

Halbleitertechnologie von A bis Z

Digitaltechnik

Statische und dynamische Leistungsaufnahme, Leckströme	3
<i>Dynamische Verlustleistung</i>	<i>3</i>
<i>Statische Verlustleistung und Leckströme</i>	<i>3</i>
<i>Techniken zur Leistungsreduktion</i>	<i>4</i>

Statische und dynamische Leistungsaufnahme, Leckströme

Dynamische Verlustleistung

Die dynamische Verlustleistung entsteht beim Umladen der parasitären Lastkapazität eines Gatters während eines Schaltvorgangs und wird durch $P_{\text{dyn}} = \alpha \cdot C_L \cdot V_{\text{DD}}^2 \cdot f$ beschrieben, wobei C_L die Lastkapazität, V_{DD} die Versorgungsspannung, f die Taktfrequenz und α die Schaltaktivität ist, also der Anteil der Taktzyklen, in denen das Gatter tatsächlich umschaltet und die in typischen Logikschaltungen bei etwa 0,1 bis 0,3 liegt. Für ein Gatter mit $C_L = 1 \text{ fF}$, $V_{\text{DD}} = 0,8 \text{ V}$, $f = 2 \text{ GHz}$ und $\alpha = 0,2$ ergibt sich eine dynamische Verlustleistung von rund $0,26 \text{ }\mu\text{W}$ – multipliziert mit mehreren hundert Millionen Gattern eines modernen Chips wird daraus schnell eine Gesamtleistung im Watt-Bereich. Da die Verlustleistung quadratisch mit der Versorgungsspannung wächst, ist die Absenkung von V_{DD} eines der wirksamsten Mittel zur Reduktion der dynamischen Leistungsaufnahme, wird jedoch durch die notwendige Mindestspannung für zuverlässiges Schalten begrenzt.

Zur dynamischen Verlustleistung zählt neben der eigentlichen Umladeleistung auch die Kurzschlussleistung, die während der endlichen Anstiegs- und Abfallzeit eines Signals entsteht, wenn PMOS- und NMOS-Pfad eines CMOS-Gatters kurzzeitig gleichzeitig leiten und die je nach Flankensteilheit 10 bis 20 % der gesamten dynamischen Verlustleistung ausmachen kann. Sie lässt sich durch möglichst steile Signalfanken, also eine ausreichende Treiberstärke der vorangehenden Stufe, gering halten.

Statische Verlustleistung und Leckströme

Auch wenn ein CMOS-Gatter nicht schaltet, fließt ein kleiner, aber bei kleinen Strukturgrößen nicht mehr vernachlässigbarer Reststrom, der die statische Verlustleistung bestimmt. Der dominante Beitrag ist der Subthreshold-Leckstrom, der auch bei einer Gate-Source-Spannung unterhalb der Schwellenspannung noch fließt, da der Kanal nicht vollständig ausgeschaltet ist; er wächst exponentiell mit sinkender Schwellenspannung und steigender Temperatur, charakterisiert durch die Subthreshold-Steigung (Subthreshold Swing), die bei realen MOSFETs bei etwa 70 bis 100 mV pro Dekade Stromänderung liegt, gegenüber einem theoretischen Minimum von rund 60 mV/Dekade bei Raumtemperatur.

Weitere Beiträge sind der Gate-Leckstrom durch das immer dünner werdende Gate-Oxid sowie Sperrstrom über die in Sperrrichtung betriebenen Drain- und

Source-pn-Übergänge. Da moderne Technologien mit niedrigeren Schwellenspannungen von oft nur 0,2 bis 0,3 V arbeiten, um bei sinkender Versorgungsspannung ausreichend Schaltgeschwindigkeit zu erhalten, steigt der Anteil der Leckleistung an der Gesamtverlustleistung mit jeder Technologiegeneration und kann in fortschrittlichen Prozessen bereits 20 bis 40 % der Gesamtverlustleistung eines Chips im Ruhezustand ausmachen.

Techniken zur Leistungsreduktion

Clock Gating unterbricht gezielt die Taktversorgung von Schaltungsteilen, die im aktuellen Betriebszustand nicht benötigt werden, und verhindert so unnötige Schaltaktivität und damit dynamische Verlustleistung, ohne den gespeicherten Zustand der betroffenen Register zu verändern; in typischen Designs lassen sich dadurch 10 bis 30 % der dynamischen Gesamtverlustleistung einsparen. Power Gating geht einen Schritt weiter und trennt ganze Schaltungsblöcke über Schalttransistoren (Header- oder Footer-Switches) vollständig von der Versorgungsspannung, wodurch zusätzlich die statische Leckleistung dieser Blöcke um ein bis zwei Größenordnungen sinkt, allerdings der gespeicherte Zustand verloren geht, sofern er nicht zuvor in einem separaten Retention-Flip-Flop gesichert wurde.

Beim Multi- V_t -Design werden auf demselben Chip Transistoren mit unterschiedlicher Schwellenspannung eingesetzt: Zellen auf zeitkritischen Pfaden erhalten eine niedrige Schwellenspannung für hohe Schaltgeschwindigkeit, während unkritische Zellen mit hoher Schwellenspannung und entsprechend deutlich geringerem, oft um eine Größenordnung reduziertem Leckstrom ausgeführt werden. Dynamic Voltage and Frequency Scaling (DVFS) passt Versorgungsspannung und Taktfrequenz zur Laufzeit an die tatsächlich benötigte Rechenleistung an: Eine Halbierung von Spannung und Frequenz reduziert die dynamische Verlustleistung dabei theoretisch auf ein Achtel, was DVFS besonders im Teillastbetrieb mobiler Prozessoren zu erheblichen Energieeinsparungen macht.