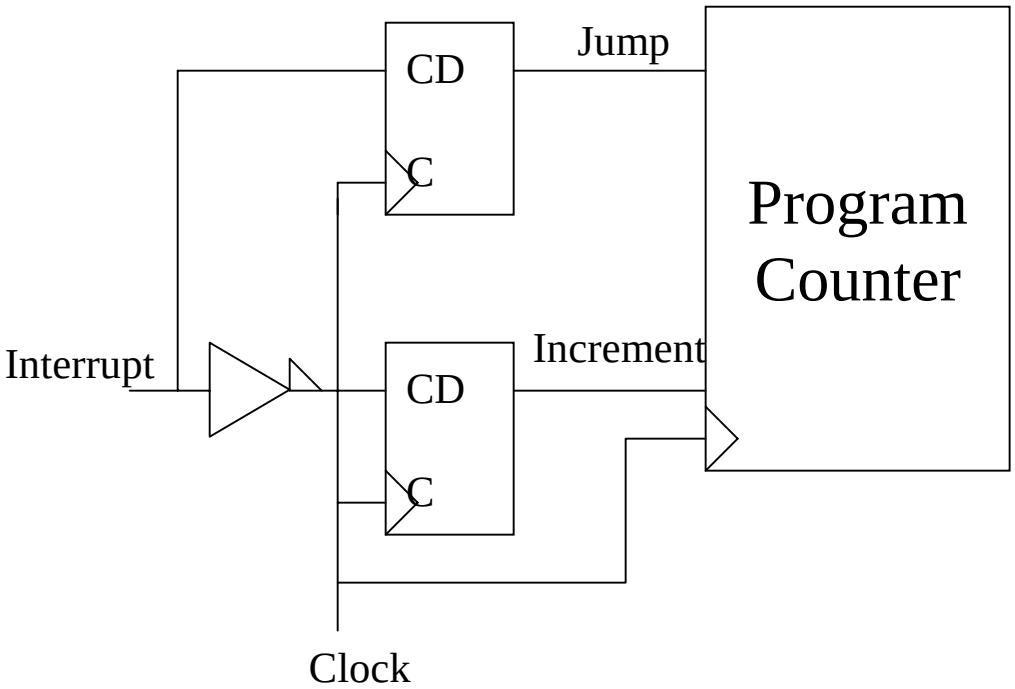
Metastability

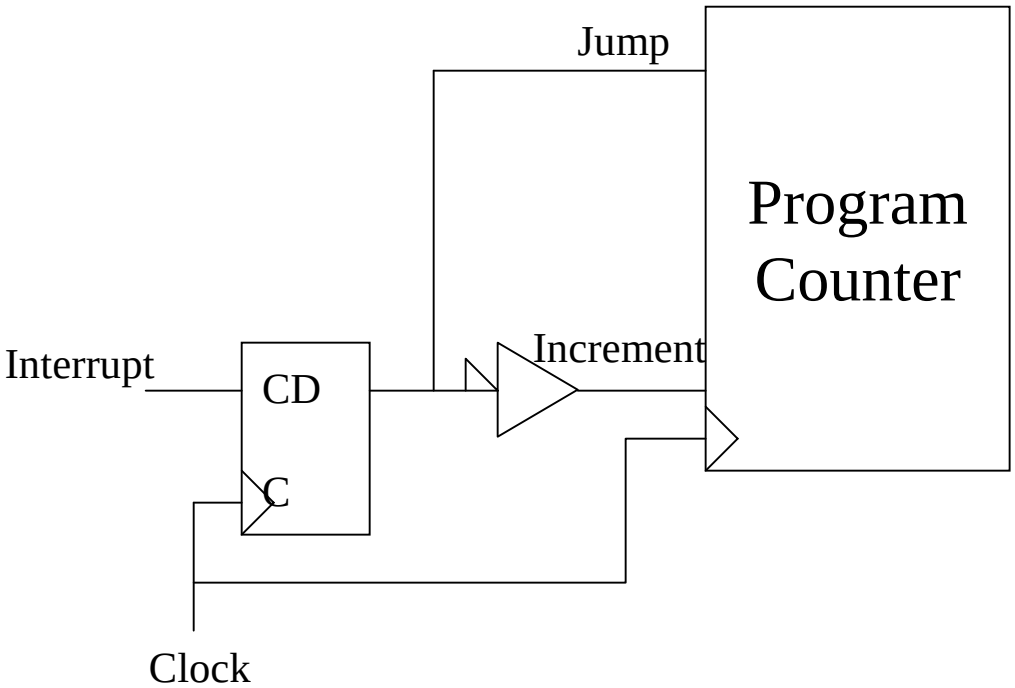
EE463
F. U. Rosenberger
Washington University
Electrical Engineering

What Problems Lurk; 1?



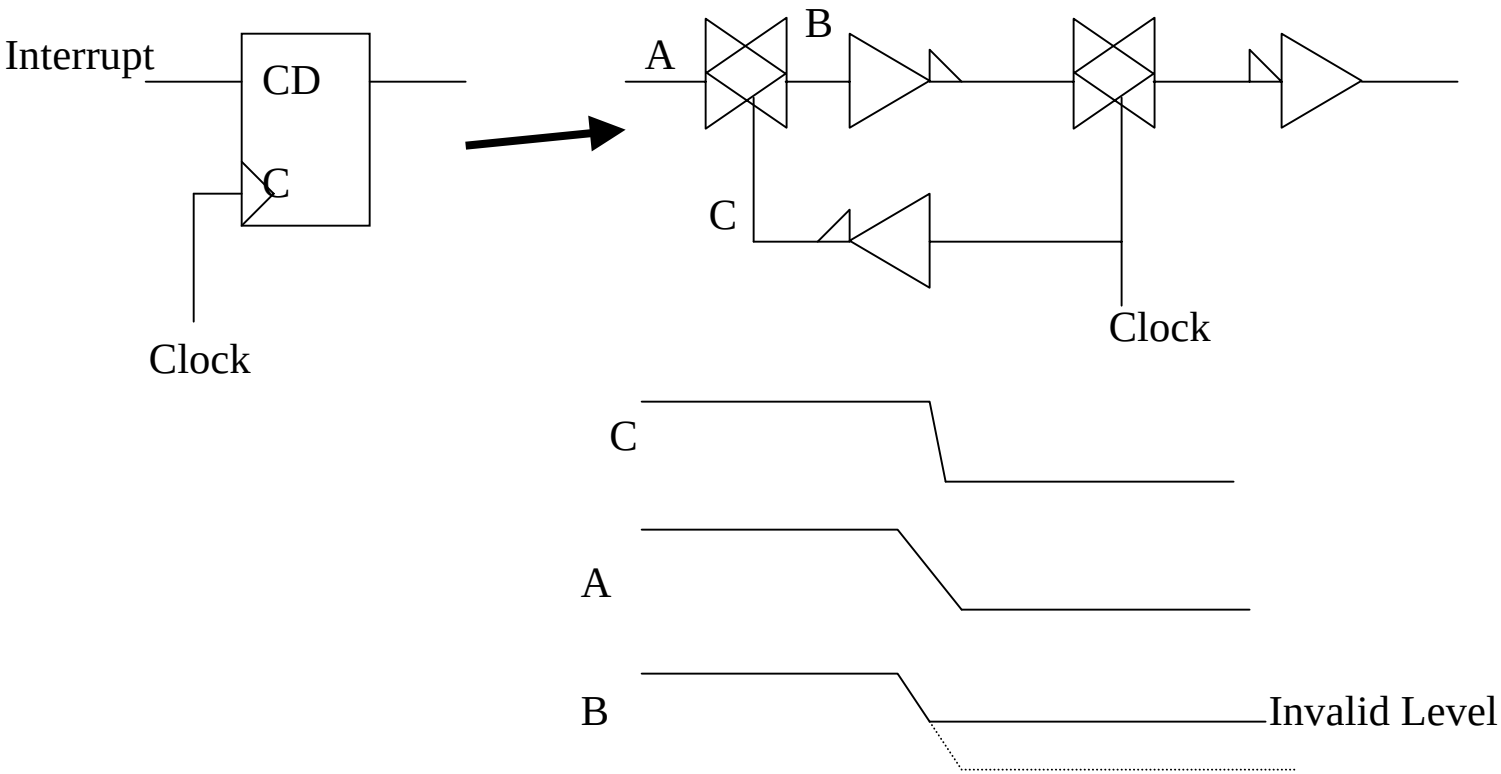
- Simple case of clocking, normal operation is to increment, jump to interrupt address on interrupt.
- Any problems here? (Ignore resetting and preparation for next interrupt, all this can be grafted onto the given circuit and does not affect the point to be made here)

What Problems Lurk; 2?



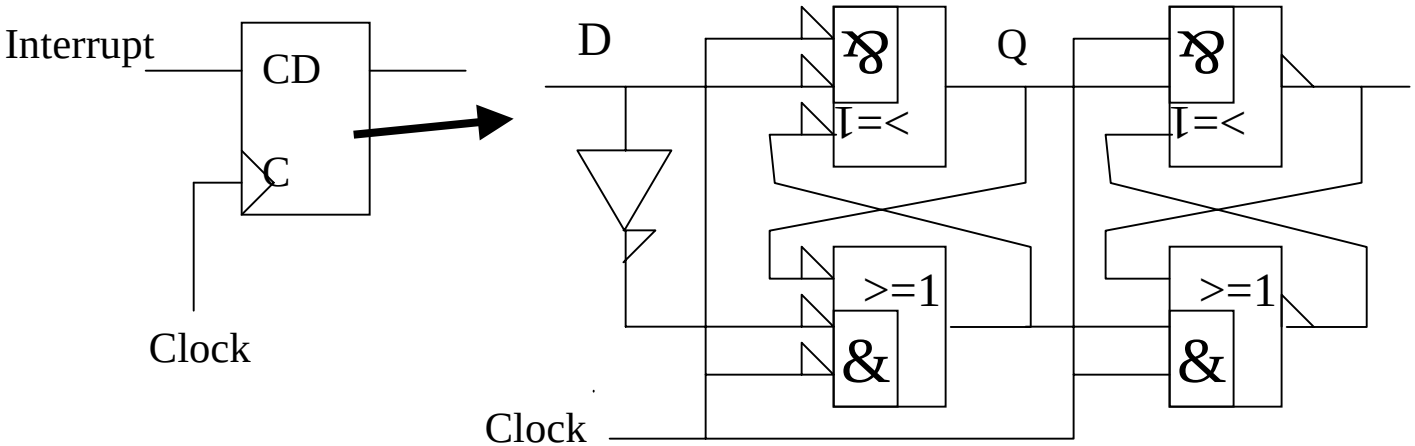
- Much better, only two states, both ok (previous circuit had 4 states, was a terrible implementation with disastrous behavior)
- Is this good enough?
- If not, why not?

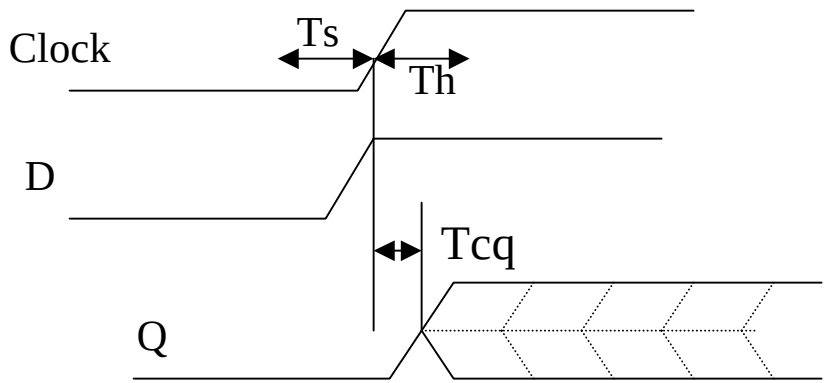
What Problems Lurk; 3?



- How is flop implemented, two dynamic latches?
- Cannot guarantee to have good level with dynamic latch and unknown data arrival time
- What now?

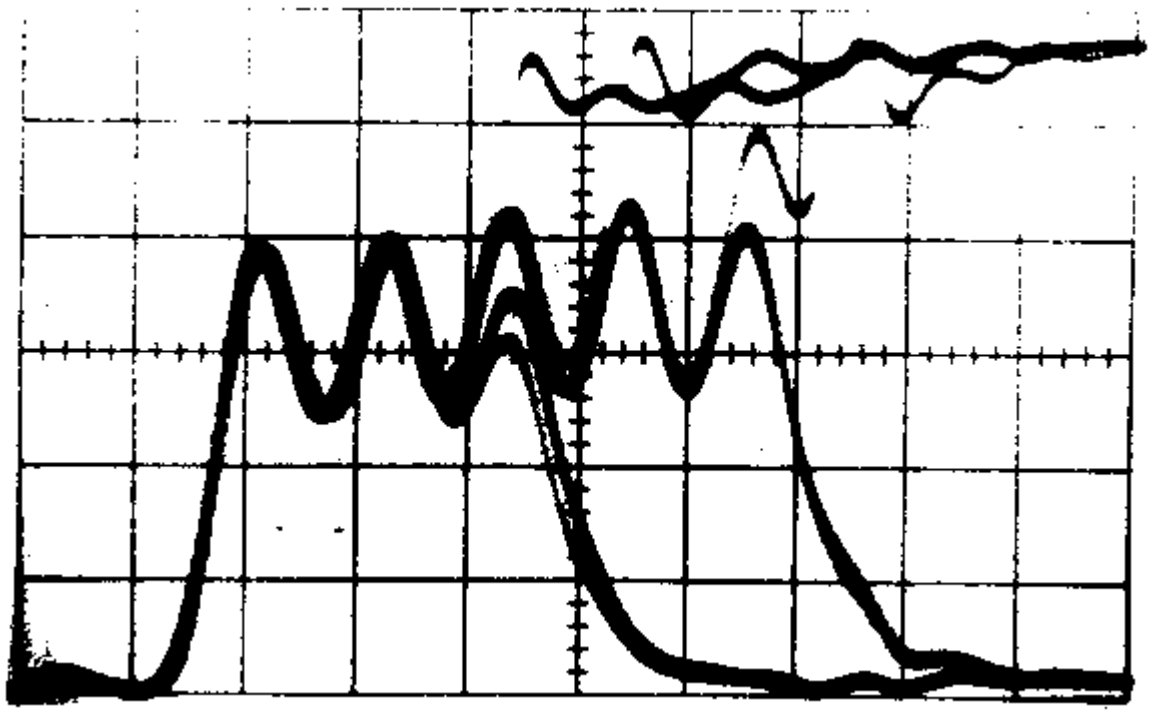
What Problems Lurk; 4?





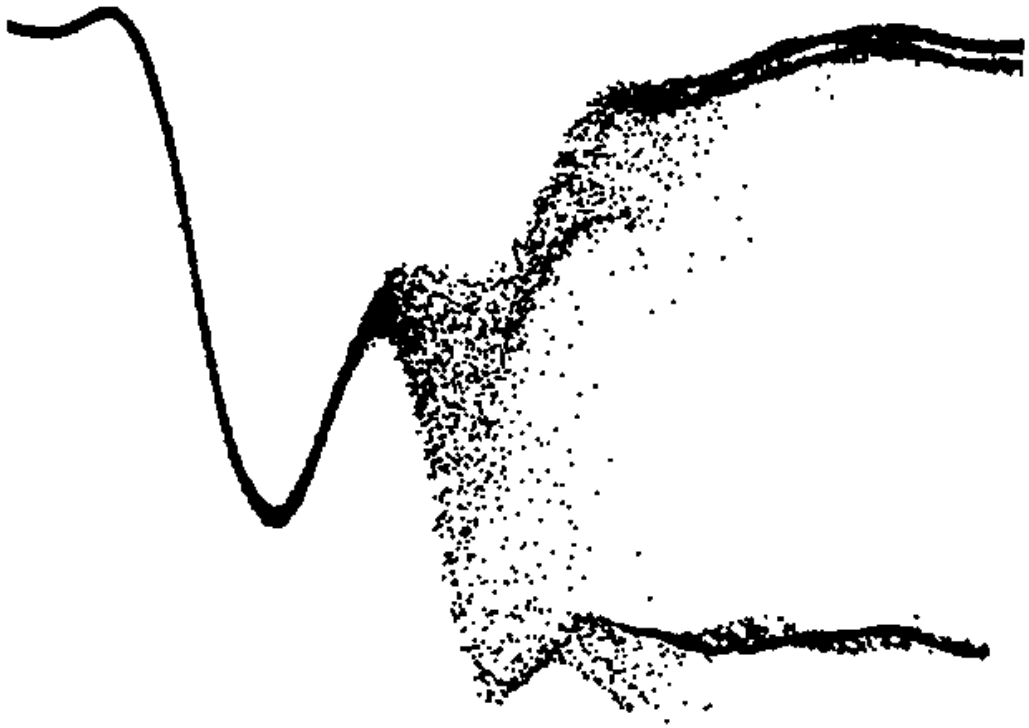
- Almost OK
- What if clock and data change at (almost) same time?
  - Adjust clock to data time until result is ambiguous
- Do real flip-flops behave this way????

Real-Time Scope Photo



- Single event traces from real-time scope. Scope was triggered on delayed transition, and signal was delayed by coax delay line so it could be viewed after determining whether to trigger or not.
- Photo by Thomas J. Chaney, Washington U.

Sampling Scope Photo



- Each dot represents one sample from one event. Dot density represents probability of events (traces) that pass through that region
- Photograph by Thomas J. Chaney, Washington U.

Fundamental

- This is fundamental problem
- Can't guarantee *correct* operation with arbitrary clock and data phase
- That is, given two events with unknown relative time, its impossible to *unambiguously* make a decision as to which arrives first, even if either decision is acceptable
- Examples of two asynchronous events, outcome does not matter as long as a single clean decision is made
  - Modular IMP (ARPANET/Internet)
  - NIH Dual-Processor with shared memory
  - Any mysterious computer crash (although its *usually* software)
  - Yellow traffic light
  - Two people meet in hallway
  - Dog midway between two food dishes may starve
  - Computer system
  - Wristwatch
  - Design decision (include feature? Simple vs complex? ...

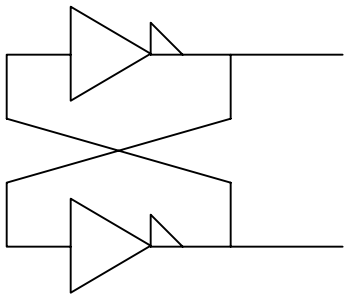
Fundamental (Continued)

- What to do?
  - Calculate failure rate
  - Engineer adequate reliability (mean time to failure)
    - Improved flip-flop, surrounding circuitry, or environment
  - Avoid “aggravated glitch”, two flops with intended same state
  - Understand and be able to identify trouble spots in design
  - Don’t look for “solution” there is none. You need faith
  - Remember worst-case design (or worst case probability)
  - Don’t believe everything you read, especially:
    - Manufacturer data sheets, “Metastable hardened”, … Nonsense
    - *An Engineering Approach to Digital Design*, William I. Fletcher, Prentice-Hall, 1980. Page 484 purports to give a solution to the metastability problem; Wrong.
    - E. G. Wormald, “A Note on Synchronizer or Interlock Maloperation,” *IEEE Trans. on Computers*, C-26, No. 3. Pp. 317-318, Mar. 1977.
    - “Metastability of CMOS Latch/Flip-Flop,” Lee-Sup Kim, Robert W. Dutton, *Journal of Solid-State Circuits*, V 25, No. 4, pp. 942-951, August 1990.
    - F. J. Hill, G. R. Peterson, *Introduction to Switching Theory and Logical Design*, Third Ed., John Wiley and Sons, 1981. Pages 509-512 have an incorrect circuit.
    - Others
    - Synopsys Designware Variable Input Synchronizer: DW04_sync

Why is Metastability a “Special” Problem, Charles E. Molnar

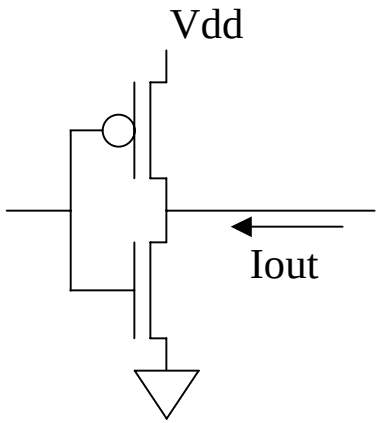
- Because *it* “breaks most of the conceptual and computational tools that we use from day to day (e.g. binary or two-state circuits)
- *It* defies careful and accurate measurements
- *It* can produce failures that leave no discernable evidence
- *It* can cause failures in systems whose software is “correct” and whose hardware passes all conventional tests
- *It* involves magnitudes of time and voltage far removed from our daily experience

Latch in Metastable Region of Operation



- We have removed the components for setting/resetting the latch
- Both outputs (and thus inputs) at approximately the same voltage
- It is just two inverters
- We need to analyze CMOS inverter operation in this region

Inverter in Linear Region (Vo ≅ Vin)



- Both transistors are in Saturation
- Ignore channel length modulation (output R)

$$I_{out} = I_{DSN} + I_{DSP}$$

$$= K_N \cdot \frac{W_N}{L_N} \bullet \frac{1}{2}(V_{IN} - V_{TN})^2 - K_P \cdot \frac{W_P}{L_P} \bullet \frac{1}{2}(V_{DD} - V_{IN} + V_{TP})^2$$

$$g_m = \frac{\partial I_{out}}{\partial V_{IN}} = K_N \cdot \frac{W_N}{L_N} \bullet (V_{IN} - V_{TN}) + K_P \cdot \frac{W_P}{L_P} \bullet (V_{DD} - V_{IN} + V_{TP})$$

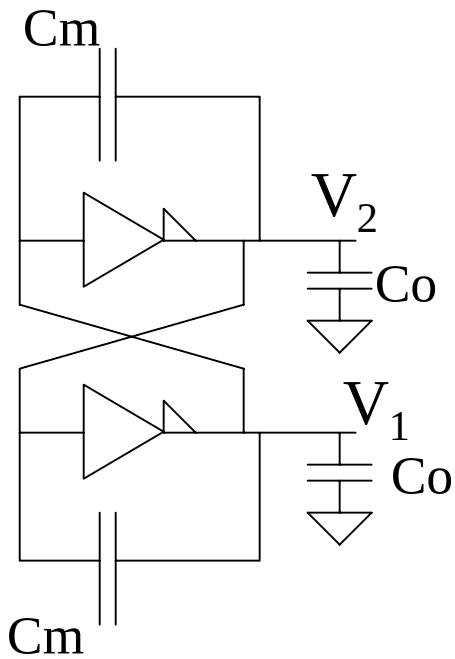
Inverter in Linear Region (Vo ≅ Vin), Continued

- Vinv is voltage for which Vin and Vout are equal
- From previous analysis of inverter

$$V_{inv} = \frac{V_{dd} + V_{TP} + V_{TN} \cdot \sqrt{\beta_N / \beta_{P1}}}{1 + \sqrt{\beta_N / \beta_{P1}}}$$

- Substitute this into equation for gm, and one can find that gm is maximum for Wn=Wp. Ref: "Synchronization Reliability in CMOS Technology", Stephen T. Flannagan, *JSSC*, SC-20, No. 4, pp. 880-882, August 1985.
- Analysis is much better than simulation for finding g_m, faster, more accurate, easy to carry out.

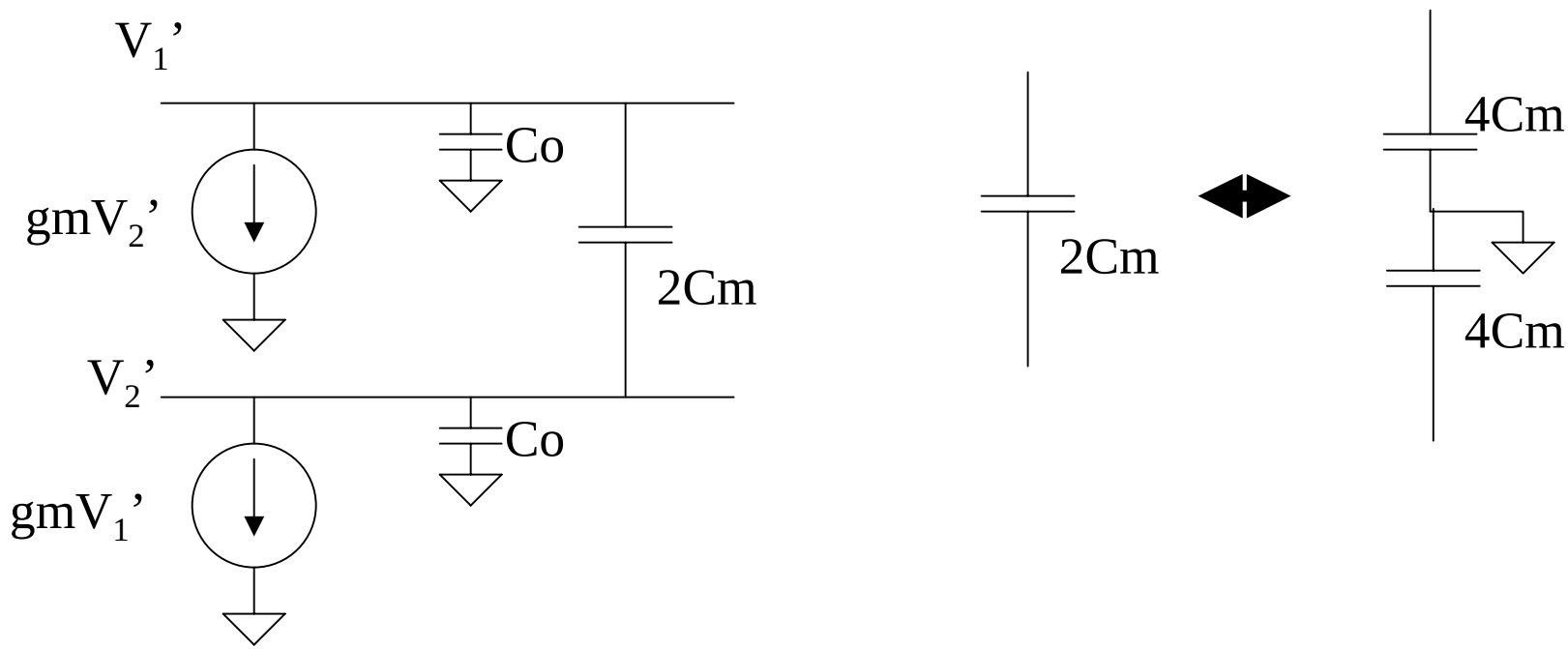
Symmetric Case, Simple Analysis



- Inverters with capacitance
- Cm is capacitance from inverter input to output, primarily FET gate to drain. Co is capacitance from input and output to ground

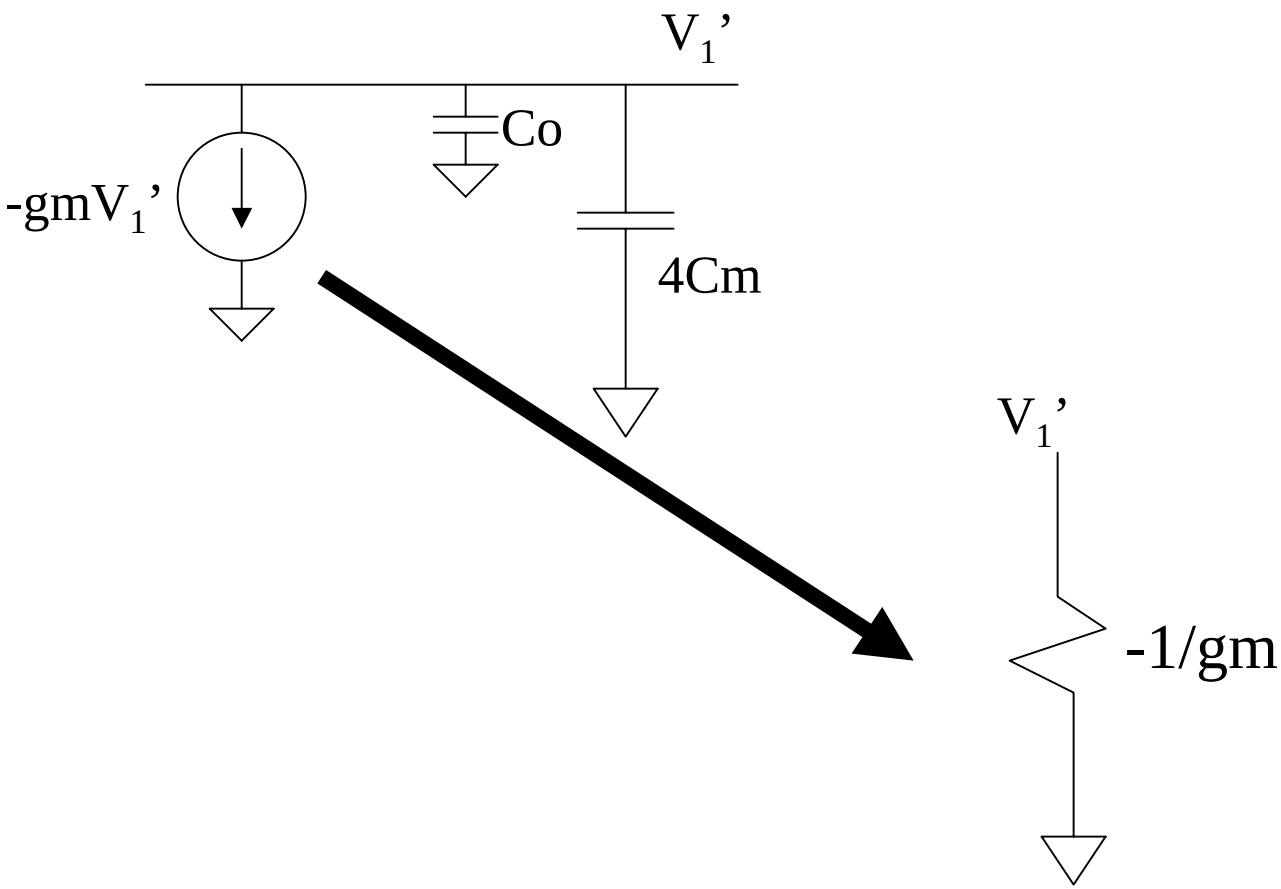
Linear Equivalent Circuit

- Small signal equivalent circuit about Vinv
- Let V_1'=V_1 - Vinv; V_2'=V_2 - Vinv



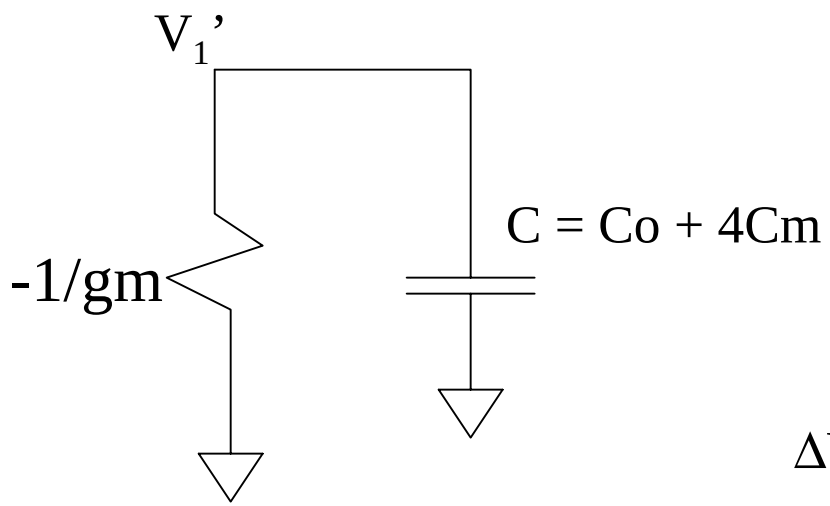
Use Symmetry

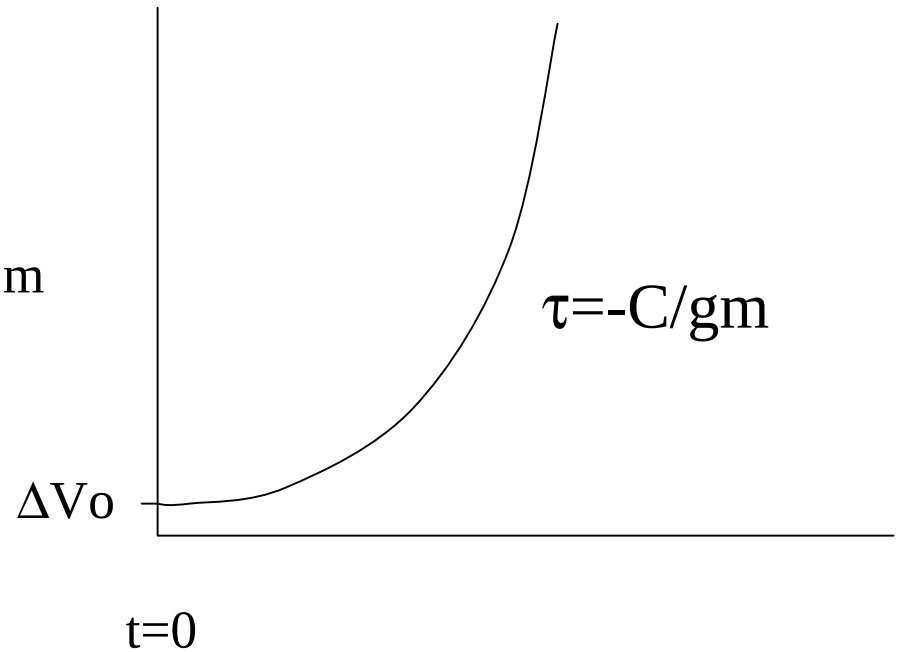
- For symmetric case, let V1’ = -V2’ = ΔVo at t=0



-RC Circuit

- Initial offset voltage grows exponentially





$$V_1' = \Delta V_0 e^{t/\tau}$$

$$\tau = C / gm$$

Example Values

- gm = 200uA/V
- C = 0.05pF
- Then τ=0.25ns
- ΔVo of 1pV grows to 1V in 7ns

- We want large gm, small C
- gm is maximum when $W_N = W_P$
- If we wait long enough, arbitrarily small initial voltage grows to significant value
- How can we use this to predict behavior?
  Need a little more analysis first!
  What is distribution of initial voltage?

Sum and Difference Voltages

- More completely, we can deal with the sum and difference voltages in the linear region, V_1 and V_2 can be calculated from these sum and difference voltages.

$$V_D = V_1 - V_2 \qquad V_\Sigma = V_1 + V_2 - 2V_{inv}$$

$$V_D = V_{D_0} e^{t/\tau_D} \qquad V_\Sigma = V_{\Sigma_0} e^{-t/\tau_\Sigma}$$

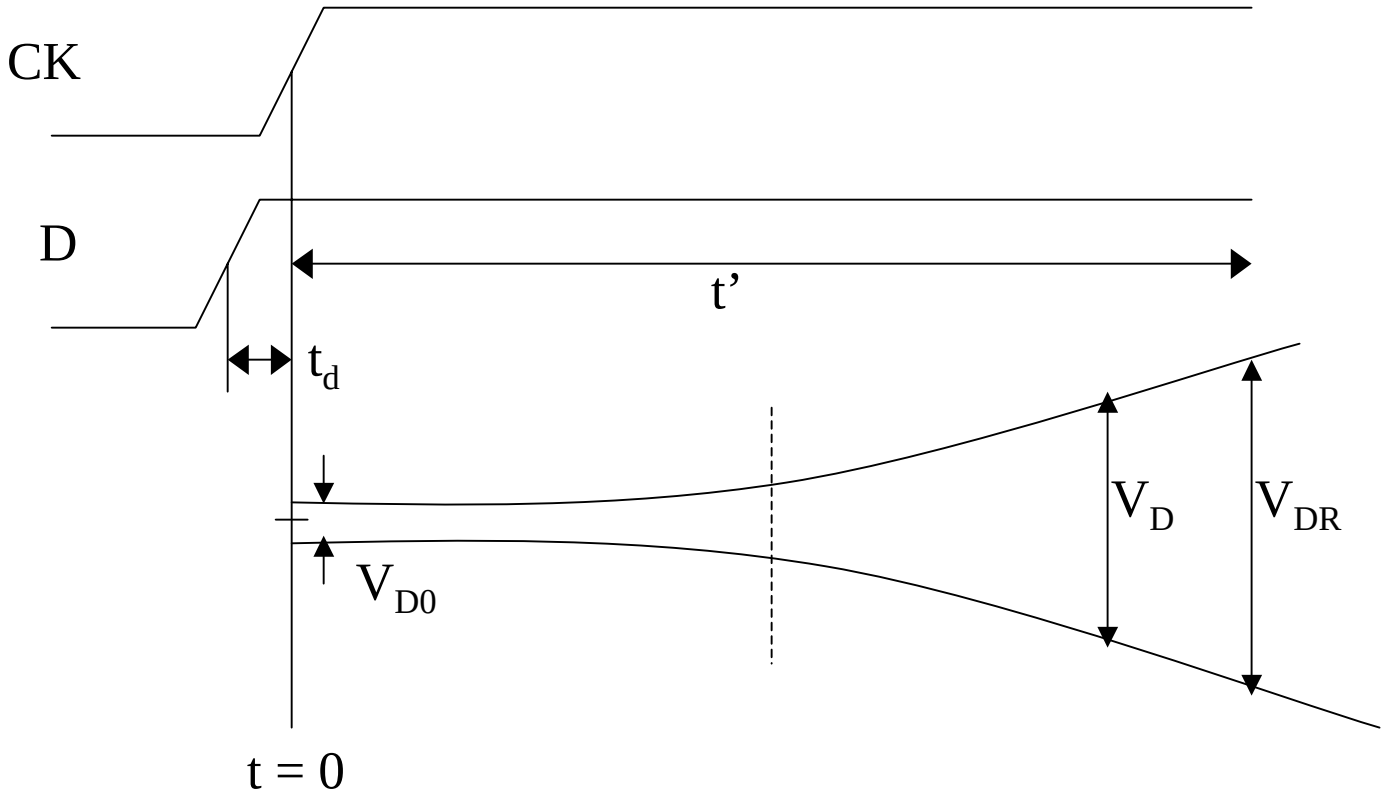
$$\tau_D = \frac{C_0 + 4C_m}{g_m} \qquad \tau_\Sigma = \frac{C_0}{g_m}$$

$$V_{D_0} = V_D @ t = 0 \qquad V_{\Sigma_0} = V_\Sigma @ t = 0$$

- Note that the exponent for the sum voltage has a negative sign, it decays with time. The difference voltage increases with time
- We can usually ignore the sum term, it has a smaller time constant, and goes to zero

Time-to-Voltage Gain

- The difference voltage at t = 0 is amplified
- If we can calculate the difference voltage at t = 0, or its equivalent, then we can calculate the difference voltage as a function of time.
- Consider what happens when the clock is deasserted as the data is changing: Initial offset voltage is dependent on data transition time with respect to time that gives perfectly balanced metastability
- G_{tv} is the time to voltage gain, change in initial voltage with respect to clock to data time. It is the efficiency in converting a small change in input time to a change in initial voltage, larger is better.



$$t_{d\infty} = t_d\big|_{V_{D0}=0}$$

$$t_d' = t_d - t_{d\infty} \qquad G_{tv} = dV_{D0}/dt$$

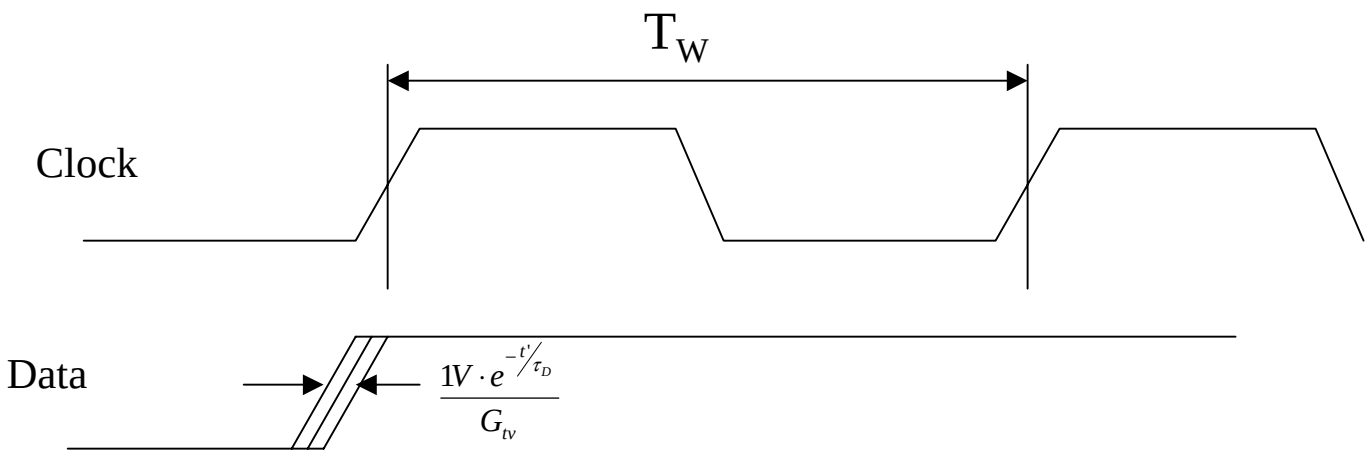
$$F(t') = \text{Probability flop unresolved at time } t'$$

$$F(t') = \text{Prob}\left\{-\frac{1}{2}V_{DR} < V_D(t') < \frac{1}{2}V\right\}$$

$$F(t') = \text{Prob}\left\{-\frac{1}{2}V_{DR} < V_{D0}e^{t'/\tau_D} < \frac{1}{2}V\right\}$$

$$F(t') = \text{Prob}\left\{-\frac{1}{2}V_{DR}e^{-t'/\tau_D} < t_d'G_{tv} < \frac{1}{2}V_{DR}e^{-t'/\tau_D}\right\}$$

$$F(t') = \text{Prob}\left\{-\frac{1}{2G_{tv}}V_{DR}e^{-t'/\tau_D} < t_d' < \frac{1}{2G_{tv}}V_{DR}e^{-t'/\tau_D}\right\}$$



- Output is metastable at time t’ if data transition time is within region shown above
- Probability of this is width of region, divided by T_w

$$F(t') = \frac{1V \cdot e^{-t'/\tau_D}}{G_{tv}} \cdot \frac{1}{T_W}$$

$$MTBF(Unresolved) = \frac{1}{Data - Event - Rate \cdot F(t')}$$

$$= \frac{G_{tv} \cdot T_W \cdot e^{t'/\tau_D}}{f_D}$$

$$= \frac{G_{tv} \cdot e^{t'/\tau_D}}{f_D \cdot f_{CK}}$$

- f_D is data rate, f_{CK} is clock rate.
- How to calculate G_{tv}?
  - It can be *approximated* by dv/dt for the flip-flop voltages.
  - Voltage is changing at this rate when clock is deasserted, capturing present value
  - Actual operation is much more complicated, flop is operating in non-linear region during interval about the clock transition, actual G_{tv} is much worse
  - Exact value not too critical, its τ that really matters

Using SPICE (simulation) to find tau and G_{tv}

- Simulate flip-flop with data and clock changing, vary clock to data time to get metastable operation (tedious process).
  - τ can be easily determined from the waveforms as metastability is resolved, difference voltage increases exponentially with time constant τ
  - G_{tv} can be determined by plotting voltage offset at t=0, vs clock to data offset. Offset voltage is found from time output crosses threshold times $e^{-t/\tau}$
  - τ is much easier to find from analysis, SPICE serves as a check.
  - G_{tv} is very difficult to find from analysis, SPICE is helpful here.
- Because of the very small voltage and time differences of interest, SPICE may give anomolous results. In fact, output vs input may not be monotonic. This is not a problem with tau, but results for G_{tv} must be examined with care to insure valid interpretation.
  - Plot over wide range to check for non-monotonic behavior

Design for Good Resolving Time

$$\tau_D = \frac{C}{g_m}$$

- Minimize C
  - Buffer output with low-Cin inverter
  - Minimize internal C, maximize gm (folded FET)
- For an amplifier, Gain-Bandwidth = g_m/C
  - Gain-bandwidth is figure of merit for transistor or amplifying device
- If we include output resistance (finite gain),

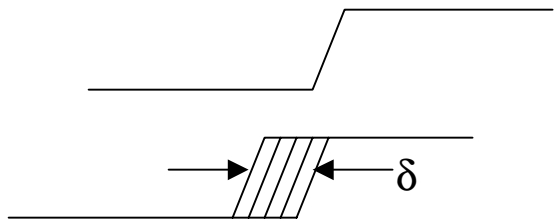
$$\tau_D = \frac{1}{(G-1)\cdot BW} = \frac{1}{(g_m \cdot R - 1)\,1/RC} = \frac{RC}{g_m \cdot R - 1} \approx \frac{C}{g_m}$$

- This is gain-bandwidth for a single stage, not multiple stages
- Two stages with gain of 10 and bandwidth of 10^9 are not equivalent to one stage of gain 100 and bandwidth of 0.707×10^9
- It is unity-gain bandwidth, not gain-bandwidth that matters

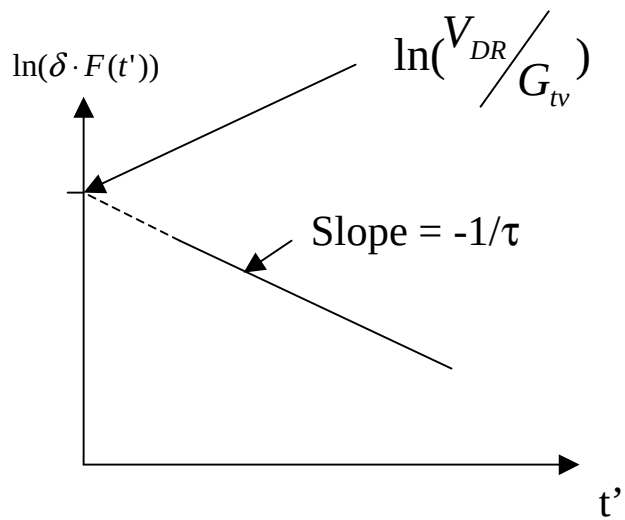
Measuring τ and G_{tv}

- Estimate F(t') for several values of t' by counting unresolved events at t' and dividing by total number of trials
  - Extreme care is required in experimental setup
  - Trials must be independent (previous state, etc)
  - Fine control of timing is required

$$F(t') = \frac{V_{DR} \cdot e^{-t'/\tau_D}}{G_{tv}} \cdot \frac{1}{\delta}$$



$$\ln(\delta \cdot F(t')) = -\frac{1}{\tau} t' + \ln(V_{DR}/G_{tv})$$



Examples

Alternate Analysis/View

- Phase-plane plots, dynamical-systems theory
- Plot V vs dV/dt

Conclusion

- Be careful!!!
- Know what you are doing!!!
- Identify circuits where metastability may be a problem
- Only one Flop receives asynchronous signal
- Allow adequate settling time for desired reliability
  - Don’t forget worst case, can be much much worse
- Don’t believe everything (or perhaps anything) you read, especially data sheets
- Text uses To rather than Gtv (To = 1/Gtv)
- Asynchronous circuits may “overcome” metastability by waiting until it is resolved, and detecting this, before using flop output
  - Wait only what is required, don’t wait max amount every time
  - Finite (but very small) probability of waiting so long result is no longer useful, thus not complete theoretical solution (although it is practical)
- Scaling, what happens as feature size becomes smaller?