# Out of Power, Out of Cooling:

## Solutions for High Density Data Centers

**By**

**Dr. Kenneth Uhlman, Director of Data Center Business Development,**

**Eaton Corporation**

**And**

**Jim Vanderpas, Director of Business Development, Mission Critical,**
**UTC Power**

## Table of Contents

## 1) Executive Summary

The relentless demand from society for instantaneous information, video, and knowledge continues to drive the global growth of the IT segment. This growth and the energy required to sustain it have stressed the limits of today's data centers. To meet these demands, IT continues to consolidate data centers, deploy blade servers and adopt virtualization to increase asset utilization and provide the ability to transfer loads across a data center or across the globe within minutes.

While new technologies that improve IT utilization, such as blade servers and virtualization, solve some issues, they also create new ones. Increased power densities from blade servers free up floor space in the legacy data center, but they also speed the process of running out of power and cooling. This results in underutilized raised floor space, especially within data centers located in major metropolitan markets such as New York City, San Francisco, or London where incremental power is not easily obtainable from the local utility grid.

Electric grid congestion limits the available power in these markets, especially with the long lead times and reluctance of the public to add power plants or transmission lines in their neighborhoods. The latest generation of service-based applications has minimum response time criteria which limit the physical location of remote data centers hosting those applications and further compounding the power availability problem.

With the U.S. energy consumption by servers and data centers expected to double in the next 5 years (EPA Report to Congress, 2007), organizations are now at a tipping point. They must develop strategies that meet their growth objectives, increase energy efficiency, and incorporate new technologies while maintaining or increasing system availability. This paper will discuss solutions that are available to meet these pressing needs, including freeing stranded capacity from existing buildings and adding incremental power and cooling solutions. Organizations that implement these solutions may be able to delay or eliminate construction of new data centers, thereby saving millions of dollars.

## 2) **Legacy Volume Server Data Centers**

Historically, many data centers used volume servers in a dedicated, single-application model, meaning they were not capable of reprovisioning themselves with multiple software applications. For practical purposes, these servers were also physically constrained to operating in their original racks, and rarely moved. This resulted in average server utilization rates in the 10-20% range. Many of these servers are dual corded, which enable them to operate on an "A" or "B" power system for increased availability.

Unfortunately, this architecture resulted in very inefficient power systems, since an idle server still consumes approximately 50% of its full load power. Likewise, in a dual bus power distribution configuration, UPS systems are loaded less than 50% to enable them to assume the full data center load in the event of losing systems "A" or "B." At these low utilization rates, legacy UPS systems are typically less than 85% efficient, resulting in energy losses and additional stranded capacity in these data centers. Inefficient legacy computer room air conditioning (CRAC) systems compound the inefficiency issues since they typically run at 100%, regardless of true heat load. The net result is data centers using only one-third to one-half of their input power for the IT load with the rest being consumed by ancillary support systems.

For example, Figure 1 shows a data center with a 1 MW utility substation with 400 kW of UPS output power feeding a farm of servers rated at 4 kW/rack.
To increase server utilization and flexibility while reducing IT costs, organizations are incorporating blade servers and virtualization. These technologies enable rapid server hardware deployment and the ability to

reprovision applications and services from one server to the next anywhere on the network within minutes. Blade servers are much more compact and have much higher power densities than yesterday's volume servers.

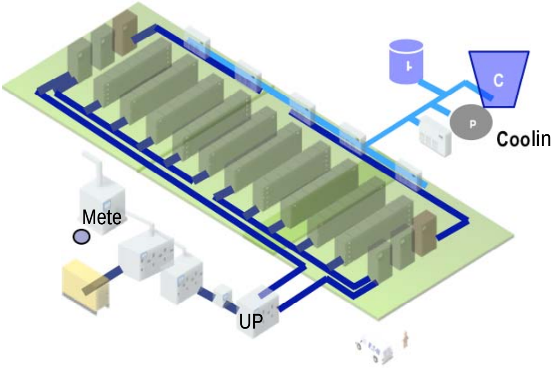

Figure 1. Legacy Volume Data Center with 4 kW/Rack

**Blade Server Data Centers**

To increase server utilization and flexibility while reducing IT costs, organizations are incorporating blade servers and virtualization. These technologies enable rapid server hardware deployment and the ability to reprovision applications and services from one server to the next anywhere on the network within minutes. Blade servers are much more compact and have much higher power densities than yesterday's volume servers. Today's blade server rack power density can easily be 25 kW/rack with higher densities being planned for in the near future, in the range of 50 kW / rack. Figure 2 illustrates how the original legacy data center from Figure 1 was reprovisioned for 25 kW/rack blade servers. Each row in the redesigned data center contains 10 racks, for a total of 250 kW. After two rows, the data center is out of UPS power and cooling capacity.

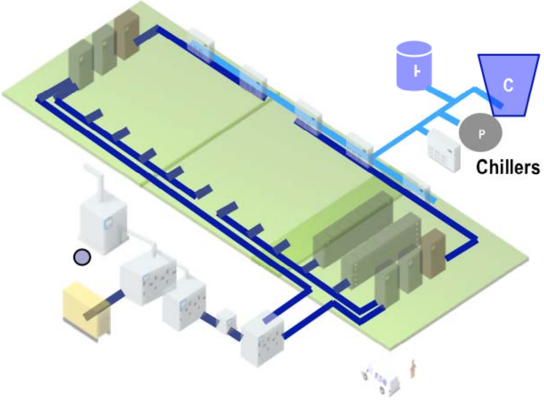

Figure 2. Blade Server Data Center

## 3) Data Center Efficiency Metrics

While the industry endorses availability metrics like The Uptime Institute's Tier rating, the industry has not had standardized metrics to measure data center efficiency. To address this issue, The Green Grid, a non-profit global organization (www.thegreengrid.org) focused on solving the energy issues in data centers, has introduced two metrics: Data Center infrastructure Efficiency (DCiE) and its inverse, Power Utilization Effectiveness (PUE). DCiE calculates the data center efficiency which is often how facilities personnel view equipment while PUE calculates the data center infrastructure overhead, which is preferred by the financial community.

The formula for DCiE is:

DCiE = Power Used by IT ($P_{dc}$) / Total Power used by the Data Center (including IT) ($P_{fac}$)

The formula for PUE is:

PUE = Total Power In ($P_{fac}$) / Power used by IT($P_{dc}$)

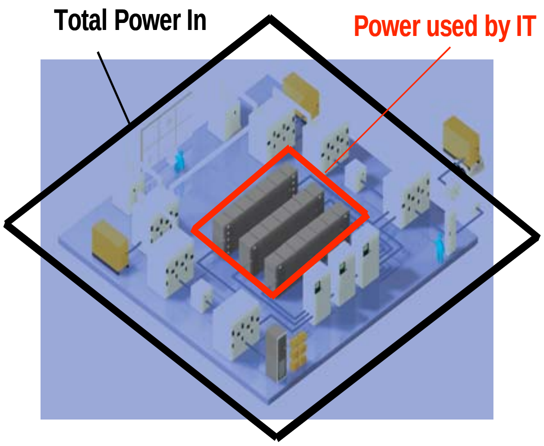

Figure 3. Data Center Metrics

Stranded power and cooling capacity can be freed by using liquid-cooled racks or computer room in-row cooling solutions and more energy-efficient UPS systems. Such methods may provide short-term solutions, but may still fall short by not enabling IT to utilize all of the data center floor space with the increased power densities required by blade servers. This is especially important for data centers that are tied to geographic locations such as the financial or telecom markets in New York City, San Francisco, London, or elsewhere.

## 4) Freeing Stranded Capacity

### IT

IT vendors have made significant advances in performance per watt of computer, storage, and network systems with significant reductions in costs. As a result, organizations continue to have an insatiable demand for growth and, therefore, continue to leverage these new technologies. Legacy IT equipment has consumed an average 4-6 kW per rack. New technologies such as blade servers consume in excess of 25 kW per rack resulting in significant power density along with power and cooling capacity issues.

Current technologies such as high-efficiency power supplies, specified under the Department of Energy's *Energy Star* program, result in reduced power and cooling loads. This enables more IT equipment to be installed. Cooperation between IT and Facilities may be required since in many organizations, IT supplies the equipment and Facilities pays the energy bills.

### Power Quality Systems

Historically, UPS systems were designed for maximum reliability, with little regard for energy efficiency, especially at light loads. The advent of dual-corded servers to increase availability has resulted in UPS systems operating below 50% load to enable "B" system to assume the full load if the "A" system fails. These legacy UPS systems are typically 85% efficient at these low utilization rates. These losses are doubled when incremental cooling is utilized to remove the heat.

High-efficiency, high-availability AC UPS systems that incorporate double conversion on demand operate at 97% efficiency across the entire load spectrum while still maintaining maximum protection. These loss reductions, when coupled with the reduced cooling load, may free up to 25% of the UPS system enabling incremental IT equipment to be installed. Other savings are available, including using high-efficiency transformers in power distribution units.

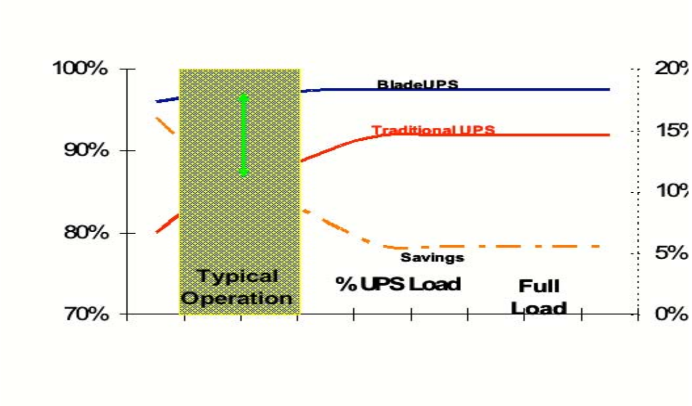

Figure 4. High Efficiency UPS Systems

### Cooling Systems

Traditional computer room air conditioning (CRAC) systems cool the entire data center by moving mass amounts of air. These cooling systems were satisfactory for legacy 4-6 kW racks, but may not be adequate to cool high-density loads. Likewise, specific areas of a data center may have higher density loads than others, which may result in hot spots with legacy CRAC cooling systems.

Addressing the inefficiencies in existing data center cooling systems may result in incremental cooling capacities being freed up. More efficient examples include hot aisle / cold aisle isolation, in-row cooling systems such as self-contained cabinets or rear-door heat exchangers. By placing the cooling directly at the heat sources, less energy is consumed and that frees up additional cooling capacity.

## 5) Incremental Power and Cooling Solutions

While new IT technologies such as blade servers use considerable less floor space than legacy IT equipment, their power densities result in retrofitted data centers with raised floor space available but which are unable to add additional equipment within the free space because the facility is out of power and cooling capacity. When all options to free stranded capacity are exhausted, IT and Facility professionals are forced to look at alternative solutions to meet the growth objectives of the organization. One alternative, the construction of a new data center, involves many millions of dollars and more than a year for deployment.

Adding utility substations, emergency generators and cooling plants is seldom practical due to the complexity involved. For many facilities, onsite combined cooling, heating and power systems (CCHP) may be a cost-effective solution. Pre-engineered fuel cell and microturbine solutions enable data center operators to:

- Add incremental power and cooling quickly. This additional power can be connected directly to the low voltage (480VAC) feeders since grid paralleling functions are contained within the microturbine package. System protection and power import/export control is easily implemented with relay panel near the utility service entrance.
- Generate electricity at 40-60% below the cost of equivalent grid-purchased energy.
- Obtain "free" cooling from absorption chillers that can be tapped into existing chilled water loops. This cooling is generated using the waste energy from the CCHP system electricity generation stage and requires no additional natural resources in the process.
- Augment absorption chilling with series flow configurations, allowing the absorption machines to divide the work load with high efficiency, VFD-controlled electric chillers. This results in expanded cooling production at reduced energy consumption (in a range of .15 - .35 kW / ton). Such technology is under development at UTC Power under the trade name Active Redundant Cooling™.
- During periods when "free cooling" is available, divert surplus thermal energy to the production of heating that can interface to non-data center hydronic or ducted-air heating systems.
- Mitigate the risk of rising energy prices. Power and cooling production costs may be "locked" for long periods through various natural gas supply contract mechanisms.
- Minimize the effects of power outages by designing your on-site generation system with redundancy that improves the overall availability of the data center.
- Capitalize, lease, or use third party companies to outsource the operation and maintenance of the CCHP system.

**Microturbine Solutions:**

Figure 5 shows a CCHP system with four 200 kW microturbines providing nominal 750 kW of net electricity and 535° F waste heat fed into 2 absorption chillers that can deliver about 350 RT nominal output of chilled water. Grid-connected operation is handled by onboard system components, eliminating complex site wiring. The power output of the microturbine generators is fed directly into the 480 VAC electrical system through any open 480V 3-phase breaker position.

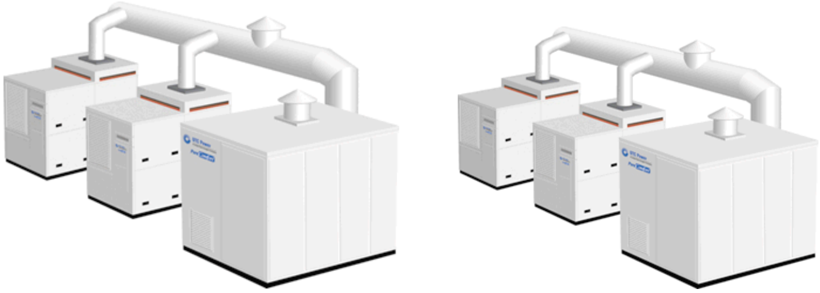

**Figure 5. Conceptual design for 800 kW gross (or ~ 750 kW net) microturbine power plant**

**Cooling**

Chilled water is a by-product of the electricity generation process. The turbine exhaust is captured by an absorption chiller to produce large amounts of cooling. In a perfectly sized system, the cooling output of the CHP system will eliminate 95-100% of the electric power required to run conventional cooling equipment. In extreme ambient conditions, the conventional equipment remains available to augment the cooling produced by the CCHP system. Therefore, the total available cooling is the sum of conventional cooling and CCHP system cooling. Absorption chillers can be integrated with Active Redundant Cooling™ which adds a high efficiency electric chiller to provide backup cooling that can be made to work for economic gain during periods of high ambient or high electricity costs.

## Heating

In installations where heat can be used outside the data center, the waste heat of the microturbines can be utilized to augment building heating systems, further increasing overall system efficiency. Thermal priority is given to cooling; however, remaining energy can simultaneously produce hot water at 175°F (79.4°C). Or, surplus exhaust energy can be diverted to create steam at up to 100 psi.

## New and Retrofit Solutions

For new data centers, CCHP systems can be integrated into the design with sufficient redundancy that they can achieve 99.9985% (or Tier III) critical load availability when working in concert with the normal utility source. This provides the equivalent contingency reliability of diesel systems working in concert with the grid. It eliminates the need for some of the diesel redundancy in the facility design. Cooling redundancy can be completely eliminated since CCHP systems designed for grid-independent load service will typically provide 2N to 3N of mission critical cooling capacity.

To achieve the best system efficiency, these surplus power and thermal capacities must be put to good economic use. This is accomplished by equipping the CCHP system for Dual Mode operation. Dual Mode allows the CCHP system to be paralleled with the utility under normal conditions. Through paralleling, surplus power can flow beyond the data center, to the surrounding building, and even be exported to the grid for economic gain (in areas where net metering is permitted). Dual mode allows electric baseloading, which is the condition at which the system runs constantly at 100% electrical output. When base loaded, turbine exhaust mass flow is at its greatest, thus maximizing the production of useful cooling (or heating). In emergency conditions, the Dual Mode control system activates the fast transfer capabilities of the turbine system, switching the critical load to be powered only by the CCHP system. Cooling systems are designed to ride through the momentary transfer outage and be fully operative during grid down conditions.

With data center retrofits, these same capabilities can be achieved provided the CCHP system is also sized for the full capacity of the IT load expansion, with its own inherent redundant power and cooling. Just as with new data centers, system redundancy can be put to full use through Dual Mode configurations. Where the utility does not have capacity for expansion, CCHP becomes the “utility” and conventional backup systems provide the second contingency needed for high tier operation. Such cases are not common, but in this scenario, CCHP systems configured for high power and cooling availability are mandated.

Microturbine systems not only provide higher efficiency and greater reliability for data centers, they are also environmentally green compared to central power plants. Using an installation in New Jersey as an example, Figure 6 illustrates the advantages in terms of $CO_2$ and Oxides of Nitrogen ($NO_x$) as well as the amount of water conserved. This is equated to the number of acres of trees planted, number of cars taken off the road annually and gallons of water conserved each year.

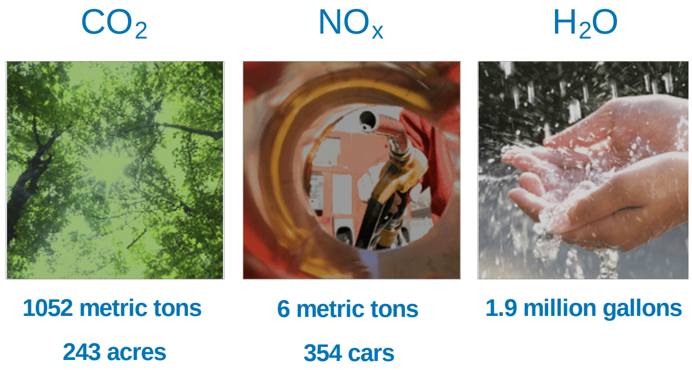

Figure 6*

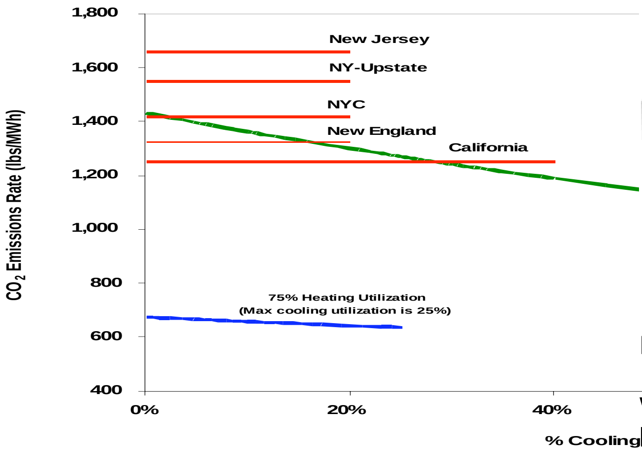

Figure 7. $CO_2$ emissions rates.

Figure 7 shows how the $CO_2$ emissions rate decreases with higher utilization of cooling. For reference the $CO_2$ emissions levels of key regions is also shown.
Figure 8 shows the cumulative carbon savings* of this system over a 10 year period. As can be seen, this is a fairly significant number, over 32,000 metric tons.

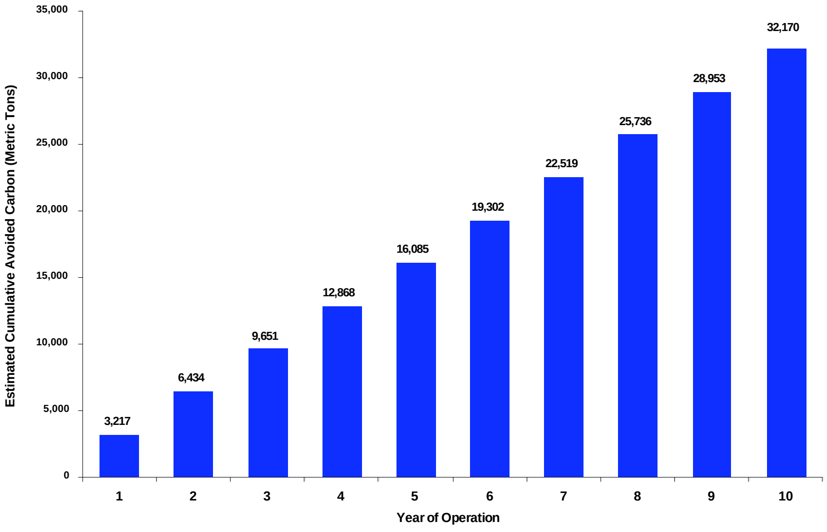

Figure 8. Average carbon savings.*

**Fuel Cell Solutions:**

Figure 9 shows a 400 kW fuel cell designed and manufactured by UTC Power in South Windsor, Connecticut.

Figure 9. UTC fuel cell.

Where fuel cell-based systems formerly were limited in their ability to produce cooling, fuel cells today can be augmented with combinations of absorption chilling and Active Redundant Cooling$^{TM}$ (under development). This combination allows thermal-to-power ratios in keeping with data center power and cooling load requirements. Fuel cells in the 400 kW class are expected to be much lower in cost ($/kW) compared to the earlier 200 kW versions. Additionally, improvements in technology have increased cell stack life to 10 years so they can last for 20 years with an overhaul at the end of year 10. The improvements in the cell stack life and the tight integration of the balance of plant, coupled with remote monitoring capabilities have driven down the O&M cost of a fuel cell to about a half of previous costs.

The capital and operating costs of fuel cells are expected to make the 400 kW unit compete as an economical alternative to grid power in many parts of the U.S.. Coupling this with their excellent emissions performance, fuel cells show good promise for data center operations, especially as data centers become increasingly environmentally and 'green' conscious. Figure 10 captures some of these distinct and significant environmental advantages in terms of carbon and Oxides of Nitrogen for a data center located in New Jersey. These emissions reductions are once again equated to number of acres of trees planted and number of cars taken off the road annually. And similar to the microturbine solution, the fuel cell on-site solution also conserves a significant amount of water as well, about 1.9 million gallons annually for an installation located in New Jersey.

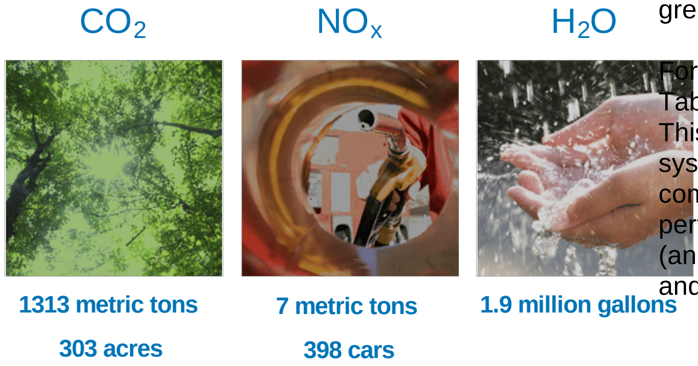

Figure 10. Carbon savings.

* In Figures 6, 7, 8, and 10, the carbon calculation methodology is consistent with the guidance of the Environmental Protection Agency (EPA) CHP Partnership Emissions Calculator.

## Financial Analysis

It is clear that on-site power systems can be designed to displace all or part of a data center's metered energy reliably. The question then becomes: can CCHP systems generate power for less than what I am paying today?

Since these systems can achieve efficiencies in the 85% to 90% range, the total input fuel energy per kilowatt hour of displaced grid energy is very low (in the range of 7000-8000 Btu). This is excellent by any power generation standard. The answer to the question is, therefore, "Yes."

On-site generation of electricity and cooling can produce energy at lower cost. The magnitude of savings is dependent on the relative cost of fuel versus electric power purchased from the electric grid. This cost difference is often called "spark spread" and Table 1 shows typical savings at various spark spreads and illustrates the importance of applying these systems where the cost differential is greatest. These examples are for a 2 x 400 kW PureCell® conceptual design (as shown in Figure 5) with the addition of an Active Redundant Cooling™ System (under development). The system works in concert with the absorption cooling, with the combination producing a total tonnage in a range of 500 T - 575 T depending on ambient conditions. Using the ambient conditions of Newark, NJ as an example, this results in a time-weighted annual average of about 750 kW of electric production and 600 T of cooling. This system is optimized at 350 T output. Such a system would be integrated to a data center with about 1.0 MW of IT load (or greater).

For a 800 kW fuel cell system installation, Table 1 illustrates the potential annual savings. This assumes a gas price of $11/MMBtu and a system-installed price of about $4.1M. The comparison is to time temperature weighted performance of electric compression chilling (annualize 0.74 kW / ton average). No federal and state incentives are assumed here.

| Electric ($/kWh) | Gas ($/MMbtu) | Spread ($) | 1st Year Savings* ($K ) |
|---|---|---|---|
| 0.20 | 11 | 9 | 948 |
| 0.19 | 11 | 8 | 889 |
| 0.18 | 11 | 7 | 801 |
| 0.17 | 11 | 6 | 713 |
| 0.16 | 11 | 5 | 625 |
| 0.15 | 11 | 4 | 537 |
| 0.14 | 11 | 3 | 449 |
| 0.13 | 11 | 2 | 361 |
| 0.12 | 11 | 1 | 273 |

Table 1. Potential savings are sensitive to spark spread.

Given the cost assumptions for such a project, state and federal incentives can result in net cost of under $2M. Paybacks with incentives can easily be under 3 years. This payback is possible without any federal or state incentives. Paybacks will improve where Federal and State incentives are available. Federal incentives are available in all States whereas State incentives, particularly for fuel cells are very significant in selected States. Figure 11 shows a summary*. For a detailed list of these State by State incentives, please refer to www.dsire.org.

## PUE Impact

A CCHP system sized to match the cooling requirement will provide maximum return on investment (shortest payback). It will also provide best efficiency. For example, each 1MW of IT load normally requires about 350 tons of cooling. If two PureComfort® Solution

Model 400M's are matched to that cooling requirement, they will supply about 350 tons and eliminate the need for conventional compression chilling. By displacing 350 tons of conventional chilling, metered power is reduced by 250 kW. The electric output of such a system becomes coincidental. The prime movers (generators) will displace about 750 kW of metered power. The total reduction in metered load will be 1,000 kW (250 kW cooling plus 750 kW input power to UPS). Total displacement is generally slightly more, given the displacement of transformer losses upstream of the point of injection. At 1,000 kW displaced energy, the CCHP system is 100% base loaded both electrically and thermally (the ideal).

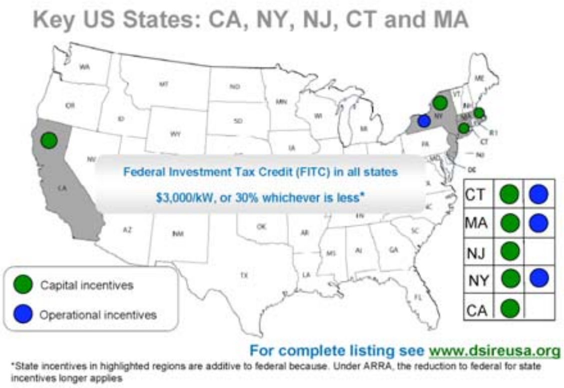

Figure 11.* California does not have any incentives for microturbines at the present time.

Economic design provides exceptional financial performance, but, by design is not intended to backup any critical load. Thus, such a design will not eliminate a layer of diesel or chiller redundancy. If the grid is down, the system Figure 11would separate from its grid connection and pause while traditional backup systems handle the emergency. The only loss during these periods is a fraction of annual savings; no data is threatened.

A system of this design results in the lowest installed cost because it is not necessary to take special precautions on elimination of SPF's, nor to provide redundancy.

The PUE effect is as follows (assuming a 1,000 kW IT load and a 1.85 PUE before CCHP):

1850 kW original metered load
\- 1000 kW displaced by CCHP
(Capacity: 250 kW cooling and 750 kW power; 100% utilized)
850 kW new metered load
\+ 750 kW of metered CCHP output
1600 kW revised numerator in PUE equation

Thus

$$\frac{1600\text{ kW metered}}{1000\text{ kW IT load}} = 1.60\text{ PUE}$$

(13.5% improvement)

## IT Capacity Increase

Figure 12 shows a comparison of data centers with and without CCHP systems being utilized. In this example a smaller CCHP system is used to increase the available power and cooling for 400 kW data center.

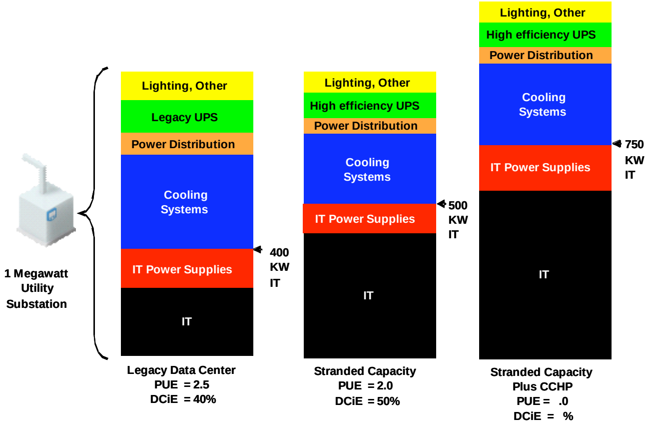

Figure 12. Data Center Comparison

## 6) Conclusions

IT and Facilities personnel must address data center IT, power, cooling, and energy efficiency issues collaboratively in order to meet the demand for increased IT services and the organization's sustainability goals. The technology market is stepping forward to offer the industry new and innovative products and solutions to allow IT and facilities managers to take control of their energy future – *today*.

## Authors:

Dr. Kenneth Uhlman, PE, is the director of Data Center Business Development for Eaton. He can be reached at KennethLUhlman@eaton.com.

James VanderPas is the Director of Business Development, Mission Critical for UTC Power. He can be reached at jim.vanderpas@utcpower.com