

Dynamic Voltage and Frequency Scaling

Energiespartechnik bei Recheneinheiten

Silvio Nießner



Sommerakademie in Leysin, August 2016

Abstract

Dynamic Voltage and Frequency Scaling, kurz: DVFS, ist eine Energiesparstrategie für Recheneinheiten, wie CPUs und GPUs. Dabei wird die am Prozessor angelegte Spannung und Taktfrequenz der Recheneinheit variiert und an die aktuell benötigte Rechenleistung angepasst. Dabei kann eine Energieeinsparung von bis zu 46,5% erreicht werden. Im Folgenden werden ausgehend von Spannungs- und Frequenzanpassung das Zusammenspiel dieser Parameter erläutert und unterschiedliche Software sowie Hardwareumsetzungen vorgestellt.

Contents

1	Dynamic Voltage and Frequency Scaling [Softwareumsetzung]	3
1.1	Frequenzanpassung	3
1.2	Spannungsanpassung	4
1.3	Spannungs- und Frequenzanpassung mittels Aufgabenstaffelung	4
2	Dynamic Voltage and Frequency Scaling [Hardwareumsetzung]	4
2.1	Spannungsanpassung mittels Spannungsinseln	4
2.2	Spannungsreduktion mittels Strukturverkleinerung	5
2.3	Herstellungsbedingter Spielraum für Spannungsreduktion	5
3	Anwendungsbereich und Zusammenfassung	7

1 Dynamic Voltage and Frequency Scaling [Softwareumsetzung]

1.1 Frequenzanpassung

Ein Baustein dieses Verfahrens bzw. dieser Strategie ist eine Frequenzanpassung des Prozessors an die aktuell benötigte Rechenleistung. Dabei wird zum Energiesparen eine Frequenzreduktion vorgenommen, die auch als Untertakten/ Underclocking bezeichnet wird. Dadurch nehmen die Wärmeemissionen ab, sodass eine geringer dimensionierte Kühlung ausreicht und z.B. eine kompaktere Bauweise ermöglicht.

Die Energieeinsparung bei einer Frequenzreduktion ergibt sich allerdings nicht nur aus einer verkleinerten Kühlung sondern gleichwohl durch den Energieverbrauch der Recheneinheit selbst.

Dieser lässt sich nach der folgenden Formel in einen konstanten und einen variablen Teil untergliedern:

$$P_{ges} = P_{const.} + P_{var.}$$

Für eine Anpassung verschiedener Parameter, wie Frequenz und Spannung, betrachtet man nur den variablen Energieverbrauch, da nur dieser veränderbar ist:

$$P_{var.} = A * C * V^2 * f$$

A = Aktivitätsfaktor

C = Kapazität

V = Betriebsspannung

f = Frequenz

Reduziert man die Frequenz und untertaktet dadurch die Recheneinheit, kann eine lineare Reduktion des Energieverbrauchs erreicht werden, was mit einer geringeren Kühlleistung einhergeht. Diese Anpassung hat zur Folge, dass sich das System unter Umständen verlangsamt, dafür allerdings eine größere Systemstabilität bietet.

1.2 Spannungsanpassung

Alternativ lässt der dynamische Energieverbrauch weitere Optimierungsoptionen offen. So kann zum Beispiel auch eine Spannungsreduktion erfolgen, die im Gegensatz zu der Frequenzanpassung sogar einen quadratischen Einfluss auf den Energieverbrauch hat. Reduziert man also die Spannung, kann ebenfalls Energie und Kühlleistung eingespart werden. Durch die Spannungsreduktion ergibt sich allerdings der Nachteil, dass das System mit sinkender Spannung zunehmend an Stabilität verliert.

1.3 Spannungs- und Frequenzanpassung mittels Aufgabenstaffelung

Kombiniert man nun die beschriebene Frequenzreduktion mit einer Spannungsreduktion kommt es zu einer Überlagerung der Vor- und Nachteile. Die Energieeinsparung erfolgt mit der dritten Potenz und die Instabilität durch die Spannungsreduktion wird durch die stabilisierende Eigenschaft der Frequenzreduktion kompensiert. Als einziger Nachteil verbleibt eine mögliche Verlangsamung des Systems. Genau diesem Nachteil widmet sich das Verfahren DVFS. Es wird bei einer nicht vollausgelasteten Rechenaktivität eine Spannungs- und Frequenzreduktion vorgenommen, sodass die nötigen Berechnungen in dem Umfang verlangsamt werden, dass sie genau dann ihr Ergebnisse liefern, wenn diese benötigt werden. Es werden also alle Rechenoperationen auf Rechenaufwand geprüft und so gestaffelt, dass sie pünktlich aber nicht früher als benötigt zur Verfügung stehen. Diese Art der Berechnung ermöglicht eine langsamere Arbeitsweise und somit eine Energieeinsparung der Recheneinheit, ohne eine merkbare Verlangsamung des Systems zu verursachen. Ein langsames aber kontinuierliches Arbeiten einer Recheneinheit verbraucht dabei deutlich weniger Energie, als eine schnellstmögliche Berechnung und ein anschließendes Warten der Recheneinheit auf weitere Operationen.

2 Dynamic Voltage and Frequency Scaling [Hardwareumsetzung]

2.1 Spannungsanpassung mittels Spannungsinseln

Das DVFS Verfahren kann wie beschrieben durch eine Aufgabenstaffelung softwaretechnisch oder durch das bilden von Spannungsinseln hardwaretechnisch umgesetzt werden. Das Konzept der Spannungsinseln sieht mehrere auf einem Chip befindliche Inseln mit unterschiedlichen Arbeitsspannungen vor. Dabei werden kritische Bereiche, bei denen keine Rechenfehler unterlaufen dürfen, mit einer höheren Spannung versorgt, als andere, nicht so zentrale bzw. relevante Chipregionen. Durch die geringere Spannungsversorgung bestimmter Chipregionen

ergibt sich erneut eine Energieeinsparung. Ein Nachteil ist dabei die hardwaretechnische Herausforderung, die Signale zwischen diesen Spannungseinseln auf die jeweils benötigte Spannung anzupassen.

2.2 Spannungsreduktion mittels Strukturverkleinerung

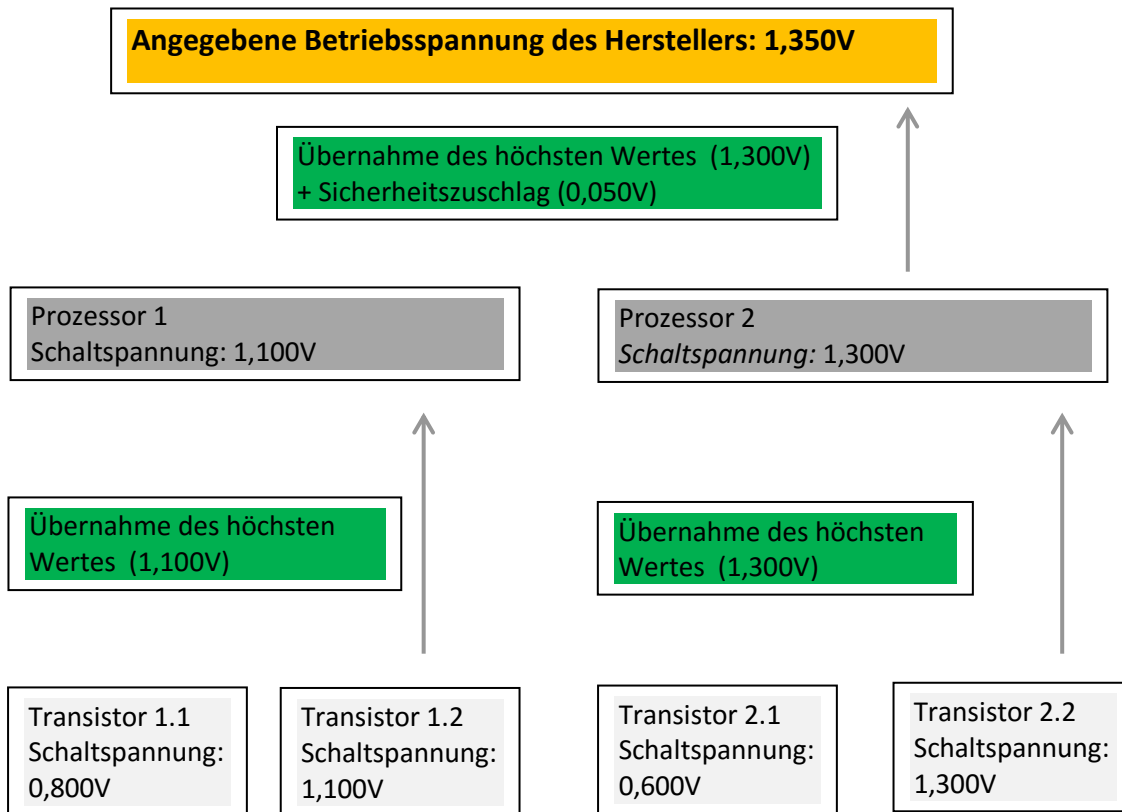
Ein weiterer Trend ist die kontinuierliche Verkleinerung der Chiptransistoren nach dem Mooreschen Gesetz. Dieses trifft zwar allmählich an seine physikalischen Grenzen, bewirkte bisher allerdings eine kontinuierliche Verkleinerung der Transistorkapazitäten.

Aus der oben aufgeführten Formel für den dynamischen Energieverbrauch einer Recheneinheit geht hervor, dass die Reduktion der Transistorkapazität ebenfalls eine lineare Reduktion des Energieverbrauchs verursacht. Die Verkleinerung der Chiptransistoren verursacht also ebenfalls eine hardwaretechnische Optimierung.

2.3 Herstellungsbedingter Spielraum für Spannungsreduktion

Prozessoren arbeiten oft auf einer höheren Spannung, als sie Fertigungsbedingt müssten. Der Grund dafür liegt bei den Schaltspannungen der Transistoren, die innerhalb eines Prozessors eine gewisse Streuung aufweisen. Das bedeutet einige Transistoren schalten schon bei geringeren Spannungen als andere. Die für einen Prozessor benötigte Spannung orientiert sich also an der Spannung des Transistors, der die höchste Spannung benötigt, wobei auf diese noch ein Sicherheitswert aufgeschlagen wird. Die benötigten Spannungen für die Schaltvorgänge variieren allerdings nicht nur innerhalb eines Prozessors, sondern auch unterhalb der Prozessoren einer Serie. Da der Hersteller die Funktionalität aller Prozessoren gewährleisten will wird die Spannungsreferenz nicht anhand des Transistors mit der meistbenötigten Spannung eines Prozessors sondern anhand des Transistors mit der meistbenötigten Spannung aller Prozessoren gewählt. Berücksichtigt man erneut einen Sicherheitszuschlag, ergibt sich unter Umständen eine vom Hersteller angegebene Betriebsspannung, die weit über der tatsächlich benötigten Betriebsspannung liegen könnte.

Anhand eines Beispiels soll dieser Sachverhalt näher verdeutlicht werden:
(Das angegebene Beispiel ist fiktiv und erdacht!)



Das Beispiel beschränkt sich auf die Betrachtung zweier Prozessoren mit jeweils zwei Transistoren. In der Realität müssten jedoch alle Transistoren eines Prozessors und alle Prozessoren einer Serie für die Festlegung der allgemein gültigen Betriebsspannung betrachtet werden. Der Hersteller würde bei diesem Beispiel also eine Betriebsspannung von 1,350V angeben, weil bei dieser Spannung eine Funktionalität aller Prozessoren gegeben wäre. Wäre man nun allerdings der Verwender des ersten Prozessors würde man ohne den Sicherheitszuschlag lediglich eine Betriebsspannung von 1,100V anstelle der vom Hersteller angegebenen 1,350V benötigen. Hieraus ergibt sich ein Spielraum für Energieeinsparungen, die bei standartmäßiger Nutzung des Prozessors nicht ausgeschöpft würden. Hier setzt das Konzept des sogenannten „Undervoltings“ an, bei der dem Prozessor gerade so viel Spannung zugeführt wird, dass er noch ohne Rechenfehler operieren kann. Dadurch fließt weniger Strom, es wird weniger Energie verbraucht, ein geringer dimensioniertes Kühlsystem wird benötigt und dadurch insgesamt Platz sowie Geld eingespart.

3 Anwendungsbereich und Zusammenfassung

Allgemein kann man sagen, dass das DVFS Verfahren in sämtlichen mobilen Geräten, wie Smartphones, Laptops, Tablets etc. sowie lokalen Systemen, wie Standrechnern, Servern etc. Anwendung findet. Besonders wichtig ist dieses Verfahren für Systeme mit limitierter Energieversorgung (z.B. Akku) und/ oder Platzverfügbarkeit (kleineres Kühlsystem). Zumeist wird ein Rechensystem größer dimensioniert als benötigt und nur auf einem Teil der Leistung betrieben, da sich dies als kostenoptimaler herausgestellt hat, als ein Rechensystem mit geringerer Leistung auf Volllast laufen zu lassen.

So findet Dynamic Voltage and Frequency Scaling eine breite Anwendung von Verbrauchersystem bis hin zu kommerziellen Anwendungsbereichen in größeren Maßstäben. Diese Energieeinsparungsmethode ist somit eine der elementarsten Grundlagen für die heute erreichte Effizienz von Recheneinheiten.

Bibliography

(28.07.16)

- http://www.ijeee-apm.com/Uploads/Media/Journal/20150617165534_ECE%20211.pdf
- <http://www.itwissen.info/definition/lexikon/DVFS-dynamic-voltage-and-frequency-scaling.html>
- https://www.academia.edu/4186102/A_survey_of_techniques_for_improving_energy_efficiency_in_embedded_computing_systems
- http://www.eecs.harvard.edu/~meeta/my_papers/hpca08_onchipvrms.pdf
- "Power-efficient System Design", Preeti Ranjan Panda, Springer 2010