WHAT ACTUALLY IS TDP, AND WHY IS IT IMPORTANT?

For electronic devices to function, a passage of electric current is required. A portion of this energy is dissipated as heat due to inefficiencies in the device, or the nonzero resistances to electric current. For most digital devices (processors, microcontrollers) it is usually assumed that all the power consumption is dissipated as heat [1].

TDP (Thermal Design Power) is the worst-case steady state thermal power (waste heat) dissipated by the device, and should be used for processor thermal solution design targets [2]. If the heat generated is not effectively removed, its temperature will rise. This rise in temperature may lead to temperature dependent mechanical and electrical failures such as die cracking due to CTE mismatch and thermal runaway [1].

Processor manufacturers specify a maximum allowable operating temperature, above which they do not guarantee its intended performance or expected lifetime [1]. Processors may throttle themselves to stay under a target temperature or even initiate an immediate shutdown if temperature exceeds certain thresholds [3].

TDP is the maximum thermal power the processor will dissipate, but not the same as the maximum power the processor can consume. It is possible for the processor to consume more than the TDP power for a short period of time that isn't "thermally significant" [2]. Intel® Turbo Boost technology, opportunistically, and automatically, allows the processor to run faster than the marked frequency if the part is operating below power, temperature and current limits [4]. TDP and core frequency can be measured with software applications such as Core Temp®.

THE PHYSICS BEHIND TDP IN PROCESSOR TECHNOLOGY

Modern processors are built with complementary metal-oxide semiconductor (CMOS) technology (Figure 1). CMOS technology provides two types of transistors (also called devices): an n-type transistor (nMOS) and a p-type transistor (pMOS) [5]. Transistor operation is controlled by electric fields, so the devices are also called Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) or simply FETs [5]. Modern CMOS processor circuits have nanometre sized devices (see Figure 2) and can host billions of FETs.

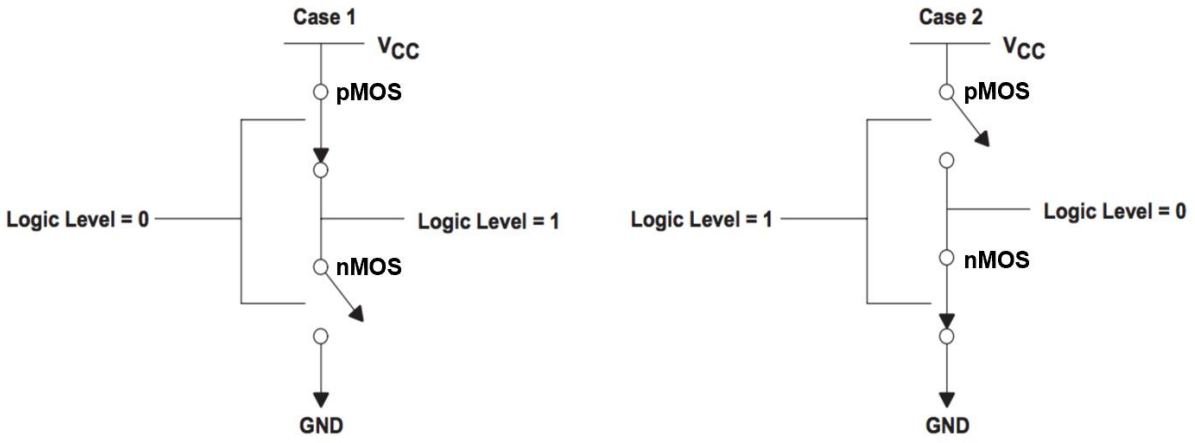

FIGURE 1 - CMOS INVERTER MODE FOR STATIC POWER CONSUMPTION [6]

The power dissipation in a CMOS integrated circuit is both dynamic and static. Dynamic (active) power dissipation is a combination of switching and short circuit power dissipation. Static (stand-by) power dissipation is the continuous power drawn even when the system is not actively switching. Static power dissipation is a combination of subthreshold, gate-oxide, junction and contention leakage [5].

The sum of the dynamic and static power components is the total power dissipated P_T:

$$P_T = \sum (P_{SW}, P_{SC}, P_L) \qquad (1)$$

Where,
P_{SW} = switching power dissipation
P_{SC} = short circuit power dissipation
P_L = Leakage power dissipation

The switching power dissipation, P_{SW}, is due to the charging and discharging of the capacitive loads of the devices while the gates switch state between low (0) and high (1) logic levels, and is determined by equation (2) below:

$$P_{SW} = \alpha C {V_{DD}}^2 f \qquad (2)$$

The activity factor, α, is the probability that the circuit node switches state because most gates do not switch every clock cycle. CMOS logic has been empirically determined to have activity factors close to 0.1 [5]. The accumulative capacitance of the devices is represented as *C*, and V_{DD} is the supply voltage [5].

The short-circuit power dissipation, P_{SC}, is caused by power rushing from V_{DD} to *GND* when both the pullup and pulldown networks are partially ON during state transitioning (Figure 1). This is normally less than 10% of the whole, so it can be conservatively estimated by adding 10% to the switching power.

P_L is the leakage power dissipation determined by devices in the circuit, and is represented by equation (3) below:

$$P_L = V_{DD} I_L \qquad (3)$$

V_{DD} is the supply voltage and I_L is the accumulative leakage current [5].

Until mid-2000's, P_{SW} has been the most significant source of power consumption [7]. Moore's law predicted that the number of transistors on a chip would double roughly every year (in 1975, revised to every 2 years) [8]. He stated that this rate of growth was the product of transistor size shrinkage, die size increase and more efficient use of silicon wafer space [9].

The consequences of transistor shrinkage (Dennard scaling) were first laid out in a paper written in 1974 by Robert H. Dennard of IBM. These changes are represented in terms of the dimensionless scaling factor κ. He postulated the following:

a) The resistance of each device (R) is unchanged by scaling [10].
b) Due to the reduction in dimensions, all circuit elements (i.e., interconnection lines as well as devices) will have their capacitances (C) reduced by a factor of κ [10].
c) The power density of the devices remains constant [10].

$$\tau = RC \qquad (4)$$

This allowed processor manufacturers to raise clock frequencies from one generation to the next without significantly increasing overall circuit power consumption. This held true from 1970's-1990's (see "Dennard scaling" Figure 3). However, looking back on the last 40 years of data, there were some inaccuracies with Dennard's postulations.

Dennard's scaling law assumed that MOSFET threshold voltage would scale along with operating voltage. This means subthreshold leakage was not considered, which has since emerged as a major contributor to static power dissipation (Figure 2). Another key assumption in Dennard's scaling law was the ability to scale gate-oxide thickness. The SiO_2 gate-oxide dielectric is only about 5 silicon atomic layers thick and represents what is likely the limit to which SiO2 can be scaled. This has resulted in increasing gate-oxide leakage power dissipation (Figure 2) [11].

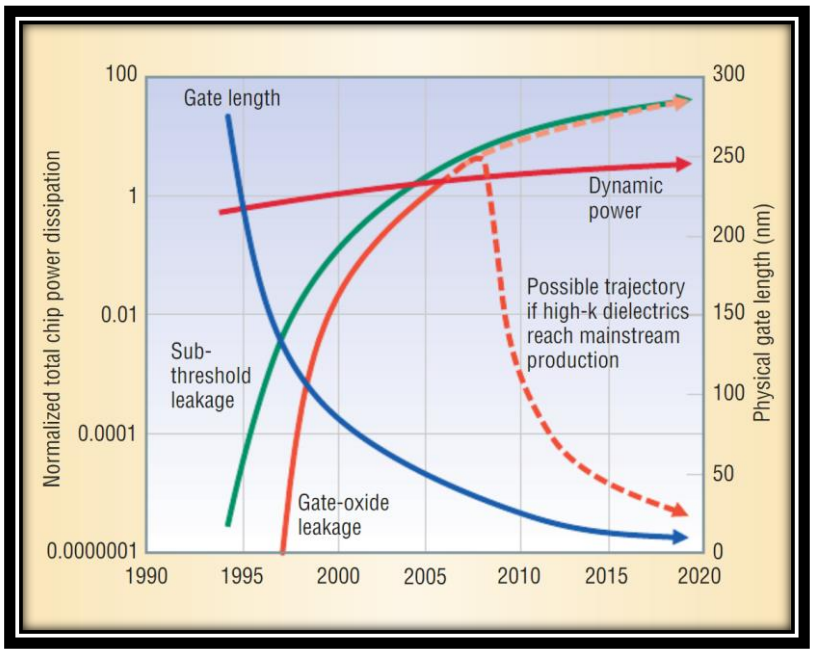

FIGURE 2 - TOTAL CHIP DYNAMIC AND STATIC POWER DISSIPATION TRENDS ITRS [7]

By about 2004, leakage power dissipation reached a level where further reduction in voltage was no longer feasible and power density reached unsustainable levels. This was referred to as the "Power Wall" (see Figure 3) [12]. The industry's response was using multi-core processors to increase performance. As transistor sizes are still decreasing these power constraints have now led to the concept of "dark silicon" i.e., silicon that is underutilized. In CPUs using 22nm technology, about 20% is dark [12].

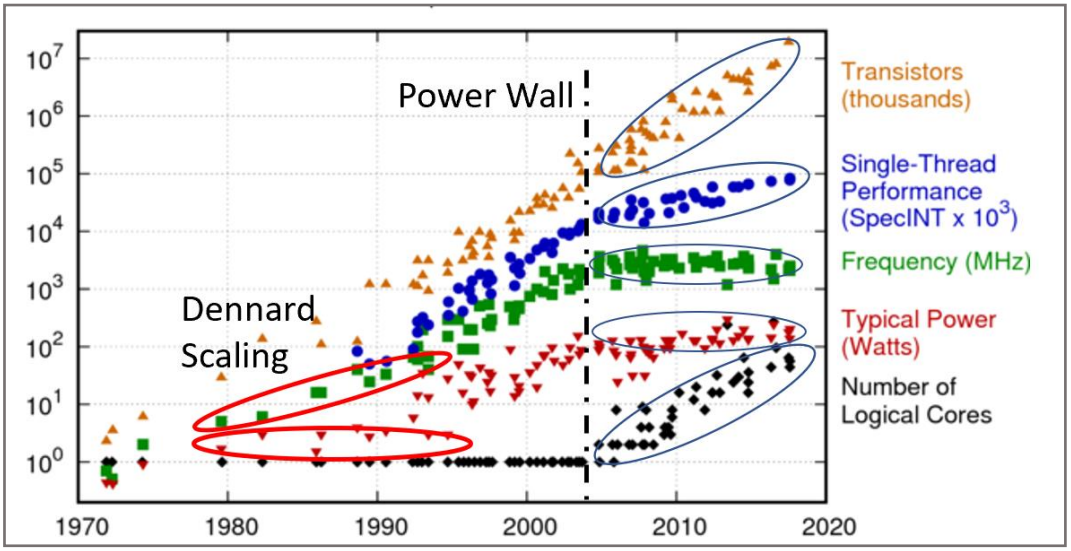

FIGURE 3 - 42 YEARS OF MICRO PROCESSING DATA [13]

REFERENCES

[1] Y. Shabany, Heat Transfer: Thermal Management of Electronics, CRC Press, 2009.

[2] Huck, Scott, “White Paper Measuring Processor Power: TDP vs ACP,” Intel Corporation, 2011.

[3] I. S. W. S. Corey Gough, CPU Power Management, APRESS, 2015.

[4] Division, Intel Enterprise Platforms and Services, “Intel® Server Board and System Products Update on Intel® Turbo Boost Technology Support with Low Power Intel® Xeon® Processor 3400/5500/5600 Serie,” Intel Corporation, 2010.

[5] D. M. H. Neil H. E. Weste, CMOS VLSI Design: A Circuits and Systems Perspective, Addison-Wesley, 2010.

[6] Texas Instruments Incorporated, “CMOS Power Consumption and Cpd Calculation,” Texas Instruments Incorporated, 1997.

[7] T. A. Nam Sung Kim, “Leakage Current:Moore’s Law MeetsStatic Power,” IEEE Computer Society, 2003.

[8] Intel Corporation, “OVER 6 DECADES OF CONTINUED TRANSISTOR SHRINKAGE, INNOVATION: Intel’s 22 Nanometer Technology Moves the Transistor into the 3rd Dimension,” Intel Coroporation, California, 2011.

[9] D. C. Brock, Understanding Moore's Law: Four Decades of Innovation, Chemical Heritage Foundation, 2006.

[10] M. I. F. H. G. H.-N. Y. M. I. ROBERT H. DENNARD, “Design of Ion-Implanted MOSFET’s with Very Small Physical Dimensions,” *IEEE,* vol. 87, no. No. 4, p. 676, 1999.

[11] M. Bohr, “A 30 year Retrospective on Dennard's MOSFET Scaling Paper,” *IEEE Solid-State Circuits Newsletter ,* vol. 12, pp. 11-13, 2007.

[12] Z. E. B. K. H. A. Zeinab Seifoori, “Introduction to Emerging SRAM-Based FPGA Architectures in Dark Silicon Era,” *Advances in Computers,* vol. 110, pp. 259-294, 2018.

[13] F. O. S. K. O. L. H. C. B. K. M.Horowitz, “42 years of micro processing data,” 2017. [Online]. Available: https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/.