

第 9 回 MOSFET の概略と MOS 構造における反転層の形成

MOSFET の歴史と地位

MOS トランジスタはバイポーラより遅い 1960 年に発明(1963 年に CMOS) されたものの、80 年代にはバイポーラより強くなり(デジタル回路における主流が入れ替わったのは 90 年位)、いまではほとんどの集積回路は MOS で出来ている。

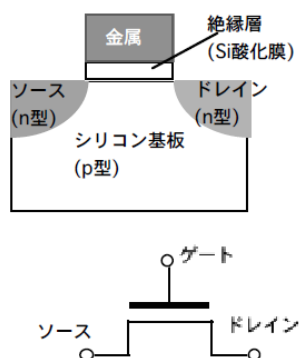
現在は集積回路と呼ばれるものの殆どすべてが CMOS と呼ばれる MOSFET の組み合わせた回路で構成されている。集積回路には非常に多くのトランジスタが入っている。Wikipedia で”transistor count”とすると、例えば iPhone11Pro に入っている Apple A13 はワンチップで 85 億個のトランジスタが入っている。GPU や FPGA だと 500 億個くらい入っているものもある。

7nm プロセスルールなどという言葉や IC の最小寸法は小さくすると速くなるという話を聞いたことが有るかと思う。これも MOSFET の特性から来ている。明らかに他より安く作れることから、アナログ回路にも最近はどんどんその分野を広げつつある。そこで、まずは、なぜ MOS トランジスタが動作するかを説明した後に、なぜ小さくすると速くなるか、そしてその流れを阻害するのは何かを理解するべく、オーバークロックなど、時々 CPU の性能を上げるのに使われるジャーゴンなどを入れつつ行いたい。

MOSFET の構造

概略構造からまず示そう。中央にシリコン基板上に酸化膜と金属が載っている、この金属(Metal)-酸化膜(Oxide)-半導体(Semiconductor)構造から MOS トランジスタと呼ばれている。

そして、電極は 3 カ所に取られている。MOS 構造の両脇にある強くドーピングされた領域であるドレイン(Drain: 下水管等の意味)とソース(Source: 源の意味)、上部の金属のゲート(門: Gate)である。回路図も示そう。英語の意味から推測がつくように、電子や正孔等のキャリアは、ゲート直下の半導体部分の流れ、ソースからドレインへ流れる。そしてその量をゲートの電圧で制御する。



また、バイポーラトランジスタに対して、ユニポーラトランジスタと呼ばれる場合もある。これは、バイポーラトランジスタが電子と正孔の両方を使っているのに対し、MOS トランジスタが電子、または正孔のどちらか一方のみのキャリアで電流を運ぶからである。FET と呼ばれる。これは電界効果トランジスタ(Field Effect Transistor)の意味である。外部からの電圧印加による電界の効果で電流を支配するからである。なお、ここからの話は原則として、電子をキャリアとする n-MOS の話のみとする。正孔をキャリアとする p-MOS の場合は、電圧を逆方向に、伝導帯を価電子帯に読みかえれば良い。(現在の IC 技術の帝王である CMOS は p-MOS と n-MOS の両方からなる。CMOS の有利さは低消費電力性にあり、回路構成に起因している。)

MOS 構造でのキャリア数

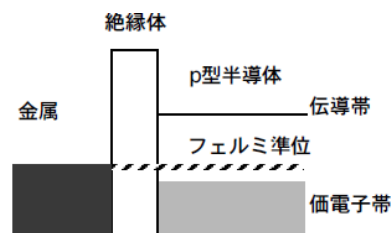
さて、続けて、電圧が掛かったときの MOS 構造の電子の数を考えよう。ここで pn 接合は整流性(すなわち p 形の電位が高いと電流が流れ(順バイアス)、低いと電流が流れない(逆バイアス))を持つと考える。

まず、ゲートには電圧を加えない状態で、ソースとドレインの間に電圧をかけても電流が流れない様になっているか、先に示した MOSFET 構造の断面図のドーピングを考えよう。ソースは電子を供給するので n 型に強くドーピングされている。またドレインから電子を電極に取り出すので、やはり n 型に強くドーピングされている。ソースとドレインの間が n 型では、いつでも電流が流れてしまう。そこで MOS 構造を形成しているゲート下の半導体は p 型となる。この p 型

はまずは接地されていると考えてよい。(このゲート下の半導体の電位によって、アナログ電子回路などで重要となるしきい値の変化が起こるが、これは 12 回で示す。)従って npn 構造となり、電圧をかけてもどちらかの pn 接合が逆バイアスされ、電流は流れない。

次に、ゲートの断面でのバンド図を描こう。初めに半導体にゲート電圧がかかっておらず、電界が印加されていない状態では、半導体は p 型なので正孔で満たされた中性状態であり、バンドはフラットである。そして絶縁体は非常に大きなバンドギャップ(約 8V)を持ち電流が流れないように大きなバリアを持っている。そして金属はフェルミ準位まで電子が満たされている状態である。電圧がかかっていないのですべてのフェルミ準位は flat である。

(これは当然のようにみえるが、じつは、理想フラットバンド状態と呼ばれる理想的な状態の一つである。普通は少しゲートに電圧を印加しないと、半導体中にフラットな状況は作れない。これも詳しくは第 12 回で)

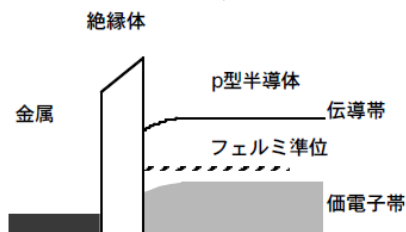


次に流れる状態をつくる為にゲートに電圧をかけて、電子が流れるようにしよう。このとき半導体中に電子がいる状態にしたいので、半導体に負電荷ができるように電圧をかける。すなわち、ポテンシャルとしては下向きにメタルのフェルミレベルを動かす。以下特に注意しない限り、基板は接地されていてかつ平衡状態にあると考える。

始めは、電子を誘起できない。なぜならば、始め酸化膜下には、アクセプタと正孔がいて、電圧をかけたときに流れてゆき、負電荷であるアクセプタが残って、電気力線を終端させる。不純物のみが残っている状態は、pn 接合時の空乏層と同じなので、空乏化していると呼び、ここでポテンシャルの曲がりが生じている。

pn 接合のときに電圧をかけないときには、電流が流れず、フェルミ準位は flat だった。いまも、絶縁層があるため、電流が流れることはない。キャリアが存在し、電流が流れない状態では、バンド(伝導帯や価電子帯)のフェルミ準位は flat であり、MIS 構造での半導体中での垂直方向は、常にフェルミ準位が flat である。絶縁体では電荷がまったく存在しないので、フェルミ準位が曲がって良い。(19 乗のキャリアが、半導体中にいたとしても、1.1eV 程度上に行けば、1 立方 cm 当たりのキャリアは 1 まで減る。一方、絶縁体のコンダクションバンドは、伝導帯より 3V 程度上なので、電荷の存在は、(特殊なイオンによるもの以外)無視して良い。)

そこで、電圧印加時には、絶縁層内でフェルミ準位が変化するが、半導体中はフラットであるとしてよい。



まだドレインソース間に電流が流すための電子が生成できなかった、さらに、ゲート電圧を印加すると、絶縁物近傍の伝導帯底がフェルミ準位と近づき、電子が誘起される。

ここで、p 形半導体の中性時のキャリア濃度を N_A としよう。どれだけ電子がでたら、その電子で電気力線が終端されるかと言えば、目安は、アクセプタの数である N_A となる。

電子の数は、バンドの曲がりによって依存した位置の関数である真性フェルミ準位 とフェルミ準位の差で決まる。式で表せば真性フェルミ準位はバンドの曲がりにより変化する位置の関数なので、

$$n(x) = n_i \exp\left(\frac{E_f - E_i(x)}{kT}\right)$$

ここで、 n_i は真性キャリア密度である。また正孔は同様に

$$p(x)=n_i \exp\left(-\frac{E_f-E_i(x)}{kT}\right)$$

と表される。

p 形半導体の中性時(即ち $x=\infty$)の真性フェルミ準位 $E_i(\infty)$ で正孔の数、即ちアクセプタの数を表すと、

$$N_A=n_i \exp\left(-\frac{E_f-E_i(\infty)}{kT}\right)$$

である。そこで、この関係で真性キャリア濃度を表すと

$$n_i=N_A \exp\left(\frac{E_f-E_i(\infty)}{kT}\right)$$

となる。絶縁体/半導体の界面を $x=0$ とすると、絶縁体/半導体の境界での電荷の数は、

$$n(0)=n_i \exp\left(\frac{E_f-E_i(0)}{kT}\right)=N_A \exp\left(\frac{2E_f-E_i(0)-E_i(\infty)}{kT}\right)$$

となる。従って、

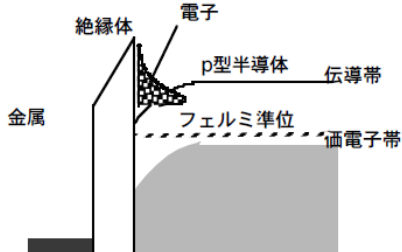
$$2E_f-E_i(0)-E_i(\infty)=0$$

となるあたりから、電子が電気力線終端の主役となる。このときのポテンシャルの曲がった量の全体は $E_i(\infty)-E_i(0)$ であり、この絶対量を電圧で表すと

$$\frac{E_i(\infty)-E_i(0)}{q}=\frac{E_i(\infty)-(2E_f-E_i(\infty))}{q}=\frac{2(E_i(\infty)-E_f)}{q}=2\phi_F$$

となる。ここでフェルミポテンシャル $\phi_F \equiv \frac{2(E_i(\infty)-E_f)}{q}$ を定義した。

これは大まかに言えば、バンドギャップより少し小さいくらいの電位変化が界面で起こったことになる。



このあとのポテンシャルの形は、式から判るように、60meV の変化で界面での電子の量は 10 倍になる早い変化である。従って、この後、ポテンシャルの形は、いままでの大きな変化から見ると充分無視できる小さな変化だけとなる。

正確に言えば、反転後の電子の総量は、反転電子の分布(厚み)を考えてポアソンの方程式を解いて出す必要がある。その場合は 120meV で 10 倍になるもう少し遅い変化である。ただ、これを解くには 8 つくらい式を解くので、電子デバイス第二で扱う。ここでは、電荷は指数関数的に非常に鋭い形で分布しており、表面の電位分布が変わっても、ほとんど奥の空乏層に変化は無い。また、非常に反転電子層が薄くなり、電子が量子化すると、この鋭い変化は、すこしゆるくなる。

ここで、p 形半導体において、主な電荷が電子となる状態となっている。この状態を**反転**とよぶ。

厳密には、正孔より電子が増えたときは反転であり、ここでいう反転の前でも起こっている。そこで弱反転(アクセプタ濃度より電子の数が少ない)と強反転(ここでいう反転)という言い方もある。

さらに、この反転が起こり始めるゲート電圧をしきい値電圧: V_T としよう。反転で形成される電子は非常に薄い領域でのみ生成され、反転層は一次的に扱う(面電荷として扱って良い)ことができる。ゲート電圧 V_G を V_T とより大きくすることにより生成される電気力線はすべて界面にある面電荷で終端されると考えて良い。

そこでそのときの誘導される電子による面電荷は、酸化膜の面積あたりの容量を C_{ox} 、厚さを d 、誘電率を ϵ_{ox} とすれば、電荷が負であることを勘案すれば単位面積当たり

$$Q_1=-C_{ox}(V_G-V_T)=-\frac{\epsilon_{ox}}{d}(V_G-V_T)$$

となる。

実際は、 $V_G=V_T$ でも、 $V_G<V_T$ でも電子による電荷は存在するが、反転層形成後の変化に較べると、とりあえず無視して良いほど小さい。

電界と移動度・電流

MOSFET は、一般に電界によって電子や正孔が流れるドリフトによって電流が支配される。電界 E がかって電荷が加速され、ある平均時間で原子や不純物にぶつかり散乱される。そのあいだの平均速度は電界に比例する。このドリフトによる電流が MOSFET の動作原理である。バイポーラトランジスタでは、電極抵抗等に効いてくる。平均速度に電子数を掛ければ電流なので

$$J_n=qnv=qn\mu_n E$$

ここで、 μ_n は電子の移動度である。

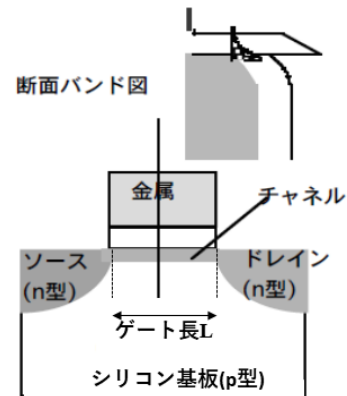
(正孔では $J_p=qpv=qp\mu_p E$ で μ_p は正孔の移動度。)

なお、確認すると、バイポーラトランジスタで使った拡散係数と移動度はアインシュタインの関係式 $D=kT\mu$ で比例関係にある。

ドレイン電圧が小さいときの電圧-電流特性

さて、本来電子がいなかったゲート下の p 形半導体中に電子が誘起できたので、ドレイン-ソース間に電圧をかければ電流が流れることになる。この電子が流れている状態では、この部分をチャンネル(Channel:水道管の意味)と呼ぶ。

もしも、ドレイン-ソース間の電圧がゲート電圧に較べて非常に小さければ、チャンネルの電子密度は一定として良い。その近似の元で、ソースの電圧を原点として、ドレインの電圧が V_D とする。またゲート下で電流が流れるチャンネルの幅を W 、長さを L (ゲート長またはチャンネル長) とする。



すると、チャンネルの横方向の電界は V_D/L となり、移動度が μ とすると、

単位幅当たりの電流は $vQ_1=-\mu EQ_1=-\mu \frac{V_D}{L} Q_1$ となる。

チャンネル幅まで入れると電流になりトランジスタのドレインに流れこむ電流は

$$I_D=-\mu \frac{W}{L} V_D Q_1=\mu \frac{W}{L} V_D C_{ox}(V_G-V_T)=\mu \frac{W}{L} V_D \frac{\epsilon_{ox}}{d}(V_G-V_T)$$

となる。

すなわちゲート電圧によって変化する抵抗として考えることができる。