6.012 - Microelectronic Devices and Circuits
Lecture 16 - CMOS scaling; CCDs - Outline

• Announcements
  PS #9 - Will be due Friday next week; no recitation tomorrow.
  Postings - CMOS scaling (multiple items)
  Exam Two – Tomorrow night, April 6, 7:30-9:30 pm, 52-141

• Review - CMOS gate delay and power
  Lecture 15 results:  Gate Delay = $12 \, n \, L_{\text{min}}^2 \frac{V_{\text{DD}}}{\mu_n (V_{\text{DD}} - V_T)^2} \propto \tau_{\text{tr}}$
  $P_{\text{dyn}@f_{\text{max}}} \propto C_L \frac{V_{\text{DD}}^2}{G_D} = K_n V_{\text{DD}} (V_{\text{DD}} - V_T)^2 / 4$

  Velocity Saturation

• CMOS scaling rules
  Power density issues and challenges
  Approaches to a solution:  Dimension scaling alone
  Scaling voltages as well

• Charge Coupled Devices (CCDs)
  Inversion layer build-up with no adjacent n+ region
  Moving charge along a chain of adjacent gates
**CMOS: transfer characteristic**

Complete characteristic w.o. Early effect:

\[ V_{DD} \]

\[ V_{IN} \]

\[ V_{OUT} \]

\[ V_{Tp} \]

\[ V_{Tn} \]

**NOTE:** We design CMOS inverters to have \( K_n = K_p \) and \( V_{Tn} = -V_{Tp} \) to obtain the optimum symmetrical characteristic.
We found from an LEC analysis that the slope in Region III is not infinite, but is instead:

\[
A_v \equiv \frac{v_{out}}{v_{in}} = \frac{\partial v_{OUT}}{\partial v_{IN}} \bigg|_{Q(= V_{DD}/2, V_{DD}/2)} = -\frac{2\sqrt{2K_n}}{[g_{on} + g_{op}]}
\]

Quick approximation: An easy way to sketch the transfer characteristic of a CMOS gate is to simply draw the three straight line portions in Regions I, III, and V:
**CMOS**: switching speed; minimum cycle time

**The load capacitance: \( C_L \)**

- Assume to be linear
- Is proportional to MOSFET gate area
- In channel: \( \mu_e = 2\mu_h \) so to have \( K_n = K_p \) we must have \( W_p/L_p = 2W_n/L_n \)
  
  Typically \( L_n = L_p = L_{\text{min}} \) and \( W_n = W_{\text{min}} \), so we also have \( W_p = 2W_{\text{min}} \)

\[
C_L \approx n \left[ W_n L_n + W_p L_p \right] C_{ox}^* = n \left[ W_{\text{min}} L_{\text{min}} + 2W_{\text{min}} L_{\text{min}} \right] C_{ox}^* = 3nW_{\text{min}} L_{\text{min}} C_{ox}^*
\]

**Charging cycle**: \( v_{\text{IN}} \): HI to LO; \( Q_n \) off, \( Q_p \) on; \( v_{\text{OUT}} \): LO to HI

- Assume charged by constant \( i_{D,\text{sat}} \)

\[
i_{\text{charge}} = -i_{Dp} \approx \frac{K_p}{2} \left[ V_{DD} - |V_{Tp}| \right]^2 = \frac{K_n}{2} \left[ V_{DD} - V_{Tn} \right]^2
\]

\[
q_{\text{charge}} = C_L V_{DD}
\]

\[
\tau_{\text{charge}} = \frac{q_{\text{charge}}}{i_{\text{charge}}} = \frac{2C_L V_{DD}}{K_n \left[ V_{DD} - V_{Tn} \right]^2} = \frac{6nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}}{W_{\text{min}} \mu_e C_{ox} \left[ V_{DD} - V_{Tn} \right]} = \frac{6nL_{\text{min}}^2 V_{DD}}{\mu_e \left[ V_{DD} - V_{Tn} \right]^2}
\]

Clif Fonstad, 4/5/11
**CMOS:** switching speed; minimum cycle time, cont.

**Discharging cycle:** $v_{\text{IN}}$: LO to HI; $Q_n$ on, $Q_p$ off; $v_{\text{OUT}}$: HI to LO

- Assume discharged by constant $i_{\text{D,sat}}$

$$ i_{\text{Disch.arg.e}} = i_{Dn} \approx \frac{K_n}{2} [V_{DD} - V_{Tn}]^2 $$

$$ q_{\text{Disch.arg.e}} = C_L V_{DD} $$

$$ \tau_{\text{Disch.arg.e}} = \frac{q_{\text{Disch.arg.e}}}{i_{\text{Disch.arg.e}}} = \frac{2C_L V_{DD}}{K_n [V_{DD} - V_{Tn}]^2} $$

$$ = \frac{6nW_{\text{min}}L_{\text{min}}C_{ox}^* V_{DD}}{W_{\text{min}} \mu e C_{ox}^* [V_{DD} - V_{Tn}]^2} = \frac{6nL_{\text{min}}^2 V_{DD}}{\mu e [V_{DD} - V_{Tn}]^2} $$

**Minimum cycle time:** $v_{\text{IN}}$: LO to HI to LO; $v_{\text{OUT}}$: HI to LO to HI

$$ \tau_{\text{Min.Cycle}} = \tau_{\text{Ch.arg.e}} + \tau_{\text{Disch.arg.e}} = \frac{12nL_{\text{min}}^2 V_{DD}}{\mu e [V_{DD} - V_{Tn}]^2} $$
**CMOS:** switching speed; minimum cycle time, cont.

**Discharging and Charging times:**

What do the expressions tell us? We have

\[ \tau_{\text{Min Cycle}} = \frac{12nL_{\text{min}}^2V_{DD}}{\mu_e[V_{DD} - V_{Tn}]^2} \]

This can be written as:

\[ \tau_{\text{Min Cycle}} = \frac{12nV_{DD}}{(V_{DD} - V_{Tn})} \cdot \frac{L_{\text{min}}}{\mu_e(V_{DD} - V_{Tn})/L_{\text{min}}} \]

The last term is the channel transit time:

\[ \tau_{\text{Ch Transit}} = \frac{L_{\text{min}}}{s_{e,Ch}} = \frac{L_{\text{min}}}{\mu_e E_{Ch}} \propto \frac{L_{\text{min}}}{\mu_e(V_{DD} - V_{Tn})/L_{\text{min}}} \]

Thus the gate delay is a multiple of the channel transit time:

\[ \tau_{\text{Min Cycle}} = \frac{12nV_{DD}}{(V_{DD} - V_{Tn})} \tau_{\text{Channel Transit}} = n' \tau_{\text{Channel Transit}} \]
**CMOS:** power dissipation - total and per unit area

**Average power dissipation**

Only dynamic for now

\[ P_{\text{dyn,ave}} = E_{\text{Dissipated per cycle}} \cdot f = C_L V_{DD}^2 \cdot f = 3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2 \cdot f \]

**Power at maximum data rate**

Maximum \( f \) will be \( 1/\tau_{\text{Gate Delay Min.}} \)

\[ P_{\text{dyn}@ f_{\text{max}}} = \frac{3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2}{\tau_{\text{Min.Cycle}}} = 3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2 \cdot \frac{\mu_e [V_{DD} - V_{Tn}]^2}{12n L_{\text{min}}^2 V_{DD}} \]

\[ = \frac{1}{4} \frac{W_{\text{min}}}{L_{\text{min}}} \mu_e C_{ox}^* V_{DD} [V_{DD} - V_{Tn}]^2 \]

**Power density at maximum data rate**

Assume that the area per inverter is proportional to \( W_{\text{min}} L_{\text{min}} \)

\[ PD_{\text{dyn}@ f_{\text{max}}} = \frac{P_{\text{dyn}@ f_{\text{max}}}}{\text{InverterArea}} \propto \frac{P_{\text{dyn}@ f_{\text{max}}}}{W_{\text{min}} L_{\text{min}}} = \frac{\mu_e C_{ox}^* V_{DD} [V_{DD} - V_{Tn}]^2}{L_{\text{min}}^2} \]
**CMOS:** design for high speed

**Maximum data rate**

Proportional to $1/\tau_{\text{Min Cycle}}$

$$\tau_{\text{Min Cycle}} = \tau_{\text{Ch arg e}} + \tau_{\text{Dis ch arg e}} = \frac{12nL_{\text{min}}^2V_{DD}}{\mu_e[V_{DD} - V_{Tn}]^2}$$

Implies we should reduce $L_{\text{min}}$ and increase $V_{DD}$.

**Note:** As we reduce $L_{\text{min}}$ we must also reduce $t_{ox}$, but $t_{ox}$ doesn't enter directly in $f_{\text{max}}$ so it doesn't impact us here.

**Power density at maximum data rate**

Assume that the area per inverter is proportional to $W_{\text{min}}L_{\text{min}}$

$$PD_{\text{dyn @ f_{max}}} \propto \frac{P_{\text{dyn@f_{max}}}}{W_{\text{min}}L_{\text{min}}} = \frac{\mu_e\varepsilon_{ox}V_{DD}[V_{DD} - V_{Tn}]^2}{t_{ox}L_{\text{min}}^2}$$

Shows us that PD increases very quickly as we reduce $L_{\text{min}}$ unless we also reduce $V_{DD}$ (which will also reduce $f_{\text{max}}$).

**Note:** Now $t_{ox}$ does appear in the expression, so the rate of increase with decreasing $L_{\text{min}}$ is even greater because $t_{ox}$ must be reduced along with $L$ to stay in the gradual channel regime.

How do we make $f_{\text{max}}$ larger without melting the silicon?

By following CMOS scaling rules, one topic of today's lecture.
CMOS: the evolution of gate length over 35+ years

Technology Nodes:
A node every 2 yrs
$L_n = 0.7 \ L_{n-1}$
$A_n = 0.5 \ A_{n-1}$
Density doubles every two years.

From the Road Map - more Lec 26

Figures: Prof. D. Antoniadis
Clearly the velocity of the electrons and holes in the channel will be saturated at even low values of $v_{DS}$!
What does this mean for the device and inverter characteristics?

**CMOS**: velocity saturation

Sanity check before looking at device scaling

CMOS gate lengths are now under 0.1 μm (100 nm). The electric field in the channel can be very high: $E_y \geq 10^4$ V/cm when $v_{DS} \geq 0.1$ V.
**MOS:** Output family with velocity saturation

\[
i_D(v_{GS}, v_{DS}, v_{BS}) \approx \begin{cases} 
0 & \text{for } v_{GS} < V_T, \ 0 < v_{DS} \\
W s_{sat} C_{ox}^* [v_{GS} - V_T(v_{BS})] & \text{for } V_T < v_{GS}, \ E_{crit} L < v_{DS} \\
\frac{W}{L} \mu_e C_{ox}^* [v_{GS} - V_T(v_{BS})] v_{DS} & \text{for } V_T < v_{GS}, \ 0 < v_{DS} < E_{crit} L
\end{cases}
\]

This simple model for the output characteristics of a very short channel MOSFET (plotted above) provides us an easy way to understand the impact of velocity saturation on MOSFET and CMOS inverter performance.
CMOS: Gate delay and $f_{\text{max}}$ with velocity saturation

Charge/discharge cycle and gate delay:
The charge and discharge currents, charges, and times are now:

$$i_{\text{Disch arg e}} = i_{\text{Ch arg e}} = W_{\text{min}} s_{\text{sat}} C_{ox}^* (V_{DD} - V_{Tn})$$

$$q_{\text{Disch arg e}} = q_{\text{Ch arg e}} = C_{L} V_{DD} = 3 W_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}$$

$$\tau_{\text{Disch arg e}} = \tau_{\text{Ch arg e}} = \frac{q_{\text{Disch arg e}}}{i_{\text{Disch arg e}}} = \frac{3 W_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}}{W_{\text{min}} s_{\text{sat}} C_{ox}^* (V_{DD} - V_{Tn})} = \frac{3 n L_{\text{min}} V_{DD}}{s_{\text{sat}} (V_{DD} - V_{Tn})}$$

CMOS minimum cycle time and power density at $f_{\text{max}}$:

$$\tau_{\text{Min. Cycle}} = \tau_{\text{Ch arg e}} + \tau_{\text{Disch arg e}} = \frac{6 n L_{\text{min}} V_{DD}}{s_{\text{sat}} [V_{DD} - V_{Tn}]}$$

Note: $\tau_{\text{ChanTransit}} = \frac{L_{\text{min}}}{s_{\text{sat}}}$

$$\tau_{\text{Min. Cycle}} \propto \frac{L_{\text{min}} V_{DD}}{s_{\text{sat}} [V_{DD} - V_{Tn}]} = n' \tau_{\text{ChanTransit}}$$

Lessons: We still benefit from reducing L, but not as quickly. Channel transit time, $L_{\text{min}}/s_{\text{sat}}$, is still critical.
CMOS: Power and power density with velocity saturation

Average power dissipation

All dynamic

\[ P_{\text{ave}} = E_{\text{Dissipated per cycle}} \cdot f = C_L V_{DD}^2 = 3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2 \]

Power at maximum data rate

Maximum f will be \(1/\tau_{\text{Gate Delay Min.}}\)

\[ P_{\text{dyn} @ f_{\text{max}}} = \frac{3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2}{\tau_{\text{Min. Cycle}}} = 3nW_{\text{min}} L_{\text{min}} C_{ox}^* V_{DD}^2 \cdot \frac{s_{\text{sat}} [V_{DD} - V_{Tn}]}{6n L_{\text{min}} V_{DD}} \]

Power density at maximum data rate

Assume that the area per inverter is proportional to \(W_{\text{min}} L_{\text{min}}\)

\[ PD_{\text{dyn} @ f_{\text{max}}} = \frac{P_{\text{dyn} @ f_{\text{max}}}}{\text{Inverter Area}} \propto \frac{P_{\text{dyn} @ f_{\text{max}}}}{W_{\text{min}} L_{\text{min}}} = \frac{s_{\text{sat}} C_{ox}^* V_{DD} [V_{DD} - V_{Tn}]}{L_{\text{min}}} \]

Lesson: Again benefit from reducing L, but again not as quickly.

Clif Fonstad, 4/5/11
CMOS: Collected results

**Maximum data rate:**

No velocity saturation:

\[ \tau_{\text{Min.Cycle}} \propto \frac{L_{\text{min}}^2 V_{DD}}{\mu_e \left[ V_{DD} - V_{Tn} \right]^2} \]

With velocity saturation:

\[ \tau_{\text{Min.Cycle}} \propto \frac{L_{\text{min}} V_{DD}}{s_{\text{sat}} \left[ V_{DD} - V_{Tn} \right]} \]

**Power density at maximum data rate:**

No velocity saturation:

\[ PD_{\text{dyn @ } f_{\text{max}}} = \frac{\mu_e \varepsilon_{\text{ox}} V_{DD} \left[ V_{DD} - V_{Tn} \right]^2}{t_{\text{ox}} L_{\text{min}}^2} \]

With velocity saturation:

\[ PD_{\text{dyn @ } f_{\text{max}}} = \frac{s_{\text{sat}} \varepsilon_{\text{ox}} V_{DD} \left[ V_{DD} - V_{Tn} \right]}{t_{\text{ox}} L_{\text{min}}} \]

Smaller is faster

Smaller also dissipates more power per unit area

Smaller also dissipates more power per unit area
Scaling Rules - making CMOS faster without melting Si

**General idea:**
Reduce dimensions by factor $1/s$: $s > 1$
Evaluate impact on speed, power, power density
Assume no velocity saturation for now

**Scaling dimensions alone:**

$$L_{\text{min}} \rightarrow L_{\text{min}} / s$$
$$W \rightarrow W / s$$
$$t_{\text{ox}} \rightarrow t_{\text{ox}} / s$$
$$N_A \rightarrow sN_A$$

This yields

$$C_{\text{ox}}^* = \frac{\varepsilon_{\text{ox}}}{t_{\text{ox}}} : C_{\text{ox}} \rightarrow sC_{\text{ox}}$$

and thus

$$\tau \propto \frac{L_{\text{min}}^2 V_{DD}}{\mu_e \left[V_{DD} - V_{\text{Tn}}\right]^2} : \tau \rightarrow \tau / s^2$$

$$P_{\text{dyn}} = 3nW_{\text{min}} L_{\text{min}} C_{\text{ox}}^* V_{DD}^2 f : P_{\text{dyn}} \rightarrow sP_{\text{dyn}}$$

$$PD_{\text{dyn} @ f_{\text{max}}} = \frac{\mu_e \varepsilon_{\text{ox}} V_{DD} \left[V_{DD} - V_{\text{Tn}}\right]^2}{t_{\text{ox}} L_{\text{min}}^2} : PD_{\text{dyn} @ f_{\text{max}}} \rightarrow s^3PD_{\text{dyn} @ f_{\text{max}}}$$

**Scaling dimensions alone can yield melted silicon!!**
Scaling Rules, cont. - constant E-field scaling

Observation:
Reducing dimensions alone won't work.
Reduce voltage in concert (constant E-field scaling)

Scaling dimensions and voltages by 1/s:

\[ L_{\text{min}} \rightarrow L_{\text{min}} / s \quad W \rightarrow W / s \quad t_{\text{ox}} \rightarrow t_{\text{ox}} / s \quad N_A \rightarrow sN_A \]
\[ V_{DD} \rightarrow V_{DD} / s \quad V_{BS} \rightarrow V_{BS} / s \quad V_T \rightarrow V_T / s \]

We still have
\[ C_{ox}^* \rightarrow sC_{ox}^* \quad K \rightarrow sK \]

but now we find

\[ \tau \propto \frac{L_{\text{min}}^2 V_{DD}}{\mu_e [V_{DD} - V_{Tn}]^2} : \quad \tau \rightarrow \tau / s \]
\[ P_{\text{dyn}} = 3nW_{\text{min}}L_{\text{min}}C_{ox}^* V_{DD}^2 f : \quad P_{\text{dyn}} \rightarrow P_{\text{dyn}} / s^2 \]
\[ PD_{\text{dyn}} @ f_{\text{max}} = \frac{\mu_e \varepsilon_{ox} V_{DD} [V_{DD} - V_{Tn}]^2}{t_{\text{ox}} L_{\text{min}}^2} : \quad PD_{\text{dyn}} @ f_{\text{max}} \rightarrow PD_{\text{dyn}} @ f_{\text{max}} \]

When we scale dimension and voltage we get higher speed and lower power, while holding the power density unchanged.
Scaling Rules, cont. - constant E-field scaling

Threshold voltage:
We've said $V_T$ scales, but this merits some discussion*:

$$V_T(v_{BS}) \equiv V_{FB} + |2\phi_{p-Si}| + \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2\varepsilon_{Si} q N_A \left[ |2\phi_{p-Si}| + |v_{BS}| \right]}$$

Small because with n$^+$-poly Si gate, $\phi_m \approx -\phi_p$ and $V_{FB} \approx -|2\phi_p|$

Dominated by $|v_{BS}|$ if $|v_{BS}| >> |2\phi_p|$

Thus:

$$V_T(v_{BS}) \approx \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2\varepsilon_{Si} q N_A |v_{BS}|} \rightarrow \frac{t_{ox} |s|}{\varepsilon_{ox}} \sqrt{2\varepsilon_{Si} q sN_A |v_{BS}| |s|} \rightarrow V_T / s$$

It works.

Subthreshold leakage and static power:
Including $V_{BS}$, $I_{D,\text{off}}$ is:

$$I_{D,\text{off}} \approx \frac{W}{L} \mu_e V_t^2 \sqrt{\frac{\varepsilon_{Si} q N_A}{2\left[-|2\phi_p| + |V_{BS}|\right]}} e^{-V_T/nV_t} \approx \frac{W}{L} \mu_e V_t^2 \sqrt{\frac{\varepsilon_{Si} q N_A}{2 |V_{BS}|}} e^{-V_T/nV_t}$$

Scaling all the factors, we find that $I_{D,\text{off}}$ and $P_{\text{static}}$ scale poorly!

$$I_{D,\text{off}} \rightarrow s I_{D,\text{off}} e^{\left\{(1-\frac{1}{s})V_T\right\}/nV_t} \quad P_{\text{Static}} = V_{DD} I_{D,\text{off}} \rightarrow P_{\text{Static}} e^{\left\{(1-\frac{1}{s})V_T\right\}/nV_t}$$

* We’re talking n-channel here, but similar results are found for the p-channel MOSFETs.
Scaling Rules, cont. - static power scales badly, but...

Static power density's scaling is even worse:

\[
PD_{\text{static}} = \frac{I_{D,\text{off}} V_{DD}}{W_{\text{min}} L_{\text{min}}} \rightarrow sI_{D,\text{off}} e^{(s-1)V_T / s n V_t} \frac{V_{DD}}{W_{\text{min}} L_{\text{min}} / s^2} \rightarrow s^2 e^{(s-1)V_T / s n V_t} PD_{\text{static}}
\]

A typical \( V_T / nV_t \) is \( \sim 10 \). If \( s = \sqrt{2} \), the exponential factor is \( \sim e^3 \), or about 20!

Bottom Line: Static power can no longer be neglected.

Figure source: Intel Web Site
Scaling Rules, cont. - What about velocity saturation?

Do the same constant E-field scaling by 1/s:

\[ \frac{L_{\text{min}}}{s} \rightarrow \frac{L_{\text{min}}}{s}, \quad \frac{W}{s} \rightarrow \frac{W}{s}, \quad \frac{t_{\text{ox}}}{s} \rightarrow \frac{t_{\text{ox}}}{s}, \quad \frac{N_A}{s} \rightarrow \frac{sN_A}{s} \]

\[ \frac{V_{\text{DD}}}{s} \rightarrow \frac{V_{\text{DD}}}{s}, \quad \frac{V_{\text{BS}}}{s} \rightarrow \frac{V_{\text{BS}}}{s}, \quad \frac{V_T}{s} \rightarrow \frac{V_T}{s} \]

so \[ C_{ox}^* \rightarrow sC_{ox}^*, \quad K \rightarrow sK \]

Examining our expressions when velocity saturation is important we find:

\[ \tau \propto \frac{L_{\text{min}}V_{\text{DD}}}{s_{\text{sat}}[V_{\text{DD}} - V_{Tn}]} : \quad \tau \rightarrow \frac{\tau}{s} \]

\[ P_{\text{dyn}} = 3nW_{\text{min}}L_{\text{min}}C_{ox}^* V_{DD}^2 f : \quad P_{\text{dyn}} \rightarrow P_{\text{dyn}}/s^2 \]

\[ PD_{\text{dyn}} \frac{V_{DD}}{f_{\text{max}}} = \frac{s_{\text{sat}} \epsilon_{ox} V_{DD}[V_{DD} - V_{Tn}]}{t_{\text{ox}} L_{\text{min}}} : \quad PD_{\text{dyn}} \frac{V_{DD}}{f_{\text{max}}} \rightarrow PD_{\text{dyn}} \frac{V_{DD}}{f_{\text{max}}} \]

Amazingly, the scaling behavior of the gate delay, average power, or power density is the same with or without velocity saturation!

Note: Velocity saturation is not a factor in \( I_{D,\text{off}} \).
Charge Coupled Devices (CCDs)

Nobel Prize in Physics 2009
"Two Revolutionary Optical Technologies"

Charles K. Kao - for initiating the search for and the development of the low-loss optical fiber

Willard S. Boyle and George E. Smith - for inventing the charge coupled device

"CCDs are widely used in digital cameras and in advanced medical and scientific instrumentation." Nobel Committee*

"And they are something you can understand in 6.012." me

The two-terminal n-MOS capacitor

Right: Basic device

For $v_{GB} \leq V_T$ nothing is different, but when $v_{GB} > V_T$, where do the electrons for the inversion layer come from?

They diffuse to the edge of the depletion region from the bulk. This is like reverse bias diode saturation current and it takes a long time to build up the inversion layer charge.
The MOS light detector -

What if we shine light on our biased MOS capacitor?

Electrons optically generated in and near the depletion region will be populate the inversion layer. The number collected in a frame time (clock period) is proportional to the light intensity.
Two adjacent MOS capacitors:

\[ V_{G2S} > V_{G1S} > V_T \]

\[ V_{G1S} > V_{G2S} > V_T \]

The charge can be passed back and forth between them.
Charge-coupled devices, CCDs: basically shift registers

An array of closely spaced 2-terminal MOS capacitors can be used to shift data along in a serial bit stream in the form of packets of electrons.
Charge-coupled devices: CCD shift registers

$\phi_1 > \phi_3 > \phi_2 > V_T$

$\phi_2 > \phi_1 > \phi_3 > V_T$

$\phi_3 > \phi_2 > \phi_1 > V_T$
The charge is shifted along and read serially using a reverse biased diode and MOS source followers.

Lifted from a Kodak website:
CCD imagers - 1-d and 2-d arrays

A linear CCD imaging array is made by placing MOS sensor pixels next to a CCD shift register, which collects their outputs and sends them out in a serial stream.

To make a 2-d CCD imager, 1-d CCD imaging arrays are integrated as adjacent columns that are coupled into a horizontal CCD shift register to combine their outputs into a row-by-row serial bit stream of the image.

2-d array figure lifted from Nobel website:
CMOS gate delay and power

Three key performance metrics: (We want to make them all smaller)

- Gate Delay = $12 \times 10^{-9} L_{\text{min}}^2 V_{\text{DD}} / \mu_e (V_{\text{DD}} - V_T)^2$
- $P_{\text{dyn}}@f_{\text{max}} \propto C_L V_{\text{DD}}^2/GD = (W_n/L_{\text{min}}) \mu_e C^*_{\text{ox}} V_{\text{DD}} (V_{\text{DD}} - V_T)^2/4$
- $P_{\text{PD, dyn, max}} \propto P_{\text{dyn}}@f_{\text{max}}/W_n L_{\text{min}} = \mu_e \varepsilon_{\text{ox}} V_{\text{DD}} (V_{\text{DD}} - V_T)^2/4 t_{\text{ox}} L_{\text{min}}^2$

CMOS scaling rules

Summary of rules: Constant E-field - scale all dimensions and all voltages by 1/s

Scaling as: $L_{\text{min}} \rightarrow L_{\text{min}}/s$ Results in: $K \rightarrow s K$
- $w \rightarrow w/s$
- $t_{\text{ox}} \rightarrow t_{\text{ox}}/s$
- $N_A \rightarrow s N_A$
- $V_T, V_{BS}, V_{DD} \rightarrow V_T/s, V_{BS}/s, V_{DD}/s$
- $P_{\text{dyn}} \rightarrow P_{\text{dyn}}/s^2$
- $PD_{\text{dyn}} \rightarrow PD_{\text{dyn}}$

Charge coupled devices (CCDs)

Shifting charge from gate to gate: Exploiting deep depletion