One Dell Way
Round Rock, Texas 78682
www.dell.com

THERMAL DESIGN OF THE DELL™ POWEREDGE™ T610™, R610™, AND R710™ SERVERS

Dominick Lovicott
Enterprise Thermal Engineering

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

For more information, contact Dell.
Information in this document is subject to change without notice.

CONTENTS

1 Introduction

Driven by rising energy cost and environmental initiatives, server and data center power consumption has become one of the foremost concerns for IT professionals and data center managers. Understanding this concern, Dell has been focused on reducing the energy consumption of its servers. Because power and airflow consumed to cool a server can be significant contributors to the overall energy consumption of the data center, Dell engineers designed a cooling solution that minimizes these needs. This is accomplished in the Dell PowerEdge R710, R610, and T610 servers by incorporating sophisticated fan control, strategic component placement, airflow management and power efficient fans. These servers are the 11th generation of Dell servers and include evolutionally improvements in Dell's thermal design and control.

This guide provides insight into the cooling design and thermal management of the PowerEdge R710, R610, and T610 monolithic servers. Please visit www.DELL.com for a more comprehensive overview of these servers.

Table 1.   R710, R610, and T610 Servers Overview

Dell PowerEdge	R610	R710	T610
Front View			
Height	1U	2U	5U
Processors	Intel® Xeon® 5500		
Number of CPU sockets	Dual		
Memory Type	DDR3		
Memory Slots	12	18	12
PCI Slots	2	4	5
Hard drives (Hot Plug)	6 - 2.5" Small form factor	6 - 3.5" Large form factor 8 - 2.5" Small form factor	8 - 3.5" Large form factor 8 - 2.5" Small form factor
Number of Fans	6 - Dual CPUs 5 - Single CPU	5 - Dual CPUs 4 - Single CPU (Hot Plug)	2 - non-redundant 4 - redundant
Power Supplies (Hot Plug)	n+1 redundancy		

2 Supported Server Environments

Cooling typically consumes a significant portion of the overall data center power budget. Because of this, some data centers are adopting higher temperature operation to enhance data center cooling efficiency. These higher ambient (server inlet or supply) temperatures enable energy reductions in the refrigeration process, whether the facility uses Computer Room Air Conditioners (CRACs) or Computer Room Air Handlers (CRAHs). Additional fan power is consumed to support the elevated ambient and is a tradeoff to the cooling power saved at the data center level. Dell has optimized the PowerEdge R610 and R710 monolithic servers for both traditional data center infrastructures (<25 °C) as well as higher ambient applications (≤28 °C).

The IT hardware deployed in high ambient data centers must have a robust thermal design to operate reliably at elevated temperatures. The Dell PowerEdge R710, R610 and T610 servers can operate from an ambient temperature of 10 to 35°C for altitudes up to and including 10,000 ft (3,048 m). Reliable operation across this range is achieved through dynamic fan speed response based on ambient temperature. At higher server inlet temperatures, fan speeds are increased to maintain component temperature requirements. Table 2 includes the supported operating and storage environmental conditions.

Table 2.   Operating and Storage Conditions

CONDITION	METRIC UNITS	STANDARD UNITS	NOTES
Operating Temperature	10° to 35°C	50° to 95°F	For altitudes above 2950 ft the maximum operating temperature is de-rated 1°F/550 ft
Storage Temperature	-40° to 65°C	-40° to 149°F	-
Operating Altitude	-16 to 3048 m	-50 to 10,000 ft	For altitudes above 2950 ft the maximum operating temperature is de-rated 1°F/550 ft
Storage Altitude	-16 to 10,600 m	-50 to 35,000 ft	-
Operating Relative Humidity	20% to 80% (non-condensing)		Maximum humidity gradiation of 10% per hour
Storage Relative Humidity	5% to 95% (non-condensing)		Maximum humidity gradiation of 10% per hour

3 Hardware Design

Board layout, heat sinks, shrouding, chassis venting and fan development are examples of server hardware that incorporate thermal design elements aimed at reducing power and airflow consumption. This section highlights some of the building blocks used in the latest generation of PowerEdge cooling solutions.

Motherboard layout and component placement can play a significant role in the thermal characteristics and cooling requirements of a server. Component placement for the R610 and R710 motherboards were chosen specifically to separate the processor and memory cooling paths. This parallel cooling path for memory reduces the airflow required to keep components within their operating limits. When components are placed in series the upstream component preheats the air for the downstream component requiring more airflow to cool the downstream component. The following illustration shows the separate airflow paths for the R710 memory and processors.
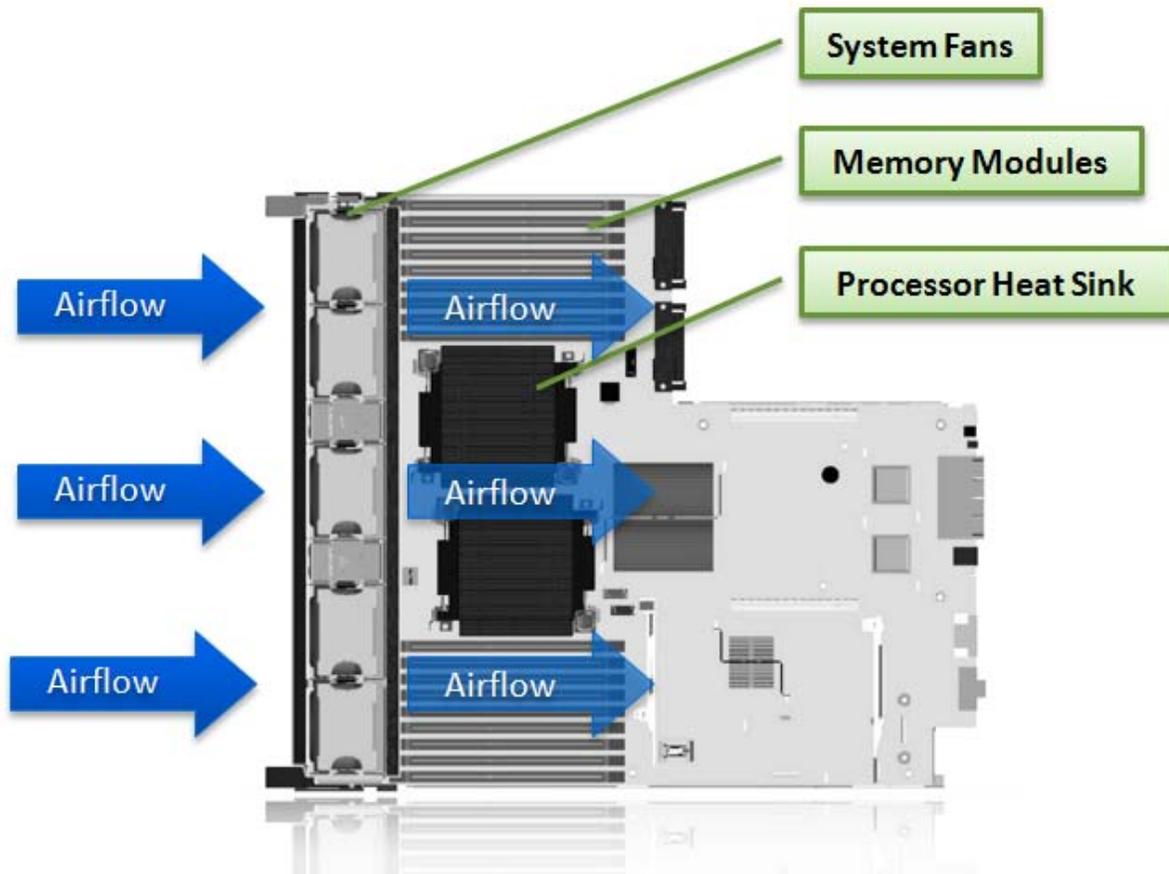

Figure 1.   R710 Airflow Paths

3.1 Airflow Management

Dell thermal solution uses airflow shrouds, baffling, sealing, and fan placement for the important act of efficiently directing airflow over hard to cool components and heat sinks. In sub-optimized systems, unmanaged airflow bypasses critical components, requiring higher fan speeds to compensate.

When the servers are racked with another system populated above it, there is typically a small gap in-between systems. The top of the R610 and R710 systems include additional inlet and exit venting that allows for more airflow when there is space above the chassis. The combination of this ventilation and closed loop fan control reduces the fan speeds required to cool when there is room for airflow above the system. This ventilation can be fully occluded without impact to system reliability.

Dell engineers also manage the airflow within a server with fan population. For a single CPU configuration, a fan in the R610 and R710 chassis can be removed because the depopulated CPU and memory slots no longer need the airflow. The R610 chassis comes populated with six fans for a system populated with dual processors with only five fans populated for a single processor configuration. Similarly, the R710 server has five fans populated for a dual processor configuration and four fans populated for a single processor configuration as shown in Figure 2.

PowerEdge R610		PowerEdge R710	
	5 System Fans		4 System Fans
Single CPU	CPU-Memory Blank — Fan Blank		CPU Blank — Fan Blank
	6 System Fans		5 System Fans
Dual CPU	CPU Heat Sinks — System Fans		CPU Heat Sinks — System Fans
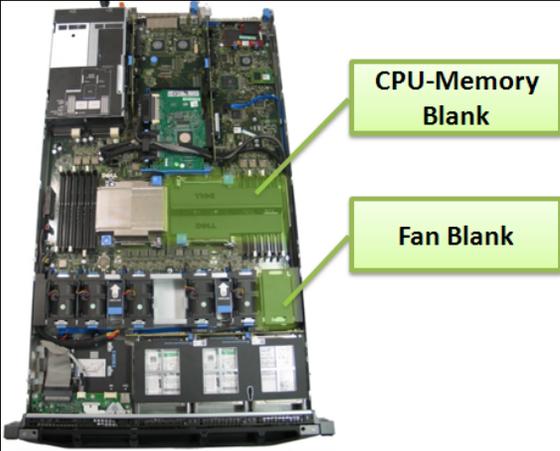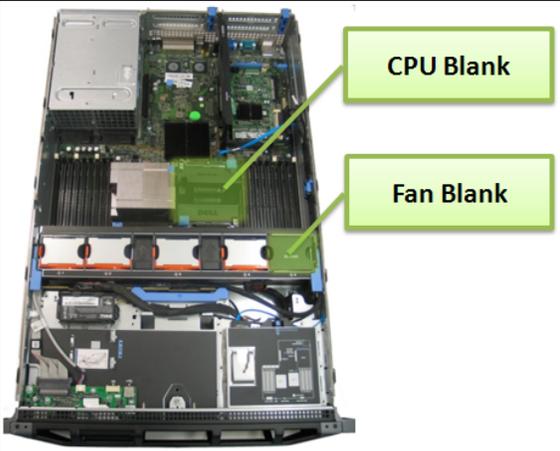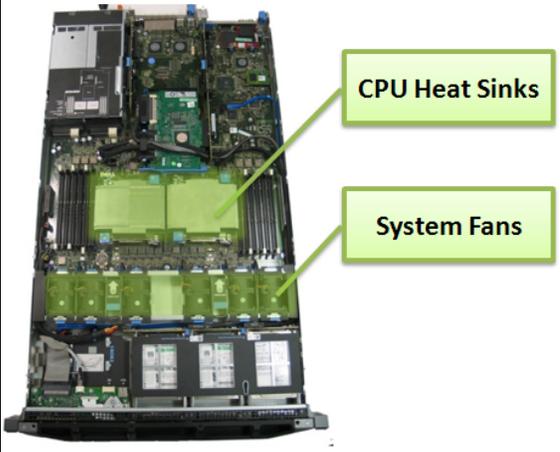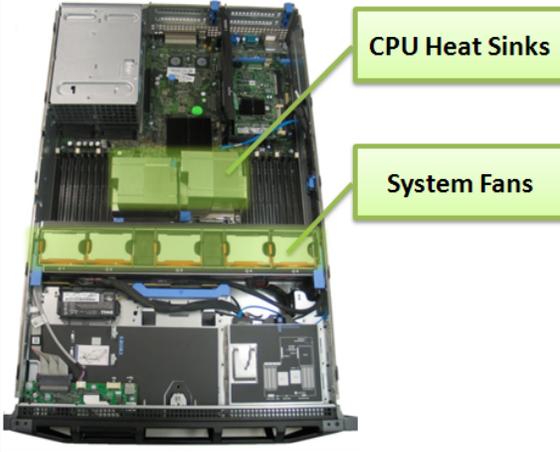
Figure 2.   R710 Fan Population

Unlike the R610 and R710, the T610 does not reduce the number of fans based on processor configuration due to its board layout. Instead, the T610 offers redundant and non-redundant cooling options. The T610 can be upgraded with an addition of two fans for cooling redundancy as shown in Figure 3. The R610 and R710 systems come with standard redundancy independent of CPU configuration.
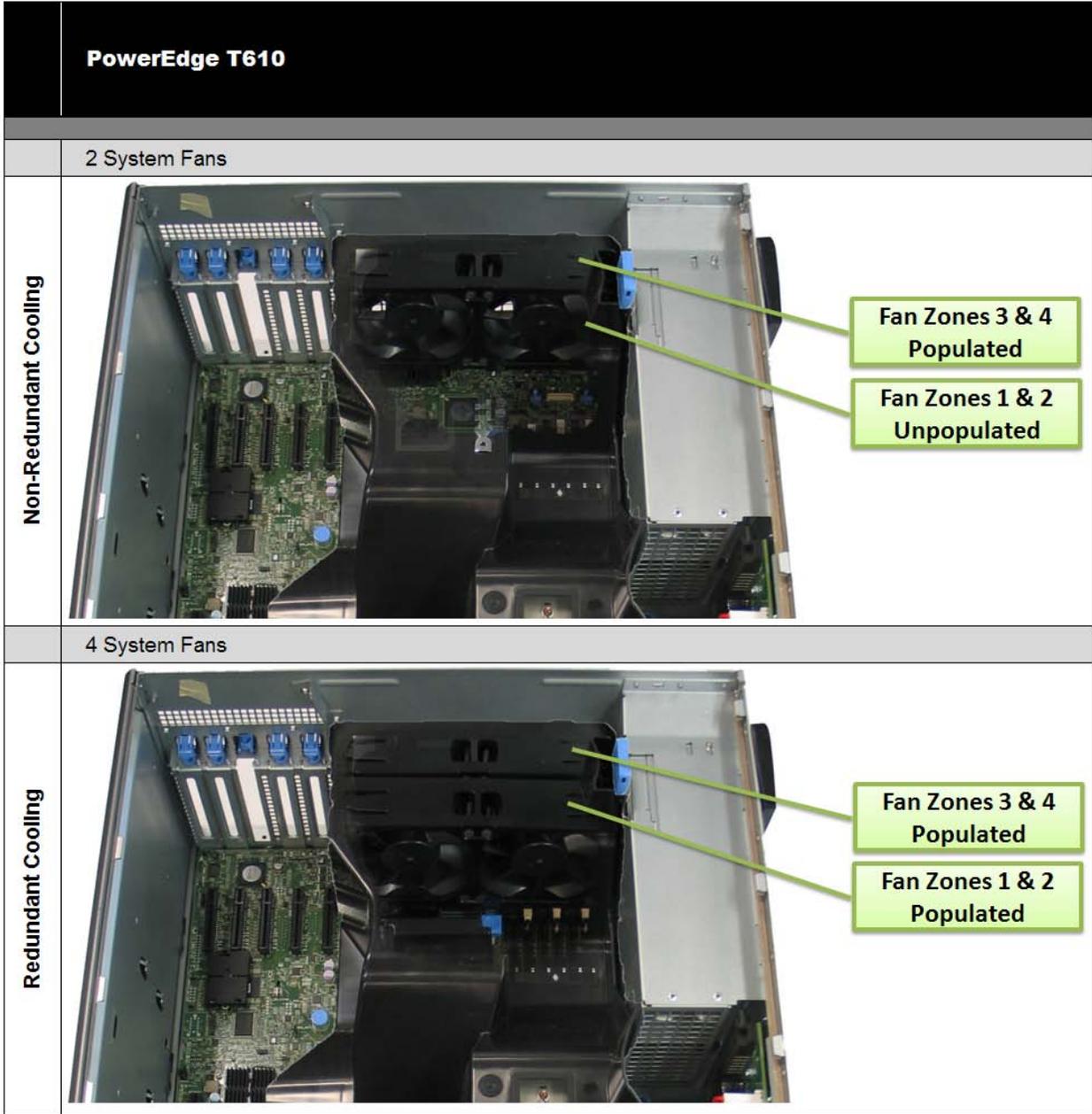
Figure 3. T610 Fan Population

3.2 Chassis Airflow and Fan Power Consumption

The airflow a fan delivers through a server is contingent upon how restrictive a server is to air movement. Airflow bottlenecks occur at high density areas of the server. The Dell PowerEdge T610, R610 and R710 servers minimize the airflow resistance by maximizing chassis inlet and exhaust perforations and optimizing the fan guards to be less restrictive.

These reductions in the airflow resistance of the server chassis allow the fan to produce more airflow movement. As shown in Figure 4, the amount of airflow delivered by a fan installed in a system is indicated by the operating point and the amount of airflow produced by a fan against no resistance is indicated by the free air operating point. Fan vendors tend to specify the free air

operating point for off-the-shelf fans and servers but using these fans without consideration of the system impedance results in a less optimized design that produces less airflow than expected. The PowerEdge R610, R710 and T610 fans were developed based on the measured chassis impedance of the respective system to achieve optimal fan performance.
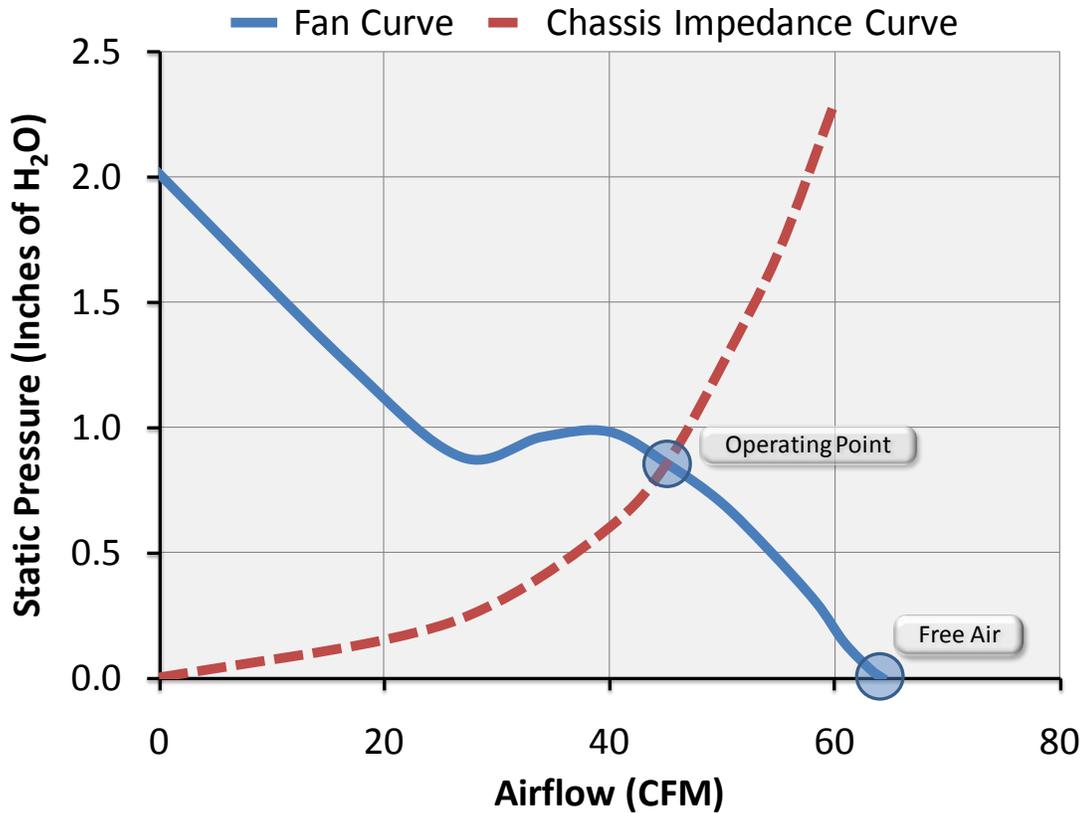

Figure 4.   Amount of Air Delivered

When developing energy efficient fans, considering fan/system airflow interactions is only half the story. Fan efficiency can be defined by the amount of electrical power consumed to move airflow through the server. Low-efficiency fans require more electricity to produce less airflow. Conversely, high efficiency fans consume less electrical power to move the same or more airflow. Fan efficiency also plays into the amount of necessary heat removed by the data center infrastructure. Energy not converted into air movement is generally dissipated as heat, vibration and noise. Reducing the amount of power consumed by the fan not only reduces the system power draw, but also reduces the load on data center cooling.

Fans can consume a significant portion of system power. For this reason, Dell has been focused on increasing fan efficiency. The generational improvements in the Dell 1U and 2U fan efficiencies are provided in Figure 5.
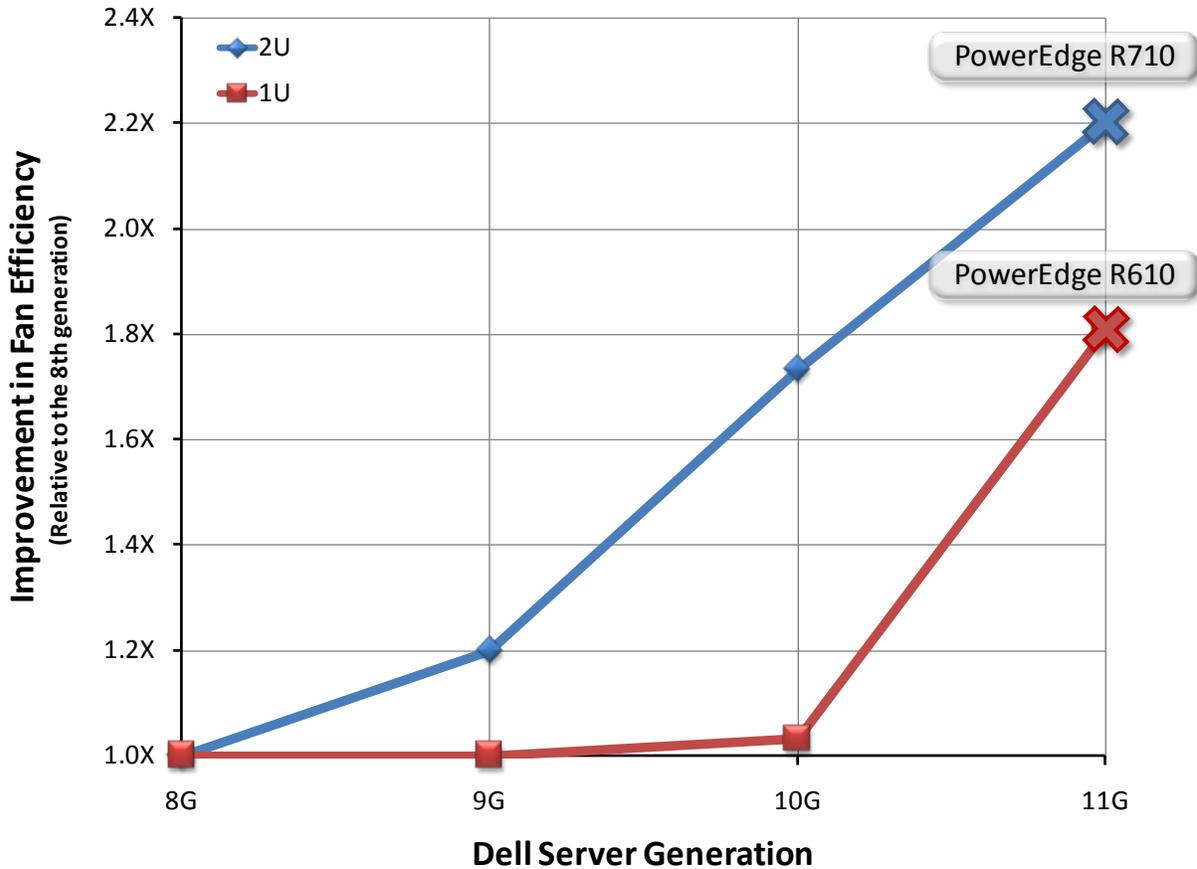
Figure 5.   Generational Improvements in Dell 1Ua2U Fans

The variable speed fans deployed in PowerEdge T610, R610, and R710 servers operate at low speeds to reduce power consumption. Figure 6 shows how fan power increases with fan speed at a cubic rate for a single versus five fans. The R610 fans have the additional power-saving capability to turn off when not needed.
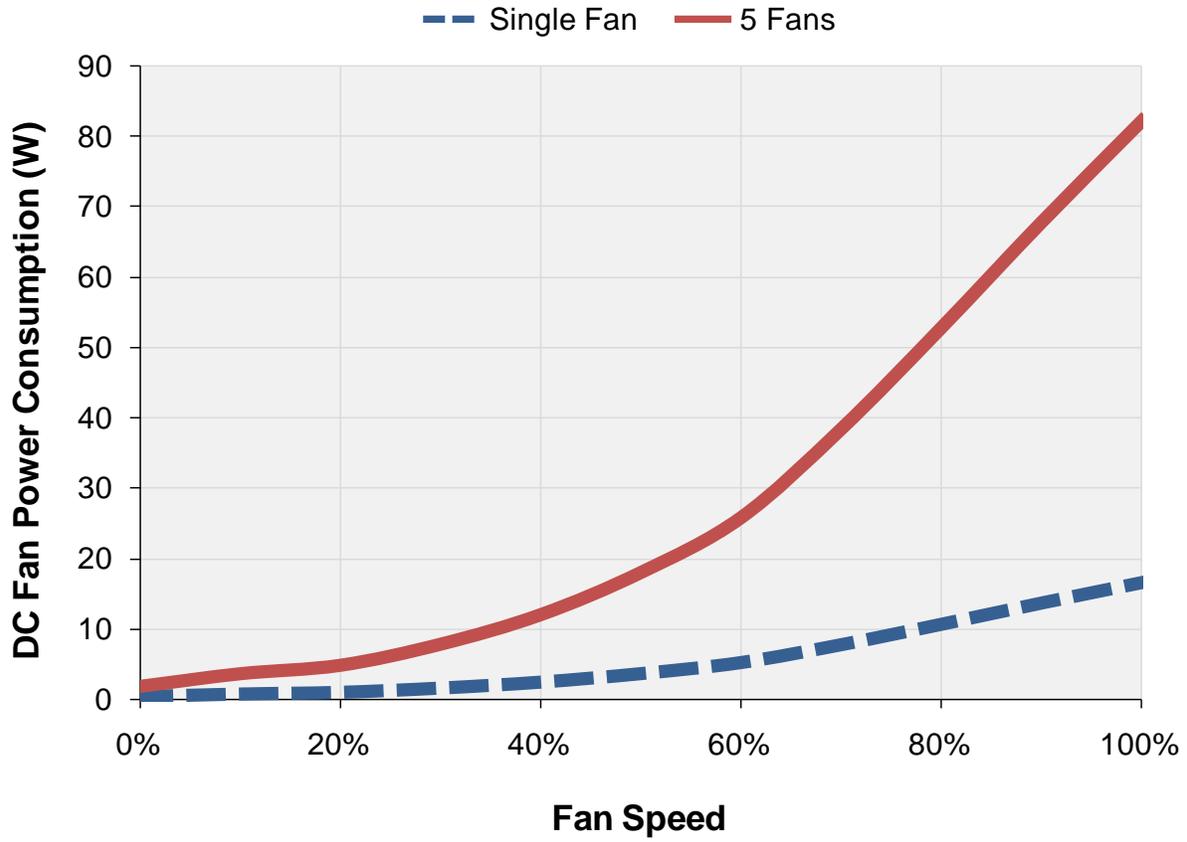
Figure 6.   Single Fan vs. Five Fans Power Rate Difference

4 iDRAC Thermal Management and Fan Control

The Dell PowerEdge R710, R610, and T610 fan control and thermal management reside on the integrated Dell Remote Access Controller (iDRAC), which ensures that the appropriate fan speeds create an airflow to properly cool the system. The iDRAC thermal management and fan control utilize multiple static and dynamic inputs to determine the appropriate output responses as illustrated in Figure 7. The iDRAC uses a pulse width modulated (PWM) signal to communicate the desired speed to the fan.
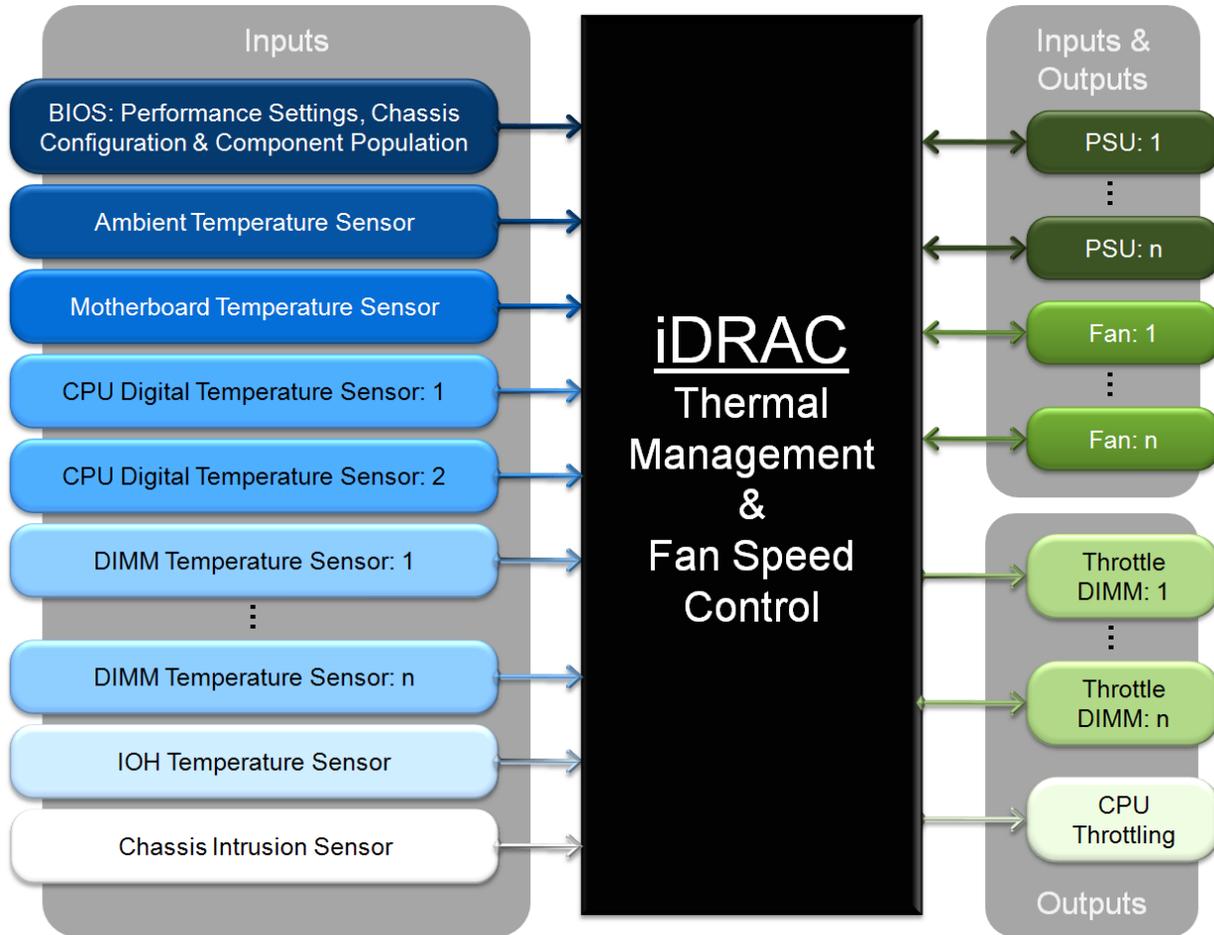

Figure 7.   Output Responses

4.1 PSU Thermal Management and Fan Control

The Power Supply Unit (PSU) has integrated intelligence that allows it to communicate with the iDRAC as well as determine its own cooling requirements based on a multitude of PSU temperature sensors. The inputs and outputs of the PSU thermal management are illustrated in the Figure 8.
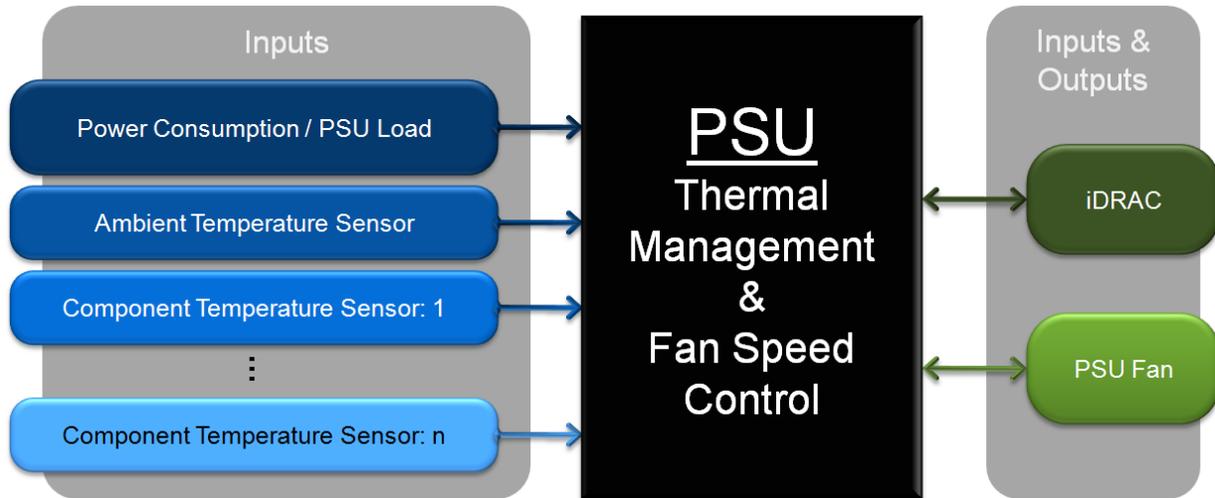
Figure 8.   Inputs and Outputs of PSU Thermal Management

4.2 Fan Zones

In the Dell PowerEdge R710, R610 and T610 servers, fans are separated into individual fan zones. The iDRAC saves fan power and airflow consumption by mapping these fan zones to specific components and sensors to provide airflow when and where it's needed. This allows only fans coupled to a component to react to component cooling requirements, which allows targeted cooling reducing fan power and system airflow consumption.

4.3 Configuration Based Thermal Management

The number, type, capacity, and speed of components are collected by the BIOS and reported to the iDRAC, which uses this information to determine the minimum fan speeds required for a given ambient temperature. Without this information, the fans would need to run at speeds to cool a worst-case maximum configuration even for lightly configured systems. This penalizes most configurations with higher airflow and power consumption.

4.4 Processor Thermal Management

The Dell PowerEdge R710, R610 and T610 servers support Intel® Xeon® 5500 processors with Thermal Design Powers (TDP) from 65W to 95W. Each CPU has a specific temperature threshold called Tcontrol (defined by Intel®) that must be maintained to guarantee reliability and performance. The iDRAC thermal management reads the processor's Digital Temperature Sensor (DTS) through the Platform Environmental Control Interface (PECI) bus and the Tcontrol temperature threshold. The Dell fan control utilizes the CPU DTS and Tcontrol values to maintain proper cooling of the processor.

The Dell PowerEdge R710, R610 and T610 offer user-configurable BIOS options that alter the target temperature for CPU cooling to take advantage of higher performance for cooler CPU temperatures or fan power savings realized by letting CPUs operate at higher temperatures. This is achieved for the performance optimized BIOS mode by subcooling the CPU below the Tcontrol threshold, and cooling the CPU to the threshold for the power optimized BIOS mode.

Dell has incorporated PID control into the iDRAC for CPU cooling to optimize the thermal control. PID is an advanced control algorithm that utilizes predictive calculations based on a measured process value compared against a target value to determine the appropriate response. The iDRAC reads the temperature of the processor and the PID algorithm uses it as the process value to determine the appropriate fan response. The following sub-components of PID control work with one another to determine the optimal fan speed:

Proportional control responds to the current temperature based on its relation to the target temperature. When the temperature is below the target the proportional control requests a reduction in fan speed and requests an increase in speed when the temperature exceeds the target.

Integral control looks at errors between the target temperature and current temperature over time. Integral control requests fan speed changes based on how long a temperature has been above or below the target value.

Derivative control looks at the increase or decrease rate in temperature to decide how to respond with fan speeds. An increased or decrease in fan speed is requested based on how fast the temperature of the processor is rising or falling independent of how close the temperature is to the target temperature.

There are several advantages of utilizing PID control for processor cooling. The correct combination of the proportional, integral and derivative control terms allows for the following:

Reduced fan speed overshoot: PID control can reduce the behavior of fans ramping up to speeds greater than required to cooling the processor.

Minimizes fan speed oscillation: PID control can correct troubling oscillations in fan speeds.

Prevents temperature overshoot: PID control can prevent the processor temperature from exceeding its limits.

4.5 Memory Thermal Management

The iDRAC fan control utilizes memory temperature sensors to provide optimal cooling to the memory. Fan zones, mapped to memory cooling, are controlled based on the hottest memory module in a specific bank of memory. When the memory temperatures exceed predefined thresholds, a combination of fan speed increases and bandwidth throttling maintain cooling requirements. This differs from the industry approach of using Closed Loop Thermal Throttling (CLTT), where the only mechanism for thermal control is memory bandwidth throttling.

In addition to the CPU cooling, the user defined BIOS settings also tune the memory cooling for performance or power optimized modes. The performance optimized mode increases fan speeds to prevent memory throttling and maximize bandwidth. The power optimized mode imposes a limit on how high fan speeds run and throttles memory bandwidth to reduce power consumption to manage memory temperatures above the fan speed limitation.

4.6 IOH Thermal Management

Dell took a unique approach to thermal management of the IOH due to the demanding thermal requirements of this chip. Combined with the custom heat sink design, the iDRAC monitors the IOH's temperature sensor and increases speeds in specific fan zones when the temperature rises above a defined threshold to ensure that the IOH is supplied with its cooling needs. This differs from the intended use of the IOH temperature sensor where it triggered throttling and temperature warnings.

4.7 Event Based Thermal Management

Aside from the management features used for normal operation, the iDRAC also includes event-based features to protect the hardware. The Dell PowerEdge T610, R610 and R710 servers offer redundant cooling solutions that incorporate hardware and software designed to maintain adequate cooling in the event of a failure. When a fan fails in a system the amount of airflow delivered to a particular subsystem is reduced. This potentially subjects components to severe temperatures that can compromise hardware reliability, data integrity and/or system functionality. If any of the fans in the system fail, the iDRAC senses the failure and can increase the remaining operational fans to compensate.

In the event that a cover/lid is removed, the airflow produced by the fans can escape into the ambient surroundings, thus bypassing the components it was intended to cool. Server airflow management and component cooling rely on the chassis cover/lid to keep air within the chassis and flowing over the system components. Dell servers include chassis intrusion switches that alert the iDRAC thermal management when the chassis lid or cover has been removed. The iDRAC can increase the fans' speeds or reduce system throughput to ensure system cooling integrity.

5 Power-to-cool

Power-to-cool ratio is a metric used by Dell thermal engineering to quantify the efficiency of a thermal solution for a particular system and the percentage of system power consumed by the fans. This metric relates the amount of heat dissipated to the amount of power used to remove it. Dell is continually improving its designs and cooling efficiency as illustrated Table 3. The budgeted power-to-cool ratio is the percentage of theoretical maximum system power consumed by the fans at full speed and the nominal power-to-cool is the percentage of actual system power consumed by the iDRAC controlled fans for a typical configuration.

The three-year energy cost-to-cool is an estimate of the cost of energy to cool the system running the spec benchmark in a 25°C data center for three years. An assumed average US energy price of 11¢/kWh and a data center Power Usage Effectiveness (PUE) of 2.0 is used to calculate this cost.

Table 3.   Improved PowerEdge Designs for Cooling

System	Server Generation	Budgeted Power-to-Cool Ratio (Percentage of system power budget allocated to fans)	Nominal Power-to-Cool Ratio (Percentage of system power used to cool a nominally configured system)*	3yr Energy Cost-to-Cool (US 11¢/kWhr / 2.0 PUE 25C Spec Benchmark Application)
PowerEdge 1950	10th	12%	6.5%	$173
PowerEdge R610	11th	11%	2.5%	$66
PowerEdge 2950	10th	12%	4.0%	$66
PowerEdge R710	11th	10%	2.0%	$47

6 Conclusion

Dell wants to simplify IT for its customers by developing energy-efficient servers that are easier to integrate into power and cooling challenged environments. The custom thermal design of our latest server generation has reduced the power-to-cool ratio over our last generation by at least 50%. Flexible BIOS options, dynamic fan speed control, custom heat sinks, efficient fans, and airflow management are some of the ways Dell provides an efficient, reliable thermal solution.