Spring 5-2-2013

Supercomputer Design: An Initial Effort to Capture the Environmental, Economic, and Societal Impacts

Marshall T. McDonnell
University of Tennessee - Knoxville, mmcdonn1@utk.edu

Supercomputer Design: An Initial Effort to Capture Environmental, Economic, And Societal Impacts

M. T. McDonnell*

May 3, 2013

Abstract

Currently, the Green500 is possibly the largest effort to direct the evolution of supercomputing towards sustainable design but does not address life-cycle impacts of a supercomputer. Supercomputer assessment with economical, environmental, and societal impacts considered would provide an optimal future supercomputer design that is sustainable. Some of the benefits of such an assessment would help determines the complete impact for bringing a supercomputer online, determine impact "hot spots" in design, and determine optimal locations for construction. By combining process-LCA and EIO-LCA methods, an initial model for in-depth sustainability analysis of supercomputer design has been developed and the results presented. The model provides comparison of economic, environmental, and societal impacts for manufacturing and operation of not only a single supercomputer but a range of designs. The ultimate goal of this model is to assist government, academia, and industry in supercomputer development for existing and future supercomputers. The advantages of the model are determining the entire impact before the initial construction begins, determine high-impact sections of the supercomputer's design, and to direct construction to optimal locations.

---

1 Introduction

Supercomputers are massive computing machines used for complex calculations for a wide variety of scientific applications. Supercomputers around the world are ranked for performance, based on floating point operations per second (flops), on the Top500 list. This list has been maintained since 1993. [1]

*mmcdonn1@utk.edu
Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN 37996-2200

Yet, with increasingly larger supercomputers being built to compete for the top rank on the list and a performance-at-any-cost attitude, vast amounts of power are being consumed to both provide computational calculation and also cooling from the heat generated for the supercomputer. From 2000 to 2006, the energy consumed by the large-scale computing sector (including supercomputers, data centers, etc.) doubled and accounted for 1.5% for the global energy consumption in 2007. [2, 3] With this increase in energy consumption, one must wonder, what are the environmental impacts of supercomputers?

Large-scale computing's environmental impact has become a growing interest for the EPA recently. [4] Metrics for the environmental impact of large-scale computing tend to focus on the energy-efficiency of the operational computing and the cooling infrastructure. In a work by Shah et al [5], a rigorous life cycle assessment (LCA) of data centers provides an overall environmental impact analysis to signify the need for more than just operational energy use (for computation and cooling) to aid in future sustainable data center design. Yet, data centers differ significantly in their internal computational components and function from supercomputers. Data centers are geared towards server-side computing, large-scale storage, and database applications. Supercomputers are designed for computational performance and speed of calculation applications. Boyd et al in IEEE in May 2011 [6] show that the environmental impact and overall LCA of high-performance CPUs (HPCs) used in supercomputers is much higher compared to intermediate-performance CPUs of other machines, such as data centers. This difference implies manufacturing and operational environmental impact for a supercomputer would be higher compared to a data center.

Efforts to capture the sustainability of supercomputing has not yet provided a rigorous LCA where beginning-to-end life-cycle impacts are determined (more in depth discussion will be in section 4). In November of 2007, the first publishing of the Green500 list began, a list of the Top500 list supercomputers only with the metric of flops per watt used.[7] Currently, the Green500 is possibly the largest effort to direct the evolution of supercomputing towards sustainable design. The flops per watt metric provides a comparison of energy-efficiency for supercomputers but does not address life-cycle impacts. Further, with increasing incorporation of general-purpose graphical processing units (GP-GPUs or just GPUs) and co-processors (such as Intel's Xeon Phi MIC) in supercomputers, the flops per watt has increased tremendously. Supercomputers incorporating GPUs and co-processors are at the top of the list for November 2012 (the top four spots all have GPUs or co-processors). [7] Yet, from the previous study mentioned [6], more advanced CPU processors using 32nm processing have a much larger manufacturing and operational environmental impact than any other CPU processors, such as 45nm processing. GPUs and co-processors of current manufacturers use anywhere from 28nm to 20nm processing for their processors,a much more material and energy intensive process than the 32nm process. [8, 9]One could reason that GPUs and co-processors have a much larger environmental impact. Thus, supercomputers ranked at the top for energy-efficiency during operation use the GPUs and co-processors with the largest environmental impact!

Supercomputer assessment with economical, environmental, and societal impacts considered would provide an optimal future supercomputer design that is sustainable. Some of the benefits of such an assessment would help determines the complete impact for bringing a supercomputer online, determine impact “hot spots” in design, and determine optimal locations for construction.

2 Background

The definition of a supercomputer is one that “...leads the world in terms of processing capacity, speed of calculation, *at the time of its introduction*.” [10] Emphasis on the time of introduction is because supercomputers, although ten to hundred million dollar assets and made up of hundreds of thousands of processor cores, only remain valid for about 5 years. From 2008 to 2009, Los Alamos National Lab’s supercomputer, Roadrunner, held the number 1 spot on the Top500 supercomputers in the world and was the first petaflop machine. Now, Roadrunner is now obsolete and being decommissioned. [1, 11] Supercomputers have short life times for such large amounts of semiconductor material used for manufacturing (notorious for high environmental impact [12]) and have a very large energy consumption. To begin an environmental impact analysis of a supercomputer, one must know what makes up a supercomputer.

At the heart of a supercomputer is hundreds of thousands of processors (This is for our current supercomputers. In 1993, the number 1 supercomputer, CM-5/1024, had only a thousand processors and the number 5 supercomputer, SX-3/44R, had only four!). [1] The processors of today that appear in supercomputers are mainly 32nm processed chips (32nm process refers to the lithographic process of manufacturing, where 32nm refers to the half-pitch, the distance in a memory cell from contact to the “bridge”, the gate that bridges between two contacts) but 45nm process chips still exist in supercomputers on the Top500. Yet, only taking into account the supercomputers with Intel and AMD 32nm processed chips gives us 81% of supercomputers on the Top500 list. Thus, as a majority, we can accurately look at current supercomputers’ processors as mainly consisting of 32nm processed processor chips (CPU chips).

GPUs and co-processors (or accelerators) are a new, emerging technology that currently has been introduced to the supercomputer community. According to the November 2012 Top500 list, less than 13% of all supercomputers incorporate some kind of GPU or co-processor into its hardware architecture. The benefit of both is the high level of multi-threading (the smallest level of program instructions to carry our computational arithmetic). A higher level of dissemination for computational calculation is accomplished using this emerging technology. This appears to be the direction for future supercomputer architecture. Yet, again, we note the fact that these new GPUs and co-processors are predicted to have a higher level of environmental impact even though they provide greater energy-efficiency for higher ranking on the Green500 list.

A supercomputer places CPUs (and GPUs or co-processors) on a node card. The number of processors on a node card depends on their size, which is usually dependent on the number of cores on the processor. These node cards are then placed in a blade. A blade is a structure that can hold multiple node cards, allowing them to be connected to each other using cables for intercommunication. This is where the dissemination of calculation becomes important. For supercomputers of today, calculations are assigned to individual processors and then to cores on the processor. This provides a much shorter latency (delay in time experienced by a system) for computation, the reason supercomputers are in existences in the first place. Thus, node cards are placed in close proximity on a blade so as to reduce latency via interconnect.

Blades are then placed in racks that make up cabinets (the size of a small closet). If a supercomputer is so large (as most today on the Top500) that it requires more than one cabinet, then multiple cabinets are interconnected. Cabinets are simply structures that are able to hold blades and are interconnects via cables to also allow dissemination of information for computational calculation. The interesting difference between node card structure and blade structure compared to cabinet structure is that cooling capabilities must be taken into account at this spatial-level of design. At this level, we must consider removing waste heat inside via a working fluid that comes in to cool the processors performing computations. More details on this spatial design will be discussed in the model design section. Again, these cabinets are connected via many feet of cable to provide this interconnection.

This infrastructure described makes up the computational component of our supercomputer but, we also have additional infrastructures that must be included to the supercomputer to make it operational. First, we must be able to store the data that is generated from computation. A storage infrastructure must be in place to provide data storage. Surprisingly, even though laptop computers provide a half to a full terabyte of data storage, supercomputers only have 10 to 50 petabytes (10,000 to 50,000 terabytes) of data storage. The key is usually in the data transfer from the computational nodes to the storage drives. Even though solid-state drives provide a faster transfer rate, this is not usually the method for transfer. A series of hard-disk drive transfers are carried out to transfer data from a compute node to a storage system/node. Once on the storage system or storage node, data transfer begins to the user's remote storage. Thus, there is not a large data storage need on the supercomputer itself, just enough for temporary storage.

Second, the operation of a supercomputer puts off large amounts of waste heat. This heat must be removed for continuous operation of a supercomputer. Therefore, large cooling infrastructures must be put into place in order to keep the supercomputer's internal processors at a tolerable temperature. For smaller supercomputers, or clusters, passive cooling may be utilized. In passive cooling, no working fluid is passed with an external flux across the processors to cool them, only natural diffusion of air. Usually in passive systems, the cabinets are completely open in design to allow air to move freely in and out. Once a given power per cabinet has been surpassed, closed cabinets must be used in order for proper cooling. This requires there to be a cool side (where the processors intake the cool working fluid) and a hot side (where the working fluid that has absorbed the waste heat is pushed) for the cabinet. Typically, the cabinets are set up in an aisle fashion to easily have access to the cabinets for maintenance and also separate the hot and cold aisles. The cabinets are closed so that the waste heat in the hot aisle does not mix with the incoming cold working fluid in the cold aisle, thus reducing the incoming temperature and capacity to remove heat. When we begin to utilize a separation of aisles and use closed cabinets, we must choose our working fluid that will be used to remove the waste heat. Air is the simplest and cheapest method for small to medium power loads. This system can be as simple as passing room temperature air over the processors and cooling the air via a small heat exchanger or using much larger heat exchangers, called computer room air conditioners (CRACs). CRACs normally use water as a working fluid and are usually coupled to other heat exchanger systems, where the final heat transfer is to an outside heat sink. With either the air or air/water systems, other cooling infrastructure supplements can be installed. Floor brushes around wires allow for less heat to be passed into the room from the outside. Blanking panels for blade slots that are empty in cabinets can be installed as well to further isolate the hot and cold aisles. Raised floors that carry the working fluid to the cabinets from underneath can be installed in the room, allowing further isolation of the cold working fluid. This also helps with humidity control, which is also a major issue in housing supercomputers. Chimneys can be fitted onto the top of cabinets to completely isolate and remove waste heat and reduce the need for a hot aisle. Finally, walls and doors can be installed around the hot and cold aisles to create hot and cold rooms, used to further isolate the two different temperature regions.

Finally, other infrastructures or components of infrastructures that are important are the building that houses the supercomputer and also the heat exchangers, generators, back-up generators, and power distribution needed. These are normally comparable to the same equipment needs of data centers and therefore are not taken into account in this study.

With such a large amount of materials that go into producing a supercomputer that only will be relevant for 5 years, question are raised about what the environmental impacts are from manufacturing. This study hopes to encompass not only the operation of the supercomputer but also take into account the manufacturing components of a supercomputer. A detailed account of the total environmental impact on manufacturing could help direct attention to materials that could be replaced or supplemented for more sustainable materials.

3 Metrics

The basis of this study is to determine life cycle inventories (LCIs) of individual components that make up a supercomputer during manufacturing, determine their LCA, and to use economic input-output LCA (EIO-LCA) for dollars spent in operation to determine economical, environmental, and societal impacts (or just sustainable impact). This is done by implementing a model that allows for complete design of the supercomputer (to be discussed in more depth in section 5) and calculates the entire sustainable impact. The metrics of this model include:

- Cost of individual components and collective end-item
- Energy needs
- Pollutants
  - Greenhouse Gases
  - Emissions
    - Air
    - Water
    - Soil
- Radioactivity
  - Air
  - Water
- Toxic Releases
- Societal Impacts

All of the metrics above are available in the model outputs but not all will be presented in this paper.

Another metric that is implemented implicitly in the model is determining the power usage effectiveness (PUE) of the entire supercomputer infrastructure (the building that houses it, lights, water, etc.). This is used to determine how much of the power going into the supercomputer infrastructure is actually used for computational work. [13] This metric is used to determine the overall power from the computational power need and the cooling infrastructure chosen to cool the supercomputer. This is an approximation of the model but hopes to capture the overall cost within reason based on the guidelines set forth by Pentair Industrial Technologies in design of data center cabinet layouts. [13]

4 Life Cycle Assessment Introduction

Life cycle assessment (LCA) tries to capture the "cradle-to-grave" environmental impact of the system of interest. One must first compile an inventory of all products and processes related to the system, usually defined as the life cycle inventory (LCI). This includes raw materials extracted, manufacturing materials added along the life cycle chain, and carries out until the end-of-life of the system. LCA also tries to capture end-of-life scenarios, such as direct disposal with no recycling to complete recycling of every material possible. Using certain software and databases, predetermined impact factors for environmental impacts can be used for analysis throughout all different phases of the life cycle for the system. [14, 15, 16] Yet, different ways to carry our life cycle assessments exist. Each have advantages and disadvantages, usually related to the detail of the the analysis. Very rigorous LCAs are time-consuming and can seem infinite in their inventories, providing troublesome compilation of data and analysis. Less rigorous detail in the LCA can group a sector rather than a specific product, giving up differentiation between specific products, but are less time and work intensive and can sometimes be extended to capture a much larger boundary than the detailed LCA.Various LCA types are provided:

PROCESS LCA Most rigorous, compile a life cycle inventory, determine environmental impact factors for material inputs and outputs, and weight results

ECONOMIC INPUT-OUTPUT LCA (EIO–LCA) Document monetary relationships between different sectors of economy. Average environmental impact given for a good within a given economic sector. Economic impact determines the environmental and societal impacts.

HYBRID EIO-LCA Manually adjust the impact of different sectors to model a specific good within the economic sector more accurately

The two LCAs used for this study will be process LCA and EIO-LCA. Further discussion of the EIO-LCA can be found elsewhere. [5, 17]

5 Model for Supercomputer Design

5.1 Model Overview

We have already stated that there is no supercomputer LCA currently available and that the data center LCA conducted by Shah et al [5] does not capture the differences between data centers and supercomputer internals. Thus, presented here is an initial effort to develop a model that incorporates the three important aspects of sustainability for supercomputer design. This model hopes to be a tool that allows ease-of-update to stay current with the ever-changing semiconductor industry as new data on the materials becomes available to refine the model and capture future design. This could be useful for government and industry who plan to implement a supercomputer and would like to know

- The complete impact of bringing a supercomputer online
- Highlights “hot spots” for high-impact in a specific sector/infrastructure
- Address these “hot spots” by redesign
- Determine optimum location for construction based on needs/impacts of the supercomputer

For our model, we incorporate three separate infrastructures or boundaries for our LCA analysis and economic modeling. These three are the compute, cooling, and building infrastructures, which are further broken down into individual components that make up that infrastructure. Fig. 1 shows a diagram for the three boundaries considered in this study. Components are located directly underneath their respective infrastructures with arrows annotating linked components between infrastructures. For example, a raised floor may need to be constructed in the room (building infrastructure) due to the cooling needs (cooling infrastructure).

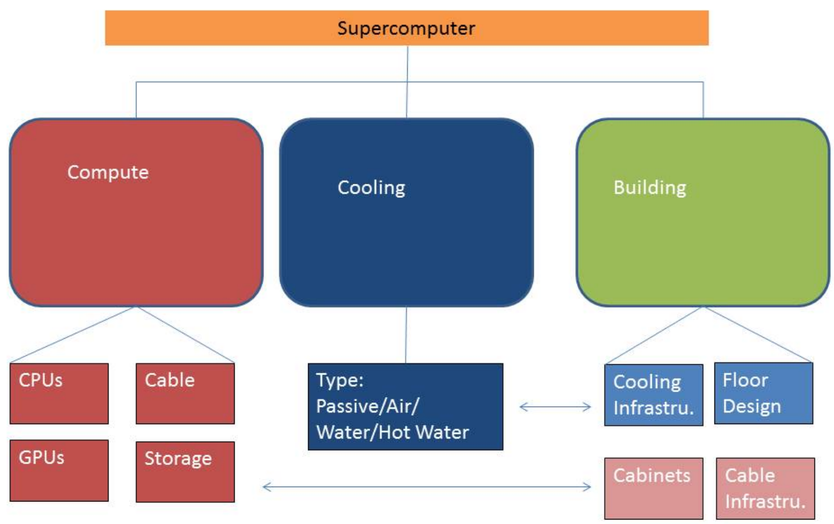

Figure 1: The infrastructures that make up the supercomputer of our model

For the model, we only have a subset of the products available for the supercomputer components. Thus, Fig. 2 presents the choices available for the model for processors, GPUs, and co-processors. The economic impact is determined based on prices found from the vendors websites. [8, 9, 18] Yet, the literature and professional contacts have confirmed that processors are sold from the vendors to the supercomputer developer for discount prices for buying in bulk. [19] Based on literature and comparing the model outputs to actual supercomputer costs, a 45% discount for the processors, co-processors and storage gives an accurate price for the compute infrastructure. The environmental impacts for the processors are based on the most current microprocessor process LCA available. [20] This detailed LCA allows the differentiation between the data center compared to the supercomputer for manufacturing-based environmental impacts. The GPU and co-processor's environmental impacts are not available in the literature because no LCI or LCA study has been carried out for GPUs or co-processors. Using literature to determine the markup on manufacturing of 22nm and 28nm processing for the GPUs co-processors [21, 22] and using a best, middle, worst case scenario forecasting, the GPU and co-processor's environmental impact is based on the 32nm processing with a 3%, 15%, and 30% markup on all materials, respectively. This is a gross approximation due to the fact that core materials may be the only ones that are used more abundantly but, unable to determine which materials needed the markup specifically, all were increased for simplicity.

Processors		
Manufacturer	**Model**	**Fabrication (nm)**
Intel		
	Xeon 5000 Series	32
	Xeon E5	32
	Xeon E7	32
AMD		
	Intergalos	32

GPUs and Co-processors		
Manufacturer	**Model**	**Fabrication (nm)**
Nvidia		
	K20/K20x	28
	2000 series	40
Intel		
	Xeon Phi	22

Figure 2: Processors and Co-processors chosen for options in the model

The storage choices are 1 terabyte versions of solid state drives (SSD) and hard disk drives (HDD). Both are priced for server-style variants of the drives. The SSD is priced at $3/gigabyte and the HDD is priced at $1.6/gigabyte. The environmental impact for both comes from the LCA study carried out by Boyd et al. [23]

The cable network selections are copper and fiber optic cables. The cables where priced using the Infiniband variants [24] of each, which makes up about 45% of all supercomputer cable networks on the Top500. [1]The fiber optic and copper cables were built using the ecoinvent database [14] and modeled after the LCA provided in the article by Unger and Gough. [25] The jacketing of the cable also used the EPA's LCA data when not available in the ecoinvent database or article. [26]

The cooling infrastructure is given the four choices of cooling types, given in Fig. 1, as well as size choices of CRACs based on power load per cabinet. The building infrastructure choices are closely correlated to the cooling infrastructure design and size. Choices include floor design, layout of the floor for area and space between aisles, and various cooling components mentioned in section 2 (floor brushes, chimneys, etc.). [13] The cooling and building infrastructure's environmental impacts are modeled using the EIO-LCA. [17]

The phases of the life cycle that the model captures are the manufacturing, the usage/operation, but not the end-of-life scenario. This will hopefully be updated after more research to determine different scenario options and their impacts. The manufacturing captures the infrastructures of:

- Compute
  - Processors and Co-processors
  - Storage
  - Cable Network
  - Cabinets
- Cooling
  - CRACs
  - Extraneous Environmental Control Equipment (piping, thermostats, wiring, etc.)
- Building
  - Room design (Raised floors, separated aisles, etc.)

The usage or operation phase consists of:

- Compute
  - Processors and Co-processors
  - Storage
- Cooling

The usage or operation phase environmental impact is determined using the EIO-LCA for both the computer and cooling operations.

Just as important as it is to say what a model does include is to say what it does not include. This model does not take into account the transportation of the components to the location of construction for the supercomputer. This is due to the model of the supercomputer not taking into account location dependence. This can be considered after the sustainable impacts are determined without location dependence. The information needed would be the location of the supercomputer, the location and path taken for the components shipped, and the various forms of transportation used to deliver the components of the supercomputer. None of this information is in the model currently. It has already been stated but no disposal evaluation is incorporated in the model either. The model takes into account the total number of years of operation but not the impact of demolition, dumping, or recycling of the materials.

Also important are major assumptions that the model is based on. First, the EIO-LCA database uses 2002 data for both the manufacturing and usage/operational phases. This implies that the energy provided is based on 2002 energy percentages from given sectors. This may cause a higher environmental impact due to a cleaner energy source being provided to today's current electric grid. Second, some of the materials for the cables have been left out due to not being included in the ecoinvent database. This is a small assumption since no more than 10% of material for either type of cable was left out, implying that a good approximation of environmental impact is captured. Third, the co-processors have only been approximated based on literature markups for manufacturing costs. This is a gross approximation that has already been discussed but, in order to give an accurate range, we have provided a best, middle, and worst case scenario. Fourth, discounts for the compute infrastructure are not exact and only approximations from literature and from fitting data to the actual prices for the entire supercomputer. If no discount is applied, the economic model over-approximation is far too large to justify no discount. Fifth, the fact that an EIO-LCA is used implies that averaging for a given sector is provided and not the exact product we are trying to model. Sixth, the data sources can quickly become out-of-date and some may already be out-of-date. The model was designed from the beginning to accommodate new data sets being included and has a very modular structure. Therefore, simply adding a few lines of code to the model to access the new data provided allows the model to stay current. A key note is that this model intends to provide a large-scale comparison to signify sections of high impact. We believe that relaxing these assumptions will not significantly impact the data and that our model is still of high fidelity.

5.2 Model Flow

We have explained the basis of the model and now we look at the actual flow for design of the supercomputer within the model. The overview of the model is shown in Fig. 3. The double-arrow to the cooling infrastructure implies that input information is provided by the user but is guided by information from the model, coupling the cooling infrastructure input with the original user input. Here we hope to not only explain the inputs and outputs shown in the figure but, the inner-workings of the how the inputs are used to design and build a supercomputer. One important note is that other outputs are generated due to the process-LCA and the EIO-LCA giving different information for environmental impacts. These are cross-referenced to determine the total output for the user.

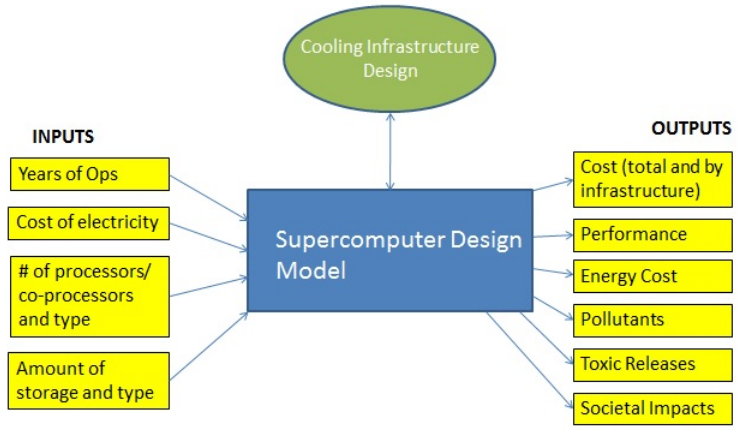

Figure 3: Model Overview

The user initially inputs the years of operation of the supercomputer (life span), the cost of electricity in the area, number and type of processors, the percentage of these that will be GPUs or co-processors, and the amount and type of storage. From this, the model first determines the total cost of the processors (and GPUs/co-processors, if included) with and without a discount (the discount can be adjusted as well). Then, the number of node cards is determined. The type of processors and GPUs/co-processors determines the amount located on a node card due to different numbers of cores per type.The number of blades and cabinets is determined based on a Cray cabinet setup. [27] This determines the floor space needed and the power per cabinet. The number of unoccupied blades and total power of the system is computed as well. The cable needs are calculated based on the floor space area to interconnect all of the cabinets and the percentage of copper and fiber optic cable is determined based off the length. Copper cable is optimal up to three feet for interconnection but needs a booster for greater lengths. Fiber optic cable is used for greater lengths. The model assumes no boosters are used and that all copper cable is kept under three feet in length. The reasoning is that the booster would not only decrease performance but also increase the environmental impact. [25] Similar reasoning is used to not use fiber optic cable throughout the entire cable network. Fiber optic cable is more expensive, has a higher environmental impact, and does not perform as well under 3 feet as copper cable does. [25] The total cost of the cable network is determined from the total length needed and percentage of copper and fiber optic cable. The storage cost is calculated directly from the user's input. The cooling infrastructure takes into account the power per cabinet to determine design requirements and limitations. Based on these requirements, the user is given questions for various cooling infrastructure additions to reduce the PUE. The additional cooling infrastructure increases the capital cost of the supercomputer but helps reduce the operational cost for cooling. The final PUE is based on literature for designing cooling infrastructures for data centers. [13] These user inputs determine the final cost of manufacturing and operation of the cooling infrastructure. The PUE is then used to determine the overall power usage and cost for operations of the supercomputer.

The final outputs are then provided in various forms (on-screen output, tables, graphs, etc.) to give the user the outputs noted in Fig. 3.

6 Application of Model

Three current top ten supercomputers have been modeled, specifically #1 (Titan), #6 (SuperMUC), and #8 (Tianhe-1A), to determine the cost, energy input, global warming potential, and multiple societal impacts for the life-cycle of the supercomputer for 1 year of operation. Fig. 4 shows the results of the model outputs, separated into the manufacturing and use/operation phase for each. For validation, multiple variables are taken into account. The Titan model manufacturing cost of $100 million matches almost exactly with the $100 million for actual capital cost. This is due to the discount fitting was bench-marked for the Titan, initially. The model output for operational cost of $14.5 million over-approximates the true operational cost of around $10 million. This could be due to the model assigning a higher PUE than the actual PUE for Titan. From a study by Lawrence Berkeley National Lab, Titan has a surprisingly low PUE of 1.83, lower than the model could predict. [28] The model currently using the cooling infrastructure to determine the PUE but, other energy-saving alterations can be implemented in the buildings infrastructure to offset the indirect power cost to reduce the PUE (solar panels, lower-energy lighting, "smart" thermostats, etc.). The predicted power is 8.57 MW and the total cabinet count is 194 from the model, close to the 8.2 MW and 200 cabinets for the true Titan.

The model predicts that the SuperMUC will cost approximately $45.5 million for manufacturing and $6 for annual operation, compared to the $68.5 million for manufacturing and approximately $8 million for actual annual operation. [29] The model over-approximates the the power of SuperMUC by 1.5 MW, suggesting that the model's PUE for the hot-water cooling system is not quite accurate yet. Further research and data to compare with model outputs is needed to accurately capture the total power consumption in a reasonable fashion.

The Tianhe-1A costs $88 million to build and approximately $20 million for annual run time. This dwarfs the model predictions and even when the discounts are relaxed, there is still a large gap. Yet, the model matches the number of cabinets (112 compared to the true 120) and the power consumption very well (4.09 MW compared to the true 4.04 MW). [30]

Fig. 5 shows the percentage of the cost, energy, and global warming from fig. 4 based on the infrastructure and then also the phase of the life-cycle for the cooling and compute infrastructures. We see that the largest cost comes from the purchase of compute components but the largest environmental impacts are coming from the cooling and compute operation phase. This implies that the

		Titan			SuperMUC			Tianhe-1A		
		Best	Middle	Worst	Best	Middle	Worst	Best	Middle	Worst
Cost	manufacturing	$99,679,395.03			$45,472,939.32			$34,929,646.66		
	operations (1 yr)	$14,446,599.55			$6,318,098.77			$6,889,817.09		
Energy (MJ)	manufacturing	256,017,990.89	256,053,461.60	256,097,799.98	152,809,039.49			114,143,383.67	114,156,988.87	114,173,995.37
	operations (1 yr)	1,603,572,550.27			701,308,963.64			764,769,696.77		
GWP (metric tons CO	manufacturing	16,848.93	16,849.20	16,849.53	10,015.25			7,498.14	7,498.24	7,498.36
	operations (1 yr)	135,364.64			59,200.59			64,557.59		

Figure 4: Cost, Energy Consumption, and Global Warming Potential for 3 Supercomputers: Titan, SuperMUC, and Tianhe-1A

source of electricity is important for providing a sustainable supercomputer. This could help determine which part of the national grid a supercomputer should be constructed near, to reduce the total environmental impact. Also, this could help motivate our grid to provide cleaner energy in the future. We see that the SuperMUC increases its cost for the cooling infrastructure and in return, reduces its percentage of energy and global warming potential for the cooling operations. This implies that using a hot-water cooling system like the SuperMUC's Aquasar [31] could also help provide a sustainable supercomputer design. Other positive aspects of having a hot-water cooling infrastructure are not currently included in the model. Such aspects are being able to reduce the temperature difference in the inlet for the working fluid and thus the working fluid can be cooled at a higher temperature as well. This waste heat can be used to heat near by buildings and still reduce the temperature of the fluid to be used as a coolant. [31] Future development of the model could add energy credit for using waste heat for another purpose, thus reducing the overall cost of the power needs.

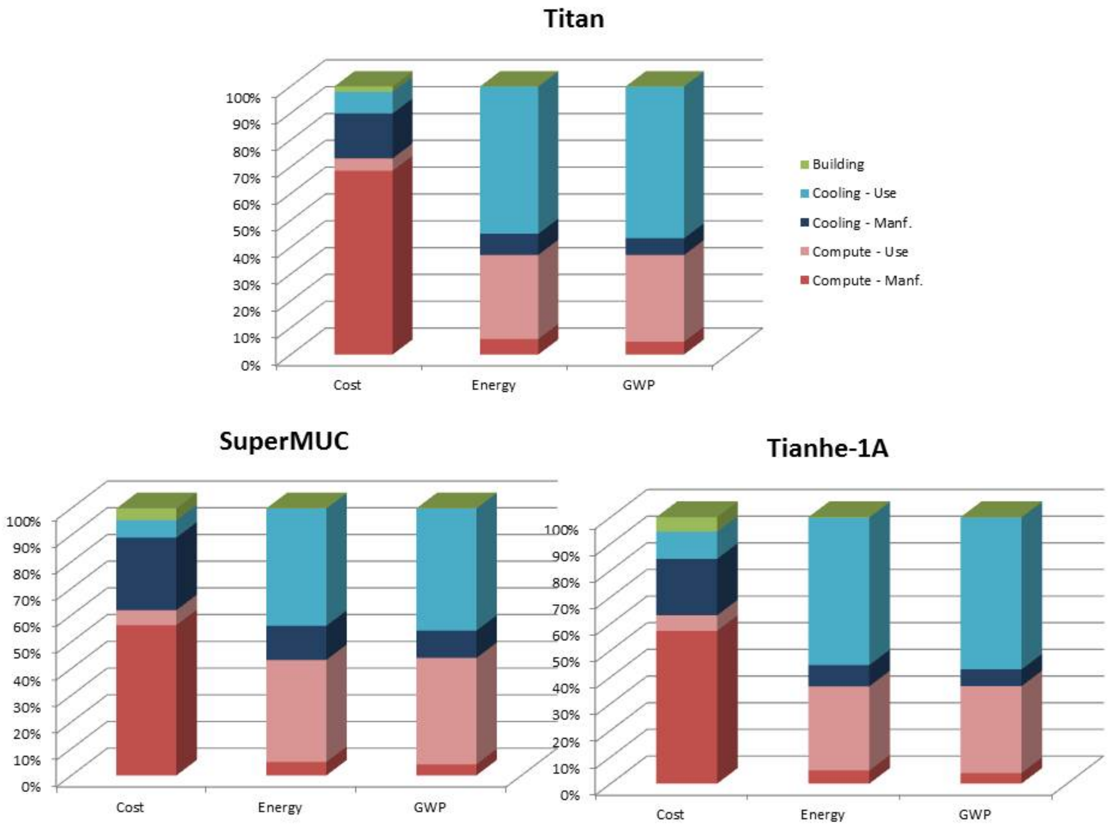

Figure 5: The Percentage of Cost, Energy, and Global Warming Potential by Infrastructure and Phase for 3 Supercomputers

Fig. 6 provides a look a the societal aspects of supercomputer design. Fig. 6 gives a log plot of three societal effects from a supercomputer's life-cycle for our three supercomputers of interest. The three societal effects are increased air acidification (given as kilograms of SO_2 equivalent), ozone depletion (given as kilograms of trichlorofluoromethane equivalent), and ranged approximations for increase in cancer for human health (given as kilograms of benzene equivalent). For both Figs. 6 and 7, compute manufacturing and the building infrastructures are not included because the process-LCA data has not been integrated to give a head-to-head comparison. Further data research and integration must be completed before any comparison can be shown. The figures below still allow accurate comparison of the the cooling use and manufacturing and compute use phases and infrastructures.

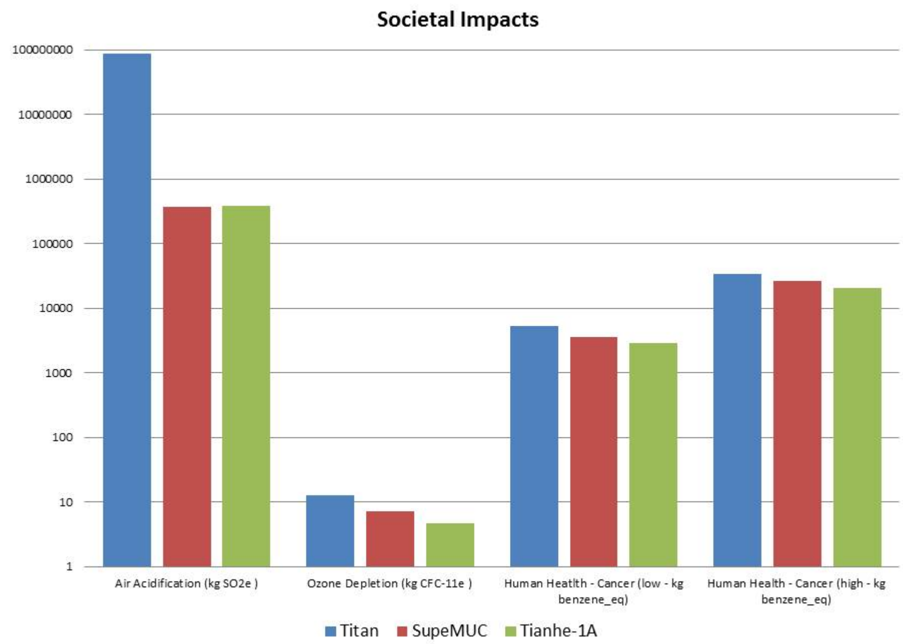

Figure 6: Societal Impacts of 3 Supercomputers for the Cooling Manufacturing and Use Phase and Compute Use Phase: Log Plot

Fig. 7 shows the percentage breakdown of fig. 6 for the two infrastructures and both phases of the cooling infrastructure. We see that the cooling and compute use phase give rise to the largest percent of air acidification, where the cooling manufacturing phase gives rise to the largest contributions to ozone depletion and increase in cancer.

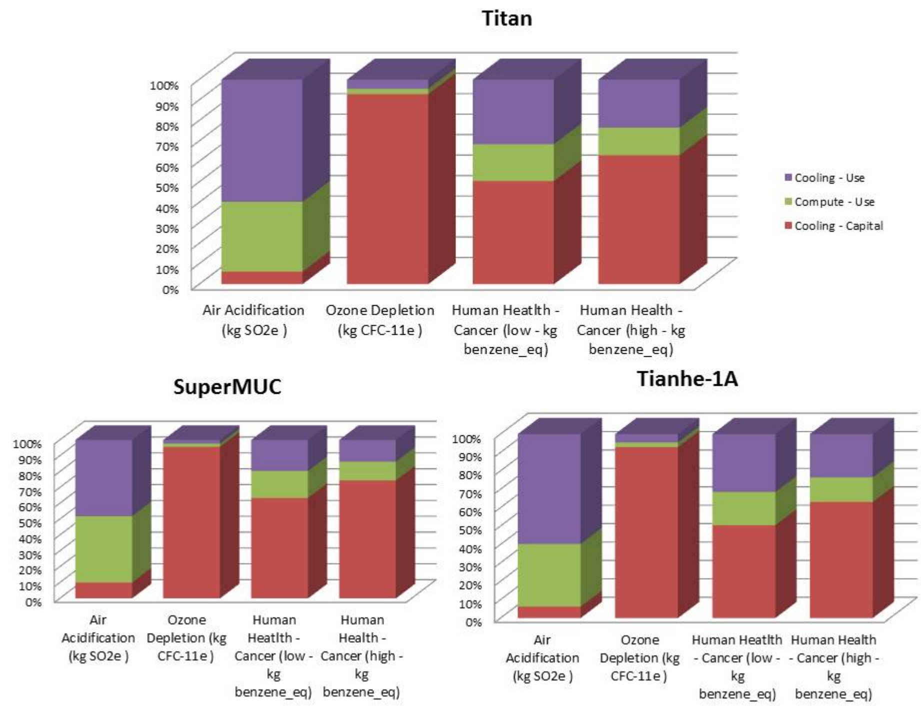

Figure 7: Percentage of Societal Impact per Infrastructure and Phase for 3 Supercomputer

7 Conclusion

By combining process-LCA and EIO-LCA methods, an initial model for in-depth sustainability analysis of supercomputer design has been developed and the results shown above. The model provides comparison of economic, environmental, and societal impacts for manufacturing and operation of not only a single supercomputer but a range of designs. This model has expanded the effort to direct supercomputers towards a sustainable design using motivation from the Green500 initiative and the data center LCA study by Shah et al. [5] The model hopes to be an ease-of-use, ease-of-update, and comprehensive tool that can keep up with the ever-changing semiconductor industry that is vital to future supercomputer design. The ultimate goal of this model is to assist government, academia, and industry in supercomputer development for existing and future supercomputers. The advantages of the model are determining the entire impact before the initial construction begins, determine high-impact sections of the supercomputer's design, and to direct construction to optimal locations.

8 Acknowledgments

I would like to acknowledge Dr. David Keffer and Dr. Paul Frymier for guidance, construction criticism, and instruction that led to the completion of this project. MM was supported by a grant from the National Science Foundation (DGE-0801470).

References

[1] The Top500 List. www.top500.org

[2] Taheri J., Zomaya, A. Y. Energy Efficiency Metrics for Data Centers. Wiley & Sons, pp. 245, 2012.

[3] Koomey, J. G. "Estimating Total Power Consumption by Servers in the U.S. and the World," Technical Report, available hightech.lbl.gov/documents/data_centers/svrpwrusecompletefinal.pdf

[4] EPA. "Energy Star Program Requirements for Computer Servers," US Environmental Protection Agency, Washington D. C. 2008.

[5] Shah, A., Bash, C., Sharma, R., Christian, T., Watson, B. J., Patel, C. "Evaluating Life-Cycle Environmental Impact of Data Centers," J. of Electronic Packaging, 133, pp. 031005-(1-9). 2011.

[6] Bol D., S. Boyd, D. Dornfeld, Application-Aware LCA of Semiconductors: Life-Cycle Energy of Microprocessors from High-Performance 32nm CPU to Ultra-low-power 130nm MCU, IEEE 2011.

[7] The Green500 List. www.green500.org

[8] Intel. Xeon Phi data sheet. Retrieved from Intel website – ark.intel.com.

[9] Nvidia. Tesla and Kepler data sheet. Retrieved from Nvidia website - http://www.Nvidia.com/object/tesla_product_literature.html

[10] Thorne, M. Lecture on Geophysical Supercomputing, University of Utah, August 2010.

[11] Brodkin, J. "World's top supercomputer from '09 is now obsolete, will be dismantled". Ars Technica. Retrieved May 1, 2013.

[12] Williams, E., Energy Intensity of Computer Manufacturing: Hybrid Assessment Combining Process and Economic Input – Output Methods, Environ. Sci. Technol. 2004, 38: pp. 6166-6174.

[13] Pentair Technical Products. "Overview and Design of Data Center Cabinets". Retrieved April 3, 2013 from http://www.graybar.com/documents/hoffman-overview-design-data-center-cabinets.pdf.

[14] Frischknecht, R., Jungbluth, N., Althaus, H. J., Doka, G., Dones, R., Heck, T., Hellweg, S., Hischier, R., Nemecek, T., Rebitzer, G., Spielmann, M. "The ecoinvent Database: Overview and Methodological Framework," Int. J. Life Cycle Assess. 10(1), pp. 3-9, 2005.

[15] Jolliet, O., Margni, M., Charles, R., umber, S., Payet, J., Rebitzer, G., Rosenbaum, R. "Impact 2002+: A New Life Cycle Impact Assessment Methodology," Int. J. Life Cycle Assess., 8(6), pp. 324-330, 2003.

[16] Goedkoop, M. and Spresnsma, R. "The Eco-Indicator 99: A damage oriented method for life cycle assessment," Technical Report, PRe Consultants, The Netherlands, 2001.

[17] Hendrickson, C. T., Lave, L. B., and Matthews, H. S., Environmental Life Cycle Assessment of Goods and Services: An Input-Output Approach, Resources for the Future, Washington D. C., 2006.

[18] AMD. AMD Interlagos data sheet. Retrieved from AMD website – products.amd.com

[19] Scalable Computing Laboratory website. "Designing a Computer Cluster,", Ames Laboratory and Iowa State University, Retrieved May 1, 2013.

[20] Schmidt, M., Hottenroth, H., Schottler, M., Fetzer, G., Schluter, B. "Life Cycle Assessment of Silicon Water Processing for Microelectronic Chips and Solar Cells," Int. J. Life Cycle Assess., 17, pp. 126-144. 2011.

[21] Farber, R. "Positioning x86 Petascale Performance with MIC Architecture: Intel's reputation is now on the Line to Demonstrate that Many Integrated Core architecture can compete against GPUs," Sci. Computing, Feb. 2012.

[22] Tan, W. L. "What You Need to Know to Establish a Digital Methodology for 28nm ARM Core-based SoCs," Strategies & Methodologies, pp. 50-52, 2011.

[23] Boyd, S., Horvath, A., Dornfeld, D. "Life-Cycle Assessment of NAND Flash Memory," IEEE, 24(1), pp. 117-124, 2011.

[24] HP website. Intel/Mellanox's Infiniband price listings. Retrieved from http://h30094.www3.hp.com

[25] Unger, N., Gough, O. "Life Cycle Considerations About Optic Fibre and Copper Cable Systems: A Case Study," J. of Cleaner Prod., 16(14), pp. 1517-1525, 2008.

[26] EPA. "Wire and Cable Insulation and Jacketing: Life-Cycle Assessments for Selected Applications," US Environmental Protection Agency, Washington D. C. 2008.

[27] Cray Supercomputer Company. Cabinet data sheet. Retrieved from http://www.cray.com/Products/Computing/

[28] Oak Ridge Leadership Computing Facilities website. Retrieved from http://www.olcf.ornl.gov/computing-resources/ of May 1, 2013.

[29] Aventin, M. L. "SuperMUC Progress Presentation," Technical University of Munich website Retrieved from http://www.lrr.in.tum.de/~gerndt/home/Teaching/Parallel%20Programming/Student12/AventinSuperM 2012.

[30] Yutong, L. "HPC Systems in NUDT," National University of Defense Technology, China, 2010.

[31] Zimmermann, S., Meijer, I., Tiwari, M. K., Paredes, S., Michel, B., Poulikakos, D. "Aquasar: A hot water cooled data center with direct energy reuse," Energy, 43, pp. 237-245, 2012.