» Whitepaper «

Software Optimization Advances Power Efficiencies in Global Telecom Solutions

Maximize Hardware Investment by Achieving 32 Percent Power Savings

Software Optimization Advances Power Efficiencies in Global Telecom Solutions

Sophisticated telecommunications systems offering higher bandwidth capacities, more intelligent processing architectures and more complex implementations result in greater power demands. And more power means higher operating costs, as well as more complex engineering to accommodate thermal management. As a result, power efficiency has emerged as one of the key areas for long-term improvement in telecom applications. Reduced energy usage means lower costs and diminished environmental impact; in turn, potential savings for carriers are significant when evaluated against the 'always on' central office or data center.

CONTENTS

Introduction

Hardware is commonly the starting point when evaluating telecom power efficiencies, given current silicon advances that provide capabilities for effectively managing a server's power consumption. Kontron's Communications Rack Mount Servers (CRMS) incorporate these advances and are standard building blocks used in a variety of telecom and network applications; these high-performance systems effectively satisfy the demanding requirements and limited space of the telecom central office and data centers. However, software is considered less often and is routinely overlooked in the quest for power savings. Dramatic energy savings can in fact be achieved by focusing attention on the operating system, its configuration and the application itself. The software optimization techniques and options outlined in this paper add significant value to hardware investments, and illustrate up to a 32 percent reduction in CRMS systems' power consumption under various workloads.

Figure 1: Communications Rackmount Servers from Kontron are designed for demanding telecom central office applications. Certified NEBS Level 3 and ETSI-compliant, these Carrier Grade systems are ideal for unified messaging, SoIP, video on demand (VOD), call control, media and signaling gateways, operational system support, SIP server, IMS, mobile location service, media servers, subscriber billing, and service provisioning.

Industry Perspective

Industry-wide focus on energy savings, driven by Verizon's initiative targeting an aggressive 20 percent annual power reduction on deployed systems, illustrates the urgency carriers are placing on power efficiency policies. Consider telcos' need to manage costs including not only the cost to supply energy, but also the cost to remove it again as heat. The resulting thermal management requirements can double the cost of the energy usage alone. Moreover, waste heat limits equipment density, consuming valuable space and limiting service capacity, especially in well-established central offices with fixed building outlines.

Verizon's initial poll of telecom vendors and manufacturers indicated confidence in achieving a ten to 15 percent reduction in power consumption for new equipment; the resulting initiative was intended to push that envelope by setting a 20 percent goal. The initiative is based on formulas designed to test the power consumption of equipment in various operating conditions, and includes a specific measurement process and series of Telecommunications Equipment Energy Efficiency Ratings. To determine power consumption in broadband, video, data center, network and equipment based on customers' premises, test data relative to each piece of equipment is entered into the formula and evaluated against the required energy rating. Equipment included in this initiative includes optical and video transport systems, switches and routers, DSLAM (Digital Subscriber Line Access Multiplexer) high-speed internet equipment and optical line termination solutions, as well as switching power systems, data center servers and power adapters that operate and monitor customer operations.

Telcos are challenged to meet these energy protocols, committing technological expertise to reducing costs and protecting the environment. Future goals will likely set an even higher bar, intended to continually improve energy savings on a global basis. As a result, system architects must understand the range of hardware and software options for meeting and exceeding energy efficiency standards today. Software optimization can differentiate significantly greater results than achieved with hardware alone, ideally driving telcos to leverage both resources for the right combination of bandwidth, performance and reliability within the most competitive power threshold.

Opportunities to Find Power Savings

Telecom equipment is typically deployed adequately for expected peak traffic plus headroom. As a result, portions remain partly idle and the system rarely operates at peak load. For telcos, this creates a unique opportunity to increase power savings by effectively matching power consumption to server workload. Applying software techniques to control CPU power usage, for example, creates different levels of usage by defining a performance cycle and a sleep cycle.

P-states, or the level of CPU performance, represent particular CPU frequencies. This refers to how fast the CPU and its various cores process data, along with its corresponding power requirement. C-states represent sleep states achieved when portions of the processors are directed to remain inactive. Deeper sleep states consume less power but require more time to return back to work. Since higher speeds consume more power, system architects would logically assume that reducing processing speeds will save power. However, occasionally P-states and C-states work against each other, requiring deeper knowledge of the application itself. For example, applying C-states may be a particularly prudent option given the high number of cores that can be found in enterprise servers or data center systems. A server may be implemented with eight cores but only require one to complete a particular task.

An installed operating system would make some of these decisions by default; however system expertise is often required to define the ideal settings for performance and power. Optimization techniques address this conflict, matching the workload to the best hardware management scheme and evaluating P-states and C-states for ideal performance. Extensive hardware design expertise may be ideal in these instances, supporting OEMs by providing uniquely deep engineering insight and performance analysis. This can be a vital step to achieving power efficiencies early and for the long term, especially critical considering that most telecom systems are locked into their defined performance settings for long-term operation. Adaptive decisions on performance vs. power are anticipated in the future, however today's system architects must not only evaluate power/performance schemes upfront but also understand how the application itself impacts chosen software techniques and options.

Software Optimization

A fully-loaded server is going to require greater amounts of power. When a server is not at full capacity, it takes a combination of both hardware and software controls to bring it down to less power utilization, essentially correlating the power to match the server load.

Not long ago, servers were largely unaware of power as a strategic asset in achieving top performance. Servers were always on, or at best turned on and off to match usage patterns – and idle servers used as much power as servers under load. Recent hardware generations have included power reduction circuitry that cooperates with software enhancements to reduce idle power as well as power under load. This hardware has power savings built-in, however benefits are only realized if the software implements power saving algorithms.

Gradations between on and off are analogous to turning the lights off when exiting each room as you walk around the house. The parts of the chip not being used can be turned off automatically through hardware and software. Unlike power management schemes which turn entire servers on and off, these power transitions take milliseconds instead of minutes and the OS remains alive and operational during the process.

Upgraded Kernels

Each generation of silicon improves, delivering more useful features and sophisticated performance options. Newer silicon for example provides deeper sleep states with lower return latency. This in turn provides finer control over frequency and ultimately, power consumption. However, hardware features only function properly if control is exercised through software optimization; at minimum, a recent operating system version should be installed that takes full advantage of the hardware features.

Coupling new hardware with a recent OS is a great step forward for many telco systems. In Linux for example, more recent kernels have an improved scheduler which makes better use of the hardware's power and sleep states. While all recent Linux kernels contain some support for sleep states, 2.6.21 introduced the "tickless" kernel. Prior to the tickless kernel, x86 Linux kept time and scheduled events by counting ticks from the legacy timer. It awakened a CPU many times a second only to increment a counter and go back to sleep. In contrast, the tickless kernel leverages the High Precision Event Timer (HPET) found on today's chipsets to schedule events. Processors sleep longer and considerable power is conserved. This simple advantage is not necessarily common to every OS distributor; each has a different policy for releasing new kernels and several major distributors in the server space do not yet include the tickless kernel.

Figure 2: The Kontron CG2100 Carrier Grade Server combines performance, ruggedness, reliability, and long life in a NEBS-3 and ETSI-compliant 2U chassis, ideal for telecommunications environments. It provides dual socket support for the Intel® Xeon® Processor 5600 series, coupling high performance with power efficiency to provide improved performance-per-watt over previous-generation rack-mount servers.

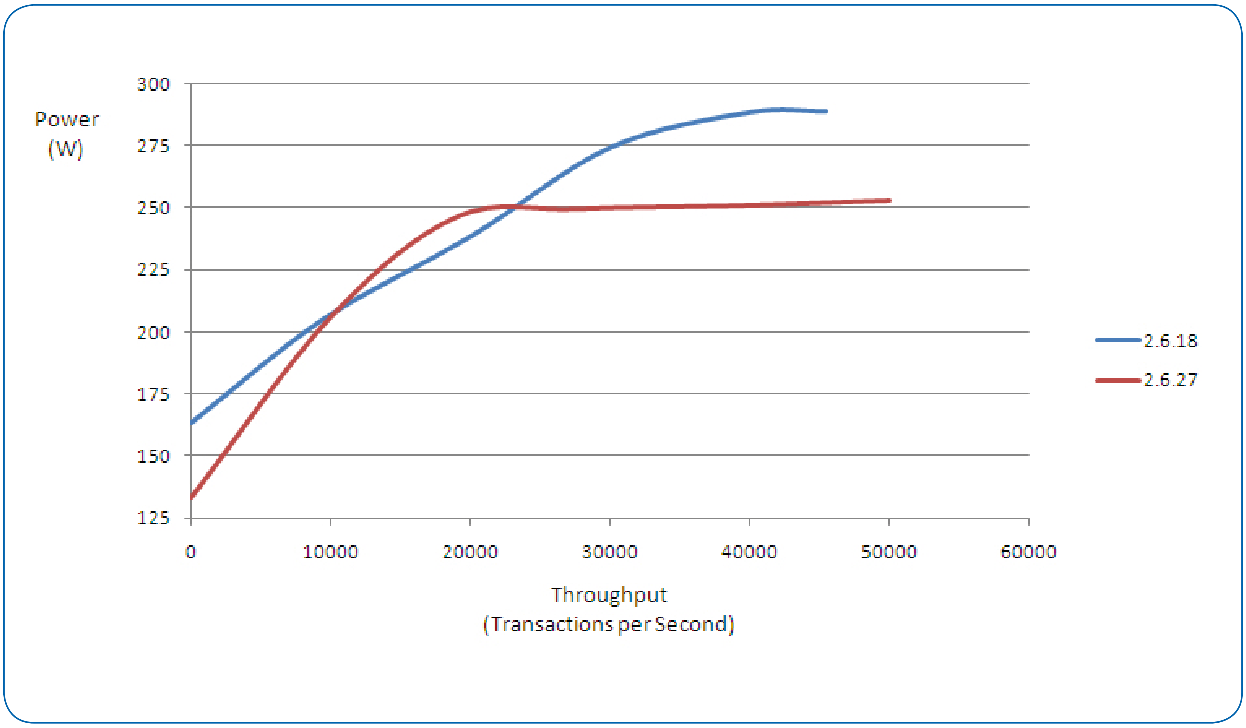

Figure 3 shows a sample workload using two popular releases of the Linux kernel, 2.6.27 and 2.6.18. In particular, 2.6.27 adds the tickless kernel and 2.6.18 does not. Using the less sophisticated timing mechanism on the earlier kernel, the idle machine consumed 163 W versus 133W with the tickless kernel, which delivered an 18 percent savings in power. Even with a significant workload, savings topped 12 percent by using the more sophisticated timing feature of the current Linux OS.

Power Governors

Servers need a strategy for how fast to process data and how often to sleep, i.e. controlling the P-states and C-states to achieve the largest energy and performance advantage. Policies such as these are implemented in the Linux power governors, and often start by asking some tough questions. Since processors consume more power at higher frequencies and minimal power while sleeping, is it better to finish a task quickly and sleep more or is it better to sleep less but consume less power while awake?

For some workloads, system administrators may determine that it is ideal to have the processor running as quickly as possible. Although consuming greater power, it completes its task quickly and returns to C-state. Other workloads, however, perform to improved power settings by letting the CPU run as slowly as possible. Even though a particular core is kept awake longer, it consumes less power during the task.

The answer is workload dependent and requires tradeoffs between throughput, latency and power consumption. Three different types of workloads must be considered, including processes that are CPU-bound, memory-bound or I/O-bound.

For example, some workloads are CPU-bound for brief spikes of activity, such as when new packets come in to be processed. In these cases, the processors run at high frequencies to complete their work quickly and then immediately return to sleep until the next spike, maximizing the amount of sleep time and minimizing power consumption. The Linux “on-demand” governor implements this particular policy and it is the default in most distributions.

In contrast, Figure 4 illustrates a memory-bound application. Changes in processor frequency affected it slightly but increasing cache size improved throughput greatly. As a result, the sample workload showed greatest power savings when run at the lowest frequency because much of the processor’s time was consumed waiting for data to return from the memory controller.

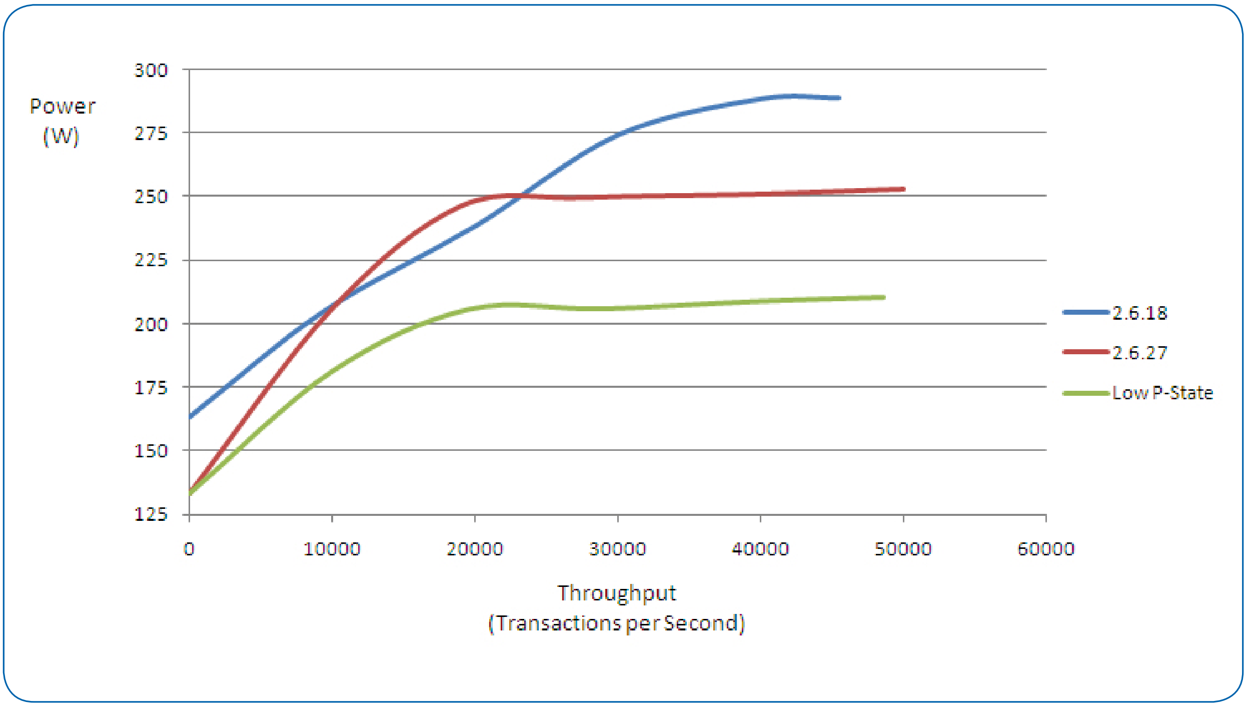

Figure 4: shows the power savings achieved by choosing a lower power state for our particular workload. The top two curves are copied from Figure 1 which used only the "on-demand" kernel. The bottom line shows the power savings achieved by taking the tickless kernel (2.6.27) and applying the "user space" governor to place the processors in the lowest power state (i.e. lowest frequency).

Overall, telecom applications driven by I/O present an interesting challenge. If the thread begins with the arrival of a packet, the best strategy depends on what is happening to that packet. A packet compared against an in-memory lookup table might benefit from the lower processor speed whereas a mathematical operation on the packet might benefit from a higher one. Further, none of this considers cache locality. In all cases, the answer can only be known by characterizing the workload on a real machine or suitable simulator. Moreover, power efficiency is only one goal, and must be considered within quality of service, and the metrics of throughput and latency.

Interrupt Handlers

Interrupt handlers present telcos with tradeoff options between power and performance. Dispersing hardware interrupts as widely as possible may maximize throughput, however at less than peak load this merely wakes processors that could otherwise sleep. Consider a packet forwarding application that receives incoming packets on multiple network interfaces. At peak load, it often makes sense to assign each interrupt handler to a separate core. At less than peak load, it is possible to achieve the requested throughput and latency while consolidating interrupt handlers on a smaller number of cores.

The OS makes no attempt to optimize this sequence for ideal power usage. Achieving power reduction here requires continual re-balancing of the interrupts based on quality of service measurements such as throughput and latency. A software daemon would consolidate or disperse interrupt handlers to achieve the desired balance. (For consistency, these tests pushed all interrupts to a single core. A real-world telecom application would need to spread interrupt handlers more widely when quality of service required better throughput or latency.)

Application Tuning

One non-power-aware application in the mix can spoil overall power savings. For example in a packet inspection application, worker threads might be dispatched to perform the actual decoding, analysis and lookup as new packets arrive. Without optimization for power awareness, the application could let the worker threads sit in a polling loop while waiting for new work items to appear in the queue. The processor handling the thread would be fully awake, consuming full power. A power-aware application would allow these threads to block, returning to the scheduler while waiting for the new event. In this instance, the processor would sleep until needed, again saving significant power.

Core selection is another power vs. performance tradeoff that can be controlled by the user. If coded correctly, idle cores consume minimal power. As threads are assigned to cores however, performance tradeoffs may arise because certain resources are shared. For instance, hyperthreaded core siblings share most of the same CPU resources, and cores within a single CPU share input/output (I/O) and cache. Adding a second CPU doubles the cache, and sharing cache may or may not be preferred depending on the application.

By default, the OS scheduler will dispatch threads as widely as possible although this can be adjusted through CPU affinity. If threads do share data, cache locality suggests that threads should be kept as close together as possible, for example using cores in the same package behind the same cache. In contrast, many applications benefit from sharing as little hardware as possible. In a dual-processor server, bringing the second package online also doubles the amount of cache, a real benefit to performance in most cases.

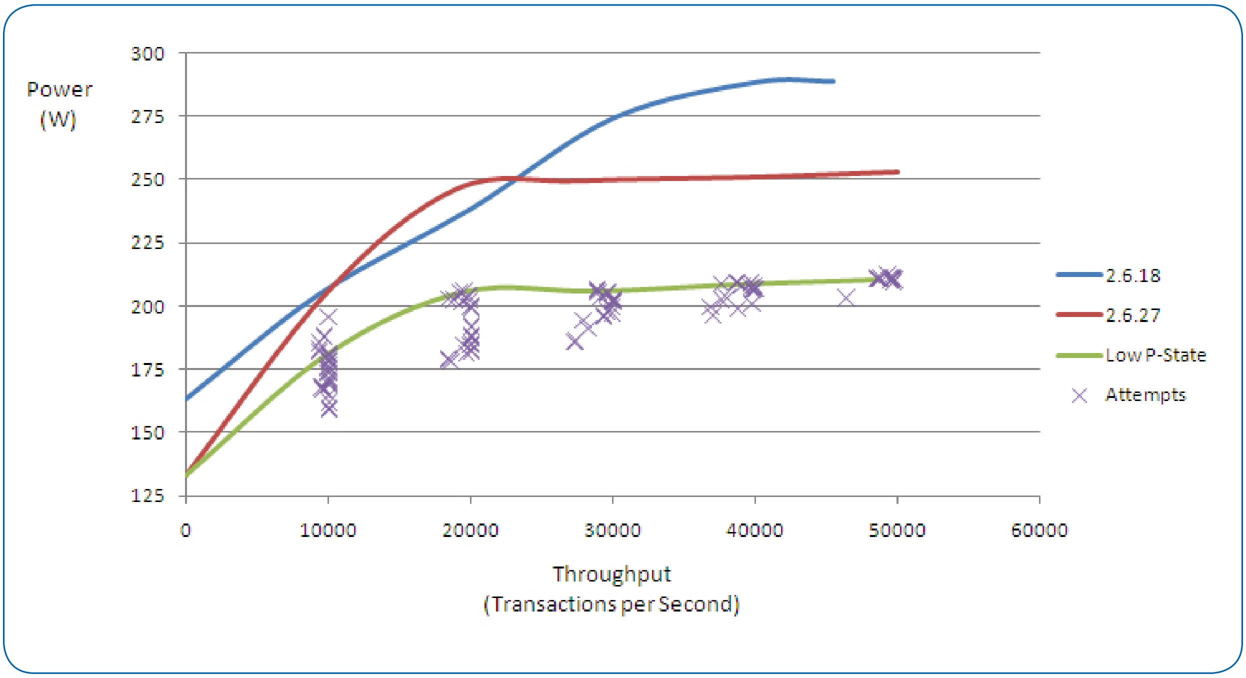

Figure 5 shows a variety of results achieved by changing the number of threads and the number of active cores. Each data point represents the power and performance achieved by adjusting these parameters. Data points below the original three lines indicate a combination of parameters that outperformed the built-in options, illustrating the power savings opportunity by implementing an optimized power governor.

Putting it all Together

The most competitive power efficiencies result from a well-written application running on the latest hardware and software. By adding greater levels of software optimization, power savings are advanced even further with a daemon that adaptively adjusts CPU affinity, interrupt handlers and CPU frequencies or power states.
By using a workload generator and tuning each system, a dramatic 18 to 32 percent power savings was realized at various workload levels when compared to the original power/performance curve with the out-of-box (2.6.18) kernel. A truly adaptive policy would monitor incoming requests and quality of service metrics to determine if additional hardware resources would benefit the workload presented at any given time.

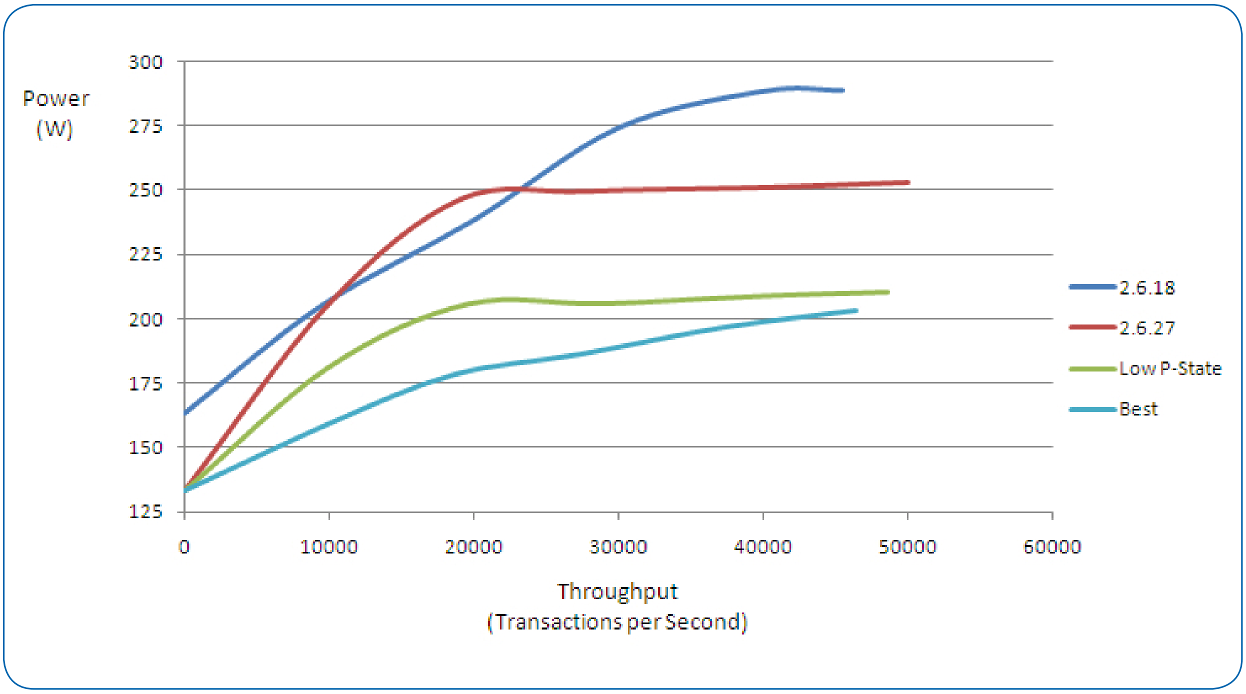

Figure 6 shows the final results.

Manufacturer Insight on Strategic Software Choices

Rugged, carrier grade servers offer the performance, long life and reliability integral to secure telecom applications. Manufacturer expertise in developing these systems is essential in validating their ability to meet and exceed demanding performance requirements – but designers must understand that power policies and actual energy savings depend on workload and application architecture. Hardware may be the first line of defense to manage these industry priorities, however greater efficiencies result when software optimization is addressed as part of the solution. In fact, using a simple workload generator and intelligent software adjustments based on manufacturer insight, the peak power of a CRMS system was reduced by nearly a third.

Blending the know-how of hardware development with extensive software expertise provides the fine tuning that distinguishes an efficient CRMS system from one optimized for long-term, application-specific power awareness. Software optimization techniques make the most of CRMS hardware investments, adding value and representing an area where deep manufacturer expertise can be a significant resource for designers. Implemented with the latest generation of server hardware, software changes can reduce power consumption significantly and help telcos achieve aggressive improvements in energy conservation.

About Kontron

Kontron, the global leader of embedded computing technology, designs and manufactures embedded and communications standards-based, rugged COTS and custom solutions for OEMs, systems integrators, and application providers in a variety of markets. Kontron engineering and manufacturing facilities, located throughout Europe, North America, and Asia-Pacific, work together with streamlined global sales and support services to help customers reduce their time-to-market and gain a competitive advantage. Kontron’s diverse product portfolio includes: boards & mezzanines, Computer-on-Modules, HMIs & displays, systems & platforms, and rugged & custom capabilities.

Kontron is a Premier member of the Intel® Embedded Alliance and has been a VDC Platinum Vendor for Embedded Computer Boards 5 years running.

Kontron is listed on the German TecDAX stock exchange under the symbol "KBC". For more information, please visit: **www.kontron.com**

CORPORATE OFFICES

Europe, Middle East & Africa

Oskar-von-Miller-Str. 1
85386 Eching/Munich
Germany

Tel.: +49 (0)8165/ 77 777
Fax: +49 (0)8165/ 77 219

info@kontron.com

North America

14118 Stowe Drive
Poway, CA 92064-7147
USA

Tel.: +1 888 294 4558
Fax: +1 858 677 0898

sales@us.kontron.com

Asia Pacific

17 Building,Block #1,ABP.
188 Southern West 4th Ring Road
Beijing 100070, P.R.China

Tel.: + 86 10 63751188
Fax: + 86 10 83682438

kcn@kontron.cn

intel Embedded Alliance Premier

#Whitepaper - Software Optimization Advances Power Efficiencies in Global Telecom Solutions# 01052011BFR
All data is for information purposes only and not guaranteed for legal purposes. Subject to change without notice. Information in this datasheet has been carefully checked and is believed to be accurate; however, no responsibilty is assumed for inaccurancies. All brand or product names are trademarks or registered trademarks of their respective owners.