Operation of Pulse Train FM detector

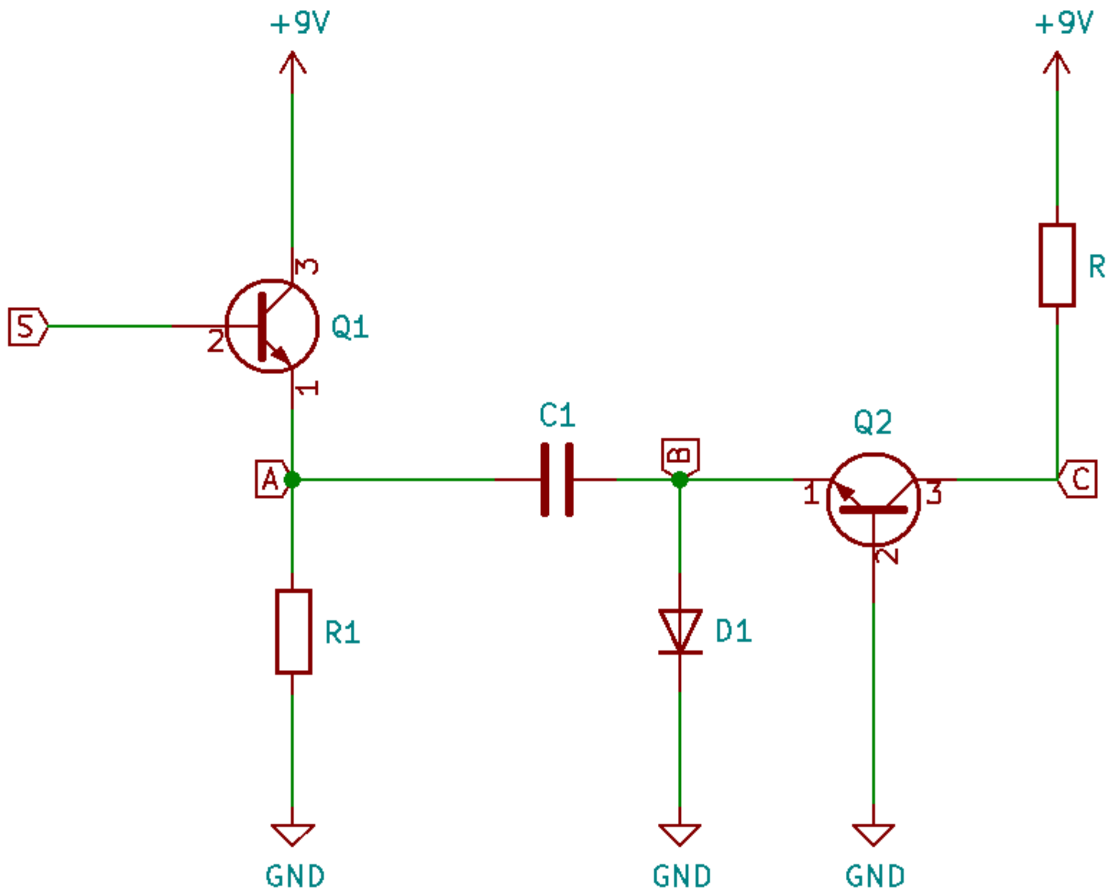
Koen van Dijken. march 2021.

1. Principles of a pulse train FM detector

A pulse train FM detector translates a message which is encoded as a frequency deviation of a carrier into the encoding message. Once every period of the (frequency deviated) carrier a pulse with constant duration is generated. As a result the duty cycle of the resulting pulse train is a direct result of the instantaneous frequency of the input. Low pass filtering the pulse train returns the encoding message.

The pulse width of the output pulses should be constant. ie. independent of the incoming signal.

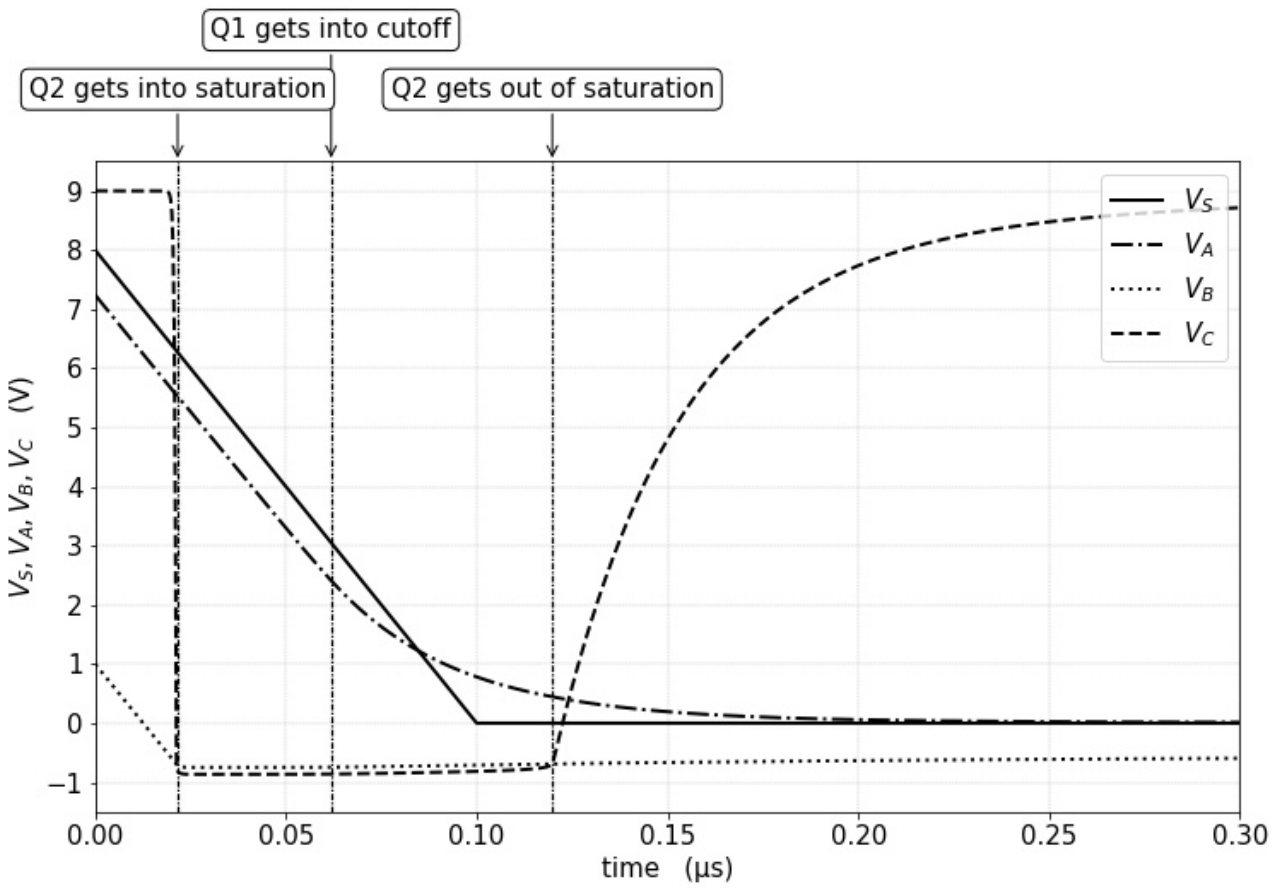
The circuit diagram of a pulse train FM detector looks like this:

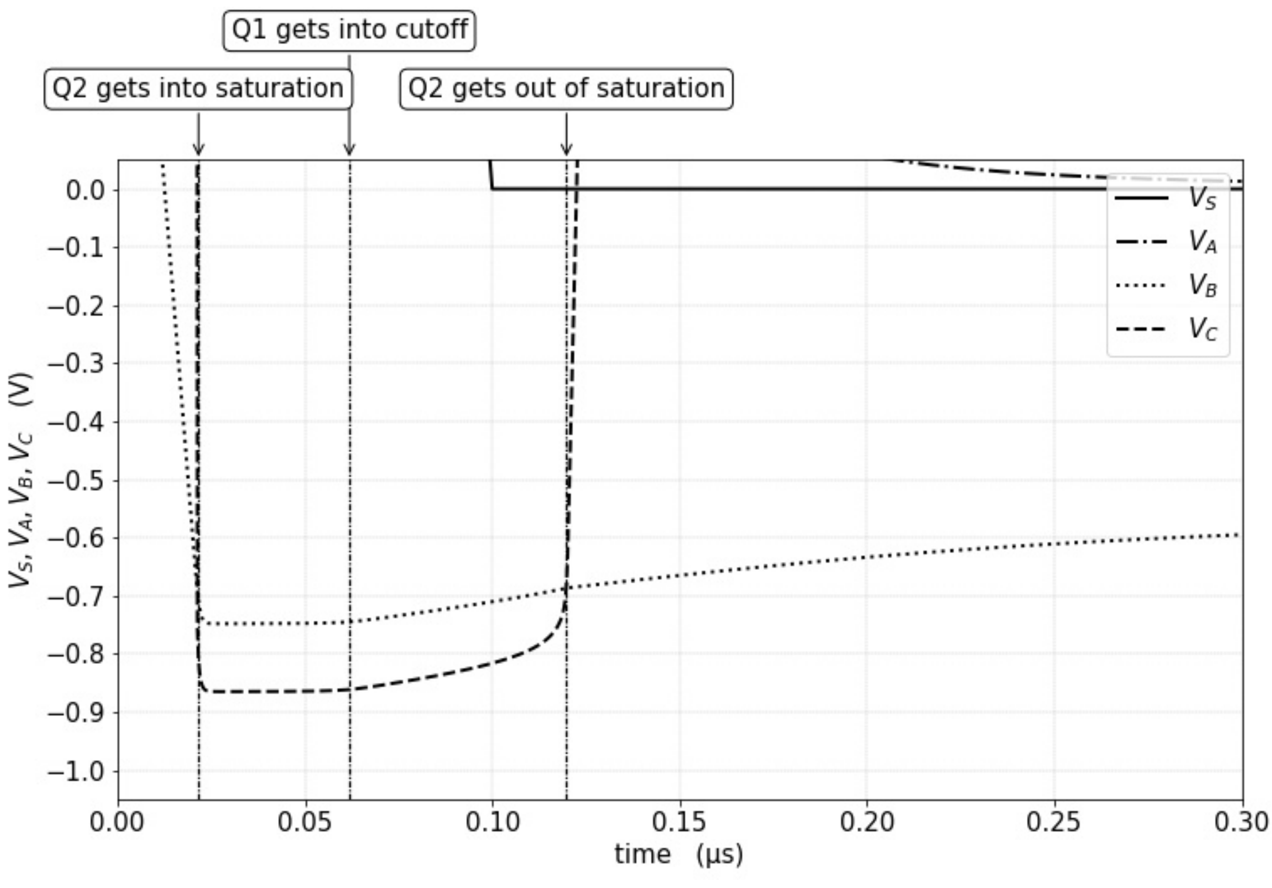


The input signal gets into the circuit at S. and the output is taken from C.

In very general terms one could describe operation of the circuit as follows. V_B will never be higher than a diode drop above because of D_1. and never lower than a diode drop below ground because of the base-emitter junction of Q_2. When C_1 is charged. V_B will be about +0.6V. On a falling input signal V_B will drop down to about -0.6V. This will get Q_2 into conduction. The Q_1 emitter current will discharge C_1. This current will pull V_C down and get Q_2 into saturation. This same current will create a voltage over R_1 which will get Q_1 into cutoff. When C_1 is sufficiently discharged to get V_B above -0.6V. Q_2 will stop conducting and V_C will rise to 9V. When V_S rises again. V_A will follow and charge C_1.

This ideal behaviour is shown in figures 1 and 2.





In this document a model is used which describes the pulse generating part of the circuit. Diode D_1 has been removed from the simulation as its only function is prevent V_B to rise above about +0.6V. Removal of D_1 has been resolved by setting the initial condition for V_B. see chapter 7.

2. General description of operation of pulse train detector

This chapter describes the operation of the pulse train FM detector. The detector operates in four distinct phases. which are separated in the below table by the red rows. The last column in this table mentions certain conclusions which can be drawn from the content in the second and third columns. These four phases can be recognised in figures 1 and 2.

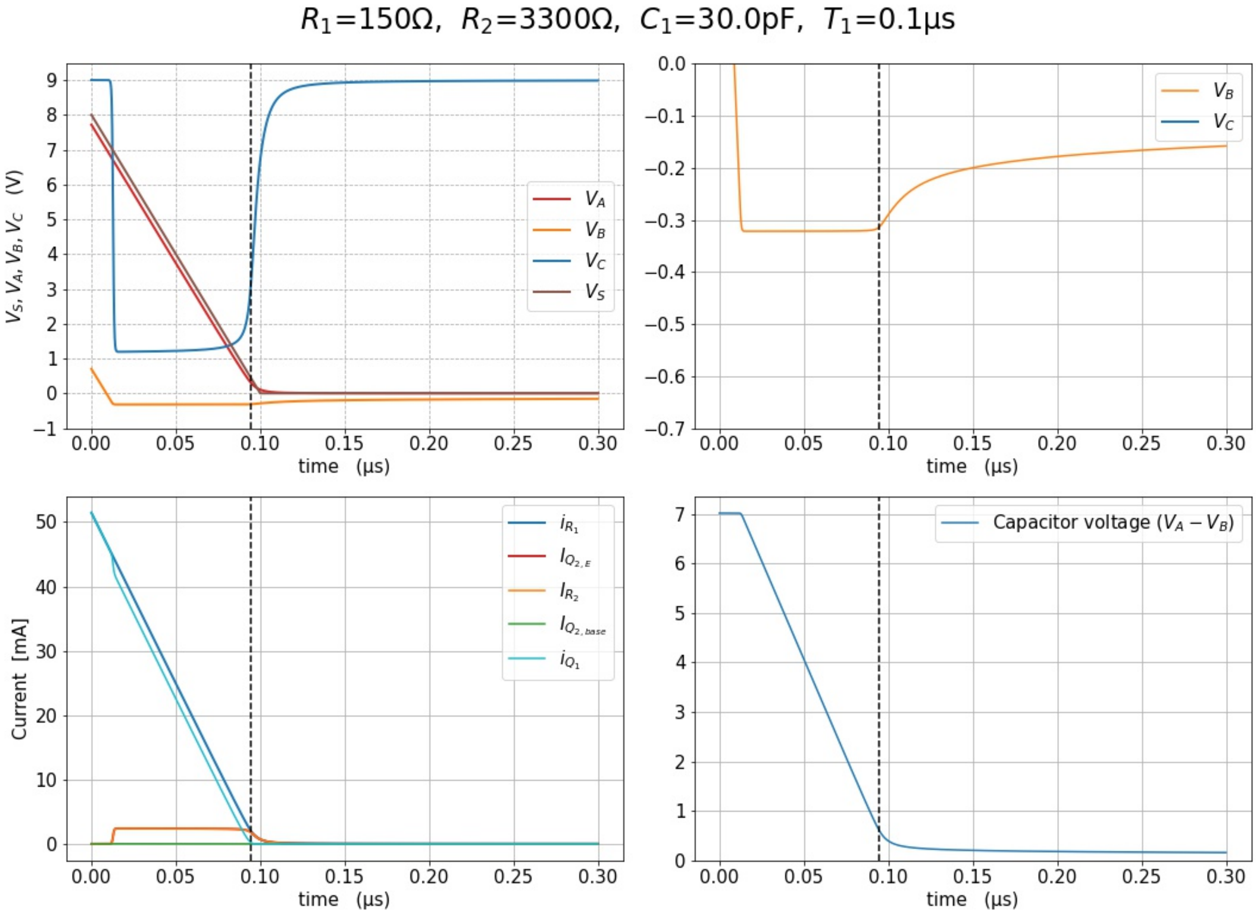
1	Q_1 in conduction. Q_2 in cutoff or active region		
a	V_A is one diode drop below V_S	In this phase Q_1 works as an emitter follower.	
b	V_B drops from initial value to about -0.4V	This negative voltage at V_B depends on the value of C_1. When Q_2 gets into its active region i_{Q_2} discharges C_1. A larger C_1 needs a larger i_{Q_2} to discharge the same amount as a small C_1. Because V_A is dictated by V_S (Q_1 in active region) for a larger C_1. V_B needs to drop lower for for more discharge current i_{Q_2} to keep up with V_A. With a small enough value for C_1. Q_2 may not even reach saturation. This is a degenerate case.	1. larger C_1 -> lower V_B 2. very small C_1 -> Q_2 does not saturate
c	i_{Q1} drops with falling V_S	i_{Q1} depends on R_1 $i_{Q_1} \approx \frac{V_S - 0.6}{R_1} - i_{Q_2}$	
d	i_{Q2} rises because of dropping V_B until Q_2 saturates	Ideally the Ebers-Moll model should be used. but the Shockley diode equation multiplied by β is a good approximation: $i_{Q_2} = \beta \cdot I_S \cdot \left(e^{-\frac{V_B}{V_T}} - 1 \right)$ As long as Q2 is not saturated $i_{R_2} = i_{Q_2}$	
e	after Q_2 gets into active region i_{R2} rises quickly		
f	V_C drops quickly until it gets Q_2 in saturation > enter 2	$V_C = V_{CC} - i_{R_2}$	
2	Q_1 in active region. Q_2 in saturation		
a	V_B reaches a minimum at about -0.4V	for the minimum value of V_B see 1b.	

b	any current i_{Q2} which cannot be taken from the collector $i_{Q2.base}$ is taken from the base $i_{Q2.base}$	this can be a substantial amount	
b	C_1 discharges because of i_{Q2}	$i_{Q_2} = C_1 \cdot \frac{\mathrm{d}(V_B - V_A)}{\mathrm{d}t}$	
c	i_{Q2} does not contribute enough to keep V_A one diode drop below V_S so Q_1 stays in active region	In this phase $V_A > i_{Q_2}R_1$. When $V_A = i_{Q_2}R_1$ the current i_{Q_2} suffices to get V_A on the level as dictated by V_S. At that moment $i_{Q_1} = 0$ and Q_1 gets into cutoff. At larger R_1 the voltage induced at A induced by i_{Q_2} will be larger. so Q_1 will get into cutoff sooner. And. because V_B depends on the value of C_1 (see 1b). and i_{Q2} depends on V_B. the moment at which Q_1 gets into cutoff depends on C_1.	3. larger $C_1 \Rightarrow Q_1$ gets into cutoff sooner
d	V_B stays constant at about -0.4V	for the minimum value of V_B see 1b.	
e	when i_{Q2} can supply enough current to maintain V_A. Q_1 gets into cutoff > enter 3	$V_A = \left(i_{Q_1} + i_{Q_2}\right) \cdot R_1 = i_{Q_2} \cdot R_1$	4. larger $R_1 \Rightarrow Q_1$ gets into cutoff sooner
f	V_C at minimum value. about one diode drop below $V_{Q2.base}$		
3	Q_1 in cutoff. Q_2 in saturation		
a	i_{Q2} is large enough to maintain V_A	$V_A = i_{Q_2} \cdot R_1$ i_{Q_1} is not needed to maintain V_A. so Q_1 is in cutoff.	
b	V_A is dropping because of C_1 discharge by i_{Q2}	$i_{Q_2} = C_1 \cdot \frac{\mathrm{d}(V_B - V_A)}{\mathrm{d}t}$	
c	V_A and V_B are determined by the diminishing voltage over the capacitor and the balance equation	$V_A \cdot Z_B = -V_B \cdot R_1$ where $Z_B = -\frac{V_B}{i_{Q_2}}$ The larger R_1 the more of the voltage over the capacitor is placed over R_1. and thus the higher is V_A and lower V_B (less negative). and so smaller i_{Q2}.	

d	by the diminishing voltage over C_1. and the balance equation V_B slowly rises	see 3c.	
e	i_{Q2} drops		
f	$i_{Q2.base}$ drops as Q_2 is still in saturation		
g	Q_2 gets slowly out of saturation		
h	V_C slowly rises. until Q_2 gets out of saturation > enter 4	V_C slowly rises because of rising V_B and thus decreasing i_{Q_2}. V_B relates to i_{Q_2} and V_A with $i_{Q_2} = \frac{V_A}{R_1} = I_{ES} \cdot \left(e^{-\frac{V_B}{V_T}} - 1 \right)$ (or better even. with the Ebers-Moll equation). But also $i_{Q_2} = C_1 \cdot \frac{\mathrm{d}\left(V_B - V_A\right)}{\mathrm{d}t}$ So a larger C_1 will discharge slower for the same V_B and i_{Q_2}. Slower discharge will cause it to take longer to get out of saturation which will give a longer negative pulse on the collector of Q_2. Larger R_1 puts more voltage on V_A than with smaller R_1. This gives a smaller i_{Q_2}. so a smaller discharge rate. so a longer output pulse. Because a larger R_1 causes Q_1 to get into cutoff sooner. V_B will be less deep because of the balance equation. For very large R_1 and less deep V_B. Q_2 will get out of saturation sooner. so give a shorter output pulse.	5. larger C_1 -> longer negative output pulse on Q_2. 6. larger R_1 -> longer negative output pulse on Q_2. 7. even larger R_1 -> shorter output pulse on Q_2.
i	iR2 slowly decreases		
4	Q_2 out of saturation		

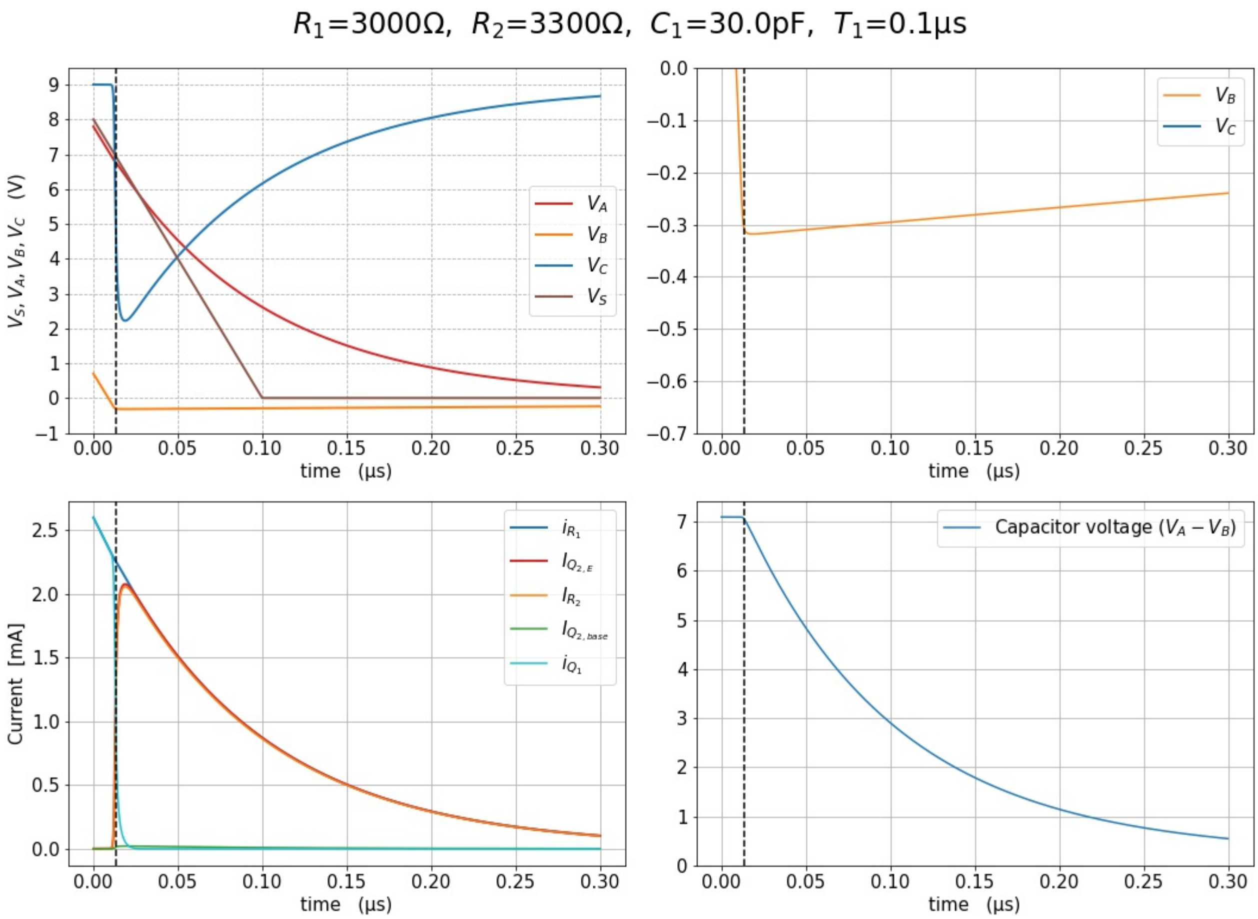
a	i_{Q2} still discharges C_1	The rate at which C_1 discharges is determined by i_{Q_2}. which in turns depends on VB. which depends on the voltage over C_1. the value of C_1 by $$i_{Q_2} = C_1 \cdot \frac{\mathrm{d}\left(V_B - V_A\right)}{\mathrm{d}t}$$ and the balance equation: $$i_{Q_2} = \frac{V_A}{R_1} = I_{ES} \cdot \left(e^{-\frac{V_B}{V_T}} - 1\right)$$ The balance equation divides the voltage over the capacitor over VA and VB. The larger R1. the smaller (in magnitude) VB. the smaller iQ2. the smaller the discharge rate. This translates to a less steep rise in VC in this phase.	8. larger R_1 -> less steep rise on output pulse 9. larger C_1 -> less steep rise on output pulse
b	because of C_1 discharge and balance equation. V_B rises and V_A drops		
c	i_{Q2} drops		
d	i_{R2} drops		
e	V_C rises	For the rate at which V_C rises see 4a	

3. Degenerate cases



The value of C_1 is too small to get Q_2 into saturation. Lowest value for V_B depends on the value of C_1 (see chapter 3.); larger C_1 gives more negative V_B. A more negative V_B gives a larger current through Q_2. If the current is too low. Q_2 will not get into saturation.

The value of R_1 is small enough to give a nice steep rising edge on the output pulse. If the value of R_1 is larger (as shown in the next figure). with the same value for C_1. The output pulse is further degenerated.

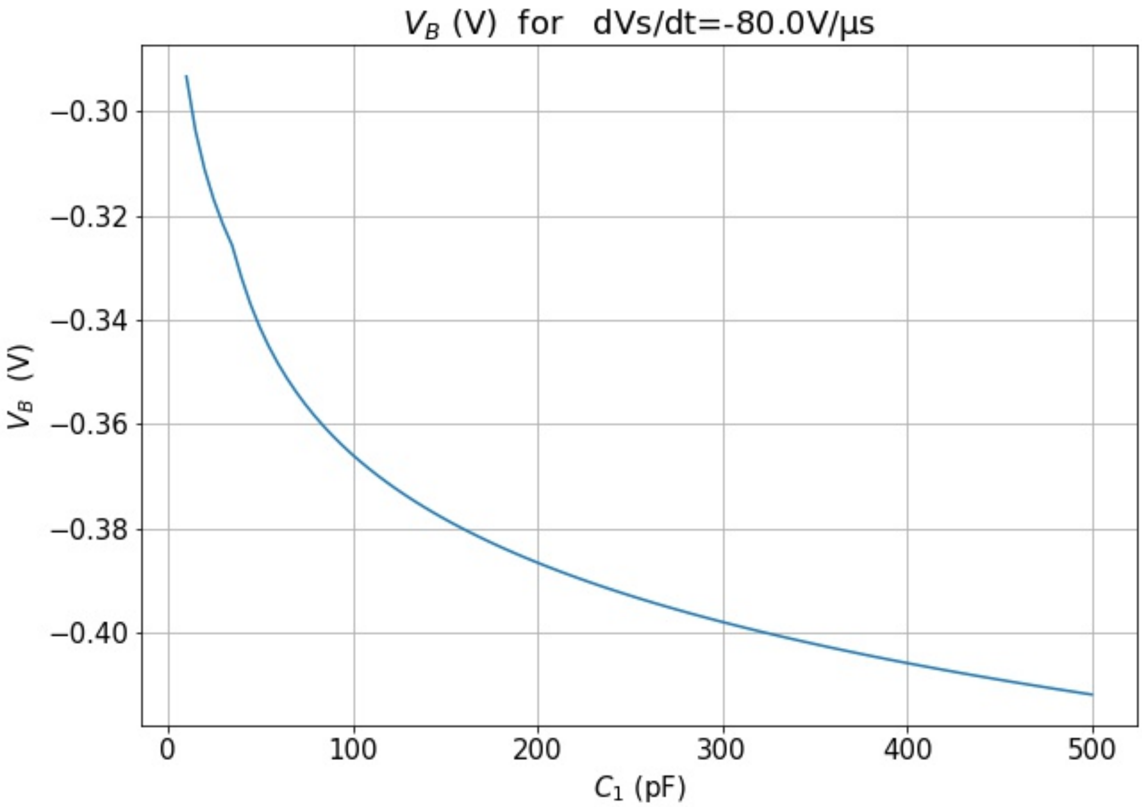


Besides a too low value for C_1. which will not let Q_2 get into saturation, the value for R_1 is too large. The too large value for R_1 causes a slow discharge of C_1 which shows as a less steep rising edge on the output pulse.

4. V_B dependency of C_1. R_1 and dV_s/dt

4.1 Q_1 in active region

The graph below shows the steady state value for V_B when Q_1 is in its active region. It depends on the value of C_1 and the rate at which V_S is dropping. It is independent of R_1. This is only valid when Q_1 is in its active region.



The capacitor discharge equation over C_1

$$i_{E,Q_2} = C_1 \left(\frac{\mathrm{d}V_B}{\mathrm{d}t} - \frac{\mathrm{d}V_A}{\mathrm{d}t} \right)$$

V_B is more or less constant when Q_1 is in its active region. With V_B constant

$$i_{E,Q_2} = -C_1 \frac{\mathrm{d}V_A}{\mathrm{d}t}$$

Because Q1 acts as a voltage follower (Q1 is in its active region)

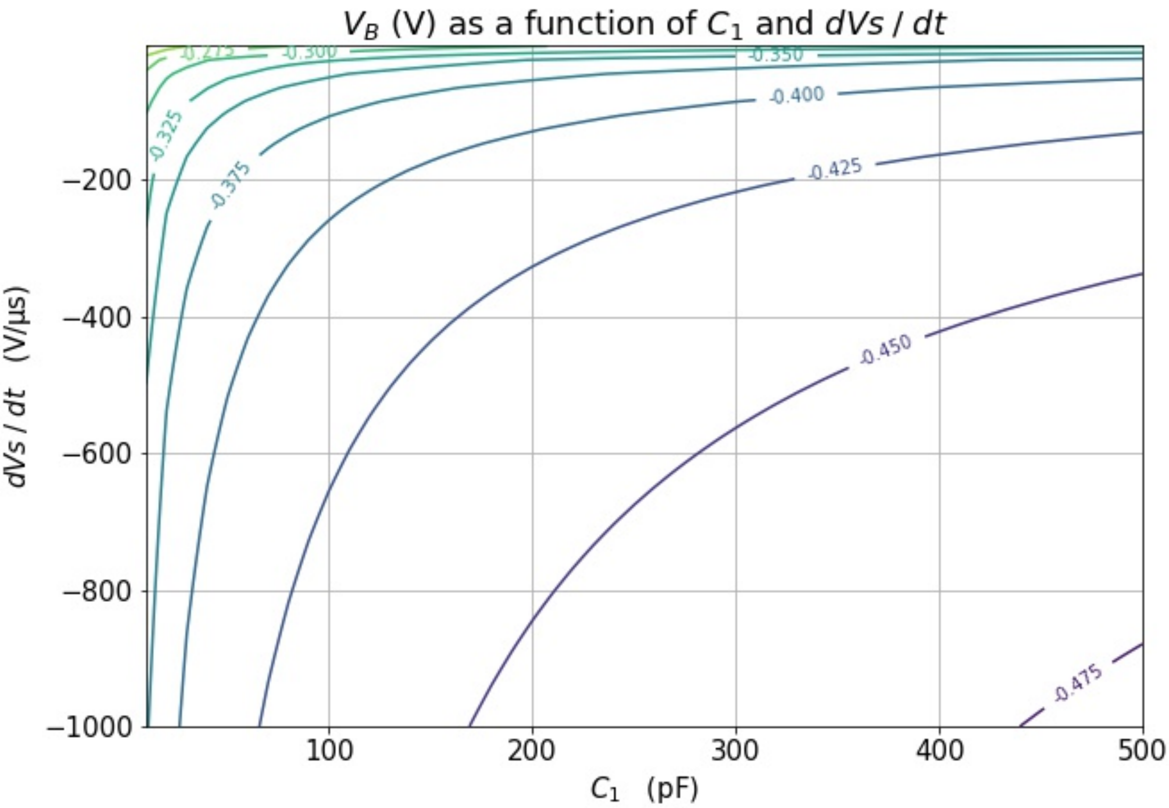
$$\frac{\mathrm{d}V_A}{\mathrm{d}t} = \frac{\mathrm{d}V_S}{\mathrm{d}t}$$

$$\Rightarrow \quad i_{E,Q_2} = -C_1 \frac{\mathrm{d}V_S}{\mathrm{d}t}$$

i_{E,Q_2} only depends on V_B (Ebers-Moll). C_1 is the variable in the above graph. To calculate the above graph a root finder is used to find V_B (using equations 2 and 3.) to fulfil the equation

$$i_{Q,E_2}\left(V_B\right) = -\,C_1\,\frac{\mathrm{d}V_S}{\mathrm{d}t}$$

It shows that with larger value for C_1. transistor Q_2 gets deeper in saturation.



The above figure shows that larger values for C_1 will give lower (more negative) values for V_B. Also larger absolute (more negative) values for $\frac{dV_S}{dt}$ gives lower (more negative) values for V_B. Larger absolute values for V_B (more negative) will get Q_2 into deeper saturation.

The value for V_B can be calculated by solving the capacitor current equation

$$i_{E,Q_2} = C_1\frac{\mathrm{d}V_{C_1}}{\mathrm{d}t}$$

If we assume that V_B is constant. we know that

$$\frac{\mathrm{d}V_{C_1}}{\mathrm{d}t} = \frac{\mathrm{d}\left(V_B - V_A\right)}{\mathrm{d}t} = -\,\frac{\mathrm{d}V_A}{\mathrm{d}t}$$

When Q_1 is in its active region. Q_1 acts as a voltage follower. and so

$$\frac{\mathrm{d}V_A}{\mathrm{d}t} = \frac{\mathrm{d}V_S}{\mathrm{d}t}$$

$$\Rightarrow \quad \frac{\mathrm{d}V_{C_1}}{\mathrm{d}t} = -\frac{\mathrm{d}V_S}{\mathrm{d}t}$$

Substitute:

$$i_{E,Q_2} + C_1 \frac{\mathrm{d}V_S}{\mathrm{d}t} = 0 \qquad (1)$$

i_{E,Q_2} can be calculated from V_B and V_C with the Ebers-Moll model:

$$i_{E,Q_2} = I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) - \alpha_R \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right) \qquad (2)$$

When V_C is known. i_{E,Q_2} can be calculated for any V_B. Next we will derive an expression for V_C in terms of V_B.

For the base current we have the following equation from the Ebers-Moll model

$$I_B = \left(1 - \alpha_F\right) \cdot I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) + \left(1 - \alpha_R\right) \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right)$$

and for the collector current we have the following equation (Ohm's law over R_2)

$$I_C = \frac{V_{CC} - V_C}{R_C}$$

Kirchhoff's current law over the transistor is

$$I_E - I_B - I_C = 0$$

where $I_E = i_{Q_2}$. Substitution of the expressions for I_E. I_B and I_C into the current law gives

$$I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) - \alpha_R \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right) - \left(1 - \alpha_F\right) \cdot I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) - \left(1 - \alpha_R\right) \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right) - \frac{V_{CC} - V_C}{R_C} = 0$$

which simplifies to

$$\alpha_F \cdot I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) - I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right) - \frac{V_{CC} - V_C}{R_C} = 0$$

From this V_C can be expressed in terms of V_B as follows

$$V_C = V_T \cdot W_0\left(\frac{R_C I_{CS}}{V_T} e^{\frac{AR_C}{V_T}}\right) - AR_C \tag{3}$$

where

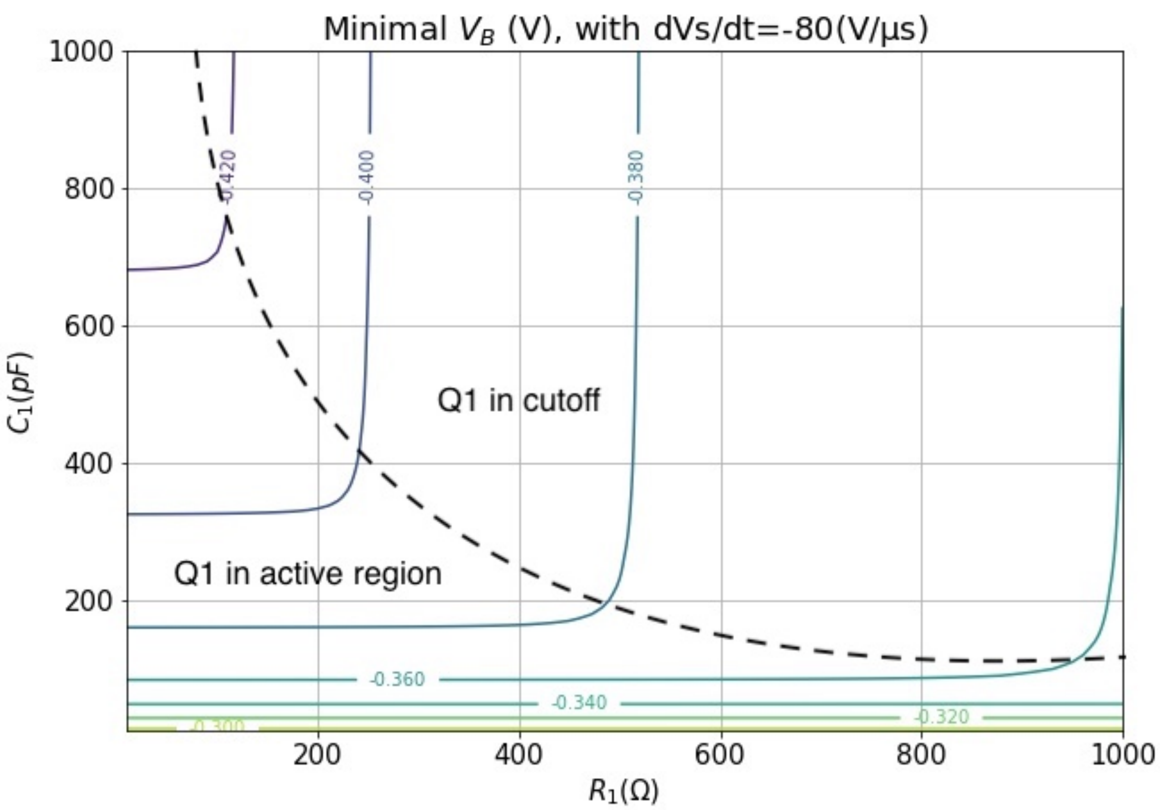
$$A = \alpha_F I_{ES}\left(e^{-\frac{V_B}{V_T}} - 1\right) + I_{CS} - \frac{V_{CC}}{R_C}$$

and W_0 is the Lambert W function.

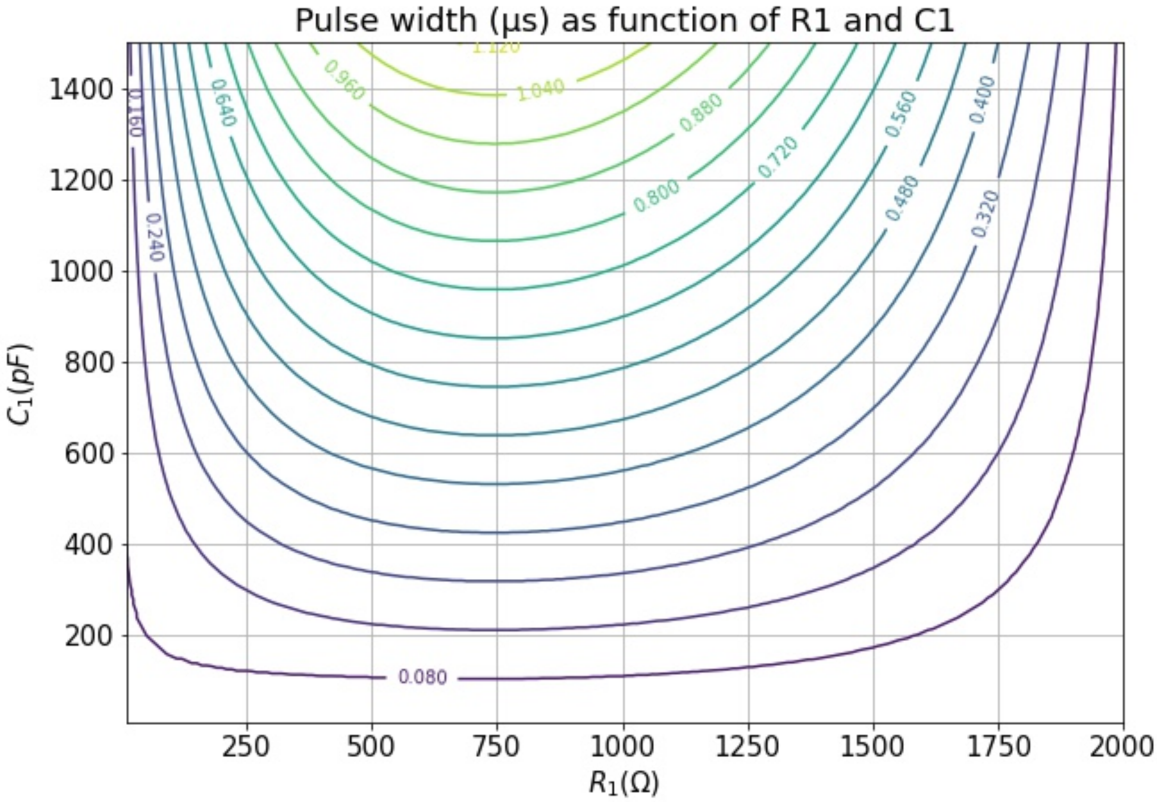
Inserting (3) into (2) allows to calculate i_{E,Q_2}. Using (1) in a root finding algorithm (eq. brentq) and solving for V_B will return V_B for any $\frac{dV_S}{dt}$ and C_1. Note that this is independent of R_1 and is only valid when Q_1 is in its active region.

4.2 Q_1 either in active region or cutoff

The graph below shows the dependency of the minimal value of V_B on R_1 and C_1 when Q_1 is in its active region or in cutoff. The horizontal part of the constant V_B lines are in the active region of Q_1 where the minimal value of V_B only depends on C1. and is independent of R_1. The vertical lines are in the cutoff region of Q_1. the minimal value of V_B depends on R_1 only.



5. Pulse width



The figure above shows lines for constant pulse width. In this context the pulse width is conveniently defined as the time between zero-crossing of the falling and rising edge of the pulse at the collector of Q_2. It should be noted that pulse width is less clear when the rising edge of the pulse is not steep. Especially in cases with large values C_1 and/or R_1 this definition does not work well.

We see that larger values for C_1 give larger pulse width. There are several counteracting effects of the value of C_1 on the pulse width.

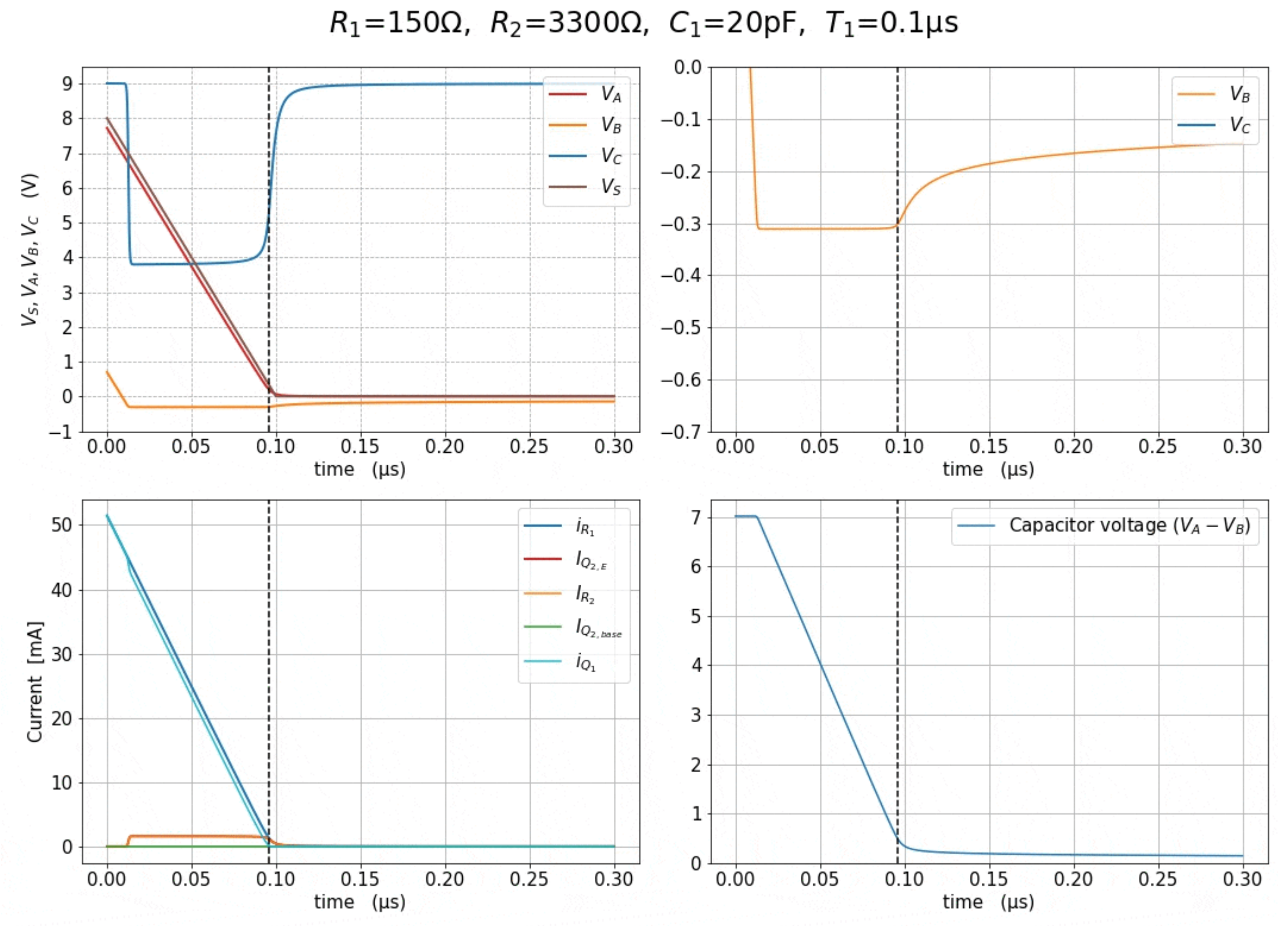
With larger C_1:

1. the value of V_B will be more negative (in times of Q_1 active region). so a larger discharge current → shorter output pulse
2. Q_1 gets into cutoff sooner. Cutoff of Q_1 means that V_A drops slower than V_S. Q_1 getting into cutoff means that C_1 discharges slower than than if Q_1 would not have been in cutoff → longer output pulse
3. more current is needed to discharge a larger value capacitor. This is especially true when Q_1 is in cutoff. → longer output pulse
4. less steep rising edge on the output pulse. although this does not play a role when using the current definition of pulse width. as this effect mainly takes place when V_C>0.

Apparently the effects which cause a longer output pulse with increasing C_1 win.

Note that for $R_1 \approx 750\Omega$ the pulse width linearly relates to C_1.

Below is an animation of increasing the value of C_1 .[1]



For increasing values of R_1 there is a maximum for the pulse width. which in the figure above seems to be between 750Ω < R_1 < 1000Ω. There are several counteracting effects of the value of R_1 on the pulse width.

With larger values of R_1:

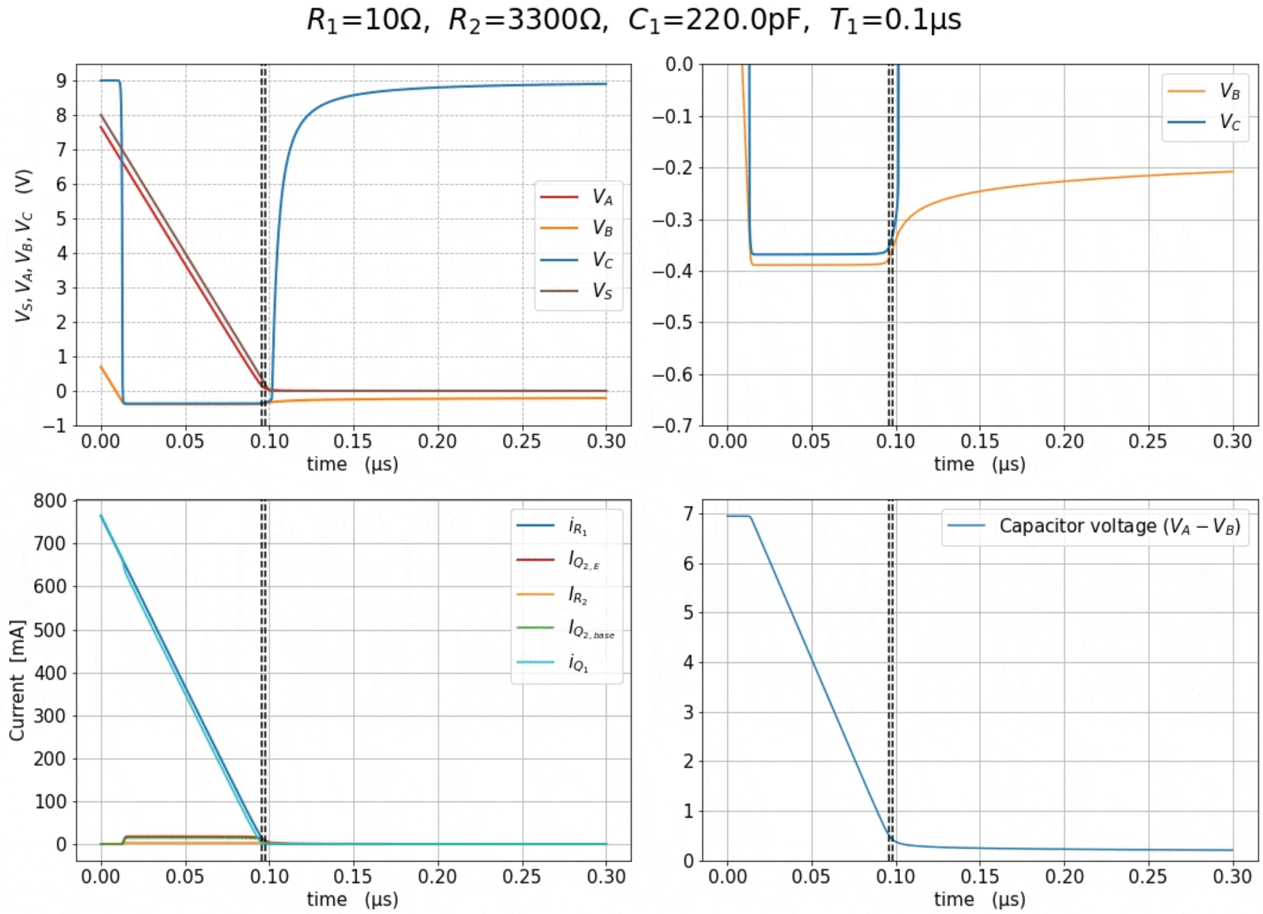
1. Q_1 gets into cutoff sooner. Cutoff of Q_1 means that V_A drops slower than V_S. Q_1 getting into cutoff means that C_1 discharges slower than than if Q_1 would not have been in cutoff → longer output pulse. This is the same effect as for a larger value of C_1.
2. with even larger R_1 cutoff of Q_1 will set in nearly immediately. As a result V_A will be dictated by the balance equation in which R_1 plays an important role. Larger R_1 will put more voltage of the capacitor on R_1 and less on the base-emitter junction of Q_2. Q_2 will get out of saturation sooner. as it has not been in deep saturation. This will give a shorter pulse width if we define it as the V_C zero crossing. But total discharging time of the capacitor will take longer. With very large values of R1 the circuit is getting into the degenerated area. Discharging of C1 in the saturated state of Q_2 will be shorter. but the overall discharging of C_1will take longer. with most of the time in the non-saturated state of Q_2. [2]
3. with even larger values for R_1 the detector gets in a degenerated state in which Q_2 does not saturate at all anymore. This starts happening for values of R_1 at which Q_1 gets into cutoff immediately. At these values for R_1. V_B does not drop as low as when Q_1 is in its active region.

---

[1] The animation shows only in mac Pages. The animation can be downloaded separately from Github. filename = "C1 series. R1=150.gif".

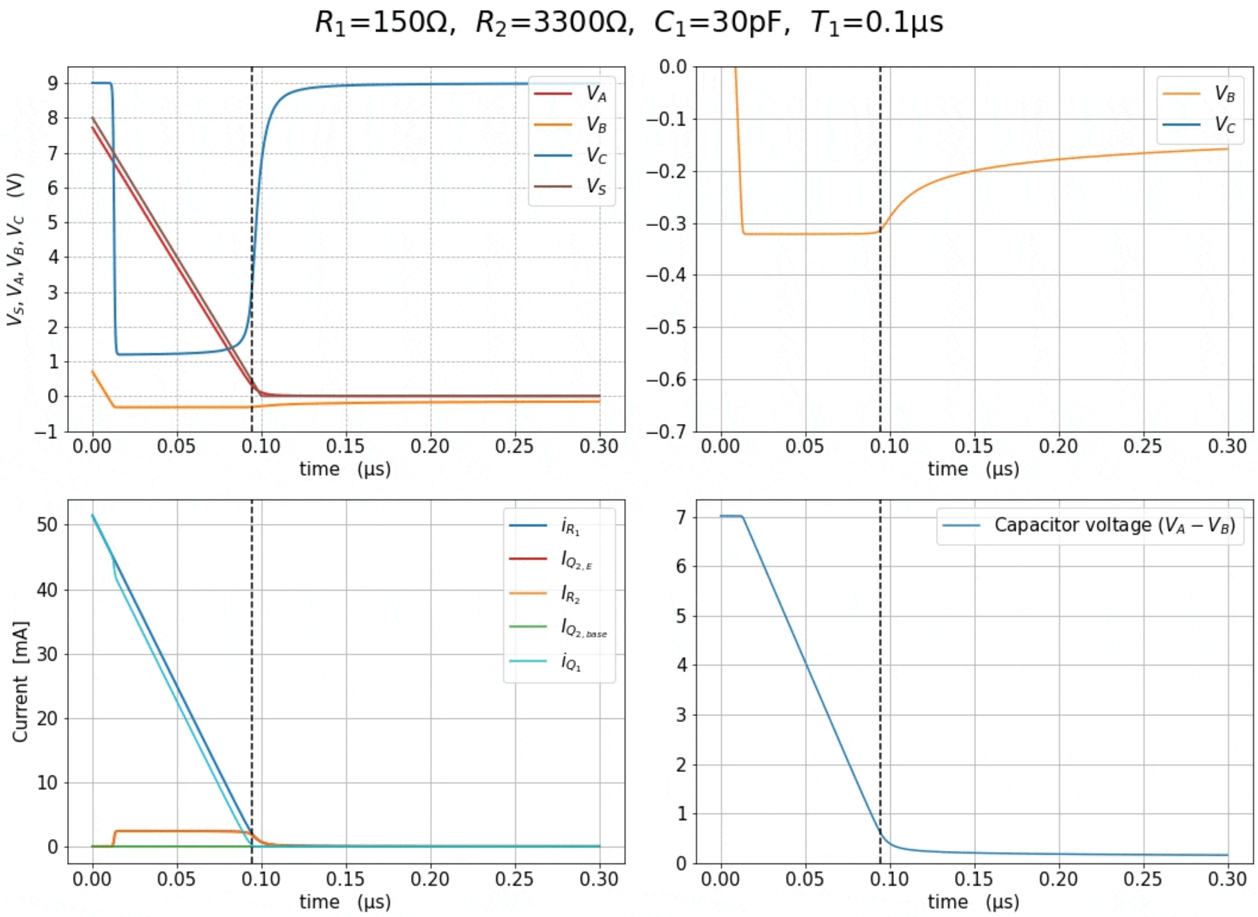
[2] In fact pulse width cannot be defined as the V_c zero-crossing for large values of R_1.

The following animation shows a series of increasing R_1 values. leading ultimately into a degenerate pulse on the output. [3]



3 The animation only shows in mac Pages. This animation can be downloaded separately from Github. filename = "R1 Series. C1=220pF.gif"

The next animation shows a degenerate case. with increasing R_1 and Q_2 not getting into saturation [4].



The definition of pulse width as the time between falling zero crossing of V_C and rising zero crossing can only be used when Q_2 gets well into saturation. If Q_2 does not saturate the zero crossing is less abrupt. and also the rising slope of the output pulse is less steep. In these cases one cannot really speak of a pulse. These are degenerated cases of the pulse train detector.

6. What are desired features of a pulse train FM detector?

1. The pulse width of the pulse at the output should be independent of frequency.
2. The pulse width should be as large as possible for maximum dynamic range. Of course the pulse width must be less than the half period of the input signal. Also there should be sufficient margin to allow for highest temporal frequency input signals to be processed well. The maximum frequency deviation in the intermediate frequency signal is equal to the maximum frequency deviation of the original FM signal at FM frequencies. For FM radio the maximum frequency deviation is ±75kHz. The period of a 75kHz signal is 13μs. This means that the maximum pulse width can be 13/2=6.5μs.
3. Each pulse in the pulse train should be independent of its predecessor. When Q_2 leaves its saturated state further discharge of C_1 takes place at a much slower rate than when Q_2 is in saturation. When either C_1 and/or R_1 have large values the rising edge on the output pulse is not steep. Before the next falling slope of the input signal. an upward swing pulls V_B high. so

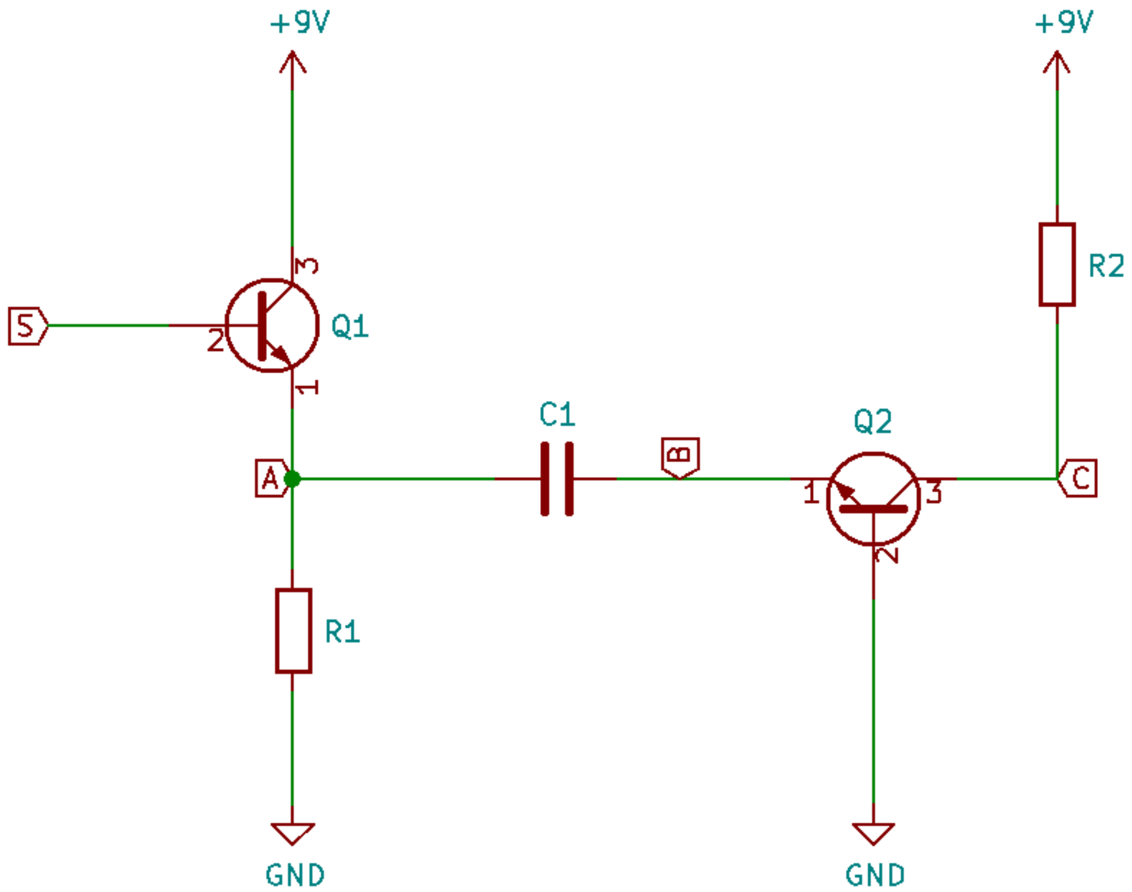
[4] The animation only shows in mac Pages. This animation can be downloaded separately from Github. filename = "R1 Series. C1=30pF.gif"

Q_2 will stop conducting. which in turn will pull V_C high. So. even if V_C has not reached V_{CC} before the input signal rises again because of high values of R_1 and/or C_1. VC will be at VCC for the next pulse. This make the pulse independent of the previous pulse.

4. The relation duty cycle vs. the input signals frequency deviation should be linear. Large R1 and/or C1 values will cause VC to be pulled abruptly high to VCC as described in the previous point. This will affect the linearity of the detector.
5. Steepest rising edges are obtained when R_1 has a small value. Smaller R1 implies more power consumption in the emitter follower stage Q1. Take care not to exceed the maximum collector current for Q_1 when using a small R_1.

7. How are the plots generated?

The plots in this document are generated with a Jupyter notebook which models the pulse train FM detector. The notebook models the following circuit:



For simplicity D_1 in the original circuit has been removed. In the real circuit D_1 is needed to limit V_B to about +0.6V. In the simulation this part of the circuit is not needed. Including D_1 in the simulation would only make the model more complicated [5].

[5] The differential equation describing the voltage at V_B would be a bit more complicated. although probably not by much. In a next revision of the model it may be implemented.

The Jupyter notebook can be downloaded from Github.

7.1 Mathematical model

The circuit is described by

- Ebers-Moll model transistor for Q_1
- Ebers-Moll model transistor for Q_2
- Kirchhoff's current law for Q_2
- Kirchhoff's current law at A
- capacitor current-voltage relation for C_1
- Ohm's law over R_1 and R_2

These simple laws are put into
- one equation to calculate V_C for any V_B and R_2
- one differential equation which describes the voltage change over C_1 in time.

These two relations entirely describe the above circuit and will (also) let us calculate the most important aspect of this circuit. V_C as a function of time.

More equations are implemented in this notebook to test the model. and to implement helper functions. They are not essential to the model however.

7.1 Transistor equations

The Ebers-Moll equations for an NPN transistor are as follows. For the Q_2 emitter current I_E we have the following:

$$I_E = I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1 \right) - \alpha_R \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1 \right)$$

For the Q_2 base current I_B we have the following Ebers-Moll equation

$$I_B = \left(1 - \alpha_F\right) \cdot I_{ES} \cdot \left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1 \right) + \left(1 - \alpha_R\right) \cdot I_{CS} \cdot \left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1 \right)$$

and for the Q_2 collector current I_C we use Ohm's law over R_2

$$I_C = \frac{V_{CC} - V_C}{R_2}$$

where $I_{ES} = \beta_F I_S$ and $I_{CS} = \beta_R I_S$. with $\beta_F = \dfrac{\alpha_F}{1 - \alpha_F}$ and $\beta_R = \dfrac{\alpha_R}{1 - \alpha_R}$.

Kirchhoff's current law over Q_2 is

$$I_E - I_B - I_C = 0$$

Substitution gives

$$I_{ES} \cdot \left(e^{-\frac{V_B}{V_T}} - 1 \right) - \alpha_R \cdot I_{CS} \cdot \left(e^{-\frac{V_C}{V_T}} - 1 \right) - \left(1 - \alpha_F\right) \cdot I_{ES} \cdot \left(e^{-\frac{V_B}{V_T}} - 1 \right) - \left(1 - \alpha_R\right) \cdot I_{CS} \cdot \left(e^{-\frac{V_C}{V_T}} - 1 \right) - \frac{V_{CC} - V_C}{R_2} = 0$$

which simplifies to

$$\boxed{\alpha_F \cdot I_{ES} \cdot \left(e^{-\frac{V_B}{V_T}} - 1 \right) - I_{CS} \cdot \left(e^{-\frac{V_C}{V_T}} - 1 \right) - \frac{V_{CC} - V_C}{R_2} = 0 \qquad (4)}$$

With equation (4) V_C can be solved for any V_B and R_2 (numerically with a root finder or with the Lambert W function. also see chapter 4.1)

7.2 Circuit equations

The Kirchhoff's current law at A tells us that

$$i_{R_1} = i_{E,Q_1} + i_{E,Q_2} \qquad (5)$$

where

$$i_{E,Q_1} = I_{ES} \left(e^{\frac{V_S - V_A}{V_T}} - 1 \right) - \alpha_R \; I_{CS} \left(e^{\frac{V_S - V_{CC}}{V_T}} - 1 \right) \qquad (6)$$

and

$$i_{R_1} = \frac{V_A}{R_1} \qquad (7)$$

The equation for i_{E,Q_2} from the Ebers-Moll model is

$$i_{E,Q_2} = I_{ES} \left(e^{-\frac{V_B}{V_T}} - 1 \right) - \alpha_R I_{CS} \left(e^{-\frac{V_C}{V_T}} - 1 \right) \qquad (8)$$

Substitution of (6). (7) and (8) into (5) gives

$$\boxed{\frac{V_A}{R_1} - I_{ES}\left(\mathrm{e}^{\frac{V_S - V_A}{V_T}} - 1\right) + \alpha_R\, I_{CS}\left(\mathrm{e}^{\frac{V_S - V_{CC}}{V_T}} - 1\right) = I_{ES}\left(\mathrm{e}^{-\frac{V_B}{V_T}} - 1\right) - \alpha_R\, I_{CS}\left(\mathrm{e}^{-\frac{V_C}{V_T}} - 1\right) \quad (9)}$$

This equation will let us calculate V_A for any given V_B and V_S (numerically. with a root finder or with the Lambert W function). Remember that V_C can be calculated from V_B with eq. (4).

Differentiating (9) with respect to time

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{ i_{R_1} = i_{E,Q_1} + i_{E,Q_2} \right\}$$

gives

$$\frac{1}{R_1}\frac{\mathrm{d}V_A}{\mathrm{d}t} = \frac{I_{ES}}{V_T}\, e^{\frac{V_S - V_A}{V_T}}\, \frac{\mathrm{d}\left(V_S - V_A\right)}{\mathrm{d}t} - \alpha_R\, I_{CS}\, \mathrm{e}^{\frac{V_S - V_{CC}}{V_T}}\, \frac{1}{V_T}\, \frac{\mathrm{d}V_S}{\mathrm{d}t} - \frac{I_{ES}}{V_T}\, e^{-\frac{V_B}{V_T}}\, \frac{\mathrm{d}V_B}{\mathrm{d}t} + \frac{\alpha_R I_{CS}}{V_T}\, e^{-\frac{V_C}{V_T}}\, \frac{\mathrm{d}V_C}{\mathrm{d}t}$$

(10)

multiplying both sides with $R_1 V_T$

$$V_T\frac{\mathrm{d}V_A}{\mathrm{d}t} = R_1 I_{ES}\, e^{\frac{V_S - V_A}{V_T}}\, \frac{\mathrm{d}\left(V_S - V_A\right)}{\mathrm{d}t} - \alpha_R\, R_1\, I_{CS}\, \mathrm{e}^{\frac{V_S - V_{CC}}{V_T}}\, \frac{\mathrm{d}V_S}{\mathrm{d}t} - R_1 I_{ES}\, e^{-\frac{V_B}{V_T}}\, \frac{\mathrm{d}V_B}{\mathrm{d}t} + R_1\alpha_R\, I_{CS}\, e^{-\frac{V_C}{V_T}}\, \frac{\mathrm{d}V_C}{\mathrm{d}t}$$

(11)

With

$$\frac{\mathrm{d}V_C}{\mathrm{d}t} = \frac{\mathrm{d}V_C}{\mathrm{d}V_B} \cdot \frac{\mathrm{d}V_B}{\mathrm{d}t}$$

and

$$\frac{\mathrm{d}\left(V_S - V_A\right)}{\mathrm{d}t} = \frac{\mathrm{d}V_S}{\mathrm{d}t} - \frac{\mathrm{d}V_A}{\mathrm{d}t}$$

Now define some variables

$$X = \frac{\mathrm{d}V_A}{\mathrm{d}t}$$

$$Y = \frac{\mathrm{d}V_B}{\mathrm{d}t}$$

$$S_1 = R_1\ I_{ES}\ e^{\frac{V_S - V_A}{V_T}}$$

$$S_2 = \alpha_R\ R_1\ I_{CS}\ \mathrm{e}^{\frac{V_S - V_{CC}}{V_T}}$$

$$Q = \alpha_R\ R_1\ I_{CS}\ e^{-\frac{V_C}{V_T}}$$

$$B = R_1\ I_{ES}\ e^{-\frac{V_B}{V_T}}$$

Now we express eq. (11) as

$$V_T X = S_1 \left(\frac{\mathrm{d}V_S}{\mathrm{d}t} - X \right) - S_2 \frac{\mathrm{d}V_S}{\mathrm{d}t} - R_1\ I_{ES}\ Y\ e^{-\frac{V_B}{V_T}} + \alpha_R\ R_1\ I_{CS}\ Y\ e^{-\frac{V_C}{V_T}}\ \frac{\mathrm{d}V_C}{\mathrm{d}V_B} \quad (12)$$

$$\Rightarrow\ V_T X = S_1 \left(\frac{\mathrm{d}V_S}{\mathrm{d}t} - X \right) - S_2 \frac{\mathrm{d}V_S}{\mathrm{d}t} + Y \left(Q \frac{\mathrm{d}V_C}{\mathrm{d}V_B} - B \right) \quad (13)$$

With the capacitor charging equation

$$i_{E,Q_2} = C_1 \frac{\mathrm{d}V_{C_1}}{\mathrm{d}t} = C_1 \frac{\mathrm{d}\left(V_B - V_A\right)}{\mathrm{d}t} = C_1 \left(\frac{\mathrm{d}V_B}{\mathrm{d}t} - \frac{\mathrm{d}V_A}{\mathrm{d}t} \right) = C\,(Y - X) \quad (14)$$

So

$$X = Y - \frac{i_{E,Q_2}}{C_1} \quad (15)$$

Substitute (15) into (13)

$$V_T \left(Y - \frac{i_{E,Q_2}}{C_1} \right) = S_1 \left(\frac{\mathrm{d}V_S}{\mathrm{d}t} - \left(Y - \frac{i_{E,Q_2}}{C_1} \right) \right) - S_2 \frac{\mathrm{d}V_S}{\mathrm{d}t} + Y \left(Q \frac{\mathrm{d}V_C}{\mathrm{d}V_B} - B \right)$$

From this extract Y and substitute $Y = \frac{\mathrm{d}V_B}{\mathrm{d}t}$

$$\boxed{\frac{\mathrm{d}V_B}{\mathrm{d}t} = \frac{i_{E,Q_2}\left(S_1 + V_T\right) + (S_1 - S_2)\ C_1\ \frac{\mathrm{d}V_S}{\mathrm{d}t}}{C_1 \left(S_1 + V_T + B - Q \frac{\mathrm{d}V_C}{\mathrm{d}V_B} \right)} \quad (16)}$$

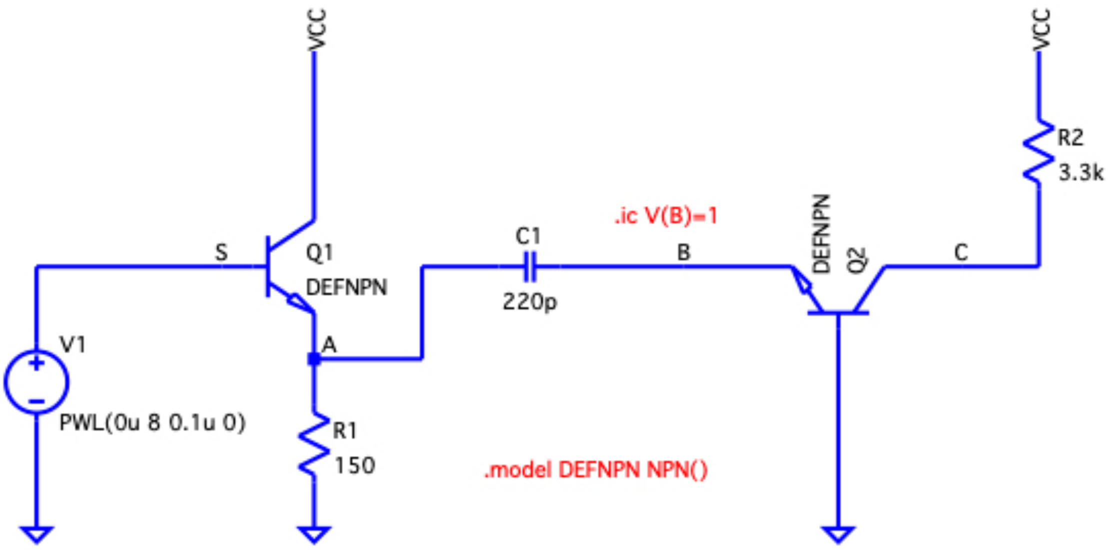
This equation for $\frac{\mathrm{d}V_B}{\mathrm{d}t}$ is a first order differential equation described in terms of V_A. V_B. V_C. i_{E,Q_2} and $\frac{\mathrm{d}V_S}{\mathrm{d}t}$ with V_B the only independent variable. V_C can be calculated from V_B using equation (4). V_A can be calculated from V_B using equation (9). i_{E,Q_2} can be calculated from V_B using equation (8). $\frac{\mathrm{d}V_S}{\mathrm{d}t}$ is a known given from the input signal. $\frac{\mathrm{d}V_C}{\mathrm{d}V_B}$ can be calculated numerically from equation (4). Solving $\frac{\mathrm{d}V_B}{\mathrm{d}t}$ numerically solves the entire circuit.

7.3 Jupyter Notebook Implementation details

- using the Lambert W functions allows for simpler code than with a root finder. Care must be taken to prevent numerical overflows. In Python use the mpmath module for increased precision / extended range.

8. Accordance with LTSpice model

With a default NPN transistor for Q_1 and Q_2 using SPICE directive `.model DEFNPN NPN()` very similar results are obtained between the Jupyter notebook and the LTSpice model [6].



Note that in the LTSpice a bare NPN model is used. Apart from the Ebers-Moll model none of the second order parasitic capacitances and Early voltage etc. are implemented [7].

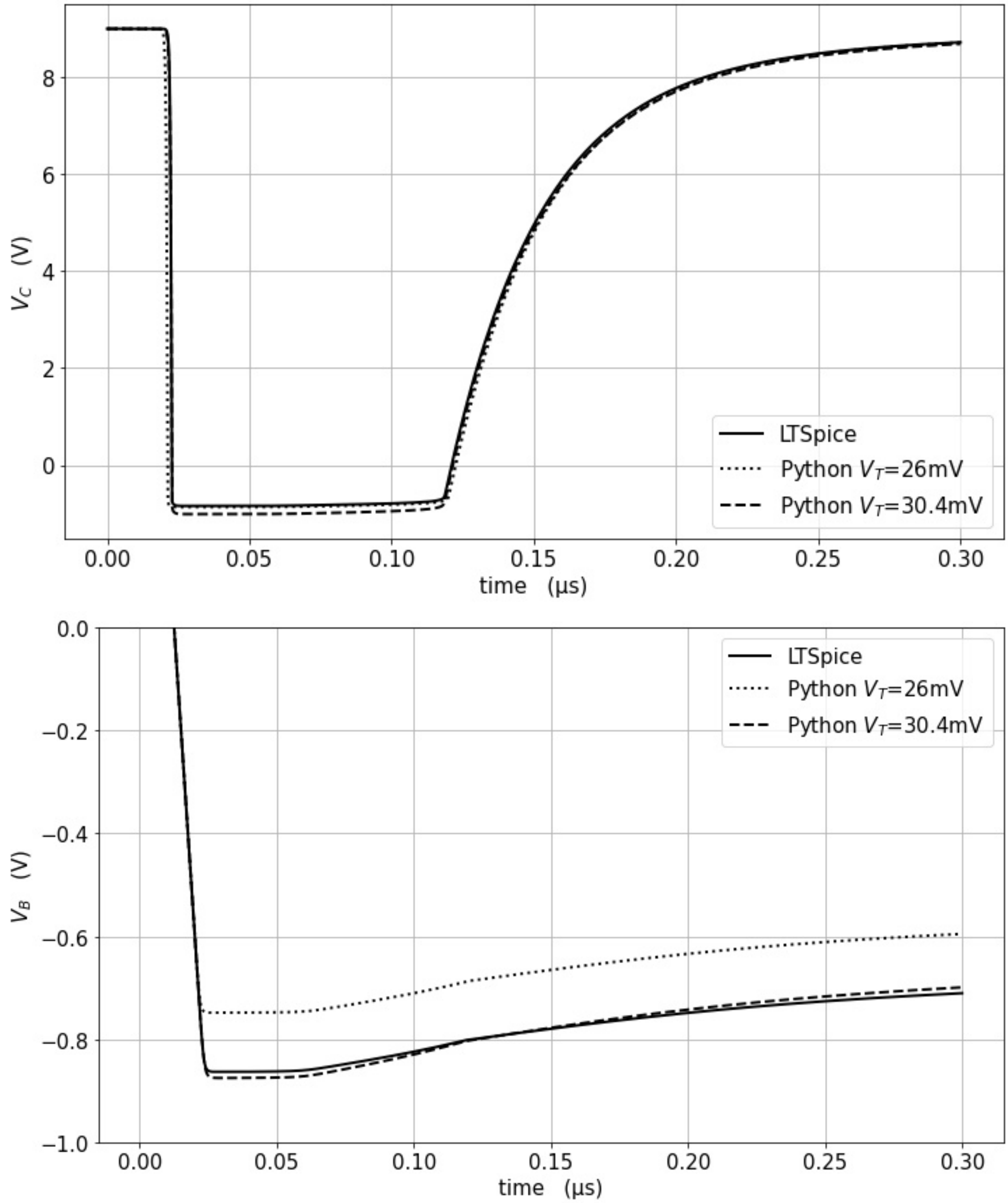
The graphs below show that the model as implemented in the Jupyter notebook agrees fairly well with the LTSpice model. Agreement can be even made more close by adjusting the thermal voltage in the Jupyter notebook. The default model uses 26mV. but agreement is closer when V_T=30.4mV is used. A thermal voltage of V_T corresponds to a temperature of

[6] Differences may be caused by different numerical methods, step sizes etc.

[7] Such a bare model can be used to find the cause of small effects in a real world circuit by adding these effects to the bare model one by one.

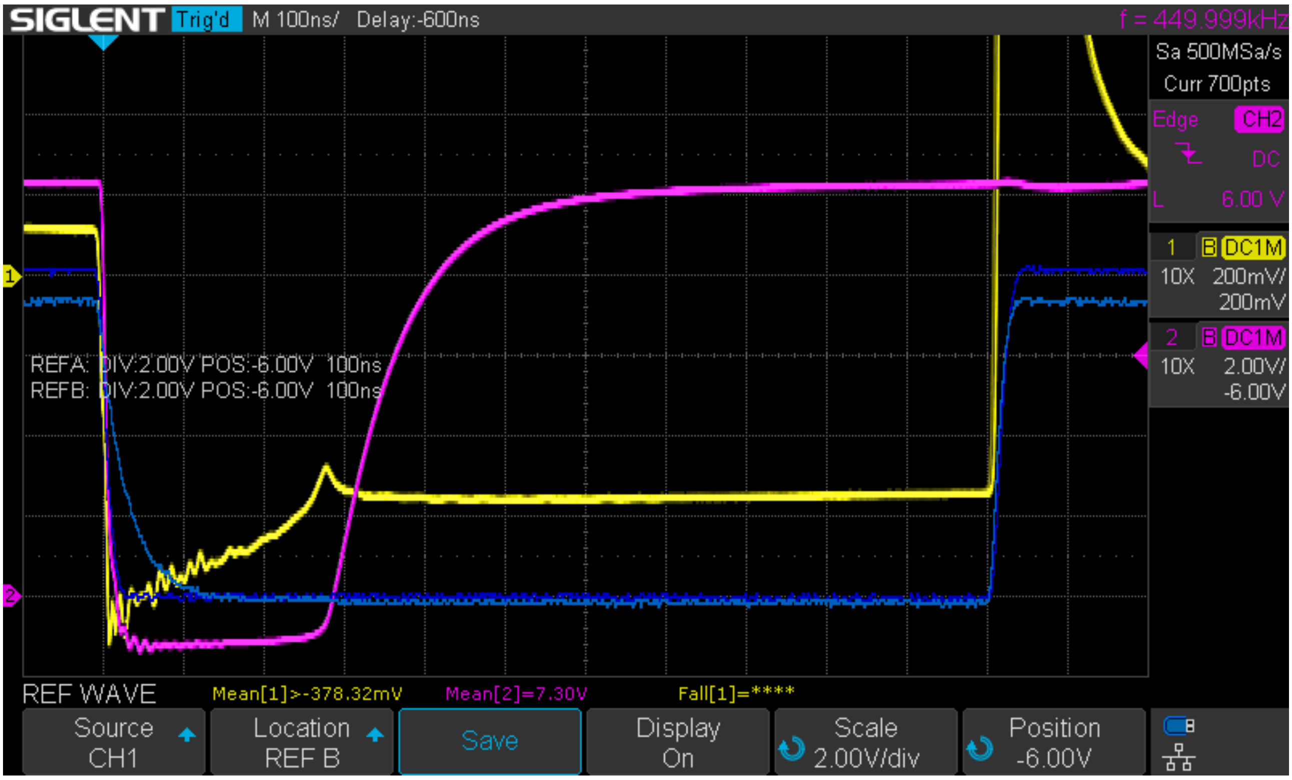
$$V_T = \frac{kT}{q} \Rightarrow T = \frac{qV_T}{k} = \frac{1.60 \cdot 10^{-19} \cdot 30.4 \cdot 10^{-3}}{1.38 \cdot 10^{-23}} = 353\text{ K} = 80\text{ °C}$$

with k the Boltzmann constant. and q the elementary electron charge.



Comparison of Jupyter Notebook model and LTSpice model. Both with C_1=220pF and R_1=150Ω. Top plot show V_C. bottom plot shows V_B.

9. Checking with a real world circuit



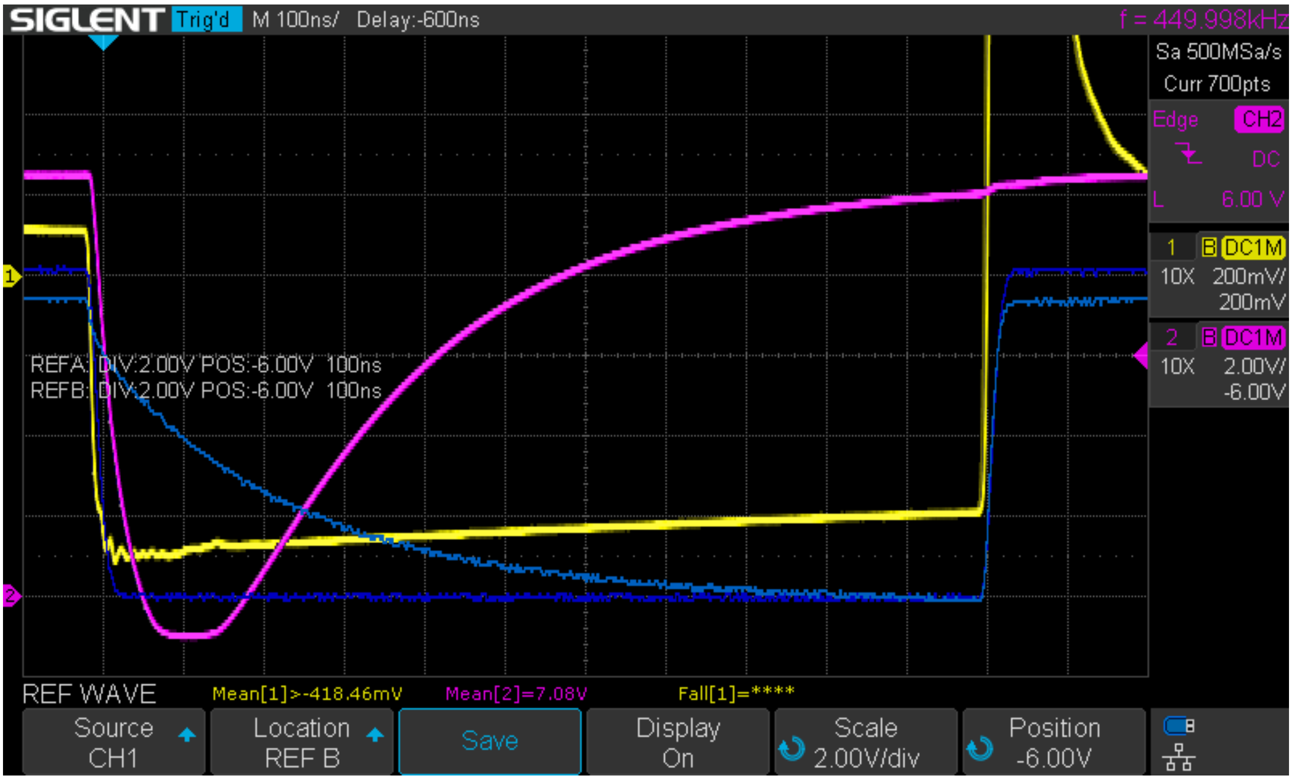
V_B: yellow. V_C: pink. V_A: light blue. V_S: dark blue with C_1=220pF and R_1=150Ω.

In the above oscilloscope plot we see an output pulse (V_C) of about 200ns. This does not agree very well with the Jupyter Notebook model which gives an output pulse of about xxxns. The LTSpice model produces an output pulse of about 350ns when using 2N3904 transistors.

The table below shows the pulse width (C_1=220pF. R_1=150Ω)

model	pulse width (ns)
LTSpice MYNPN	95
Jupyter	100
real world circuit	200
LTSpice 2N3904	350

Probably second order effects in the transistors play a big deal in the exact pulse width.



V_B: yellow. V_C: pink. V_A: light blue. V_S: dark blue with C_1=220pF and R_1=1100Ω.

In the above oscilloscope plot R_1 prevents Q_2 to get well into saturation. Q_1 gets into cutoff too early. Q_1 gets into cutoff before it has had a chance to push V_B down enough to get Q_2 into saturation. This leads to a badly defined output pulse.

10. Duty cycle

Define some variable

f_c: carrier frequency
f_Δ: frequency deviation
$f_{min} = f_c - f_\Delta$: minimal instantaneous frequency
$f_{max} = f_c + f_\Delta$: maximal instantaneous frequency
p_{max}: maximal pulse width
p_{min}: minimal pulse width (=0, but irrelevant)
d: duty cycle of output signal
d_{max}: maximal positive duty cycle of output
d_{min}: minimal positive duty cycle of output

The maximal pulse width $p_{max} = \frac{1}{2f_{max}}$
By design the minimal positive duty cycle $d_{min} = 0.5$.

The duty cycle $d = \frac{\frac{1}{f} - p}{\frac{1}{f}} = 1 - p \cdot f$

Pulse width p is fixed by circuit design, f is variable by nature (frequency modulated signal).

$$f_{min} \ < \ f \ < \ f_{max}$$

$$f_c - f_\Delta \ < \ f \ < \ f_c + f_\Delta$$

$$0 \ < \ p \ < \ \frac{1}{2f_{max}} = \frac{1}{2\left(f_c + f_\Delta\right)} = p_{max}$$

$$d = 1 - p \cdot f$$

$$d_{max} = 1 - p_{min} f_{min} = 1$$

$$d_{min} = 1 - p_{max} f_{max}$$

$$\Delta d = d_{max} - d_{min} = 2 \cdot p \cdot f_\Delta$$

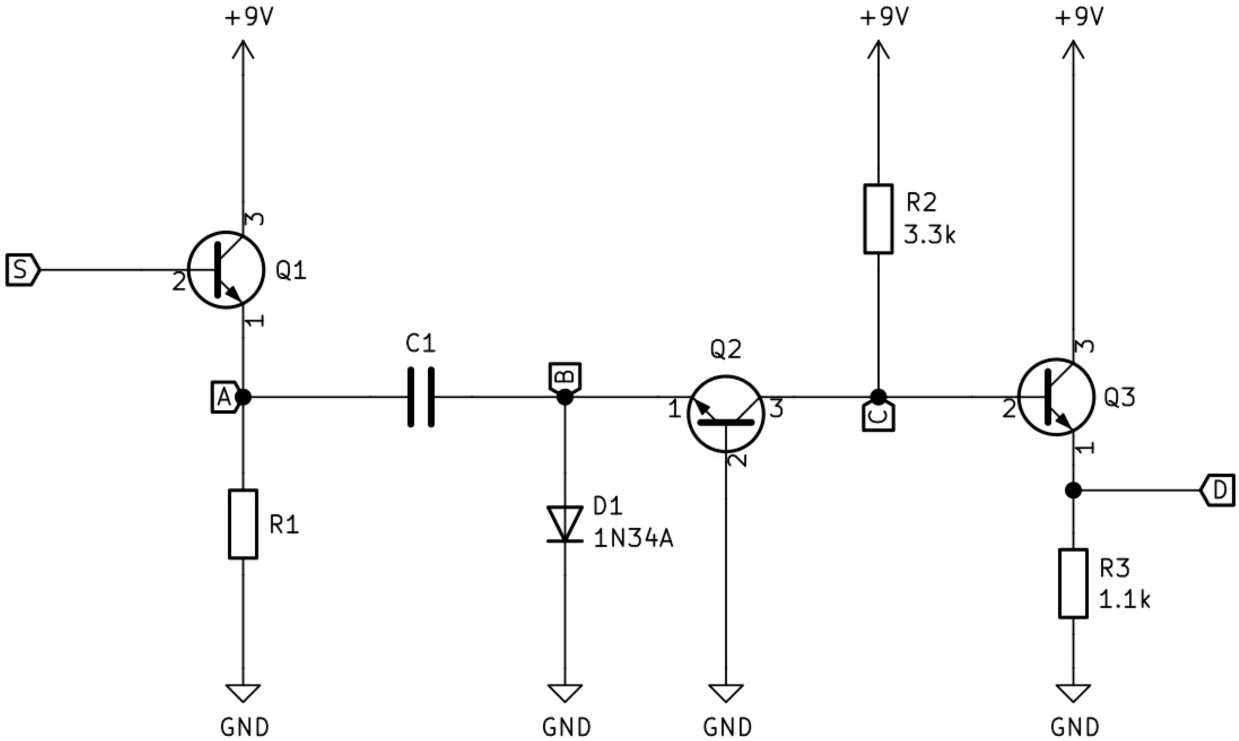
Δd is the duty cycle difference between maximal and minimal instantaneous frequency. It should be maximised to get the highest output from the detector.

f_Δ cannot be changed as it is dictated by the frequency modulated input signal. That leaves the pulse width p to be maximised. With $p_{max} = \frac{1}{2(f_c + f_\Delta)}$ only the carrier frequency f_c can be minimised to maximise Δd. Of course the carrier frequency f_c should still be higher than the maximal frequency deviation f_Δ.

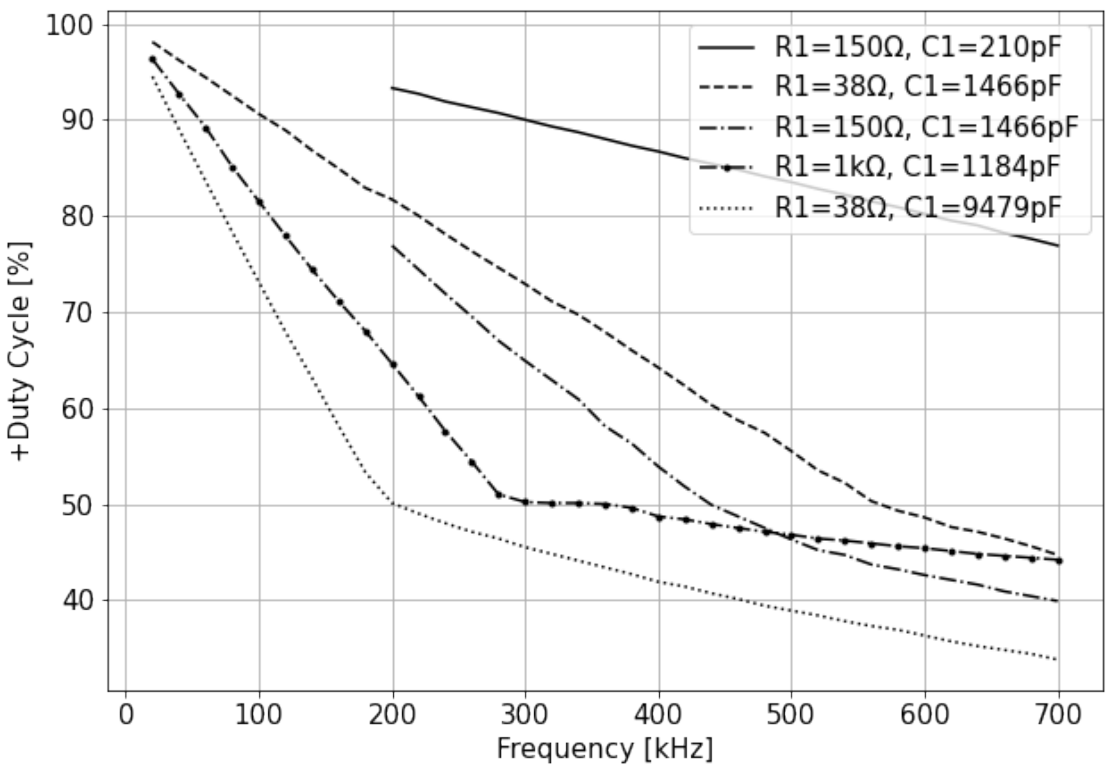
If we choose $f_c = f_\Delta$, and then choose $p = p_{max} = \frac{1}{4f_\Delta}$, we will get a maximum output signal with $\Delta d = 2 \cdot p \cdot f_\Delta = 2\frac{f_\Delta}{4f_\Delta} = 0.5$. This is as can be expected. The duty cycle can never be higher than 100% (with $p = 0$), and never be lower than 50% (with $p = p_{max} = \frac{1}{2(f_c + f_\Delta)}$).

11. Measurements

11.1 Linearity



Circuit to measure the duty cycle of the pulse train detector. Input signal a 8Vpp square pulse at S. Output measured at D with Siglent SDS1202X-E oscilloscope, +duty cycle measurement. All transistors are 2N3904.



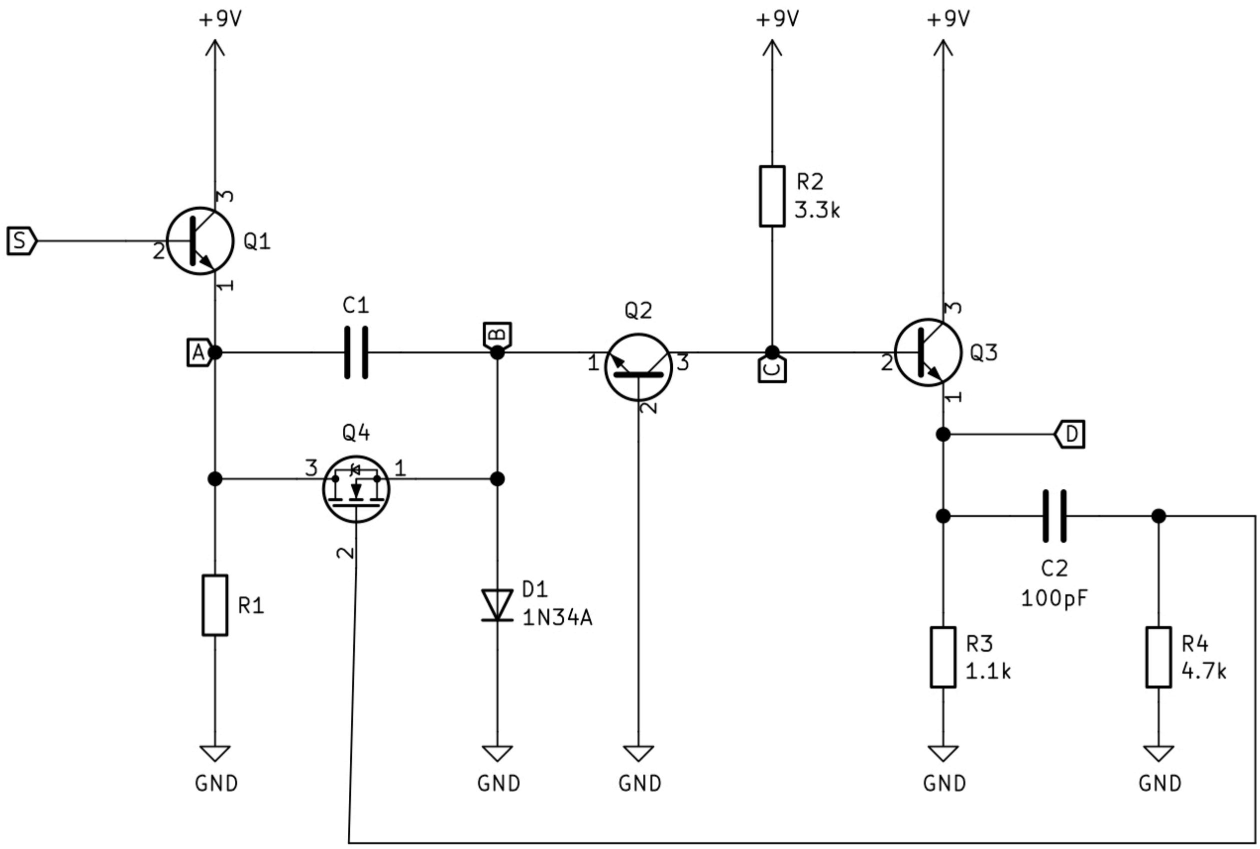
In the above plot it is visible that below a positive duty cycle of 50% we are running into problems (as predicted in chapter 10). Also, a lower intermediate frequency allows for larger values of R_1 and C_1 (with their, in general, larger pulse widths). That will give larger duty cycle swings for the same frequency deviation, and by this a higher output (audio) signal once the pulse train has been lowpass filtered.

12. Improvements

12.1 Squaring up the output pulse

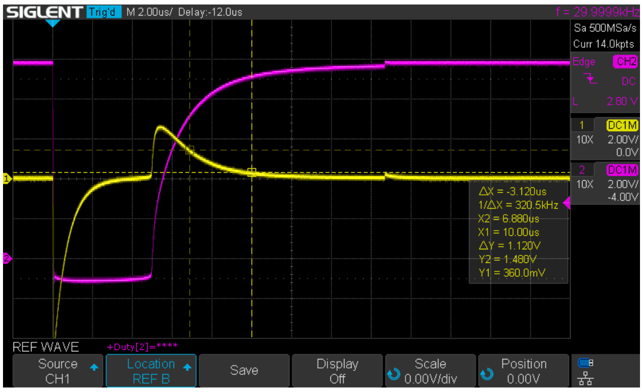
With large values for C_1 and/or R_1 the rising edge of the output pulse is not very steep. This is because at a certain moment Q_2 gets out of saturation, when C_1 has not been discharged a reasonable amount [8]. The remaining charge on C_1 will leak away through R_1. The voltage over C_1 will be distributed over R_1 and the base-emitter junction of Q_2. The part over the b-e junction of Q_2 is what decides the collector voltage of Q_2. For large values of R_1 and/or C_1, discharge will take place slowly and as such the rising edge of the pulse at the Q_2's collector will not be steep.

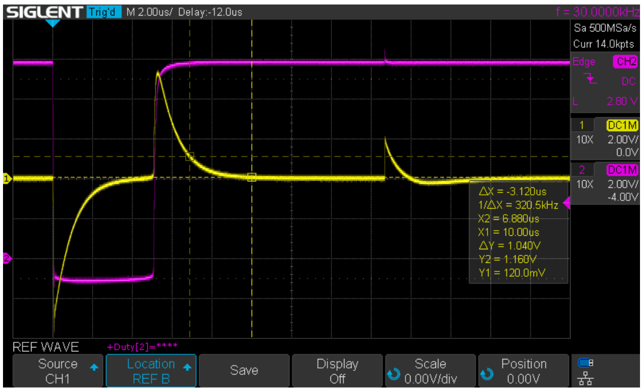
Several preliminary attempts at steepening up the output pulse have been made. The most promising at this moment seems to be to detect when Q_2 gets out of saturation, and at that moment short the terminals of C_1. The following circuit is used:



C_2, R_4 and Q_4 have been added. C_2 and R_4 act as a high pass filter to detect the rising edge of the output pulse. The positive voltage pulse created at that moment opens Q_4, which shorts C_1 and causes it to discharge at once. This discharge pulls up V_B, which stops the current through Q_2, and pulls V_C to V_{CC}.

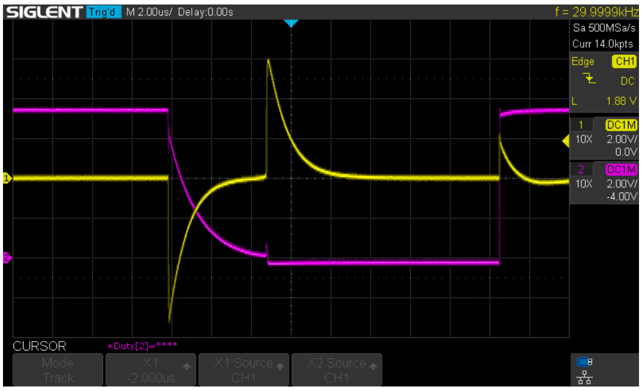
[8] This happens for any value for R_1 and C_1, only for large values it is more notable.

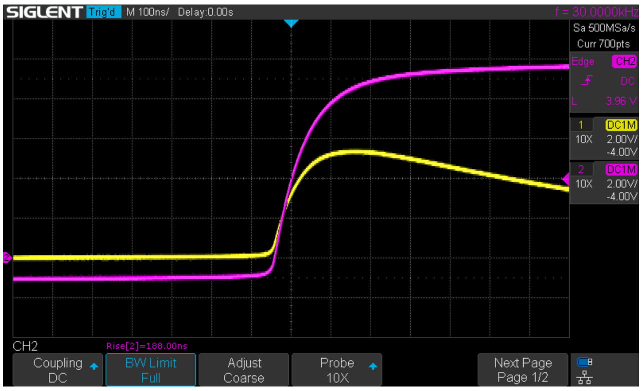




R_1=150Ω, C_1=10nF, without C_2, R_4 and disconnected Q_4. Violet = output pulse on collector Q_3, yellow = gate of Q_4.

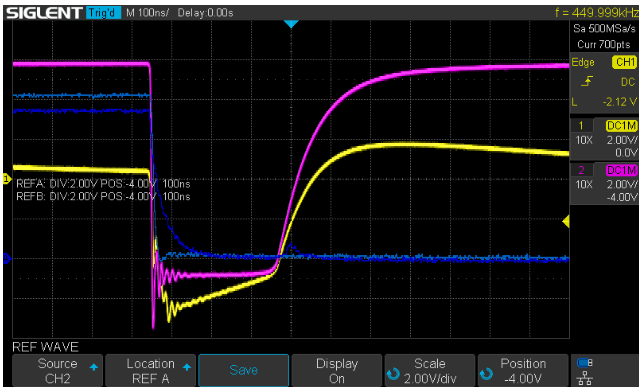
R_1=150Ω, C_1=10nF, with C_2, R_4 and Q_4. Violet = output pulse on collector Q_3, yellow = gate of Q_4.

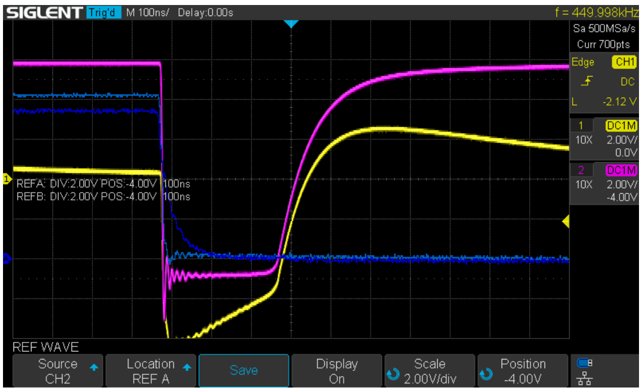




R_1=150Ω, C_1=10nF, with C_2, R_4 and Q_4. Violet = V_A, yellow = gate of Q_4. The violet trace show the sudden drop in voltage at V_A, because of the sudden discharge of C_1.

R_1=150Ω, C_1=10nF, with C_2, R_4 and Q_4. Violet = output pulse on collector Q_3, yellow = gate of Q_4. Compare with the plots below, for C_1=220pF. The rise time of the output pulse is the same.





R_1=150Ω, C_1=220pF, with C_2, R_4 and Q_4. Pink = V_C, yellow = gate of Q_4, dark blue = V_A, light blue = V_S.

R_1=150Ω, C_1=220pF, with C_2, R_4 and disconnected Q_4. Pink = V_C, yellow = gate of Q_4, dark blue = V_A, light blue = V_S.

This is with Q_4 disconnected. For C1=220pF this makes no difference with the speedup circuit installed. Compare with the plot above, for C_1=10nF. The rise time of the output pulse is the same.

Not much research has been done yet on this circuit. Several notes:

- C_2 and R_4 have been chosen as large as possible, to get phase shift as low as possible
- C_2 and R_4 should have an RC time constant times 2.2 less than half the period of the intermediate frequency. If it is larger, it may still keep Q_4 in conduction, which hampers the charging of C_1. In this case $2.2R_4C_2 = 2.2 \cdot 4700 \cdot 100e^{-12} = 1\mu s$.In the above oscilloscope plots the fall time, (90%-10% discharge time of C_2) is 2μs. This still has to be investigated. With the current 450kHz intermediate frequency of the FM receiver, with a period of 2.2μs, a RC discharge time of 1μs should suite well. The intermediate frequency used in the above oscilloscope plots is much smaller, 30kHz, for illustration purposes.
- for Q_4, a BJT has been tried instead of a MOSFET. A BJT needs a lower voltage to trigger into conduction, but it loads the output circuit too much.
- output rise times for C_1=10nF (with speedup circuit) and C_1=220pF (with or without speedup circuit) are the same. This shows that the rise time for C_1=220pF is not influenced by the speedup circuit. It also shows that this rise time is the shortest obtainable with this circuit. The limiting factor is (probably) the parasitic capacitances of Q_2 and Q_3.
- ideally we want C_2 to be discharged completely before Q_2 gets out of saturation to get the maximum upswing on the gate of Q_4.

12.2 Lowering the intermediate frequency

Lowering the intermediate frequency allows larger values for R_1 and C_1 (see 11.1, Linearity). This in turn will create a larger duty cycle variation (see 11.1, Linearity). A larger variation in duty cycle will create a larger output signal of the pulse train detector once the pulse train has been lowpass filtered. When using larger values for R_1 and C_1, the rising edge of the output pulse will get less steep. This can be solved by techniques as described in the previous paragraph 12.1.

The following plots show an example.

R_1=1kΩ, C_1=1nF, C2=100pF, R4=4.7kΩ. Yellow trace shows V_A, pink trace shows V_D. The yellow trace clearly shows the discharge of C_1 once Q_2 gets out of saturation and opens Q_4. This plot is taken from an actual FM modulated input signal of a radio station.

These plots are for larger values of R_1 and C_1 than in the preceding examples. Although the values are large, due to the speedup circuit a steep output pulse is generated with a rise time of only 84ns. The output pulse is about 1.6μs. The large value for R_1 implies less current in Q_1 and R_1, with less power dissipation. In this setup the intermediate frequency has been lowered from 450kHz to 250kHz. Thus duty cycle variation is much larger now, with a consequently larger audio signal after the pulse train has been lowpass filtered.