AI.2. Passive Components in AI Systems

Slavomir Pala,
Kyocera-AVX, Lanskroun, Czech Republic, slavomir.pala@kyocera-avx.com

Ron Demcko,
Kyocera-AVX, Fountain Inn, S.C., USA, ron.demcko@kyocera-avx.com

ABSTRACT

Capacitors play multiple critical roles in the world of AI systems ranging from improving power integrity, EMI control and optimizing signal processing. The type of AI system impacts the importance of the capacitors performance and features relative to the overall type of system. For instance, AI systems involved in deep learning or machine learning have different power needs and processing intensity when compared to Edge AI. In high-power/high-performance AI systems, low profile sizes are becoming preferred (due to liquid cooling layout constraints) along with high capacitance values becoming critical. In Edge AI the needs tend towards smaller field deployed boards that will require small physical size capacitors with low loss, and broad-spectrum responses. This paper is aimed at providing a high-level overview into the types of capacitors various AI platforms might use and the performance features of the different capacitor types.

The Evolving State of AI

AI performance has improved dramatically over the past few years driven by advancements in large language models (LLMs), generative AI evolution and increased efficiency architectures[1]. Those software and architectural advancements were made possible through progress made in increased computational power within semiconductors and simplified scale up of hardware.

Improvements in semiconductor performance were achieved in GPUs, TPUs and the emergence of specialized AI accelerators. AI accelerators such as Amazon Trainium2 offer ~ 4x computational throughput at a power reduction of better than 50%.

Accelerators are tailored for specific tasks to reduce processing loads on other hardware. In this scenario, they can potentially reduce energy use and improve overall system performance. The accelerator trends indicate a concentration of work on reduced energy consumption of the system through efficient processing load sharing and partitioning. Power consumption is a key consideration at every stage within a design process given the immense increase of power used in a general server (~5000W) to that of a next generation AI server (~50,000W). Exact power consumption numbers vary but it's safe to say that AI servers use 5 or 10x more power.

Regardless of the specific IC type, the semiconductor scaling process and the associated reduced line widths are essential in creating highly complex, high gate count ICs needed in the AI arena. However, scaling impacts IC supply voltage ratings, allowable ripple voltage on power rails and di/dt demands. The reduced supply voltages require higher currents needed to achieve power levels demanded by high gate count ICs. Those higher power ICs also are clocked at higher frequencies and in turn higher di/dt draws exist on the supply rails resulting in a complex power integrity scenario[2].

An example of reduced power supply voltage as a function of semiconductor process line widths is shown in Figure 1. The typical allowable ripple voltage by supply voltage level is of great importance to be mindful of in design. As this graph indicates the ripple voltages must be controlled to tighter limits as the ICs supply voltage drops. This implies an ever more complex power integrity scenario for the end system given that larger acreage of Silicon gates create higher di/dt loads on the power supply. High frequency decoupling capacitors and bulk capacitors play an important role in stabilization of power supplies and achieving ripple voltage targets required by high-performance ICs.

The trend of AI servers using finer geometry features suggests a further tightening of the allowable ripple voltage requirements to the IC. Further, larger numbers of high power, low voltage ICs networked reinforces the general thought that AI servers will use about 10x the number of capacitors to meet the power integrity goals required in advanced AI chipsets[3].

Though the ICs themselves will operate at lower voltages the power incoming to those rails has progressed upwards from 12 volts to mostly 48 volts due to the advantages high voltages give in power conversion efficiency and reduced copper cabling costs. An ever more common approach is to perform power conversion as close to the load as possible thus achieving cost reductions through smaller traces/wire/cabling and reducing I^2 R losses.

As power demands continue to increase, designers begin to consider GaN based semiconductors for more efficient power conversion. GaN could provide higher conversion efficiency & therefore generate less heat. Heat is a critical concern as power per moderately powerful AI racks approach 90kW[4].

The use of GaN in power conversion would enable power conversion switching frequencies of up to 2 MHz and would place increased importance on the parasitic inductance and ESR values in bulk capacitors.

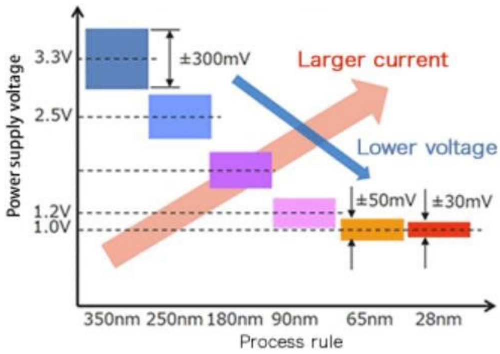

Figure 1, Trend in IC Supply Voltage by line width (Source: techweb.rohm)

CAPACITOR USE & OVERVIEW

Capacitor use within AI have many roles to play regardless of the specific subsystem. A high-level example of the capacitor type and capacitance value used within a specific AI block is shown in Figure 2. This table breaks out the use and roles of capacitors along with common capacitance ranges used.The technology used follows conventional design thinking of bulk capacitors used for large amounts of charge retention and supply that 'feed' high frequency decoupling capacitors.

As the gates and speeds of switching of ICs increase, the overall capacitance needed to supply the di/dt increases, as not to experience Vdroop on the IC. Capacitors are placed as closely to the power sensitive loads in order to maintain maximum efficiency. Close capacitor proximity to the load reduces the inductive loop of the capacitor and traces relative to the load (IC).

Additionally, low inductance capacitors are also employed and placed close to the load to provide even higher levels of power quality.

To this point in time, low inductance capacitors were exclusively based upon Multi-Layer Ceramic Capacitor technology (MLCC). The three types of low inductance ceramic capacitor are:

1) Reverse geometry MLCCs – capacitors with two terminations, one each on opposing long sides of the MLCC.
2) Interdigitated Capacitors – capacitors with alternate termination connections which can be placed on all 4 sides of the MLCC package. Three terminal capacitors are a sub family of Interdigitated Capacitors.
3) Land Grid Array – a capacitor with terminations on the longer side of the MLCC with electrodes arranged vertical to the PCB.

Application	MLCC	Aluminum Electrolytic	Tantalum Polymer	Typical Range in uF
VRM, Power Delivery	x	x	x	0.1 - 2200
GPU / CPU	x		x	0.01- 680
Memory Modules	x			0.1 - 47
PSU Bulk Power		x	x	470 - 10,000
NIC/ SSD	x		x	0.1 - 470

Figure 2, Application, capacitor technology and typical range

We can generalize and state that high frequency decoupling is usually approached by using low inductance capacitors placed as close to the **IC** load as possible with values between 10nF and 2.2uF. Mid frequency decoupling is addressed with capacitance values of 1uF to 10uF. Bulk capacitors ranging from 100uF to 470uF[5] are commonly used to achieve power supply stabilization and are placed as close to the **regulators** as possible. Capacitor proximity to the IC load or regulator is critical for optimized charge delivery response.

There are multiple locations that MLCCs can placed to reduce the distance from the load and therefore drop overall loop inductance.

These areas can be grouped into three areas[6]:

- Main board capacitor – designated by M
- Die or Land side capacitor – designated by D / L
- Embedded capacitor – designated by E

Advances in MLCC case size reduction make moving capacitors closer to the load more practical given the progress that high C/V ceramics have made. An example of capacitor placement options is shown in Figure 3.
Currently, small case size MLCCs and thin designed MLCCs meet embedded applications target thicknesses of 300 µm, 220 µm, and 150 µm. These MLCCs can easily be used in the tightest of dimensional constraint application.

However, MLCCs can exhibit capacitance instability under DC & AC bias as well as variations in capacitance due to capacitor age and temperature. Those instabilities[7] though known and accepted in the industry, must be taken into account as designers create power tree minimum capacitance calculations. For example, it could not be uncommon for a MLCC to lose half of its nominal value once in actual use. Given that core voltages are so low (Figure 1), MLCCs are the best choice for high frequency capacitors. Low voltage rail applications potentially experience smaller voltage bias effects due to sub 3.3V bias.

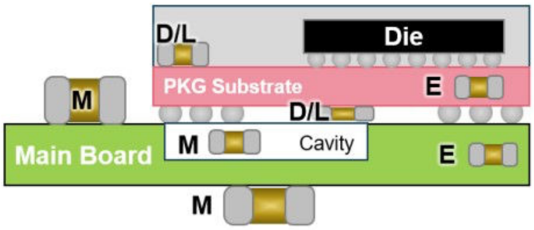

Figure 3, MLCC placement options relative to processing core

TANTALUM POLYMER BULK CAPACITORS

There are many choices for bulk capacitors ranging from various configurations of Aluminum Electrolytic capacitors to Tantalum & Tantalum Polymer capacitors. Tantalum polymer capacitors are a near ideal choice for bulk capacitors due to their capacitance density, multiple small physical sizes, excellent reliability and electrical performance.

A key feature of tantalum conductive polymer capacitors is improved ESR over traditional MnO_2 tantalum capacitors[8]. Tantalum Polymers exhibit approximately one-eighth the ESR of standard tantalum devices and therefore can handle about 8 times the current of MnO_2 devices. Tantalum Polymers also exhibit improved energy density. This feature creates a flexibility for designers to find reasonably sized bulk capacitors in small case sizes with reduced component height options. Typically, case size dimensions can range from 0402 to 2924 and heights can be as low as 0.50 mm. The small case size allows close placement of the capacitor to either the IC load or the regulator. In fact, tantalum polymers offer such size reductions that the regulator can start being placed closer to the IC load. This helps optimise the power delivery network. The small case sizes also offer an inherently low inductance with standard case size packages ranging from 1nH to 2.5 nH. Tantalum polymer capacitors have the added advantages of:

- Wide capacitance range with low inductance in small-case size, face down/undertab package
- Wide voltage range of 2.5 V to 125 V
- High Surge Robustness
- No DC bias voltage effects
- Capacitance retention at high frequency

- No piezo effect
- High reliability: consumer, commercial, Auto grade, COTS, MIL-PRF grades, Space grades are available - Lower derating required: - use at 90% rated voltage/10% voltage derating for products rated up to 10 V
- use at 80% rated voltage/20% voltage derating for products rated >/= 16 V

NEW REDUCED INDUCTANCE TANTALUM POLYMER CAPACITOR

Given that standard capacitor configurations of Tantalum Polymers offer low ESR performance and acceptable low inductance, an effort was placed on characterizing the inductance of those capacitors and potentially reducing their inductance. A comparison of case inductance between standard lead frame tantalums (figure 4A) is compared to the inductance of multiple tantalum pellets connected in a parallel fashion on modified standard lead frames (figure 4B) and leadless tantalum capacitors (no lead frame shown in figure 4C).
The results of modified or eliminated lead frames indicate that the inductance of tantalum capacitors can be reduced significantly.

Case Size	Typical Self Inductance value (nH)	Case Size	Typical Self Inductance value (nH)	Case Size	Typical Self Inductance value (nH)
A	1.8	H	1.8	U	2.4
B	1.8	K	1.8	V	2.4
C	2.2	N	1.4	W	2.2
D	2.4	P	1.4	X	2.4
E	2.5	R	1.4	Y	2.4
F	2.2	S	1.8	5	2.4

4A - Inductance for Standard lead frame by Case Size

Case Size	Typical Self Inductance value (nH)
A	1.5
B	1.6
D	1.4
E	1.0
H	1.4
I	1.3
J	1.2
K	1.1
L	1.2
M	1.3
R	1.4
T	1.6
U	1.3
V	1.5

4B– Inductance for leadless technology by Case Size

Case Size	Typical Self Inductance value (nH)
D	1.0
E	2.5
U	2.4
V	2.4
Y	1.0

4C – Inductance for Multi-Anode Pellets in a single case size

Figure 4*, Inductance comparison by frame technology*

The reduction of parasitic inductance allows the reduced inductance capacitor to have an extended frequency response within the power distribution network. Further, there are no voltage bias effects which would reduce the actual capacitance in circuit.
MLCC has developed lower inductance designs, and it is expected that Ta-polymer could follow this trend to offer high capacitance with lower ESL to match expected performance needs of low voltage AI cores.

A real-world circuit example of a reduced inductance tantalum capacitors circuit impact is shown in Figure 5. The graph shows a comparison of the noise level on a LTC3315B buck converter standard reference design board. The Gray curve shows the noise level of the board as received. The orange line shows the impact of the low inductance, low ESR Tantalum Polymer capacitor. When using the reduced inductance tantalum polymer, the output noise is significantly reduced across the near 100 MHz spectrum of test. Further reduction of noise improvements is expected when using the low inductance tantalum polymer and an optimised PCB using compatible pad layouts.

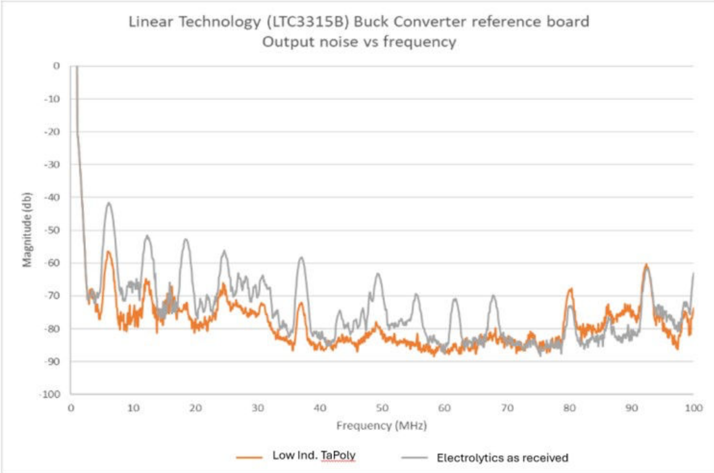

Figure 5, dB noise across 1 to 100 MHz spectrum

SUMMARY AND CONCLUSIONS

AI systems are rapidly evolving into servers, edge applications and more. Regardless of the specific AI configuration, powerful ICs are requiring higher levels of power quality. Capacitors are used to achieve power quality, and the specific configuration of capacitors is typically a cascade of capacitor solutions. High frequency decoupling capacitors are placed as close to the IC load as possibly. Mid frequency decoupling strategically placed around the board to bridge power quality issues between the regulators & the high frequency decoupling capacitors placed around the IC load. Bulk capacitors are typically placed at the PSU & regulators (at a minimum).

Ceramic capacitors dominate high frequency decoupling applications and many mid frequency applications. New reduced inductance tantalum polymer capacitors give designers a way to move large amounts of bulk capacitance close to the regulators. Their small size, large capacitance and low inductance values allow regulators to move closer to the IC loads thereby improving power quality in AI power trees. Further capacitance values are expected to increase greatly in both MLCCs and Tantalum Polymer as technology improves. Those improvements are essential to advanced ICs with reduces voltage rails and higher current transitions.

REFERENCES

[1] Ten Noteworthy AI Research Papers of 2023; Ahead of AI; https://magazine.sebastianraschka.com/p/10-ai-researchpapers-2023

[2] Metal Inductors and Trends in Power Supply's 2018; Tech Web; https://techweb.rohm.com.trend/engineer/4804/

[3] Electronic Components Challenges in AI Power Management; CMSE 2025; https://www.tjgreenllc/cmse/

[4] Can Silicon Supply Enough Power for the Future of AI Silicon?; ELECTRONICS DESIGN 2023; https://www.electronicdesign.com/blogs/the-briefing/article/21278725/electronic-design-can-silicon-supply-enoughpower-for-the-future-of-ai-silicon

[5] Block Diagram: AI Server; https://ele.kyocera.com/en/topics/block_aiserver/

[6] MLCCs Are At The Forefront Of Capacitor Miniaturization; How2Power; May 2025; How2Power.com - Answering Your Questions About Power Design: : Electronics Design Services : Voltage Regulator : Industrial Power Supplies : Semiconductors : Transformer : IGBT Module : Software : Digital Power Supply DC

[7] Class-II MLCC DC-Bias Aging Issues in Automotive Application; August 2022, European Passive Components Institute; Class-II MLCC DC-Bias Aging Issues in Automotive Application

[8] Space-Grade Tantalum Polymer Caps Deliver Lower ESR And Better Current Handling For Bulk Capacitor Applications; How2Power; August 2021; How2Power.com - Answering Your Questions About Power Design: : Electronics Design Services : Voltage Regulator : Industrial Power Supplies : Semiconductors : Transformer : IGBT Module : Software : Digital Power Supply DC