performance of the extended MOSFET model. It can be seen that the curves fit the measurement data far better than the model from figure 8. Also, the channel-length modulation effect of the p-channel MOSFET is modeled correctly. The n-channel model differs slightly from the measurement data at the transition between linear and saturation region. However taking everything into account, the extended model seems to be well suited for usage in the inverter configuration of figure 5.

To validate the suitability of the model given by equations (9) and (17), the input/output behavior of the CMOS-inverter was simulated and compared with measurement data. The supply voltage was set to 9 V. A voltage sweep from 0 V to 9 V is used as an input signal, while the output current $i_{out}$ was fixed at −2.5 mA, 0 mA, or 2.5 mA. Additionally, the same simulation was done using the simple MOSFET model from equation (8). The simulation results and the measurement data are depicted in figure 13. It can be seen that the simple model performs worse than the extended one. Especially at the end of the transition to the low state, the simple model does not fit the measurement data very accurately. Furthermore the steepness of the transition is higher than it should be. This would result into an inverter which saturates faster than its analog counterpart. However, the extended model reproduces the characteristics of the inverter very precisely. Consequently it is suited to be used in a simulation of the whole distortion circuit.

The static input current of the inverter due to leakage can usually be neglected. The input capacitance $C_{in} = 15 \, \text{pF}$ [4], on the other hand, may be relevant in certain applications. However, in the given circuit, it is small compared to the capacitors C2 and C4 connecting input and output and may therefore be neglected as well. Note that the passivity of the proposed models is only given if the polynomial for $\alpha$ remains positive for the given operating range. Additionally, the drain-source voltages are constrained to be positive for n-channel mosfets and negative for p-channel mosfets.

The passivity of the models was analyzed following the approach from [8].
In order to perform a simulation of the distortion circuit, the MOSFET model in ACME\(^1\), which previously implemented the simple model of equation (8), has been extended to correspond to equations (9) and (17). The extended model is available starting with ACME version 0.9.3. Using this MOSFET model implementation, a simulation of the whole circuit with ACME (which is an implementation of the method described in [9]) is straightforward, the source code can be found at https://lkoeper.gitlab.io/dafx-2020-cmos-llama/. The behavior of the two channels of the CD4049 is nearly identical, so the same parameters are used for both.

To evaluate the impact of the MOSFET model, simulations using both the simple and the extended model have been conducted. The used input signals are sinusoids with an amplitude of 0.2 V and a frequency of 100 Hz and 1 kHz, sampled at 96 kHz. The root-mean-squared (RMS) error between measurements and simulation results for different gain settings is given in table 4. It should be noted that the audio interface used to record the measurements includes a low-cut filter at its input. To obtain comparable results, the simulation contains a first-order high-pass filter with a cutoff frequency of 7.94 Hz as a post-processing step.

It can be seen that the extended MOSFET model always performs better than the simple one. The difference between the two models can be seen best by taking a look at the time domain representation in figures 14 and 15. The extended model matches the measurement curve almost exactly. Contrarily, the curve for the simple MOSFET model has visible differences to the measurements. The curvature of both halfwaves does not fit completely to the reference.

To evaluate the acoustical relevance of these differences, a MUSHRA listening test [10] has been conducted. It compares the recorded reference signal with an anchor signal obtained by low-pass filtering at 3.5 kHz and with simulations using both the simple as well as the extended model. Two different guitar signals are used as input, “chord strumming” and “palm muting”. The former sounds very open with a relatively large frequency content while the latter focuses more on the low frequency range with precise staccato-like stroking. Both input signals have been processed as a post-processing step.

\(^1\)https://github.com/HSU-ANT/ACME.jl

Table 4: Root-mean-squared (RMS) error between simulation and measurement of the Llama

<table>
<thead>
<tr>
<th>input freq. (Hz)</th>
<th>gain</th>
<th>simple model</th>
<th>extended model</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0%</td>
<td>33.95</td>
<td>23.18</td>
</tr>
<tr>
<td>25%</td>
<td>24.39</td>
<td>13.46</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>19.04</td>
<td>10.41</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>20.81</td>
<td>8.70</td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>27.35</td>
<td>8.39</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>33.22</td>
<td>19.74</td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>17.79</td>
<td>14.70</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>17.67</td>
<td>8.62</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>18.29</td>
<td>9.36</td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>20.08</td>
<td>8.88</td>
<td></td>
</tr>
</tbody>
</table>
inverter model which is computationally efficient and has as few and easily adjustable parameters as possible. We have chosen to model the inverter based on the two MOSFETs it comprises.

The simplest MOSFET model employed does not show satisfactory capabilities to match the measured behavior of the device. Therefore, it has been extended in two ways: Inclusion of the channel length modulation effect and introduction of a dependence of the MOSFET parameters on the gate-source voltage by low-order polynomials. While the former is a well-known physical effect, the latter is purely heuristic but has proven effective and introduces only few additional parameters.

Simulation of the whole effect device confirms that the simple MOSFET model only suffices to give results that are similar to the analog counterpart, but with differences clearly audible. With the extended model, however, it is much more challenging to distinguish original device and simulation. Thus, the modeling of the MOSFETs and the whole circuit can be considered successful.

6. REFERENCES