
Matthew Stachoni
Microsol Resources

Learning Objectives

- Understand the relative computing demands of each of the applications in Building Design Suite
- Review the current state of the art and sweet spots in processors, memory, storage, and graphics
- Learn to specify workstations for different classes of BIM usage profiles
- Learn how to best shop for complete systems and individual components

Description

Working with today's Building Information Modeling (BIM) tools presents a special challenge to your IT infrastructure. Wrestling with the computational demands of the Revit software platform—as well as other BIM-related applications such as 3ds Max, Showcase, Navisworks Manage, Lumion, Rhino, and others—means that one needs the right knowledge to make sound investments in workstation and server hardware. Get inside the mind of a certified hardware geek to understand the variables to consider when purchasing hardware to support the demands of Autodesk, Inc.’s, suite of BIM and 3D applications.

Fully updated for 2015, this class will give you the scoop on the latest advancements in workstation processors, motherboards, memory, and graphics cards. This year we will pay special attention to the latest advancements in graphics technologies to meet the demands of high-end rendering, animation, and visualization across a wide variety of platforms.

Your AU Experts

Matt Stachoni has over 25 years of experience as a BIM, CAD, and IT manager for a variety of architectural and engineering firms, and has been using Autodesk software professionally since 1987. Matt is currently a BIM Specialist with Microsol Resources, an Autodesk Premier Partner in New York City, Philadelphia, and Boston. He provides training, BIM implementation, specialized consultation services, and technical support across a wide array of Autodesk AEC applications.

Previously, Matt was the BIM and IT Manager for Erdy McHenry Architecture LLC and a BIM specialist for CADapult Ltd. (now Applied Software), where he provided on-site construction BIM modeling and coordination services for construction managers, HVAC, and MEP trade contractors for a number of projects. He is a contributing writer for AUGIWorld Magazine, and this is his 12th year speaking at Autodesk University.

e-mail: matt@stachoni.com
mstachoni@microsolresources.com

Twitter: @MattStachoni
I. Introduction

“Wow, this workstation is way too fast for me.”
– Said no one. Ever.

Specifying new BIM / 3D workstations tuned for Autodesk’s 3D and BIM applications found in the Building Design Suite (BDS) can be a daunting task with all of the choices you have. Making educated decisions as to where to spend money requires quite a bit of research wading through online reviews, forums, and working with salespeople who don’t understand what you do on a daily basis. Moreover, recent advancements in processors, graphics, and storage can challenge preconceptions of what is important.

Computing hardware had long ago met the relatively low demands of 2D CAD, but data-rich 3D BIM and visualization processes will tax any workstation to some extent. Many of the old CAD rules no longer apply. You are not working with small project files, and Autodesk applications do not exactly fire up on a dime. Individual project assets can exceed a Gigabyte as the BIM data grows and modeling gets more complex.

Today there is a large variety of ways to showcase BIM projects using high end but accessible visualization tools, so those users will require the most powerful workstations you can find. The capability and complexity of the tools in the Building Design Suite improves with each release. Iterating through adaptive components in Revit, or using the advanced rendering technologies such as the iray rendering engine in 3ds Max can bring even mightiest systems to their knees. Knowing how these challenges can best be met in hardware is a key aspect of this class.

Taken together, this class is designed to arm you with the knowledge you need to make sound purchasing decisions today, and to plan for what is coming down the road in 2015.

What This Class Will Answer

This class will concentrate on specifying new systems for BIM applications in the Autodesk® Building Design Suite, namely Revit®, 3ds Max®, Navisworks®, and Showcase®.

Specifically, we want to answer these fundamental concerns and questions:

- Identify the different user requirements which have specific hardware demands
- What are the latest AEC and IT industry trends and how do they affect your purchasing decisions?
- What are the different hardware platforms found in today’s BIM and 3D workstations?
- How much system RAM is appropriate? Where does it make a difference?
- What is the difference between a workstation graphics card and a “gaming” graphics card
- Are solid state drives (SSDs) worth the extra cost? What size is currently appropriate?
- Do I build my own machine or do I buy a complete system from a vendor?

To do this we will cover the five critical subsystems and review specific components to look for in every workstation:

- Central Processing Units (CPUs)
- Chipsets and motherboard features
- System memory (RAM)
- Graphics processors (GPUs)
- Storage

In addition we will look at what is new in operating systems; laptop and mobile platforms; peripherals like keyboards, mice, and monitors; and include an updated buying guide.
**Disclaimer**

In this class I will often make references and tacit recommendations for specific system components. This is purely my opinion, stemming largely from extensive personal experience and research in building systems for myself, my customers, and my company. Use this handout as a source of technical information and a buying guide, but remember that you are spending your own money (or someone you work for). Thus, the onus is on you to do your own research when compiling your specifications and systems. I have no vested interest in any manufacturer and make no endorsements of any specific product mentioned in this document.

**Identifying Your User Requirements**

The first thing to understand is that one workstation size does not fit all user needs. Whenever you specify a workstation for someone, you must understand their typical computing requirements. In general you can classify users into one of three use case scenarios which largely points to specifying a particular workstation class.

1. **The Grunts**: These folks use Revit day in and day out, but rarely step outside of that to use more sophisticated software. They are typically tasked with the mundane jobs of project design, documentation, and project management, but do not regularly create complex families, high end renderings, or produce extended animations. Revit is clearly at the top of the process-consumption food chain, and nothing else they do taxes their system more than that. However, they will grow over time, so their workstations need to handle at least some higher-order workloads without choking.

2. **The BIM Champs**: These are your BIM managers and advanced users who not only use Revit all day for production support, but delve into the nooks and crannies of the program to help turn the design concepts into modeled reality. They develop project content, create Dynamo scripts, manage models from a variety of sources, update and fix problems, and so on. They may regularly interoperate with additional modeling software such as 3ds Max, Rhino, and SketchUp, and pull light to medium duty creating visualizations.

3. **The Viz Wizards**: These are your 3D and visualization gurus who spend as much time in visualization applications as they do in Revit. They routinely need to bring Revit models into programs like 3ds Max, Maya, Infrastructure Modeler, Unreal Engine 4, and Lumion. These users will extensively use real-time ray tracing, specialize in photorealistic renderings and animations, and develop other advanced visualizations. The Viz Wiz will absolutely use as much horsepower as you can throw at them.

Ideally, each one of these kinds of users would be assigned a specific kind of workstation that is fully optimized for their needs. Given that you may find it best to buy systems in bulk, you may be tempted to specify a single workstation configuration for everyone without consideration to specific user workloads. I believe this is a mistake, because a Viz Wiz will quickly tire of waiting on hardware more appropriate for a Grunt, and it is too expensive to specify the highest-end machine for everyone.

Given the voluminous information in this handout, you should be able to effectively specify hardware that is targeted to these three user profile types and not worry about over- or under-allocating resources for new systems. Later in this handout we will establish three separate workstation configurations that target these specific user requirement profiles.
II. Industry Pressures and Key Trends

In building out any modern workstation or IT system, we need to first recognize the size of the problems we need to deal with, and understand what workstation subsystems are challenged by a particular task. Before we delve too deeply into the specifics of hardware components, we need to review five key industry trends which shape today’s state of the art and drive the future of BIM computing:

- Maximizing Performance per Watt (PPW);
- Realizing the promise of parallelism, multithreading, and multiprocessing;
- Leveraging GPU-accelerated computing and the importance of computer gaming;
- Increased adoption of virtualization and cloud computing;
- Tight price differentials between computer components

Taken together these technologies allow us to scale workloads up, down, and out.

Maximizing Performance per Watt and the Impending End of Moore’s Law

Every year Intel, NVIDIA, and ATI release new iterations of their hardware, and every year their products get faster, smaller, and cooler. Sometimes by a little, sometimes by a lot. A key design criteria used in today’s microprocessor fabrication process is to maximize energy efficiency, measured as Performance per Watt.

For years, the rate of improvement in integrated circuit design has been predicted quite accurately by Gordon E. Moore, a co-founder of Intel. Moore’s Law is the observation that, over the history of computing hardware, the number of transistors in an integrated circuit has doubled approximately every two years\(^1\).

\[ T = 10 \times 2^{(t/2 \times \text{years})} \]

This is done by shrinking the fabrication process, effectively decreasing the size of each transistor. Transistor size is ultimately determined by the semiconductor technology node using a particular lithography fabrication process. A node/process is measured in nanometers (nm), or millionths of a

\(^1\) Moore’s Law, [https://en.wikipedia.org/wiki/Moore%27s_law](https://en.wikipedia.org/wiki/Moore%27s_law)
millimeter. That is essentially the distance between the current source side and current drain side of a transistor. To give a sense of scale, a single human hair is about 100,000 nm, or 100 microns wide.

Moore's law is an exponential function, meaning the rate of change is always increasing, and this has largely been true until just recently. Every two to four years a new, smaller technology node makes its way into semiconductor production, from about 10 microns (or 10,000 nm) wide in 1971 to only 14 nm wide today. That would be the same as shrinking the average 5'-6” tall person down to about 3/32” high.

Accordingly, transistor count has gone up from 2,300 transistors to between 1.35 - 2.6 billion transistors in today’s CPU models, which is what ultimately drives how powerful the chip is. Put into perspective, Boston Symphony Hall holds about 2,370 people during the Pops season. The population of China is about 1.357 billion people. Now imagine you could fit the population of China into Boston Symphony Hall. That’s the scale of Moore’s Law.

As the fabrication process shrinks, integrated circuits use less energy and produces less heat, which allows more transistors on a chip. In the late 1990s into the 2000s the trend was to increase on-die transistor counts and die sizes, but the fabrication process was still in the 600nm to 90nm range, so CPUs got larger. Energy consumption and heat dissipation became serious engineering challenges, and led to a new market for exotic cooling components such as larger fans, air coolers with heat pipes, complete closed-loop water cooling solutions with pumps, reservoirs and radiators, and even submerging the workstation in a vat of mineral oil.

Today’s 14nm processors and GPUs are not only physically smaller, but also have advanced internal power management optimizations that reduce power (and thus heat) when it is not required. Increasing PPW allows higher performance to be stuffed into smaller and more mobile packages and platforms, which opened the floodgates to the vast development of mobile technologies that we all take for granted.

This has two side effects. First, the development of more powerful, smaller, cooler running, and largely silent CPUs and GPUs allows you to stuff more of them in a single workstation without it cooking itself. At the same time, CPU clock speeds have been able to go from about 2.4GHz to 4GHz and beyond.

Secondly, complex BIM applications can now extend from the desktop to more mobile platforms, such as performing 3D modeling using a small laptop during design meetings, running clash detections on a construction site using tablets, or using drone-mounted cameras to capture HD imagery.

While breakthroughs in transistor technology at small scales have enabled us to get down to a 14nm process, we are starting to see the end of Moore’s law on the horizon. 14nm is only about 70 silicon atoms wide. At these small scales, the ability to control current flow across a transistor without leakage becomes a significant problem. Transistor design has had to change drastically to keep the weird effects of quantum mechanics at bay. 2 In 10 years we expect to get down to a 5nm process, which is only about 25 atoms wide. The physical limitations in process technology using classical transistor-based computing design is sealing the fate of Moore’s Law and concurrently pushing the development of Quantum Computing. To paraphrase Intel Fellow Mark Bohr, we are quickly running out of atoms to play with. 3

Parallel Processing, Multiprocessing, and Multithreading

Key problems associated with BIM and 3D visualization, such as energy modeling, photorealistic imagery, and engineering simulations, are simply too big for a single processor core to handle efficiently. Many of these problems are highly parallel in nature, where large tasks can often be neatly broken down into smaller ones that don’t rely on each other to finish before the next one can be worked on.

---

2 https://www.youtube.com/watch?v=rtl5wRyHpTg
3 http://www.theregister.co.uk/2013/08/27/moore\_law\_will\_be\_repealed\_due\_to\_economics\_not\_physics/
This lead to the use of multiple CPU systems. All modern processors and operating systems fully support both multiprocessing - the ability to push separate processes to multiple CPU cores in a system - and multithreading, the ability to execute separate threads of a single process across multiple processors. Processor technology has evolved to meet this demand, first by allowing multiple CPUs on a motherboard, then by introducing more efficient multi-core designs in a single CPU package. The more cores your machine has, the snappier your overall system response is and the faster any compute-intensive task such as rendering will complete.

These kinds of non-sequential workloads can be distributed to multiple processor cores on a CPU, multiple physical CPUs in a single PC, or even out to multiple physical computers that will chew on that particular problem and return results that can be aggregated later. Over time we've all made the mass migration to multi-core computing even if we aren't aware of it, even down to our tablets and phones.

In particular, 3D photorealistic rendering lends itself very well to parallel processing. The ray tracing pipeline used in today's rendering engines involves sending out rays from various sources (lights and cameras), accurately bouncing them off of or passing through objects they encounter in the scene, changing the data “payload” in each ray as it picks up physical properties from the object(s) it interacts with, and finally returning a color pixel value to the screen. This process is computationally expensive as it has to be physically accurate, and can simulate a wide variety of visual effects, such as reflections, refraction of light through various materials, shadows, caustics, blooms, and so on.

You can see this parallel processing in action when you render a scene using mental ray. Mental ray renders scenes in separate tiles called buckets. Each processor core in your CPU is assigned a bucket and renders it before moving to the next one. The number of buckets you see corresponds to the number of cores available. The more cores, the more buckets, and the faster the rendering.

Autodesk recognized the benefits of parallelization and provides the Backburner distributed rendering software with 3ds Max. You can easily create your own rendering farm where you send a rendering job out to multiple computers, each of which would render a little bit of the whole, send their finished portion back, which then gets assembled back into a single image or animation. With enough machines, what would take a single PC hours can be created in a fraction of the time.

Just running an operating system and multiple concurrent applications is, in many ways, a parallel problem as well. Even without running any applications, a modern OS has many background processes running at the same time, such as the security subsystem, anti-virus protection, network connectivity, disk I/O, and the list goes on. Each of your applications may run one or more separate processes as well, and processes themselves can spin off separate threads of execution. In Revit, the rendering process is separate from the host Revit.exe process. In AutoCAD, the VisualLISP subsystem runs in its own separate thread.

While today you can maximize efficiency for highly parallel CPU workloads by outfitting a workstation with multiple physical CPUs, each with multiple cores, this is significantly expensive and a case of diminishing returns. Other advancements may point to other directions instead of trying to pile on CPU cores.
The Road to GPU Accelerated Computing and the Impact of Gaming

Multiprocessing is not limited to CPUs. Recognizing the parallel nature of many graphics tasks, graphic processor unit (GPU) designers at ATI and NVIDIA have created microarchitectures that are massively multiprocessing in nature and are fully programmable to boot. Given the right combination of software and hardware, we can now offload compute-intensive parallelized portions of a problem to the graphics card and free up the CPU to run other code. And those new GPU-compute tasks do not have to be graphics related, but could focus on things like modeling weather patterns, acoustical analysis, protein folding, and other complex problems.

Fundamentally, CPUs and GPUs process tasks differently, and in many ways the GPU represents the future of parallel processing. GPUs are specialized for compute-intensive, highly parallel computation - exactly what graphics rendering is about - and are therefore designed such that more transistors are devoted to data processing rather than data caching and flow control.

A CPU consists of a few – from 2 to 8 in most systems - relatively large cores which are optimized for sequential, serialized processing, executing a single thread at a very fast rate. Conversely, today’s GPU has a massively parallel architecture consisting of thousands of much smaller, highly efficient cores (termed Stream Processors) designed to execute many concurrent threads more slowly. The physical chip is also larger. A GPU can have 3 to 4 times as many transistors on the die than a CPU has. Indeed, it is by increasing the Performance per Watt that the GPU can cram so many cores into a single die.

It wasn’t always like this. Back in olden times traditional GPUs used a fixed-function pipeline, and thus had a much more limited scope of work they could perform. They did not really think at all, but simply mapped function calls from the application through the driver to dedicated logic in the GPU that was designed to support them in a hard-coded fashion. This led to all sorts of driver-related issues and false optimizations.

A graphics data pipeline is really a rasterization pipeline. It is composed of a series of steps used to create a 2D raster representation from a 3D scene in real time. The GPU is fed 3D geometric primitive, lighting, texture map, and instructional data from the application. It then works to transform, subdivide, and triangulate the geometry; illuminate the scene; rasterize the vector information to pixels; shade those pixels; assemble the 2D raster image in the frame buffer; and output it to the monitor.

Real Time Rendering in Gaming

In games, the GPU needs to do this as many times a second as possible to maintain smoothness of play. For example, a detailed dissection of a rendered frame from Grand Theft Auto V$^4$ reveals a highly complex rendering pipeline. The 3D meshes that make up the scene are culled and drawn in lower and higher levels of detail depending on their distance from the camera. Even the lights that make up an entire city nighttime scene are individually modeled - that’s tens of thousands of polygons being pushed to the GPU.

The rendering pipeline then performs a large array of multiple passes, rendering out many High Dynamic Range (HDR) buffers, which are bitmaps of various types - diffuse, specular, normal, irradiance, alpha, shadow, reflection, etc. Along the way it applies effects for water surfaces, subsurface scattering, atmosphere, sun and sky, and transparencies. Then it applies tone mapping – also called photographic exposure - which converts the HDR information to a Low Dynamic Range (LDR) space. The scene is then anti-aliased to smooth out jagged edges of the meshes, a lens distortion is applied to make things more film-like, and the user interface (e.g. health, status, the mini-map of the city) is drawn on top of the scene. Then post effects such as lens flares, light streaks, anamorphic lenses, heat haze, and depth of field to blur out things that are not in focus are applied.

---

$^4$ http://www.adriancourreges.com/blog/2015/11/02/gta-v-graphics-study/
A game like GTA V needs to do all of this about 50 to 60 times a second to make the game playable. But how can all of these very highly complex steps be performed at such a high rate?

**The Importance of Shaders**

Today’s GPUs have a programmable graphics pipeline which can be manipulated through small programs called Shaders, which work on scene data to make complex effects happen in real time. OpenGL and Direct3D (part of the DirectX multimedia API for Windows) are 3D graphics APIs that went from the fixed-function hard-coded model to supporting a newer shader-based programmable model.

Shaders work on a specific aspect of a graphical object and pass it on. For example, a Vertex Shader processes vertices, performing transformation, skinning, and lighting operations. It takes a single vertex as input and produces a single modified output vertex. Geometry shaders process entire primitives consisting of multiple vertices, edges, polygons. Tessellation shaders subdivide simpler meshes into finer meshes allowing for level of detail scaling. Pixel shaders compute color and other attributes, such as bump mapping, shadows, specular highlights, and so on.

Shaders are written to apply transformations to a large set of elements at a time, which is very well suited to parallel processing. This led to the creation of GPUs with many cores to handle these massively parallel tasks, and modern GPUs have multiple shader pipelines to facilitate high computational throughput. The DirectX API, released with each version of Windows, regularly defines new shader models which increase programming model flexibilities and capabilities.

**Modernizing Traditional Professional Renderers**

Two of the primary 3D rendering engines in the Building Design Suite 2016 are NVIDIA’s mental ray and the new Autodesk Raytracer. Both use real-world materials, photometric lights, and global illumination (GI) methods (e.g. ray tracing and Final Gather) for producing photorealism. mental ray is owned and licensed by NVIDIA, to which Autodesk pays a licensing fee with each application it ships with. Autodesk is prohibited from making changes to mental ray, although historically they have worked closely with NVIDIA on its development.
mental ray is almost 30 years old whereas the Autodesk Raytracer is brand new. Accordingly, for end users the primary difference between using the two is in simplicity and speed, where the new Autodesk Raytracer can produce images much faster, more efficiently, and with far less tweaking than mental ray. Autodesk Raytracer also produces images that are arguably of better rendering quality⁵. Autodesk Raytracer is currently in use in AutoCAD, Revit, Navisworks, and Showcase, and available in a plug-in for 3ds Max as an Autodesk Labs project⁶.

However, neither mental ray, Autodesk Raytracer, or other popular 3rd party renderers like V-Ray Advanced use the computational power of the GPU to accelerate rendering tasks. Rendering with these engines is almost entirely a CPU-bound process, so a 3D artist workstation would need to be outfitted with multiple (and expensive) physical multi-core CPUs. Of course, you can lower render times in 3ds Max by setting up a render farm using the included Backburner software, but each node on the farm needs to be pretty well equipped. That has a huge impact on how you manage rendering workloads and deadlines.

Designed for rasterizing many frames of simplified geometry to the screen per second, GPUs were not meant for performing ray-tracing calculations. This is rapidly changing as most of a GPU’s hardware is now devoted to 32-bit floating point shader processors. NVIDIA exploited this in 2007 with an entirely new GPU computing environment called CUDA (Compute Unified Device Architecture), which is a parallel computing platform and programming model established to provide direct access to the massive number of parallel computational elements in their CUDA GPUs. Non-CUDA platforms (that is to say, AMD / ATI) can use the Open Computing Language (OpenCL) framework, which allows for programs to execute code across heterogeneous platforms – CPUs, GPUs, and others.

Using the CUDA / OpenCL platforms we now have the ability to perform non-graphical, general-purpose computing on the GPU (often referred to as GPGPU), as well as accelerating graphics tasks such as calculating game physics.

One of the most compelling areas GPU Compute directly comes into play for Building Design Suite users is the NVIDIA iray rendering engine built into 3ds Max. Similar technologies such as V-Ray RT and BIM IQ are going to bring GPU rendering directly to Revit as add-ons. We’ll discuss this in more depth in the section on graphics. Given the nature of parallelism, I would not be surprised to see GPU compute technologies to be exploited for other uses across all future BIM applications.

**Using Gaming Engines for Architectural Visualization**

So now that we have advanced shaders and commercially available (and cheap) GPU hardware that harnesses them, we can now apply them to the task of design visualization and try to reach that holy grail of navigating, in real-time, a photorealistic project environment.

The advancements made in today’s game engines is quickly competing with, and sometimes surpassing, what dedicated rendering engines like mental ray, v-ray and others can create. A game engine is a complete editing environment, not unlike 3ds Max or Maya, for working with 3D assets, developing lifelike materials, adding photometric lighting, animations, and writing custom programming code that turns a collection of static 3D models into full-fledged action titles. Instead of the same old highly post-processed imagery or “sitting in a shopping cart being wheeled around the site” type animations, the result is a free running “game” that you and your clients can explore and interact with and that renders in real time. The difference is that the overall image quality in these game engines is incredibly high.

Currently the most well-known gaming engines available are Unreal Engine 4 and Unity 5, and are quickly becoming co-opted by the building design viz community. In 2014 Autodesk acquired the Bitsquid game

---

⁵ http://www.aecbytes.com/tipsandtricks/2015/issue72-revit.html
⁶ https://beta.autodesk.com/callout/?callid=%7B56FF59F6-8E30-440E-B8A0-8015AFF45AED%7D
engine which has been evolved into their new Stingray\(^7\) product. Lumion, a very popular architectural visualization application, is also underpinned by a gaming engine, on top of which they add value by creating a user interface, prebuilt content, and other assets.

What’s truly great about Unreal Engine 4 and Unity 5 is their cost to the design firm. They’re \textit{free}. Both Unreal and Unity charge publishers a percentage of revenue for games, but for design visualizations, there is no charge. The user community is huge and growing every day, and independent folks develop add-ons, materials, models, and environments that you can purchase and drop into your project. Even Stingray is cheap, which is included with Maya LT can be purchased for a low $30/month fee.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ue4.png}
\caption{This is not a photograph. It's Unreal Engine 4. And it's free. Image courtesy ue4arch.com.}
\end{figure}

\textbf{Virtualization and Cloud Computing}

One of the more compelling side-effects of cheap, fast processing is the (re)rise of virtual computing. Simply put, Virtual Machine (VM) technology allows an entire computing system to be emulated in software. Multiple VMs, each with their own virtual hardware, OS, and applications can run on a single physical machine.

VMs are in use in most businesses today in some fashion. Companies may host multiple server VMs on a single high-end box, which allows them to deploy fewer physical boxes to host file storage servers, Microsoft Exchange servers, SQL database servers, application servers, web servers, and others. Revit Server, which allows office-to-office synchronization of Revit files, is often put on its own VM.

This is valuable because many server services don’t require a lot of horsepower, but you don’t usually want to combine application servers on one physical box under a single OS. You don’t want your file server also hosting Exchange, for example, for many reasons; the primary one being that if one goes down it takes the other out. Putting all your eggs in one basket usually leaves you with scrambled eggs when you hit a bump in the road.

\footnote{7 http://www.engadget.com/2015/08/03/autodesk-stingray/}
VMs also allows IT a lot of flexibility in how these servers are apportioned across available hardware and allows for better serviceability. On the outside, VMs are just single files that contain the OS, files, and applications. As such a VM can be shut down independently of the host box or other VMs, moved to another machine, and fired up within minutes. You cannot do this with Microsoft Exchange or your company accounting system installed on a normal server.

IT departments will use VMs to test new operating systems and applications, or to use a VM for compatibility with older apps and devices. If you have an old scanner that won’t work with a modern 64-bit system, don’t throw it out; simply fire up a Windows XP VM and run it under that.

Today’s virtualization extends to the workstation as well. Companies are building out their own on-premise VM data centers in their IT closets, delivering standardized, high performance workstation desktops to in-house and remote users working with modest client hardware. By providing VMs to all users, IT can easily service the back-end hardware, provide well over 99% uptime, and instantly deploy new applications and updates across the board (a surprisingly huge factor with the 2016 releases).

Traditionally. The primary impediment to deploying VMs of high-end applications like Revit, Navisworks, and 3ds Max has been in how to share graphics resources between virtual machines. Older VMs could not provide the kind of dedicated “virtual” graphics capabilities required by these applications to run well. This is now largely alleviated with new capabilities in VM providers such as VMWare and others, where you can install multiple high-end GPUs in a server host box that can be shared between workstation VMs.

The Cloud Effect
VM technologies have migrated from you hosting multiple virtual machines on your server to someone else doing it for you, me, and everyone else. Data centers are everywhere and accessible any time. That is why they call it the Cloud, and no information technology discussion today would be complete with some reference to cloud computing.

By now, it’s taken for granted that processing speed increases over time and price per process drops. This economy of scale is coupled with the ubiquitous adoption of very fast Internet access at almost every level. The mixing of cheap and fast computing performance with ubiquitous broadband networking has resulted in easy access to remote processing horsepower. Just as the cost of 1GB of disk storage has plummeted from $1,000 to just a few pennies, the same thing is happening to CPU cycles as they become widely available on demand.

This has manifested itself in the emerging benefit of widely distributed cloud-based computing services. The cloud is quickly migrating from the low hanging fruit of simple storage-anywhere-anytime mechanism (e.g., Dropbox, Box.net), to providing users with massive remote access capabilities to fast machines. This will soon become on-demand, essentially limitless, and very cheap computing horsepower.

Accordingly, the entire concept of a single user working on a single CPU with its own memory and storage is quickly being expanded beyond the box in response to the kinds of complex problems mentioned earlier, particularly with BIM. This is the impetus behind Autodesk’s A360 series of large-scale distributed computing projects, such as Revit’s Cloud Rendering, Green Building Studio energy analysis, and structural analysis capabilities.

Today you can readily tap into distributed computing cycles as you need them to get a very large job done instead of trying to throw more hardware at it locally. You could have a series of still renders that need to get out the door, or a long animation whose production would normally sink your local workstation or in-house Backburner render farm. Autodesk’s Cloud Rendering service almost immediately provided a huge productivity boon to design firms, because it reduced the cost of getting high quality renderings from hours to just a few minutes. Autodesk isn’t the only one doing this - there are hundreds of smaller
dedicated render farm companies which will provide near-zero setup of dozens of high-performance CPU+GPU combinations to get the job done quickly and affordably.

And you can build this out yourself. Amazon’s EC2 (Elastic Compute Cloud) service is a general-purpose web service that provides compute capacity in the cloud, and provides the ability to build a temporary virtual rendering farm for very little money, starting at about $0.65 cents per core hour for a GPU+CPU configuration. This is exactly what Autodesk did, as their Cloud Rendering service runs on Amazon’s EC2.

Once signed up you have a whole host of machines at your disposal to chew on whatever problem you need to send. A cost comparison of using Amazon EC2 for iray rendering\(^8\) and a tutorial on how to set up an EC2 account is provided in the footnotes\(^9\).

Interesting note: At AU 2014 I assisted in a Families lab taught by Paul Aubin. As part of an experiment, Autodesk worked with Amazon EC2 engineers to set up the lab to be run with Revit hosted on the EC2 platform. None of the 80 or so workstations were doing any heavy lifting at all; they were connected to the cloud machines running Revit. Aside from working around the terrible Internet connection at Mandalay Bay’s conference center, the labs ran with very few issues. Models could be spun around in 3D fairly smoothly using the Realistic model display style.

We can see where the future is leading; that is, to “thin” desktop clients with just enough computing horsepower to turn themselves on, accessing major computing iron that is housed somewhere else. Because most of the processing happens across possibly thousands of CPUs housed in the datacenter, your local machine may no longer need be a powerhouse. At some point this will become more and more prevalent, perhaps to where we reach a stage where the computing power of your desktop, tablet, or phone will almost be irrelevant, because it will naturally harness CPU cycles elsewhere for everyday computing, not just when the need arises due to insufficient local resources.

**Price vs. Performance Compression**

One of the side effects of steadily increasing computing power is the market-driven compression of prices. At the normal end of the scale, the pricing differences between any two similar components of different capacities or speeds has shrunk dramatically, making the higher end option often the more logical buy. For example, a high quality 1TB drive is about $70, a 2TB drive is about $117, and a 4TB drive is about $200. The 4TB drive nets you 4x the storage for under three times the price.

For system memory, conventional wisdom once decreed 8GB as a starting point for a BIM workstation. Back in olden times this typically meant outfitting a system with (4) 2GB 240-pin DDR3 memory modules, as 4GB modules were much more expensive. Today, a 2GB module is about $13 ($6.50/GB), and 4GB modules have dropped to about $25 ($6.25/GB), making it obvious to go with (2) 4GB modules.

When 8GB modules came on the scene they similarly carried a price premium, but today they have dropped to under $35 or so, or only about $4.50/GB. Because you need to install RAM in pairs, it makes sense to install 16GB as (2) 8GB modules. Modern desktop motherboards have 4 memory slots, so there is an argument to be made to simply go for 32GB as (4) 8GB modules from the start. This all changes a bit under the new DDR4 RAM specification as we will see layer in Section V, but the idea is the same.

Note that 16GB RAM modules are available, but they about twice as expensive – about $8.75/GB. While it may seem logical to shoot the works and try to outfit your system with (4) 16GB for 64GB total, mainstream desktop CPUs and chipsets do not support more that 32GB of RAM. Plus, most Building Design Suite applications will not see any real performance increase with 64GB of RAM unless you are rendering


very large scenes. Additionally, to move to a system with more than 32GB of RAM requires a higher end computing platform which we will discuss later.

For any subsystem or component, it typically doesn’t pay to go for the low end except when you know you won’t need the extra capability. For example, in a business-class graphics workstation scenario, most of the data is held on a server, so a “lowly” 1TB drive is more than adequate to house the OS, applications, a user’s profile data, and pretty much their entire iTunes library. In the case of DVD/RW, which no one uses anymore, it makes sense to leave it out completely and move on.

Processors have a different story. CPU pricing is based upon capability and popularity, but price curves are anything but linear and work in the opposite direction. A 3.2GHz CPU might be $220 and a 3.4GHz is incrementally higher at $300, but a 3.5GHz CPU – 100MHz faster - could be $600. This makes for plenty of “sweet spot” targets for each kind of CPU lineup.

Graphics cards are typically set to price points based on the GPU and the amount of memory on the card. Both AMD (which owns ATI) and NVIDIA may debut 5 or 6 new cards a year, typically based on the latest GPU architecture. Models of a particular GPU are differentiated by base clock speed, onboard memory, or number of internal GPU cores present or activated. Both companies issue reference boards that 3rd party card manufacturers use to build their offerings. The result being that prices between different manufacturer’s cards using the same GPU may only be between $0 and $20 of each other, with more expensive variations available that have game bundles, special coolers, or have been internally overclocked by the manufacturer.

A Problem of Perception
Shrinking prices for components that are good enough for the mainstream can skew the perception of what a BIM and 3D workstation should cost. Accounting usually balks when they see workstation quotes pegging $4,000 when they can pick up a mainstream desktop machine for $699 at the local big box store. Don’t be swayed and don’t give in: your needs for BIM are much different.
III. Building Design Suite Application Demands

Within any workstation there are four primary components that are stressed to some degree by the applications in the Building Design Suite and directly determine overall performance in some key area: the processor (CPU), system memory (RAM), the graphics card (GPU), and the storage subsystem. While the chipset is an important component which glues the CPU and RAM together on the motherboard, it is not by itself stressed by applications directly. If anything, the chipset in conjunction with the CPU defines a computing platform and largely determines how expandable and capable the system is overall.

Different applications stress different parts of your system. Given the current state of hardware, today’s typical entry-level workstation may perform well in most, but not all of the applications within the Suite, due to specific deficiencies in one or more subsystems. You need to evaluate how much time you spend in each application - and what you are doing inside of each one - and apply that performance requirement to the capabilities of each component.

Application / Demand Matrix

The following table provides a look at how each of the major applications in the Building Design Suite are affected by the different components and subsystems in your workstation. Each value is on a scale of 1-10, where 1 = low sensitivity / low requirements and 10 = very high sensitivity / very high requirements.

<table>
<thead>
<tr>
<th>Application / Demand Matrix</th>
<th>CPU Speed / Multithreading</th>
<th>System Ram - Amount / Speed</th>
<th>Graphics Card GPU Capabilities</th>
<th>Graphics Card Memory Size</th>
<th>Hard Drive Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revit</td>
<td>10 / 9</td>
<td>9 / 7</td>
<td>7</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>3ds Max</td>
<td>10 / 10</td>
<td>9 / 7</td>
<td>7 / 5 /10</td>
<td>6 / 10 (Nitrous / mr / iray™)</td>
<td>10</td>
</tr>
<tr>
<td>Navisworks Simulate Navisworks Manage</td>
<td>7 / 7</td>
<td>6 / 6</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Showcase</td>
<td>9 / 8</td>
<td>8 / 6</td>
<td>9</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>AutoCAD (2D &amp; 3D)</td>
<td>6 / 6</td>
<td>5 / 5</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>AutoCAD Architecture AutoCAD MEP</td>
<td>8 / 6</td>
<td>7 / 5</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>ReCap</td>
<td>10 / 10</td>
<td>9 / 5</td>
<td>8</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

To give a sense of direction to the above chart, any low-end workstation you could purchase off the shelf at Best Buy will likely perform adequately well in most of these applications, up to a rating of around 6. For example, such a baseline system would be adequate for AutoCAD 2D and 3D drawing tasks for most projects. However, it would likely need specific tweaking in the storage subsystem to run something like Navisworks Manage, and is really inappropriate for Revit, ReCap, or 3ds Max.

Not that those applications will not run in such a baseline system; but rather, that system is not optimized for them. As you can see from the chart above, most of the Building Design Suite applications have at least one aspect which requires careful consideration for a particular component.

---

10 Base machine specs: Core i5 /AMD A8 series processor, 8GB RAM, built-in video, 1TB mechanical hard drive
Application Notes: Revit

The Autodesk Revit platform stresses most major components in a workstation because of the nature of the datasets you work on in building design. Most Suite users spend more time in Revit than most other applications, so tuning your workstation specifically for this application is a smart choice.

Because of the size and complexity of most BIM modeling efforts, it requires a fast CPU, a decent amount of system RAM, and a fast storage system. However, Revit has rather mundane graphics demands, but will make use of fast GPUs. But you can get by with relatively basic graphics cards, even on large projects.

Revit is, at its heart, a database management system (DBMS) application. As such, it takes advantage of certain technical efficiencies in modern high-end CPUs, such as multiple cores and larger internal L1, L2, and L3 high-speed memory caches. Modern CPUs within the same microarchitecture lineup have similar multiple cores and L1/L2/L3 caches, with the differences limited primarily to core clock speed. Differentiations in cache size and number of cores appear between the major lines of any given microarchitecture. This is particularly evident for CPUs specifically designed for database servers, which have more cores per CPU, allow for multiple physical CPU installations, and increased L1/L2/L3 cache sizes.

Revit’s high computing requirements are primarily due to the fact that it has to track every element and family instance as well as the relationships between all of those elements at all times. Revit is all about relationships; its Parametric Change Engine works within the framework of model 2D and 3D geometry, parameters, constraints of various types, and hosted and hosting elements that understand their place in the building and allow the required flexibility. All of these aspects of the model must respond to changes properly and update all downstream dependencies immediately. Let’s see how each component is specifically affected by Revit:

**Processor (CPU):** Revit requires a fast CPU because all its work is computationally expensive. There are no shortcuts to be had; it has to do everything by the numbers to ensure model fidelity. This is particularly noticeable when performing a Synchronize with Central (SWC) operation, as Revit first saves the local file, pulls down any model changes from the Central Model, integrates them with any local changes, validates everything, and sends the composite data back to the server. Revit 2016 supports multi-threading in certain operations:\(^{11}\):

- Progressive display (new in 2016) used by all models all the time
- Vector + Raster printing
- 2D Vector Export such as DWG and DWF
- Rendering. Autodesk Raytracer in particular overcomes the 16-core limitation seen in mental ray.
- Wall Join representation in plan and section views
- Loading elements into memory, reducing view open times when elements are initially displayed
- Parallel computation of silhouette edges, accelerating navigation of perspective 3D views
- Translation of high level graphical representation of model elements and annotations into display lists optimized for graphics cards, engaged when opening views or changing view properties
- File Open and Save
- Point Cloud Data Display

Autodesk will continue to exploit multi-core capabilities in other areas in future releases.

---

\(^{11}\) [http://insidethefactory.typepad.com/my_weblog/2015/05/revit-performance-scalability.html](http://insidethefactory.typepad.com/my_weblog/2015/05/revit-performance-scalability.html)
All modern CPUs are 64-bit and meet or exceed the minimum recommended standard established by Autodesk. But with everything else, you want to choose a CPU with the very latest microarchitecture platform, the most cores you can afford, the fastest core clock speed, and the most L2 cache available. We will discuss these specific options in the Processor section of this handout.

**System Memory (RAM):** The need to compute all of these relational dependencies is only part of the problem. Memory size is another sensitive aspect of Revit performance. According to Autodesk, Revit consumes 20 times the model file size in memory, meaning a 100MB model will consume 2GB of system memory before you do anything to it. If you link large models together or perform a rendering operation without limiting what is in the view, you can see where your memory subsystem can be a key bottleneck in performance.

The more open views you have, the higher the memory consumption for the Revit.exe process. Additionally, changes to the model will be updated in any open view that would be affected, so close out of all hidden views when possible and before making major changes.

With operating systems getting more complex and RAM being so inexpensive, 16GB (as 2x8GB) is today’s minimum recommended for the general professional level. 32GB or more would be appropriate for systems that do a lot of rendering or work in other Building Design Suite applications simultaneously.

**Graphics:** With Revit we have a comprehensive 2D and 3D design environment which requires decent performance graphics capabilities to use effectively. However, I have found that Revit performs adequately well on most projects under relatively mainstream graphics cards between $200 and $350.

This is mostly because Revit views typically contain only a subset of the total project geometry for the sake of clarity. Even in 3D views, one typically filters out and limits the amount of data which enables the system to respond quickly enough for most GPUs can handle with aplomb. But the graphics card can get a real workout as we demand more use of shaded and realistic views complete with material appearances. Toss in sketchy lines, anti-aliasing, ambient shadows, lighting, and so on, and view performance can slow down dramatically. The better the graphics card, the more eye candy can be turned on and performance levels can remain high.

Your graphics performance penalties grow as the complexity of the view grows, but Autodesk is helping to alleviate viewport performance bottlenecks. In 2014, Revit viewports got a nice bump with the inclusion of a new adaptive degradation feature called Optimized View Navigation. This allows Revit to reduce the amount of information drawn during pan, zoom and orbit operations and thus improve performance.

In 2015 we gained the ability to limit smoothing / anti-aliasing operations on a per-view setting using the Graphics Display Options dialog. Anti-aliasing (AO) is the technology that eliminates jagged pixels on diagonal geometry by blending the line pixels with the background. It looks great but is computationally expensive, so view performance can be increased by only turning it on in the views that require it.

These settings are found in the Options > Graphics tab and in the view’s Graphic Display Options:
Revit 2015 improved performance in the Ray Trace interactive rendering visual style, and in other views, Revit 2015 improved drawing performance such that many elements are drawn simultaneously in larger batches using fewer drawing calls; specifically, that release added some new enhancements to increase GPU use, namely drawing multiple geometries of the same material to the GPU as a batch, and drawing multiple identical objects to the GPU as a batch.

Additionally, a newer, faster process is used for displaying selected objects, and the underlying technology used for displaying MEP elements in views improves performance when opening and manipulating views with many MEP elements.

Autodesk has made even more strides to improve the efficiency of navigating around the viewport without waiting for it to finish drawing the model. Revit 2016 now makes use of GPU processing for drawing all graphics via DirectX 11, giving Revit approximately a 30% increased drawing operation speed.

Note that while Revit can use a decent graphics card foundation for upper order operations, it is actually agnostic about specific video card makes or models. All cards manufactured over the past five years will support Revit 2016’s minimum system requirements which specify a DirectX 11 / Shader Model 3 capable card running under Windows 7 64-bit. This allows for all viewport display modes, adaptive degradation, ambient occlusion effects, and so on. The general rule is that the faster (and more expensive) the card is, the better it will be for Revit, but only up to a point. We discuss this in the section on graphics.

**Storage:** It is no secret that the size of the files Revit creates are huge, particularly compared to traditional AutoCAD DWG files and represents a bottleneck in opening and saving projects. 60MB Revit .RVT files are typical minimums for smaller projects, with file sizes of 100MB and more being common. MEP models with typically start around 60-80MB for complete basic projects and go up from there. Today, models topping 1GB are not uncommon.

For Workshared projects Revit needs to first copy these files off of the network to the local drive to create your Local File, and keep that file synchronized with the Central Model. While we cannot do much on the network side (we are all on Gigabit networks these days), file I/O operations can take their toll on your local storage subsystem.

Also, don’t forget that Revit itself is a large program and takes a while just to fire up, so you want a fast storage subsystem to comfortably use the application with large models. Revit is certainly an application where Solid State Drives (SSDs) shine.

**Modeling Efficiently is Key**

Overall, Revit performance and model size is as much directly tied to implementing efficient Best Practices in your company as it is the hardware. A badly executed 200MB model will perform much worse than a solid 300MB model. With such inefficient models, Revit can consume a lot of processing power in resolving things that it otherwise would not.

Two primary ways of improving performance is to limit the amount of work Revit has to do in the views. Create families with 3D elements turned off in plan and elevation views, and use fast Symbolic Lines to represent the geometry instead. This minimizes the amount of information Revit will need to process in performing the hidden line mode for 2D plan, elevation, section and detail views. In 3D views, the goal is to minimize the number of polygons to deal with, so use the Section Box tool to crop the model to only the area you want to work on at any one time. The use of Filters to turn off large swaths of unnecessary
geometry can be a huge performance boon, particularly in Revit MEP, where you can have lots of stuff on screen at one time.

Fortunately Autodesk provides a very good document on modeling efficiently in Revit. Autodesk’s Model Performance Technical Note document has been updated to 2015 and is an invaluable resource for every Revit user.13

Application Notes: 3ds Max

**Autodesk 3ds Max** has base system requirements that are about the same as they are for Revit. However, 3ds Max stresses your workstation differently and exposes weakness in certain components. With 3ds Max there isn’t any BIM data interaction to deal with, unless you link RVT / FBX files (which does add a lot of overhead). Overall, 3ds Max is all about having high end graphics capabilities that can handle the display and navigation of millions of polygons as well as large complicated textures and lighting. You also have to contend with CPU-limited and/or GPU-limited processes in rendering.

For typical AEC imagery which doesn’t require subobject animation, the problems that Max has to deal with are related to the following:

- **Polgons** - Interacting with millions of vertices, edges, faces, and elements on screen at any time;
- **Materials** - Handling physical properties, bitmaps, reactions to incoming light energy, surface mapping on polygonal surfaces, and procedural texture generation;
- **Lighting** - Calculating physical and non-physical lighting models, direct and indirect illumination, shadows, reflections, and caustics;
- **Rendering** - Combining polygons, materials, lighting, and environmental properties together to produce final photorealistic imagery; ray tracing under the mental ray and iray rendering engines; performing post-rendering effects

Each component affects performance thusly:

**CPU:** 3ds Max is a highly tuned and optimized multi-threaded application. Geometry, viewport, lighting, materials, and rendering subsystems can all be computationally expensive and 3ds Max will take full advantage of multiple cores / processors. Having many fast cores allows for fast interaction with the program even with very large scenes. As discussed, the standard scanline and mental ray rendering engines are almost wholly CPU dependent and designed from the ground up to take advantage of multiple processors, and scale pretty linearly with your CPUs capabilities. For those engines, having CPUs with multiple cores and/or moving to multiple physical processor hardware platforms will shorten rendering times considerably. In addition, Max includes distributed bucket rendering with Backburner, which allow you to spread a single rendering task across physical machines, even further reducing rendering times.

Like Revit, 3ds Max can make whole use of the best CPU you can afford. If you spend a lot of time in 3ds Max and render high resolution images, you owe it to yourself to look at more highly-powered workstations that feature a high number of cores on the CPU. The Return on Investment (ROI) for high end hardware is typically shorter for Max than any other program in the Building Design Suite, because the effects are so immediately validated.

**RAM:** 3ds Max 2016 also requires a good deal of system memory, particularly for large complex scenes with Revit links as well as for rendering operations. The application itself will consume about 540MB with just an empty scene and with minimal plug-ins loaded. If you regularly deal with large animation projects with complex models and lots of textures, you will find installing more RAM to be compelling in your

---

specification decisions. As we discuss in the section on CPUs and chipsets, the choice of CPU decides how much RAM your system can address. Baseline desktop CPUs top out at 32GB, but higher-order machines can support 64GB on up to 192GB on very high end workstations. Most AEC scenes can readily work fine within 16GB, after which there is a case of diminishing returns. However, for those who regularly work with large complex scenes, and multiple applications / datasets open at once, moving to a more aggressive hardware platform with multiple physical CPUs will, as a side benefit, result in more addressable RAM and provide that double benefit to the Max user.

Note that this is true for any machine used in a rendering farm as well; rendering jobs sent to non-production machines with a low amount of RAM can often fail. The best bet is to ensure all machines on a farm have the required amount of RAM to start with and, as much as possible, the same basic CPU capabilities as your primary 3ds Max machine.

**Graphics:** With 3ds Max we have a continually improving viewport display system (Nitrous) which is working to take more direct advantage of the graphics processing unit (GPU) capabilities in various ways. The Nitrous viewport allows for a more interactive, real-time working environment with lighting and shadows, which requires higher-end graphics hardware to use effectively.

In 2014 Nitrous got a nice bump in viewport performance with support for highly complex scenes with millions of polygons, better depth of field, and adaptive degradation controls that allow scene manipulation with higher interactivity. In 2015 viewports were made faster with a number of improvements accelerating navigation, selection and viewport texture baking. Anti-aliasing could be enabled with minimal impact on performance but real-world experience says this largely depends on the graphics card. Nitrous improvement continue in 3ds Max 2016, where users report much better viewport performance.

A big differentiator in graphics platform selection is the rendering engine used. Unlike mental ray, the iray rendering system can directly use NVIDIA GPUs to accelerate rendering tasks to a very high degree. In addition, iray performance scales directly with the number of GPUs in your system. This obliquely plays into the choice of CPU /chipset platform, as this determines the number of PCI Express slots. So if you want 3, 4, or even 5 graphics cards to leverage in iray productively, you necessarily need to specify a CPU hardware platform that can handle multiple graphics cards at high speeds. We specifically discuss the needs of iray users in 3ds Max in the section on graphics hardware.

Lastly, 3ds Max users on Subscription can now enjoy A360 cloud rendering support, allowing machines with minimal CPU or GPU resources to quickly render images without tying up their desktop or configuring rendering farms.

**Storage:** The 3ds Max program itself can be notoriously slow to load, particularly if you use a lot of plugins. Factor in the large .max files you create (particularly if you link Revit files), a fast local storage system will pay off greatly.

Finally, remember that Viz Wizards will often work simultaneously in other programs, such as Photoshop, Mudbox, Revit, Inventor, and AutoCAD, so make sure your workstation specification can cover all of these bases concurrently.

**Application Notes: Navisworks Manage / Simulate**

**Autodesk Navisworks Manage** and **Autodesk Navisworks Simulate** are primarily used by the construction industry to review, verify, and simulate the constructability of a project. Its three main features are the Clash Detective (in Navisworks Manage only) that identifies and tracks collisions between building elements before they are built; the TimeLiner which applies a construction schedule to the building
elements, allowing you to simulate the construction process; and integrated 2D and 3D quantification for performing easy takeoffs for estimating purposes.

As such, Navisworks is all about fast viewpoint processing to allow you to navigate very large and complex building models. In most cases, these originate as design intent endpoint models and then are extended by the contractors to include more accurate information for fabrication and installation. Additionally these construction models can be from various sources outside of Revit, such as Fabrication CADmep+ for ductwork and piping, structural steel fabrication models from Tekla Structures, IFC files, site management and organization models from SketchUp, and so on.

The key ingredient that makes this happen is an optimized graphics engine which imports CAD and BIM data and translates it into greatly simplified “shell” geometry, in the form of .NWC files, which minimizes the polygons and allows for more fluid interaction and navigation.

One of the biggest criticisms with Navisworks was that, while it will easily handle navigation through a 2 million SF hospital project with dozens of linked models, the graphics look bland and not at all lifelike. Realistic imagery was never intended to be Navisworks’ forte, but this is getting a lot better with each release. In 2016 we now have the multi-threaded Autodesk Raytracer rendering engine, cloud rendering using the Autodesk 360 service, and improvements in using ReCap point cloud data. Viewports have been improved with better occlusion culling (disabling obscured objects not seen by the camera) and improved faceting factor with Revit files.

**Processor:** Navisworks was engineered to perform well on rather modest hardware, much more so that Revit or 3ds Max. Any modern desktop processor will handle Navisworks just fine for most construction models. Larger models will demand faster processors, just as it would in Revit and 3ds Max. But because Navisworks works on much smaller and more efficient files, performance on even very large projects does not suffer in the same way.

Surprisingly, Navisworks-centric operations, such as Time Liner, Quantification, and Clash Detective, do not require a lot of horsepower to run fast. Clash tests in particular run extremely fast even on modest hardware because the geometry is simplified and the testing algorithm is very efficient. However, the new Autodesk rendering engine will demand higher performance systems to render effectively, so if you are planning to do rendering from Navisworks, target your system specifications similar to Revit and 3ds Max.

**RAM:** Navisworks 2016 by itself consumes a rather modest amount of RAM - about 180MB without a model loaded. Because the .NWC files it uses are rather small, additional memory required with your construction models is also pretty modest. We have found that standard 8GB systems will work well with Navisworks and moderately sized projects.

**Graphics:** The geometric simplification from the source CAD/BIM file to .NWC allows for more complex models to be on screen and navigated in real time. In addition, Navisworks will adaptively drop out geometry as you maneuver around to maintain a minimum frame rate, so the better your video subsystem the less drop out should occur. Since there are far fewer polygons on screen, Navisworks won’t test your graphics card’s abilities as much as other applications. Most decent cards that would be applicable for the rest of the Building Design Suite will handle moderately complex Navisworks models without issue.

**Storage:** The files Navisworks creates and works with (.NWC) are a fraction of the size of the originating Revit/CAD files. NWCS store the compressed geometry of the original application file and strip out all of the application specific data it does not need, e.g. constraints. A 60MB Revit MEP file will produce a Navisworks NWC file that might be 1/10th the size. This lowers the impact on your storage and network systems, as there isn’t as much data to transfer.
**Bottom Line:** Overall, Navisworks has some of the more modest requirements of the applications in the Building Design Suite in terms of system hardware. Because most Navisworks users are Revit users as well, outfitting a workstation suitable for Revit will cover Navisworks just fine.

**Application Notes: Recap Studio / Recap Pro**

*Autodesk ReCap Studio*, found in the Building Design Suite, as well as ReCap Pro are designed to work with point cloud files of several billions of points. ReCap allows you to import, index, convert, navigate, and edit point cloud files, saving them to the highly efficient .RCS file format which can then be linked into AutoCAD, Revit, Navisworks, and 3ds Max with the appropriate Point Cloud extension installed. Once linked into a design application, you can snap to and trace the points in the cloud file to recreate the geometry to be used downstream.

The user interface for ReCap is quite unlike anything else Autodesk has in the Building Design Suite, and can be rather confusing and sluggish to respond to user input. Once the UI is learned, interacting with the point cloud data itself is relatively quick and straightforward, because editing operations are minimal.

**Processor:** Probably the biggest single operation that affects performance is going to be in re-indexing the raw point cloud scan files into the .RCS format. Processing massive raw point cloud scans can take a very long time depending on how many there are. The indexing operation is heavily reliant on the CPU and disk as it writes out very large .RCS files. CPU utilization can be peg to 100% when indexing files which can reduce performance elsewhere. Having a very fast modern processor at your disposal will definitely make the index operation faster.

Once the scans are indexed and in ReCap, however, CPU utilization goes down quite a bit. A test project of 80 .RCS files that total about 18GB was not a problem for the average workstation with 8GB of RAM to handle. Typical operations, such as cropping point cloud data, turning individual scans on and off, and so on were fairly straightforward without an excessive performance hit.

**Memory:** ReCap’s memory consumption is pretty lightweight, around 150MB by itself. When indexing point cloud scans RAM utilization will jump to between 500MB and 1GB. Loaded up with 18GB of .RCS files, memory consumption only hit about 900MB, demonstrating the effectiveness of the indexing operation. Modestly equipped workstations will probably handle most ReCap projects without issue.

**Graphics:** This is one area that needs special attention for heavy ReCap use. The ability to navigate and explore point clouds in real time is a very compelling thing - it’s like walking through a fuzzy 3D photograph. To do this effectively means you need a decently powered graphics card. ReCap has some controls to optimize the display of the point cloud, but a marginal workstation without a fast card will definitely suffer no matter how small the project.

**Storage:** ReCap project files (.RCP) are small, in the 1-5MB range. They simply reference the large .RCS scan files and add data, much like Navisworks .NWF files reference .NWC files which contain the actual geometry. For most scan projects you’ll be dealing with many large individual point cloud scan files that are between 100 and 300MB, so a ReCap project of 50 or so scans will consume many GB of disk space. Working locally, Solid State drives will definitely help ReCap operations as it can suck in that volume of data very quickly. If you work with point clouds on the majority of your projects, expect to add disks to your server’s storage arrays.

**Application Notes: AutoCAD / AutoCAD Architecture / AutoCAD MEP**

*Autodesk AutoCAD 2016* is the industry standard bearer for 2D and 3D CAD. Because it has been around for so long, its hardware requirements are pretty well understood and can be handled by modest entry level workstations. For 2D drafting and design, any modern PC or workstation should suffice. For AutoCAD
Architecture (ACA) and AutoCAD MEP (AMEP) your hardware requirements go up because of the complexity of these vertical applications as well as the increased use of 3D.

**Processor:** Modern CPUs will largely handle AutoCAD, ACA, and AMEP tasks without issue. As your projects get larger and you work with more AEC objects, CPU usage will climb as AutoCAD Architecture and MEP needs to calculate wall joins, track systems, schedule counts through external references, and other more CPU intensive operations.

**System Memory:** Most systems with equipped with 8GB will handle base AutoCAD just fine. AutoCAD consumes 165MB by itself without any drawing files loaded. ACA weighs in at 200 MB, and AMEP at 242MB. In use, the verticals can and will consume a lot more memory that base AutoCAD because of the additional AEC specific information held in each object, as well as keeping track of their display configurations. Drawings with many layout tabs and tabs with many viewports will also consume more RAM because AutoCAD will cache the information to make switching between tabs faster.

**Graphics:** The needs of 2D CAD have been well handled by moderately priced graphics cards for some time. However, for 3D CAD, ACA and AMEP work, a higher-end graphics card will pay off with faster 3D operations such as hide, orbit, and display representation operations. If you only do 2D CAD in AutoCAD but also do 3D work in other Suite programs like 3ds Max, ensure your graphics capabilities can adequately match the higher demand of the other applications.

**Storage:** All AutoCAD based applications work with comparatively small .DWG files, so storage requirements are easily met on baseline systems. As with all Building Design Suite applications, AutoCAD and particularly the verticals can take a long time to load, and thus will benefit from fast disk subsystems in that regard.

**Application Notes: Autodesk Showcase**

**Autodesk Showcase** provides real-time interaction with ray-traced lighting and materials, allowing you to fluidly visualize your design and make comparative, intelligent decisions faster.

Showcase is essentially a DirectX-based gaming engine used for presenting models created elsewhere. Applications typically export out to the .FBX format and are imported into Showcase for refinement in materials and lighting. You can then develop and assign materials, lighting, and environmental settings; set up alternatives for review; create still shots, transition animations, and storyboards; and essentially create an interactive presentation right from the design models.

**Processor:** Showcase very much likes a fast CPU to import / load files and handle its primary operations.

**RAM:** Showcase consumes a mundane 328MB of system RAM without any loaded scenes. But load up the “House.zip” sample model (about 55MB, including textures), and memory consumption grew to a about a little over 1GB. Expect even higher high memory usage with your models.

**Graphics:** As it relies on DirectX technology to display and work with 3D data, Showcase is very heavily reliant on the GPU for its display operations and almost all tasks depend on the fast display of fully shaded views. Because DirectX 9 is so well supported across all graphics cards, probably any choice you make will run Showcase but it will definitely favor faster gaming cards. As with everything else the more advanced the graphics card the fluid and responsive your work within Showcase will be.

**Storage:** Showcase has the same storage requirements as other applications in the Building Design Suite. Fast subsystems help with application and project load times. Data files can be large but typically not as large as Revit projects.
IV. Processors and Chipsets

Selecting a processor sets the foundation for the entire system and is all about comparing capabilities, speed, and cost. Two processors can be of the same microarchitecture and differ only by 100MHz - which is inconsequential on a 3.4GHz processor - but differ in cost by hundreds of dollars. The microarchitecture of the chip and the process by which it was made advances year after year, so your attention will naturally focus on the latest and greatest models when specifying a workstation.

While there are dozens of CPU models out there, most are differentiated by tiny yet very important details. Use this guide when shopping for workstations to understand just what CPU the vendor has dropped into the system. All of the technical information can easily be compared at http://ark.intel.com.

This section will discuss four primary kinds of Intel CPUs: The latest 5th-generation “Skylake” line of mainstream desktop CPUs, the existing Haswell-E “High End Desktop” (HEDT) lineup, the Haswell EP Xeon E3 / E5 v3 families, and the latest 5th generation Core i7 mobile lineup. Along the way we’ll discuss how Intel develops CPUs over time, what each kind of CPU brings to the table.

We will also look at deciding factors such as chipsets, memory architectures, and expansion capabilities that will segregate workstation classes for the three user profile types discussed at the end of Section I.

By the way, we are focusing only on Intel processors. AMD also makes x86 compatible CPUs, but they do not compete at all against the kinds of CPUs we are looking at for Building Design Suite loads. AMDs CPUs generally do well at the low end of the scale but even there have a tough time competing with Intel on speed, value, and especially power consumption.

Intel’s Microarchitectures and Processes

Before we talk about the specifics in today’s CPU models, we should discuss how Intel develops their chips. This will let you understand what’s under the hood when making processor and platform choices.

First some definitions: The term “microarchitecture” refers to the computer organization of a particular microprocessor model. It is defined as “the way a given instruction set architecture is implemented on a processor.” Microarchitectures describe the overall data pipeline and the interconnections between the components of the processor, such as registers, gates, caches, arithmetic logic units, and larger elements such as entire graphics cores. The microarchitecture decides how fast or slow data will flow through its pipeline and how efficient that pipeline runs. Microprocessor engineers are always looking to ensure no part of the CPU is left unused for any length of time; an empty pipeline means that data somewhere is waiting to be processed and precious cycles are being wasted as nothing gets done.

Every release of a new microarchitecture is given a code name. From the 286 onward we’ve had the i386, Pentium P5, P6 (Pentium Pro), NetBurst (Pentium 4), Core, Nehalem (Core i3, i5, i7), Sandy Bridge, Haswell, and in 2015 we have Skylake. Future microarchitectures will be called Larrabee, Bonnell, and Silvermont. Within each microarchitecture we also get incremental improvements which get their own code names, so keeping each one straight is in itself a big hurdle.

As previously discussed, the term “technology node,” “manufacturing process,” or simply “process” primarily refers to the size of the lithography of the transistors on a CPU, and is discussed in terms of nanometers (nm). Over the past 10 years we’ve gone from a 65nm process in 2006 with the P6 and NetBurst microarchitectures (which included the Pentium 4, Pentium M and Celeron lines); down to a 45m process in 2008 with Core and Nehalem; to a 32nm process with Sandy Bridge in 2010; and to a 22nm.

http://en.wikipedia.org/wiki/Microarchitecture
process with Haswell in 2012. In 2015 we are now down to a 14nm node with Skylake. We are then expected to go down to 10nm in 2016, 7nm in 2018 and 5nm in 2020.  

The Tick-Tock Development Model
To balance the work between microarchitecture and process advancements, Intel adopted a “Tick-Tock” development strategy in 2007 for all of its future processor development cycles. This strategy has every introduction of a new microarchitecture be followed up with a die shrink of the process technology with that same microarchitecture.

In short, a **Tick** shrinks the process technology used in the current microarchitecture. As discussed in the Maximizing Performance per Watt discussion in Section II, shrinking a process is very hard and a big deal, because if it were easy we’d already be at the smallest process possible. Each time Intel wants to shrink the size of the process, they pretty much has to invent tons of new lithography technology along the way to get a laser beam small enough to physically etch the silicon. Remember we are down to 14nm today with roadmaps down to 10nm and then to 7nm. The technical hurdles to get there and still maintain cohesion and stability in a CPUs operation are becoming absolutely huge problems.

In addition to the die shrink, Ticks include small but important tweaks to the CPU cores as well, but usually nothing Earth-shattering. With a Tick you essentially get the same CPU design as last year but, with a smaller process comes lower power consumption / heat / noise, along with bug fixes, new instructions, internal optimizations, and slightly higher performance at the same or lower prices.

Of course, each die shrink Tick also gets a new code name as well. For example, the Westmere “tick” was a 32nm die shrink of the Nehalem microarchitecture which added several new features. Ivy Bridge was a 22nm die shrink of 32nm Sandy Bridge, and Broadwell was a (very short lived) 14nm die shrink of Haswell.

Conversely, a **Tock** is the introduction of an entirely new microarchitecture CPU design based on that now existing smaller process. This is introduced after Intel formally vets the smaller process with the previous tick and has everything working. Every year we can expect one tick or one tock, with some variations.

Legacy Microarchitectures and CPUs: Nehalem to Haswell
Let’s look at a brief history of CPU microarchitectures over the past few years so you can understand where your current system fits into the overall landscape. Then we will dive into the current lineups in greater detail in the next sections.

**1st Generation Tock: 45nm Nehalem in 2008**

Starting in 2008, we had the introduction of the Nehalem microarchitecture (as a Tock), based on the 45nm process introduced the series prior. The new Core i7 lineup of CPUs were the first quad-core processors that provided a noticeably large jump in performance, mostly due to the inclusion of several key new advances in CPU design.

First, there was now a memory controller integrated on the CPU itself running at full CPU speed. Nehalem CPUs also integrated a 16-lane PCIe 2.0 controller. Taken together, these integrations completely replaced the old Front Side Bus and external Northbridge memory controller hub that was used to communicate with system memory, the video card, and the I/O controller hub (also called the Southbridge). This bringing of external functionality onboard to increase performance closer to CPU speeds is something Intel would increase in the future.

Next, Nehalem introduced Turbo Boost, a technology that allows the chip to overclock itself on demand, typically 10-15% over the base clock. We’ll look at Turbo Boost in detail.

Nehalem / Core i7 also reintroduced **Hyper-Threading**, a technology first debuted in the Pentium 4 that duplicates certain sections of the processor allowing it to execute independent threads simultaneously. This effectively makes the operating system see double the number of cores available. The operating system will then schedule two threads or processes simultaneously, or allow the processor to work on other scheduled tasks if the processor core stalls due to a cache miss or its execution resources free up.

Imagine you are in line at the grocery store and the person in front of you has to write a check, or needs someone to look up the price on their Doritos. You are experiencing the same kind of blockages CPUs do. Hyper-Threading effectively is when another cashier opens up their lane and lets you go through. It simply makes the processor more efficient by keeping the lanes of data always moving.

Mainstream Nehalem CPUs in this era were the quad-core Bloomfield i7-9xx series and the Lynnfield i7-8xx series, which were and are still quite capable processors. Bloomfield CPUs were introduced first and carried a triple channel memory controller. This alone increased costs as you had to have memory installed in threes, not twos, and motherboards now required six DIMM slots instead of four. The lower-powered Lynnfield i7-8xx series was introduced later which had a dual-channel memory controller and we were back to four DIMM slots and inexpensive motherboards.

**1st Generation Tick: 32nm Westmere in 2010**

Although not trumpeted as a huge deal at the time, in 2010 we had a Tick (die shrink) of Nehalem to 32nm with the Westmere architecture. Not many people remember this because it was limited to peripheral CPUs and not very many mainstream desktop models. Westmere introduced dual-core Arrandale (mobile) and Clarkdale (low-end desktop) CPUs, the six-core, triple-channel Gulftown desktop and Westmere-EP server variants, and ten-core, quad-channel Westmere-EX, typically found on high-end Xeon CPUs meant for database servers.

In addition to the Core-i7 introduced in Nehalem, Westmere introduced the Core-i3 and Core-i5 variants, each of which targets a specific market segment. We still see them today. Core-i3 CPUs are typically low powered, dual core versions most often seen in ultraportables and very inexpensive PCs, so they are out of contention in any BIM / viz workstation. Core i5 CPUs are quad-core but do not include Hyper-Threading, so they can be considered out of the running as well. Core i7 CPUs are quad-core and include Hyper-Threading, and are the baseline CPUs you should focus on for the purposes of this discussion.

**2nd Generation Tock: 32nm Sandy Bridge in 2011**

In 2011 things got very interesting with new microarchitecture called Sandy Bridge, based on the same 32nm process as Westmere, but with many dramatic internal improvements. Sandy Bridge represented...
an impressive increase in overall performance. Improvements to the L1 and L2 caches, faster memory controllers, AVX extensions, and a new integrated graphics processor (IGP) included in the CPU package made up the major features.

Sandy Bridge was important because it clearly broke away from past CPUs in terms of performance. The on-chip GPU - one of the first to be considered even marginally good - came in two flavors: Intel HD Graphics 2000 and 3000, with the latter being more powerful. This was important for the mainstream user as it finally allowed mid-size desktop PCs (not workstations you or I would buy) to forego a discrete graphics card. Of course, BIM designers and visualization artists require decent graphics far above what an IGP can provide.

Specific processor models included the Core i3-21xx dual-core; Core i5-23xx, i5-24xx, and i5-25xx quad-core; and the Core i7-26xx and i7-27xx quad-core with Hyper-Threading lines. In particular, the Core i7-2600K was an immensely popular CPU of this era, and there are still plenty of Revit and BIM workstations out there based on this chip.

**Sandy Bridge-E in 2011**

In Q4 2011 Intel released a new “Extreme” variant of Sandy Bridge called Sandy Bridge-E. Neither a Tick nor a Tock, it was intended to stretch the Sandy Bridge architecture to higher performance levels with more cores (up to 8) and more L3 cache. The high-end desktop-oriented lineup included the (largely ignored) 4-core Core i7-3820 with 10MB of L3 cache, and the 6-core Core i7-3930K ($550) and the i7-3960X ($1,000) with 12/15MB cache respectively. This introduction of an “extreme” or high-end desktop variants will carry forward with each new microarchitecture.

SB-E technology was incorporated into the Xeon E5-16xx series with 4-6 cores and 10-15MB of L3 cache. The Sandy Bridge-EN variant in the E5-24xx family allowed dual-socket physical CPUs on the motherboard. While the EN product line was limited to at most 2 processors, the Sandy Bridge-EP variant in the Xeon E5-26xx and E5-46xx were slower 6-8 core versions that allowed two or four physical CPUs in a system.

In fact, the 6-core desktop Sandy Bridge-E is really a die-harvested Sandy Bridge-EP Xeon. While the EP-based Xeon will have all 8 cores enabled, the 6-core Sandy Bridge-E simply has two cores fused off.

In particular, these 6-core i7-39xx Sandy Bridge-E’s and Xeon E5s made excellent workstation foundations in their day. Sandy Bridge-E CPUs did not include the IGP – considered useless for workstation use anyway - but did have a quad-channel memory controller that supported up to 64GB of DDR3 system RAM and provided massive memory bandwidth. A quad-channel controller meant memory has to be installed in fours to run most effectively, which required more expensive motherboards with 8 slots.

Another plus for the emerging GPU compute market was the inclusion of 40 PCI Express (PCIe) 3.0 lanes on the CPU, whereas normal Sandy Bridge CPUs only included 16 PCIe 2.0 lanes. The PCIe 3.0 specification basically doubles the bandwidth of PCIe 2.1, where a single PCIe 3.0 8-lane x8 slot runs as fast as a PCIe 2.1 16-lane x16 slot. However, a single modern GPU is pretty tame, bandwidth wise, and you would not see much of a performance delta at all between PCIe 3.0 x8 and x16.

Incidentally, SB-E’s PCIe 3.0 was implemented before the PCIe 3.0 standard was officially ratified, meaning that it was never fully validated for that CPU. This caused technical issues where in some cases cards would default back to PCIe 2.0 speeds, such as NVIDIA’s Kepler series. You could sometimes force PCIe 3.0 mode on SB-E in many cases, but in others you would experience instabilities.

PCIe 3.0’s additional headroom is suited very well to GPU compute as it allows more GPUs to be installed in the system without degrading all of them to the constricting x4 bandwidth. For people who needed
additional graphics card(s) for dedicated GPU compute tasks, the lack of PCIe 3.0 became a deal breaker. See the section on PCI Express for a fuller explanation.

Sandy-Bridge E was important in that it often traded top benchmarks with the later Ivy Bridge due to the addition of two cores and higher memory bandwidth, and represented a solid investment for heavy Building Design Suite users.

3rd Generation Tick: Ivy Bridge in 2012
Hot on the trail of Sandy Bridge-E, we got a Tick die shrink to 22nm with Ivy Bridge in April 2012. Backwardly pin-compatible with Sandy Bridge’s LGA 1155 socket, most motherboards required a simple BIOS update. Ivy Bridge brought some new technologies, such as the 3-dimensional “Tri-Gate” transistor (required because at this tiny scale we need to build better electron gates). We also got a 16-lane, fully validated PCIe 3.0 controller, and relatively small improvements in speed (roughly ~5-10% over Sandy Bridge), but with a remarkably lowered power draw. At this time you started to see the emergence of the “mini-ITX” form factor emerge with very small, very silent, but fully powered desktop designs.

The onboard Intel HD Graphics 4000 GPU was upgraded with full DirectX 11, OpenGL 3.1, and OpenCL 1.1 support. While better than the 3000, it was not fast enough for intense gaming when compared to the discrete card competition, which is why the graphics card market still remained so vibrant.

Overall, the HD Graphics 4000 compared to the ATI Radeon HD 5850 and NVIDIA GeForce GTX 560, both respect able cards for BIM given Revit’s fairly mundane video requirements. For real BIM and 3ds Max users, however, it was best to avoid the IGP and get a dedicated card.

The Ivy Bridge lineup included the dual-core Core i3-3xxx CPUs; the quad-core Core i5-33xx, i5-34xx, and i5-35xx CPUs; and quad-core Core i7-3770K with Hyper-Threading.

Ivy Bridge-E in 2013
2013’s Ivy Bridge-E was the follow-up to Sandy Bridge-E, using the same core as 22nm Ivy Bridge but aimed squarely at the high-end desktop enthusiast (and Building Design Suite user). As with SB-E it has 4 and 6 core variants, higher clock speeds, larger L3 caches, no IGP, 40 PCIe 3.0 lanes, quad-channel memory, and higher prices. It’s typically billed as a desktop version of the Xeon E5.

Unlike SB-E, there is no bat to the knee handicapping here – the 6-core CPUs are truly 6 cores, not 8. IVB-E was great for workstations in that it has fully validated 40 PCIe 3.0 lanes, more than twice that of standard desktop Sandy Bridge, Ivy Bridge, and Haswell parts. This means you can easily install three or four powerful graphics cards and get at least x8 speeds on each one.

The Ivy Bridge-E lineup included three versions: Similar to SB-E, at the low end we had the $320 4-core i7-4820K @ 3.7GHz which was largely useless. The $555 i7-4930K represented the sweet spot, with 6 cores @ 3.4GHz and 12MB of L3 cache. The $990 i7-4960X, which gets you the same 6 cores as its little brother and a paltry 200MHz bump in speed to 3.8GHz, was just stupidly expensive.

One big consideration for IVB-E was the cooling system used. Because of the relatively small die area - the result of 2 fewer cores than SB-E - you have a TDP (thermal design power) of 130W, which is similar to the high-end hot-running CPUs of yesteryear. None of the IVB-E CPUs shipped with an air cooler, and most chose closed loop water cooling for the silence and capability. Closed loop water coolers are pretty common these days, and even Intel offered a specific new water cooler for the Ivy Bridge-E.

4th Generation Tock - Haswell in 2013
June 2013 introduced the new Haswell microarchitecture. Composed of 1.6 billion transistors (compared to 1.4 billion on Ivy Bridge), and optimized for the 22nm process, the CPU was only slightly larger than Ivy
Bridge, even though the graphics core grew by 25%. Internally we got improved branch prediction, improved memory controllers that allow better memory overclocking, improved floating-point and integer math performance, and overall internal pipeline efficiency as the CPU can now process up to 8 instructions per clock instead of 6 with Ivy Bridge. Workloads with larger datasets would see benefits from the larger internal buffers as well.

Haswell was interesting because, while it was clearly faster than Ivy Bridge, it got a refresh halfway through 2014 with the “Devil’s Canyon” i7-4790K, which offered an improved thermal interface material that allowed for a modest increase in clock speed to 4.0GHz, and was supported by a new 9-series chipset (Z97). Coupled with the issues with Broadwell’s release date, the i7-4790K quickly became the go-to CPU in 2014 through almost all of 2015. A year later and it is still a very popular CPU for BIM workstations.

5th Generation Tick – Broadwell in 2014 2015 ...

Broadwell is Intel’s 14nm die shrink (tick) of the Haswell microarchitecture. Broadwell is where we really start to see Moore’s Law start to break down. Intel simply could not execute the 14nm process quickly enough to keep Broadwell on schedule for a widespread 2014 release. It ended up severely delayed until early 2015, which was closing in on the new Skylake microarchitecture release date. Also, Broadwell did not produce the full range of desktop CPUs that could supersede Haswell, and most publications recommended users to hold off on Broadwell and wait for Skylake, which is what just about everyone did.

Turbo Boost Technology Explained

When comparing clock speeds, you will notice that it is no longer given as a single number, but represented as a core clock speed and a “Max Turbo” frequency. Intel’s Turbo Boost Technology 1.0 was introduced in Nehalem processors, and improved single-threaded application performance by allowing the processor to run above its guaranteed base operating frequency by dynamically controlling the CPU’s clock rate. It is activated when the operating system requests higher performance states of the processor.

The clock rate of any processor is limited by its power consumption and temperature, which is driven by the number of cores currently in use and the maximum frequency of the active cores. When the OS demands more performance and the processor is running below its power/thermal limits, a single core’s clock rate can increase in regular increments of 100MHz to meet demand up to the upper Max Turbo frequency. When any of the electrical limits are reached, the clock frequency drops in 100MHz increments until it is again working within its design limits. Turbo Boost technology has multiple algorithms operating in parallel to manage current and temperature levels to maximize performance and efficiency.

Turbo specifications for a processor are noted as a/b/c/d/... notation, where each number is the maximum turbo bin for n, n-1, n-2, n-3 ... n-n-1 active cores respectively, where n is the total number of cores in the processor. For a 4-core CPU, a notation of 8/8/9/10 means that with 3 or 4 cores active, the turbo bin is 8, with 2 cores active the turbo bin is 9, and with one core active the turbo bin is 10.

These bins are multiplied by the standard increment (100MHz in Sandy Bridge and later) and added to the base clock frequency in MHz. For example, the i7-6700 Skylake CPU has a base frequency of 3.4GHz (3400 MHz) and a max Turbo frequency of 4.0GHz. Its Turbo bins are 3/4/5/6, which breaks down as follows:

<table>
<thead>
<tr>
<th>No. of cores active</th>
<th>No. of Turbo bin steps</th>
<th>Calculation</th>
<th>Max frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>3400 + (3 x 100) = 3400 + 300 = 3700</td>
<td>3.7 GHz</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3400 + (4 x 100) = 3400 + 400 = 3800</td>
<td>3.8 GHz</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3400 + (5 x 100) = 3400 + 500 = 3900</td>
<td>3.9 GHz</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3400 + (6 x 100) = 3400 + 600 = 4000</td>
<td>4.0 GHz</td>
</tr>
</tbody>
</table>
Turbo bins for processors are model dependent and are provided at http://ark.intel.com and http://en.wikipedia.org/wiki/List_of_Intel_Core_i7_microprocessors. In general, normal desktop processors have low bin numbers for the first two digits because the core base clock is already set relatively high, leaving lower headroom at the top when three or four cores are active. Last year’s hotness, the Haswell i7-4790K, has a bin of 2/3/4/4 (4.0 to 4.4Ghz) and today’s Skylake i7-6700K has a bin of 0/0/0/2, meaning it has a performance range of 4.0Ghz to only 4.2Ghz.

What is important to note is that the “Max Turbo” rate you see advertised will only be reached for instances when one or at most two processor cores are active and can be boosted. For tasks which are heavily multithreaded and require as much CPU speed as possible (e.g., rendering), all cores will be active and thus the clock speed is only going to ramp up a little bit - if at all.

This is why the Building Design Suite user should concentrate more on the core clock (worst case) speed rather than the Turbo speed when selecting a processor. While the Turbo bin number can vary slightly between models of a given series, 100MHz differences between the active core speeds within a single CPU lineup aren’t anything to get excited about.

Intel has officially supported overclocking by users only on the enthusiast-grade K-series parts with their unlocked multipliers. It is worth noting that with Sandy Bridge and Ivy Bridge processors with “locked” (non-K) models, overclockers soon discovered that they were able to squeeze out a little more performance by increasing the Turbo bins by 4 above stock, effectively adding (at most) a 400MHz overclock. This was a BIOS setting generally found in higher end enthusiast motherboards and not in something you would see from a Dell or HP.

From Haswell moving forward, Intel has removed all Turbo bin tweaking shenanigans, leaving the K-series as the only true overclockable processor.

**Sockets and Chipset Basics**

No discussion of any processor is complete without including both the microprocessor socket and the Platform Controller Hub - or PCH, commonly referred to as the chipset - that needs to marry the CPU to the motherboard. For any CPU line, Intel typically creates a socket standard and a matching PCH to go with it, in various flavors that cater to a certain market.

Today’s CPU provides its own dedicated connections directly to system RAM, up to 40 lanes of direct PCIe-Express (PCIe) connectivity for multiple graphics cards in several configurations, and video I/O for the IGP. The PCH’s contains a high-speed link to the CPU and its primary function is to provide second tier I/O to devices such as additional PCIe lanes, 6Gbs SATA, m.2 NVMe ports for SSDs, high definition audio, more USB ports, and Ethernet.
Historically chipsets have been the most disappointing aspect of any CPU release. Up until recently Intel had a habit of providing somewhat brain damaged chipsets which did not meet the relative power of the CPU it supported and, by extension, the needs of the enthusiast buying the system. Typically they lacked enough latest-gen features such as enough USB 3.0 ports, SATA 6Gbps ports, or limited how its PCIe lanes could be provisioned to support multiple graphics cards.

Because the CPU, in conjunction with the chipset, drives the features exposed on the motherboard in terms of expansion capabilities, your choice of CPU is really not just about raw processor speed. Rather, it is about what the entire platform and its expandability bring to the table to suit your purpose. In fact, that is really what differentiates the builds we will create later in the handout.

If, for example, you are a Viz Wiz using the latest in GPU acceleration for high end renderings and animations, you necessarily need more than one high end graphics card. Perhaps more than a typical desktop CPU + chipset can provide, so you may need to look at the “extreme” side of things with higher-end CPU platforms such as found in Haswell-E and Xeon. We’ll discuss these in depth in our discussion on CPUs later in this section.
Today's Desktop CPU: Skylake Core i7-6700K

6th Generation Tock
Skylake is Intel’s 6th generation, new microarchitecture designed for the same 14nm process as Broadwell. With Broadwell launched months late in mid-2015 and being a case of too little too late, Intel shortened Broadwell’s release cycle to just 60 days and was never really put into circulation on the desktop, instead relegated mostly to mobile platforms. With Skylake’s launch in August 2015, Intel hopes to recover from stumbling in their tick-tock cadence and get them back on track for the next round of new hotness.

Before we get into the specifics, let us first review how Intel names their desktop CPUs:

The Skylake 4-core CPUs use the i7-67xxx and i5-66xx name to distinguish them as their 6th generation processors. The K suffix designates an unlocked CPU multiplier, allowing the processor to be overclocked by the end user by tweaking the core clock multiplier in the BIOS. The i7 series is a 4-core CPU with Hyperthreading (8 threads of execution). The i5 series is a 4-core without Hyperthreading (4 threads).

Other model suffixes available are the S for low-power models with a 65W TDP and T for ultra-low power models with only a 35W TDP. We can remove the T, S, and i5 series from consideration. At the time of this writing this leaves us with only one Skylake CPU model to choose from: the Core i7-6700K.

Skylake i7-6700K Specifics
Skylake is architecturally dissimilar to previous Core designs, and brings slightly greater performance over the previous generation while reducing power consumption. Major changes include the new LGA 1151 socket and the elimination of the fully integrated voltage regulator (FIVR) that was introduced in Haswell only two years prior. This reduces chip complexity and overall power draw, and allows motherboard manufacturers to provide their own voltage regulation implementations on the motherboard. Skylake features the DMI 3.0 bus - a faster 4GB/s connector to the 100-series chipset that brings the expandability enthusiast have been waiting for. The CPU features a 16-lane PCI Express 3.0 controller which will drive a single video card at x16 speed or 2 cards and x8 each.

Skylake’s on-board GPU, the Intel HD Graphics 530, is a rather mundane affair with still too-low performance to be worthy of consideration. In fact, it underperforms the IGP on Broadwell by quite a margin. However, it is fully DirectX 12 compatible ensuring it will work with Windows 10 gaming titles which leverage it and beyond. As it drops support for VGA output, HD 530 can, by itself, to run up to three 4K resolution (4096 x2304) screens using HDMI 1.4, DisplayPort, and DVI interfaces.

Perhaps the biggest change that affects buying decisions is dual-channel DDR4 memory support, along with backward compatibility with existing low-voltage (1.2V) DDR3L memory. DDR4 memory, as discussed later, is a new memory specification promising better performance. Being new, it is slightly more expensive at this time. The type of RAM you would ultimately use depends on the choice of motherboard. Note: Using normal 1.5V DDR3 in a Skylake system can actually damage the CPU.

Other enhancements include support for Thunderbolt 3.0. With a new USB Type-C connector that combines PCI Express, DisplayPort, and DC power all in one cable, Thunderbolt 3.0 allows up to 6 peripherals, including displays, to be supported by one connector.
Speed Shift
Up until Skylake, power management was a task that was split between the CPU and the operating system. When the CPU is operating at its highest guaranteed frequency (the base frequency given in the spec sheets), turbo boost is enabled and performed by the CPU’s power management hardware. When operating at frequencies below the guaranteed rate, the power management is done by the operating system. The system’s firmware tells the OS a range of frequencies to choose from, and the OS picks one based on current workload, power settings, and system temperature. The issue is that power-down and power-up cycling could take a long time reducing efficiency.

With Skylake, power management is more cooperative, with a new power saving feature called Speed Shift technology. The OS still has some say in the matter – the local power setting can force a lowered frequency to save battery life – but the CPU handles the rest. This makes power management far more responsive, allowing the system to “wake up” from a low frequency to high frequency in response to new workloads almost immediately, without waiting for the OS to give the go ahead.

TSX Instructions
The biggest potential news with last year’s Haswell CPU was that it originally was designed with a new extension to the x86 instruction set called Transactional Synchronization Extensions, or TSX. TSX adds transactional memory support, which speeds up the execution of multi-threaded software. It does this by monitoring threads for conflicting memory accesses, aborting or rolling back transactions that cannot be successfully completed.

This is important for developers of multi-threaded applications (such as Autodesk) because writing stable, fast multi-threaded code is hard work. Programmers have to lock areas of memory in order to ensure concurrent threads do not overwrite the same memory location. They do this with complicated programming techniques that lock a wide region of memory (called a table) at a time. Called coarse-grain locking, this practice tends to cause delays and introduces more overhead to multi-threaded code execution and can stall out pipelines.

Consider two people trying to edit the same Excel spreadsheet at the same time. Even though they may be working on different parts of it, a coarse-grained approach is to lock the whole thing down while one users works on it until they finish, during which time the other user is left with nothing to do.

TSX instructions allows programmers to easily write fine-grain locks that lock smaller regions of memory and allow multiple threaded applications to perform much better. In our Excel example, this technique would only lock down individual rows or columns in the spreadsheet. This could potentially increase performance and efficiency, but with a risk of error. TSX-able software would eliminate those errors by moving the evaluation of data integrity from software, which is slow, into the hardware, which is fast. Initial tests indicate TSX-enabled software will perform about 40% faster in specific workloads and provide 4-5 times more database transactions per second.

TSX was intended to be made available in certain Haswell CPUs last year. Unfortunately, in August 2014 Intel confirmed a serious erratum (bug) found with the TSX instruction set that could result in unpredictable system behavior, forcing Intel to disable TSX entirely in all Haswell CPUs. However, the bug has been fixed and today, TSX is fully enabled in the Skylake i7-6700K.

Of course, Autodesk would have to code in TSX support for this to be of great use to Building Design Suite user. However, given that Revit development has pretty well kept pace with advancements in multi-core CPUs, I would expect that now, with TSX issues out of the way, Autodesk would implement TSX capable code in future releases. Of course, how this affects overall performance remains to be seen.
Intel’s High End Desktop Platform with Haswell-E – The Desktop CPU that Goes To 11

As discussed previously, Intel offers what they call the “High End Desktop” or HEDT platform of CPUs that amps up the desktop CPU platform specifically for the enthusiast crowd. Although now formally two generations behind, the Haswell-E remains the best all-around high-end desktop-centric CPU available today for those that make full use of multithreaded applications. Haswell-E is a 22nm processor class and offered in three models: the 6-core Core i7-5820K, the 6-core i7-5930K, and a true 8-core i7-5960X. In many cases it competes well with its Xeon older brothers, as you will see in the next section.

At 2.6 billion transistors, Haswell-E supports only DDR4 memory with its quad-channel memory controller, meaning your RAM needs to be installed in quads instead of twos. Models include up to 20MB L3 cache, 28-40 lanes of PCIe 3.0 connectivity, and of course Hyperthreading. In addition, all Haswell-E processors are unlocked, meaning they are open to overclocking by enthusiasts.

<table>
<thead>
<tr>
<th>Processor Name</th>
<th>Core™ i7-5820K</th>
<th>Core™ i7-5930K</th>
<th>Core™ i7-5960X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cache</td>
<td>15 MB</td>
<td>15 MB</td>
<td>20 MB</td>
</tr>
<tr>
<td>Pricing</td>
<td>Box: $396.00</td>
<td>Box: $594.00</td>
<td>Box: $1059.00</td>
</tr>
<tr>
<td></td>
<td>Tray: $389.00</td>
<td>Tray: $583.00</td>
<td>Tray: $999.00</td>
</tr>
<tr>
<td># of Cores</td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td># of Threads</td>
<td>12</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Processor Base Frequency</td>
<td>3.3 GHz</td>
<td>3.5 GHz</td>
<td>3 GHz</td>
</tr>
<tr>
<td>Max Turbo Frequency</td>
<td>3.6 GHz</td>
<td>3.7 GHz</td>
<td>3.5 GHz</td>
</tr>
<tr>
<td>TDP</td>
<td>140 W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max Memory Size</td>
<td>64 GB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Types</td>
<td>DDR4-1333/1600/2133</td>
<td></td>
<td></td>
</tr>
<tr>
<td># of Memory Channels</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max Memory Bandwidth</td>
<td>68 GB/s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCI Express 3.0 Lanes</td>
<td>28</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>
Notice that the top of the line 8-core i7-5960X is only running with a base clock of 3 GHz, which is 25% slower than the 4GHz Skylake i7-6700K. This is directly due to the TDP (Thermal Design Power) limitation; to get all 8 cores cooking along any faster would produce too much heat, so the highest base clock that would provide lowest TDP was chosen. Water cooling is a must, and Intel offers the same TS13X closed loop solution it did for Ivy-Bridge E in the previous generation.

But it’s FAST. At stock speeds and benchmarking multithreaded applications, the Haswell-E i7-5960X is a great deal faster than the Skylake i7-6700K, despite the 1GHz deficiency in clock speed. In most benchmark tests that stress multi-threading, there is roughly a consistent 45 percent to 50 percent improvement going to the Haswell-E part over its simpler 4-core desktop cousin.

Note the low end i7-5820’s inclusion of only 28 PCIe 3.0 lanes to the others’ 40 lanes. At first glance this may seem like a serious hindrance, but in fact this would only be limiting if you planned to stuff more than 3 GPUs into the system.

With 28 lanes, dual GPU setups would run at x16/x8 speeds, rather than x16/x16 as you could with 40 lanes. Three-GPU setups on a 28-lane chip would run at x8/x8/x8 instead of the x16/x16/x8 they would normally run with a full 40 lanes. However, given that any video card will run at PCIe 3.0 x8 speeds with very little to no degradation in throughput, this still provides solid GPU performance for low money.

Haswell-E uses the LGA 2011-3 socket and is supported by the X99 chipset. This modern PCH fixes some of the sins of the past by supporting up to six USB 3.0 ports and up to ten 6 Gb/s SATA ports. The weak link is that it only provides 8 PCIe 2.0 lanes, whereas Skylake’s Z170 PCH provides an additional 16 PCIe 3.0 lanes. Gigabit networking and audio round out the package. New X99 motherboards support the new m.2 specification for PCIe-based SSDs.

Now that Skylake is launched, the (leaked) roadmap for Intel’s high-end desktop CPU class will make 2016 a very interesting year. We expect to see a Broadwell-E in Q1 2016, followed by a Skylake refresh in Q3 2016 similar to that of Haswell in 2014. Skylake-E should follow in Q3 2016 as well, along with a “tick” (die shrink) of Skylake to 10nm with Cannon Lake.

---

Xeon for Professional 3D Workstations
The Xeon line of CPUs specifically targets the professional 3D workstation and server market, and there is a broad range of cores, speeds, and cache to fit almost any need. Xeons are typically found in high end machines such as the Dell Precision and HP Z series of workstations.

However, understand that the term “workstation” is really just semantics. You can work just fine in all Autodesk Building Design Suite applications on a decent enough desktop machine, and play GTA V all day on a Dell Precision Xeon machine with 6 cores and gobs of RAM (I know this from personal experience). However, there are benefits to Xeon-based workstations to those that require its additional headroom and capabilities. The question to answer is this: Do you push the software in ways such that it will take advantage of a Xeon processor, or are you better served overall with desktop – oriented CPUs like Skylake and Haswell-E. It’s not an easy question.

While you can build a Xeon-based system yourself, it will be a little tougher. Xeon-capable motherboards are not as sold in the open market in the same volume as typical desktop components, and they have fewer differentiating features between them. You as a single buyer do not have the purchasing power of a Dell or HP, so they will most likely be more expensive than what you could get in a packaged workstation. For the DIYer you may find building a system based on the Haswell-E a better proposition.

Xeons traditionally live much longer shelf lives than their desktop cousins and there are many models out there, new and old. You need to review the CPU specifics in your workstation quote, as it may be outfitted with a Xeon from yesteryear, and you end up with older technology instead of the current generation.

Xeon - Key Technical Differences to Desktop CPUs
Xeons offers a few important internal differences over desktop CPUs:

1. Xeons have better virtualization performance, with decreased time to needed to enter and exit a virtual machine. Xeons allow the shadowing of VM control structures, which should improve the efficiency of VM management. In addition, built-in hooks can monitor cache usage by thread so that “noisy neighbors” that affect cache contention can be isolated from other VMs.

2. Xeons are qualified to handle heavier, more intensive loads over a sustained time.

3. Xeons cannot be overclocked easily. Most Xeon-based motherboard are built for stability, not speed.

4. Only the lowest-end quad-core Xeons come with an integrated graphics processor (IGP).

5. Xeons have larger (sometimes much larger) L3 cache, e.g. 15-20MB vs. 8MB L3 cache on Skylake.

6. Higher-order Xeons can have 6 / 8/ 10 / 12 / 18 cores on a single die but will run slower.

7. High-end Xeons support multi-CPU configurations. Desktop / Haswell-E are limited to one CPU only.

8. Xeons support ECC (Error Correcting Code) memory

9. Xeons generally support more installed system RAM (over 64GB) than desktop CPUs.

10. The Quick Path Interconnect (QPI) between CPUs or a CPU to the PCH is faster than the DMI 3.0 interconnect used in Skylake (Haswell-E uses QPI links).

11. Xeon E3 models have dual-channel memory controller, but E5 Xeons, like Haswell-E, utilize quad-channel memory controllers, requiring RAM DIMMs to be installed in fours. However, this has little to no effect in real-world performance.\(^{17}\)

The Xeon Product Lines

Xeons are split into three major product lines. The E3 series is at the low end, basically 4-core Xeon versions of desktop Skylake processors, and are meant for entry-level workstations. The E5 has the broadest appeal, available in a wide variety of configurations. You will find the E5 in single and multi-CPU, medium to high-end graphics workstations. The E7 is in its own weird little world and is strictly for enterprise-level computing, such as database servers, Citrix servers, and in general powering up the cloud. While the Facebook crowd isn’t buying E7 based systems, Facebook itself probably has more than a few.

The naming convention for Xeons is markedly different from desktops but is highly informative at a glance. It follows the nomenclature described in this graphic:

The E# tells you the product code and in general what performance range you are in. The first digit in the 4-digit code after the dash tells you how many you can install in a system (1, 2, 4, and even 8). The second digit indicates the socket type it uses. The last two numbers are the specific product SKU, where the numbers ending in zero or one do not have an integrated graphics processor (IGP), and models ending in 5 or 6 have an IGP. “L” suffixed models are low power and thus are of no concern. The trailing version number (v3, v4, v5) is arguably one of the most important codes as it tells you what microarchitecture the processor is based on. Version-less parts are based on Sandy Bridge, v2 parts were based on Ivy Bridge, v3 parts are based on Haswell, v4 parts are Broadwell, and v5 parts are Skylake.

Xeon Phi

Along with the typical tried and true Xeon family of processors, Intel recently came out with the Xeon Phi coprocessor family. With the push for parallelized computing at an all-time high, the Xeon Phi works to match this workload with providing up to 61 cores, 244 threads, and 1.2 teraflops of performance. (For comparison, a NVIDIA GeForce GTX 980 Ti has a peak compute score of 5.63 TFLOPS). There are a variety of models available as well as form factors, including a standard PCIe x16 card which can be piggybacked onto an existing computer system. As it is brand new with no real track record behind it, the Xeon Phi is beyond the scope of this class. Perhaps in coming years it will be a factor to consider, and perhaps not.

Xeons for Workstations

Xeons serve the high end workstation and in particular the virtual machine / database server markets well because they can handle many more VMs and database transactions without choking on the workload. But the demands on a BIM workstations are very different from that of heavy database and VM servers. For one thing, a single user is not going to peg the processor to the degree that a hundred thousand database transaction every second will. Much of your workload may be graphics related and not be entirely CPU-bound.
To that end, we are going to limit ourselves to a select subgroup of what is currently available. To make the cut for our needs, we are only going to review Xeons that meet the following criteria:

- We are only going to consider CPUs in the E3 and E5 product lines;
- We will consider only models based on the newest Skylake (v5) and Haswell (v3) series;
- Limit the maximum cores per processor to 8 and a minimum core clock speed of 3.2GHz;
- Maximum support of (2) physical processors;
- Models with no integrated graphics processor (IGP)

We limit ourselves to 2 physical CPUs / 8 cores max / 3.2Ghz because this represents the most practical (and affordable) BIM workstation configuration. While you can certainly stuff as many as 8 CPUs into a box with each CPU having upwards of 18(!) cores, the cost to do so is quite prohibitive. Those sorts of configurations are more appropriate for VM servers who host many virtual machines at one time.

Perhaps more importantly, the more cores you have on a CPU the lower the core clock speed must be to handle the thermal stresses, which will hurt single-threaded application performance that would otherwise scream on a desktop 4GHz i7-6700K. Much of what we do in BIM and 3D is still single threaded, where raw speed reigns over the number of cores. Most designers simply never spend enough time crunching rendering problems that would warrant more than 8 cores per CPU. This is particularly true today, now that we have very good cloud-based rendering services, in-house rendering farms, and GPU-based rendering solutions available at a very low cost that can provide imagery far faster than localized CPU-bound rendering solutions can.

**Xeon E3-12xx v5 Family**

The Xeon E3-12xx family actually includes five variants you may see configured in a workstation:

- The original E3-12xx family, released in Q2 ‘11, based on Sandy Bridge;
- E3-12xx v2, released in Q2’12, based on Ivy Bridge;
- E3-12xx v3, released in Q2’13, based on Haswell;
- E3-12xx v4, released in Q2’15, based on Broadwell;
- E3-12xx v5, released in Q4 ‘15, based on Skylake

For the purposes of this discussion, we are going to limit ourselves to considering only E3-12xx v5 Xeons, based on the Skylake microarchitecture. As stated previously, members of the Xeon E3 family are meant for entry level workstations and are really Xeon versions of corresponding desktop CPUs, with very similar specifications. The difference is that while there may only be one worthy desktop model (i.e., the i7-6700K), Intel builds four usable Xeons in this class, separated by 100 MHz clock speeds.

The E3-12xx v5 CPU works in single CPU only configurations and uses the FLGA 1151 socket. As with their desktop cousin, they come with 4 cores, all support Hyper-Threading, have a dual-channel memory controller supporting up to 64GB of both DDR4 and DDR3 RAM, 8MB of L3 cache, and support for the new TSX instructions. Aside from the lower available clock speeds and support for ECC RAM, they are almost identical to the desktop Core i7-6700K. Given the minimum criteria listed above, the contenders are:

<table>
<thead>
<tr>
<th>Processor Name</th>
<th>Cores / Threads</th>
<th>Base Frequency</th>
<th>Max Turbo Frequency</th>
<th>L3 Cache</th>
<th>TDP</th>
<th>PCI-E Lanes</th>
<th>Bus Type</th>
<th>Bulk Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3-1230 v5</td>
<td>4 / 8</td>
<td>3.4 GHz</td>
<td>3.8 GHz</td>
<td>8 MB</td>
<td>80 W</td>
<td>16</td>
<td>DMI 3</td>
<td>Tray: $250.00</td>
</tr>
<tr>
<td>E3-1240 v5</td>
<td>4 / 8</td>
<td>3.5 GHz</td>
<td>3.9 GHz</td>
<td>8 MB</td>
<td>80 W</td>
<td>16</td>
<td>DMI 3</td>
<td>Tray: $282.00</td>
</tr>
<tr>
<td>E3-1270 v5</td>
<td>4 / 8</td>
<td>3.6 GHz</td>
<td>4 GHz</td>
<td>8 MB</td>
<td>80 W</td>
<td>16</td>
<td>DMI 3</td>
<td>Tray: $339.00</td>
</tr>
<tr>
<td>E3-1280 v3</td>
<td>4 / 8</td>
<td>3.7 GHz</td>
<td>4 GHz</td>
<td>8 MB</td>
<td>80 W</td>
<td>16</td>
<td>DMI 3</td>
<td>Tray: $612.00</td>
</tr>
</tbody>
</table>

Source: [http://ark.intel.com/compare/88182,88176,88174,88171](http://ark.intel.com/compare/88182,88176,88174,88171)
**Xeon E3-12xx v5 Notes**

1. All models end in a 0 (zero) which indicates they do not have an IGP.
2. Pricing is given in 1000-bulk trays to system integrators. Box prices are usually $11 more.
3. Any 100MHz increase in speed is at best a paltry 2-3% increase in performance, which is negligible.
4. Note the very high price jump of $273 from the E3-1270v3 @ 3.6 GHz to the E3-1280v3 at 3.7 GHz. Paying that much extra for a tiny 100MHz speed increase is just dumb.

**Xeon E3-12xx v5 Analysis and Conclusion**

For the E3 family, I think the **E3-1270 v5** @ 3.6 GHz at $339 holds the sweet spot right now, but even then it’s not all that attractive. Remember that the latest Skylake i7-6700K is running at a base speed of 4 GHz for the exact same price. Given that the other specifications are almost exactly the same between the two, you should not pay a premium for a “workstation” class machine that is, internally, physically the same but slightly slower that a mainstream desktop. Clearly, there’s no advantage to buying a Xeon E3-12xx workstation over a more mundane desktop i7-6700K.

**Xeon E5-16xx v3 Family**

The Xeon E5 represents the broadest implementation of Intel’s Haswell-based workstation / server class CPU and, with their increased number of cores, is generally recommended for BIM workstations that have to do a lot of multithreaded processing such as rendering.

The Xeon E5-16xx v3 is built on the Haswell 22nm microarchitecture and very similar to the Haswell-E discussed previously. For the E5-16xx v3 we have four choices that are appropriate:

<table>
<thead>
<tr>
<th>Processor Name</th>
<th>Cores / Threads</th>
<th>Base Frequency</th>
<th>Max Turbo Frequency</th>
<th>L3 Cache</th>
<th>TDP</th>
<th>PCI-E Lanes</th>
<th>Bus Type</th>
<th>Bulk Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5-1620 v3</td>
<td>4 / 8</td>
<td>3.5 GHz</td>
<td>3.6 GHz</td>
<td>10 MB</td>
<td>140 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $294.00</td>
</tr>
<tr>
<td>E5-1630 v3</td>
<td>4 / 8</td>
<td>3.7 GHz</td>
<td>3.8 GHz</td>
<td>10 MB</td>
<td>140 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $372.00</td>
</tr>
<tr>
<td>E5-1650 v3</td>
<td>6 / 12</td>
<td>3.5 GHz</td>
<td>3.8 GHz</td>
<td>15 MB</td>
<td>140 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $583.00</td>
</tr>
<tr>
<td>E5-1680 v3</td>
<td>8 / 16</td>
<td>3.2 GHz</td>
<td>3.8 GHz</td>
<td>20 MB</td>
<td>140 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $1,723.00</td>
</tr>
</tbody>
</table>

Source: [http://ark.intel.com/compare/82763,82764,82766,82767,82765](http://ark.intel.com/compare/82763,82764,82766,82767,82765)

**Xeon E5-16xx v3 Notes**

1. The Xeon E5-16xx v3 family very closely resembles the Haswell-E high end desktop CPUs in almost every respect. For example, the Xeon E5-1650 v5 almost exactly matches the specs for the Haswell-E i7-5930K; the Xeon E5-1680 v3 matches specs with the Haswell i7-5960X.
2. Xeon E5-16xx processors are for uniprocessor systems only.
3. Being based on the Haswell microarchitecture, these CPUs do not have TSX instructions enabled. To date there is no concrete information as to how important this is currently, or how critical it could become in the future.
4. All Xeon E5-16xx v3 series CPUs provide a 40 lane PCIe 3.0 controller.
5. Notice the variations in clock speed. As more cores and L3 cache is added, the clock speed either stays the same or falls slightly. Note that the E5-1660 v3 (8 cores with 20MB cache) is not included because its clock speed is only 3.00 GHz and below the threshold.
6. Xeon E5-16xx v3s have a 4-channel memory controller and support up to 768GB of DDR4 RAM.
Xeon E5-16xx v3 Analysis and Conclusion
Comparisons to the high-end desktop Haswell-E are inevitable, as the Xeon E5-1650 v3 is almost spec for spec identical to the Haswell-E i7-5930K – exactly the same number of cores, clock speed, L3 cache, TDP, PCIe lanes, and even price. This 6-core Xeon is my pick of the litter in this class.

At $1,723, the 8-core E5-1680 v3 @ 3.2GHz is almost $700 more than the Haswell-E i7-5960X @ 3GHz. While the Haswell-E’s i7-5960X, at 3GHz, is slower than our 3.2GHz minimum cutoff for consideration, the i7-5960X can be (and often is) overclocked to around 4.5GHz on average\(^4\). In fact, some vendors provide i7-5960X based systems already overclocked out of the box, complete with water cooling solutions. Bottom line: If you need 8 cores, build a Haswell-E i7-5960X system and forego the $700 Xeon tax.

Xeon E5-26xx v3
The Xeon E5-26xx v3 can be implemented on a single or dual-CPU system. While there many models to choose from, only three contestants in this class meet our minimum specs:

<table>
<thead>
<tr>
<th>Processor Name</th>
<th>Cores / Threads</th>
<th>Base Frequency</th>
<th>Max Turbo Frequency</th>
<th>L3 Cache</th>
<th>TDP</th>
<th>PCI-E Lanes</th>
<th>Bus Type</th>
<th>Bulk Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5-2637 v3</td>
<td>4 / 8</td>
<td>3.5 GHz</td>
<td>3.7 GHz</td>
<td>15 MB</td>
<td>135 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $996.00</td>
</tr>
<tr>
<td>E5-2643 v3</td>
<td>6 / 12</td>
<td>3.4 GHz</td>
<td>3.7 GHz</td>
<td>20 MB</td>
<td>135 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $1,552.00</td>
</tr>
<tr>
<td>E5-2667 v3</td>
<td>8 / 16</td>
<td>3.2 GHz</td>
<td>3.6 GHz</td>
<td>20 MB</td>
<td>135 W</td>
<td>40</td>
<td>QPI</td>
<td>Tray: $2,057.00</td>
</tr>
</tbody>
</table>

Source: [http://ark.intel.com/compare/83361,81900,83358](http://ark.intel.com/compare/83361,81900,83358)

Xeon E5-26xx v3 Analysis and Conclusion
Clock speeds for the E5-26xx series are a little slower than E5-16xx, between 3.2-3.5GHz, and their prices are twice as high on average. The strength of the lineup is the ability to work in dual-CPU configurations, so they should only be considered for such systems. Two 8-core E5-2667s with 16 cores / 32 threads will smack down everyone else when it comes to CPU-bound rendering tasks. Of course, it will cost over $4,100 for the CPUs alone, which would easily purchase another entire machine that could be used in a rendering farm. With the advent of cloud computing and distributed rendering, the need for a single 3D workhorse to handle what are typically intermittent high workloads is getting smaller.

Choosing a Mobile CPU and Platform
When talking about choosing a mobile platform (laptop, tablet, two-in-one, mobile workstation, etc.) you have many choices and more things to consider. Mobile architectures are pretty closed; there is not much leeway in choosing components. Your first task is to identify how the machine will function in your life. Will it be a mobile ancillary machine to your desktop, where you do light duty modeling, take it on the road or into meetings to show off models, and make presentations? Or will be it a truly mobile workstation and desktop replacement that needs to crunch through large Revit and 3ds Max models all the time?

Regardless of any other consideration, you will want as much power in the most mobile friendly platform available. Balancing power and portability is a catch-22 situation for BIM / 3D visualization purposes, as faster laptops are traditionally bulkier than slower ones, so the faster the machine you have, the larger, heavier, noisier, and generally more unpleasant it can be to use on a daily basis.

In the case of the ancillary machine, you would probably choose the form factor first. It is light and nice to carry around? Does it have enough juice to run the Building Design Suite applications at least marginally well for the purpose of design review and presentation?

\(^4\)[http://www.anandtech.com/show/8426/the-intel-haswell-e-cpu-review-core-i7-5960x-i7-5930k-i7-5820k-tested/3](http://www.anandtech.com/show/8426/the-intel-haswell-e-cpu-review-core-i7-5960x-i7-5930k-i7-5820k-tested/3)
In the case of a mobile workstation, it needs to be powerful enough to run the software all the time – portability is secondary. Of course, you also need to take into account how much production, modeling and rendering you would actually need to do on it to justify the mobility in the first place.

Luckily, the increase in Power per Watt means all components are getting smaller, cooler, and more efficient while still maintaining or increasing performance. With the latest Broadwell and Skylake 14nm generation combined with advancements in storage and graphics, our mobile workstations are getting more mobile all of the time.

Intel's overall strategy is to build out a broad, scalable Skylake platform that provides a microprocessor for every form factor and type of device out there on the market. On the desktop we have the i3, i5, and i7 families each of which caters to a particular power profile. On the mobile side we have several families which will serve everything from tablets and detachable 2-in-one devices, to thin and light notebooks and portable all-in-ones, to powerful mobile performance laptops and workstations.

6th Generation Mobile Core i7 Processors
To that end we are going to focus on the needs for the mobile workstation class of machines which can serve as desktop replacements for users of the Building Design Suite.

Similar to its desktop counterpart, Intel’s 6th generation of mobile processors includes the Core i7-6xxxHQ (high performance graphics, quad-core) series, which is considered the most appropriate for a mobile BIM workstation. This 45W CPU offers Hyperthreading (8 threads) and the same Intel Graphics 530 IGP as is on the desktop i7-6700K.

That said, here are the following Skylake models used in typical mobile workstation class machines:

<table>
<thead>
<tr>
<th>Processor Name</th>
<th>Cores / Threads</th>
<th>Base Frequency</th>
<th>Max Turbo Frequency</th>
<th>L3 Cache</th>
<th>TDP</th>
<th>IGP</th>
<th>Bulk Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>i7-6700HQ</td>
<td>4 / 8</td>
<td>2.6 GHz</td>
<td>3.5 GHz</td>
<td>8 MB</td>
<td>45 W</td>
<td>HD Graphics 530</td>
<td>Tray: $378.00</td>
</tr>
<tr>
<td>i7-6820HQ</td>
<td>4 / 8</td>
<td>2.7 GHz</td>
<td>3.6 GHz</td>
<td>8 MB</td>
<td>45 W</td>
<td>HD Graphics 530</td>
<td>Tray: $378.00</td>
</tr>
<tr>
<td>i7-6920HQ</td>
<td>4 / 8</td>
<td>2.9 GHz</td>
<td>3.8 GHz</td>
<td>8 MB</td>
<td>45 W</td>
<td>HD Graphics 530</td>
<td>Tray: $568.00</td>
</tr>
<tr>
<td>i7-6820HK</td>
<td>4 / 8</td>
<td>2.7 GHz</td>
<td>3.6 GHz</td>
<td>8 MB</td>
<td>45 W</td>
<td>HD Graphics 530</td>
<td>Tray: $378.00</td>
</tr>
</tbody>
</table>

Source: [http://ark.intel.com/compare/88969,88967,88970,88972](http://ark.intel.com/compare/88969,88967,88970,88972)

I7-6xxxHQ Notes, Analysis, and Conclusion

1. Any mobile workstation that is going to be a daily driver is going to have a separate GPU installed in the system, so the IGP is of no real concern.
2. Right off the bat, we see that the i7-6700HQ and i7-6820HQ are the same price, but the i7-6820HQ is slightly faster, making it the better buy.
3. Note the range between the base frequency and the turbo frequency. For single threaded, low end workloads, the CPU can crank up into the mid 3GHz range, competing with desktop CPUs. With anything more stressful the additional cores get busy and the clock frequency drops way down.
4. There is one nice surprise in the lineup: an unlocked mobile CPU. The Core i7-6820HK enables the overclockable laptop. I can’t imagine anyone would want to overclock a laptop, but, well, there it is.
5. The i7-6920HQ is the performance king, but only by 200 MHz - only 7.4% faster. Put another way: The i7-6920HQ costs $196 per GHz. The i7-6820HQ costs $140 per GHz, making it the best deal here.
A Word on Evaluating CPU Benchmarks

This handout is not providing exhaustive benchmarks for processors. That’s primarily because it’s been difficult to impossible to find a steady state of independent benchmark tests that are applied to the several models that we would be most interested in. Most enthusiast web sites benchmark a CPU’s gaming prowess over everything else, which emphasizes certain properties in the CPU that are much different than what is stressed in Building Design Suite applications. These sites do not pay much attention to the workstation market and typically do not benchmark Xeon models. The ones that do benchmark Xeons attempt to benchmark gaming performance, so we’re back to apples to oranges.

One point of warning. While you can check out a large CPU benchmark comparison chart at www.cpubenchmark.net, understand what it is: It is an average of user-submitted benchmarks from Passmark’s CMU Mark benchmarking tool. The problem is that we don’t know what motherboards, installed RAM, overclocking status, or even operating systems are involved with any of the results, so they can be skewed.

For example, cpubenchmark.com lists the following results for the most popular processors of yesteryear as well as those discussed in this handout:

<table>
<thead>
<tr>
<th>Processor Model</th>
<th>Cores / Threads</th>
<th>CPU Speed (Base - Max Frequency)</th>
<th>Passmark – CPU Mark Score¹⁹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynnfield Core i7-860 (c.2009)</td>
<td>4 / 8</td>
<td>2.8 – 3.5 GHz</td>
<td>5,098</td>
</tr>
<tr>
<td>Sandy Bridge Core i7-2600K (c.2010)</td>
<td>4 / 8</td>
<td>3.4 – 3.8 GHz</td>
<td>8,584</td>
</tr>
<tr>
<td>Ivy Bridge Core i7-3770K (c.2012)</td>
<td>4 / 8</td>
<td>3.5 – 3.9 GHz</td>
<td>9,628</td>
</tr>
<tr>
<td>Haswell “Devil’s Canyon” Core i7-4790K</td>
<td>4 / 8</td>
<td>4.0 – 4.4 GHz</td>
<td>11,291</td>
</tr>
<tr>
<td>Skylake i7-6700K</td>
<td>4 / 8</td>
<td>4.0 – 4.2 GHz</td>
<td>11,928</td>
</tr>
<tr>
<td>Haswell-E Core i7-5930K</td>
<td>6 / 12</td>
<td>3.5 – 3.7 GHz</td>
<td>14,202</td>
</tr>
<tr>
<td>Haswell-E Core i7-5960X</td>
<td>8 / 16</td>
<td>3.0 – 3.5 GHz</td>
<td>15,984</td>
</tr>
<tr>
<td>Xeon E3-1230 v5</td>
<td>4 / 8</td>
<td>3.4 – 3.8 GHz</td>
<td>10,735</td>
</tr>
<tr>
<td>Xeon E3-1240 v5</td>
<td>4 / 8</td>
<td>3.5 – 3.9 GHz</td>
<td>10,477</td>
</tr>
<tr>
<td>Xeon E3-1270 v5</td>
<td>4 / 8</td>
<td>3.6 – 4.0 GHz</td>
<td>11,332</td>
</tr>
<tr>
<td>Xeon E3-1280 v5</td>
<td>4 / 8</td>
<td>3.7 – 4.0 GHz</td>
<td>12,807</td>
</tr>
<tr>
<td>Xeon E5-1650 v3</td>
<td>6 / 12</td>
<td>3.5 – 3.8 GHz</td>
<td>13,659</td>
</tr>
<tr>
<td>Xeon E5-1660 v3</td>
<td>8 / 16</td>
<td>3.0 – 3.5 GHz</td>
<td>23,168</td>
</tr>
<tr>
<td>Xeon E5-1680 v3</td>
<td>8 / 16</td>
<td>3.2 – 3.8 GHz</td>
<td>15,532</td>
</tr>
<tr>
<td>Xeon E5-2637 v3</td>
<td>4 / 8</td>
<td>3.4 – 3.7 GHz</td>
<td>10,519</td>
</tr>
<tr>
<td>Xeon E5-2643 v3</td>
<td>6 / 12</td>
<td>3.4 – 3.7 GHz</td>
<td>14,202</td>
</tr>
<tr>
<td>Xeon E5-2667 v3</td>
<td>8 / 16</td>
<td>3.2 – 3.6 GHz</td>
<td>17,884</td>
</tr>
<tr>
<td>Xeon E5-2687W v3</td>
<td>10 / 20</td>
<td>3.1 – 3.5 GHz</td>
<td>15,848</td>
</tr>
<tr>
<td>Xeon E5-2699 v3</td>
<td>18 / 36</td>
<td>2.3 – 3.6 GHz</td>
<td>35,409</td>
</tr>
</tbody>
</table>

Sources:
https://www.cpubenchmark.net/high_end_cpus.html

¹⁹ Note: Some scores from cpu-monkey.com are calculated, not actual, and are not endorsed by Passmark
Passmark’s *CPU Mark* is a processor benchmark and thus will naturally favor more cores, but it also oddly favors the desktop models as well. For example, the 8-core Haswell-E Core i7-5960X scored faster than the 10 core Xeon E5-2687W v3, which could be explained away by the relatively slow clock speed of the Xeon. But the 6-core Haswell-E Core i7-5930K scores slightly better than the 6-core Xeon E5-1650 v3 with the same base clock speed.

When looking at these benchmarks, bear in mind that Xeons are integrated into vendor specific workstations which are primarily concerned with stability over speed. They are not overclocked and have system BIOSes with voltage settings tuned very conservatively. Whereas the enthusiast gamer is probably tweaking out their DIY desktop systems, which can easily be put together from the open components market.

**The Final Words on Processors**

The latest Skylake desktop and Haswell-E high-end desktop processors perform very well, coming close to or downright beating the Xeons in several tests, particularly ones that stress multimedia and single-threaded performance. Where the high-end E5 Xeons will rule in Building Design Suite applications is in highly multithreaded applications and scenarios, such as working on very large Revit models and the 3ds Max dedicated graphics workstation, particularly with dual CPUs.

**For the Grunt:**

For most mainstream applications, as well as the probably 90% of your BIM Grunts, the Skylake i7-6700K CPU will do nicely at an appropriately moderate price point. The i7-6700K is the next evolution beyond last year’s already excellent Haswell i7-4970K, and about 10% faster per clock. It is technically and financially equivalent to its Xeon cousin the E3-1270 v5, but is clocked a little higher. Both provide excellent baseline performance for designers of all kinds.

**For the BIM Champ:**

For advanced Revit users who push the software with Dynamo and integrate it with various other applications, a faster machine is in order. For these folks I like the Haswell-E 6-core i7-5930K with all 40 PCIe lanes for multi-GPU setups. If the user does any serious rendering in traditional renderers, the 8-core i7-5960X is a great but expensive alternative. Both processors will push performance higher than the Skylake into Xeon territory without breaking the bank.

**For the Viz Wiz**

For high-end Revit and 3ds Max Design use, where your need for fast local rendering trumps most everything else, the Haswell-E i7-5960X is a great value proposition, with dual 8-core Xeon E5-2667 v3 representing a great 3D workstation. Outfit it with 64GB of RAM with a great video card or three, and watch the images fly out of the machine.

Beyond this, if you really need as much horsepower as you can throw at your design and visualization problems, look at systems with dual physical CPUs and/or very high core counts. Near the absolute top of the line is the Xeon E5-2697 v3, with 35MB of L3 cache and 14 cores, which rockets away with a Passmark score of 22,147, about twice as fast as the Skylake i7-6700K. At $2,702 per CPU, it’s a tough one to swallow, until you need those renderings out the door by the end of business.

If you are thinking of buying dual-CPU systems to shave down rendering times, you should consider three alternatives: (a) Distributed rendering across a farm using Backburner; (b) Try GPU based rendering with iray, and / or (C) Cloud based rendering using Amazon EC2 or other services. Any and all of these solutions are intrinsically cheaper and potentially much more powerful than a very expensive single machine.
V. System Memory

Memory has always been the least exciting thing to worry about when specifying a new workstation, but not anymore. Back in olden times (2012 or so) you bought 8 or 16GB or so of standard off the shelf cheap DDR3 DRAM and were happy, dammit. Conventional wisdom is to specify 16GB for a basic “Grunt” level Revit workstation that doesn’t do a lot of rendering, 32GB for a BIM Champ machine, and 64GB + for a high-end 3D Viz Wiz machine that is expected to take on constant rendering. For most common workloads, there is a case of diminishing returns after 16GB.

With today’s latest CPUs on tap, purchasing RAM in a new system is not so simple, as there are various memory chip speeds and capacities that factor into the equation.

**DDR4 SDRAM DIMMs**

Today’s workstation memory standard for Skylake, Haswell-E X99, E3, and E5 Xeon systems is a 288-pin Double Data Rate Type Four, Synchronous Dynamic Random Access Memory, Dual Inline Memory Module, or DDR4 SDRAM DIMM. Because the choice of CPU is always paramount, and the RAM used is dependent on the CPU, you really don’t have much choice is whether you use DDR3 or DDR4. Your choices are in capacity and DIMM speed.

DDR4 has been a long time in the making and market acceptance has been slow, primarily because only Haswell-E and newer Xeons required it. Now that Skylake is out on the desktop, DDR4 will begin to gain much more market share and prices will naturally drop to DDR3 levels.

There are several key differences between DDR3 and DDR4. First, DDR4 operates at a lower voltage at 1.2V than DDR3 at 1.5V. This equates to about 18W in power savings over DDR3 in a typical system with 4 DIMMS. Not much, until you factor in server farms and data centers and large scale architectures with thousands of DDR4 modules, and it adds up.

The other key difference is speed. DDR3 starts at 800 MT/s (mega-transactions per second) up to as high as 2133 MT/s. DDR4 picks up where DDR3 leaves off at 2133 MT/s and goes up from there, massively increasing memory bandwidth. However, real-world benchmarks reveal that increasing memory bandwidth by itself does not improve application performance to the degree you would think.

This increase in speed results in an increase in latency as well. You will see this listed as CAS (Column Access Strobe) Latency, or CL, in memory specs. It is the delay time between the moment a memory controller requests a particular memory column on a RAM module and the moment the data from the given location is available on the module’s output pins. CL is given as a simple integer and the lower the CL, the better the module performs.

While a DDR3 2133 module may have a CL of 9 or 10, DDR4 2133 will have a CL of 15 or higher. However, the importance of this in overall response time is somewhat negligible overall. If you are purchasing a packaged system you will have no say in the specific modules, but if shopping for a DIY system, you want modules with the lowest CL value. Note that when purchasing RAM for a BIY system or upgrades, you need to ensure the CL number is the same for all DIMMs in the system.

The DDR4 DIMM module is subtly different from DDR3 in a few ways. First, they are pin-incompatible, meaning you cannot physically install the wrong module in a system. DDR4 are 288-pin DIMMs whereas DDR3 are 240 pin modules, so the pitch between the pins is reduced from 1.0 mm to 0.85 mm.

Physically there is a change in the bottom edge connector for DDR4, which is curved down in the middle area somewhat that lowers the insertion force required for installation in a motherboard slot.
DDR RAM Channels and Configuration Rules

When configuring memory for your system, it is important to understand some of the internals of your particular CPU. Because different CPUs have different memory controllers, memory configurations aren’t all the same.

Modern CPUs have embedded memory controllers which directly talk to the memory DIMM through two, three, or four separate channels that double, triple, or quadruple the communication bandwidth between the CPU and system memory. Each channel is a 64-bit pathway to one or more memory modules. The DIMMs on these channels are talked to by the controller simultaneously in parallel and are treated as a single entity, as far as the CPU is concerned.

By doing so your system memory can minimize the effects of high latencies and timing delays that occur with fewer memory modules. Otherwise, RAM falls back to single data rate speeds and system performance suffers. This is why memory slots are color coded to indicate the channels, so you know to install RAM to fill the same-colored slots before moving on to fill others.

The next constraint is the number of slots available. Skylake and Xeon E3 systems have four RAM slots for the four channels. Haswell-E and Xeon E5 typically have at least 8 slots to provide a pair of quad channels. Thus, if you aren’t maximizing your RAM capacity when you purchase the system, you want to determine the module size such that it fills up 4 slots (channels) with the largest capacity at a time, so that you don’t populate all of your slots with smaller modules and leave it problematic to add more memory later.

---

**DDR Density and Pricing**

DDR4 memory has the advantage that it can support single modules of higher densities, so 16GB, 32GB, on up to 512MB modules are theoretically possible. DDR3 has a theoretical maximum of 128GB per DIMM. Checking Newegg.com prices, 4GB DDR4 modules are about $28, or $7/GB. 8GB DDR4 modules are around $50, or $6.25/GB, and 16GB modules are going for $130, or $8.13/GB. Thus, 8GB modules seem to be the sweet spot between capacity and $/GB. Given the constraints of the number of DIMMS to install (governed by the memory controller) and number of slots, it makes the most sense to do the following:

1. Skylake systems have a dual memory controller and support up to 32GB of RAM, so max out the available slots by installing (4) 8GB modules. Use (2) 8GB modules for a 16GB system.
2. Like Skylake, Xeon E3-12xx v5 CPUs have a dual-channel memory controller but can address up to 64GB, so (4) 16GB modules would max out your capacity and slots.
3. Haswell-E can address up to 64GB and X99 motherboards have 8 slots, so you would install (8) 8GB modules to max out the RAM. If you choose to only install 32GB you would install (4) 8GB modules and still enable quad-channel bandwidth, with 4 slots open for a future upgrade.
4. For Xeon E5 systems which can address up to 768GB (!) of RAM, memory configuration is limited to the number of slots on the motherboard and available DIMM module size. If you wanted 128GB of RAM – probably the most you would ever need – either (8) 16GB modules or (4) 32MB modules would work.

**DDR DRAM Speeds**

DDR4 DIMMs can be specified and purchased at different speeds, provided as two standard name formats: DDR4 XXXX where XXXX = the data transfer rate of the raw chips expressed in mega-transactions per second; or PC4 NNNN where NNNN is the total theoretical bandwidth of the entire module in MB/s. Note: None of these numbers refers to the modules speed in MHz, GHz, or whatever. Enthusiast web sites get this wrong all of the time.

DDR4 memory starts at a transfer rate of DDR4 2133 and ramps up to DDR4 4000. Popular speeds are DDR4 2400 (PC4 19200), DDR4 2800 (PC4 22400), and DDR4 3000 (PC4 24000).

However, understand that with increased latency, higher speed DDR4 RAM may not provide much if at all speed benefit by itself. RAM throughput doesn’t scale well to increase performance in applications, so paying more for special very fast RAM, after a point, is of no real benefit. Today’s benchmarking indicates that DDR4 2666 is likely the sweet spot, with faster modules producing inconsequential gains.

**ECC Memory is not Required in a Workstation**

On Xeon workstations, you will most likely find options to get Non-ECC or ECC memory. ECC stands for Error Correcting Code, a standard specification that can detect and correct most common kinds of internal data errors and corruption. ECC memory is specified in systems where data corruption cannot be tolerated under any circumstances, e.g. scientific or financial computing programs.

Support for ECC memory is limited strictly to the Xeon and not supported on Core i7 desktop or Haswell-E platforms. ECC memory is more expensive than non-ECC and the error correction mechanism lowers memory performance by about 2-3 percent. Workstation system integrators will often default to ECC memory in their preconfigured systems. However, in many cases you can specify non-ECC RAM even on high-end workstations without issue. DRAM manufacturers have much tighter tolerances on their fabrication processes and the propensity for memory errors is much lower than it was just five years ago.
VI. Graphics Cards

Selecting the right graphics card for the Building Design Suite can be problematic, to say the least. There is a lot of uncertainty about how GPUs work in general and how Autodesk applications use them to make decisions difficult. Once you understand some basics, it’s actually not that hard, and this section will attempt to dispel some misconceptions you may have about what to get.

To a large degree, the capabilities and performance of the card is proportional to the cost of the card, but only to a point. Today, simpler operations such as 2D CAD are universally covered adequately across all cards that range from the low end to well over $1,000.

What separates the better graphics cards from the pack is how much time they can save you in wholly GPU-bound operations, such as orbiting in shaded 3D views, rendering using the iray renderer, or pushing out complex animations in Lumion. These are the differentiating factors that go into a decision; if you do those things all of the time, you need to pick a strong performer in that area. Otherwise, you could perhaps pick a more moderately priced card. Or pick something perhaps up the scale that you can grow into as you adopt more GPU-bound operations in your daily work. The key point is that you need to let your computing needs determine the class of card you need.

Professional and Gaming Video Cards

The choice of graphics solution for any professional machine starts by looking at the two primary kinds of cards being offered. In one corner you have expensive, professional workstation-class cards meant for the “design professionals” in the tight blue jeans and Daniel Libeskind glasses. In the other corner you have a plethora of cheap gaming cards meant for the dude living in his parent’s basement.

It may surprise you, but it is no secret that the GPU architecture used in professional cards like ATI’s FirePro and NVIDIA’s Quadro lines are almost always identical to those used in their Radeon and GeForce gaming cards. The GPUs may have different code names but internally they are remarkably the same. For example, the architecture used in the top shelf AMD FirePro W family is the same design used in the Radeon HD 7xxx series desktop boards. The GM204 GPU, with 2,048 CUDA cores used in the $2,000 Quadro M5000 is the same GPU used in the $510 GeForce GTX 980. AMD and NVIDIA may fine tune the innards of each for a particular target price point (typically by tweaking core clock rates), but the vast majority of the plumbing is the same. Carefully examining the benchmarks proves this out.

Even then, NVIDIA and ATI still make a lot of money by selling professional level cards with professional level markups attached to them. There are several reasons why professional cards cost what they do:

1. Professional graphics cards have drivers that are certified by Autodesk, Catia, Solidworks, and a dozen other application developers, to ensure the card meets the minimum specification under each application. This is expensive and costs NVIDIA / ATi money. However, there is clear evidence to suggest that this can potentially means quite a bit, particularly for applications who are built around OpenGL instead of DirectX. Cards which are not certified may not be able to run those applications well or at all because their drivers are terrible.

2. Professional-class drivers may take advantage of specific features in those cards meant for high-end content creation, such as special antialiasing, z-buffering, and others. Gaming card drivers may or may not expose that functionality because those features are considered the domain of professional content developers, not consumers (gamers).
3. Workstation class cards may include more onboard RAM, although today this is somewhat ambiguous as modern game titles (and therefore gaming enthusiasts) demand more video memory. Currently, the amount of onboard video RAM is most critically important for iray rendering on NVIDIA-based graphics cards. However, NVIDIA is in the process of changing how this works under the hood which may alleviate these memory bottlenecks in future microarchitectures.

4. ATI and NVIDIA oversee the fabrication of their professional line of graphics cards, with only a small number of manufacturers. This small number of vendors reduces competition which raises prices. Gaming cards are all built by independent manufacturers like ASUS, Gigabyte, and EVGA, from a reference design by NVIDIA and ATI.

5. Professional cards have tighter tolerances for build quality. These are generally built to run under more rigorous conditions, like 24 x 7 x 365 inside of server racks. To that end they are clocked slower to lower TDP and lengthen their lifespans. Gaming cards are meant for speed, first and foremost, and aren’t built to run full tilt all day long. However, the demands of intense GPU based rendering can take its toll on the silicon, and causes temperatures to rise and cooling fans to work overtime. Cheap gaming cards have been known to go up in smoke - literally - when pushed to render many scenes.

While it’s a given that professional cards can and do well in Autodesk applications, the strategy by Autodesk, workstation vendors, and NVIDIA/AMD to promulgate professional-class cards is hindered by two facts: First, the vast majority of applications in the Building Design Suite are mostly graphics card agnostic. They have to be, in order for them to work across most PCs. All Autodesk applications use the traditional Windows DirectX 9 API as a minimum standard, and most applications now use features in DirectX 11. All of today’s graphics cards have passed that standard long ago and will work fine.

That’s not to say that a professional card won’t work as well or better, but the idea that you “have” to run a professional card in any Autodesk applications is pure nonsense. Professional cards have vetted drivers which may reduce the risk of seeing graphic anomalies in Revit, Navisworks, or 3ds Max, but nothing is guaranteed.

Sometimes someone makes the argument that, because Autodesk has not certified a particular card, it should be discounted immediately as not appropriate. Remember that Autodesk is not in the business of making sure your card works. It has limited resources to test new hardware and the hundreds of drivers that appear almost daily, so it is impossible for Autodesk to keep on top of this. Autodesk develops its software using the industry standard DirectX API, and both NVIDIA and ATI understand how to code their drivers for the Windows environment very well.

On the other hand, there is no guarantee that you will or won’t have issues with gaming cards as well. Many gaming cards and driver combinations work fine across the board for Building Design Suite. There are others I have found which don’t play well with Revit. It may be because of a specific driver, a particular manufacturer, model, manufacture lot, or combination of factors. That said, in every case where I have seen issues I’ve been able to find older or newer drivers that fix the issue. It may cost some time in troubleshooting, but you can safely gamble that it will be nowhere near as much as a professional card would cost. And if a particular gaming card simply doesn’t work no matter what, sell it and get something else. At least you aren’t overspending on something that is comparatively underpowered.

In terms of performance, conventional wisdom states that just about any card you buy priced from $250 on up will be fine for Suite use. Spend less than that and you could suffer lag in dense point clouds, high polygon 3ds Max models, Revit views with lots of geometry and shadows turned on, etc. Spend more to get more usable eye candy on screen, hardware-accelerated anti-aliasing, or if you simply need to push a lot of polygons around on a regular basis.
**Professional vs Gaming Card Value**

The second issue with professional cards is that in many cases, the relative cost/performance ratio of professional level cards gets seriously out of whack, and you end up paying a lot more money for the same or even lower performance that you would have with an unwashed gaming card.

For example: A late model NVIDIA GeForce GTX 970 card costs around $350 and by all accounts will work extremely well for all Building Design Suite application work, including 3ds Max Design. Based on the newer second generation Maxwell 2 architecture, it features the GM204 GPU with 1664 CUDA cores, a 256-bit memory interface, a 1050 MHz clock, and has 4GB of onboard RAM. It has a dual-link DVI-I, HDMI 2.0, and 3 DisplayPort 1.2 ports, so it will plug into anything on your desk. By all accounts it represents a solid medium range value. Not the absolute best for everything, but very good nonetheless.

On the other hand, an entry-level workstation class NVIDIA Quadro K2200 is based on the 1st generation Maxwell architecture GPU GM107 with a paltry 640 CUDA cores, a 128-bit memory interface, a 1000 MHz clock, 4GB of RAM, 1 DVI, no HDMI, and only two DisplayPort 1.2 ports. It currently runs around $399 retail. Although core clock rates are very close, its fill and texture rates are a fraction that of the GTX 970. Why would anyone pay $50 more for absolutely no benefit?

**GPU Rendering: The iray Factor**

The choice of video card narrows and the picture focuses much more clearly if you want to render using a GPU-accelerated rendering engine such as iray in 3ds Max Design. iray, like mental ray, is developed by NVIDIA’s Advanced Research Computing (ARC) group. It is licensed by Autodesk for use in 3ds Max and Maya. iray is available as a standalone rendering engine that can plug into other applications as well. NVIDIA currently has the beta of BIMIQ, an iray rendering engine plugin for Revit, available at www.bimiq.com.

---


Traditionally, 3D visualization artists made heavy use of mental ray, which was acquired by NVIDIA when it bought mental images back in 2007. Autodesk licenses both mental ray and iray, takes the code from NVIDIA and builds application-centric user interfaces around it. This is why mental ray rendering in Revit is much simpler than that in 3ds Max, because much less of the renderer is exposed to end user tweakery.

As an older but proven rendering engine, mental ray can be extremely complex to work with and has lots of legacy code that keeps things slow. Mental ray solely uses the CPU for its rendering tasks. Mental ray is available in many applications in Building Design Suite 2016, but is being supplanted by the new Autodesk Raytracer (ART) engine (also a CPU-only renderer).

Conversely, iray was designed by NVIDIA from the ground up to do two major things. First, make it pushbutton simple to create stunning photorealistic imagery. Physically accurate lighting and materials are designed to work together to produce stunning results – so good that it can be used for all sorts of real-world analysis. Instead of fiddling with literally hundreds of little knobs and controls all throughout the interface to fake realism, the idea is to model up the scene, install the lighting, set up materials, and push the Render button. That’s it. Come back a little later and you have yourself a very, very good photorealistic image that arguably beats what you can get out of mental ray and is far easier to boot.

Second, iray was designed to take advantage of the new CUDA GPU hardware platform and programming model found in NVIDIA graphics cards (and only NVIDIA graphics cards) to perform the rendering on the card itself, not on the CPU. A single GPU is typically 4-12x faster than a quad-core CPU for these kinds of parallel tasks, and will complete these kinds of images in record time.

One of these is a photograph, the other an iray rendering. I still don’t know which one is which.
One very cool feature of iray is that you can create timed renderings. The nature of iray is that it renders by converging on the image as a whole instead of rendering separate buckets. Renders start out grainy and fine details sharpen over time. You either tell iray how many iterations to go through, or you set a hard time limit. You literally tell iray how long to work on the image instead of waiting for it to finish – a boon to overall productivity and meeting deadlines.

**NVIDIA’s Maxwell Architecture in Depth**

Just as Intel and AMD have microarchitectures for CPUs, so do graphics companies have microarchitectures for GPUs as well. In 2014 NVIDIA debuted the Maxwell architecture, the successor to the highly successful Kepler architecture that powered the GTX 7xx series as well as the GeForce Titan.

Maxwell introduced an all-new design for the Streaming Multiprocessor (SMM) that focuses on increasing power efficiency. Maxwell’s SMM replaces Kepler’s SMX with a newer, more efficient design. NVIDIA states that a 128 CUDA core SMM has 90% of the performance of a 192 CUDA core Kepler-era SMX.

Oddly, NVIDIA’s first Maxwell implementation was the GM107 GPU, which was on the lower end of the scale, suitable for more entry level graphics. Later in 2014 NVIDIA introduced “Maxwell 2” with the GM20x series. This second generation Maxwell architecture is designed to drive efficiency and raw performance with new features and internal optimizations.

Looking under the hood at a block diagram of the GM204 a popular GPU used in the GTX 9xx series, you can see that it is built as a series of components nested into each other. It is composed of 4 Graphics Processing Clusters, or GPC units, each housing 4 Streaming Multiprocessors (16 total) and a dedicated raster engine. Each SMM has 128 CUDA cores, a PolyMorph Engine, and eight texture units. With 16 SMMs, GM204 provides a total of 2048 CUDA cores and 128 texture units.

[Image: The GM204 Block Diagram and SMM]

---

A particular graphics card’s performance revolves around the layout and implementation of multiple SMMs in the GPU. A full GM204 implementation as shown above is reserved for the high end GTX 980 gaming card. Because not everyone needs to have the best performing GPU, lower end GTX 9xx cards still use the GM204 but will have executions units disabled in order to differentiate performance and price points for different market segments. By the way, this is not just something NVIDIA does. ATi makes their high end GPU then cuts it down to create lower end cards.

Problems with Maxwell 2’s GTX 970
An unintended side-effect of Maxwell 2’s improvements was found in that disabling individual units in the GPU – a typical move to take one GPU and make several different lower-end SKUs from it – comes at the cost of dividing the memory bus into high speed and low speed segments that cannot be accessed at the same time for reads and writes. This was famously the case with the GTX 970, which is the same GM204 GPU as in the GTX 980 but with some execution units disabled. The GTX 980 has a full 2046 CUDA cores, whereas the GTX 980 disables some internal execution units to provide only 1664 CUDA cores.

This upset many GTX 970 users who now have 3.5GB of video RAM running on a fast segment on a 224-bit bus, and another 0.5GB in a low-speed segment executing on a 32-bit bus. This means that the reported peak figures cannot be obtained in the real world, because it requires all memory able to be able to read and write at the same time.

Tesla
The other lineup in NVIDIA’s stable is the Tesla, NVIDIA’s lineup of GPUs specifically dedicated to compute tasks. Based on the Kepler microarchitecture, Tesla cards aren’t really graphics cards as they do not physically connect to a monitor. They are meant to be dedicated processing plants for GPU accelerated operations for engineering and scientific demands. Current models from fastest to slowest are the K80, K40, K20X, K20, and K10.

While NVIDIA sometimes promotes Tesla for rendering tasks, they do not provide nearly the cost / performance benefit of a GeForce or even a Quadro. They are clocked much slower – well under 1000 MHz - as they are often housed in 4U rackmount cases where temperatures need to be lower. They are most often compared against CPUs which are orders of magnitude slower. Their main benefit is the high amount of onboard RAM available (12GB on the K40) which is required for GPU rendering in iray. However, they are also insanely expensive for what they can do.

Gaming vs. Workstation Graphics Card Benchmarks, Cost Comparisons and Other Factors
Recall the $50 difference between the middle-of-the-road GTX 970 gaming card and the more expensive but less capable K2200. Now let’s look at the super high end, explicitly for CUDA / iray / V-Ray RT rendering: A top of the line Quadro M6000 retails for an eye popping $5,000. It runs the full implementation of the newer Maxwell GM200 CPU, with all 3,072 CUDA cores and all 24 SMM units operational, at a core clock of 988 MHz, and has 12GB on board. Being that CUDA performance scales linearly, we can look at the $/CUDA core as a metric. At $5K the M6000 has a metric of $1.63/CUDA core.

Then let’s look at a flagship “gaming” version of this, the GeForce GTX Titan X. It is also based on the full implementation of the GM200 GPU, with 24 SMMs and 3,072 CUDA cores, clocked at a slightly faster 1000 MHz and also with 12GB of video RAM. With a price of $1,000, the GTX Titan X 1/5th the cost of the M6000. Not cheap but has a CUDA metric of only $0.33 / CUDA core.
Comparing a series of benchmark scenes in 3ds Max 2016, the GTX Titan X beat the M6000 by a small amount, thanks to the slightly higher clock speed. Interestingly, both cards beat the previous Kepler generation by a mile.

In other words, you could conceivably install a GTX Titan X, get the same results as running an M6000, run it until it blows up, get another one, run it into the ground until that one explodes, and repeat the process 3 more times before you’ve spent the equivalent of one M6000 card. Or, even better, you could install 4 GTX Titans in a single system and get 4x the performance and save $1000 to boot.
NVIDIA GPU Comparison Chart
Below is a comparison chart which enumerates all of NVIDIA’s latest model gaming and professional graphics cards based on the Maxwell 2nd generation technology:

<table>
<thead>
<tr>
<th>GPU Code Name</th>
<th>GTX 970</th>
<th>GTX 980</th>
<th>GTX 980 Ti</th>
<th>GTX Titan X</th>
<th>Quadro M5000</th>
<th>Quadro K6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM Units</td>
<td>13</td>
<td>16</td>
<td>22</td>
<td>24</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>CUDA Cores</td>
<td>1664</td>
<td>2048</td>
<td>2816</td>
<td>3072</td>
<td>2048</td>
<td>3072</td>
</tr>
<tr>
<td>Core Clock (MHz)</td>
<td>1050</td>
<td>1127</td>
<td>1000</td>
<td>1000</td>
<td>1050</td>
<td>988</td>
</tr>
<tr>
<td>Pixel Rate (GPixels/s)</td>
<td>58.8</td>
<td>72.1</td>
<td>96</td>
<td>96</td>
<td>67.2</td>
<td>94.8</td>
</tr>
<tr>
<td>Texture Rate (GT/s)</td>
<td>109</td>
<td>144</td>
<td>176</td>
<td>192</td>
<td>134</td>
<td>190</td>
</tr>
<tr>
<td>Floating Point (GFlops)</td>
<td>3494</td>
<td>4616</td>
<td>5632</td>
<td>6144</td>
<td>4301</td>
<td>6070</td>
</tr>
<tr>
<td>Onboard RAM</td>
<td>4GB</td>
<td>4GB</td>
<td>6GB</td>
<td>12GB</td>
<td>8GB</td>
<td>12GB</td>
</tr>
<tr>
<td>Memory Bus Width</td>
<td>256-bit</td>
<td>256-bit</td>
<td>384-bit</td>
<td>384-bit</td>
<td>256-bit</td>
<td>384-bit</td>
</tr>
<tr>
<td>Memory Bandwidth (GB/s)</td>
<td>224</td>
<td>224</td>
<td>337</td>
<td>337</td>
<td>212</td>
<td>317</td>
</tr>
<tr>
<td>TDP</td>
<td>148W</td>
<td>165W</td>
<td>250W</td>
<td>250W</td>
<td>150W</td>
<td>250W</td>
</tr>
<tr>
<td>DX / OGL / SM</td>
<td>12.0/4.5/5.0</td>
<td>12.0/4.5/5.0</td>
<td>12.0/4.5/5.0</td>
<td>12.0/4.5/5.0</td>
<td>12.0/4.5/5.0</td>
<td>12.0/4.5/5.0</td>
</tr>
<tr>
<td>Power Connectors</td>
<td>2x 6-pin</td>
<td>2x 6-pin</td>
<td>1x 6-pin + 1x 8-pin</td>
<td>1x 6-pin + 1x 8-pin</td>
<td>2x 6-pin</td>
<td>1x8-pin</td>
</tr>
<tr>
<td>Outputs</td>
<td>1x DVI</td>
<td>1x DVI</td>
<td>1x DVI-I</td>
<td>1x DVI-I</td>
<td>1x DVI-I</td>
<td>1x DVI-I</td>
</tr>
<tr>
<td></td>
<td>1x HDMI</td>
<td>1x HDMI</td>
<td>1x HDMI</td>
<td>1x HDMI</td>
<td>0x HDMI</td>
<td>0x HDMI</td>
</tr>
<tr>
<td></td>
<td>3x DisplayPort</td>
<td>3x DisplayPort</td>
<td>3x DisplayPort</td>
<td>3x DisplayPort</td>
<td>4x DisplayPort</td>
<td>4x DisplayPort</td>
</tr>
<tr>
<td>Newegg Price (Nov. 2015)</td>
<td>$349.00</td>
<td>$510.00</td>
<td>$660.00</td>
<td>$999.00</td>
<td>$1,999.00</td>
<td>$4,999.00</td>
</tr>
</tbody>
</table>

Notes on NVIDIA Maxwell 2 GPUs

1. The GTX 970 and GTX 980 can be considered entry level cards for GPU rendering tasks. Both are rather underwhelming compared to much more capable models. I consider these excellent cards for Building Design Suite BIM users who may only dabble in iray rendering here and there.

2. Overall the GTX 980 Ti has the highest specifications for the price and is considered hands down the best value across the board. Its performance rivals the $5,000 M6000 and it is relatively cheap at $660. It is built on the same GM200 GPU as the Titan X and M6000 but only slightly cut down, with 22 SMMs activated for a total of 2816 active CUDA cores. It is also clocked at 1000 MHz, same as the GTX Titan X, but has only 6GB of video memory. With a price of $650, the GTX 980 Ti has a CUDA metric of only $0.23 / CUDA core. While benchmarks put it only slightly slower than the GTX Titan X and M6000, the price / performance ratio of the GTX 980 Ti cannot be beat, and it represents the best card available today for use in CUDA-based rendering engines.
3. The Titan X proves you don’t need a $6,000 card to be the fastest one around. Its only problem is the GTX 980 Ti is so close in performance and available for a fraction of the price.

4. The M5000 is in a terrible spot. Essentially the same card as the GTX 980 using the older GM204, slightly lower clocked but with twice the memory, it is simply outclassed from above with the M6000, and below with the GTX Titan X and GTX 980 Ti.

5. Although not included in the table, the GTX 960 use the GM206 GPU, which has half of the internal resources than GM204 (1024 CUDA cores, 2GB RAM, and a 128-bit memory interface width). In benchmarks it performs slower than the GTX 970 but respectfully well for its low $200 price point.

**Multiple GPU Considerations**

The thought of putting more than one GPU in your workstation may at first seem a little nuts, but the truth is that it is becoming the norm for people who perform GPU rendering on a regular basis.

You may have heard of existing technologies for running multiple GPUs together to aggregate performance, such as SLI (NVIDIA) and Crossfire (ATI). These are technologies meant for gamers who can tie two cards together and combine their pixel pushing power for maintaining high frame rates. **Neither SLI nor Crossfire works for iray or any other Autodesk product.** In fact, enabling SLI actually decreases iray performance quite a bit.

When you render on iray on a single GPU system, the graphics card will be running full tilt until it is finished. That means any movement by the mouse or trying to switch applications in Windows will temporarily stall the rendering process and system responsiveness will grind to a halt. Most people doing GPU rendering have learned early on to install at least one other card in the system just to handle the Windows OS operations. This doesn’t have to be a very high end card - in fact, any decent card will do. In 3ds Max Design’s iray renderer rollout, you can assign the specific card you want to dedicate to handling Windows and which one you want to handle the rendering task. You can also instruct it to limit the number of available CPU cores as well to maintain system responsiveness.

The other consideration is how many physical cards you can fit into a single workstation. This is limited by the number and type of PCIe slots you have available. All systems have a limited number of PCIe slots, some may run faster than others, and your computer’s case may be too small or not ventilated well enough to support more than two full length cards at one time. Or, you may have a laptop that obviously cannot be expanded. For these cases, you can invest in external PCIe enclosures which can house 1 to 4 or more PCIe cards. These additional cards aren’t connected to any monitors - they are simply graphics coprocessors which are churning on your renderings. You can interface this with your computer via an adapter card or a Thunderbolt connection.

The main reason to use multiple GPUs is that iray rendering performance scales almost directly with the total number of CUDA cores available. Put simply, for a single card the number of CUDA cores x the base clock frequency = relative performance. Additionally, iray can leverage and automatically use all of the CUDA cores across separate video cards. Add more CUDA-compatible GPUs to a system and watch your render times drop - fast.
GPU scalability is an important factor to analyze when specifying graphics cards. Is it more cost effective to use one very large and expensive GPU, or have several lower-end ones that all work together? Below is a chart showing how multiple GPUs in a single system scale up pretty linearly. 2 GPUs is roughly 2x faster; 4 GPUs is 4x faster, etc. Given the relatively affordable prices for GTX 970 and GTX 980 cards, it perhaps makes sense to scale performance by putting more of them in your case instead of a faster one.

<table>
<thead>
<tr>
<th>GPU</th>
<th>CUDA Cores</th>
<th>Expected Render Time(seconds)*</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) GTX 970</td>
<td>4992</td>
<td>189</td>
<td>$1,050.00</td>
</tr>
<tr>
<td>(2) GTX 980</td>
<td>4096</td>
<td>250</td>
<td>$1,200.00</td>
</tr>
<tr>
<td>(2) GTX 970</td>
<td>3328</td>
<td>283</td>
<td>$700.00</td>
</tr>
<tr>
<td>GTX 980</td>
<td>2048</td>
<td>500</td>
<td>$600.00</td>
</tr>
<tr>
<td>GTX 970</td>
<td>1664</td>
<td>566</td>
<td>$350.00</td>
</tr>
</tbody>
</table>


Here you can see that, when calculating value in terms of $ / CUDA performance, an affordable card like the GTX 970 really can’t be beat in a multi-GPU setup. Given single-card benchmarks you can extrapolate what multiple cards will provide and see where the economy of scale lies. The limitation of using the GTX 970 would be the fairly small 4GB of onboard system RAM which needs to contain the entire scene. Individual experience will provide insight as to how much of a limitation this actually is.

**Waiting for Pascal in 2016**

While NVIDIA, like Intel and AMD, always strive to increase Performance per Watt, shrinking the process from 28nm has proved to be very difficult. The next step was planned to go to 20nm in 2015, but neither NVIDIA nor AMD could make it happen. This is the primary reason why there were no real new developments in GPU design in 2015.

By all accounts, 2016 should be a banner year across the board for CPUs, GPUs, and storage. For their part, NVIDIA is coming out with their new Pascal architecture. This new architecture will be another order of magnitude more powerful than Maxwell, featuring 3D memory, unified memory, and NVLink:

- **3D Memory**: Stacks DRAM chips into dense modules with wide interfaces, and brings them inside the same package as the GPU. This lets GPUs get data from memory more quickly – boosting throughput and efficiency – allowing us to build more compact GPUs that put more power into smaller devices. The result: several times greater bandwidth, more than twice the memory capacity and quadrupled energy efficiency.
- **Unified Memory**: This will make building applications that take advantage of what both GPUs and CPUs can do quicker and easier by allowing the CPU to access the GPU’s memory, and the GPU to access the CPU’s memory, so developers don’t have to allocate resources between the two.
- **NVLink**: Today’s computers are constrained by the speed at which data can move between the CPU and GPU. NVLink puts a fatter pipe between the CPU and GPU, allowing data to flow at more than 80GB per second, compared to the 16GB per second available now. Pascal Module: NVIDIA has designed a module to house Pascal GPUs with NVLink. At one-third the size of the standard boards used today, they’ll put the power of GPUs into more compact form factors than ever before.

---

What about AMD?

AMD / ATI produces great graphics cards, and most middle to high-end Radeon HD cards will work perfectly fine in the Building Design Suite.

However, ATI does not implement the CUDA architecture so their GPGPU capabilities are limited to OpenCL, an open-source framework for writing applications that execute on GPUs as well as CPUs. The bad news is that not many folks develop GPU renderers for OpenCL. With V-Ray RT and iray being such popular GPU renderers, and the Optix plumbing available for others to easily create GPU accelerated renderers, there is little impetus for someone to develop the kind of widespread OpenCL libraries you need to gain traction.

However, OpenCL is supported by high end renderers like Indigo (http://www.indigorenderer.com) that can plug into 3ds Max, Maya, Revit, SketchUp, Blender, and others. OpenCL acceleration is a hallmark of the ATI Radeon HD series of cards – and they absolutely blow NVIDIA cards out of the water in OpenCL specific benchmarks; see http://bit.ly/1BgJBdi

Inside PCI Express 3.0

Peripheral Component Interconnect Express, or PCI Express for short, or PCIe for shorter, is a high-speed serial expansion bus standard that replaced the older PCI, PCI-X, and AGP standards. Unlike the old standard slots dedicated for graphics, a PCIe standard slot can host any kind of compatible expansion card, whether it is for graphics, sound, storage, networking, or whatever. In fact there is a large push today to move Solid State Drives from SATA to the PCIe bus directly, which will be discussed in the next section.

PCIe is based on a point-to-point serial bus topology that uses serial links (point to point communication channels) connecting every device to the host controller, which is either on the CPU itself or on the motherboard’s chipset. A single link between two devices is comprised of 1 to 32 lanes, which are the physical signaling traces that run to each PCIe slot. (Each lane is actually composed of 4 wires or signal traces on the motherboard). PCIe defines slots and connectors supporting multi-lane links from one to 32 lanes in powers of 2 (1, 2, 4, 8, 16, or 32) increments. Links are expressed with an ‘xN’ prefix, so x16 (‘By 16’) represents a 16-lane card or slot which is the largest commonly used size.

Given this, slots come in different sizes given in standard x1, x4, x8, x16 notation, which represents the largest physical card that will fit. The lane size is automatically negotiated during device initialization, so smaller cards can fit in larger form factor slots; e.g., you can put an x8-size card in an x16-size slot without issue and it just works.

However, size alone does not necessarily refer to the slot’s bandwidth; that is determined by the lanes going to the slot regardless of size. When there is a difference this is usually specified as “xsize @x speed,” for example, “x16 @ x8” means an x16 size physical slot that is configured to only run at x8. Pay attention to your motherboard manual to understand which slot can do what. On some motherboard, the PCIe slots are colored to indicate their capacity, but that is solely up to the manufacturer / model.

When considering a system that you know you are going to install multiple GPUs, review the motherboard specifications. You will often see the PCIe expansion slots listed as follows: →

<table>
<thead>
<tr>
<th>Expansion Slots</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCIe Express 3.0 x16</td>
</tr>
<tr>
<td>PCIe Express 2.0 x16</td>
</tr>
<tr>
<td>PCIe Express x1</td>
</tr>
</tbody>
</table>

This refers to the number of physical PCI Express slots provided, which ones are PCI Express 2.0 and PCI Express 3.0 compliant, how large the slots are (i.e., how wide are the lanes serving it), and how the slots will use the available lanes in different configurations. Much of this is actually determined by the chipset used, which in turn is reliant on the CPU so there is not a lot of usable flexibility here per CPU platform.
The history of PCIe introduced several standards, each of which improves throughput and internal optimizations for signaling and data integrity over the previous version. PCIe 1.0 had an effective bandwidth of 250 MB/s maximum data rate per lane, resulting in a 4 GB/s aggregate rate per x16 connector. PCIe 2.0 bumped this up to a 500 MB/s / lane rate (8 GB/s for an x16 connector) and is still commonly used today. PCIe 3.0 doubled this to 984.6 MB/s per lane (almost 16 GB/s per x16 connector) and is the latest standard. The PCIe 3.0 standard was first fully supported by the Ivy Bridge CPU introduced in 2012 and is used by all graphics cards today. PCIe 4.0 is expected to double this again to 1969 MB/s per lane or 31.5 GB/s per x16 slot.

As noted in the section of processors, CPUs have their own embedded PCI controller which provides anywhere from 16 to 40 PCIe 3.0 lanes depending on model. These lanes service the x16 expansion slots used for graphics cards, allowing for up to dual GPU configurations for 16-lane GPUs, and support up to 5 graphics cards on a 40-lane GPU.

The latest Z170 PCH, with its DMI 3.0 connection to the CPU, provides 20 PCIe 3.0 lanes for greater peripheral connectivity such as networking, audio, and storage. In fact, storage is one reason why Intel upgraded DMI 2.0 to DMI 3.0, effectively doubling the bandwidth between the CPU and PCH. As discussed in the next section, SATA as an interface to high end SSDs is coming to an end, and PCIe based SSDs are going to roam the Earth next year. The expected massive increase in SSD bandwidth requires a faster connection to the CPU.

While the total number of lanes is a physical constant, the configuration of PCIe slots is always a compromise between throughput and number of devices at that speed. The motherboard shown below is an ASUS Z170 Deluxe LGA 1151, typical of current Skylake system. The allotment of PCI lanes across the 7 PCIe slots is a variable depending on what is plugged in where.
7 PCIe slots divide up the allotment of PCIe lanes:

- There are two PCI Express 3.0 x16 slots that will run a single GPU at x16 or dual GPUs at x8/x8. These PCIe slots use the 16 lanes from the CPU.
- 1 PCI Express 3.0 x16 slot which runs at maximum x4 mode, compatible with PCIe x1, x2 and x4 devices. However, this PCIe x16 slot shares bandwidth with SATA connectors. The PCIe x16 is by default set at x2 mode.
- 4 PCI Express 3.0 x1 slots, typically for audio cards, modems, or other low-speed adapters.

You can see above that you need to pay particular attention to the configuration of the PCIe slots on the Skylake platform and make sure you plug in the right things to the right slots, or you end up with something that needs more bandwidth than the slot can provide. For example, NVIDIA cards require at least a PCIe 3.0 x8 slot to even work. Put it into a x16 slot running x4 and nothing happens. Furthermore, often if you plug something into one slot, another slot will go from x16 to x8 speeds because the lanes are all shared. The motherboard manual will go into excruciating detail on what you can plug into what slot.

**PCIe 2.0 vs PCIe 3.0**

It is interesting to note that the throughput provided by PCIe 2.0 was underutilized by even the most powerful graphics cards, so the doubling of bandwidth in PCIe 3.0 by itself provided little to no improvement performance when comparing the same card on both slots. Rather, PCIe 3.0 provides plenty of headroom for multiple GPUs to be installed in the same system and each use more available lanes. Because PCIe 3.0’s throughput is roughly double that of PCIe 2.0, the throughput for a PCIe 3.0 card at x8 will equal that of a PCIe 2.0 card running at x16, which is fine. Even running at x8 mode, a powerful graphics card is not bandwidth limited by the PCIe bus, so there is no problem in running two PCIe 3.0 x16 cards at x8. In fact, running a PCIe 3.0 x16 card at x4 speed (at 25% of the available bandwidth) reduces overall performance by only about 14%.

**Maximizing GPU Rendering on Advanced Platforms**

Because of the 16 PCIe lane limitations of Skylake and Xeon E3 CPUs, high-end Viz Wizards who need more than two GPUs necessarily need to move to a more advanced CPU / motherboard platform. With 40 available PCIe 3.0 lanes provided by the Haswell-E and Xeon E5 platforms, users of GPU renderers have the ability to use 3-way, 4-way, or even 5-way PCIe 3.0 cards, lowering render times and increasing viewport interactivity by leaps and bounds.

For example, the ASUS x99-Deluxe motherboard for the Haswell-E platform (which can support CPUs with either 28 or 40 PCIe 3.0 lanes) has five x16 slots and one x4 sized slot. The expansion capabilities are:

**PCI Express 3.0 x16**
- 40-Lane CPU - 5 x PCIe 3.0/2.0 x16 (x16, x16/x16, x16/x16/x8, x8/x8/x8 mode)
- 28-Lane CPU - 3 x PCIe 3.0/2.0 x16 (x16, x16/x8, x8/x8/x8)
  - Note: The 5th PCIe x16 slot shares bandwidth with m.2 x4. Triple PCIe 3.0/2.0 configuration is default set at x8/x8/x8. Adjust PCIEX16_5 Slot Bandwidth in BIOS.

**PCI Express 2.0 x16**
- 28-Lane CPU - 2 x PCIe 2.0 x16 (x1 mode)

**PCI Express x4**
- 40-Lane CPU - 1 x PCIe 2.0 x4 (max at x4 mode)
- 28-Lane CPU - 1 x PCIe 2.0 x4 (max at x4 mode)
  - Note: The PCIe4_1, USB3_E12 and SATAEXPRESS_E1 connectors share the same bandwidth. The SATAEXPRESS_E1 will be disabled when there is a device installed on PCIEX4_1 slot. Set this option to X2 Mode or X4 Mode when the installed PCIe device is higher than X4 interface.
VII. Storage

Solid State Drives
Today, Solid State Drives (SSDs) are now considered mainstream components and are a necessity for any Building Design Suite user, regardless of where they fall in the user profile range discussed in Section 1. Having no moving parts, all of the data is held in solid state non-volatile memory. There’s no noise or heat, and they draw much less power than a mechanical hard disk.

The performance improvement of an SSD is truly dramatic over typical hard drives. An SSD is arguably the single best upgrade you can make to any system of any speed. Random access times are about 0.1ms, compared to 5-10ms for hard drives. Read times are not affected by file fragmentation or where the data is stored, compared to HDDs where data written to the inner cylinders are read more slowly than the outer cylinders. Driver support under Windows - especially Windows 8 and 10 - is excellent.

In particular, large applications such as in the Building Design Suite take on a new life under SSDs. Every application in the Building Design Suite takes a while to load on mechanical drives, but are very quick to open from an SSD. The same goes for loading up large Revit models; the speed at which an SSD can deliver a 300MB+ file into system memory is pretty amazing.

The first issue with SSDs is obviously their cost. While their prices have dropped considerably over the years (they are well under $0.50/GB for even the best models) they are nowhere near as cheap as mechanical drives, where a typical 1TB HD is running around $90, about 9 cents/GB. But that’s OK because the higher price is completely justified given the amazing performance increase.

Think about this: Running two 512GB SSDs in a RAID 0 array (to get 1TB) would run about $450 total. A single 1TB mechanical drive is about $90, so the premium you are paying is $360 for SSDs. Taken over a relatively short 3-year lifespan, that premium is only $10/month, or about $0.32 per day for amazingly fast performance. You could easily cost your company more than $0.32 per day just waiting for Revit to open your project models, so the ROI is instantaneous.

The Death of SATA
All of your disks, SSD or not, plug into SATA ports on the motherboard. This Serial ATA III interface uses the Advanced Host Controller Interface (AHCI) as the “software” interface which allows SATA to use advanced features such as hot plug and native command queueing (NCQ) to optimize transfers.

Today’s problem with SSDs is that they are now too fast for the traditional storage subsystem in most PCs. SATA / AHCI were originally intended to host mechanical hard disks, and are unprepared to take on the throughput of modern SSDs. The SATA III 6Gb/s interface limits us to 500+MB/s reads and sub-500MB/s writes. New SSDs have sequential read/write speeds upwards of 2.8 GB/s and 1.9 GB/s respectively, so the SATA interface itself is limiting how fast our SSDs can perform.

Over the past few years there has been a push to get SATA out of the way by utilizing the PCI Express bus. Instead of trying to double the bandwidth of SATA – which would take forever to catch up with the progress of SSD technology - it was realized that the PCIe bus provides the bandwidth required, which can easily be added to by adding additional lanes.

First there were attempts to fuse the two together in a backward-compatible format, called SATA Express. This was adopted on quite a few late model motherboards last year but has a rocky future, as it’s not that much faster than SATA and today’s SSD performance even outstrips that. The industry has collectively made up its mind that, if the move is clearly to go to PCIe based storage interfaces, let’s figure out how it should look and function on its own instead of trying to shoehorn PCIe and SATA together.
Exit AHCI, Enter NVMe

The main problem with SSDs on today’s hardware is AHCI, the software protocol designed to specify the operation of SATA host bus adapters to work with mechanical hard drives in mind. It is more optimized for high latency rotating media rather than low-latency non-volatile storage in SSDs. As a result AHCI cannot take full advantage of SSDs, and with PCIe the potential bandwidth increases dramatically. It’s like driving a race car on a dirt road. We need a new software interface optimized for both SSDs and PCIe.

Non-Volatile Memory Express, or NVMe, has been designed from the ground up specifically to capitalize on the low latency and parallelism of PCI Express SSDs, and complementing the parallelism of contemporary CPUs, platforms and applications. At a high level, primary advantages of NVMe over AHCI relate to NVMe’s ability to exploit parallelism in host hardware and software, based on its design advantages that include data transfers with fewer stages, greater depth of command queues, and more efficient interrupt processing.²⁵

<table>
<thead>
<tr>
<th></th>
<th>NVMe</th>
<th>AHCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>2.8 µs</td>
<td>6.0 µs</td>
</tr>
<tr>
<td>Maximum Queue Depth</td>
<td>Up to 64K queues with 64K commands each</td>
<td>Up to 1 queue with 32 commands each</td>
</tr>
<tr>
<td>Multicore Support</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>4KB Efficiency</td>
<td>One 64B fetch</td>
<td>Two serialized host DRAM fetches required</td>
</tr>
</tbody>
</table>

Source: Intel

Say Hello to your Little Friend

The move to a faster interface is not the only thing that is changing on the SSD format. Meet your new hard drive:

This Samsung 950 Pro is a 512GB SSD using the new m.2 interface, and at 80mm long it’s a little over half the length of a DDR4 DIMM module. This new PCI Express Mini Card form factor plugs directly onto your motherboard and right into the PCIe 3.0 bus (up to four lanes). PCIe adapter cards are available for cheap if you have a legacy machine with a free x4 PCIe slot.

²⁵ https://en.wikipedia.org/wiki/SATA_Express#Features
The new m.2 interface is a replacement for mSATA, but it is far more versatile in that it combines PCI, SATA, and USB standards into one connector. m.2 supports a wider variety of hardware, including wi-fi cards, sat-nav cards, audio cards, and of course SSDs. m.2 supports NVMe as the logical device interface, while also maintaining backward compatibility with AHCI (at the cost of not delivering optimal performance). The cards that use this interface can range from full length 80mm boards used for SSDs down to 42mm cards usually designed for wireless chips. The full m.2 experience (connecting to 4 PCIe lanes) is available on almost all new Skylake Z170 motherboards, as well as the Haswell-E X99 platform. Note that last year’s Z97 chipset for Haswell / LGA 1150 socket reserved only two PCIe lanes, limiting bandwidth to 1GB/s.

However, it’s not as rosy as it seems. Early problems with m.2 are typical of other similar new technologies. First, there is currently a price premium on these drives, making them much more expensive per GB than typical 2.5” drives of the same size. Second, m.2 drives are currently limited to 256 and 512GB sizes, although 1TB drives are promised sometime next year.

Third, while benchmark tests of raw data throughput show the shiny new m.2/NVMe drives trouncing their backwater SATA/AHCI cousin, real-world benchmarks show their improvements are more modest, perhaps only around 10% faster. Still, that is a lot considering how flat the difference is between the current crop of SATA SSDs, who all perform within fractions of a percent of each other.

Another issue at the moment is that some benchmark testing indicates that performance of m.2 drives is getting throttled due to the buildup of heat. In the above synthetic benchmarks you see all of the PCIe NVMe drives well ahead of typical 2.5” SSD running on the SATA interface. The top performing Intel drives are PCIe NVMe drives as well, but are enterprise-class drives on a traditional x4 PCIe card. The problem with these enterprise cards is that, while they are the fastest SSDs of them all, they are very expensive ($700 - $1,000) and require a PCI-e 3.0 x4 slot. A consumer-oriented PCs may not the PCIe x4 bandwidth available with the video card installed, and multi-GPU workstations may have issues fitting one in as well.

While performance of m.2/NVMe drives is a good deal greater than SATA SSDs, taken together with the added cost and possible heat-related performance issues, opting for an m.2 NVMe drive for a new build may not make that much sense right now. Later in 2016 this will be a different story. Check with your system vendor to see what the premium upcharge is going to be for essentially the same storage size – it could be considerably higher to be an early adopter. However, it is inevitable that performance will go up and prices will go down as their market share grows.
Sizing SSD Sweet Spots
In terms of capacity, the go-to sweet spot was typically 240-256GB range simply because it was affordable. This year we see the 480-512GB segment drop to essentially 2013-era 256GB prices and claim the new Sweet Spot crown.

In terms of Building Design Suite requirements, while 256GB would hold the OS, Microsoft Office, the Adobe CS Suite, and the Building Design Suite, it held little else. Users would constantly have to purge out all Revit local files and folders, and watch their %TEMP% folder like a hawk and clean it out at least once a week. Lumion in particular can also pose a problem because it saves everything locally, amounting to gigabytes of data taking up space for every project.

As such, it is typical that systems with 256 or 512GB SSDs also have a 1TB mechanical drive for near-line storage. This makes it convenient to store user data that is not performance critical, such as their iTunes library, pictures, personal data, and so on.

Reliability was of great concern when SSDs first came out, because we simply didn’t know how much you could read and write off of them before they started to die. The NAND technology behind the non-volatile RAM will inevitably wear down over time. However, TechReport.com did an SSD endurance experiment and found that consumer-grade SSDs can take a serious data pounding, with some passing the 1.5 Petabyte write mark, which is far more data than users will ever need to write during the useful lives of the drives. Or their own lives, for that matter: 1.5PB equates to 50GB writes per day for over 80 years.

RAID Arrays for SSDs
One of the ways people try to improve performance on mechanical disks is with a RAID (Redundant Array of Inexpensive Disks) system, comprising of two or more hard disks with the data striped across them. This can dramatically improve read and writing performance because the system multi-threads I/O to happen asynchronously (independently at the same time) across multiple devices.

Today, RAID’s main use case is for data reliability and redundancy in servers. Depending on the RAID level configuration of disks and redundancy, it offers the ability to stripe data and parity information across multiple drives. If a single drive goes down, the other two can operate with no data loss. In fact, the 3rd drive can be rebuilt in the background, a new drive hot swapped out, and the RAID array rebuilt without inconveniencing any users or turning the server off, minimizing downtimes.

RAID 0 is often used on enthusiast systems because it combines two or more drives into one logical volume, and stripes the data across both, increasing storage performance dramatically. However, for SSDs the performance is already there, and benchmarks show little to no real-world performance benefit. The appeal of SSDs in RAID 0 is more to aggregate the space of smaller, cheaper drives into one larger volume.

With most data being held on the server instead of the workstation, business machines typically do not implement RAID due to increased costs and, in the case of RAID 0, increased liability in case of a single drive failure, which knocks out the entire array.

Windows 7 and 8 can implement RAID in software, allowing you to create logical volumes of combinations of available drive space (using any drive size), making it unnecessary to purchase special hardware.
 VIII. Mobile Computing

When purchasing a laptop or mobile workstation, you have different sets of parameters to think about before you make a decision. Those decisions have to be made on how you plan to use it on a daily basis more than anything else. Mobile computing becomes a compelling lifestyle addition and typically trumps raw performance. It’s very easy to become seduced by the freedom of sitting in a coffee shop working on a Revit model rather than sitting in some fabric covered box all day.

Portability vs. Power

As mentioned in Section IV on Processors, the first consideration in shopping for a laptop is to understand how you will use it. Is it to be your one and only daily driver, or are you working between a laptop and a desktop on a regular basis? The amount of time you need to carry around your laptop will definitely carry more weight - excuse the pun - in your decisions about form factors, power requirements, and other factors we will discuss.

This question is largely between portability and power, because you really can’t get too much of both in one unit. Ultraportable 13” machines are small and light and lovely to travel with, but not much use for running Revit unless you can remote desktop into big iron back at the office. On the other hand, solid machines that you would want to Revit on all day start at 15” screen sizes up to 17”, and heft is definitely a factor. Laptops with 17” screens are quite heavy to lug around, and forget about using one on an airplane.

This is changing, particularly in the system thickness. Today’s latest processors require less power and put out less heat, so new laptop models are getting much thinner now. Clearly the legacy of the MacBook Air lives throughout a lot of new impossibly thin devices.

For moderate Revit usage with enough power to get through the entire day every day, chances are you can get by with a 15” model, particularly if you drive it while in a docking station and have large displays. No matter what, desktops with their higher-clocked CPUs and more RAM are much more powerful and are what you want for Revit heavy lifting and renderings. They just aren’t as much fun to take down to Starbucks.

Screen Choices

A major factor in deciding the base form factor is the screen size and resolution. Choose wisely, because this largely determines almost everything else about the system as it drives the overall usability of the system. Within each model lineup, you have only a few options for screen resolution, anti-glare options, and so on.

Because UI elements are defined in pixels, it’s a balance to find a resolution that provides a decent size for icons and other UI elements without being too tiny. For example, the default in a 15.6” screen may be 1366x768, which is just too coarse for design applications. The next step up may be 1920x1080 which many would consider perfect for that size form factor.

You may be tempted to go with higher resolutions such as QHD 3200x1800, but beware of scaling issues. 3200x1800 offer a lot of screen real estate if left at its native resolution, but even those with 20/20 vision will have problems adjusting to tiny UI elements. Windows 8.1 works a lot better than Windows 7 at providing a good scaling factor of 150% to handle such high resolutions, which helps make font, icons, symbols and windows easier to read. However, you may hit on a few stubborn programs that yield less desirable results. Especially games, which often do not work as intended, or do not work at all with the resolutions, especially on 15.6” or smaller screens.
All in all this higher high resolution may cause more problems than benefits, especially when you consider that there was nothing really broken about a 1920x1080 resolution on a 15.6” screen. I would not recommend a QHD panel for anything smaller than a 17” screen, and I have a hard time recommending a 17” form factor in the first place.

The other consideration is screen technology. Most mobile laptop manufacturers offer several different screens; you will see either IPS (In Plane Switching) screens or Twisted Nematic (TN) panels offered. Go for the IPS displays – they are brighter, have more vibrant colors, and have wider viewing angles. They do cost more, but you are going to live with the display for the life of the laptop. It’s best not scrimp in the really important stuff like what you stare every day.

With Windows 8 / 10 being so popular the choice of touch screens is now here. Of course, if you are like me, anyone touching my screen with their greasy fingers is subject to the death penalty, so this may not be a worthwhile option. Antiglare versus glossy screens are entirely a personal preference. Here, a field trip to the local big box store to look at various differences in screens may be in order.

**System Memory**

With any mobile workstation you are limited on the amount of RAM you can install, typically to only 2 slots in mainstream models and 4 slots in “mobile workstation” class machines (sometimes with 2 slots under the keyboard and out of easy reach). As a desktop replacement for a Building Design Suite user, I typically recommend 32GB of RAM as either 2x16GB or 4x8GB modules right out of the gate. That covers all bases and means you won’t have to open the laptop up again or pull out memory you cannot use if you ever want to upgrade.

The same warnings for purchasing RAM upgrades from system vendors applies to laptops as well as desktops. Dell and HP will have no issue charging you hundreds of dollars for a meager 8GB RAM upgrade. If your vendor won’t match the price and you know how to work a screwdriver, buy the laptop with the bare minimum you can get. Pick up your RAM elsewhere and install it yourself. It literally takes minutes on most laptops. Just don’t lose the tiny screws.

**Graphics Choices**

Your choice of graphics card is going to be severely limited within any particular line, so it may end up driving your form factor and model decision. What goes for desktop cards goes for laptops; stay away from the integrated Intel HD graphics unless you rarely touch Building Design Suite applications.

Workstation class laptops are going to come with workstation class graphics, and there’s not many options to substitute it out after the fact. In any particular model, I typically recommend upgrading to the fastest option available within reason, because you never know what you may work on next year that really takes advantage of it. Do a quick ROI: a $300 graphics board upcharge works out to only about $8 a month over the life of the laptop.

However, be cognizant of the differences between GPUs. On the NVIDIA front the Quadro GPUs available for a 15.6” platform are the M1000M and M2000M – pretty good mobile GPUs. On the 17.3” platform you have the higher-order GPUs in the M3000M, M4000M, and K5000M. A Quadro M1000M may only be $120 less than a Quadro M2000M, but the M2000M has 67% more CUDA cores (640 vs 384) which could get you through a rendering hump. Of course, it also consumes 15W more power. For everything except GPU rendering under 3ds Max, both GPUs will perform pretty much exactly the same.

For NVIDIA cards in workstations, I typically look for Optimus technology in the BIOS and immediately turn it off. Note that this option will usually not be found in the BIOSes of non-workstation class laptops. Optimus is supposed to use both the IGP on the CPU for mundane graphics tasks as well as a discrete GPU.
for the heavy lifting. This helps battery life as the IGP consumes a small fraction of the power an M1000M GPU draws. Optimus automatically detects when an application requires the GPU and switches it on for that app. You can also direct the video driver to use the GPU or integrated graphics in the NVIDIA control panel or via a shortcut context menu.

While Optimus has gotten better over the years, it is still not perfect. Apps like Steam will require the GPU and draw lots of power even when minimized. It comes down to how you use the laptop. If you keep it plugged in most of the time, power draw isn’t an issue. Disable Optimus and enjoy high end graphics all of the time. If you travel a lot and work off the battery and/or only need the GPU sometimes, then keep it on and see if it works for you.

Storage
In a typical mobile workstation, storage is going to be limited to a few choices. Some systems may offer storage as a single 2.5” drive or in an m.2 slot, or in combination with a mechanical drive that is housed in the optical drive bay. You may also have the option for a 1.8” “Mini-Card” that can typically be a SSD, so your choices may allow for one SSD only, two SSDs, or an SSD and a 2.5” mechanical disk. As with desktop systems, ensure you get an SSD as your boot drive in all cases.

Regardless of use, you will probably hit the 256GB C:\ ceiling pretty early on, so I recommend you opt for a 512GB SSD instead of the smaller SSD and an additional drive. Remember that a second mechanical drive will consume power and heat and lower battery life.

If you think you will store more than 500GB worth of OS, apps, and data on it, go for the secondary mechanical disk. You can probably opt for a 7200 RPM drive for fast access, but you can save a lot of power and heat by going for a slower 5200RPM drive. If all it does is store near-line data, the slower speed will be fine. It goes without saying to avoid 5200 RPM drives for your OS and applications at all cost.

Larger 17” models usually provide more options with regards to configuring multiple drives, and most come with two full 2.5” drive bays.

Docking Stations – Yes, You Need One
If you work on a laptop at a desk for any amount of time, don’t get a laptop without a docking station, some dedicated monitors, and a keyboard. They make living with your laptop between work and home easy, and you won’t need to lug the power brick around.

Most docking stations can drive two monitors, and may come with a combination of DVI, HDMI, DisplayPorts, audio, and USB ports. Double check the video ports against your intended monitors before purchasing. Depending on the model of your laptop, connecting to the docking station is either easy or not. If you are buying the laptop from Dell, you want to get Dell’s docking stations. Aftermarket units can be a crapshoot.
The best configuration is to have (at least) two large 24”+ screens connected to a docking station. By having multiple large screens you can keep the laptop closed in the dock, perhaps tucked away under the desk. If you have only one screen, you can still use the laptop screen for dual-screen use but the two screens will be different sizes, making it odd to work with.

Part of the pain of using a docking station is in how much room they take up. You have to physically affix the laptop to the dock and it needs to stay there, so if you are working on a small or cluttered desk the laptop + dock can take up a lot of room. I like to have mine on a small side table or on a rolling pedestal that slides under the desk.

New wireless docks stations are available which cut the cable clutter even more. A wireless dock looks more like a small brick with the cables out to your monitor and keyboard, but all you have to do is put the laptop somewhere near the dock to have it connect. They use the newer WiGig standard, so your laptop specification will need a WiGig card included to use it. However, they do have performance and capability restrictions which may be a deal killer, so review the specific unit carefully.

**Stuff You Want**

Laptops are incredibly personal devices. When evaluating a laptop model it pays to pay attention to the things that could make life easy or difficult. In particular you want as many USB 3.0 ports as possible, because you will use external drives and memory sticks more.

Look for a USB-C port if possible. USB-C is a new connector standard that is poised to take over all other form factors for connectivity and power. Like the micro-USB connector on your phone, it’s small but elegantly designed, in the way that USB A and B are not. It is reversible and can be used in any orientation.

USB-C fully supports the new USB 3.1 specification. USB 3.1 is a revised version of the USB spec that is twice as fast as USB 3.0, and delivers much more power. Overall USB-C as a connector does just support USB 3.1; it also supports DisplayPort, HDMI, and VGA. Meaning the same connector form factor could be used for a wide variety of devices from video to power to external storage.

You also want a full size HDMI out, and mini DisplayPort if you can as well. Forget VGA. RJ-45 rounds out the standard ports you need.

As important as the ports are, you need them in the right place. If you aren’t using a docking station you are probably going to plug in a mouse, power cord, and network cable to the unit itself. Are the ports out of the way of your hands when you are typing? I had a laptop where all of the ports were in exactly the wrong spot: nothing was along the back, everything was along the sides closest to the front of the laptop, right where my hands naturally rest. Stupendously bad design.

The keyboard can make or break a laptop. This is subjective so I highly encourage you to sample the machine in a store somewhere, and extensively type on it. One of the laptop keyboard features I used to have was a numeric keypad. This shifts the keyboard to the left a little and it took some time to get used to it, but it got to be a real necessity. Also: Look for a dedicated Calculator button – it will gets a lot of use. It’s a given that you will have media controls for volume and mute.

**Stuff You Want, You Just Don’t Know It Yet**

Speaking of keyboards, a backlit keyboard is a fantastic feature. It’s also a cheap upgrade. It sounds dumb, but pay for it - you will not believe how handy it is when working at night.

If you plan on working a lot remotely without power, e.g. on airplanes, you may want to opt for a larger, 9-cell battery. They are heavier to lug around all of the time, so I suggest a smaller 6-cell one instead which, with today’s power-sipping processors, should last enough to get some things done. Best bet: get a 6-cell for everyday use and a larger 9-cell battery for when you are on the road.
Get a wireless mouse; preferably one with a very little USB transmitter that can stay inserted in the USB port without breaking off. I personally like the Logitech MX Master in large part because of the tiny transmitter. I’ve wrecked more than a few USB ports with longer dongles getting bumped around.

Get a good carrying case with pockets for stuff. I use one, but am not a fan of backpack cases – the straps never last very long. A soft, professional looking briefcase style is nice, and you won’t look like a college student at client meetings. But if you have a huge power cord brick, make sure it fits without bulking out.

Get the most advanced wireless options available that support the widest array of standards, e.g. dual band, 802.11ac / a / b / g / n. It is only pennies more. Bluetooth 4.0 is a must; if you have a Smartphone or Bluetooth in your car it’s a real benefit to transfer contact information or other near-field communications.

If you are on the go without the ability to connect to Wi-Fi, you may want to invest in a 4G cellular modem. Optionally you may be able to tether your phone, turning it into a Wi-Fi hotspot, or invest in a personal Wi-Fi device from your cellular network provider.

Stuff You Don’t Want
Laptops in general are notorious for looking good when you first unpack them, but after using them for a day you see which design decisions were winners and which were crap.

We used to have a fleet of HP laptops for our training lab. From almost every aspect, they were annoying to use and slow. The screen wasn’t a high enough resolution so the icons in Revit were squished. The keys were in the wrong places and just felt lousy. The network and power ports were all wrong and in the way.

Then we purchased some Dell XPS laptops for our training lab. In general, they were good machines – they were fast enough and the screen was quite nice. But the keyboard has smaller keys which are easy to miss, making typing a real chore. There were 4 special flat buttons in a plastic touch strip along the top right side that control things like Wi-Fi and audio volume. The problem is that every time you go to hit the delete key – which you do a million times in Revit - one of these buttons got pressed by accident and a dialog box pops up. That’s quite annoying. But we couldn’t disable the keys because there isn’t any other way to control those functions otherwise.

Even on newer systems dumb things crop up, like misplaced key layout. The takeaway here is that you should spend some time with the desired unit to get a feel for what design features are going to detract from your day to day workability.

IX. Peripherals
The peripherals you outfit in your system may not be critical for overall performance, but they do affect day to day usage in important ways and thus should be a focus in specification.

Monitors
Monitors are, like hard drives, something that many people don’t think twice about much when purchasing a new system, and that’s rather unfortunate. They go for cheap and that is it. The choice of screen has a dramatic effect on your comfort and convenience level.

Today, large LCDs are incredible bargains compared to their ancestors. It used to be that a good 21” CRT monitor with a resolution of 1280x1024 was about $1,900. Today’s common 24” LCDs have almost double the resolution, use a third as much power, generate almost no heat, are uniformly bright with no distortion, and really good ones cost about a quarter of that to boot.
If one large LCD is great, two are better. The advantages of dual screen systems are hard to ignore. With the cost per monitor so low compared to the improvement in efficiency, it doesn’t make any sense not to have two large screens at your disposal. Within a single application like Revit you can move the Project Browser and Properties Palette off to the other monitor, freeing up space.

For mundane tasks, the biggest boon multiple monitors have to productivity is in managing multiple applications, such as browser windows, email, office applications, Photoshop, etc. No one stays in only one program for very long; juggling multiple windows on one screen is maddening.

With 3ds Max Design, the use of three monitors (which may require >1 graphics card) is entirely warranted. That way you can have the application window on one screen, the scene explorer and material editor on another, and a curve editor on the right, all at the same time – that’s efficiency.

With multiple monitors comes the need to properly manage the windows. The ability to spread an application across both screens or to bounce a maximized window around is highly productive. To that end I recommend investing in a multi-monitor utility such as UltraMon or DisplayFusion. Each offers the ability to put additional buttons in the top right corner to manage windows across screens. If you have an NVIDIA card, the driver includes an “Nview” utility which can offer up some similar functionality as well.

Monitor Technology
There are differentiators between monitor models which do make a difference. First and foremost is technology used for the panel itself. There are three main technologies: TN (Twisted Nematic), VA (Vertical Alignment), and IPS (In-Plane Switching). Of the three, look for and demand an IPS panel. At every metric, IPS panels deliver sharper, almost reference like quality.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Color Reproduction</th>
<th>Viewing Angle</th>
<th>Response Time</th>
<th>Price</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Plane Switching (IPS)</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Good</td>
<td>Expensive</td>
<td>Slight color tinges may be visible at an angle</td>
</tr>
<tr>
<td>Vertical Alignment (VA)</td>
<td>Good</td>
<td>Good</td>
<td>Average</td>
<td>Reasonable</td>
<td>Colors shift when viewed at an angle</td>
</tr>
<tr>
<td>Twisted Nematic (TN)</td>
<td>Average</td>
<td>Average</td>
<td>Excellent</td>
<td>Inexpensive</td>
<td>Limited to 6-bit color; Technology improving.</td>
</tr>
</tbody>
</table>

There are several iterations of IPS panels, such as S-IPS, H-IPS, e-IPS, and P-IPS. They are all relatively similar to each other and not really worth getting into detail. Until the holy grail of flexible OLED monitors are commercially viable, an IPS screen is what you want.

The second concern is size and screen resolution. Probably the most practical size in a dual screen configuration is a 24” LCD, although 27” screens are becoming more popular as prices drop. 30” LCDs are still considered a luxury as their prices are over $1,000. However, if you can swing one (or two), they are definitely worth it. Monitors will last several machine generations, so it pays to remember it will be around for a while so buy the largest one you can.

While TN is inexpensive it did have a reputation for being cheap but with poor to average color fidelity and overall quality. That’s changing a little as TN is improving around the edges; look at specific reviews before jumping on a TN panel.

4K Displays
The resolution (number of pixels horizontally and vertically) determines how much of your application and user interface is available to you at one time. Currently the Holy Grail of hype today are so-called “4K” displays, which is often and incorrectly used to refer to the more common standard UHD, which is 3,840 x 2,160. That’s 4x as many pixels as a typical 1080p 1920x1080 monitor.
They can and do look fantastic, but can have drawbacks. First, you have lots of pixels, so your 4K will need to be pretty large – at least 24” or larger. The main problem is that graphics cards need to be good enough to spit out that kind of resolution. This is not as much of a concern as it once was, but plugging a 4K display on a 4-year old machine won’t be without issue. Windows itself is a problem too. At that resolution, icons and screen elements get so tiny that you need to scale them up, and that never looks good.

The bottom line on 4K resolution is that while display technology is improving to incorporate 4K as a new standard, it has some shakedown cruises to complete. You need to check all of the specifications carefully and ensure it matches your graphics card’s capabilities. Remember too that lower resolution displays are just fine for anything you need to do in the Building Design Suite and you won’t have any hassles in trying to marry a particular high resolution monitor with a video card.

**Resolutions You Want**

Because UI elements are metered out in pixels, the more pixels you have on screen the more stuff you see at one time, and the more you can get done without fiddling with pans or scrolling windows. That time all adds up. With the Ribbon being so prevalent in Revit, smaller resolutions chop the icons down to smaller bits making the interface slightly harder to use; you may also end up collapsing it down because of the lacking vertical dimension.

For a 24” monitor, look for a resolution of no less than 1920x1200. Most cheap 24” screens max out at 1920x1080, which is acceptable to most but you lose over a hundred pixels in height. Furthermore, those wide aspect ratio screens aren’t the most productive for Revit, 3ds Max and other Autodesk apps, which appreciates the height to handle the Project Browser, property palettes, the command panel, and other vertical UI elements.

Other niceties on good LCD panels are USB ports and memory card readers, which put things closer to your desk rather than being in your PC, which might be on the floor or otherwise hard to reach.

I’ve seen people using wall mounted 48” LCD televisions with success. You just need to ensure you can run a completely digital signal out to the TV. You’ll probably need to invest in a DVI-D->HDMI adapter if you do not have an HDMI Out port on your machine. Don’t fall back to VGA. Remember that TVs will probably max out at 1280x1080 resolution, which is smaller than you can get on good quality 24” and larger monitors. Because they are physically larger, the pixels will also be larger, so ensure you are far enough away from the TV that the coarseness doesn’t become an issue.

**DisplayPort**

Another thing to look out for are DisplayPort ports. These are similar to HDMI ports but slightly bigger and are offered on more high-end monitors and graphics cards. DisplayPort is the interface meant to replace VGA and DVI, and can be used to carry audio, USB, and other forms of data. It packets the data, much like Ethernet and PCI Express. The current standard is DisplayPort 1.3, released in September of last year, which doubles the bandwidth of DisplayPort 1.2 which is enough to allow for two 4K (3840x2160) monitors at 60Hz in 24-bit RGB mode. Today you will mostly find DisplayPort 1.2 ports on video cards and monitors. DisplayPort itself is a drop-dead simple form factor to use. Just push and click, no knobs or screwdrivers needed.

**Mice and Keyboards**

Along with the monitor the most important thing to choose wisely is your mouse and keyboard. Don’t settle for the cheap plastic things given out by your system vendor. They almost all universally stink while the options out there are truly great. You are going to handle each of these every day for years, so it pays to find ones you like. That said, I’m going to provide my personal recommendations.
For both keyboards and mice, I highly recommend wireless devices. They can be integrated as one solution with some manufacturers. They can either be radio controlled (RF) or implemented via Bluetooth. Logitech and others offer Unifying Receiver technology that allows more than one compatible device to be hooked into a single USB receiver, so you can mix and match their wireless products pretty easily.

**Personal Recommendations:**

Mouse: Logitech MX Master. I graduated from four or five Logitech Performance MX mice, and it’s better in almost every way. For me the killer function is the thumb button, which is perfectly placed and molded into the mouse body, not a separate physical button, so you aren’t tripping over it.

Through the driver, I configure it as a Middle Button. This allows me to pan/rotate around in applications using my thumb instead of moving my index finger from the left button to the wheel. Because of this, I can select an object in Revit, start to drag it by holding down the left button, and while still holding the LMB down, use my thumb to pan the screen at the same time. It’s an incredibly efficient way to drive the software and quite unwieldy to do (for me anyway) with the middle mouse button and left mouse buttons alone.

Keyboard:

Keyboards, like mice, are very personal devices. As a primary instrument to interfacing with your computer, spending time with a good one will make your daily life much easier. Typing for extended periods of time on bad ones can be debilitating. If you play games you need one that is ready to take on the abuse.

In my view a good keyboard offers the following things: excellent key feel, a separate numeric keypad, and multimedia and mini-application controls. Other may appreciate remappable keys which can fire off specific commands. I also highly appreciate backlighting for working in low light. Feel is purely subjective, so it pays to shop at a brick and mortar store and play with the keys for a while. The inclusion of multimedia keys is a no-brainer, where you can control playback and volume control in one place, as well as launch applications with a single button. Many keyboards offer programmable keys which allow you to set up application-specific functions.

Manufacturers have clued into this aspect of the enthusiast market, and built high-end mechanical keyboards for discriminating users. Some of you may remember the old IBM keyboards, which were built like tanks, weighed about as much as one, and had a very heavy feel to the key switches giving off very audible clicks when depressed. This went away with the introduction of cheap keyboards and laptop chicklet keys. Now manufacturers are bringing heavy back with the use of Cherry MX switches.

Cherry Corporation, founded in 1953, is actually the oldest keyboard manufacturer in the world. They produce a line of keyboard switches, the Cherry MX series, which come in several color-coded varieties that vary in actuation force (measured in centi-Newtons, or cN), tactile feel, and audible click noise.

Linear switches are the simplest type, moving straight up and down without any additional tactile feedback or loud clicking noise. In this lineup, the Cherry MX Black switches have a medium to high actuation force (60 cN), making them the stiffest of the lineup. They are most often used in point-of-sale equipment. Cherry MX Red switches have a low actuation force at 45cN, marketed for gamers where rapid actuation is important.

Linear switches are the simplest type, moving straight up and down without any additional tactile feedback or loud clicking noise. In this lineup, the Cherry MX Black switches have a medium to high actuation force (60 cN), making them the stiffest of the lineup. They are most often used in point-of-sale equipment. Cherry MX Red switches have a low actuation force at 45cN, marketed for gamers where rapid actuation is important.

Other types of switches add tactile feedback and varying levels of click noise. Cherry MX Brown switches are a popular switch with a tactile but non-clicky action. It provides excellent feedback without the noise associated with most tactile keyboards. Cherry MX Blue switches are tactile and clicky, favored by heavy typists due to their tactile “bump” and click sound, but do have a higher actuation force (50 cN).
X. Operating Systems – Windows 10

As far as choice of operating system goes, luckily this is a pretty easy decision. Your choices are narrowed down to just three. You could go with either Windows 7 or Windows 8, but in the end you’ll probably end up with Windows 10 on the hard drive - whether you want it or not. See the Links section at the end of the handout for the download link.

Regardless, in my opinion it’s time for most users to forget Windows 7 /8 and move to Windows 10. I appreciate the improved elegance over 8.1. Some of the new features and benefits to the user include:

- Cost is free for almost all Win 7/8 users. Windows 10 is very easy to download and drop dead easy to install. No license key is required – it picks it up from your existing Windows 7/8 installation.
- Although not officially supported, I have personally seen no incompatibilities with anything in the Building Design Suite Ultimate, even when installed with all of the goodies. However, I may just be lucky in that regard. I have gotten service calls from Windows 10 users that were not.
- Consistent interface regardless of device type. Tablets operate just like desktops with a touch UI.
- DirectX 12 should bring a 20-40% increase over DirectX 11 in applications that make heavy use of the API, such as those in the Building Design Suite. Windows 7/8 will never have access to DX 12.
- Solid performance all around with support for more devices and newer technologies such as NVMe SSDs. All of my computers’ devices just seemed to work, from laptops to desktops.
- Virtual Desktops allow single-monitor users to access multiple virtual desktops handy for splitting up usage between work and play, or different projects, or whatever.
- System requirements are the same as Windows 7/8.
- Redesigned Start Menu, a combination of Windows 8 Live Tiles and Windows 7 functionality.
- Enhanced Store provides access to music, videos, games, and apps. Apps from the Store will run on every device – PC, laptop, tablet, or phone.
- Cortana, a personal assistant, much like Siri.
- Unprecedented (and potentially problematic) new policies to enforce updates via Windows Update when you may not want them. However, you will get security patches applied right away which should improve security inside the firewall.
- Microsoft Edge browser - very fast, very lean, boatloads of new features.
- No more weird charms bar. All settings and applications are accessed in the Start Menu.
- All “apps” run as foreground windowed applications.
- Windows Snap – snap windows to one half of your screen. Works great on dual monitor systems. Use Windows Key + arrows to move windows around the screen.

If you already have Windows 10, make sure you get the new massive Windows Update called “Threshold 2” that came out in November 2015. It could be considered Windows 10’s first Service Pack, and included a ton of bug fixes as well as some new features:

- Includes a new “Find my Device” utility. Probably not needed for desktops, but for tablets and laptops, it could be invaluable.
- Bigger and more flexible Start Menu, allowing a 4th column of medium sized tiles
- Side-by-side applications can automatically resize by sliding the shared edge between them.
- Greater installation flexibility – specify locations for new apps, pictures, music, etc.
- Upgraded cloud Edge syncing for passwords and favorites.
XI. Build or Buy

For most people who specify BIM and 3D workstations for their company, the question of buying a packaged workstation from a system vendor like HP or Dell is a no-brainer. For others they may want to build it out themselves. The question comes down to cost per unit performance, availability of specific components, warranties, and servicing levels. This section provides some insight as to the pros and cons of building verses buying complete systems.

**Buy it!**

For many people, particularly those in business environments, purchasing an assembled machine is clearly the way to go. You probably already have a preferred hardware vendor, internal IT support, and a host of corporate IT policies and bylaws, so building an out-of-spec workstation which is not blessed by your IT folks or backed up by a system-wide warranty is considered risky and a non-starter.

**The Pros:**

Buying a workstation from a vendor is largely about minimizing risk and maximizing uptime. The onus is on the vendor to test all of the components, install the OS and the drivers, and make sure it works before handing it off to you. All you should have to do is unpack it, turn it on, and install your software.

Vendors design and build professional-grade workstations as complete systems that can be tested and certified by Autodesk and other software vendors. They specify components which are manufactured under stricter quality control than ones in typical desktops.

By establishing a solid business to business relationship with your vendor, they will work with you to configure systems more to your liking and provide benefits such as free shipping and notifications of sales and special offers.

Warranties, especially next day repair services, are crucial for keeping things humming along and are highly recommended. If you laptop’s video card goes south you can’t just pop down to best Buy to pick another one up. Having a certified mechanic come to your place of business and replace the card on the next day justifies the warranty.

**The Cons:**

The primary issues with packaged workstations are:

- Typically lousy choices in preconfigured models, especially if you know what you want.
- Unreal upcharges for trivial upgrades or changes from preconfigured models.
- Proprietary components.
- Inflexibility in configuration.

System vendors love creating packaged system configurations, building a bunch of them, and shipping them on demand. When you have customers that need 20 workstations that are already sitting in the warehouse, that’s easy to deal with. When you have to build 20 custom machines that have non-standard parts, shipping will be delayed and errors and omissions can be more frequent. But that’s part of what they do on a daily basis.

When a vendor like Dell or HP designs a line of workstations, they configure several models that align with certain price points and requirements. The basis of each new model has a range of available processors, RAM configurations, graphics cards, and hard drive sizes and configurations. The components are easy for them to get and provide a decent range of performance and price points.
The problem is that typically their pre-packaged systems have non-optimal configurations. They default to the lowest-end / last year’s CPUs, incorrect amounts of RAM, terrible video cards, and slow hard disks. This is solely to make the advertised price artificially low. Once you start swapping out components and making upgrades that actually function for your needs, the price quickly skyrockets. You’ll often have ridiculous upcharges to go from 4GB to 16GB or swap out a 1TB hard drive for a 250GB SSD. They may entice you with a low prices of $999 but after you configure it for real-world use it’s like $2,999.

Vendors like Dell and HP custom design their cases, motherboards, and power supplies, all meant to provide adequate reliability for the lowest price to the vendor. Above all, the system must be stable, so you won’t find any knobs to tweak in the system BIOS that you would in an enthusiast motherboard from ASUS, for example. The power supply won’t have the nice fan controls or modular cables or gold-encrusted capacitors and such. The memory modules won’t be painted neon pink with little finger-like heat sinks on them. In short the system components are meant to do a job, not look snazzy doing it.

One pretty popular vendor who caters to the design world is BOXX, who can configure systems to your exact specifications, but you will find these are built from off the shelf parts you could put together yourself for much less.

Finally, the ability for you to swap out items is limited, depending on vendor. Remember that the completed system needs to be certifiable, and Autodesk won’t certify systems with “gaming” graphics cards, even if they run rings around the professional cards. Whether this is a conspiracy between Autodesk and NVIDIA/AMD to artificially keep their workstation lines viable is up for debate. For the base workstation platform that’s not much of an issue - vendors will usually make available all appropriate CPUs for that particular machine. Traditional vendors like Dell nor HP will not swap out the Quadro M2200 for a GeForce GTX 980 Ti because it can’t get certified. Others like BOXX will.

Tips and Tricks for Purchasing Packaged Systems

When purchasing a machine from an established vendor, it helps to understand some tricks of the trade.

1. **Always talk to a real person and establish a professional corporate relationship with the vendor.** In the cases of Dell and HP, you will likely make use of their online stores to do research and initial pricing, and perhaps save a cart of your prospective machine. Look to make them a partner. You will always save money by doing so. The online store is designed to allow people easy ways to purchase machines, but more components and options are available if you have an account and go through your salesperson. Try to establish the highest tier of service that you can. For example, Dell has a “premier” customer status that provides decent service and a good online support portal for tracking your machines, their warranties, and service histories. Use these things to your benefit. Good corporate customers often get freebies such as free shipping or deals on specific components.

2. **Once you narrow down your system, do not pay more for any upgrade than absolutely necessary.** Disregard the online upcharges. They are fictional if you are working through a sales rep. For example, you will typically find Dell and HP have exorbitantly high RAM upgrade prices. Price out RAM from someplace like Crucial.com, which has the exact same module specification for your particular workstation. See if your salesperson can negotiate the RAM upgrade charge. If they cannot match Crucial.com’s price, buy the system with the least amount of RAM, and purchase the RAM upgrade elsewhere. This is one time where installing the RAM yourself can save a lot of money, particularly if it is more than one system. RAM is RAM - Dell doesn’t sprinkle magic dust on it that makes it work better with their systems. They get it from the same place you would, so don’t pay more for it.

3. **Get the physical media for all of the software you are purchasing.** At a minimum this would include the OS, but also Office and the driver disks as well. I once purchased a number of systems for a
customer to find that one of them shipped without anything installed on the hard disk. No OS, nothing. It was completely blank, and having the resource DVDs on hand was a godsend.

4. **When you receive a quote from your vendor, check it over – twice.** It’s easy to for them to key in a slightly wrong thing. The wrong CPU, video card, RAM configuration – you name it, it can get screwed up. Out of a hundred quotes that I have received over the years, easily 40% of them had some sort of tiny mistake, such as extra shipping costs, wrong RAM configuration, and so on.

5. **Evaluate and formally approve the extended warranties that can and will be tacked on to the machine.** However, vendors will often slip an extended warranty (with an extended price tag) into your system quote and hope you will overlook it. Depending on your IT policies and expected lifespan of the machine, this may or may not be something to consider. When disaster strikes it’s nice to have their servicing agent come onsite and replace the part for free. Generally speaking the 3 year onsite warranty that comes with most vendor systems is worth it for the corporate machine. It’s usually inexpensive and pays for itself if the machine blows a fuse. After three years, consider the warranty cost carefully against the expected replacement time for the machine. The exception to this is laptops, where I recommend you carry over warranties after the original one expires.

6. Lastly, **seek out your Autodesk reseller and see if they resell HP or Dell equipment.** They just might and can provide good deals on high-end workstation class machines. Equipment resellers access special sales portals where they can build out high end hardware and peripherals at lower prices. The reseller then tacks on their profit margin and sells it to the end customer. If you purchase your software and hardware in one package, the reseller may be able to provide low to zero cost financing as well. Note that this is limited to professional level workstations, not consumer models. Systems resellers typically cannot provide any discounts on consumer-grade hardware because the pricing is the same as the prices you could get online.

**Build It Yourself!**

On the other hand, you can build your system from scratch. If you have done your research and are looking to ensure your system has specific components, building your new workstation makes sense. For small office / home office machines this is a great way to get what you want without spending a lot of money, depending on the configuration.

**The Pros:**

The more you know about specific components and your specific workstation needs, the more a BIY approach makes sense. As we’ve seen in the graphics benchmarks, the best card for a particular job is possibly not on the model’s available list. If a system vendor doesn’t provide a particular option you like, there is little reason to pay money for what you do not.

If you are an overclocker obsessed with squeezing out every bit of performance potential, a BIY route is definitely the way to go, as you can outfit it with the right motherboard and cooling options required to ramp up the CPU and memory timings. Enthusiast motherboards that have less than conservative timings have solid quality capacitors and cooling solutions to enable higher performance levels. ASUS and Gigabyte hate RMAs as much as Dell and HP do.

If you are looking for alternative builds, such as small form factor (SFF) cases or even building your own custom case - and people do this quite a bit - then obviously you are looking at a BIY approach. Even with an off the shelf case, it gives you the opportunity to physically put it all together and optimize where everything is, so you can route the cables or place the drives in the best location to afford the best airflow.
The Cons:
Building systems comes with its own issues and risks. Researching the parts usually takes longer than picking from a no-name list on a configuration page. You also have to pick the motherboard, RAM, case, and power supply, all of which would come by default in a vendor’s machine and all of which have tons of candidates to choose from.

Second, you can’t be sure the parts you get will actually work when you get them assembled. Defective parts could need to be RMA’d back to the store or manufacturer, delaying when you can start using the system. Components could fail early after the initial build and need to be sent back. Bad batches of components can make this happen more than once - it can seriously be a real pain.

Third, upscale parts like Xeon CPUs and motherboards that support them are not commodity parts and thus are not cheap. Some models are not available as boxed units to consumers, they are only sold to system integrators, so there is a ceiling as to what you can build yourself.

Lastly, vendor workstations come with at least three year system-wide warranties, and repair parts can be overnighted to you. You usually get one-year warranties on most parts (drives may be up to 5 years) but won’t get the same quick turnaround with a bad part from an online retailer, increasing downtime.

Tricks for Buying Build It Yourself Systems
Most online retailers have good RMA programs in case something doesn’t work, so seek out reputable companies with good return policies.

If you are able to get everything from one place, you can save on overall total shipping costs.

Most online stores have “wish lists” and shopping carts, so configure a few of these for research. Check it every so often as prices fluctuate day to day. If you see a nice rebate happening at the moment, strike quickly because they will be gone at some point.

When researching components, pay strict attention to the reviews for each one. Often it is the little things that turn a great product into a problem due to combinations of factors. Look at components with many reviews, as this points to their popularity and should provide a more even evaluation of quality.

They’re a dying breed these days, but check out any local computers shows in your area. I’ve built several personal machines from parts bought from a tiny booth in a local gym for less than I could get at Newegg.com. Because the selection may be lower, expect to make some compromises if expediency is an issue. What I like about the shows is that you can pick up weird smaller parts like removable drive caddies, replacement fans, cables, drive covers, and so on for almost nothing.

Make sure you have enough cables. Hard drives come either as a full kit with a cable, or just the bare drive. Motherboards will typically come with enough cables and brackets for most builds, but you may need extras.

Get a package of small zip-ties to tuck away the power and drive cables to make things neat and improve air flow. Get a small “flush cutter” tool which easily and cleanly cuts small gage wires and plastic zip-ties.

Check the specific processor for the inclusion of the heat sink and fan: Haswell-E and Xeon boxed CPUs typically do not come with a cooler. Air coolers will work within the processors temperature envelope, or you may opt for a different cooling solution altogether like a closed loop water cooling system and radiator to take the temperatures even lower. This is critical for overclockers and people who run their systems at higher levels for extended periods of time.

Lastly, give yourself time to put it together right. A typical system build, from the unboxing of parts to turning on the power button is a few hours. It’s not rocket science but a little bit of planning helps.
Hybrid Builds
It is possible you may get the best of both worlds by buying a bare-bones workstation from your vendor to get the warranty and base build quality, and purchasing the stuff you really want separately, such as video cards and peripherals.

A vendor’s base workstation machine will come with a motherboard, case, CPU, a drive, and at least some memory, and a chintzy graphics card. Get the fastest CPU you can afford with the least RAM, the storage subsystem, and just built-in graphics if available, since those are the parts you want to be choosy about. Then purchase the components you really want elsewhere. You may initially be paying a little extra for extra parts, but keep them on hand as spares in case anything goes south.

Tips for Setting Up New Systems
Whether you buy a system or build it from scratch, there is a standard series of startup tasks I recommend to make sure everything is in working order.

1. Once the machine is initialized and you get to the desktop for the very first time, confirm all components and benchmark the machine in CPU, memory, and graphics using standard free benchmarks such as SiSoft Sandra 2015, 3D Mark11, PCMark8, Cinebench, and others. Anything you do to the machine from here on out - adding software / hardware, tweaking registry settings, running applications, and so on - will change your machine’s performance profile.

2. Once initially benchmarked, immediately check the machine for “craplets” and uninstall them. Depending on the system this could be included backup software, update utilities, printing utilities, or picture management apps which are often worthless. Go to msconfig.exe in Windows 7 or the Task Manager Startup tab in Windows 8 to see what applications are starting up with the system and disable them. If you know your way around Windows Services, optimize the machine further by disabling application specific services that do nothing but consume memory space and CPU cycles. Benchmark the system again and see if removing unnecessary apps helps things out.

3. Install and run the latest version of CCleaner and perform a cleanup of the Registry. Also root out unnecessary startup items and uninstall anything that looks sketchy.

4. Run Windows Update and Application Updates. Every system I’ve ever received was outdated out of the box. Make sure all of your patches are applied before starting any application installations. Do not install any Microsoft-recommended drivers at this point.

5. Update the drivers. This is something a surprising number of people do not do, particularly with new workstations. Immediately check the component manufacturer’s or system vendor’s support & driver site for motherboard-specific driver updates for the BIOS, chipset, audio, USB, and network. Go directly to NVIDIA or AMD for video card drivers. Dell and HP typically update 3rd party board drivers once in a blue moon, whereas NVIDIA and AMD will have new drivers every month. Just updating drivers can have a significant effect on performance, so benchmark again.

6. Finally, install your applications and tweak any post-install settings. For Autodesk apps, reboot the system, launch each one and accept the EULA agreement. Then launch the Autodesk Application Manager to see how many updates are available. With a clean install of the Building Design Suite 2016, expect to spend an hour simply downloading and installing service packs and updates.

Once all of your apps are installed, again check what application specific startup processes or running services are unnecessary. Lastly, right before you really start using it, benchmark it again.

In this section my goal was to build three separate classes of machines for three classes of Building Design Suite. I originally intended to build three classes based on the three user types identified earlier:

1. A Grunt machine, based on the Skylake desktop platform;
2. A BIM Champ machine, based on the Intel HEDT (High End Desktop) Haswell-E platform;
3. A Viz Wiz workstation, based on the Xeon E5 platform.

**Build Philosophy and Methodology**

For each build class, I was going to strive to specify two machines: a Dell Precision Workstation and a BIY (Built It Yourself) system with components priced from Newegg.com. We would then compare prices and declare a winner. I selected the Dell Precision for two reasons: First, they are very popular with the design works and many people are familiar with their systems and have accounts with them. Secondly, they have configurations which are similar for other similar top-tier vendors like HP.

For the BIY systems, I use Newegg.com because they have aggressive pricing and can be a single source for all parts. This should save on shipping costs and general pain, because if anything goes wrong there is one place to go. Newegg.com has a pretty good reputation for low prices and low hassles if something needs to be RMA’d.

However, it becomes apparent that while it’s easy to find desktop and HEDT based parts and build out those two workstations, it’s much more difficult to build out a Xeon E5-based workstation that would be considered cost effective from parts. For that, I simply priced out a single hybrid build with a Dell Precision Workstation along with upgrade components from Newegg.

As mentioned earlier, specifying a workstation from a vendor is full of compromises, which I hate. They usually don’t offer the right graphics card, so you either buy a lower-end one or a more expensive “workstation” class card that performs the same or worse than the one you wanted.

Due to these limitations, I decided to build the Dell workstations as hybrid builds, purchasing as much from Dell as I could and supplementing it with the right parts from Newegg.com. In other words, I did not let my specification become compromised because Dell did not allow me to purchase a particular part.

**Base Specifications and Terms:**

- I’m pricing everything anonymously from online sites. For Dell, if I requested actual quotes from human beings, even salespeople, the price would probably be lower.
- I’m including dual 27” Dell’s U2715H IPS flat panel monitors in all builds.
- No MS Office, productivity software, security software, or specific vendor tweaks are included.
- For the Dell systems, all DVD resource disks for the OS and drivers are included in the price.
- Shipping prices are not included on any builds.
- Newegg prices include any current as-of-this-writing rebates and special offers.
- For the video card, I’m specifying as a standard the GeForce GTX 980 Ti.
- Unless forced to do so, I did not specify optical storage. Who uses DVDs anymore?
**The Grunt: Dell Precision T3620**
First up is a Dell Precision T3620 Workstation in a mini tower.

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Xeon E3-1270 v5 Quad-Core @ 3.6GHz</td>
<td>$2,249.26</td>
</tr>
<tr>
<td>Memory</td>
<td>32GB (4x8GB) DDR4 2133 Non-ECC</td>
<td></td>
</tr>
<tr>
<td>Graphics</td>
<td>NVIDIA Quadro NVS 315, 1GB</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>512GB 2.5&quot; SATA High Performance Solid State Drive (model unknown)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1TB 3.5&quot; SATA 7200 RPM Hard Drive (model unknown)</td>
<td></td>
</tr>
<tr>
<td>Power supply</td>
<td>365W Up to 92% Efficient Power Supply</td>
<td></td>
</tr>
<tr>
<td>Keyboard</td>
<td>Dell KB216 Wired Keyboard Black</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Windows 10 Professional, 64-bit</td>
<td></td>
</tr>
<tr>
<td>Warranty</td>
<td>3 Year Hardware Service with Onsite/In-Home Service after Remote Diagnosis</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>None (see below)</td>
<td></td>
</tr>
<tr>
<td>Resource Disk</td>
<td>Windows 10 OS Recovery and Resource DVDs</td>
<td></td>
</tr>
<tr>
<td>Monitor</td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$549.99</td>
</tr>
<tr>
<td></td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$549.99</td>
</tr>
<tr>
<td>Dell Subtotal</td>
<td></td>
<td>$3,349.24</td>
</tr>
</tbody>
</table>

**Additional items purchased separately from Newegg.com**

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Card</td>
<td>EVGA GeForce GTX 980 Ti w/6GB</td>
<td>$649.99</td>
</tr>
<tr>
<td>Mouse</td>
<td>Logitech MX Master Mouse</td>
<td>$99.99</td>
</tr>
<tr>
<td>Newegg Subtotal</td>
<td></td>
<td>$749.98</td>
</tr>
<tr>
<td><strong>System Total</strong></td>
<td></td>
<td><strong>$4,099.22</strong></td>
</tr>
</tbody>
</table>

**Notes**

1. Dell provides this model with either a Xeon E3-1270 v5 or a Core i7-6700 processor for the same price. However, the Xeon E3-1270 v5 is base clocked at 3.6GHz whereas the i7-6700 is only 3.4GHz. Note: The i7-6700 is not the “K” model and cannot be overclocked.
2. The Xeon does not have an IGP so a default video card must be purchased.
3. The 512GB SATA drive was selected, as the 512GB m.2 PCIe NVMe drive was $80 more.
4. The U2715H monitor is a great price at $549.99, but only if you purchase the monitor as part of a system, as it includes a $150 rebate. Dell’s retail price for the screen is $699.99.
5. There are no advanced USB 3.1 ports on the system.
The Grunt: Newegg Edition

Next up is a comparable home-brewed BIY system completely from Newegg:

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Fractal Design Define R4 Black Pearl Mid-Tower Computer Case</td>
<td>$99.99</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Core i7-6700K Skylake 4.0GHz</td>
<td>$399.99</td>
</tr>
<tr>
<td>Motherboard</td>
<td>ASUS Z170-PRO LGA 1151 motherboard</td>
<td>$184.99</td>
</tr>
<tr>
<td>Memory</td>
<td>Crucial 32GB (4 x 8GB) 288-Pin DDR4 SDRAM DDR4 2133 (PC4 17000) Desktop Memory Model CT4K8G4DFD8213</td>
<td>$169.99</td>
</tr>
<tr>
<td>Graphics</td>
<td>EVGA GeForce GTX 980 Ti w/6GB</td>
<td>$649.99</td>
</tr>
<tr>
<td>Storage</td>
<td>Samsung 850 Pro 512GB 2.5&quot; Solid State Drive</td>
<td>$219.99</td>
</tr>
<tr>
<td></td>
<td>Western Digital Black 1TB 3.5&quot; 7200 RPM 6GB/s Hard Drive</td>
<td>$69.99</td>
</tr>
<tr>
<td>Power supply</td>
<td>Corsair AX Series AX 860 860W 80+ Platinum PSU</td>
<td>$164.99</td>
</tr>
<tr>
<td>Mouse</td>
<td>Logitech MX Master Mouse</td>
<td>$99.99</td>
</tr>
<tr>
<td>Keyboard</td>
<td>Corsair Vengeance K70 RGB Keyboard Cherry MX Red switches</td>
<td>$161.89</td>
</tr>
<tr>
<td>OS</td>
<td>Windows 10 Professional, 64-bit, OEM</td>
<td>$139.99</td>
</tr>
<tr>
<td>Monitor</td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td></td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td>System Total</td>
<td></td>
<td>$3,556.81</td>
</tr>
</tbody>
</table>

**Notes**

1. Because motherboards for the Xeon E3 are difficult to get on the open market, this system was specified with a common Skylake i7-6700K processor @ 4GHz with a Z170 motherboard.
2. The ASUS motherboard comes with two USB 3.1 ports for newer USB high speed devices.
3. The Samsung 850 Pro SSD is regarded as top of its class in benchmarks.
4. The Corsair mechanical keyboard is definitely high end and could be considered a luxury.
5. Windows 10 Pro was specified, but you could save money by purchasing Windows 8.1 and upgrading for free. However, you would need to install Windows 8.1 then upgrade in place, which is time wasted and would negate any savings.

**The Grunt BIM Workstation: Build Analysis**

Don’t use the term “entry level” for this workstation; it’s a professional-class build meant for people using everything in the Building Design Suite, and will handle the vast majority of the tasks with aplomb.

Overall the Newegg.com system was about $540 less than the Dell for the same basic configuration. However, the Newegg machine has a faster CPU (4GHz vs 3.6GHz), a better keyboard, and a known performer for the SSD. The Fractal Design case is serviceable but not as high quality as the Dell Precision mid-tower. Assume the Dell system to be nearly silent but the Newegg system to emit some fan noise.

The Dell was more expensive primarily due to the 32GB of RAM, which was a staggering $459 more than specifying it with a single 4GB stick, and $289 more than the entire 32GB kit from Newegg.

As I wasn’t willing to accept the “workstation” class video cards Dell offered, I specified a cheap Quadro NVS card and picked up the GTX 980 Ti from Newegg. I could have saved this cost by specifying the i7-6700 CPU, which has an IGP. However, it is clocked much slower than the i7-6700K by 600MHz – ouch.

Downsides are what you would expect of BIY systems: No on-site service warranties, higher propensity to RMA back a defective component, and the time it takes to build it and install the OS.
The BIM Champ 6-Core Xeon Workstation: Dell Precision 5810
Stepping up to 6 CPU cores, we have the following Dell Precision 5810 based on the Xeon E5-1650 v3 using the Haswell microarchitecture. Internally, the Xeon is very closely aligned with the Haswell-E i7-5930K. Both have the same core speed and feature set.

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Xeon E5-1650 v3 6-Core @ 3.5GHz</td>
<td>$3,643.50</td>
</tr>
<tr>
<td>Memory</td>
<td>32GB (4x8GB) DDR4 2133 ECC</td>
<td></td>
</tr>
<tr>
<td>Graphics</td>
<td>NVIDIA Quadro M4000 8GB card</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>512GB 2.5” SATA High Performance Solid State Drive (model unknown)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8x Slimline DVD-ROM</td>
<td></td>
</tr>
<tr>
<td>Keyboard</td>
<td>Dell KB212-B QuietKey USB Keyboard Black</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Windows 8.1 Professional, 64-bit, w/DVD Recovery</td>
<td></td>
</tr>
<tr>
<td>Warranty</td>
<td>3 Year Hardware Service with Onsite/In-Home Service after Remote Diagnosis</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>None (see below)</td>
<td></td>
</tr>
<tr>
<td>Resource Disk</td>
<td>Windows 8.1 OS Recovery and Resource DVDs</td>
<td></td>
</tr>
<tr>
<td>Monitor</td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$549.99</td>
</tr>
<tr>
<td></td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$549.99</td>
</tr>
<tr>
<td>Dell Subtotal</td>
<td></td>
<td>$4,743.48</td>
</tr>
<tr>
<td>Additional items purchased separately from Newegg.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>Logitech MX Master Mouse</td>
<td>$99.99</td>
</tr>
<tr>
<td>Newegg Subtotal</td>
<td></td>
<td>$99.99</td>
</tr>
<tr>
<td>System Total</td>
<td></td>
<td>$4,843.47</td>
</tr>
</tbody>
</table>

Notes
1. Dell’s online configurator does not allow me to not specify a DVD-ROM, so there is a 8x unit included.
2. Dell’s online configurator does not allow a 2.5” SSD boot drive and a 3.5” mechanical hard drive, so this system is being specified without the additional 1TB of storage as the Grunt machine has.
3. The 5810 does not offer Windows 10 at this time, so Windows 8.1 will be specified (you can assume you can upgrade to Windows 10 later).
4. As with the Grunt Precision build, you must specific a video card. However, the lowest you can go is a 4GB AMD graphics card. Because we want to stay with NVIDIA, we will upgrade to a Quadro M4000 8GB card.
5. Although the Xeon platform supports up to 768GB of RAM, Dell limits the 5810 to only 32GB.
The BIM Champ 6-Core Haswell-E: Newegg

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Fractal Design Define R4 Black Pearl Mid-Tower Computer Case</td>
<td>$99.99</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Core i7-5930K Haswell-E 6-Core 3.5 GHz</td>
<td>$459.99</td>
</tr>
<tr>
<td>Processor Cooler</td>
<td>Corsair Hydro H100i GTX Extreme Performance Water/Liquid CPU Cooler</td>
<td>$119.99</td>
</tr>
<tr>
<td>Motherboard</td>
<td>MSI X99A Raider LGA 2011-v3 Motherboard</td>
<td>$189.99</td>
</tr>
<tr>
<td>Memory</td>
<td>Crucial 32GB (4 x 8GB) 288-Pin DDR4 SDRAM DDR4 2133 (PC4 17000) Desktop Memory Model CT4K8G4DFD8213</td>
<td>$169.99</td>
</tr>
<tr>
<td>Graphics</td>
<td>EVGA GeForce GTX 980 Ti w/6GB</td>
<td>$649.99</td>
</tr>
<tr>
<td>Storage</td>
<td>Samsung 850 Pro 512GB 2.5&quot; Solid State Drive</td>
<td>$219.99</td>
</tr>
<tr>
<td></td>
<td>Samsung 18X DVD-ROM 48X Model SH-118CB/BEBE</td>
<td>$19.99</td>
</tr>
<tr>
<td>Power supply</td>
<td>Corsair AX Series AX 860 860W 80+ Platinum PSU</td>
<td>$164.99</td>
</tr>
<tr>
<td>Mouse</td>
<td>Logitech MX Master Mouse</td>
<td>$99.99</td>
</tr>
<tr>
<td>Keyboard</td>
<td>Corsair Vengeance K70 RGB Keyboard Cherry MX Red switches</td>
<td>$161.89</td>
</tr>
<tr>
<td>OS</td>
<td>Windows 10 Professional, 64-bit, OEM</td>
<td>$139.99</td>
</tr>
<tr>
<td>Monitor</td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td></td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td>System Total</td>
<td></td>
<td>$3,691.80</td>
</tr>
</tbody>
</table>

Notes
1. Because the Haswell-E CPU does not come with a cooler, a Corsair water cooling system was added.
2. The GTX 980 Ti was specified to maximize value. The Quadro M4000 retails for $889.00 and is built on the same GM204 as the GTX 980 and Quadro M5000. It has only 1664 CUDA cores active but has 8GB of video RAM.
3. The 1TB hard disk was replaced by an 18x DVD-ROM to maintain parity with the Dell 5810.

The BIM Champ Build Analysis
For this system I decided to step things up to the next logical level; a 6-core Haswell-based Xeon E5 CPU in the Dell Precision 5810 and a Haswell-E i7-5930K in the Newegg build. Comparing the CPUs, the Xeon E5-1650 v5 has a Passmark score of benchmarks at 13,470. The Core i7-5930K has a score of 13,666. Given the disparities that can exist with this benchmark, you can safely call the two equal in performance.

However, as usual, the issues with specifying Dell workstation components brings on the headaches. Because the list of video cards is limited to professional level cards, I had to specify a Quadro M4000 for the Precision. Otherwise we would have paid about $275 more total to drop down to an AMD Fire Pro card and buy the GTX 980 Ti separately. However, that could provide a way to run both at the same time, and designate the AMD to Windows tasks while the GTX 980 chews on GPU-accelerated rendering tasks.

The M4000 is based on the GM204 GPU, the same one on the GTX 980 / 970, but not the GTX 980 Ti, which is built on the later GM200 GPU. The M4000 has 1664 active CUDA cores which puts it equal with the GTX 970, which is a very entry-level GPU for the Building Design Suite. Clearly the Precision is going to be hamstrung in the graphics department, but to take it higher would make the price disparity much higher.

As with the Grunt build, the DIY system wins in the price / performance category. It is over $1,100 less than the Dell Precision but has way more graphics power, better cooling, a better DVD-ROM, a better OS, and a better keyboard.
The Viz Wiz 8-Core Haswell-E Newegg Build

The BIM Champ 6-Core Haswell-E: Newegg

<table>
<thead>
<tr>
<th>Component</th>
<th>Item</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Corsair Obsidian 750D Black Aluminum / Steel ATX Full Tower Computer Case</td>
<td>$124.99</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Core i7-5960X Haswell-E 8-Core 3.0 GHz</td>
<td>$1,049.99</td>
</tr>
<tr>
<td>Processor Cooler</td>
<td>Corsair Hydro H100i GTX Extreme Performance Water/Liquid CPU Cooler</td>
<td>$119.99</td>
</tr>
<tr>
<td>Motherboard</td>
<td>ASRock X99 OC Formula/3.1 Extended ATX Intel Motherboard</td>
<td>$319.99</td>
</tr>
<tr>
<td>Memory</td>
<td>CORSAIR Dominator Platinum 64GB (8 x 8GB) 288-Pin DDR4 SDRAM DDR4 2400 (PC4 19200) C14 Memory Kit Model CMD64GX4M8A2400C14</td>
<td>$609.99</td>
</tr>
<tr>
<td>Graphics</td>
<td>EVGA GeForce GTX 980 Ti w/6GB</td>
<td>$649.99</td>
</tr>
<tr>
<td>Storage</td>
<td>Samsung SM951 M.2 512GB PCIe 3.0 SSD MZHVP512HDGL-00000 - OEM</td>
<td>$338.50</td>
</tr>
<tr>
<td></td>
<td>Western Digital Black 1TB 3.5” 7200 RPM 6GB/s Hard Drive</td>
<td>$69.99</td>
</tr>
<tr>
<td>Power supply</td>
<td>CORSAIR AXi series AX1200i 1200W ATX12V 80 PLUS PLATINUM Certified Full Modular Active PFC Power Supply</td>
<td>$299.99</td>
</tr>
<tr>
<td>Mouse</td>
<td>Logitech MX Master Mouse</td>
<td>$99.99</td>
</tr>
<tr>
<td>Keyboard</td>
<td>Corsair Vengeance K70 RGB Keyboard Cherry MX Red switches</td>
<td>$161.99</td>
</tr>
<tr>
<td>OS</td>
<td>Windows 10 Professional, 64-bit, OEM</td>
<td>$139.99</td>
</tr>
<tr>
<td>Monitor</td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td></td>
<td>Dell UltraSharp U2715H 27-inch Widescreen Flat Panel</td>
<td>$597.51</td>
</tr>
<tr>
<td>System Total</td>
<td></td>
<td>$6,480.29</td>
</tr>
</tbody>
</table>

Notes
1. Because we can use install up to 5 graphics cards in this system, I specified the Corsair 750D full ATX tower case. It has plenty of room for an Extended ATX motherboard and three video cards.
2. I chose the ASRock motherboard for its 5 PCIe x16 slot expandability for hosting multiple graphics cards as well as the inclusion of USB 3.1 ports.
3. The Samsung SM951 is an m.2 format SSD that has better performance than the 850 Pro.

The Viz Wiz Build Analysis
This system has two requirements: Burn through any multithreaded workload across the board in all Building Design Suite applications, and render iray scenes as fast as possible. To that end I went with the Haswell-E 8-core i7-5960X running at 3.0 GHz. It should be overclockable to between 4 - 4.5 GHz.

In looking at what Dell offered in the high end, their Precision 7810/7910 tower models had too many limitations to be competitive. They come with an 8-core or 10-core Xeon E5-26xx series CPU in a tight package. While these CPUs will burn through a mental ray rendering (especially with two in a system), these CPUs run well under 3.0GHz and would not be competitive in everyday application usage. And with Dell’s conservative BIOS, you likely won’t be able to overclock them. Because mental ray / traditional CPU renderers are being outpaced by newer GPU-accelerated renderers (not to mention cloud rendering) the requirement for increasing CPU cores is waning in favor of installing more GPUs dedicated to the task.

Oddly enough, even though these are dual-CPU capable, Dell only configures the system with a single processor and the online configurator does not allow you to add the second CPU. When specifying a dual-CPU machine you need to purchase both at the same time, to ensure both CPUs are of the same stepping. Otherwise if you add a second CPU later, you could have stability issues.
Dell was too limiting in other areas as well. It also does not offer Windows 10 - probably because many applications are not certified for it yet - and weirdly limits your graphics card choices to last year’s Kepler Quadro lineup, which are trounced by today’s Maxwell based GPUs. This is obviously a case where driver certification is falling behind schedule.

For some weird reason Dell also limits how you can allocate storage, in that you cannot configure a system online with a 2.5” SSD and a 3.5” mechanical drive. Finally, Dell is charging over $870 for the 64GB upgrade over 16GB - $200 more than the entire higher-speed 64GB Corsair kit from Newegg.

While some of this may be alleviated by talking with your Dell sales representative, in the end I do not see the Dell being worth the effort. If you are going to spend over $6,000 for a system, you should be able to get exactly what you need.

To that end I went with a complete BIY system based on a water-cooled Haswell-E i7-5960X. Its 8 cores will burn through multithreaded Revit tasks such as scene regeneration and Autodesk Raytracer rendering. While the base clock is only 3.0 GHz, this CPU can be overclocked to about 4.5GHz pretty easily.

The Corsair Dominator 64GB memory kit is expensive but has a stellar reputation for stability and runs cool with the added heat sinks. Speed was upgraded from DDR-2133 to DDR4-2400, which will give a small uptick in performance. After that, improvements in performance drop to unnoticeable levels in application use.

To handle graphics and rendering, I went with three EVGA GTX 980 Ti cards. This maximizes the total number of CUDA cores (2,816 x 3 = 8,448) in as few cards as possible without costing thousands more. For heavy iray use, adding a fourth GTX 970 card can be assigned to handle Windows graphics tasks in 3ds Max and let the GTX 980 Tis handle the GPU rendering part. In particular, the EVGA sticks out from the crowd in that over 80% of the reviews were 5 star, and 11% were 4 star.

It’s a more expensive, but the Samsung SM951 m.2 PCIe NVMe SSD removes the bottlenecks of SATA and AHCI to provide clearly better performance than the Samsung 850 Pro. It is a great fit for the m.2 PCIe x4 slot on the motherboard. The 1TB WD hard drive is included to provide more near-line storage.

For this many GPUs, you need a motherboard than can handle it. The choice of the ASRock X99 Extended ATX Intel Motherboard was based on its reputation for easy overclocking and its 5 PCIe x16 slots allowing a fully 5-way x8/x8/x8/x8/x8 configuration. It also includes 3 forward-compatible USB 3.1 ports (two Type-A and one Type-C connectors), dual Gigabit network ports, as well as two m.2 slots for SSDs.

To handle the sheer mass of all components and be able to keep things cool, the Corsair Obsidian 750D case was specified. Although a little odd, it has the size and flexibility to house the Extended ATX motherboard and all of these components without issue, as well as provide the necessary overhead room for the Corsair H100i water cooler and radiator.
XIII. Links

**Industry Pressures**
Setting up an Amazon EC3 render farm with Backburner

**Revit Specific**
Revit Model Performance Technical Note 2015:

**Processors:**
Intel’s Comparative CPU Database
http://ark.intel.com/

Intel Tick-Tock Model

Intel’s Skylake CPU Reviews
http://www.anandtech.com/show/9483/intel-skylake-review-6700k-6600k-ddr4-ddr3-ipc-6th-generation
http://arstechnica.com/gadgets/2015/08/intel-skylake-core-i7-6700k-reviewed/
http://www.maximumpc.com/intel-skylake-review/

Intel Skylake (6th Generation) and Haswell (4th Generation) CPUs Compared
http://ark.intel.com/compare/88195,80807,77656,75123

Haswell-Based Xeons E3-12xx v5 Series Compared
http://ark.intel.com/compare/88182,88176,88174,88171

Haswell-Based Xeons E5-16xx v5 Series Compared
http://ark.intel.com/compare/82763,82764,82766,82767,82765

Transactional Synchronization (TSX Instructions) in Haswell:

Intel Xeon E5-2600 v3 Review
http://www.tomshardware.com/reviews/intel-xeon-e5-2600-v3-haswell-ep,3932.html

**Graphics:**
Autodesk Certified Graphics Hardware
http://usa.autodesk.com/adsk/servlet/systcert?id=18844534&siteID=123112

GPU Database
http://www.techpowerup.com/gpudb/

NVIDIA GTX 980 (Maxwell 2) white paper:
http://international.download.nvidia.com/geforce-com/international/pdfs/GeForce_GTX_980_Whitepaper_FINAL.PDF

NVIDIA GTX 980 Ti Review
http://www.anandtech.com/show/9306/the-nvidia-geforce-gtx-980-ti-review
BOXX Blogs: GeForce GTX Rendering Benchmarks and Comparisons

BOXX Blogs: Do you need NVIDIA Maximus for GPU Rendering?
http://boxxblogs.blogspot.com/2013/05/NVIDIA-maximus-and-gpu-rendering.html


Arion Benchmark: http://www.randomcontrol.com/arionbench

iray Performance Tips (from 2011)

GTC 2015 on demand - GPU Ray Tracing presentation

NVIDIA mobile GPU comparison sheet:

Memory
Quad-channel RAM vs. Dual-channel RAM

DDR4 2133 to 3200 Memory Scaling

Peripherals:
Cherry MX Switches:

Windows 10:
Download link for in-place upgrade and media creation tool